

# chapitre3 : Réseau de flots et mesures physiques

Wilfried Ehounou

*July 5<sup>e</sup> 2018*



# Contents

<b>1</b>	<b>Réseau de flots et mesures physiques</b>	<b>5</b>
1.1	Réseau électrique d'un datacenter : un graphe de flots . . . . .	5
1.1.1	Topologie du réseau électrique du datacenter . . . . .	5
1.1.1.0.0.1	Les sources . . . . .	5
1.1.1.0.0.2	Les tableaux . . . . .	6
1.1.1.0.0.3	Les baies (ou racks) de serveurs . . . . .	6
1.1.1.0.0.4	Les câbles . . . . .	6
1.1.2	Réseau électrique : un graphe de flots . . . . .	7
1.2	Modélisation du réseau électrique du datacenter . . . . .	7
1.2.1	Description du graphe de flots . . . . .	7
1.2.2	Grandeurs, flots et contraintes physiques . . . . .	8
1.2.2.1	Grandeurs physiques . . . . .	8
1.2.2.2	Capacité d'un arc . . . . .	8
1.2.2.3	Description d'un flot physique . . . . .	9
1.2.2.4	Contraintes sur les flots . . . . .	9
1.2.2.5	Description de Verif-correl . . . . .	10
1.3	État de l'art sur la découverte de topologie . . . . .	11
1.3.1	Découverte des topologies par des sondes . . . . .	11
1.3.2	Tomographie des réseaux . . . . .	12
1.3.2.0.0.1	l'estimation des paramètres (caractéristiques) d'un lien . . . . .	12
1.3.2.0.0.2	La prédiction de topologie . . . . .	13
1.3.2.0.0.3	La densité du trafic entre émetteur/recepteur . . . . .	13
1.3.3	Reconstruction de la topologie par les séries temporelles . . . . .	13
1.4	Problème de découverte de topologie électrique . . . . .	14
1.5	Conclusion du chapitre . . . . .	15



# Chapter 1

## Réseau de flots et mesures physiques

Les réseaux énergétiques ont pour rôle de fournir une énergie à des entités consommatrices sans interruption de service. Ces réseaux fonctionnent en courants continu et en alternatif.

Notre étude est limitée au courant alternatif parce que la demande d'électricité des équipements varie constamment et la quantité d'électricité est connue pour chaque équipement de ce réseau. Notre objectif est de regrouper les équipements qui ont la même source d'alimentation. Ces sources d'alimentation subissent régulièrement des maintenances et le schéma électrique n'est pas mis à jour, ce qui entraîne des problèmes dans le système de surveillance de ce réseau et aussi la planification de nouvelles maintenances.

Dans ce réseau électrique, nous connaissons les liens et les mesures potentielles sur ces liens mais nous ignorons les extrémités de ces liens. Notre problème est d'identifier ces extrémités à partir des mesures électriques. Nous faisons de la *découverte de topologie*. Le modèle proposé a été établi notamment à partir de la description du réseau électrique d'un datacenter d'un opérateur téléphonique appelé *Champlan*.

Notre chapitre comprend quatre parties. La première partie montre la relation entre un réseau électrique et un graphe de flots. La seconde partie modélise le réseau de flots et la troisième partie présente les travaux existants sur la découverte de topologie. La dernière partie présente l'approche retenue pour résoudre notre problème.

### 1.1 Réseau électrique d'un datacenter : un graphe de flots

#### 1.1.1 Topologie du réseau électrique du datacenter

Un Datacenter ou centre de données est un site physique regroupant des installations informatiques interconnectées (serveurs, réseau informatique, système de sauvegarde) et une infrastructure énergétique adéquate (un système de distribution électrique, un commutateur électrique, des réserves d'énergie, des générateurs dédiés à la sauvegarde de données, un système de ventilation et de refroidissement).

Le réseau électrique est le coeur de datacenter car il alimente à la fois les infrastructures informatique et énergétique. Il se compose de quatres entités :

**1.1.1.0.0.1 Les sources** : ces équipements sont les points d'entrée de l'électricité dans le datacenter et sont directement rattachés au gestionnaire de réseau régional ou national. Ils sont de deux types :

- Ceux qui alimentent le réseau en cas de dysfonctionnement du gestionnaire de réseau. Ce sont les accumulateurs, les groupes électrogènes.
- Ceux qui transforment l'électricité reçue du gestionnaire pour des puissances utilisables dans le datacenter. Ce sont les transformateurs basse tension communément appelés *Transformateur Général Basse Tension (TGBT)*.

Généralement, ces deux types d'équipements ne fonctionnent pas concomitamment.

