Rapport INRIA 1994 — Programme 1

Systèmes informatiques répartis temps réel tolérant les fautes

PROJET REFLECS

3 mai 1995

PROJET REFLECS

Systèmes informatiques répartis temps réel tolérant les fautes

Localisation : *Rocquencourt*

Mots-clés : accès multiple (1), algorithme d'ordonnancement en-ligne (1), algorithme de contrôle de concomitance (1), algorithme réparti (1), application critique (1), architecture répartie (1), consensus réparti (1), évaluation de performance (1), modélisation analytique (1), parallélisme asynchrone (1), preuve de conception correcte (1), protocole de communication (1), raisonnement d'adversité (1), réseau local câblé (1), réseau local radio (1), simulation (1), système complexe (1), temps réel (1), tolérance aux fautes (1).

1 Composition de l'équipe

Responsable scientifique

Gérard LE LANN, Directeur de Recherche

Secrétariat

Dominique POULICET

Personnel INRIA

Philippe JACQUET, partagé avec le projet ALGO - Chargé de Recherche
Pascale MINET, Chargée de Recherche
Paul MUHLETHALER, Chargé de Recherche

Professeur invité

Sam TOUEG, Cornell University

Ingénieur expert

Nicolas RIVIERRE

Chercheur post-doctorant

Emmanuelle ANCEAUME, Cornell University

Chercheurs doctorants

Bruno di GENNARO, Université Paris 6
Laurent GEORGE, Université Versailles-Saint-Quentin, boursier INRIA
Jean-François HERMANT, Université Paris 6

Collaborateur Extérieur

Bernadette Charron-Bost, CNRS/ LIX, Palaiseau

Ingénieur industrie

Olivier HULO, Thomson-CSF/SCTF

Stagiaires

Sonia METTALI, ENSI, Tunis

2 Présentation du projet

Les travaux du projet Reflecs concernent les systèmes informatiques complexes servant à mettre en œuvre des applications critiques c'est-à-dire des applications dont les défaillances sont inacceptables ou catastrophiques. Ces travaux portent sur les trois thèmes "répartition", "temps réel" et "tolérance aux fautes", pour lesquels sont élaborées des solutions (algorithmes, protocoles) et des preuves de propriétés ainsi qu'une méthodologie de conception et de dimensionnement exploitant ces preuves.

Les propriétés logiques étudiées sont la sûreté logique ("safety"), la vivacité ("liveness"), la ponctualité ("timeliness") et la confiance ("dependability"). Les propriétés physiques étudiées sont diverses mesures de ponctualité (temps de réponse, gigues) et de confiance (disponibilité, fiabilité).

Les domaines couverts sont définis comme suit :

- Répartition : un système est dit réparti si les fonctions qu'il assure sont obtenues par le biais d'algorithmes explicitement conçus pour être exécutés en parallèle, de façon asynchrone, par plusieurs processeurs, sans connaissance de l'état global du système (propriétés de sûreté et vivacité).
- Temps réel : un système est dit temps réel si ses spécifications comportent des références directes ou indirectes au temps physique et si son comportement est déterminé par des algorithmes d'ordonnancement utilisant directement ou indirectement des attributs temporels dérivés des spécifications (propriétés de ponctualité).
- Tolérance aux fautes : un système est dit tolérant aux fautes si son comportement reste conforme à ses spécifications malgré la présence d'états erronés (logiciel, matériel, données) et malgré l'occurrence de défaillances partielles d'origine interne ou dues à l'environnement (propriétés de confiance).

Nos préoccupations concernent essentiellement l'algorithmique parallèle asynchrone, tolérante aux fautes et garantissant, sous certaines hypothèses aussi peu restrictives que possible, la terminaison des processus en temps borné. Nous faisons une nette distinction entre les cinq domaines suivants :

- l'identification des solutions algorithmiques et des modèles de systèmes pour un problème informatique donné (par exemple, consensus réparti en modèle à délais bornés inconnus),
- la preuve d'existence de propriétés logiques et physiques, par raisonnements d'adversité le plus souvent (par exemple, expression des bornes supérieures des temps de réponse dans un système temps réel réparti, des bornes inférieures de redondance dans un système réparti devant tolérer les fautes),
- l'évaluation de performances, par modélisation analytique essentiellement (par exemple, expression des délais moyens et des écarts-types des temps de réponse dans un système réparti),
- la spécification formelle des solutions algorithmiques,
- la réalisation vérifiable (en matériel, en logiciel) des solutions algorithmiques.

