TIMING ANALYSIS OF THE ARINC 629 DATABUS FOR REAL-TIME APPLICATIONS

Dr. N. C. Audsley & A. Grigg :
BAe. Dependable Computing Systems Centre,
Dept. of Computer Science, University of York, UK

1. Introduction

This paper presents an outline description of the ARINC 629 data communication standard from the perspective of timing behaviour and proposes an approach to predict communication delays under the various protocol configurations offered by the standard.

The communication delays associated with ARINC 629 are described in terms of :

- potential sources of delay and associated conditions likely to induce worst case behaviour,
- possible worst case durations associated with each of these delays, given certain assumptions about implementation.

This study was carried out at the British Aerospace Dependable Computing Systems Centre, University of York.

2. Background

The ARINC 629 civil aircraft databus standard [1] has been developed as a successor to ARINC 429, intended to provide general purpose data communications between avionic sub-systems (LRUs - line replaceable units). The 629 standard may be employed in place of or in conjunction with[1] 429 and is applicable to traditional, *federated* systems as well as future civil IMA systems.

The original 429 standard, defining a uni-directional broadcast databus (single transmitter, multiple receivers), has been used extensively in the civil[2] aircraft field, notably by Boeing on the 757 and 767 and also by Airbus. Research and development work carried out by Boeing in the form of the DATAC (Digital Autonomous Terminal Access Communication) programme has, however, resulted in a new standard being agreed for use on modern civil aircraft demanding more data processing and transfer at higher speeds. This was applied on the Boeing 777 [2] and became the ARINC 629 standard in 1989.

3. Outline Description

The ARINC 629 standard defines a multi-level protocol for inter-LRU communications across a common, multiple access databus.

The following paragraphs give a basic description of each of the levels defined (see also Figure 1) :

- Upper layer - This provides a sub-set of the standard ISO-OSI (Open System Interconnection) reference model [3] application, presentation, session and transport functions.
- Network layer - This deals primarily with data structuring issues.

  A *message* has variable length and is comprised of up to 31 *wordstrings*. Each wordstring has variable length and contains one (20 bit) *label word* and up to 256 (20 bit) *data words*. Interestingly, the ARINC 653 application software / run-time executive interface, APEX [4], does not impose a maximum message length restriction and so, if 629 is used in conjunction with some APEX-compliant run-time executive, message partitioning must be provided as a function of the executive.

[1] ARINC 629 and 429 are not directly compatible but may still be used on the same aircraft.
[2] There are also a number of military applications of ARINC 429, particularly for helicopters.

Broadcast and directed messages (to a specific terminal set) are permitted.

- Data link layer - The main responsibility of this layer is to co-ordinate access to the single channel medium across all terminals. The adopted approach aims to provide fair distribution of access and employs a collision-avoidance philosophy. This is achieved without the stipulation of any global time source, *ie.* each terminal has its own local time source which may drift relative to others.

- Physical layer - Within each LRU, a single 629 terminal and the main computational component(s) interact via an area of shared memory. Whilst the standard does not define any mechanisms to govern such interactions, it does not assume that shared memory access is contention-free.

  A 2 Mbps serial data transmission rate is specified for twisted pair conductors. The use of fibre optics is expected to permit higher rates to be specified some time in the future, but there are still some concerns over how the various levels of protocol will scale. To this end, Boeing is currently involved in a number of ground-based and airborne evaluation programmes with other industrial partners and airline customers [2].

  To reinforce the approach of collision avoidance adopted at the data link level, multiple timers and redundant circuitry are employed within each terminal to prevent single hardware faults causing multiple terminals to transmit simultaneously.

  The maximum number of terminals permissible on the bus is 120.

The rest of this paper will concentrate primarily on the data link layer, which has the most significant influence on the timing behaviour of the databus under normal operating conditions.

4. Data Link Layer Protocols - Outline Description and Basic Timing Considerations

The standard supports two alternative data link level protocols, termed the *basic protocol* and the *combined protocol*, which cannot co-exist on the same bus due to fundamental differences in the way the general bus access control strategy is applied (as described in sub-sections below). However, there are many similarities between the two, as the rest of this section indicates.

