(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

COURBEVOIE

(11) **N° de publication :** **3 040 575**

(à n'utiliser que pour les commandes de reproduction)

(21) **N° d'enregistrement national :** **15 58131**

(51) Int Cl[8] : ***H 04 L 12/751*** *(2017.01), H 04 L 12/46*

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 02.09.15.

(30) **Priorité :**

(43) **Date de mise à la disposition du public de la demande :** 03.03.17 Bulletin 17/09.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule*

(60) **Références à d'autres documents nationaux apparentés :**

( ) **Demande(s) d'extension :**

(71) **Demandeur(s) :** *BULL SAS Société par actions simplifiée* — FR.

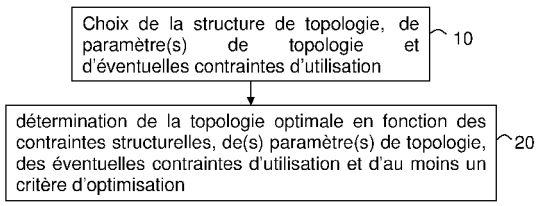
(72) **Inventeur(s) :** VIGNERAS PIERRE et REMY KEVIN.

(73) **Titulaire(s) :** BULL SAS Société par actions simplifiée.

(74) **Mandataire(s) :** CABINET CAMUS LEBKIRI Société à responsabilité limitée.

(54) PROCEDE DE DETERMINATION AUTOMATISEE DE TOPOLOGIES D'INTERCONNEXION OPTIMALES DE SYSTEMES COMPRENANT DES NOEUDS DE SERVICE, ET DISPOSITIF DE TRAITEMENT ASSOCIE.

(57) Un procédé permet de déterminer une topologie d'interconnexion optimale d'un système comprenant des noeuds de service interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau. Ce procédé comprend une première étape (10) dans laquelle on choisit une structure pour une topologie d'interconnexion optimale à déterminer parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, puis on choisit au moins un paramètre de topologie, et une seconde étape (20) dans laquelle on détermine au moyen d'un solveur la topologie d'interconnexion optimale en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi et d'au moins un critère d'optimisation choisi.



FR 3 040 575 - A1

PROCÉDÉ DE DÉTERMINATION AUTOMATISÉE DE TOPOLOGIES D'INTERCONNEXION OPTIMALES DE SYSTÈMES COMPRENANT DES NŒUDS DE SERVICE, ET DISPOSITIF DE TRAITEMENT ASSOCIÉ

L'invention concerne les systèmes comprenant des nœuds de service interconnectés entre eux via des liaisons, et plus précisément la détermination de topologies d'interconnexion de tels systèmes.

On entend ici par « nœud de service » un équipement électronique comprenant des ressources pouvant être de tout type dès lors qu'elles assurent au moins un service configurable utile au système auquel elles appartiennent ou à une application tournant dans ce système. Par conséquent, il pourra par exemple s'agir d'un équipement de calcul comprenant au moins un processeur, ou d'un équipement d'entrée/sortie, ou d'un routeur (ou « switch »), ou encore de moyens de stockage de données.

Par ailleurs, on entend ici par « topologie d'interconnexion » l'architecture (ou le réseau d'interconnexion) définissant l'intégralité des liaisons (ou interconnexions) entre tous les nœuds de service d'un système. Il s'agit d'un graphe non orienté où les sommets représentent, par exemple, des routeurs et les arcs représentent les liaisons (ou câbles). On notera qu'il existe plusieurs structures de topologie, et notamment celles dites hiérarchiques, celles en anneau (à topologie directe), et celles à maillage. Parmi les structures hiérarchiques, on peut notamment citer celles appelées en anglais « Fat tree », « Generalized Fat Tree » (ou GFT), « Extended Generalized Fat Tree » (ou XGFT), et « Parallel Generalized Fat Tree » (ou PGFT). Parmi les structures à maillage, on peut notamment citer celle appelée en anglais « HyperX ».

Compte tenu des deux définitions qui précèdent, un système pourra être notamment un calculateur de type dit « haute performance » (ou supercalculateur), ou un centre de données (ou « data center »).

La détermination (ou le dimensionnement) de la topologie d'interconnexion d'un système, comme par exemple un supercalculateur, est

une tâche complexe en raison des très nombreuses contraintes qui doivent être priseS en compte, et notamment le niveau de performance, le coût, la faisabilité technique, les contraintes logistiques (et notamment l'encombrement), et les contraintes thermiques.

En raison de cette complexité, la détermination (ou le dimensionnement) de la topologie d'interconnexion d'un système est fréquemment réalisée à la main après que la structure de cette topologie d'interconnexion ait été choisie, ce qui induit plusieurs inconvénients. En effet, la topologie d'interconnexion déterminée est rarement optimale sur tous les plans, voire infaisable techniquement, et donc on est contraint de réaliser de nombreuses modifications qui induisent fréquemment un surcoût important et/ou d'éventuels retards de mise à disposition du client et/ou une éventuelle diminution du niveau de performance souhaité initialement.

