
Sujet : Délai de bout-en-bout dans les réseaux commutés

Il est rappelé que le jury attend un exposé de 35 minutes, pédagogique et structuré, fondé sur ce sujet. Ce texte se conclut par une liste de pistes de réflexion pour guider la conception de l'exposé, mais toute initiative personnelle pertinente est appréciée. L'exposé doit contenir une ou plusieurs illustrations informatiques ainsi qu'une discussion autour d'une dimension éthique, sociétale, environnementale, économique ou juridique. Une piste particulière est proposée à cet effet.

1 Introduction

Ce sujet s'intéresse à la modélisation du délai de transmission de trames dans un réseau Ethernet commuté. Les réseaux Ethernet commutés sont utilisés pour interconnecter des ordinateurs (ou stations) de nos jours. Comme décrit plus loin, ils permettent, via des commutateurs, l'échange de données via des câbles Ethernet. Leur principal avantage est d'offrir un débit de transmission garanti sur le câble qui relie la station au commutateur. Plus spécifiquement, le sujet de cette épreuve traite du calcul de ce délai dans le but d'obtenir la distribution des délais observés pour une séquence de trames transmise par une station vers son destinataire. La modélisation proposée est exploitée principalement dans le domaine des systèmes embarqués critiques du monde automobile ou aéronautique. Ces systèmes se composent de calculateurs répartis qui sont interconnectés par un réseau Ethernet commuté. Par exemple, sur une automobile, le calculateur qui gère le moteur est relié à des calculateurs qui gèrent les fonctions de freinage avancées ou des fonctions qui opèrent les équipements de l'habitacle. Ainsi, les applications sont distribuées sur les calculateurs et communiquent entre elles ou vers des actionneurs ou des capteurs dans le but de contrôler le système global. Ces systèmes sont dits critiques car leur dysfonctionnement peut causer des pertes humaines, ce qui est notamment le cas s'ils ne réagissent pas en temps réel. Dans la suite, le terme *station* est utilisé pour désigner un calculateur ou un ordinateur relié au réseau.

2 Ethernet commuté

On s'intéresse dans ce sujet aux réseaux Ethernet commutés. Ces réseaux sont déployés localement pour interconnecter des stations avec des liens filaires communément appelés câbles Ethernet. Ils présentent des débits de transmission qui varient de 10 Mbit/s à 400 Gbit/s en fonction des technologies de transmission (paire torsadée en cuivre ou fibre optique). Les équipements d'interconnexion sont appelés des commutateurs ou ponts Ethernet (ou switchs en anglais). On supposera que chaque station est directement connectée à un et un seul commutateur. Les ports physiques d'un commutateur sont numérotés. Les commutateurs peuvent être déployés en cascade et former un réseau de commutateurs interconnectant des stations éloignées physiquement tel que représenté sur la Figure 1.

Chaque station possède une file d'attente des trames prêtes à être émises vers leur destination. La file d'attente présente une discipline FIFO (First In First Out). On se place ici dans le contexte des réseaux

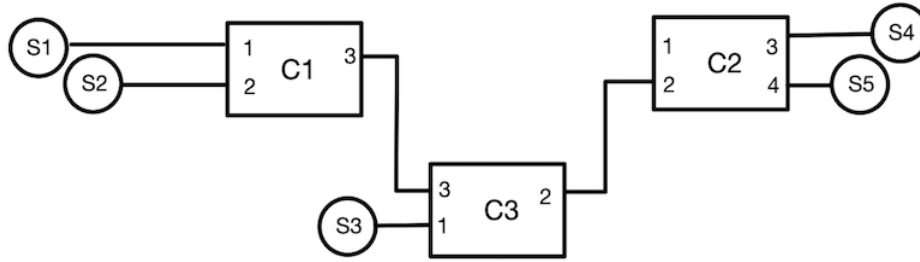


FIGURE 1 – Un exemple de réseau commuté composé de 3 commutateurs $C1$ à $C3$ et 5 stations $S1$ à $S5$.

dits *full-duplex* où chaque sens de transmission (de la station vers le commutateur et réciproquement) possède un support dédié. Un lien full-duplex peut donc être vu comme deux liens unidirectionnels indépendants et de sens opposés. Ainsi, dès qu'une trame est en tête de file d'attente, elle est émise *immédiatement* sur le support.