In general, bus access is co-ordinated at two levels :

1. System level decomposition of bus time into cycles, minor and major, which may be of fixed duration (TDMA-like (Time Division Multiple Access) [5]) or variable depending on protocol configuration.
2. Medium access contention resolution across multiple terminals within each minor cycle, by a combination of carrier sensing and observation of pre-assigned waiting times and common *bus quiet* periods, *ie.* periods of inactivity. This level of control is, therefore, CSMA-like (Carrier Sense Multiple Access) [5] but with a collision avoidance philosophy.

The basic and combined protocols both support periodic and sporadic transmissions. Periodic messages are scheduled automatically at the 629 terminal level, according to pre-defined message tables (defining message contents, destinations, *etc.*) and transmission control parameters. It is the manner in which sporadics are handled which provides the main difference between the two protocols.

The medium access control policy associated with each protocol will now be described. This includes a basic timing assessment which does not specifically consider the initialisation phase of bus operation, which can be relatively uncontrolled, depending on the sequence in which local clocks are initialised across multiple terminals. Due to certain synchronisation mechanisms, however, bus operation will normally settle down to some steady state after a very short time, at which point it is appropriate to consider the co-ordinated behaviour of terminals in the form of minor cycles.

4.1 Basic Protocol (BP)

Two sub-modes are defined for BP; the periodic mode, where all transmission times are fixed within each minor cycle, and the aperiodic mode, where individual terminal transmission times may vary between cycles.

In both modes, access is granted to each terminal in a fixed order and without pre-emption within each minor cycle. In the aperiodic mode, sporadic messages may be serviced at a terminal by adding time to the current access period for that terminal - minor cycles are permitted to over-run, if necessary, until all terminals have been granted their pre-defined, 'periodic' access times. The two modes are not exclusive and the periodic mode will revert to aperiodic in the event of transient overload.

For a given terminal, i, from a set of n on the bus, TG_i represents a unique time (known as the *terminal gap*) for which a terminal must wait after any bus activity before starting its own transmission. Tx_i represents the transmission time of that terminal and may vary between minor cycles, depending on the mode of operation (with worst case Tx_i denoted by TX_i).

For all n terminals, TI represents a common time interval (known as the *transmit interval*) for which each terminal must wait between its own successive transmissions. Further, SG (*synchronisation gap*) is a common quiet time, longer than any individual terminal gap on the bus and is intended to cater for variations in actual transmission times of individual terminals between successive minor cycles.

Three timers are utilised at each terminal as follows :

- A TI timer which starts immediately every time the terminal commences transmission.
- An SG timer which starts immediately every time the bus is sensed quiet. The timer may be reset either before it has elapsed, if any bus activity is detected, or after it has elapsed, the next time the local terminal begins transmitting.
- A TG timer which starts only after SG has elapsed. Like SG, TG starts immediately every time the bus is sensed quiet. Unlike SG, however, the timer is reset only after it has elapsed and upon detection of any bus activity.

These timers have different effects on behaviour, depending on the operating mode and the manner in which they are assigned values.

The assignment of timer values is done on a per-system basis and must address a trade-off between bus cycle time and terminal transmission frequency, *ie.* the cycle time must allow all terminals to transmit exactly once within a cycle, whilst terminals must also be permitted to transmit at a frequency suitable to meet message update rate requirements.

4.1.1 Basic Timing Assessment for BP

Figure 2a[3] shows the timing behaviour associated with a single minor cycle in BP periodic mode. The first part of our assessment applies to this mode.

Each terminal has its own concept of a minor cycle, since clocks are independent. At the 629 bus level, however, encompassing all terminals, a minor cycle may be defined as the time between successive synchronisation events to occur. This is comprised of a bus idle time until some (any) TI timer elapses, followed by a sequence of terminal transmissions separated only by TG delays, followed by the SG delay required to achieve re-synchronisation.

The order in which terminals are granted bus access is the same for all minor cycles but initial ordering is determined by the initialisation sequence and relative drift between local clocks (only for

[3] Figures 2a, 2b and 3 are intended only to highlight timing properties relevant to this study - it is acknowledged that more detailed timing diagrams exist, *eg.* in the standard itself.

synchronous initialisation of clocks and zero drift does this ordering equate to increasing *TG* values, $TG_1 < TG_2 < ... < TG_n$).