**1.1.1.0.0.2 Les tableaux** : aussi appelé tableau de répartition, ce sont des équipements passifs dont la fonction est celle de commutateur. Ils représentent l'organe centrale de l'installation dans la mesure où ils regroupent tous les circuits électriques et systèmes de protection vers les baies de serveurs. Une baie est une armoire contenant plusieurs serveurs. Chaque baie possède un disjoncteur sur ce tableau afin d'interrompre l'alimentation en cas de danger. Les tableaux sont considérés comme des équipements passifs car l'électricité qui traverse ces équipements a une perte négligeable. Ces pertes sont les *pertes par effets joules*. Un exemple de tableau est présenté dans le table 1.1 (à gauche).

**1.1.1.0.0.3 Les baies (ou racks) de serveurs** : les baies distribuent la puissance nécessaire au fonctionnement de chaque serveur qui lui est rattaché. La baie a un rôle de multiprise pour tous les serveurs. Dans le système de supervision électrique, les baies sont les consommateurs de l'électricité. Elles sont des équipements actifs dans le réseau électrique. Le table 1.1 (à droite) présente une baie de serveurs.

**1.1.1.0.0.4 Les câbles** : les câbles ont pour rôle de rattacher les 3 entités précédemment citées afin de transporter l'électricité vers les baies de serveurs. Ils sont caractérisés par les résistances que nous considérons constantes.



Table 1.1: À gauche : tableaux électriques Okken PCC, à Droite : baie de serveurs (source: news.pixelistes.com).

### 1.1.2 Réseau électrique : un graphe de flots

L'électricité, acheminée par le gestionnaire de réseau, arrive aux transformateurs basse tension qui généralement fonctionnent en mode triphasé. Ces transformateurs vont convertir la puissance *HTA* reçue (20KVA chez Enedis) en une puissance *BT* (400KVA) et cette puissance est envoyée sur chaque phase. Une phase est un canal de transport de courant et le courant d'une phase est exprimé en fonction du sinus et d'un décalage de  $2\pi/3$  par rapport au courant d'une autre phase. Chaque phase transporte cette énergie aux divers tableaux. Chaque tableau peut être rattaché à deux phases pour éviter les micro-coupures d'électricité. Pendant ces micro-coupures, les accumulateurs et les groupes électrogènes prennent le relai dans le but d'éviter une interruption de service. Les tableaux sont rattachés aux phases et aux baies par des câbles électriques. Chaque équipement mesure la quantité d'électricité qui le traverse. Les câbles sont unidirectionnels. Aucun équipement ne s'alimente soit-même et les équipements de même nature ne sont pas rattachés entre eux. Par exemple, il n'existe aucun câble entre des tableaux et aucune baie n'alimente une autre baie. L'électricité suit un sens : des sources vers les baies. Chaque équipement mesure la quantité d'électricité qui le traverse. Par convention, ces mesures sont portées par les câbles incidents entrants dans chaque équipement.

Notre réseau électrique se modélise avec un réseau de flots dont le graphe est un graphe orienté sans circuit *Directed Acyclic Graph (DAG)* dans lequel chaque sommet représente un équipement, un arc représente des câbles électriques et qu'aucun équipement ne s'alimente soi-même (absence de circuits dans le réseau).

**Conclusion** : le réseau électrique d'un datacenter comprend des sources, des tableaux, des baies de serveurs et ils sont tous reliés par des câbles électriques. Les équipements de ce réseau ne s'alimentent pas soi-même et les équipements de même nature ne possèdent pas de câbles entre eux. Les câbles sont unidirectionnels et l'électricité a toujours le même sens : de la source aux baies. Nous en concluons que la topologie du datacenter est un *graphe orienté sans circuit* induit par le sens de la circulation du courant électrique sur ces liens. Ainsi ce graphe est la topologie du réseau électrique.

## 1.2 Modélisation du réseau électrique du datacenter

### 1.2.1 Description du graphe de flots

Soit  $G = (V, A, CAP)$  le graphe orienté sans circuit modélisant le réseau électrique. Chaque sommet de  $G$  est soit une source, soit un tableau ou soit un serveur. L'ensemble des sommets  $V$  est composé des équipements sources  $V_S$ , intermédiaires ou passifs  $V_I$  et charges ou serveurs  $V_C$ . Il est une union disjointe deux à deux de  $V_C$ ,  $V_I$  et  $V_S$  de cardinalité  $n$  dans laquelle :

- Les sommets  $V_S$  sont des sommets de degré entrant nul  $d^- = 0$ .
- Les sommets  $V_I$  sont des sommets de degrés entrant et sortant non nuls  $d^- \neq 0, d^+ \neq 0$ .
- Les sommets  $V_C$  sont des sommets de degré sortant nul  $d^+ = 0$ .

$$V = V_S \cup V_I \cup V_C \text{ et } V_S \cap V_I = \emptyset \text{ et } V_S \cap V_C = \emptyset \text{ et } V_C \cap V_I = \emptyset$$

L'ensemble des arcs  $A$  modélise  $m$  câbles électriques. Chaque arc a un flot, une capacité et une résistance considérée constante. La capacité de chaque arc est fonction de la grandeur associée à

cet arc.

