We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Token Passing Techniques for Hard Real-Time Communication

Gianluca Franchino*, Giorgio C. Buttazzo* and Tullio Facchinetti**.

*Scuola Superiore Sant’Anna, Italy
**University of Pavia, Italy

1. Introduction

Distributed computing platforms are increasingly used to develop critical embedded systems, like control applications, sensors networks, telecommunication, and robotics systems. In such distributed applications, the correct behaviour depends on the timely execution of the tasks running on different nodes, which may frequently exchange shared data. In particular, since the nodes are typically connected through a common channel without the need of multi-hop communication, it turns out that a timely communication mainly depends on how the nodes access the channel. Even when a multi-hop communication is needed, a timely message delivery is not feasible without the support of a predictable channel access mechanism, which is implemented in the Medium Access Control (MAC) sub-layer of the communication stack.

There exist several MAC protocols designed for providing a timely communication among distributed nodes, mainly in the factory communication domain. One of the most effective solutions is given by token passing protocols, which have some nice characteristics that make them suitable for real applications. For instance, because of the token passing mechanism, the nodes do not need to be synchronized and they have an implicit bandwidth reclaiming mechanism that allows other nodes to exploit the unused bandwidth. Moreover, such protocols can serve both real-time and best-effort (non real-time) traffic.

Among token passing protocols, the timed token policy is a channel scheduling approach first proposed by Grow in (Grow, 1982). Since then, it has received a substantial attention and several relevant results have been derived (Zhang et al., 2004), which make timed token protocols suitable for the real-time communication in industrial applications. Timed token policies are used as channel access mechanism in several standard protocols such as, for instance, PROFIBUS (Profibus, 1996) and FDDI (FDDI, 1987). However, the application domain of the timed token policy is not restricted to the cited communication standards, and some examples on their use can be found in (Lenzini et al., 2004) and (Cicconelli et al., 2007).

To improve the ability of timed token protocols of managing real-time traffic, Shin and Zheng (Shin & Zheng, 1995) proposed a modification of the timed token protocol, which can guarantee a greater bandwidth for the real-time traffic with respect to the classic timed token protocol; however, under certain conditions, it cannot manage the best-effort traffic (Franchino et al., 2008). The Budget Sharing Token (BuST) protocol has been proposed as an improvement with respect to existing timed token protocols (Franchino et al., 2008). In
particular, BuST improves the bandwidth guaranteed for the real-time traffic with respect to the timed token protocol, and it can manage best-effort traffic in those situations where the modified timed token protocol fails on serving this kind of traffic.

This chapter introduces a few token passing protocols, presenting their characteristics and illustrating the main results available in the literature. The BuST protocol is discussed in detail, illustrating its main advantages by means of analytic and simulation comparisons. New and well known properties of the protocols for managing both hard real-time and best-effort traffic will be analyzed, describing drawbacks and strengths of each protocol. The analysis is carried out considering the main budget allocation schemes available in the literature, thus contributing to the comparative characterization of the several schemes nowadays available.

2. Network Model

The communication network is composed by \( n \) nodes sharing a common medium, e.g. a bus, where each node can transmit both real-time and best-effort traffic. The former kind of traffic is modelled by assigning each node \( i \) a synchronous message stream \( S_i \), which is described by three parameters \((C_i, D_i, T_i)\), where:

- \( C_i \) is maximum amount of time necessary to transmit a message generated in the stream \( S_i \). This includes the time to transmit both the message payload and the message headers/footers;
- \( D_i \) is the relative deadline of a message generated by stream \( S_i \) that is, the maximum amount of time that a message can wait in transmission queue before its transmission is completed. Hence, the transmission of the \( j \)-th message, which is queued at time \( t_{i,j} \), must be completed no later than its absolute deadline \( d_i = t_{i,j} + D_i \);
- \( T_i \) is the generation period of messages in stream \( S_i \). If the \( j \)-th message in the stream \( S_i \) is put in the transmission queue at time \( t_{i,j} \), then the \((j+1)\)-th will arrive in the transmission queue at time \( t_{i,j+1} = t_{i,j} + T_i \).

Without loss of generality, only a stream per node it is considered, since the case of a network with more streams per node can be represented with a logical equivalent one with a stream per logical node, as showed in (Agrawal et al., 1994). We consider that each node \( i \) is assigned a bandwidth \( H_i \) also called time budget, used to bound the transmission of the node.

The channel utilization of each message in stream \( S_i \) is:

\[
U^S_i = \frac{C_i}{\min(T_i, D_i)}
\]  \hspace{1cm} (1)

The total channel utilization of a periodic stream set is then:

\[
U^S = \sum_{i=1}^{n} U^S_i
\]  \hspace{1cm} (2)

which measures the channel bandwidth utilized by the real-time (periodic) traffic.

Before describing the protocols, the following definitions are needed:
**Definition 1.** \( \tau \) is the time needed to transmit the token between nodes, including the overhead introduced by the protocol.

**Definition 2.** The Target Token Rotation Time (TTRT) represents the expected time needed by the token to complete an entire round-trip of the network.

Any assignment of the time budgets \( H_i \) must satisfy the following two constraints:

**Definition 3.** (Protocol Constraint) The total bandwidth allocated to the nodes, during one complete token rotation, must be less than the available network bandwidth, that is,

\[
\sum_{i=1}^{n} H_i \leq 1 - \frac{\tau}{TTRT} \tag{3}
\]

The Protocol Constraint is necessary to ensure a stable operation of the protocols.

**Definition 4.** (Deadline Constraint) If \( s_{i,j} \) denotes the time at which the transmission of the \( j \)-th message in stream \( S_i \) is completed, the deadline constraint requires that for \( i=1, \ldots, n \), and \( j=1, 2, \ldots, \)

\[
s_{i,j} \leq t_{i,j} + D_i \tag{4}
\]

where \( t_{i,j} \) is the message arrival time and \( D_i \) is its relative deadline.

Meeting the Deadline Constraint ensures that every periodic message is transmitted before its absolute deadline. Note that in Inequality (4), while \( t_{i,j} \) and \( D_i \) are defined by the application, \( s_{i,j} \) depends on the bandwidth (budget) allocation and on the TTRT value.