Nos travaux concernent les trois premiers domaines.

3 Actions de recherche

3.1 Transactionnel Réparti Temps Réel : TR^2

Participants : E. Anceaume, L. George, G. Le Lann, P. Minet

Un système Transactionnel Réparti Temps Réel (noté TR^2) est un système composé de clients et de serveurs, interconnectés par un système de communication. Des requêtes extérieures sont reçues par les clients et doivent être exécutées par les clients et les serveurs. Chaque requête est modélisée par une transaction composée d'actions, une action devant, par hypothèse, être exécutée complètement par un seul serveur. Une transaction est représentée par le graphe acyclique (GA) de ses actions. Les contraintes temporelles sont modélisées par une échéance relative stricte affectée à chaque transaction, l'échéance étant non connue à l'avance.

La difficulté principale posée par un système TR^2 est de trouver les algorithmes grâce auxquels sont satisfaites de manière certaine les propriétés combinées de sérialisabilité (une propriété de sûreté), de ponctualité et de confiance. La sérialisabilité est la propriété d'exécution des transactions équivalente à une exécution sérielle ; la ponctualité est la propriété de terminaison des transactions avant leur échéance ; la confiance est la propriété de tolérer des défaillances simultanées de serveurs.

Les connaissances préalables disponibles lors de la conception d'un système TR^2 se réduisent à la connaissance, pour chaque transaction, de son GA et de la liste des serveurs devant exécuter les actions de cette transaction. Par contre, les lois d'arrivées et les échéances des transactions soumises au système sont inconnues. A la non connaissance du futur s'ajoute le problème de l'absence de vue globale des transactions à exécuter. Les délais d'acheminement entre clients et serveurs sont compris entre deux bornes min et max. Il en est de même pour la durée de chaque transaction. On s'impose d'exprimer les bornes supérieures des temps de service des transactions sous ces hypothèses minimales, ainsi que des conditions de faisabilité. Pour ce faire, on utilise des résultats de la théorie de l'ordonnancement, de l'algorithmique répartie et des techniques classiques en théorie des jeux. Nous avons établi ces bornes, les conditions de faisabilité et les conditions de stationnarité pour une combinaison algorithmique particulière qui repose sur la combinaison d'un consensus inter-serveurs et d'un ordonnancement local de type EDF non

préemptif (Earliest Deadline First). Les valeurs numériques des paramètres entrant dans l'expression de ces bornes ne doivent être fournies que lors du dimensionnement du système. La solution proposée est donc générique.

3.2 Consensus Distribué, Concomitance et Défaillances : CD^2

Participants : E. Anceaume, B. Charron-Bost (LIX), G. Le Lann, P. Minet, S. Toueg

Le problème CD^2 est celui de l'obtention d'une décision unique en temps fini par les processeurs corrects d'un groupe de processeurs qui ont des avis initiaux différents, en présence de défaillances de sémantique donnée. De nombreux résultats d'impossibilité et de possibilité ont été établis au cours des dix dernières années, selon les modèles de systèmes et les modes de défaillances considérés. Cependant, il n'existe pas à ce jour de résultats établissant clairement ce que sont les hiérarchies de modèles, de problèmes et d'algorithmes existants.

Nous avons examiné les modèles partiellement synchrones les plus connus (une borne Delta sur les délais existe mais est inconnue ou bien Delta est connue mais ne devient vraie qu'à partir d'un instant inconnu) et obtenu certains résultats d'équivalence de modèles.

Des raisonnements simples d'adversité ont également permis de montrer que certains algorithmes de diffusion atomique (ceux d'ISIS par exemple) sont bloquants, bien que le modèle asynchrone considéré soit "enrichi" de réveils par rapport au modèle asynchrone classique. D'autre part, sachant que "Diffusion Atomique" et "Consensus" sont des problèmes équivalents, nous avons commencé à examiner certaines expressions possibles du problème d'"Appartenance à un Groupe" ("group membership"). L'expression de ce problème doit être suffisamment "faible" pour n'être pas équivalente à Consensus, ce qui transformerait en tautologies les preuves de Consensus à partir d'une solution à Appartenance à un Groupe.

Enfin, pour le modèle asynchrone, une structuration rigoureuse paraît possible grâce au concept de "détecteur de défaillance" (Chandra et Toueg, 1991), qui permet le découplage entre le modèle de connaissance locale à chaque processeur, le modèle reflétant le type de connaissance

partagée qui est considérée et, enfin, le ou les algorithmes destinés à résoudre les problèmes CD^2.

