

Rapport d'activités de la thèse année par année

2017

Contents

1	Présentation du sujet de thèse	1
2	Déroulement année 1	2
3	Déroulement année 2	2
4	Déroulement année 3	2
5	Travaux à achever	3

1 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 taches 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 presents 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.

2 Activités année 1

Nous nous sommes consacrés à l'étude de corrélations de mesures physiques.

Nous avons débuté par le traitement des données reçus en utilisant l'interpolation de Lagrange pour combler les mesures absentes dans les datasets.

Ensuite, nous avons observé que certains équipements ont les mêmes profils de consommation quand ils sont alimentés par le même équipement. Cela signifie que tous ces équipements dans un graphe partagent une extrémité commune. On en conclut que les mesures physiques des équipements sont corrélées si leurs profils de consommation (leur tendances) sont identiques. Partant de cette observation, les méthodes de corrélation glissantes et par fenêtre de Pearson se sont montrées très efficaces.

Toutefois la présence d'onduleurs dans le réseau électrique modifie ces profils en aval de ceux-ci en les lissant c'est-à-dire tous les équipements alimentés par l'onduleur. Cela signifie des tendances différentes en tension et dans certains cas en intensité. Cela induit que les méthodes citées plus haut ne sont plus pertinentes. Nous avons décidé de nous tourner vers le machine learning.

3 Activités année 2

La deuxième année fut consacrée à l'élaboration d'algorithme de reconstruction de graphes après une bibliographie sur la reconstruction de graphes. Nous avons utilisé la notion de line graphs [F.(1972)] pour distinguer les équipements ayant une extrémité commune. En effet, en transformant les arêtes du graphe en noeuds et les noeuds du graphe en arêtes, on obtient une clique. Différentes méthodes existent et celle choisie pour la décomposition en line graphs est celle de Philippe Lehot [Lehot(1974)]. Malheureusement cette algorithme ne traite pas les sommets dans des ambiguïtés c'est-à-dire ayant deux couvertures en cliques distincts, chose quasi impossible parce que la décomposition en cliques d'un line graphe est unique.

Nous avons inclus à l'algorithme existant la détection des sommets en situation d'ambiguïtés et trouvent la meilleure clique de couverture à partir de la loi de conservation des noeuds appliquée aux mesures physiques.

Rappelons que l'électricité dans sa propagation subit des pertes dues à la résistance des câbles. Ce phénomène est dit pertes par effets joules et nous le notons ϵ . Nous avons conclu que le meilleur intervalle des pertes par effets joules pour lequel la loi de conservation sur nos mesures était respectée, est $\epsilon \geq 0.8$.

Cependant l'algorithme peut ne retrouver aucune décomposition en cliques parce que le graphe n'est pas un line graphe. Dans ce cas de figure, nous proposons l'algorithme de correction dont le but est de trouver un line graphe ayant une distance de Hamming minimale par rapport à celui fournit en entrée.

4 Activités année 3

La simulation des algorithmes proposées sur des données concerne les travaux de la troisième année. Les premiers tests ont débuté sur des données théoriques. En effet, nous avons généré 500 line graphs avec les mesures physiques de chaque sommet et on modifie $k \in [1, 10]$ cases de la matrice d'adjacence du line graphe de la manière suivante :

- la case à 0 du line graphe est transformée en 1: on parle de corrélations fausses positives.

- la case à 1 du line graphe est transformée en 0: on parle de corrélations fausses négatives.

Enfin, pour chaque case de la matrice d'adjacence du line graphe, on attribue une valeur de corrélation selon le type de la case de cette matrice (corrélations fausses négatives, vrai négatives, fausses positives et vrai positives).

Les tests réalisés sur des données théoriques présentent de bons résultats lorsque l'on ajoute $k < 7$ erreurs de corrélations dans la matrice d'adjacence du line graphe. En d'autres termes, les algorithmes (de couverture et de correction) fournissent le line graphe du réseau électrique lorsque la matrice d'adjacence du graphe fourni en paramètres de nos algorithmes possède $k < 7$ erreurs de corrélations, soit 2% erreurs de corrélations pour 70% de corrélations fausses négatives et 30% de corrélations de fausses positives.

Rappelons que la fonction de coût pour la modification d'une case de la matrice d'adjacence utilise les valeurs de corrélations et le seuil de transformation de corrélation en valeurs binaires est $s = 0.7$. La fonction de coût unitaire (c'est-à-dire la modification est de 1) ne fournit pas de résultats probants.

Les données réelles proviennent d'un datacenter d'un opérateur de téléphonie. Le test effectué sur ces données consiste à la détermination de la matrice de corrélation, à la transformation de la matrice de corrélation en une matrice d'adjacence en appliquant le seuil s et enfin l'exécution des algorithmes de couverture et de correction.

Les premiers résultats obtenus sont très décevants car la distance de Hamming est 153. Cela s'explique par le fait que la matrice d'adjacence contient 8% erreurs de corrélations dont 85.5% et 14.5% de corrélations respectivement fausses positives et fausses négatives. Les erreurs de corrélations proviennent de la présence d'onduleurs dans le réseau dans la mesure ou l'onduleur lisse les profils de consommation des équipements qui lui sont rattachés impliquant que les équipements de branches différentes se retrouvent avec les profils de consommations c'est-à-dire une droite horizontale.

5 Travaux à achever

- Tester sur les données théoriques une fonction de coût en cloche. La particularité de cette fonction est les valeurs de corrélation fausses positives et fausses négatives très proche du seuil s . Cela entraîne que le coût d'une corrélation vrai négative ou vrai positive est très élevé et qu'on priorise la modification de corrélations fausses positives et fausses négatives.
- Trouver une méthode de corrélation qui soit indifférent à la présence d'un onduleur dans le réseau électrique.
- Déterminer le graphe root du line graphe.

References

- [F.(1972)] Harary F. Line graphs. In *Graph Theory*, chapter 8, pages 71–83. Massachusetts: Addison-Wesley, 1972.
- [Lehot(1974)] 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.