Il existe un outil informatique, de type « web », développé par la société Mellanox, et destiné à déterminer (ou dimensionner) des topologies d'interconnexion. Cet outil est adapté aux structures de type Fat Tree, et permet de choisir le nombre de nœuds de calcul ainsi que le type de routeur à utiliser sur les deux premiers niveaux de la topologie d'interconnexion, et le facteur de blocage. Le principal inconvénient de cet outil réside dans le fait qu'il ne fournit que le nombre de routeurs et de liaisons nécessaires à la mise en place du système, et donc ne donne aucune information sur la configuration topologique (comme une description formelle, par exemple), ce qui est pourtant indispensable à l'évaluation de la qualité de la topologie en termes de routage et de rapport qualité/prix.

L'invention a donc notamment pour but d'améliorer la situation.

Elle propose notamment à cet effet un procédé, destiné à permettre la détermination d'une topologie d'interconnexion optimale d'un système comprenant des nœuds de service interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau.

Ce procédé se caractérise par le fait qu'il comprend :

- une première étape dans laquelle on choisit une structure pour une

topologie d'interconnexion optimale à déterminer parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, puis on choisit au moins un paramètre de topologie parmi un nombre maximal de ports des routeurs, un nombre de niveaux de routeurs, un nombre de nœuds de service connectés à un routeur appartenant au premier niveau, un nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, un nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, un nombre de liaisons (ou liens) entre un routeur d'un niveau i et un nœud de service d'un niveau i -1, et

- une seconde étape dans laquelle on détermine au moyen d'un solveur la topologie d'interconnexion optimale en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi et d'au moins un critère d'optimisation choisi.

Ainsi, on peut désormais déterminer des topologies d'interconnexion optimales de différentes structures, et faisables techniquement, c'est-à-dire correspondant aux demandes des clients sans nécessiter de nombreuses modifications lors de leur mise en place.

Le procédé selon l'invention peut comporter d'autres caractéristiques qui peuvent être prises séparément ou en combinaison, et notamment :

- chaque critère d'optimisation peut être choisi parmi (au moins) un coût du système, un niveau de performance, une extensibilité du système, une consommation d'énergie du système, des contraintes thermiques, et un encombrement du système ;
- chaque contrainte structurelle peut être choisie parmi (au moins) les inéquations suivantes : $p_i*m_i + p_{i+1}*w_{i+1} \leq radixMax$, $p_h*m_h \leq radixMax$, et $w_1*p_1 + noeudsParFeuille \leq radixMax$, où m_i est le nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, w_i est le nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, p_i est le nombre de liaisons (ou liens) entre un routeur d'un niveau i et un nœud de service d'un niveau i -1, radixMax est le nombre maximal de ports de chaque routeur, h est le nombre de niveaux de routeurs, et noeudsParFeuille est le nombre de nœuds de service connectés à un même routeur appartenant au premier niveau ;

- dans la seconde étape on peut déterminer la topologie d'interconnexion optimale en fonction également d'au moins une contrainte d'utilisation choisie ;
  - chaque contrainte d'utilisation peut être choisie parmi (au moins) une interdiction de connecter à un routeur appartenant à un niveau i donné plus de liaisons (ou liens) vers le niveau supérieur à ce niveau donné que de liaisons (ou liens) vers le niveau inférieur à ce niveau donné, une conservation d'une bande passante en entrée et en sortie de chaque routeur, une non conservation d'une bande passante en entrée et en sortie de chaque routeur, un nombre minimum de nœuds de service, un nombre maximum de liaisons (ou liens) entre deux routeurs ;
- le système peut être choisi parmi (au moins) un supercalculateur et un centre de données (ou « data center »).

L'invention propose également un produit programme d'ordinateur comprenant un jeu d'instructions qui, lorsqu'il est exécuté par des moyens de traitement, est propre à mettre en œuvre un procédé du type de celui présenté ci-avant pour déterminer une topologie d'interconnexion optimale d'un système comprenant des nœuds de service interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau.

L'invention propose également un dispositif de traitement, destiné à déterminer une topologie d'interconnexion optimale d'un système comprenant des nœuds de service interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau, et propre à équiper un équipement électronique comprenant des ressources de calcul et une interface homme/machine.

Ce dispositif de traitement se caractérise par le fait qu'il comprend :

- une interface de programmation agencée pour permettre à un usager de choisir via l'interface homme/machine une structure pour une topologie d'interconnexion optimale à déterminer parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, et au moins un paramètre de topologie choisi parmi un nombre maximal de ports des routeurs, un nombre de niveaux de routeurs, un nombre de nœuds de

service connectés à un routeur appartenant au premier niveau, un nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, un nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, un nombre de liaisons (ou liens) entre un routeur d'un niveau i et un nœud de service d'un niveau i -1, et

- des moyens de résolution agencés pour déterminer la topologie d'interconnexion optimale en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi et d'au moins un critère d'optimisation choisi.

L'invention propose également un équipement électronique comprenant des ressources de calcul, une interface homme/machine, et un dispositif de traitement du type de celui présenté ci-avant.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après, et des dessins annexés, sur lesquels :

- la figure 1 illustre de façon schématique et fonctionnelle un ordinateur équipé d'un exemple de réalisation d'un dispositif de traitement selon l'invention,
- la figure 2 illustre un exemple d'algorithme mettant en œuvre un procédé de détermination selon l'invention, et
- la figure 3 illustre de façon schématique un exemple de topologie d'interconnexion optimisée, déterminé au moyen d'un procédé de détermination selon l'invention avec certains paramètres de topologie.