**Definition 5.** A stream set \( \Gamma = \{S_1, S_2, \ldots, S_n\} \) is said to be feasible or schedulable when both the Deadline Constraint and the Protocol Constraint are met.

To test the Protocol Constraint it is sufficient to check whether the sum of the budgets are not greater than the TTRT minus the overhead \( \tau \). However, testing the Deadline Constraint may be much more complicated. The following method is often used for testing whether the Deadline Constraint is met:

Let \( X_i \) be the minimum amount of time available for node \( i \) to transmit its \( j \)-th message during the time interval \( [t_{i,j}, t_{i,j} + D_i] \), then for a message stream set with message deadlines not greater than periods, the Deadline Constraint can be satisfied if and only if for all \( i, i=1, 2, \ldots, n, X_i \geq C_i. \)

### 3. Timed Token Protocols

The timed token protocol (TTP) is a basic channel access technique, namely a MAC protocol, which can be used to manage real-time traffic while guaranteeing a fair sharing of the unused bandwidth to the best-effort traffic. In timed token approaches, a token travels between nodes in a circular fashion and each node can transmit only when it possesses the token. An important parameter is the Target Token Rotation Time (TTRT), which represents the expected time needed by the token to complete an entire round-trip of the network. Each node \( i \) has an associated time budget \( H_i \); whenever a node receives the token, it can transmit...
its real-time messages for a time no greater than $H_i$. It can then transmit its best-effort messages if the time elapsed since the previous token arrival to the same node is less than the value of $TTRT$, that is, only if the token arrives earlier than expected. Figure 1, shows a typical timed token based network where 4 nodes sharing a common channel are arranged in a logical ring, as far as the token passing mechanism is concerned.

Fig. 1. Timed token network

3.1 TTP operation rules
To better understand the TTP operation, the protocol channel access rules are detailed below:

- During the network start up, each node $i$ declares a $TTRT$ value equal to one half of the deadline $D_i$ related to its message stream $S_i$. The minimum declared value is chosen as $TTRT$, and each node $i$ is then assigned a time budget $H_i$ that depends on $TTRT$.
- Each node uses two timers, the token holding timer ($THT$) and the token-rotation-timer ($TRT$). The $TRT$ counter always increases, whereas the $THT$ only increases when the node is delivering best-effort traffic. When $TRT$ reaches $TTRT$, it is reset to 0 and the token is signed as "late" by incrementing the node's late counter $L_c$ by one. To initialize the timers and $L_c$, no messages are sent during the first token rotation after the ring initialization.
- Only the node holding the token can transmit messages. Transmission is controlled by the timers, but an in-progress transmission of a single packet is not interrupted until its completion. When node $i$ gets the token, it performs the following operations:
  o If $L_c > 0$, it sets $L_c = L_c - 1$ and $THT = TTRT$. Otherwise, $THT = TRT$ and $TRT = 0$.
  o If node $i$ has synchronous packets, it transmits them for a time no greater than $H_i$.
- If node $i$ has best-effort packets, it transmits them until $THT$ counts up to $TTRT$, or until all the asynchronous traffic is sent, which ever comes first.
• Node $i$ passes the token to station $(i + 1) \mod (n)$.

The main drawback of TTP is the inability of guaranteeing the total available bandwidth for the real-time traffic. As Johnson and Sevciks (Johnson & Sevciks, 1987) showed, the average interval between two consecutive visits of the token at the same node, namely the average token rotation time, does not exceed $TTRT$ and the maximum rotation time does not exceed $2TTRT$. Thus, if $D$ is the minimum deadline among stream deadlines, i.e. $D = \min(D)$, it turns out that $TTRT = D/2$. From the Protocol Constraint it follows that:

$$\sum_{i=1}^{n} H_i \leq TTRT - \tau = \frac{D}{2} - \tau$$  \hspace{1cm} (5)

In the time interval defined by $D$, the bandwidth available for the real-time traffic can be obtained dividing the above equation by $D$:

$$\frac{\sum_{i=1}^{n} H_i}{D} \leq \frac{1}{2} \cdot \frac{\tau}{D}$$  \hspace{1cm} (6)

Since the total available bandwidth is $1 - \tau/TTRT$, the TTP can guarantee at most half of the total available bandwidth for the real-time traffic.

3.2 Modified TTP operation rules

To improve the bandwidth guaranteed for the real-time traffic, Shin (Shin & Zheng, 1995) proposed to modify the timed token rules, limiting the maximum time available for best-effort traffic to $T_{B}^{MAX}$, where:

$$T_{B}^{MAX} = TTRT - \sum_{i=1}^{n} H_i - \tau$$  \hspace{1cm} (7)

Notice that, under TTP, the maximum bandwidth allocated for best-effort messages is $T_{B}^{MAX} = TTRT - \tau$.

In (Chan et al., 1997), the authors showed that the maximum token rotation time for the Modified Timed Token Protocol (MTTP) is bounded by $TTRT$. This means that, under MTTP it is possible to select a value for $TTRT$ no greater than the minimum deadline $D$. Hence, being $TTRT \leq D$, from the Protocol Constraint it follows that:

$$\sum_{i=1}^{n} H_i \leq D - \tau$$  \hspace{1cm} (8)

and dividing by $D$, it gives:

$$\frac{\sum_{i=1}^{n} H_i}{D} \leq 1 - \frac{\tau}{D}$$  \hspace{1cm} (9)
That is, MTTP can guarantee all the available bandwidth for the real-time traffic. By defining $T_S = \sum H_i$, to guarantee that $T_B^{MAX} = T_{TTR} - T_S - \tau$, MTTP can be defined as follows:

1. A new token rotation time is defined as $T_{TTR}^n = T_{TTR} - T_S$ and used instead of $T_{TTR}$;
2. The counting of a node’s token rotation timer (TRT) is stopped when a real-time message is being delivered by the node.