Un commutateur présente des ports capables de recevoir et d'émettre des trames. Par exemple, sur la Figure 1, on peut voir que $C1$ a 3 ports numérotés de 1 à 3. Dès qu'une trame est reçue par un commutateur, elle est transmise vers le port de sortie en suivant une des règles définies dans sa table de commutation. Cette table associe l'adresse MAC d'une station au numéro de port qui permet de la joindre selon un routage pré-établi. On supposera cette table (et donc les routes) connue dans ce sujet. Le port de sortie comporte une file d'attente avec une discipline de service qui peut être configurée. Cette file d'attente regroupe toutes les trames en attente de transmission sur le port de sortie. La trame en tête de file est transmise immédiatement si la discipline est FIFO. Si la file comporte 2 trames, la seconde doit attendre que la transmission de la première soit terminée avant d'être émise. Comme vu dans la suite, l'émission d'une trame n'est pas instantanée et dépend du débit de transmission et de sa longueur.

3 Le délai de bout-en-bout

On nomme **délai de bout-en-bout** la durée entre l'arrivée de trame dans la file d'attente de la station source et la réception du dernier bit par la station destinataire. Dans ce sujet, on suppose que le réseau fonctionne à un débit de 10 Mbit/s. La taille des trames est exprimée en octets et varie entre 64 et 1500 octets.

3.1 Délai d'une transmission directe

Le délai de transmission directe d'une trame de taille D bits sur un câble (par exemple entre $S1$ et $C1$) est la somme de deux durées :

- la durée d'émission des D bits au débit de R bits/s ;
- la durée de propagation d'un bit sur le câble à une vitesse de 200 000 km/s.

Vue la longueur des câbles (quelques mètres au maximum) dans un tel réseau local, la durée de propagation sera négligée. La durée d'émission varie ainsi entre 51,2 μ s et 1200 μ s pour $R = 10$ Mbit/s.

3.2 Délai de commutation

Le délai de traversée d'un commutateur par une trame émise par exemple par $S1$ est la somme de 2 durées :

- la durée 'technologique' liée au transfert de la trame d'une mémoire tampon sur le port d'entrée vers la file d'attente du port de sortie. Cette durée est constante et de l'ordre de 10 microsecondes.
- la durée d'attente dans la file d'attente du port de sortie.

3.3 Délai de bout-en-bout : cas d'un seul commutateur

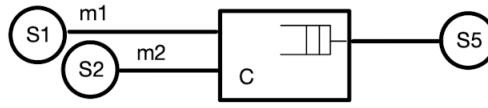


FIGURE 2 – Réseau commuté avec 1 commutateur.

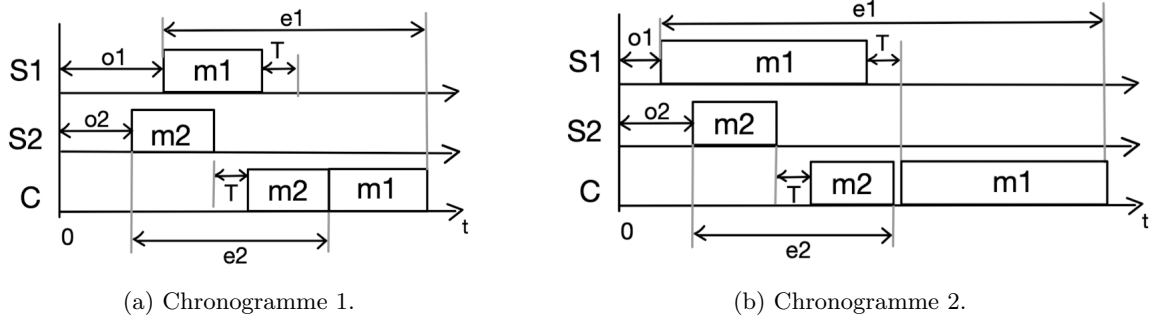


FIGURE 3 – Deux exemples de transmissions concurrentes dans le réseau de la Figure 2.