L'invention a notamment pour objet de proposer un procédé de détermination, et un dispositif de traitement DT associé, destinés à permettre la détermination d'une topologie d'interconnexion optimale d'un système S comprenant des nœuds de service N_{ij}.

Dans ce qui suit, on considère, à titre d'exemple non limitatif, que le système S est un calculateur haute performance (ou supercalculateur). Mais l'invention n'est pas limitée à ce type de système. Elle concerne en effet tout système comprenant des nœuds de service interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau, et notamment les centres de données (ou

data centers).

Comme indiqué plus haut, l'invention propose un procédé destiné à permettre la détermination d'une topologie d'interconnexion optimale d'un système S (ici un supercalculateur) comprenant des nœuds de service N_{ij} interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau.

Ce procédé comprend des première et seconde étapes, référencées respectivement 10 et 20 dans l'exemple d'algorithme de la figure 2, et pouvant être mises en œuvre par un dispositif de traitement DT selon l'invention.

Comme illustré sur la figure 1, un dispositif de traitement DT, selon l'invention, comprend au moins une interface de programmation INP et des moyens de résolution MR.

Par exemple, l'interface de programmation INP peut être de type API (« Application Programming Interface »). A titre d'exemple non limitatif, elle peut être réalisée à partir d'une librairie python, telle que celle appelée Pyomo. Il est rappelé que Pyomo est une librairie python permettant de générer, dans un format particulier (.nl) supporté par beaucoup de programmes de résolution d'équations, des systèmes d'équations préalablement déclarés et instanciés au moyen d'une API.

Dans l'exemple non limitatif illustré sur la figure 1, le dispositif de traitement DT est installé dans un équipement électronique EE agencé sous la forme d'un ordinateur et donc disposant notamment d'une interface homme/machine IH (ici un clavier et un écran d'affichage) et de ressources de calcul RC. Mais cela n'est pas obligatoire. Le dispositif de traitement DT pourrait en effet être un équipement électronique disposant d'une interface homme/machine IH et de ressources de calcul RC. Ainsi, il pourrait s'agir d'un ordinateur portable ou d'une tablette électronique, par exemple. Par ailleurs, dans une variante de réalisation non illustrée, le dispositif de traitement DT pourrait être installé dans un équipement électronique EE agencé sous une autre forme qu'un ordinateur, et notamment sous la forme d'un ordinateur portable ou d'une tablette électronique. Par conséquent, le dispositif de traitement DT peut être réalisé soit sous la forme de modules logiciels (ou informatiques, ou encore « software ») ; on est alors en présence d'un produit

programme d'ordinateur comprenant un jeu d'instructions qui, lorsqu'il est exécuté par des moyens de traitement de type circuits électroniques (ou « hardware »), est propre à mettre en œuvre le procédé de détermination, soit sous la forme d'une combinaison de modules logiciels et de circuits électroniques.

Durant la première étape 10 du procédé selon l'invention, on (un usager de l'équipement électronique EE) commence par choisir, parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, et au moyen de l'interface homme/machine IH, une structure pour une topologie d'interconnexion optimale devant être déterminée pour un système S. L'ensemble de structures est fourni par l'interface de programmation INP du dispositif de traitement DT. Par exemple, cet ensemble peut comprendre au moins une structure de type hiérarchique (comme par exemple Fat tree ou Generalized Fat Tree (ou GFT) ou Extended Generalized Fat Tree (ou XGFT) ou encore Parallel Generalized Fat Tree (ou PGFT)) et/ou au moins une structure en anneau et/ou une structure à maillage (comme par exemple HyperX).

Par exemple, les structures hiérarchiques peuvent être définies chacune par au moins une contrainte structurelle fonction d'au moins deux variables m_i et w_i, et parfois également d'une troisième variable p_i, où l'indice i indexe le niveau auquel appartient un routeur au sein de la structure. m_i est le nombre de nœuds de service N_{i-1j} d'un niveau i-1 connectés à un même routeur N_{ij} d'un niveau i. w_i est le nombre de routeurs N_{ij} d'un niveau i connectés à un même nœud de service N_{i-1j} d'un niveau i-1. p_i est le nombre de liaisons (ou liens) entre un routeur N_{ij} d'un niveau i et un nœud de service N_{i-1j} d'un niveau i -1. Ces variables m, w et p sont des vecteurs de dimension h. Avec le choix fait ci-dessus pour l'indexage, la valeur i = 0 est attribuée au niveau contenant les nœuds de service N_{0j} qui sont connectés à des routeurs N_{1j} (dits feuilles) du premier niveau (i = 1).

Ainsi, un GFT (Generalized Fat Tree) peut être défini par au moins une contrainte structurelle fonction des deux variables m et w (vecteurs de dimension égale à un), ainsi que par h, où $\forall i \in [\![1,h]\!]$, chaque routeur N_{ij} de niveau i est relié à m nœuds de service N_{i-1j} du niveau inférieur i-1, et chaque

routeur N_{i-1j} de niveau i-1 est relié à w routeurs N_{ij} du niveau supérieur i. Il peut donc être noté (h,m,w).