The full details of the protocol rules are given below:

- During the network start up, each node $i$ declares a $T_{TTR}$ value equal to the deadline $D_i$ related to its message stream $S_i$. The minimum declared value is chosen as $T_{TTR}$. Each node $i$ is assigned a time budget $H_i$ that depends on $T_{TTR}$, and it sets $T_{TTR}^n = T_{TTR} - T_S$.
- Each node uses two timers, the token holding timer (THT) and the token-rotation-timer (TRT). The TRT counter increases only when the node is not transmitting real-time traffic, whereas the THT only increases when the node is delivering best-effort traffic.
- Only the node holding the token can transmit messages. Transmission is controlled by the timers, but an in-progress transmission of a single packet is not interrupted until its completion. When node $i$ gets the token, it performs the following operations:
  - If $L_c > 0$, it sets $L_c = L_c - 1$ and $THT = T_{TTR}$. Otherwise, $THT = TRT$ and $TRT = 0$.
  - If node $i$ has synchronous packets, it transmits them for a time no greater than $H_i$.
- If node $i$ has best-effort packets, it transmits them until $THT$ counts up to $T_{TTR}$, or until all the asynchronous traffic is sent, which ever comes first.
- Node $i$ passes the token to station $(i + 1) \mod (n)$.

4. The BuST protocol

The BuST protocol has been devised to improve the ability of TTP in managing real-time traffic, and to overcome the problems MTTP can have in managing best-effort traffic (see Section 7). Unlike timed token protocols, the BuST protocol assigns each node a time budget $H_i$ for transmitting its real-time traffic. When a node receives the token, it can transmit the associated real-time traffic for a time no greater than the corresponding budget. The main difference with respect to TTP and MTTP concerns the best-effort traffic service. Under TTP, when the token arrives early, the node can transmit best-effort traffic for a time no greater than $T_A = T_{TTR} - \tau - T_{LRD}$, where $T_{LRD}$ is the time spent in the last round-trip of the token. Using MTTP, a node does the same, but with $T_A = T_{TTR} - \tau - \sum H_i$. With BuST, a node can deliver non real-time traffic each time it gets the token, early or not, using the spare budget left by real-time messages. If $H_i^{cons}$ is the budget consumed by node $i$ to deliver periodic traffic, then it can send best-effort traffic for a time no greater than $T_A = H_i - H_i^{cons}$, even if
the token is not early. Observe that, \( TTP \) and \( MTTP \) can deliver best-effort traffic only when the token is early, that is, when \( T_{LRT} < TTRT - \tau \).

The following rules specify the the \( BuST \) protocol in detail:

- During the network initialization phase, each node \( i \) declares a \( TTRT \) value equal to the deadline \( D_i \) of its periodic message stream. The minimum declared value is selected as \( TTRT \). Each node \( i \) is assigned a time budget \( H_i \) that depends on \( TTRT \).
- Each node has one timer, the token holding-rotation timer \( THRT \). The \( THRT \) counter always increases. To initialize the timers, no messages are sent during the first token rotation.
- Only the node having the token can transmit messages. The transmission is controlled by \( THRT \), but an in-progress transmission of a single packet is not interrupted until its completion. When node \( i \) gets the token, it performs the following operations:
  - It sets \( THRT = 0 \).
  - If node \( i \) has real-time packets, it transmits them until \( THRT \) counts up to \( H_i \) or until all the real-time traffic is sent, whichever comes first.
  - If node \( i \) has best-effort packets, it transmits them until \( THRT \) counts up to \( H_i \) or until all the best-effort traffic is sent, which ever comes first.
  - If a real-time message becomes ready during the transmission of best-effort packets, and \( THRT < H_i \), the transmission is stopped and the node starts delivering the real-time traffic until \( THRT \) counts up to \( H_i \) or until all the periodic traffic is sent, whichever comes first.
  - If node \( i \) completes the transmission of the periodic traffic without entirely consuming its budget, i.e. \( THRT < H_i \), it starts transmitting its non real-time traffic, if any, until \( THRT \) counts up to \( H_i \) or until all the best-effort traffic is sent, which ever comes first. Note that, in this case, the transmission is not stopped even if a real-time message becomes ready.
  - Node \( i \) passes the token to station \( (i +1) \mod (n) \).

The overhead generated when a best-effort transmission is interrupted can be easily accounted in \( \tau \). The same observation can be done for the overhead generated when an in-progress packet transmission is not interrupted until its completion. This also holds for \( TTP \) and \( MTTP \).

As we can note from the protocol rules, under \( BuST \), any node \( i \) exploits its time budget \( H_i \) to deliver both real-time and non real-time messages. When compared to \( TTP \), \( BuST \) improves (as \( MTTP \)) the bandwidth available for real-time messages and halves the bandwidth lost due to the protocol overhead. In addition, \( BuST \) is able to deliver best-effort traffic also in those cases in which \( MTTP \) fails, as it will be shown in Section 7.

As a final remark, the implementation of \( BuST \) only requires one timer, instead of the two timers needed by \( TTP \) and \( MTTP \). This can be useful when \( BuST \) is adopted in small embedded systems, where resources (e.g., hardware timers) are scarce.

5. Time properties

In this section, we introduce the main time properties of the protocols under analysis. These properties are the basis to understand the protocols timing behaviour, to verify if a given periodic stream set \( M = \{S_1, S_2, ..., S_n\} \) is feasible, i.e. both the Protocol and the Deadline
Constraints are met. Moreover, the results showed in the following can be used to analyse the protocols performance under the budget allocation schemes introduced in the next sections.

Lemmas 5.1, 5.2 and 5.3 provide an upper bound on the maximum transmission time for a real-time message under the BuST, MTTP and MTTP protocols.