La Figure 2 décrit un réseau composé d'un seul commutateur, et schématise l'envoi de 2 trames, $m1$ et $m2$ vers la station 5 depuis les stations 1 et 2. Les chronogrammes des Figures 3a et 3b décrivent deux déroulements possibles où les deux messages émis par les stations se chevauchent. On y adopte les notations suivantes pour une trame $m1$:

- d_1 le délai de transmission des $D1$ bits au débit de R bits/s ;
- o_1 la date de demande d'émission (ou offset) de la trame par l'émetteur. Dans ce cas, la trame est prête à être transmise (elle est en tête de la file d'attente de la station) ;
- T la durée technologique de commutation ;
- e_1 le délai de bout-en-bout.

Dans la Figure 3a, la station $S2$ émet en premier au bout de o_2 secondes la trame $m2$. La station $S1$ émet en second au bout de o_1 secondes la trame $m1$ et au bout de son délai de transmission d_1 puis du temps T de commutation dans C , la trame $m1$ se retrouve bloquée dans la file d'attente et doit attendre la fin d'émission de $m2$. Dans la Figure 3b, la trame $m1$ est plus grande que dans le 1er cas, mais elle arrive plus tôt et ne se retrouve pas bloquée.

Ainsi, on observe sur le chronogramme 1 que $m2$ présente le délai minimal de transmission car il est arrivé dans la file d'attente du port de sortie du commutateur en premier. Par contre, $m1$ a dû attendre que la transmission de $m2$ se termine. On dira que $m2$ interfère $m1$. Sur le chronogramme 2, par contre, même si $m2$ arrive en premier dans la file d'attente du port de sortie, il n'interfère pas la transmission de $m1$ qui présente un délai de bout-en-bout minimal.

3.4 Délai de bout-en-bout : cas d'un réseau de commutateurs

Dans une topologie où une trame parcourt plusieurs commutateurs (cf. Figure 1) pour atteindre sa destination, elle peut être interférée par les autres trames qui partagent les ports de sortie de chaque commutateur traversé. Par exemple, une trame émise par $S1$ à destination de $S4$ se verra peut-être interférée par les trames qui partagent le port 3 de $C1$, par les trames qui partagent le port 2 de $C3$ et finalement par les trames qui partagent le port 3 de $C2$.

Le chronogramme 3 illustre une telle situation. On suppose dans cet exemple que les stations $S1$, $S2$, $S3$ et $S5$ émettent une trame $m1$, $m2$, $m3$ et $m5$ vers $S4$. On observe sur ce chronogramme 4 que les trames $m2$, $m3$ et $m5$ ont interféré $m1$, ce qui a engendré un délai de bout-en-bout important. La trame

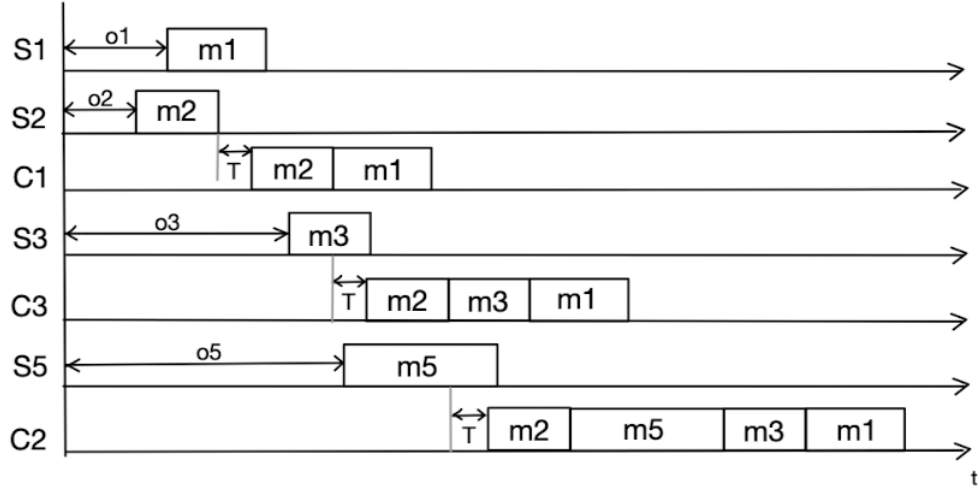


FIGURE 4 – Chronogramme 3.

$m2$ par contre n'a pas été interférée et présente le délai de bout-en-bout le plus faible possible. Les trames $m3$ et $m5$ ont été interférées par $m2$ uniquement.