Un XGFT (Extended Generalized Fat Tree) peut être défini par au moins une contrainte structurelle fonction des deux variables m et w (vecteurs de dimension égale à h), ainsi que par h, où $\forall i \in [\![1,h]\!]$, $m_i > 0$ et $w_i > 0$. Il peut donc être noté $(h,[m_1, m_2,\ldots,m_h],[w_1, w_2,\ldots,w_h])$. Par ailleurs, il permet d'avoir des liaisons différentes (nombre de liaisons montantes/descendantes) pour les différents niveaux i de la structure en arbre.

Un PGFT (Parallel Generalized Fat Tree) est un XGFT auquel on ajoute le troisième vecteur p. Il peut donc être noté $(h,[m_1, m_2,\ldots,m_h],[w_1, w_2,\ldots,w_h],[p_1, p_2,\ldots,p_h])$. Il permet d'introduire la notion de redondance de liens entre deux nœuds de service de deux niveaux voisins.

La première étape 10 se poursuit en choisissant au moins un paramètre de topologie parmi :

- un nombre maximal de ports des routeurs N_{ij} (avec $i \geq 1$),
- un nombre h de niveaux de routeurs N_{ij} (avec $i \geq 1$),
- un nombre de nœuds de service N_{0j} connectés à un routeur N_{1j} appartenant au premier niveau (i = 1),
- un nombre de routeurs N_{i-1j} d'un niveau i-1 connectés à un même routeur N_{ij} d'un niveau i,
- un nombre de routeurs N_{ij} d'un niveau i connectés à un même nœud de service N_{i-1j} d'un niveau i-1, et
- un nombre de liaisons (ou liens) entre un routeur N_{ij} d'un niveau i et un nœud de service N_{i-1j} d'un niveau i -1.

On notera que l'usager de l'équipement électronique EE peut choisir au moyen de l'interface homme/machine IH chaque paramètre de topologie parmi une liste de paramètres de topologie fournie par l'interface de programmation INP.

Cette liste peut, par exemple, être stockée dans des moyens de stockage MS du dispositif de traitement DT. Ces moyens de stockage MS peuvent, par exemple, se présenter sous la forme d'une mémoire, éventuellement de type logiciel.

L'invention propose donc de définir de façon non ambigüe la structure de chaque topologie d'interconnexion à partir d'une ou plusieurs contraintes structurelles (définies chacune en fonction d'au moins une variable), en raison de la régularité des topologies d'interconnexion théoriques. On notera que ces contraintes structurelles, et donc les structures des topologies d'interconnexion, peuvent même être exprimées sous forme d'équations mathématiques, comme indiqué ci-dessous.

A titre d'exemple pour les structures PGFT, chaque contrainte structurelle peut être choisie parmi les inéquations suivantes $\forall i \in [1,h-1]$, où h est le nombre de niveaux de routeurs de la topologie d'interconnexion :

- $p_i * m_i + p_{i+1} * w_{i+1} \leq radixMax$ (cette contrainte structurelle permet de spécifier que le nombre de liaisons connectées à un routeur ne doit pas dépasser le nombre de ports disponibles sur ce routeur),
- $p_h * m_h \leq radixMax$, et
- $w_1 * p_1 + noeudsParFeuille \leq radixMax$,

où radixMax est le nombre maximal de ports de chaque routeur N_{ij}, et noeudsParFeuille est le nombre de nœuds de service N_{0j} connectés à un même routeur N_{1j} dit « feuille » du fait qu'il appartient au premier niveau.

Egalement à titre d'exemple pour les structures XGFT, chaque contrainte structurelle peut être choisie parmi les inéquations suivantes $\forall i \in [1,h-1]$, où h est le nombre de niveaux de routeurs de la topologie d'interconnexion :

- $m_i + w_{i+1} \leq radixMax$,
- $m_h \leq radixMax$, et
- $w_1 + noeudsParFeuille \leq radixMax$.

Egalement à titre d'exemple pour les structures GFT, chaque contrainte structurelle peut être choisie parmi les inéquations suivantes $\forall i \in [1,h-1]$, où h est le nombre de niveaux de routeurs de la topologie d'interconnexion :

- $m + w \leq radixMax$,
- $m \leq radixMax$, et
- $w + noeudsParFeuille \leq radixMax$.

Durant la deuxième étape (20) du procédé selon l'invention, on détermine au moyen d'un solveur (défini par les moyens de résolution MR) la topologie d'interconnexion optimale du système S, en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi dans la première étape (10) et d'au moins un critère d'optimisation choisi.

Le choix de chaque critère d'optimisation peut, par exemple être effectué par l'usager dans la première étape 10. En variante, il peut se faire dans la seconde étape 20. Ce choix peut se faire au moyen de l'interface homme/machine IH parmi une liste de critères d'optimisation fournie par l'interface de programmation INP. Cette liste peut, par exemple, être stockée dans les moyens de stockage MS du dispositif de traitement DT.

Grâce à la définition de la structure de chaque topologie d'interconnexion à partir d'au moins une contrainte structurelle fonction d'au moins une variable, on peut désormais déterminer la topologie d'interconnexion optimale d'un système S par résolution d'un problème d'optimisation au moyen d'un solveur (ou moyens de résolution MR). En d'autres termes, la séparation entre la modélisation de la structure et la méthode de résolution permet de déléguer la recherche de solutions optimales à un outil mathématique (ou solveur) fiable et robuste.