**Lemma 5.1** Under the BuST protocol, for all budget allocation schemes, if \( T_i \geq TTRT \), \( i=1, \ldots, n \), it holds:

\[
\forall i, j: s_{i,j} \leq t_{i,j} + \left[ \frac{C_i}{H_i} \right] \left( \sum_{k=1}^{n} H_k + \tau \right)
\]

(10)

**Proof.** See (Franchino et al., 2007). □

**Lemma 5.2** Under the TTP protocol, for all budget allocation schemes, if \( T_i \geq 2TTRT \), \( i=1, \ldots, n \), it holds:

\[
\forall i, j: s_{i,j} \leq t_{i,j} + \left( \left[ \frac{C_i}{H_i} \right] + 1 \right) TTRT + C_i \left( \left[ \frac{C_i}{H_i} \right] - 1 \right) H_i
\]

(11)

**Proof.** See (Franchino et al., 2007). □

**Lemma 5.3** Under the MTTP protocol, for all budget allocation schemes, if \( T_i \geq TTRT \), \( i=1, \ldots, n \), it holds:

\[
\forall i, j: s_{i,j} \leq t_{i,j} + \left[ \frac{C_i}{H_i} \right] TTRT + C_i \left( \left[ \frac{C_i}{H_i} \right] - 1 \right) H_i
\]

(12)

**Proof.** See (Franchino et al., 2007). □

The results and proofs in (Chen & Zhao, 1992) and (Chan et al., 1997) have been used as basis to derive the properties of BuST shown in the following. For comparison, the same type of results for TPP and MTTP, available in literature, are reported as well.

**Theorem 5.1** Under the BuST protocol, for any \( i=1, \ldots, n \), and for any integer \( l > 0 \) it turns out that:

\[
t_i(l+1) - t_i(l) \geq \sum_{j=1}^{n} H_j + \tau
\]

(13)

where \( t_i(l) \) is the time the token makes the \( l \)-th visit at node \( i \).
Proof. Since a node \(i\) can use only its time budget \(H_i\) to transmit both real-time and best-effort traffic, the length of interval \([t_i(l), t_i(l+1)]\) is equal to the sum of:

1. time for real-time traffic transmission: \(t_{jrt} \leq \sum_{j=1}^{n} H_j\);
2. time for non-real-time traffic transmission: \(t_{nrt} \leq \sum_{j=1}^{n} H_j - t_{n}\);
3. time due to protocol overhead and token passing: \(\tau\).

Thus,

\[ t_i(l+1) - t_i(l) = t_{nrt} + t_{n} + \tau \leq t_{n} + \sum_{j=1}^{n} H_j - t_{n} + \tau = \sum_{j=1}^{n} H_j + \tau. \]

\[ \square \]

The following corollary generalizes the last theorem providing the upper bound on the time elapsed between \(v\) consecutive token arrivals at the same node.

**Corollary 5.1** Under the BuST protocol, for any, \(i=1, \ldots, n\), and for any integer \(l > 0\) and \(v > 0\) it turns out that:

\[ t_i(l + v) - t_i(l) \leq v \left(\sum_{j=1}^{n} H_j + \tau\right) \tag{14}\]

where \(t_i(l)\) is the time the token makes the \(l\)-th visit at node \(i\).

**Proof.** By Theorem 5.1, it follows that:

\[ t_i(l + v) - t_i(l) = [t_i(l+1) - t_i(l)] + [t_i(l+2) - t_i(l+1)] + \ldots + [t_i(l+v) - t_i(l+v-1)] \]

\[ \leq \sum_{j=1}^{v} H_j + \tau + \sum_{j=1}^{v} H_j + \tau + \ldots + \sum_{j=1}^{v} H_j + \tau = v \left(\sum_{j=1}^{n} H_j + \tau\right). \]

\[ \square \]

The following result gives the upper bound between two consecutive token visits at the same node with the timed token protocol.

**Theorem 5.2** Under the TTP protocol, for any, \(i=1, \ldots, n\), and for any integer \(l > 0\) it turns out that:

\[ t_i(l+1) - t_i(l) \leq TTRT + \sum_{j=1}^{n} H_j + \tau \leq 2 \cdot TTRT \tag{15}\]

**Proof.** See (Sevcik & Johnson, 1987). \(\square\)
The next corollary, provides the upper bound on the time elapsed between \( v \) consecutive token visits at the same node.

**Corollary 5.2** Under the TTP protocol, for any \( i=1, \ldots, n \), and for any integer \( l > 0 \) and \( v > 0 \) it turns out that:

\[
t_i(l + v) - t_i(l) \leq v \cdot TTRT + \sum_{j=1, j \neq i}^{n} H_j + \tau
\]

(16)

where \( t_i(l) \) is the time the token makes the \( l \)-th visit at node \( i \).

**Proof.** See (Chen & Zhao, 1992). \( \square \)

The bound between two consecutive token arrivals at the same node, under the modified timed token protocol, is provided by the following theorem.

**Theorem 5.3** Under the MTTP protocol, for any \( i=1, \ldots, n \), and for any integer \( l > 0 \) it turns out that:

\[
t_i(l + 1) - t_i(l) \leq TTRT
\]

(17)

**Proof.** See (Chan et al., 1997). \( \square \)

The corollary below, generalizes the last theorem giving the upper bound between \( v \) consecutive token visits at the same node.

**Corollary 5.3** Under the MTTP protocol, for any \( i=1, \ldots, n \), and for any integer \( l > 0 \) and \( v > 0 \) it turns out that:

\[
t_i(l + v) - t_i(l) \leq v \cdot TTRT
\]

(18)

where \( t_i(l) \) is the time the token makes the \( l \)-th visit at node \( i \).

**Proof.** See (Chan et al., 1997). \( \square \)

The following lemmas, give the minimum amount of time available for node \( i \) to transmit its real-time traffic before the deadline of each message. Notice that, for simplicity’s sake we consider a deadline \( D_i = T_i \). When the deadline \( D_i \) is less than the period \( T_i \), it is sufficient to replace \( T_i \) with \( D_i \) in the following results.

**Lemma 5.3** Under the BuST protocol, for any \( i=1, \ldots, n \), if at time \( t \) a periodic message with period \( T_i \) arrives at node \( i \), then in time interval \( (t, t + T_i] \) the minimum amount of time \( X_i \) available for node \( i \) to transmit the message is:
Token Passing Techniques for Hard Real-Time Communication

\[
X_i \leq \frac{T_i}{\sum_{j=1}^{n} H_j + \tau} \left( H_i + \max(0, H_i - \delta_i) \right) \\
\text{(19)}
\]

where \( \delta_i = \frac{T_i}{\sum_{j=1}^{n} H_j + \tau} \left( \sum_{j=1}^{n} H_j + \tau - T_i \right) \).