Une implantation de $\diamond$W, le détecteur de défaillance le plus faible permettant de résoudre Consensus en présence de défaillances de type "arrêt" dans un modèle asynchrone, entreprise initialement à Cornell University, est poursuivie à Rocquencourt par E. Anceaume.

3.3 Défaillances temporelles dans les systèmes répartis temps réel critiques

Participants : G. Le Lann, S. Toueg

L'un des obstacles rencontrés avec les systèmes temps réel critiques est celui de la vérification en-ligne des hypothèses de conception. Il est en effet impossible de prouver que la probabilité de non-respect des hypothèses de charge, de délais élémentaires ou de densités de fautes est "naturellement" négligeable devant, par exemple, 1.10^{-9}/heure (d'autant plus invraisemblable qu'il s'agit obligatoirement d'architectures redondantes contrôlées par des algorithmes répartis). La vérification en-ligne des hypothèses de bornes supérieures sur le nombre de défaillances de type arrêt ou omission ne pose pas de problèmes différents de ceux résolus par les algorithmes de consensus réparti (voir action CD^2). Par contre, la vérification en-ligne des hypothèses de bornes supérieures sur le nombre de défaillances de type ponctualité (dites temporelles) n'a jusqu'à présent fait l'objet que d'un petit nombre de publications.

Que ce soit à la conception (modèle synchrone) ou au dimensionnement physique (modèles partiellement synchrones ou asynchrones), des propriétés temporelles ne peuvent être garanties pour une application temps réel critique qu'à la condition que soient satisfaites certaines hypothèses d'existence de bornes supérieures et inférieures pour des délais de communication et de calcul.

La vérification en-ligne de ces hypothèses s'effectue par le biais de tests sur les délais entre processeurs. Nous avons montré que la classe la plus "populaire" de ces tests, basée sur des horloges synchronisées et la datation des émissions et des réceptions, ne peut permettre de satisfaire la double contrainte suivante :

- tous les messages à délais corrects sont acceptés

- tous les messages à délais incorrects (défaillances temporelles) sont détectés et éliminés (transformées en défaillances de type omission).

Ces tests ne peuvent être "déterministes", contrairement à ce qui est affirmé dans la littérature récente.

D'autres travaux portent sur une autre classe de tests basés sur l'idée qu'il est possible, en-ligne, d'accumuler de la connaissance sur les délais. La relation entre cette notion et le concept de "borne supérieure qui finit par exister" (Delta finit par être vraie) de certains modèles partiellement synchrones est en cours d'examen.

3.4 Ordonnancement temps réel non préemptif

Participants : L. George, P. Mühlethaler, N. Rivierre

Des travaux précédents ont permis de montrer que, dans la classe des ordonnancements non oisifs, l'algorithme "à échéance la plus proche en premier" (EDF) non-préemptif est optimal. Basés sur cette optimalité, des critères de faisabilité ont été développés pour l'ordonnancement de tâches périodiques. On montre, comme dans le cas d'ordonnancement préemptif, qu'un test nécessaire et suffisant porte sur un intervalle de longueur deux fois égale à celle du ppcm (plus petit commun multiple) des périodes des tâches périodiques considérées. Un test pseudo polynomial de faisabilité peut être dérivé dans le cas où la charge est strictement inférieure à 1. Pour des tâches non concrètes (c'est-à-dire lorsque les temps d'activation des tâches sont inconnus), un test suffisant de faisabilité a également été proposé.

Dans la classe des ordonnancements oisifs, un algorithme basé sur des optimalités partielles de l'algorithme EDF a été développé. Sa complexité est évaluée en détail et l'importance des scénarios d'arrivée est mise en évidence. Des résultats de simulation montrent l'intérêt de l'algorithme proposé.

3.5 Ordonnancement réparti temps réel non préemptif

Participants : J.F. Hermant, G. Le Lann, N. Rivierre

Des travaux précédents ont permis de montrer l'existence de solutions déterministes au problème du contrôle des accès multiples à un canal de

communication à diffusion pour lequel les messages doivent être transmis avant des échéances strictes mais non connues à l'avance. Ce problème a été montré équivalent au problème de l'ordonnancement de tâches sur un monoprocesseur, à ceci près que les ordonnanceurs (les points d'accès au canal) ont une connaissance strictement locale des messages en attente d'ordonnancement (la ressource est accédée de façon répartie) et que la ressource (le canal) n'est pas réquisitionnable. Les algorithmes déterministes proposés sont :

- CSMA-DCR, la variante Ethernet à tri binaire
- DOD/CSMA-CD, qui est CSMA-DCR couplé à un algorithme d'ordonnancement par échéance la plus proche en premier (EDF).