A titre d'exemple non limitatif, on peut utiliser le solveur appelé « Bonmin », lequel est capable de résoudre des systèmes d'équations non-linéaires à variables entières. Il permet de trouver des solutions précises en des temps réduits (typiquement quelques secondes pour des topologies de 64000 nœuds de service). En variante, on peut, par exemple, utiliser le solveur Ipopt.

Par exemple, chaque critère d'optimisation peut être choisi parmi au moins le coût du système S, le niveau de performance souhaité du système S (par exemple sa puissance de calcul ou sa rapidité de calcul), l'extensibilité du système S, la consommation d'énergie souhaitée du système S, des contraintes thermiques, notamment liées à l'environnement d'accueil du système S, et l'encombrement du système S (critère logistique). D'une manière générale un critère d'optimisation peut être n'importe quelle propriété

d'une topologie d'interconnexion.

On notera que dans certaines situations l'utilisation des seules contraintes structurelles peut aboutir à une topologie d'interconnexion juste mais non utilisable dans la pratique. Par conséquent, dans la seconde étape 20, on peut déterminer la topologie d'interconnexion optimale en fonction également d'au moins une contrainte d'utilisation choisie.

Par exemple, chaque contrainte d'utilisation peut être choisie parmi au moins :

- une interdiction de connecter à un routeur N_{ij} (avec i > 1) appartenant à un niveau i donné plus de liaisons (ou liens) vers le niveau supérieur à ce niveau donné que de liaisons (ou liens) vers le niveau inférieur à ce niveau donné (en effet les liaisons inférieures ne pourront pas générer suffisamment de trafic pour que toutes les liaisons supérieures soient utilisées). Avec les exemples de définition des variables m, w, p et h donnés plus haut, cette contrainte d'utilisation peut s'exprimer sous forme mathématique au moyen de l'inéquation $w_{i+1} \leq m_i$,
- une conservation d'une bande passante en entrée et en sortie de chaque routeur N_{ij} (avec i > 1). Avec les exemples de définition des variables m, w, p et h donnés plus haut, cette contrainte d'utilisation peut s'exprimer sous forme mathématique au moyen de l'équation $facteursMin_i = facteursMax_i = 1$, où $facteursMin_i$ et $facteursMax_i$ sont des facteurs de blocage minimal et maximal pour le niveau i considéré (un facteur de blocage étant un rapport de nombres de nœuds de service entre un niveau supérieur à un niveau i donné et un niveau inférieur à ce niveau i donné). Par exemple, on peut utiliser les inéquations $facteursMin_i \leq p_{i+1}*w_{i+1}/p_i*m_i \leq facteursMax_i$ et $facteursMin_1 \leq p_1*w_1/noeudsParFeuille \leq facteursMax_1$. Dans ce cas, si on n'a que des composantes égales à 1 pour $facteursMin_i$ et $facteursMax_i$, on a $1 \leq p_1*w_1/noeudsParFeuille \leq 1$ (pour le niveau 1, et la même chose pour les autres niveaux), et donc il y a conservation de la bande passante. En revanche, si on a par exemple pour le niveau 1 $facteursMin_1 = 0,5$ et $facteursMax_1 = 1$, on autorise jusqu'à une liaison montante pour deux descendantes, et donc il n'y a pas conservation de la bande passante. On notera que l'on pourrait ici utiliser d'autres formules, comme par exemple

facteursMin$_i$ = 1 et facteursMax$_i$ = 1 pour tous les i ou $p_{i+1} * w_{i+1} = p_i * m_i$,

- un nombre minimum de nœuds de service N_{ij}. Avec les exemples de définition des variables m, h et noeudsParFeuille donnés plus haut, cette contrainte d'utilisation peut s'exprimer sous forme mathématique au moyen de l'inéquation noeudsParFeuille*$\prod m_i$ (de i = 1 à h) ≥ minNoeuds, où minNoeuds est le nombre de nœuds de service N_{0j} connectés à des routeurs feuilles N_{1j} du premier niveau (i = 1), et
- un nombre maximum de liaisons (ou liens) entre deux routeurs N_{ij}. Cette contrainte d'utilisation permet d'éviter d'introduire une baisse de la tolérance aux pannes si un routeur est défectueux. Avec l'exemple de définition de la variable p donné plus haut, cette contrainte d'utilisation peut s'exprimer sous forme mathématique au moyen de l'inéquation $1 \leq p_i \leq$ maxRépétition, où maxRépétition est le nombre maximum de liaisons entre deux routeurs.

On notera qu'il est également possible d'utiliser des contraintes d'utilisation spécifiques pour écarter certains cas particuliers. C'est notamment le cas lorsque l'une des composantes du vecteur m et/ou l'une des composantes du vecteur w est égale à 1. En effet, dans ce cas il est possible de construire une structure PGFT similaire en supprimant un niveau (pour un prix inférieur et tout aussi résistant aux pannes). Dans ce cas, on peut utiliser une contrainte d'utilisation qui s'exprime sous forme mathématique au moyen des inéquations $w_{i+1} > 1$ et $m_i > 1$.