**Proof.** Suppose \( t \) the time at which a new message is ready at node \( i \). In the worst case, the token has just left node \( i \) when the new message is ready. As a consequence of Theorem 5.1, in time interval \( I_i = [t, t + \sum_{j=1}^{n} H_j + \tau] \) node \( i \) receives the token at least once. Let \( t + t_1 \) be the time at which node \( i \) receives the token since \( t \), it turns out that

\[
0 < t_1 \leq \sum_{j=1}^{n} H_j - H_i + \tau.
\]

Hence,

\[
H_i \leq \sum_{j=1}^{n} H_j + \tau - t_1,
\]

that is, the time available for node \( i \) to transmit its real-time traffic in \( I_i \) is \( H_i \). Thus, by Corollary 5.1, it turns out that in a time interval \( I_i = [t, m \cdot (t + \sum_{j=1}^{n} H_j + \tau)] \) node \( i \) has a minimum amount of time, to deliver its real-time traffic, equal to \( mH_i \).

If \( \Delta_i = T_i / (\sum_{j=1}^{n} H_j + \tau) \) and \( m = \lfloor \Delta_i \rfloor \), two cases are possible:

1. \( \Delta_i \) is an integer, such that \( \delta_i = 0 \) and \( I_i = [t, t + m \cdot \left( \sum_{j=1}^{n} H_j + \tau \right)] = [t, t + T_i] \), then it follows that

\[
X_i = \frac{T_i}{\sum_{j=1}^{n} H_j + \tau} H_i.
\]

2. \( \Delta_i \) is not an integer, such that \( \delta_i > 0 \). In the worst case, the \((m+1)\)-th token’s arrival at node \( i \) may be as late as

\[
t + m \cdot \left( \sum_{j=1}^{n} H_j + \tau \right) + \sum_{j=1}^{n} H_j + \tau + H_i,
\]

then, this token arrival will be not more than \( t + T_i \) if:

\[
\sum_{j=1}^{n} H_j + \tau - H_i \leq T_i + m \cdot \left( \sum_{j=1}^{n} H_j + \tau \right) = \sum_{j=1}^{n} H_j + \tau - \delta_i.
\]
In this case, the residual transmission time available for the real-time traffic is the left time until \( t+T_i \), which is equal to \( H_i - \delta_i \), that is:

\[
X_i \leq \left[ \frac{T_i}{TTRT} \right] H_i + \max \left\{ 0, H_i - \delta_i \right\}
\]

Lemma 5.4 Under the TTP protocol, for any \( i=1,...,n \), if at time \( t \) a periodic message with period \( T_i \) arrives at node \( i \), then in time interval \( (t,t+T_i] \) the minimum amount of time \( X_i \) available for node \( i \) to transmit the message is:

\[
X_i \leq \left[ \frac{T_i}{TTRT} \right] H_i + \max \left\{ 0, \min \left( \delta_i - \tau - \sum_{j \neq i} \left( H_j \right), H_i \right) \right\}
\]

(20)

where \( \delta_i = T_i - \left[ \frac{T_i}{TTRT} \right] TTRT \).

Proof. See (Chen & Zhao, 1992).

Lemma 5.5 Under the MTTP protocol, for any \( i=1,...,n \), if at time \( t \) a periodic message with period \( T_i \) arrives at node \( i \), then in time interval \( (t,t+T_i] \) the minimum amount of time \( X_i \) available for node \( i \) to transmit the message is:

\[
X_i \leq \left[ \frac{T_i}{TTRT} \right] H_i + \max \left\{ 0, H_i - \delta_i \right\}
\]

(21)

where \( \delta_i = \left[ \frac{T_i}{TTRT} \right] TTRT - T_i \).

Proof. See (Chan et al., 1997).

6. Performance analysis

The real-time guarantee of a stream set highly depends on the budget allocation scheme (BAS) adopted for budgets assignment given the stream set parameters.

In literature exist several budget allocation schemes provided for timed-token protocols. They are traditionally classified into global or local schemes, depending on whether they need global or local information to assign the budgets. Local information is, for instance, the node stream parameters. Global information is, for instance, the number of nodes and the total channel utilization \( U^S \) required by the streams.

In this work, the budget allocation schemes are classified as proposed in (Daoxu et al. 1998). They can be divided into two categories, depending on the way they assign the budgets. The first category is the set of the \( TTRT\)-partitioning schemes, where a scheme belonging to this
set assigns node budgets partitioning the time expected for the token to make a rotation of the network, i.e. $TTRT - \tau$. The $TTRT$-partitioning schemes analyzed in the following are shown in Table 1.

The second category of budget allocation schemes, is the set of $C_i$-partitioning schemes, where a scheme belonging to this class assigns each budget $H_i$ partitioning the maximum time length, $C_i$, to send a message from the stream $S_i$ among a certain number of token rotation cycles. The $C_i$-partitioning schemes analyzed in the following are shown in Table 2.

<table>
<thead>
<tr>
<th>$TTRT$-partitioning schemes</th>
<th>Assignment rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional Allocation (PA)</td>
<td>$H_i = U_j (TTRT - \tau)$</td>
</tr>
<tr>
<td>Normalized Proportional Allocation (NPA)</td>
<td>$H_i = \frac{U_j}{U} (TTRT - \tau)$</td>
</tr>
<tr>
<td>Equal Partition Allocation (EPA)</td>
<td>$H_i = \frac{TTRT - \tau}{n}$</td>
</tr>
</tbody>
</table>

Table 1. $TTRT$-partitioning budget allocation schemes.

<table>
<thead>
<tr>
<th>$C_i$-partitioning schemes</th>
<th>Assignment rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Allocation (LA)</td>
<td>$H_i = \frac{C_i T_i}{TTRT}$</td>
</tr>
<tr>
<td>Modified Local Allocation (MLA)</td>
<td>$H_i = \frac{C_i}{TTRT}$</td>
</tr>
</tbody>
</table>

Table 2. $C_i$-partitioning budget allocation schemes.