Les bornes supérieures certaines sur les temps de réponse sont obtenues par raisonnements d'adversité, en considérant un adversaire arbitrairement puissant (doté d'une quantité infinie de messages), noté Z_0. Ces bornes, notées $B^0(i,r)$, pour toute source de messages i et pour tout rang en file d'attente r, comprennent en particulier des délais de parcours d'arbres binaires. L'expression exacte de ces délais, plutôt qu'une borne supérieure, a été proposée et vérifiée pour des arbres ayant jusqu'à 2^{12} feuilles, lorsque toutes les feuilles ne pas pas occupées, c'est-à-dire lorsque l'on considère un adversaire Z_1 plus restreint que Z_0. Ce travail nous permettra d'exprimer les bornes $B^1(i,r)$.

Des applications numériques permettent de vérifier que les bornes B^0 sont "bonnes" (typiquement, quelques millisecondes ou dizaines de millisecondes pour des réseaux Ethernet pouvant être "très chargés" (8 Mbits/s par exemple)).

3.6 Routage dans les réseaux locaux radio avec stations mobiles

Participants : P. Jacquet, P. Minet

Dans un réseau local radio de type HIPERLAN (HIgh PErformance Radio Local Area Network), la portée limitée de la radio ne permet généralement pas à une station de communiquer avec toutes les autres stations de l'HIPERLAN. C'est pourquoi le routage est nécessaire. Ce routage doit intervenir au niveau MAC afin d'offrir à l'utilisateur les mêmes services qu'un réseau local filaire. Il doit permettre la coexistence de stations Relais et de stations Non Relais.

Bien que le routage dans un HIPERLAN présente de nombreuses similarités avec le routage effectué par la couche Réseau dans le modèle de référence OSI, il en diffère par la fréquence accrue avec laquelle les changements de topologie interviennent. La solution retenue est basée sur :

- un routage à chaque saut. Cette technique permet en effet d'adapter la route empruntée par un message pour prendre en compte les derniers changements topologiques. Elle permet d'avoir des entêtes de message plus courts.
- une propagation des changements de topologies basée sur la technique d'état des liens. Après un changement de topologie, il suffit d'une diffusion à travers l'HIPERLAN pour que toutes les stations Relais soient informées du changement et mettent à jour localement leurs tables.

Notre contribution a aussi consisté à définir les principes fondamentaux du routage dans un HIPERLAN :

- acquisition de la connaissance du voisinage et vérification de la symétrie de la communication radio entre stations voisines.
- propagation des changements de topologie.
- propagation d'un message multipoint.

Les différentes procédures associées ont été définies pour les stations Relais et les stations Non Relais. Les communications avec une station en dehors de l'HIPERLAN local, que cette station appartienne à un autre HIPERLAN ou à un réseau local filaire IEEE 802, se font via un pont.

3.7 Méthodologie de conception et de dimensionnement prouvables

Participants : G. Le Lann, P. Minet

Les preuves de conception et de dimensionnement corrects sont essentielles dans le cas des applications temps réel critiques complexes. De telles applications ne peuvent être mises en œuvre que sur des systèmes informatiques répartis, temps réel et tolérant les fautes. Sans preuves, il est impossible de montrer qu'un système informatique est une solution correcte aux problèmes posés par une application critique complexe. Ces

preuves ont donc pour but de permettre de démontrer qu'une spécification d'implantation, notée $\langle \Sigma, \sigma \rangle$, satisfait une spécification de besoins applicatifs, notée $\langle G, g \rangle$. La notation $\langle N, n \rangle$ signifie que l'on exige les propriétés N sous hypothèse n. Toute spécification est constituée de deux parties, une partie logique et une partie physique. Ainsi, par exemple, $\langle G, g \rangle \equiv \langle \langle \Lambda, \lambda \rangle, \langle \Phi, \phi \rangle \rangle$. Des propriétés logiques Γ peuvent être obtenues par une conception, notée Δ, qui consiste à donner un modèle de système et un algorithme, sous hypothèse γ. Une conception Δ est prouvée correcte si l'on a : $\Gamma(\Delta) \supseteq \Lambda$ et $\gamma(\Delta) \supseteq \lambda$.

