

Compte rendu d'activité pour l'année écoulée

1 Les objectifs fixés

Présentation du sujet de thèse: L'univers du datacenter se révèle être beaucoup plus complexe que l'on y pense. il regroupe plusieurs métiers dont l'interaction devient difficile au vue des tâches à accomplir complètement différentes. Toutefois ces secteurs d'activités partagent un objectif commun qui est l'accessibilité et la disponibilité des services IT hébergés dans cette infrastructure. En effet au début de leur apparition, l'intérêt était dirigé vers les données contenues dans les serveurs afin qu'ils soient accessibles 24h/24 délaissant au passage l'impact que ces bâtiments avaient sur l'environnement. On peut citer la production de chaleur issues des équipements IT, la surconsommation d'électricité pour éviter toute interruption d'intensité et une utilisation d'eau assez importante afin de refroidir les serveurs.

Notre coeur de métier est l'optimisation des énergies (électrique, thermique) produites dans un souci de réduire les couts d'exploitations et les impacts écologiques de ces bâtiments. Un des soucis rencontrés lors de la réalisation de notre projet a été l'absence de référentielles correctes sur les équipements présents dans l'infrastructure. Par exemple certains appareils, remplacés par d'autres avec des caractéristiques différentes ne figurent nulle part dans la documentation, l'historique des maintenances est décousu avec beaucoup d'incompréhension, l'organisation physique des appareils est modifiée et pas mis à jour et les équipements hors tension ne figurent pas dans le schema de fonctionnement.

Face à ces contraintes, notre préoccupation est de savoir comment trouver l'architecture physique du réseau associé au fonctionnement du système à plusieurs instants donnée. L'idée nous est alors venu d'utiliser les valeurs de grandeurs physiques constamment remontée dans le temps qui reflète l'état du système. Il s'agit des mesures physiques.

La première année a permis de définir quelles variables sont importantes dans un réseau énergétique. Ensuite nous avons effectué des correlations entre mesures issues de la meme grandeur physique dans le but de detecter les arcs ou arêtes partageant un sommet en commun. Et enfin nous avons défini une matrice de corrélation associée à toutes les grandeurs physiques en supposant qu'un arc est corrélé à un arc si toutes ses grandeurs sont corrélés entre elles.

La deuxième année a conduit à la proposition de deux algorithmes. Nous avons supposé que si plusieurs arêtes partagent un meme sommet alors ils forment un clique dans un graphe dual. Nous avons considéré la matrice de corrélation comme la matrice d'adjacence du line graphe induit par le graphe non orienté du réseau énergétique.

Le premier algorithme est *l'algorithme de couverture* proposé par Philippe Lehot [1] dont nous avons modifié afin qu'il considère les arêtes de notre graphe comme des sommets. On en a déduit la propriété suivante:

propriete 1 *Un graphe dont chaque sommet appartient à deux cliques et chaque arête à une clique admet une couverture en clique. Ce graphe est un line graphe.*

Ensuite, les sommets appartenant à plus de deux cliques, font l'objet de traitement dans l'algorithme suivant. Ces sommets proviennent d'erreurs de corrélations pendant le calcul de la matrice de corrélation. Ces erreurs sont des corrélations *fausses positives* et *fausses négatives*.

Le second algorithme nommé *algorithme de correction* a pour but de fusionner les cliques voisines d'un sommet, issu à l'ensemble des sommets n'appartenant à aucune clique, de tel sorte que

- ce sommet appartienne à deux cliques.
- le coût de cette fusion de cliques soit minimal.

L'objectif de cet algorithme est de fournir le line graphe le plus proche possible du line graphe induit par le réseau de flots, En d'autres termes, le line graphe dont la distance de Hamming soit minimale.

Les travaux de la troisième année consiste principalement à tester la robustesse des algorithmes proposés. Il s'agit:

- de définir l'ordre de traitement des sommets, n'appartenant à aucune cliques. En effet, nous avons générer 500 line graphes dans lesquels on a modifié $k \in [1, \dots, 9]$ corrélations (il y'a autant d'arêtes ajoutées et d'arêtes supprimées). On a comparé les distances de Hamming moyennes pour chaque nombre k selon quatre méthodes: degré minimum avec remise, coût minimum avec remise, degré minimum avec permutation et coût minimum avec permutation. Et la meilleure méthode est celle de degré minimum avec permutation.
- de spécifier dans quelle condition, les algorithmes fournissent de meilleurs resultats. La condition est le taux de corrélation fausses positives par rapport à celle fausses negatives dans la matrice d'adjacence. En effet, on modifie $k \in [1, \dots, 9]$ corrélations selon une probabilité p de sorte qu'ils aient plus d'aretes ajoutés dans les line graphes quand p tend vers 1 et plus d'arêtes supprimées quand p tend vers 0. On constate que les distances de hamming entre graphes fournis par les algorithmes et les line graphes générés est égale à 0 dans 60% des graphes. Cela signifie que, dans 60% des cas, les algorithmes prédisent les line graphes des graphes de base.
- de tester sur les réseaux réelles. On a remarqué que le temps d'exécution de l'algorithme de correction est infini quand le degré moyen du réseau de flots est supérieur à 9. Nous devons changer la structure de données utilisé pour constater si cela change largement les résultats déjà obtenus.

Travaux à réaliser Dans le quatrième trimestre, je compte faire les taches suivantes:

- changer la structure de données pour tester les performances de l'algorithme de correction.
- orienter les aretes du graphe car nous pouvons déjà reconstruire le réseau non orienté à partir du line graphe.
- tester les algorithmes sur ces line graphes particuliers tel que les graphes iourtes.

- début de la rédaction de la thèse.

References

- [1] Philippe G. H. Lehot. An optimal algorithm to detect a line graph and output its root graph. In *Journal of the 'ACM'*, pages 569–575. Association for Computing Machinery (ACM), oct 1974.