The budget allocation schemes showed in the tables above have been extensively analyzed in literature. For a complete survey, an interested reader can see (Zhang et al., 2004) for TTP; (Chan et al., 1997) and (Daoxu et al., 1998) for MTTP; (Franchino et al., 2007), (Franchino et al., 2008) and (Franchino et al., 2008a) for BuST.

In the following, we compare the schemes performance under the token passing protocols considered in this chapter.

### 6.1 Performance metric

For evaluating and comparing the performance of different budget allocation schemes several metrics have been proposed. One of the most widely adopted metric is the Worst Case Achievable Utilization (WCAU). The WCAU of a budget allocation scheme represents the largest utilization ($U^*$) of the network such that, for any real-time message set whose total network utilization is $U_S \leq U^*$, the budget allocation scheme can guarantee the timeliness of each single real-time message.

The WCAU test is useful to guarantee the feasibility of a periodic stream set when only an estimation of the amount of real-time traffic is known (i.e., the maximum time required to send a message) without requiring a detailed characterization of each single real-time message.
6.2 Budget allocation schemes analysis

In the following, we assume \( D_i = T_i \) for all the streams, however, the same results can be derived for the case where \( D_i < T_i \) by simply substituting \( D_i \) to \( T_i \). To make the treatment clearer, when not differently specified, let \( \beta = T_i / TTRT \), \( \beta_{\min} = \min(T_i) / TTRT \) and \( \alpha = T / TTRT \). Parameter \( \alpha \) represents the bandwidth loss due to the overhead.

Table 3 summarizes the WCAUs of the budget allocation schemes considered in this work.

With the PA scheme, the BuST protocol is the only one having a WCAU greater than 0, whereas with TTP and MTTP no stream set can be guaranteed.

With the NPA scheme, BuST and MTTP present the same WCAU which depends on \( \beta_{\min} \) hence the smaller TTRT the greater is the bandwidth that the protocols can guarantee for the real-time traffic. Since \( \beta_{\min} \leq 1 \) the minimum WCAU for the NPA, under both BuST and MTTP, is \((1-\alpha)/2\). Instead, the TTP protocols with the NPA scheme presents a WCAU which does not depends on \( \beta_{\min} \) and it is lower than that presented by the other protocols. It is worth observing that, under the NPA scheme, the MTTP protocol cannot serve best-effort traffic (Franchino et al., 2008). Conversely, the BuST protocol presents the same WCAU of MTTP and can serve also best-effort traffic.

Under the EPA scheme, BuST and MTTP have a greater WCAU with respect to TTP. However, the WCAU of all the tree protocols is very poor and it depends on the number of nodes. As for the NPA scheme, also under the EPA scheme the MTTP protocol cannot serve non real-time traffic.

BuST, MTTP and TTP present the same WCAU under the LA scheme. As for the NPA scheme, the WCAU under the LA scheme depends on \( \beta_{\min} \) i.e. on the choice of TTRT. With the LA scheme \( \beta_{\min} \geq 2 \), thus the minimum WCAU of this scheme is \((1-\alpha)/3\).

With the MLA scheme, TTP presents a null WCAU. Instead, both BuST and MTTP present a WCAU which depends on \( \beta_{\min} \) and equivalent to that presented with the NPA scheme. As shown in (Franchino et al., 2008a), with the MLA scheme MTTP can not serve best-effort traffic when \( \sum_{i=1}^{n} H_i \geq TTRT + \tau \).

<table>
<thead>
<tr>
<th>Allocation schemes</th>
<th>TTP</th>
<th>MTTP</th>
<th>BuST</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>0</td>
<td>0</td>
<td>( \frac{1-3\alpha}{2(1-\alpha)} )</td>
</tr>
<tr>
<td>NPA</td>
<td>( \frac{1-\alpha}{3} )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
</tr>
<tr>
<td>ELA</td>
<td>( \frac{1-\alpha}{3n-(1-\alpha)} )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
</tr>
<tr>
<td>LA (( \beta_{\min} \geq 2 ))</td>
<td>( \beta_{\min} ) ( (1-\alpha) )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
</tr>
<tr>
<td>MLA (( \beta_{\min} \geq 2 ))</td>
<td>0</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
<td>( \frac{\beta_{\min}}{\beta_{\min}+1} ) ( (1-\alpha) )</td>
</tr>
</tbody>
</table>

Table 3. WCAU of the considered schemes.
7. Best-effort traffic service

So far, real-time streams service have been analyzed. This section briefly describes the best-effort service of the BuST protocol and its improvements with respect to MTTP.

As shown in Section 4, under MTTP the maximum time a node can use to deliver non real-time traffic is

$$T_{A} = TTRT - \sum_{i=1}^{n} H_i - \tau .$$

It can be observed that:

- For the PA scheme

$$T_{A} = TTRT - \sum_{i=1}^{n} \frac{U_i^S}{U^S} (TTRT - \tau) - \tau = (1-U^S)(TTRT - \tau) ;$$

- For the NPA scheme

$$T_{A} = TTRT - \sum_{i=1}^{n} \frac{TTTRT}{n} - \tau = 0 ;$$

- For the EPA scheme

$$T_{A} = TTRT - \sum_{i=1}^{n} \frac{TTTRT}{n} - \tau = 0 ;$$

This means that MTTP is not able to deliver best-effort traffic under both the NPA and the EPA schemes, while $T_{A}$ depends on $U^S$ using the PA scheme. Moreover, it can be observed that MTTP can not serve non real-time traffic when $\sum H_i \geq TTRT + \tau$, that is, when the Protocol Constraint is not met. To verify this last statement, it is sufficient to note that, as stated in Section 3.1, each time a node receives the token it can transmit non real-time traffic for a time no greater than $T_{A}$, and since $TTTRT - \tau - \sum H_i < 0$, it follows that $T_{A} = 0$. This means that, when this last condition holds, MTTP can starve best effort traffic also with both the LA and the MLA schemes.