Les preuves de propriétés conférées par un algorithme imposent d'exprimer des fonctions dites caractéristiques, comme des bornes supérieures sur les temps de réponse ou des bornes inférieures sur des degrés de redondance ou sur des probabilités de respect d'hypothèses. Un sous-ensemble des variables entrant dans l'expression de ces fonctions caractéristiques sont dites d'implantation (par exemple, vitesses des processeurs, capacités mémoire, débits des canaux E/S, etc.). Soit ∇ l'opération de quantification des variables d'implantation, appelée dimensionnement. Un dimensionnement (de conception) permet de quantifier les propriétés et les hypothèses, c'est-à-dire d'obtenir des propriétés physiques. Un dimensionnement ∇ est prouvé correct si l'on a :

$\nabla(\Gamma) \supseteq \Phi$ et $\nabla(\gamma) \supseteq \phi$.

Donc, si l'on note $\langle \Sigma, \sigma \rangle$ la spécification de $\{\Delta, \nabla\}$, on prouve que $\langle \Sigma, \sigma \rangle$ implique $\langle G, g \rangle$. La réalisation du système qui implante correctement $\langle \Sigma, \sigma \rangle$ peut alors être entreprise, selon des méthodes de Génie Logiciel, de Génie Electrique, Optique, etc. appropriées.

Cette méthodologie n'impose ni n'interdit aucun formalisme ou langage de programmation particulier.

Nous avons commencé à mettre en œuvre cette méthodologie dans le cadre d'une application d'avionique militaire (contrat DRET en partenariat avec Dassault-Aviation). Cette méthodologie a également été retenue par Thomson-SCTF dans le cadre du partenariat INRIA-Thomson.

4 Actions industrielles

4.1 Projet Esprit III LAURA

Participants : P. Jacquet, P. Minet, P. Mühlethaler, N. Rivierre

Durant cette deuxième année du projet européen LAURA, l'INRIA a terminé sa tâche de spécification du protocole d'accès et du protocole de routage.

Les bases de travail dans ces deux domaines ayant été développées durant la première année du projet, cette deuxième année a été consacrée au suivi des premières phases d'implantation de nos partenaires et à l'approfondissement des principes établis. En parallèle, nous avons assuré la promotion des idées développées au sein du projet LAURA dans le comité de normalisation RES 10 de l'ETSI.

En ce qui concerne le protocole d'accès, la technique "collide and win" qui avait été définie a été affinée en rapport avec les contraintes de la transmission radio. Une optimisation de la technique en fonction des temps de transition émission-réception (TX/RX) et réception-transmission a été développée.

Pour le protocole "divide and conquer", les conditions d'activation du signal de collision ont été précisées en relation avec la partie récepteur radio.

4.2 Normalisation ETSI

Participants : P. Jacquet, P. Minet, P. Mühlethaler, N. Rivierre

Les choix faits dans le cadre du projet LAURA ont été défendus en comité RES10, principalement dans les domaines suivants : accès multiple, technique de routage, adressage pour réseaux de type HIPERLAN.

4.2.1 Accès multiple

En ce qui concerne l'accès multiple, nous avons mis en évidence la faculté du protocole "collide and win" à obtenir des fonctions de priorités hiérarchiquement indépendantes. Cette propriété est indispensable pour gérer des services à temps bornés, comme la voix ou la video. D'autre part, le protocole "collide and win" permet d'obtenir un comportement efficace et stable, contrairement à la technique classique de type back-off.

L'état actuel des choix à l'ETSI est le suivant. Le protocole d'accès comporte deux phases. La première, "priority resolution protocol", utilise des signalisations en "bursts" très proches de ceux utilisés dans le protocole "collide and win". Cette phase sera complètement définie par la norme. La seconde phase, "collision resolution protocol", peut être librement choisie pour autant que la technique utilisée n'est pas prédatrice. Ce statu quo à l'ETSI est favorable car il est compatible avec les choix effectués dans LAURA. Notons que ce statu quo a résulté d'âpres discussions et d'une volonté soutenue de notre part de promouvoir des idées nouvelles face à des techniques anciennes tel le back-off préconisé par AT&T et dont les inconvénients sont bien connus.

4.2.2 Adressage dans les réseaux locaux radio avec stations mobiles

L'utilisateur d'un réseau local radio de type HIPERLAN ne connaît que l'adressage MAC défini dans la norme internationale (voir ISO 10039). Pour des raisons d'efficacité (ex : filtrage de l'information reçue), il peut paraître souhaitable de disposer d'un adressage HIPERLAN. Cet adressage doit rester interne à l'HIPERLAN et donc totalement transparent à l'utilisateur. De plus l'adressage HIPERLAN doit permettre d'utiliser les techniques classiques de pontage/routage. Toute adresse HIPERLAN comprend deux champs :

- l'un appelé HID permettant d'identifier l'HIPERLAN, dont la procédure d'attribution est en cours de définition
- l'autre appelé NID permettant d'identifier la station dans l'HIPERLAN. Il s'agit de l'adresse MAC telle que définie dans la norme internationale.