A fixed minor cycle duration, equal to the value of *TI*, can be guaranteed (without over-run) provided that the following relationship holds across all minor cycles :

$$TI \geq (\sum_{i=1}^{n}(TG_i + Tx_i)) + SG \qquad ...(1)$$

Hence, assignment of *TI* timer value based on worst case message transmission times (*TX*) would guarantee no minor cycle over-run for normal, fault-free operation. (Remember that *TG* values are constant.)

This will still result, however, in variable bus access times, *ie.* release jitter, for individual terminals across minor cycles. To guarantee true periodic behaviour for individual terminals, it must be assured that periodic transmission lengths are constant, *ie.* Tx_i is constant (= TX_i) for all *i*. Under such a restriction, the above inequality (1) reduces to an equality relation by which a suitable *TI* value may be assigned.

If, under any circumstances, whether planned or unplanned, the following condition (2) holds[4], then the basic protocol will operate in aperiodic mode :

$$TI < (\sum_{i=1}^{n}(TG_i + Tx_i)) + SG \qquad ...(2)$$

Figure 2b shows the timing behaviour associated with a single minor cycle in BP aperiodic mode. The remainder of our assessment applies to this mode.

Each minor cycle is comprised of a sequence of terminal transmissions separated only by *TG* delays, followed by the *SG* delay required to achieve re-synchronisation. The order in which terminals are granted bus access is again the same for all minor cycles but, in the aperiodic mode, this ordering equates directly to increasing *TG* values, *ie.* it is always the case that $TG_1 < TG_2 < ... < TG_n$.

Since this mode is intended to cater for both periodic and sporadic transmissions, some other mechanism is required to achieve terminal synchronisation at the minor cycle level. It is the *SG* timer, rather than the *TI* timer, which provides such synchronisation, but only at the expense of introducing variable length minor cycles. Hence, the stability of periodic transmissions between successive minor cycles is directly dependant on the degree of sporadic behaviour in the earlier cycle, making periodic release jitter a significant problem in this mode.

It is a direct result of synchronisation on *SG* that terminals are granted bus access directly in order of increasing *TG*, since all *TI* timers will elapse before *SG* has taken effect. This is apparent from the inequality (2).

The only way to avoid variable length minor cycles is to anticipate worst case sporadic behaviour and calculate a *TI* value accordingly - effectively reverting back to periodic mode but accepting a high degree of bus under-utilisation.

It turns out, in fact, to be very difficult to make any behavioural guarantees for the aperiodic operating mode other than basic minor cycle synchronisation and consistent ordering of terminal transmissions.

[4] (1) and (2) are clearly mutually exclusive.

4.2 Combined Protocol (CP)

This protocol was originally proposed and developed by British Aerospace and Smiths Industries. The aim was to overcome the shortfalls of the basic protocol for systems which require a more effective approach to combined handling of periodic and sporadic data transmissions.

Periodic transmissions (termed *level one* in CP) are serviced in a fixed order and without pre-emption, as in the basic protocol. Bus cycle times are, however, fixed in duration (for appropriate assignment of bus control parameters).

Any sporadic which arrives during the current minor cycle may only be serviced within the cycle time available after all periodics have completed. Sporadics which are ‘shorter and more frequent’ (termed *level two*) are serviced first followed by, in whatever cycle time remains, sporadics which are ‘longer and less frequent’(termed *level three*). Level two and level three messages are limited to one wordstring in length.

Figure 3 shows the timing behaviour associated with a single minor cycle in CP.

As for BP, each terminal has a unique terminal gap (*TG*) pre-assigned. The transmit interval (*TI*) is only applicable, however, to the first (periodic) terminal transmission in each minor cycle - for all other terminals, this is over-ridden by a *concatenation event* (*CE*) which forces all, un-elapsed *TI* timers to be cancelled. This has the effect of compressing periodics into a burst of activity (separated only by *TG* delays) at the start of each cycle.

Two types of synchronisation gap are defined. Firstly, the *periodic synchronisation gap* (*PSG*) is used to achieve synchronisation at the minor cycle level. Secondly, an *aperiodic synchronisation gap* (*ASG*) is employed within each minor cycle to synchronise transition between level one and two transmissions and, in turn, level two and three transmissions. To this end, the *ASG* timer is started upon detection of bus quiet after the initial burst of periodic transmissions and reset before it has elapsed upon detection of any bus activity. The use of ASG in this manner causes behaviour internal to a minor frame to be similar that of the aperiodic mode of BP. As a result, sporadics only consume resources on an ‘as required’ basis - all terminals are offered access to the bus in levels two and three, but this is only accepted by terminals which have sporadics ready to transmit, which gives level three sporadics a better chance (on average) of being serviced.