To analyze a worst-case scenario for non real-time service, we assume that each node receiving the token has always best-effort traffic to deliver. In this case, the total channel utilization of the network, including real-time and best-effort traffic, is equal to $1-\alpha$, i.e. the channel is fully utilized.

The following theorem provides the minimum bandwidth that a node $i$ can exploit to deliver non real-time traffic under the BuST protocol.

**Theorem 7.2** Under the BuST protocol, a node $i$ can guarantee for the non real-time traffic a minimum bandwidth $U_{i}^{\text{net}}$, which depends on the budget allocation scheme adopted. In particular:

- with PA, $U_{i}^{\text{net}} = U_{i}^{S} \left( \frac{1}{U^S + \frac{\alpha}{1-\alpha}} - 1 \right)$;

- with NPA, $U_{i}^{\text{net}} = U_{i}^{S} \left( \frac{1-\alpha}{U^S} - 1 \right)$;

- with EPA, $U_{i}^{\text{net}} = \frac{1-\alpha}{n}$;
Proof. See (Franchino et al., 2007), (Franchino et al., 2008) and (Franchino et al., 2008a).

We have shown that, under the BuST protocol, non real-time traffic at each node has a minimum bandwidth guaranteed. In addition, it is worth observing that, when not all nodes of the network have to send best-effort traffic, during the round trip of the token, the value of $U_{nt}$ can increase.

8. Simulation results

In this section the performance of BuST, TTP and MTTP is compared by simulation. The simulations have been performed through a discrete-event simulator written for this purpose in C language. The simulation scenario considers a network consisting of 10 nodes. Each node has a periodic stream with a relative deadline ranging from 10 msec to 100 msec. An infinite amount of non real-time traffic is assumed: this means that, every time a node receives the token, it has some non real-time traffic to deliver. As already stated in Section 7, in this case the total channel utilization $U_{TOT} = U_{nt} + U_S$ is equal to the total available bandwidth $1 - \alpha$. Node budgets are assigned using the PA, NPA, LA, and MLA budget allocation schemes.

Performance is evaluated through the Maximum Deadline Miss Ratio (MDMR), defined as the ratio between the number of messages that miss their deadline and the total number of generated messages. The MDMR is measured as a function of the real-time channel utilization $U_S$, ranging from 0.1 to 1.0 (step of 0.1). For each value of $U_S$, up to 1000 simulation runs are performed, and the MDMR is considered among the runs. A different stream set is generated each run. In particular, for each stream set the utilizations $U_S$ have been generated randomly with a uniform distribution using the method proposed in (Bini & Buttazzo, 2004). For each value of $U_S$, a relative deadline $D_i$ is generated randomly with a uniform distribution in the interval [10, 100] msec. Periods are assumed equal to deadlines, i.e. for all $i$, $T_i = D_i$. The message lengths $C_i$ have been computed as $C_i = U_S D_i$. The overhead $\tau$ is assumed equal to 20 $\mu$sec.

In the simulations, the token is considered as never lost and no ring recovery process is implemented. In this way, a violation of the Protocol Constraint will not compromise the stability of the protocols.

For each scheme, three different scenarios have been considered in the simulations. One where $TTRT = min(D_i)$, one where $TTRT = min(D_i)/2$, and the last with $TTRT = min(D_i)$ and only real-time traffic in the network.

Note that, when there is only real-traffic BuST, TTP and MTTP operate in the same way, that is, they are in practice the same protocol. Therefore, as it will be shown in the following, in case of only real-time traffic all the three protocols present the same performance in terms of MDMR.

8.1 Results with the PA scheme

In this subsection the results with the PA scheme are analyzed. Figure 2 shows the MDMR when the PA scheme is used to assign the node budgets, and $TTRT = min(D_i)$.

As expected from the results showed in Section 6, as long as $U_S \leq 0.5$, with BuST all the messages meet their deadlines (MDMR = 0), while TTP and MTTP present a non-null
Token Passing Techniques for Hard Real-Time Communication 257

deadline miss ratio. For \( U^S = 0.6 \), BuST presents a very small MDMR (equal to 0.5%) that cannot be appreciated in the figure. For \( U^S \geq 0.7 \), BuST presents a significant MDMR which is about 76% when \( U^S = 1 \), that is, when the channel is over-utilized (remember that the available bandwidth is 1-\( \alpha \)).

While TTP presents a significant MDMR for all values of \( U^S \), MTTP presents a MDMR lower than that shown by BuST for \( U^S \geq 0.7 \).

Figure 3 shows the MDMR with \( TTRT = \min(D_i)/2 \). Notice that, the performance of both BuST and MTTP improve as expected. Instead TTP still presents a very poor performance.

Figure 4 shows the MDMR when nodes have only real-time traffic to deliver and \( TTRT = \min(D_i) \). In this case, as said before, the three protocols operate in the same way, producing the same performance. As long as \( U^S \) is not greater than 0.5, there are not deadline misses. With \( U^S = 0.6 \) we have a very small MDMR, which is equal to 0.69%. For \( U^S \geq 0.7 \) the MDMR starts increasing significantly.

### 8.2 Results with the NPA scheme

Figure 5 reports the MDMR when the NPA scheme is used to assign node budgets and \( TTRT = \min(D_i) \).

As long as \( U^S \leq 0.5 \), BuST and MTTP have no deadline miss; for \( U^S \geq 0.6 \), they start experiencing deadline misses. It is worth noticing that, for \( U^S = 0.6 \), BuST presents a MDMR close to 0%, which is not appreciable in the figure.

TTP has deadline misses for all values of \( U^S \); this is due to the fact that TTP requires that \( TTRT \leq \min(D_i)/2 \) to work properly.

For \( U^S \geq 0.9 \), MTTP performs better than BuST. However, it is worth remembering that, under the NPA scheme, MTTP is not delivering non real-time traffic. This is the reason why it can provide a better service for real-time traffic.

Figure 6 shows the MDMR with \( TTRT = \min(D_i)/2 \). Notice that, with a lower \( TTRT \), the performance of all the three protocols improves, as expected by the theoretical analysis showed in the previous sections. In particular, TTP has no deadline miss as long as \( U^S \leq 0.3 \).