On notera également que l'usager de l'équipement électronique EE peut choisir au moyen de l'interface homme/machine IH chaque contrainte d'utilisation parmi une liste de contraintes d'utilisation fournie par l'interface de programmation INP. Cette liste peut, par exemple, être stockée dans les moyens de stockage MS du dispositif de traitement DT.

On a schématiquement illustré sur la figure 3 un exemple non limitatif de système S à structure PGFT définie par (h,m,w,p) = (2, [4,2], [2,2], [1,2]). On considère ici que chaque routeur a quatre ports disponibles. Un exemple de construction de ce système S est donné ci-dessous.

Pour la construction du premier niveau de routeurs N_{1j}, on utilise la valeur de m_1. Dans cet exemple, m_1 vaut 4 et donc les routeurs N_{2j} du

deuxième niveau seront reliés à quatre routeurs feuilles N_{1j} différents du premier niveau. Puis, on utilise la valeur de w_1. Dans cet exemple, w_1 vaut 2 et donc chaque routeur feuille N_{1j} du premier niveau est relié à deux routeurs N_{2j} du deuxième niveau. Puis, on utilise la valeur de p_1. Dans cet exemple, p_1 vaut 1 et donc on utilise une unique liaison (non redondante) entre les premier et deuxième niveaux. On appelle ci-après ilot de niveau 1 un groupe comprenant quatre routeurs N_{1j} (j = 1 à 4) du premier niveau et deux routeurs N_{2j} (j = 1 à 2) du deuxième niveau. Dans cet exemple, le système comprend deux ilots de niveau 1.

Ensuite, on utilise la valeur de m_2. Dans cet exemple, m_2 vaut 2 et donc les routeurs N_{3j} du troisième niveau seront reliés à deux routeurs N_{2j} de deux ilots différents du niveau 2. Puis, on utilise la valeur de w_2. Dans cet exemple, w_2 vaut 2 et donc chaque routeur N_{2j} du deuxième niveau sera relié à deux routeurs N_{3j} différents appartenant au troisième niveau (et plus précisément à des ilots de niveau 2 différents). Chaque ilot de niveau 1 ayant deux routeurs N_{2j} du deuxième niveau, on a donc deux routeurs N_{3j} dans chaque îlot de niveau 2.

Ensuite, on complète les liaisons manquantes entre les deuxième et troisième niveaux, puis, la valeur de p_2 étant égale à 2, on double chaque liaison entre routeurs des deuxième et troisième niveaux. Enfin, dans le cas d'un supercalculateur S on ajoute des nœuds de calcul N_{0j} en les connectant aux routeurs N_{1j} du premier niveau. Ici, la liaison entre chaque nœud de calcul N_{0j} et le routeur N_{1j} correspondant est simple pour ne pas dépasser le nombre de ports disponibles sur ce routeur N_{1j}.

Cette méthode de construction peut être mise en œuvre par le solveur (ou moyens de résolution MR) pour le système S de la figure 3, avec toutes les contraintes structurelles et toutes les contraintes d'utilisation décrites plus haut, mais également avec le coût du système S comme critère d'optimisation.

L'invention ne se limite pas aux modes de réalisation de procédé de détermination, de dispositif de traitement, et d'équipement électronique décrits ci-avant, seulement à titre d'exemple, mais elle englobe toutes les variantes que pourra envisager l'homme de l'art dans le seul cadre des revendications

ci-après.

REVENDICATIONS

1. Procédé de détermination d'une topologie d'interconnexion optimale d'un système (S) comprenant des nœuds de service (N_{ij}) interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau, caractérisé en ce qu'il comprend une première étape (10) dans laquelle on choisit une structure pour une topologie d'interconnexion optimale à déterminer parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, puis on choisit au moins un paramètre de topologie parmi un nombre maximal de ports des routeurs, un nombre de niveaux de routeurs, un nombre de nœuds de service (N_{ij}) connectés à un routeur appartenant au premier niveau et ci-après dit « routeur feuille », un nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, un nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, un nombre de liaisons entre un routeur d'un niveau i et un nœud de service (N_{i-1j}) d'un niveau i -1, et une seconde étape (20) dans laquelle on détermine au moyen d'un solveur ladite topologie d'interconnexion optimale en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi et d'au moins un critère d'optimisation choisi.

2. Procédé selon la revendication 1, caractérisé en ce que chaque critère d'optimisation est choisi dans un groupe comprenant un coût dudit système (S), un niveau de performance dudit système (S), une extensibilité dudit système (S), une consommation d'énergie dudit système (S), des contraintes thermiques, et un encombrement dudit système (S).

3. Procédé selon l'une des revendications 1 et 2, caractérisé en ce que chaque contrainte structurelle est choisie dans un groupe d'inéquations comprenant $p_i * m_i + p_{i+1} * w_{i+1} \leq \text{radixMax}$, $p_h * m_h \leq \text{radixMax}$, et $w_1 * p_1 + \text{noeudsParFeuille} \leq \text{radixMax}$, où m_i est le nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, w_i est le nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, p_i est le nombre de liaisons entre un routeur d'un niveau i et un nœud de

service (N_{i-1j}) d'un niveau i -1, radixMax est le nombre maximal de ports de chaque routeur, h est le nombre de niveaux de routeurs, et noeudsParFeuille est le nombre de nœuds de service (N_{0j}) connectés à un même routeur (N_{1j}) appartenant audit premier niveau.