Within each minor cycle, each terminal is restricted to one level two transmission but may perform multiple level three transmissions if required and if there is time available. Whenever any terminal (i) transmits multiple level three messages in a single minor cycle, these will be separated by at least *ASG* plus TG_i (since terminals are granted access in order of *TG* and bursts of level three transmissions are separated by *ASG*).

In order to enforce fixed duration minor cycles, an *aperiodic access time-out* (*AT*) is used to indicate the time to next periodic (level one) transmission, so that any sporadic transmission with the potential to take longer than this time is prevented from starting in the current cycle. Level three sporadics are permitted to span multiple minor cycles; backlog level three messages always take priority over those generated in the current cycle. Level two sporadics must be transmitted within the current minor cycle, otherwise they are lost.

4.2.1 Basic Timing Assessment for CP

The considerations involved here in *TI* value assignment are the same as in the basic protocol. For *ASG* there is an equivalence to *SG* in the aperiodic mode of the basic protocol. Based on these values, the following recommendations are given by the standard for assignment of other control parameters :

$$PSG \geq 5ASG \qquad ...(3)$$ [5]

[5] This ensures that PSG is always greater than any other possible gap between successive transmissions.

$$AT = TI - (MAL + PSG) \qquad ...(4)$$

where *MAL* is the maximum time for a sporadic transmission (maximum length 257 words).

Unlike BP, fixed minor cycle duration is guaranteed by the CP protocol. Further, due to the lack of sporadic interference with periodic (level one) transmissions, release jitter does not affect periodics to the same extent as in BP; in fact, if periodic transmissions are of fixed length, release jitter can be virtually eliminated.

The order in which terminals are granted level one bus access is the same for all minor cycles but, like the periodic mode in BP, initial ordering is determined by the initialisation sequence and relative drift between local clocks.

For level two sporadics, behavioural guarantees can be provided if the level two transmission sequence is designed around worst case occurrence, *ie.* minimum inter-arrival rate and message length, although this may result in considerable under-utilisation of the bus in the 'average' minor cycle. Anything less than this will, however, result in failure to transmit all outstanding sporadic messages in the current minor cycle as sporadic behaviour approaches worst case.

The situation is not good for level three sporadics, since there is no guarantee of any minimum service time per minor cycle (nor across a number of cycles). This makes the provision of timing guarantees very difficult in the general case. In the extreme case of an implementation of CP based on constant worst case behaviour for all communications (levels one, two and three equally), level three sporadics could be guaranteed. This solution would, however, be equivalent to a traditional TDMA static solution, which can be inefficient for systems which require some degree of sporadic behaviour.

5. Assessment of End-to-end Delays

We will now consider the ARINC 629 protocols in the context of distributed, real-time applications which require guarantees of overall end-to-end timing behaviour.

The term *end-to-end message delay* refers to the time between message 'production' and 'consumption' at the application level. The descriptions of such given here are general in nature and actual delays will vary between particular implementations of the 629 standard and the nature of its interactions at each processing node with the run-time executive via local shared memory.

5.1 Basic Protocol

Under the basic protocol, any given data transfer may potentially be delayed at three stages[6], assuming some operating system (*eg.* an ARINC/APEX-compliant run-time executive) is resident between the application level and the 629 transmission system :

1. Delay at the source node - This may be expressed as the time between the application software declaring the message as 'produced', *eg.* via a call to an APEX function, and the local 629 terminal accessing the message in its area of memory shared with the host run-time executive in preparation for transmission.

   This can be broken down into the time taken to process the APEX function call by the run-time executive (A_P, which is dependant upon scheduling behaviour at the APEX level), **plus** the time that the run-time executive is blocked (on write) by the 629 terminal due to contention over the shared memory (B_P, which is dependant upon the actual shared memory access mechanisms adopted), **plus** the time spent waiting for retrieval from the shared memory in preparation for transmission by the 629 terminal (M_P, which is dependant upon scheduling behaviour at the 629 level).

[6] In practice, additional delays will be incurred due to backplane communication (*eg.* ARINC 659) - these are not addressed in this paper.