La structure des trames MAC est en cours de définition.

4.2.3 Routage

Les travaux décrits §.3.6 ont été soumis à l'ETSI.

4.3 Algorithmique d'ordonnancement en-ligne pour mémoire de masse embarquée

Participants : L. George, J.F. Hermant, P. Jacquet, G. Le Lann, P. Mühlethaler, N. Rivierre

L'étude réalisée pour Alcatel-Espace a pour objet de déterminer un dimensionnement correct d'une mémoire modulaire embarquée satellite.

Le problème posé, ainsi que ses solutions, ne peuvent être décrits de façon détaillée (clause contractuelle de confidentialité).

Les modules de mémoire jouent un rôle de tampons entre des sources d'information (capteurs) à détails hautement variables et un canal de sortie à débit constant mais partagé par les modules.

Cette étude s'articule en deux phases. Une première phase a consisté à proposer un algorithme d'ordonnancement en-ligne montré équivalent à celui proposé par Alcatel-Espace, mais potentiellement capable de conduire à de meilleurs dimensionnements. Dans une deuxième phase, différentes analyses ont été conduites, de nature stochastique ou déterministe, afin de caractériser exactement les dimensionnements induits par l'algorithme proposé (par exemple, par réseaux de files d'attente, par raisonnements d'adversité). Ce travail sera poursuivi en 1995.

4.4 Projet PHIDIAS

Participant : G. Le Lann

Participation à l'examen conduit par Thomson-SDC de l'appel d'offre PHIDIAS, de la Délégation Générale à l'Aviation Civile (nouveau système de contrôle de trafic aérien). La participation a essentiellement porté sur les choix de conception pour le traitement parallèle asynchrone et l'obtention de très hautes disponibilités pour les services critiques (indisponibilité inférieure à trois secondes par an).

4.5 Génie

Dans le cadre du programme Génie (partenariat Dassault-Aviation/ INRIA), le projet est chargé d'examiner les problèmes de sûreté de fonctionnement et de tolérance aux fautes.

5 Actions nationales et internationales

5.1 Cabernet

Participation au Research Network Cabernet "Distributed Computing System Architectures" (ESPRIT). Contribution au pôle "recherche" (temps réel réparti (G. Le Lann), diffusion atomique (P. Minet)).

5.2 Participation au projet "Construction and performance of real-time transactions"

Ce projet, dirigé par Peter van der Stok et Dicter Hammer de l'Université de Eindhoven, a pour thème les bases de données réparties devant respecter des contraintes temps réel. G. Le Lann et P. Minet interviennent en tant qu'experts invités par le Ministère de l'Industrie (Pays-Bas) qui finance le projet.

5.3 Participation à comités de programme

- CFIP'95, Colloque Francophone sur l'Ingénierie des Protocoles, Rennes, mai 1995 (P. Minet)
- Real-Time Systems, Paris, janvier 1995 (P. Minet)
- Real-Time Systems, Paris, janvier 1994 (G. Le Lann, président du comité)
- 12th IFAC Workshop on Distributed Computer Control Systems, Tolède (Espagne), septembre 1994 (G. Le Lann)
- IEEE Conference on Parallel and Distributed Real-Time Systems, Cancun (Mexique), avril 1994 (G. Le Lann)
- 5th IFIP Working Conference on Dependable Computing for Critical Applications, septembre 1995 (G. Le Lann)

6 Diffusion des résultats

6.1 Actions d'enseignement

- INSTN, DEA d'Informatique, cours réseaux locaux (B. di Genaro, P. Minet)

- Université Paris 6, DESS Téléinformatique, cours réseaux locaux temps réel (B. di Gennaro, P. Minet)
- ENST Paris, 3ème année, cours réseaux locaux temps réel et tolérance aux fautes (G. Le Lann, P. Minet, N. Rivierre)
- SUPELEC Rennes, option MEARI (post-doctorale), cours informatique répartie (G. Le Lann), cours modélisation et planification des réseaux (P. Mühlethaler)
- EURECOM (Sophia Antipolis), cours réseaux locaux radio et réseaux locaux industriels (N. Rivierre)
- EURECOM (Sophia Antipolis), cours introduction aux techniques de modélisation des techniques d'accès (P. Mühlethaler)