Par convention, les arcs incidents entrants dans chaque sommet du graphe portent les mesures des grandeurs physiques.

## 1.2.2 Grandeurs, flots et contraintes physiques

Nous présentons les grandeurs physiques dans le réseau électrique puis définissons la capacité d'un arc et enfin décrivons les flots selon chaque grandeur pour un arc donnée.

### 1.2.2.1 Grandeurs physiques

Le réseau électrique a deux modes de fonctionnement : le mode triphasé regroupant les grandeurs  $U_{12}, U_{23}, U_{13}, I_{12}, I_{23}, I_{13}$  et le mode monophasé regroupant les grandeurs  $I, U$ . Les autres grandeurs sont communes aux deux systèmes (les grandeurs  $P, Q, S, \cos\phi$ ). Les symboles  $U_{12}, U_{23}, U_{13}, U$  sont des tensions,  $I_{12}, I_{23}, I_{13}, I$  des intensités,  $P, Q, S$  des puissances actives, réactives, apparentes respectivement et  $\cos\phi$  ou  $FP$  le facteur de puissance.

Une grandeur physique sur un arc est une caractéristique physique mesurable selon une unité de mesure.

Selon le phénomène physique pris en compte, les grandeurs physiques sont prédéfinies. Dans le cas de l'électricité, les grandeurs physiques forment l'ensemble  $\mathbf{GP}$  de cardinalité finie défini comme suit :

$$GP = \{I, I_1, I_2, I_3, U, U_{12}, U_{23}, U_{13}, P, Q, S, FP\} \quad (1.1)$$

avec le facteur de puissance  $FP$  qui désigne le déphasage entre l'intensité ( $I$ ) et la tension ( $U$ ).

Ces deux modes (triphasé et monophasé) peuvent fonctionner dans le même réseau. Cela implique qu'il n'existe qu'un seul sous-ensemble de grandeurs sur un arc, soit des grandeurs monophasées soit des grandeurs triphasées. On note  $GP^{a_i}$  l'ensemble des grandeurs sur un arc  $a_i$

$$\forall a_i \in A, \quad GP^{a_i} \subset GP \quad (1.2)$$

On distingue deux types de grandeurs:

- Grandeurs à différentiel de potentiel: les tensions. On les note  $gp_{dtp} \in \{U_{12}, U_{23}, U_{31}, U\}$ .
- Grandeurs à effet calorique: l'intensité, la puissance active et réactive. On les note  $gp_{cal} \in \{P, I_1, I_2, I_3, Q\}$ .

### 1.2.2.2 Capacité d'un arc

Soit un arc  $a_i \in A$  et  $GP^{a_i} \in GP$  l'ensemble des grandeurs physiques associé à l'arc  $a_i$ . La capacité d'un arc  $a_i$  est une fonction  $Cap_{a_i}$  qui, pour chaque grandeur physique  $GP^{a_i}$  associe une valeur réelle positive  $\mathbb{R}^+$

$$Cap_{a_i} : GP^{a_i} \rightarrow \mathbb{R}^+ \quad (1.3)$$

Le vecteur  $CAP$  contient les capacités pour chaque grandeur et chaque arc.

$$CAP = (Cap_a[x])_{a \in A, x \in GP^a} \quad (1.4)$$



### 1.2.2.3 Description d'un flot physique

Les mesures physiques sont des valeurs de ces grandeurs. Le vecteur de mesures  $gp_a^x$  de norme  $T^{a,x}$  (c'est-à-dire le nombre de valeurs associées à une grandeur dans une série temporelle) est une série de mesures associée à l'arc  $a$  et à la grandeur  $x \in GP^a$  dont le  $i^{eme}$  élément est  $gp_a^x[i]$ . Le vecteur  $gp_a^x$  associé à la grandeur  $x \in GP^a$  et à l'arc  $a \in V$  est défini comme suit :

$$gp_a^x = (gp_a^x[t])_{0 < t < T^{a,x}} \quad (1.5)$$

avec  $gp_a^x[t] \in \mathbb{R}^+$  et  $t \in \mathbb{N}^+$ .