The main factors which influence M_P are the relationship between bus control parameters *TI*, *SG* and $\{TG_i, Tx_i\}$, *ie.* whether the bus is operating in periodic or aperiodic mode, and whether or not all messages are transmitted by the 629 terminal every time bus access is granted.

In the periodic mode, worst case M_P is at least *TI* (greater for messages which are not guaranteed transmission on every bus access). In the aperiodic mode, an additional factor is introduced into the worst case M_P calculation - this is the time beyond *TI* for which sporadic transmissions delay the end of the minor cycle, which must take account of all possible sporadics occurring at worst case.

2. Delay in transmission across the 629 bus **-** This may be expressed as the time between the message being retrieved from shared memory at the source node and being deposited by a remote 629 terminal into the shared memory at the destination node.

   This can be broken down into the time elapsed within the current 629 transmission window for message selection and transport (X_{PC}, which is dependant upon order and duration of transport) **plus** the time that the destination 629 terminal is blocked by the run-time executive due to contention over the shared memory (M_C, which is dependant upon the actual shared memory access mechanisms adopted).

   A worst case value for X_{PC} equates to the worst case duration of the local 629 terminal's periodic transmission (Tx_i, with worst case TX_i).

3. Delay at the destination node - This may be expressed as the time between the message being deposited by a remote 629 terminal into the shared memory at the destination node and the 'consumer' application being permitted to read the message, *eg.* via an APEX function call.

   This can be broken down into the time that the run-time executive is blocked (on read) by the 629 terminal due to shared memory contention (B_C, which is dependant upon the actual shared memory access mechanisms adopted), **plus** the time taken to process the APEX function call by the run-time executive (A_C, which is dependant upon scheduling behaviour at the APEX level).

This gives us an expression to calculate the duration of an end-to-end message delay, E^{BP}, under the basic protocol :

$$E^{BP} = A_P + B_P + M_P + X_{PC} + M_C + B_C + A_C \qquad ...(5)$$

Figure 4 illustrates this decomposition of end-to-end message delay into intermediate delays. Note that the individual delays in the diagram are not relatively scaled - in practice, service delays will dominate and shared memory access delays will be relatively small.

5.2 Combined Protocol

For the 629 combined protocol **level one and two** transmissions, the intermediate delays which contribute to the overall end-to-end message delay are the same as those described above for the basic protocol, *ie.* equation (5) also applies to CP. In terms of evaluating each of the intermediate delays, however, there are some minor differences :

1. Delay at the source node - There is no allowance required for minor cycle over-run (a characteristic of BP aperiodic mode), since this is precluded in CP. Hence, the delay M_P is equivalent to that of the periodic mode in BP. Otherwise, A_P and B_P are the same as in BP.

   Assuming that worst case sporadic behaviour is taken into account in planning the level two transmission sequence, worst case M_P is at least *TI* (greater for messages which are not guaranteed transmission on every bus access).

2. Delay in transmission across the 629 bus -Same as BP.

3. Delay at the destination node - Same as BP.

For CP **level three** transmissions, the delay is theoretically unbounded since there is no guarantee at the 629 level of any minimum time per minor cycle (nor across a number of cycles) in which level three transmissions will be serviced. This implies a problem defining worst case M_P.

6. Conclusions

ARINC 629 defines a multi-level protocol for multi-terminal access to a common databus.

There is a similarity, to a limited extent, between the ARINC 629 bus control mechanisms employed at the data link level and the TDMA approach, due to the partitioning of bus time into minor and major cycles and emphasis on the design-time assignment of bus control parameters. This is attractive from a certification viewpoint since specific implementations of the TDMA approach have been certified many times before and, indeed, an implementation of 629 itself has already been certified[7] for the Boeing 777.

There are also similarities with the CSMA approach due to emphasis on run-time detection of bus quiet prior to transmission within a minor cycle. This approach, along with support for variable length transmissions, most significantly influences the degree to which behavioural guarantees can be provided.

In general, run-time transmission ordering is deterministic, assuming a suitable pre-assignment of bus control parameters and, for certain protocol configurations, a controlled timer initialisation sequence.

Regardless of the protocol configuration adopted, worst case delays are different between each terminal pairing, due to the influence of terminal-specific parameters *TG* and *Tx*. Also, the worst case delay between a terminal pair in one direction will differ from that in the other direction.