4. Procédé selon l'une des revendications 1 à 3, caractérisé en ce que dans ladite seconde étape (20) on détermine ladite topologie d'interconnexion optimale en fonction également d'au moins une contrainte d'utilisation choisie.

5. Procédé selon la revendication 4, caractérisé en ce que chaque contrainte d'utilisation est choisie dans un groupe comprenant une interdiction de connecter à un routeur appartenant à un niveau i donné plus de liaisons vers le niveau supérieur audit niveau donné que de liaisons vers le niveau inférieur audit niveau donné, une conservation d'une bande passante en entrée et en sortie de chaque routeur, une non conservation d'une bande passante en entrée et en sortie de chaque routeur, un nombre minimum de nœuds de service (N_{ij}), un nombre maximum de liaisons entre deux routeurs.

6. Procédé selon l'une des revendications 1 à 5, caractérisé en ce que ledit système (S) est choisi dans un groupe comprenant un supercalculateur et un centre de données.

7. Produit programme d'ordinateur comprenant un jeu d'instructions qui, lorsqu'il est exécuté par des moyens de traitement, est propre à mettre en œuvre le procédé selon l'une des revendications précédentes pour déterminer une topologie d'interconnexion optimale d'un système (S) comprenant des nœuds de service (N_{ij}) interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau.

8. Dispositif de traitement (DT) pour déterminer une topologie d'interconnexion optimale d'un système (S) comprenant des nœuds de service (N_{ij}) interconnectés entre eux via des liaisons et constituant pour certains des routeurs ayant des ports et appartenant à au moins un niveau, et propre à équiper un équipement électronique (EE) comprenant des ressources de calcul (RC) et une interface homme/machine (IH), caractérisé en ce qu'il comprend i) une interface de programmation (INP) agencée pour permettre à un usager de choisir via ladite interface homme/machine (IH) une

structure pour une topologie d'interconnexion optimale à déterminer parmi un ensemble de structures définies chacune par au moins une contrainte structurelle, et au moins un paramètre de topologie choisi parmi un nombre maximal de ports des routeurs, un nombre de niveaux de routeurs, un nombre de nœuds de service (N_{ij}) connectés à un routeur appartenant au premier niveau, un nombre de nœuds de service d'un niveau i-1 connectés à un même routeur d'un niveau i, un nombre de routeurs d'un niveau i connectés à un même nœud de service d'un niveau i-1, un nombre de liaisons entre un routeur d'un niveau i et un nœud de service (N_{i-1j}) d'un niveau i -1, et ii) des moyens de résolution (MR) agencés pour déterminer ladite topologie d'interconnexion optimale en fonction d'au moins chaque contrainte structurelle correspondante et chaque paramètre de topologie choisi et d'au moins un critère d'optimisation choisi.

9. Equipement électronique (EE) comprenant des ressources de calcul (RC) et une interface homme/machine (IH), caractérisé en ce qu'il comprend en outre un dispositif de traitement (DT) selon la revendication 8.