4 Distribution des délais de bout-en-bout

Ce sujet se positionne dans le contexte des réseaux embarqués critiques où les stations émettrices sont des machines (et non des êtres humains!) : des calculateurs embarqués, des capteurs, des actuateurs, etc. Les calculateurs hébergent des applications critiques qui communiquent avec les autres éléments du réseau. Leur demande de trafic est généralement pré-établie en phase de conception du système. Ainsi, les concepteurs de réseaux embarqués peuvent s'appuyer sur ces hypothèses pour modéliser la performance des transmissions. Un des critères de performance clé dans ces systèmes est le délai de bout-en-bout des messages critiques dans le but de vérifier que le système présente des garanties temporelles temps réel. Les hypothèses énoncées dans la suite sont ainsi principalement liées au contexte des réseaux embarqués critiques.

4.1 Flux périodique de trames

On suppose qu'une station transmet une série de trames périodiquement vers une ou plusieurs destinations. Cette série de trames est caractérisée par sa période P et les tailles minimale et maximale des trames exprimées en octets.

On désigne communément cette série de trames avec le terme de **flux périodique** que l'on dénote f_i où i désigne l'indice de la station émettrice, P_i sa période, $DMIN_i$ et $DMAX_i$ les bornes sur la taille D_i des trames émises. Dans la réalité, une station peut émettre plusieurs flux, mais pour simplifier le propos, on se borne ici à ne présenter que le cas où une station n'émet qu'un seul flux. Aussi, on a $64 \leq DMIN_i \leq DMAX_i \leq 1500$ octets.

Dans les réseaux avioniques, les périodes des flux sont choisies dans l'ensemble des valeurs $\{2, 4, 8, 16, 32, 64, 128\}$ exprimées en millisecondes.

Un flux emprunte sur le réseau une route (possiblement multicast) qui relie la station émettrice à la (ou les) station(s) destinataire(s). Une route multicast est représentée par un arbre couvrant minimal en nombre de sauts (i.e. le coût d'un arc vaut 1) enraciné à la station émettrice dont les feuilles sont les

destinations du flux. Le commutation qui se situe à un branchement de l'arbre va créer autant de copies que de fils et les placer dans les ports de sortie qui leur sont associés.

Dans les réseaux commutés embarqués, les propriétés des flux (et donc leurs routes) sont définies à la conception du système et ne changent pas quand le système est en opération. Dans les systèmes avioniques, il n'y a pas de synchronisation globale des calculateurs. Ainsi, au démarrage du système complet, les applications démarrent à leur rythme avant de commencer à émettre leurs trames. On modélise ce démarrage par un offset temporel o_i qui décale le début du flux périodique f_i . La Figure 5 présente schématiquement les dates de demande d'émission de deux flux périodiques qui présentent des dates de démarrage et des périodes différentes.

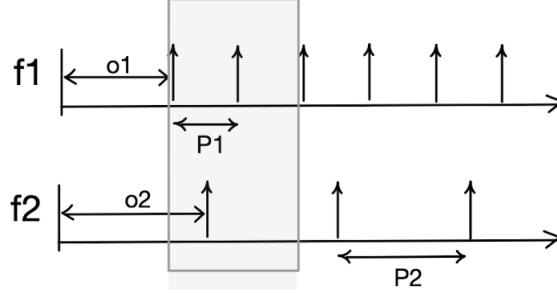


FIGURE 5 – Dates de demande d'émission de deux flux périodiques. Ici $P_2 = 2P_1$.

4.2 Distribution des délais de bout-en-bout

Pour un routage donné, les trames d'un même flux f_i présentent des délais de bout-en-bout variables. Il y a plusieurs origines possibles :

- La taille de la trame émise (entre $DMIN_i$ et $DMAX_i$) à chaque occurrence du flux ;
- Le nombre et la taille des trames émises par les autres flux qui partagent les ports de sortie des commutateurs traversés par le flux (les trames interférentes) ;
- La variation de l'offset de démarrage des flux si les stations démarrent de façon asynchrone (i.e. non concomitante).

On notera que si un flux présente une route multicast, on doit établir le délai de bout-en-bout pour les chemins entre la source et chaque destination.

Calcul du délai de bout-en-bout. Pour construire la distribution des délais de bout-en-bout d'un flux donné, il convient de pouvoir d'abord déterminer le délai de bout-en-bout pour une réalisation donnée telle que présentée par exemple sur le chronogramme 3 de la Figure 4. Dans cette réalisation, la taille des trames et leur date de demande d'émission sont connues.