The provision of true real-time behaviour depends very much on the manner by which the protocol is applied. In this vein, the following conclusions have arisen about the timing behaviour under different configurations :

- The basic protocol will only deliver true real-time behaviour for systems which solely require periodic, fixed length message transmissions. Variable length periodic messages or any sporadic activity will result in variable minor cycle durations.
- The combined protocol offers more stable periodic response than the basic protocol for systems which require some combination of periodic and sporadic message transmissions, since minor cycle duration is fixed and there is no potential for sporadic messages to interfere with periodics.
- The extent to which required sporadic behaviour can be guaranteed under the combined protocol depends on the degree of acceptable under-utilisation of the databus[8].

Other, more general conclusions are as follows :

- A maximum message length is imposed by ARINC 629 but not by 653 (the application software / run-time executive interface, APEX) and so, if 629 is used in conjunction with some APEX-compliant run-time executive, then message partitioning must be provided as part of the executive function.
- In practice, service delays will form the most significant part of overall end-to-end message delays.

[7] The Boeing 777 employs physically separate ARINC 629 buses for flight controls and general avionics functions. The 629 bus is certified as a separate sub-system itself within the flight control system.
[8] TDMA-like protocols will always suffer from this problem.

7. References

[1] Airlines Electronic Engineering Committee, “ARINC Specification 629-3 : IMA Multi-transmitter Databus,” 1994.

[2] Berger, SJ., “ARINC 629 Digital Communication System Application on the 777 and Beyond,” *Proceedings of ERA Avionics Conference and Exhibition*, 1995.

[3] Day, JD., Zimmermann, H., “The OSI Reference Model,” *Proceedings of the IEEE*, vol.71, 1983.

[4] Airlines Electronic Engineering Committee, “ARINC 653 : Avionics Application Software Standard Interface,” draft 14, 1995.

[5] Kurose, JF., Schwartz, M., Yemini, Y., "Multi-access Protocols and Time-constrained Communication," *ACM Computing Surveys*, vol.16 no.1, 1984.

8. Figures

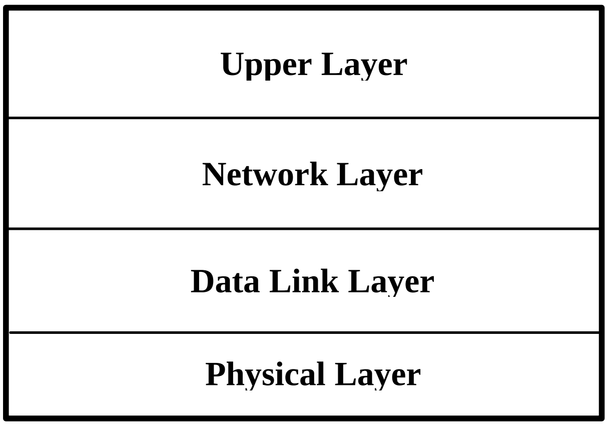

Figure 1 - Protocol Layers

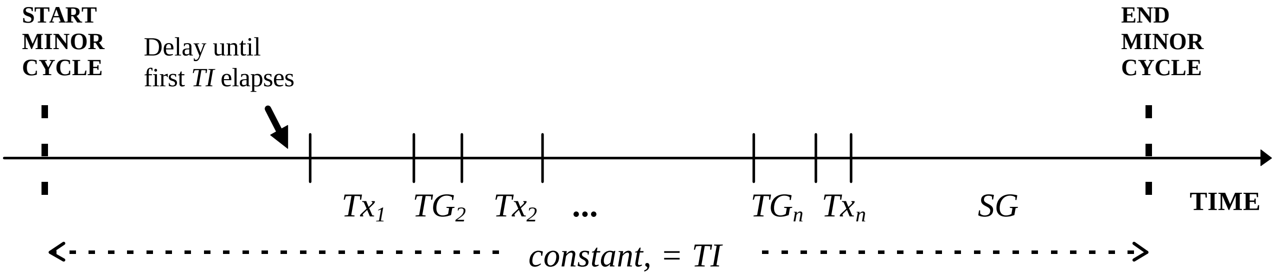

Figure 2a - Data Link Layer Basic Protocol (Minor Cycle in Periodic Mode)