6.2 Participation à des manifestations

- Real-Time Systems, Paris, janvier (G. Le Lann, P. Minet)
- Journée DRET "Temps réel et sûreté de fonctionnement dans les applications de la Défense", Paris, mars (G. Le Lann)
- Journée de présentation des travaux du groupe OFTA "Informatique tolérante aux fautes", mars (G. Le Lann, P. Minet)
- 7èmes Journées Internationales des Sciences Informatiques, Tunis, mai (G. Le Lann)
- UIMP Seminar "Real-Time Technologies and Systems", La Coruña (Espagne), juillet (G. Le Lann)
- 13th IFIP Congress, Hambourg (Allemagne), septembre (G. Le Lann)
- Workshop "Unifying Theory and Practice in Distributed Computing", Dagstuhl (Allemagne), septembre (G. Le Lann)
- 3rd International Symposium "Formal Techniques in Real-Time and Fault-Tolerant Systems", Lübeck (Allemagne), septembre, présentation (G. Le Lann), organisation d'une table ronde "Comparative Merits of Synchronous/Partially Synchronous/Asynchronous Models in the Case of Safety/Mission Critical Real-Time Systems" (G. Le Lann)

6.3 Activités extérieures

man

6.3.1 OFTA

Groupe de travail OFTA sur le thème "Informatique tolérante aux fautes" ; ce groupe réuni par l'Observatoire Français des Techniques Avancées avait pour mission d'élaborer une monographie de la série ARAGO, laquelle est publiée par Masson (ISBN 2/225-84522-0) depuis mars (G. Le Lann, P. Minet)

6.3.2 ETSI

Contributions à la normalisation internationale des réseaux locaux radio (P. Jacquet, P. Minet, P. Mühlethaler, N. Rivierre).

6.3.3 Formations internes

Trois séminaires sur les systèmes TRDF (Temps Réel, Traitement Distribué, Tolérance aux Fautes) donnés en interne chez Thomson-SCTF et chez Syseca (G. Le Lann).

6.3.4 Comités de rédaction

Le projet est représenté (G. Le Lann) dans les comités de rédaction (Editorial Boards) des revues "Dependable Computing and Fault-Tolerant Systems" (Springer-Verlag) et "Real-Time Systems" (Kluwer Academic).

7 Publications

Communications à des congrès, colloques, etc.

[1] G. Le_Lann, P. Minet, D. Powell, «Tolérance aux fautes et répartition : concepts et mécanismes», *in : RTS'94*, p. 171–190, Paris, janvier 1994.

[2] G. Le_Lann, N. Rivierre, «Real-Time Communications over Broadcast Networks : the CSMA-DCR and the DOD/CSMA-CD Protocols», *in : RTS'94*, p. 67–84, Paris, janvier 1994.

[3] G. Le_Lann, «Certifiable Critical Complex Computing Systems», *in : IFIP Congress'94*, (North-Holland), p. 287–294, Hamburg, septembre 1994.

[4] G. Le_Lann, «Scheduling in Critical Real-Time Systems : a Manifesto», Springer-Verlag LNCS n° 863, p. 511–528, Lübeck, septembre 1994.

[5] G. Le_Lann, «Systèmes transactionnels répartis temps réel», *in : JISI'94*, p. 1–9, Tunis, mai 1994.

[6] G. Le_Lann, «Why should we keep using precambrian design approaches at the dawn of the 3rd millenium ?», *in : RTS'94*, p. 123–135, Paris, janvier 1994.

Divers

[7] J.-F. Hermant, «Etude de deux protocoles de communications temps réel sur réseaux à diffusion : les protocoles CSMA-DCR et DOD/CSMA-CD», rapport de Stage de DEA, juillet 1994, Institut National des Sciences et Techniques Nucléaires, 100 p.

[8] P. Jacquet, P. Minet, «A flexible IntraForwarding», ETSI/RES10, 1994, RES10/SAG94/24.

[9] P. Jacquet, P. Mühlethaler, N. Rivierre, «The LAURA proposal», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/3.

[10] P. Jacquet, P. Mühlethaler, N. Rivierre, «Medium access control specification for HIPERLAN», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/14.

[11] P. Jacquet, P. Mühlethaler, N. Rivierre, «Proven CSMA/CD», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/11.

[12] P. Jacquet, P. Mühlethaler, N. Rivierre, «Uncertainties and novelties», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/6.