For \( U^S = 0.4 \) and \( U^S = 0.5 \), TTP presents an MDMR close to 0%. For \( U^S \geq 0.6 \), TTP presents an MDMR significantly greater than 0%. Notice that, BuST and MTTP provide more or less the same performance, with the difference that BuST serves also non real-time traffic.

Figure 7 shows the MDMR when nodes have only real-time traffic to deliver and \( TTRT = \min(D_i) \).

### 8.3 Results with the LA scheme

Figure 8 shows the MDMR when the LA scheme is used to assign node budgets and \( TTRT = \min(D_i)/2 \).

As long as \( U^S \leq 0.8 \), MTTP and BuST present a null MDMR. For \( U^S = 0.9 \), they present a MDMR not appreciable in the figure, which is less than 0.05% for both protocols. TTP presents a null MDMR as long as \( U^S \leq 0.4 \), and a MDMR less than 0.3% for \( 0.5 \leq U^S \leq 0.7 \), which is not appreciable in the figure. For \( U^S > 0.7 \), TTP presents a MDMR significantly greater than BuST and MTTP.

Figure 9 shows the MDMR when the nodes have only real-time traffic to deliver. As it can be noted, the absence of non real-time traffic improves the protocols performance.
considerably. In particular, as long as \( U^S \leq 0.8 \), the MDMR is null; for \( U^S = 0.9 \), the MDMR is still close to 0\%, and when \( U^S = 1 \) the MDMR is about 7\%.

Finally, it is important to notice that a message deadline miss is due to the Protocol Constraint violation. Furthermore, as highlighted in Section 7, when the Protocol Constraint is not met, or even if the sum of the budgets is equal to \( TTRT - \tau \), \( MTTP \) cannot serve non real-time traffic.

### 8.4 Results with the MLA scheme

Figure 10 shows the MDMR when the MLA scheme is used to assign the node budgets, and \( TTRT = \min(D_i) \).

As long as \( U^S \leq 0.7 \), \( MTTP \) and \( BuST \) present a null MDMR; for \( U^S = 0.8 \) and for \( U^F = 0.9 \), the MDMR is less than 1\% for both protocols. Under \( TTP \), the MDMR is non-null for all values of \( U^S \). This is due to the fact that, as stated in Section 6, the Deadline Constraint cannot be satisfied when the MLA scheme is used under \( TTP \).

Figure 11 shows the MDMR under the MLA scheme when \( TTRT = \min(D_i)/2 \). For \( BuST \) and \( MTTP \), as long as \( U^S \leq 0.8 \), the MDMR is null. When \( U^S = 0.9 \), the MDMR is not greater than 0.3\%, hence is not appreciable in the figure, while when the channel is overloaded, i.e. \( U^S = 1 \), the MDMR is approximately equal to 15\%.

Figure 12 depicts the MDMR when nodes have only real-time traffic to deliver and \( TTRT = \min(D_i) \). As said before, in this case, all the three protocols present the same performance, and the absence of non real-time traffic improves the protocols performance considerably. In particular, as long as \( U^S \leq 0.8 \), the MDMR is null or very small. For \( U^S = 0.9 \), the MDMR is less than 2\%, and when \( U^S = 1 \) the MDMR is close to 5\%.

As highlighted in Section 7, when the MDMR is not null it means that the Protocol Constraint is not met and, because of this, \( MTTP \) is not able to deliver non real-time traffic.

![Fig. 2. Maximum Deadline Miss Ratio with the PA scheme.](image-url)
Fig. 3. Maximum Deadline Miss Ratio with the PA scheme when $TTRT = \min(D)/2$.

Fig. 4. Maximum Deadline Miss Ratio with the PA scheme with only best-effort traffic.

Fig. 5. Maximum Deadline Miss Ratio with the NPA scheme.
Fig. 6. Maximum Deadline Miss Ratio with the NPA scheme when $TTRT = \min(D_i)/2$.

Fig. 7. Maximum Deadline Miss Ratio with the NPA scheme with only best-effort traffic.

Fig. 8. Maximum Deadline Miss Ratio with the LA scheme.
Fig. 9. Maximum Deadline Miss Ratio with the LA scheme with only best-effort traffic.

Fig. 10. Maximum Deadline Miss Ratio with the MLA scheme.
Fig. 11. Maximum Deadline Miss Ratio with the MLA scheme when $TTRT = \min(D_i)/2$.

Fig. 12. Maximum Deadline Miss Ratio with the MLA scheme with only best-effort traffic.

9. Conclusions

In this work, we introduce and analyzed the BuST protocol in comparison with timed token and modified timed token protocols. All the three protocols are discussed in detail, showing their strengths and their drawbacks under different budget allocation schemes. New and well known time properties of the protocols are presented, together with their capability on managing both hard real-time and best-effort traffic. In particular, we have seen that both BuST and MTTP are superior on serving real-time traffic with respect to the traditional
timed token protocol. However, it has been shown that when both NPA and EPA schemes are used to assign the node budgets, MTTP can starve best-effort traffic. Conversely, the BuST protocol can serve also best-effort traffic under all the analyzed budget allocation schemes.

Future work is focused on extending the performance analysis of BuST with other allocation schemes available in literature.

10. References


Factory automation has evolved significantly in the last few decades, and is today a complex, interdisciplinary, scientific area. In this book a selection of papers on topics related to factory automation is presented, covering a broad spectrum, so that the reader may become familiar with the various fields, and also study them in more depth where required. Within various chapters in this book, special attention is given to distributed applications and their use of networks, since it is one of the most relevant subjects in the evolution of factory automation. Different Medium Access Control and networks are analyzed, while Ethernet and Wireless networks are looked at in more detail, since they are among the hottest topics in recent research. Another important subject is everything concerning the increase in the complexity of factory automation, and the need for flexibility and interoperability. Finally the use of multi-agent systems, advanced control, formal methods, or the application in this field of RFID, are additional examples of the ideas and disciplines that experts around the world have analyzed in their work.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