**Remarque 1** Soient  $gp_a^x[t]$  le  $i^{eme}$  élément de la série temporelle de la grandeur  $x \in GP^a$  et  $a, a'$  deux arcs distincts.

- $gp_a^x[t]$  et  $gp_a^y[t]$ , pour  $x, y \in GP^a$  sont prises aux mêmes instants.
- pour  $x \in GP^a \cap GP^{a'}$ ,  $gp_a^x[t]$  et  $gp_{a'}^x[t]$  sont prises à des instants différents.

Certaines valeurs de  $gp_a^x$  sont indéfinies ou erronées dans certains cas.

### 1.2.2.4 Contraintes sur les flots

Un flot  $gp_a^x[t]$  est admissible s'il respecte, pour chaque arc  $a \in A$  traversé, la contrainte ci-dessous :

$$0 \leq gp_a^x[t] \leq Cap_a[x] \quad (1.6)$$

avec  $Cap_a$  la capacité de l'arc  $a$  pour la grandeur  $x \in GP^a$ .

Un flot est une fonction qui prend en entrées un arc  $a$ , une grandeur  $x \in GP^a$ , un vecteur  $gp_a^x$  associé à la grandeur  $x$  de l'arc  $a$ , un facteur de puissance  $\cos\phi$  ou  $FP$  associé à l'arc  $a$  et retourne un vecteur défini comme suit :

$$flo(a, x, gp_a^x) = f(a, x, r, \cos\phi, gp_a^x) = \begin{cases} \frac{gp_a^x}{r \times \cos\phi}, & x \in gp_{ddp} \\ gp_a^x, & x \in gp_{cal} \end{cases} \quad (1.7)$$

avec  $r$  la résistance du câble,  $FP$  ou  $\cos\phi$  le facteur de puissance.

Une valeur  $flo_t(a, x, gp_a^x)$  de  $flo(a, x, gp_a^x)$  s'obtient à un indice  $t < T^{a,x}$  donné et se définit comme suit :

$$flo_t(a, x, gp_a^x) = f(a, x, r, \cos\phi, gp_a^x) = \begin{cases} \frac{gp_a^x(t)}{r \times \cos\phi}, & x \in gp_{ddp} \\ gp_a^x(t), & x \in gp_{cal} \end{cases} \quad (1.8)$$

Soient  $a \in A$  un arc et  $x \in GP^a$  une grandeur liée à l'arc  $a$ . L'ensemble des arcs incidents à  $a$  et ayant la même extrémité initiale que  $a$  est noté  $succ(a)$  et l'ensemble des arcs incidents à  $a$  et ayant la même extrémité finale que  $a$  est noté  $pred(a)$ . Tous les éléments de  $succ(a)$  et  $pred(a)$  ont les mêmes grandeurs physiques.

La fonction  $flo$  doit respecter la contrainte de la loi de conservation  $R$  [?]. La loi de conservation  $R$  ne s'applique qu'avec les grandeurs à effet calorifique  $gp_{cal} \in GP$  et se définit comme suit:

1.

$$\sum_{a_j \in pred(a)} flo_t(a_j, x, gp_{a_j}^x) = \sum_{a_k \in succ(a)} flo_t(a_k, x, gp_{a_k}^x) + \epsilon \quad (1.9)$$

avec  $\epsilon$  les pertes par effets joules. Cette équation est la loi de conservation ou de Kirchhoff

### 1.2.2.5 Description de Verif-correl

La fonction *Verif – correl* détermine le sous-ensemble d’arcs entrants et sortants d’un sommet du réseau électrique en se basant sur les mesures physiques et les lois de conservation  $R$  définies dans le paragraphe 1.2.2.4.

Soit  $S \subset A$  l’ensemble fini d’arcs incidents à un sommet  $v \in V$  et  $x \in GP$  une grandeur physique telle que chaque arc  $a \in S$  a un flot  $gp_a^x$ . Nous partitionnons  $S$  en deux sous-ensembles  $S_1$  et  $S_2$  tels que  $S_1 \cap S_2 = \emptyset$ .

La fonction *Verif – correl* est booléenne, prend en paramètres  $S_1$ ,  $S_2$  et une grandeur  $x$ . Elle retourne 1 si :

- $S_1$  est l’ensemble des arcs entrants du sommet  $v$ .
- $S_2$  est l’ensemble des arcs sortants du sommet  $v$ .