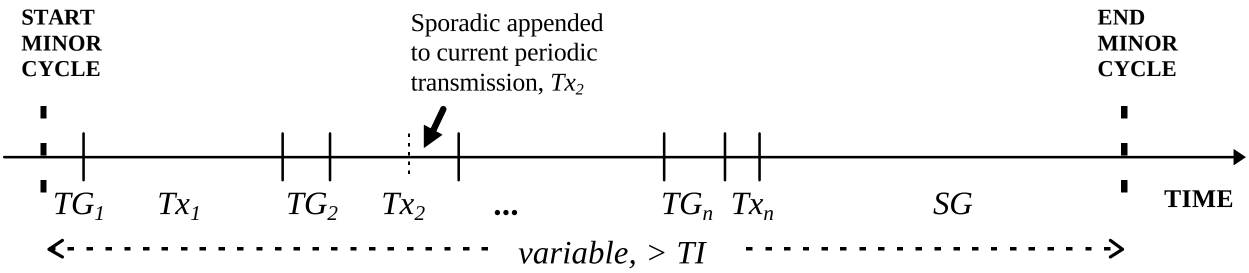

Figure 2b - Data Link Layer Basic Protocol (Minor Cycle in Aperiodic Mode)

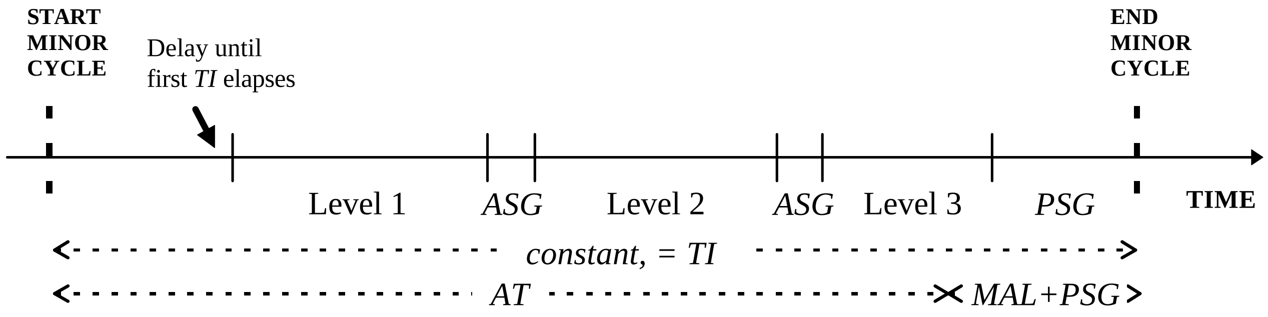

Figure 3 - Data Link Layer Combined Protocol (Minor Cycle)

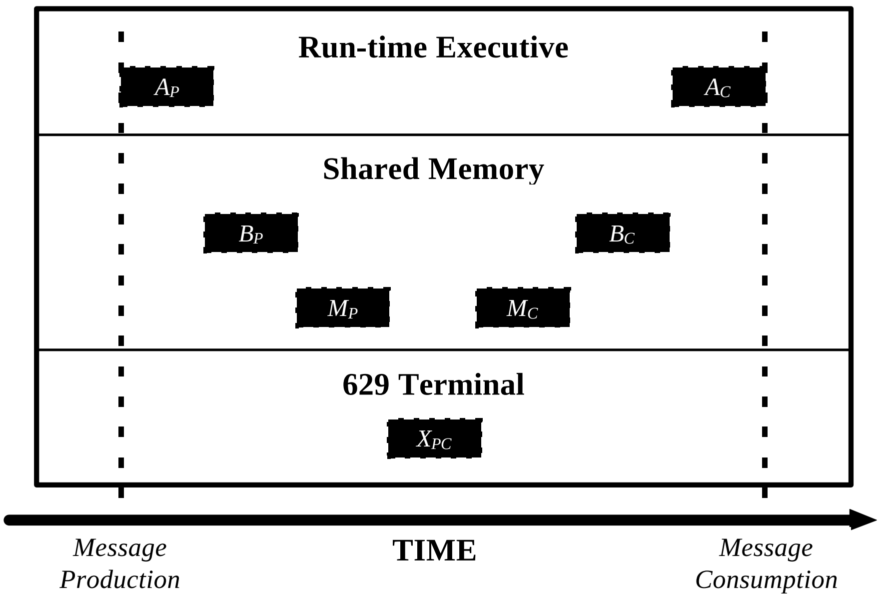

Figure 4 - Decomposition of End-to-end Message Delay