L'exemple de la Figure 4 suppose que tous les flux ont la même période P et qu'elle est suffisamment longue pour permettre à toutes les trames d'arriver à leur destination avant la fin de la période. Ainsi, à chaque période P , des trames $m1$, $m2$, $m3$ et $m5$ sont prêtes aux mêmes dates (modulo P) que celles indiquées sur le chronogramme.

Pour déterminer le délai de bout-en-bout d'une trame pour une réalisation donnée, il est nécessaire d'établir, pour chaque commutateur traversé, la date de fin de transmission de cette trame sur son port de sortie. Par exemple, sur le chronogramme 3, la date de fin de transmission de m_1 sur le port 3 de $C1$ vaut $t_1^f(C1) = d_1 + 2d_2 + T + o_2$.

Connaissant les dates de début d'émission de $m1$ et $m2$ par $C1$ et celle de $m3$ par $S3$, on peut déterminer un ordre d'émission des 3 trames dans la FIFO du port de sortie de $C3$. On peut alors en déduire une récurrence qui détermine $t_1^f(C3)$ à partir de $t_1^f(C1)$, T , d_1 et des dates de fin d'émission des trames qui précèdent $m1$ dans la FIFO de $C3$.

Calcul de la distribution des délais de bout-en-bout. La distribution (statistique) des délais représente un tableau qui associe des classes de valeurs obtenues avec le nombre d'apparitions des délais observés (ici calculés) par classe. Les classes de valeur sont ici des intervalles (de taille égale) de délais observables. On peut aussi associer la fréquence d'apparition aux classes. La fréquence est définie par le ratio entre le nombre d'apparitions et le nombre total de délais observés, et s'exprime par un pourcentage ou par un réel entre 0 et 1.

Pour un réseau et une définition des flux donnée, on peut obtenir la distribution des délais de bout-en-bout pour un flux en réalisant une recherche exhaustive sur les configurations possibles en faisant varier :

- la taille des trames émises par le flux,
- l'offset du démarrage des flux.

Si les flux présentent des périodes d'émission hétérogènes, la séquence des demandes d'émission des flux suivra un motif périodique de période égale au plus petit commun multiple des périodes *PPCM*. Sur l'exemple de la Figure 5, le motif périodique est de longueur P_2 et la séquence des 3 demandes d'émission que l'on doit étudier pour construire la distribution est illustrée par l'encadrement grisé. Ainsi, pour chaque configuration, on devra calculer :

- le délai de bout-en-bout pour la trame de f_2 et établir sa distribution en variant les configurations,
- le délai de bout-en-bout pour les deux trames de f_1 , et les cumuler dans une même distribution en variant les configurations.

En terme d'offset, on pourra aussi simplifier la modélisation en observant que la variation relative des offsets des flux provoque une inversion dans l'ordre des trames qui arrivent dans les files d'attente des commutateurs traversés.

5 Délai pire-cas et borne

Dans les systèmes critiques, le bon fonctionnement des applications distribuées sur les calculateurs est assuré par l'établissement, à priori, du délai de bout-en-bout maximal que chaque flux peut subir. La valeur maximale du délai de bout-en-bout est généralement désignée par le terme **délai pire-cas**. La certification du bon fonctionnement des traitements exploite directement la connaissance des délais pire-cas des flux déployés dans le système.

En d'autres termes, si, lors de l'exécution du système critique, une trame présente une valeur supérieure au pire-cas déterminé par calcul, l'application destinataire ne peut terminer son traitement avec les données nécessaires ou dans les temps impartis, ce qui risque de provoquer un dysfonctionnement général du système. Ainsi, il est nécessaire soit de pouvoir déterminer avec certitude le pire-cas exact (et non une valeur approchée) ou de construire une borne supérieure sur le délai pire-cas (avec des méthodes qui ne sont pas abordées dans ce sujet).

5.1 Délai pire-cas

La recherche du délai pire-cas exact peut se déduire de la distribution des délais déterminée précédemment par recherche exhaustive. Néanmoins, dans un réseau avionique réel, une telle approche ne passe pas à l'échelle. Les plus gros réseaux interconnectent plus d'une centaine de calculateurs (et donc autant d'émetteurs), 8 commutateurs et transportent un millier de flux périodiques qui se décomposent en plus de 6000 routes.