[13] P. Jacquet, P. Mühlethaler, «Analysis of separation power of the comb», Contribution à la normalisation ETSI/RES10, 1993, RES10/SAG/57.

[14] P. Jacquet, P. Mühlethaler, «Analysis of separation power of the comb», Contribution à la normalisation ETSI/RES10, 1993, RES10/SAG/64.

[15] P. Jacquet, P. Mühlethaler, «Collision detection with comb», Contribution à la normalisation ETSI/RES10, 1993, RES10/SAG/58.

[16] P. Jacquet, P. Mühlethaler, «Broadcast transmission in HIPERLAN», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/5.

[17] P. JACQUET, P. MÜHLETHALER, «Comment about contention resolution efficiency of random delay and comb resolution method», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/7.

[18] P. JACQUET, P. MÜHLETHALER, «Consensus contention CSMA protocol : LAURA updated proposal for HIPERLAN», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/51.

[19] P. JACQUET, P. MÜHLETHALER, «HIPERLAN power saving must preserve application throughput», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/20.

[20] P. JACQUET, P. MÜHLETHALER, «L-Ary comb Generalization to various physical constraint», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/62.

[21] P. JACQUET, P. MÜHLETHALER, «Performance and services of the URN protocol», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/8.

[22] P. JACQUET, P. MÜHLETHALER, «Priorities based on discrepancies of back off times», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/8.

[23] P. JACQUET, P. MÜHLETHALER, «Priorities based on discrepancies of interframe spacing», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/21.

[24] P. JACQUET, P. MÜHLETHALER, «Priorities in local area systems», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/4.

[25] P. JACQUET, P. MÜHLETHALER, «Priority functions in HIPERLAN. Consequence on architecture and compatibility», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/4.

[26] P. JACQUET, P. MÜHLETHALER, «Quantitative comparison between the window scheme and comb scheme in a dynamic model», Contribution à la normalisation ETSI/RES10, 1994, RES10/TTG/16.

[27] P. JACQUET, P. MÜHLETHALER, «Simulation results. Comparison of AT & T and LAURA CAM», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/61.

[28] P. JACQUET, P. MÜHLETHALER, «Simulation results. Comparison of AT & T and LAURA updated MAC proposals», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/63.

[29] P. JACQUET, P. MÜHLETHALER, «Simulation results. Effect of parameters variation in simulation models», Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/58.

[30] G. LE_LANN, «Algorithmique d'ordonnancement dédiée aux systèmes temps réel critiques», Journée DRET, mars 1994, Paris.

[31] G. Le_Lann, « Designing Real-Time Fault Tolerant Systems at the Dawn of the Third Millenium », UIMP Seminar "Real-Time Technologies and Systems", juillet 1994, La Coruña (Espagne).

[32] G. Le_Lann, « Real-Time Communications over Broadcast Networks : Protocols, Technology and Performance », UIMP Seminar "Real-Time Technologies and Systems", juillet 1994, La Coruña (Espagne).

[33] P. Minet, « IntraForwarding principles », ETSI/RES10, 1994, RES10/SAG94/1.

[34] N. Rivierre, « HIPERLAN Addressing II », Contribution à la normalisation ETSI/RES10, 1994, RES10/SAG/15.

8 Abstract

Research work conducted within project REFLECS concentrates on those algorithmic issues arising with Distributed Real-Time Fault-Tolerant computing systems.

The major outputs of the project are algorithms, protocols, behavioral proofs, performance proofs and design methodologies for critical and complex computerized applications.

The notion of distribution is equivalent to that of concurrent asynchronous computations in the absence of global system state. Regarding fault-tolerance, we consider crash, omission and timing failures. Under such models, we are interested in proving the existence of specific properties, such as global safety or global liveness, whenever processes conflict with each other at run-time. With respect to real-time issues, the underlying premise of our work is that on-line (or dynamic) decision making raises interesting research challenges and is needed to implement future systems. In other words, we do not assume full a priori knowledge of future run-time conditions.

A significant amount of our work is directed at critical systems. We have developed a technique to establish timeliness proofs in the presence of concurrency and failures. Such proofs, and their companion design methodology, are mandatory for critical systems. We also use analytical modeling to predict the average behavior of such systems.

We are also investing a significant amount of work in the areas of distributed consensus, atomic broadcast and group membership under partially synchronous and asynchronous models.

Our research work is applied to such technologies as wireless local area networks, real-time local area networks, real-time fault-tolerant distributed operating systems and middleware.

Table des matières