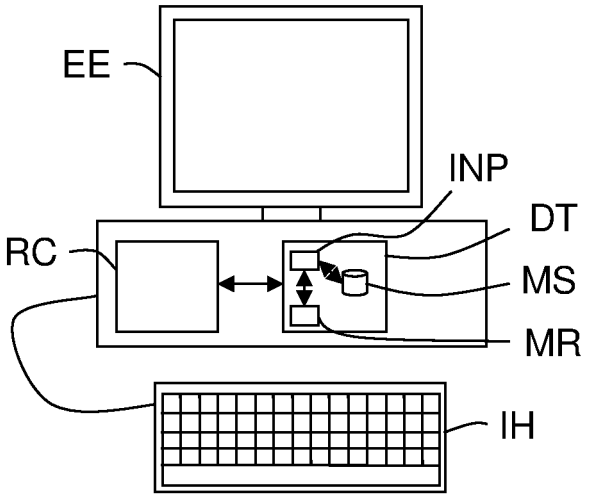


FIG.1

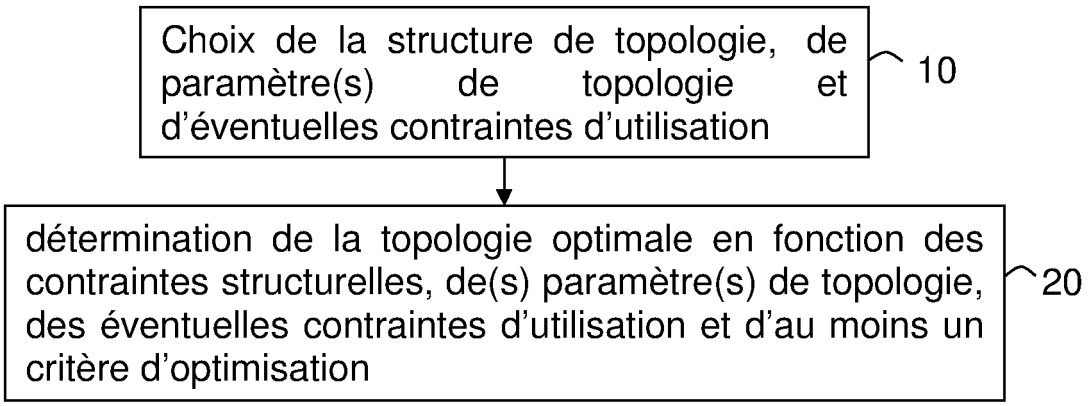


FIG.2

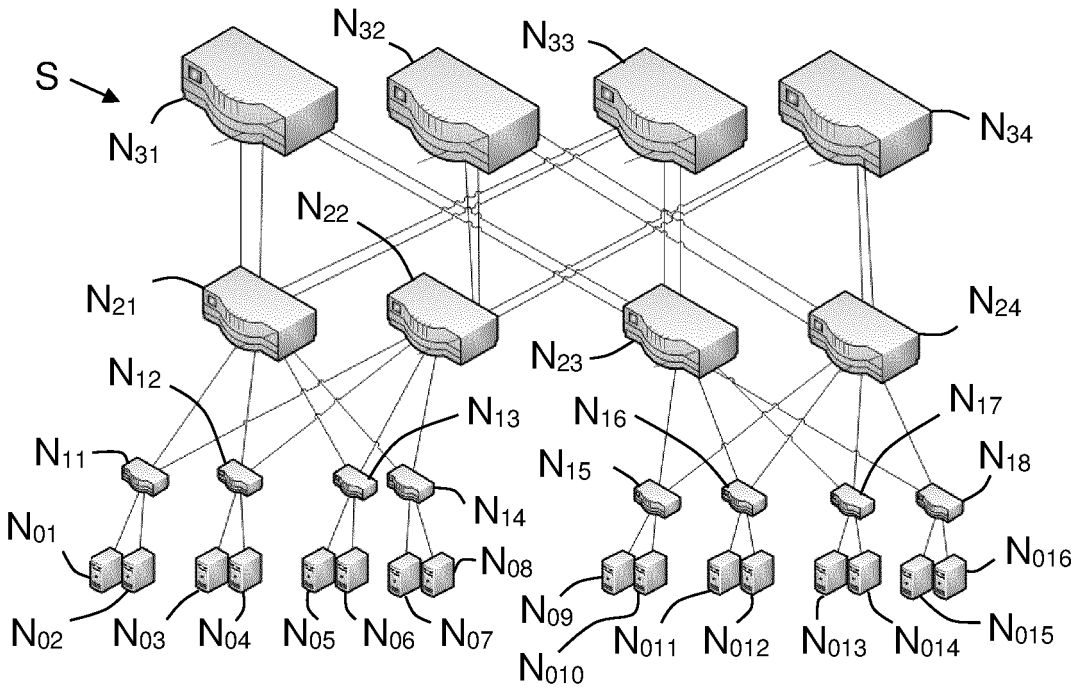


FIG.3

RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications déposées avant le commencement de la recherche

N° d'enregistrement national

FA 813915
FR 1558131

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	US 2014/023084 A1 (GOLDENBERG DROR [IL] ET AL) 23 janvier 2014 (2014-01-23) * abrégé * * alinéas [0100] - [0107] * -----	1-9	H04L12/751 H04L12/46
A	US 9 124 483 B2 (PETERSON BENJAMIN K [US]) 1 septembre 2015 (2015-09-01) * abrégé * * colonne 1, ligne 41 - colonne 2, ligne 19; revendication 1 * -----	1-9	
A	US 2013/108263 A1 (SRINIVAS ANAND [US] ET AL) 2 mai 2013 (2013-05-02) * abrégé * * alinéas [0003], [0009] - [0016], [0051] - [0071] * -----	1-9	
A	HAITAO WU ET AL: "MDCube", CONEXT '09 PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON EMERGING NETWORKING EXPERIMENTS AND TECHNOLOGIES; DECEMBER 1 - 4, 2009; ROME, ITALY, 1 janvier 2009 (2009-01-01), page 25, XP055258840, New York, NY, USA. DOI: 10.1145/1658939.1658943 ISBN: 978-1-60558-636-6 * Introduction and item 1 * -----	1-9	DOMAINES TECHNIQUES RECHERCHÉS (IPC) H04L

Date d'achèvement de la recherche	Examinateur
16 mars 2016	Le Bras, Patrick

CATÉGORIE DES DOCUMENTS CITÉS

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : arrière-plan technologique
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons

& : membre de la même famille, document correspondant

EPO FORM 1503 12.99 (P04C14)

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 1558131 FA 813915

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 16-03-2016
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2014023084 A1	23-01-2014	US 2012207018 A1	16-08-2012
		US 2014023084 A1	23-01-2014
US 9124483 B2	01-09-2015	CA 2814189 A1	10-05-2012
		CN 103299269 A	11-09-2013
		EP 2633395 A1	04-09-2013
		US 2013003605 A1	03-01-2013
		WO 2012061158 A1	10-05-2012
US 2013108263 A1	02-05-2013	US 2013108263 A1	02-05-2013
		US 2015350023 A1	03-12-2015

EPO FORM P0465