Il est possible de simplifier la recherche du pire-cas exact en restreignant la recherche aux configurations où la taille des trames est la taille maximale tolérée. Néanmoins, on ne peut supprimer les possibles offsets au démarrage des flux qui peuvent varier d'un lancement à l'autre du système.

5.2 Délai pire-cas approché

Il existe des heuristiques qui permettent de déterminer des configurations défavorables en faisant des hypothèses sur la présence de trames interférentes dans la FIFO des ports de sortie des commutateurs empruntés par la route du flux f_x que l'on étudie. On peut, par exemple, construire un scénario où les trames interférentes sont arrivées dans les FIFO avant f_x en étant :

- soit placées par ordre décroissant de taille,
- soit placées par ordre croissant du nombre de ports partagés avec f_x sur sa route.

Ces deux hypothèses peuvent être exploitées par une heuristique de construction d'une configuration défavorable. On notera qu'on cherche principalement à déterminer le délai de bout-en-bout de f_x sans forcément exhiber les dates d'émission des trames qui y ont mené.

Pour illustrer ce propos, on suppose qu'on recherche le délai pire-cas approché du flux f_1 émis par $S1$ vers $S4$ sur la Figure 1. La station $S2$ émet aussi un flux f_2 vers $S4$ qui partage ainsi 3 ports avec f_1 . On a également $S5$ qui émet f_5 vers $S4$ qui partage 1 port avec f_1 . On suppose que tous les flux ont une même période. La taille de trame maximale des flux interférents vérifie $D_2 \leq D_5$. Dans la file d'attente du port 3 du commutateur $C2$, la première hypothèse positionnera la trame de f_5 avant la trame de f_2 dans la file d'attente tandis que la seconde positionnera la trame de f_2 avant la trame de f_5 . Dans la file d'attente du port 3, la trame de f_2 se trouvera aussi devant la trame de f_1 .

6 Pistes de réflexion pour l'exposé

1. Proposer un algorithme qui calcule le délai de bout-en-bout d'une trame qui traverse un unique commutateur en fonction des caractéristiques des trames émises par les autres stations du réseau. On pourra s'appuyer sur l'exemple du sujet de la Figure 2 et des chronogrammes 3a, 3b. Illustrer avec cet algorithme comment plusieurs trames peuvent interférer le délai de bout-en-bout d'une trame donnée.
2. A partir de ce calcul de délai de bout-en-bout, proposer une implémentation qui calcule la distribution des délais pour une des trames en faisant varier soit la taille des trames, soit les offsets des trames ou les deux conjointement.
3. Etendre les deux études précédentes à une topologie qui présente plusieurs commutateurs. On pourra alors mettre en exergue l'influence sur le délai de bout-en-bout du nombre de ports de sortie partagés par la trame étudiée avec les autres trames sur son chemin. Il est conseillé de traiter le cas de la Figure 1 et du chronogramme 4 avec toutes les trames qui convergent vers une même destination.
4. Etendre l'étude du délai de bout-en-bout d'une trame à un flux périodique connaissant les caractéristiques de tous les flux du réseau et pour une topologie donnée.
5. Proposer un algorithme qui détermine un délai pire-cas approché selon une des heuristiques proposées dans ce sujet, soit pour une séquence de trames donnée, soit pour des flux. Comparer les performances d'une recherche du délai de bout-en-bout pire-cas exact et la recherche d'une configuration qui exploite une heuristique de calcul de délai pire-cas approché.
6. Le réseau Ethernet commuté de ce sujet se différencie des réseaux Ethernet utilisés dans un cadre non-critique. On peut souligner que les tables de commutations sont définies statiquement, ce qui n'est pas le cas des réseaux Ethernet commutés communément utilisés par le grand public. Les routes doivent donc être déterminées dynamiquement avec un algorithme de routage. Discuter de l'intérêt de déployer les algorithmes de routage que vous connaissez dans un tel contexte.
7. Discuter de l'impact du choix des routes des trames (ou des flux) sur la distribution du délai.
8. Discuter de l'importance de la justesse des modèles utilisés pour déterminer une borne supérieure sur le délai pire-cas quand le réseau est embarqué dans un système critique temps réel de type automobile, bus, métro automatique ou avionique.