Si les lois  $R$  ne sont pas vérifiées alors *Verif – correl* retourne 0.

$$Verif - correl(S_1, S_2, x) = 1 \Leftrightarrow \sum_{a_i \in S_1} gp_{a_i}^x(t) - \sum_{a_j \in S_2} gp_{a_j}^x(t) \leq \epsilon$$

En effet, considérons  $t$  un instant de temps et  $diff(t) = \sum_{a_i \in S_1} gp_{a_i}^x(t) - \sum_{a_j \in S_2} gp_{a_j}^x(t)$  la différence entre les flots entrants et sortants à un instant  $t$ . La différence  $diff(t)$  doit toujours être inférieure aux pertes par effets joules  $\epsilon$  quelle que soit l’instant  $t$ .

Nous décidons que la fonction *Verif – Correl* retourne 0 si plus de 10% des différences  $diff(t)$  sont supérieures à  $\epsilon$ .

Nous considérons, dans la suite du rapport, que la *décision de Verif – correl est toujours exacte*. Cela signifie que si  $Verif - correl(S_1, S_2, x) = 0$  alors les arcs de  $S$  ne partagent pas un sommet de  $V$ . Cependant, dans la pratique, nous utilisons rarement cette fonction parce que nous devons tester dans l’ordre de  $2^{|S|}$  bipartitions dans le pire des cas pour obtenir la bonne partition et cela est impossible pour  $S$  très grand.

**Conclusion** : le réseau électrique est modélisé par un graphe  $G = (V, A, CAP)$ . L’ensemble des sommets  $V$  est modélisé par des équipements sources  $V_S$ , intermédiaires  $V_I$  et serveurs  $V_C$ . Les sous-ensembles  $V_S, V_I, V_C$  sont disjoints deux à deux. Chaque arc  $a$  contient des grandeurs physiques  $GP^a$ . Nous en avons dénombré 8 regroupées en grandeurs à différentiel de potentiel  $gp_{ddp} = \{U_{12}, U_{23}, U_{31}, U\}$  et en grandeurs à effet calorifique  $gp_{cal} = \{I_1, I_2, I_3, P\}$ . Pour chaque grandeur physique  $x \in GP^a$  associée à l’arc  $a$ , une capacité  $Cap_a$  et les mesures  $gp_a^x$  lui sont associées. Le flot de mesures  $flo$  est un vecteur de mesures qui dépend de l’arc  $a$ , de la grandeur physique  $x$  et de la mesure physique  $gp_a^x$ . Une valeur de flot  $flo_t$  vérifie la loi de conservation  $R$  et est définie comme suit :

$$flo_t(a, x, gp_a^x) = f[a, x, r, \cos\phi, gp_a^x, t] = \begin{cases} \frac{gp_a^x[t]}{r \times \cos\phi}, gp \in \{U, U_{12}, U_{23}, U_{13}\} \\ gp_a^x[t], gp \in \{P, Q, I, I_1, I_2, I_3\} \end{cases} \quad (1.10)$$

Nous avons défini la fonction *Verif – correl*, qui étant donnée deux ensembles d’arcs  $S_1, S_2$ , affirme si ces arcs concourent en un sommet  $v \in V$  en attribuant  $S_1$  à l’ensemble d’arcs entrants et  $S_2$  à l’ensemble d’arcs sortants du sommet  $v$ . Nous considérons que la réponse de *Verif – correl* est toujours exacte.

## 1.3 État de l'art sur la découverte de topologie

Notre problème de découverte de topologie est d'identifier la topologie du réseau à partir des mesures c'est-à-dire les extrémités communes aux arcs dans le réseau.

Il existe un grand intérêt à la découverte de topologie notamment dans les systèmes distribués avec le déploiement de nouvelles générations de capteurs qui permettent de collecter des mesures selon différentes granularités (millisecondes, secondes, minutes, quart d'heure, etc). La communauté scientifique examine très peu comment interpréter ces mesures et l'impact de celles-ci dans le réseau. En effet, beaucoup de travaux sont orientés sur la découverte de la topologie au moyen de protocoles réseau. Des sondes sont propagées dans le réseau omettant la présence de capteurs dans les réseaux. D'autres travaux cherchent à prédire la topologie du réseau informatique à partir des lois statistiques et des modèles de files d'attentes.

Nous regroupons ces travaux en trois axes. Les deux premiers axes font principalement de la métrologie et cela leur permet de déduire la topologie en connaissant les caractéristiques de chaque lien et de chaque noeud du système. Le dernier axe porte sur la reconstruction de la topologie par les séries temporelles.

### 1.3.1 Découverte des topologies par des sondes

La découverte de topologie est un sujet important dans les réseaux informatiques. En effet, dans ces réseaux, on connaît la topologie physique et les noeuds mais on ignore l'état des noeuds. On recherche alors la topologie fonctionnelle du réseau (l'interconnexion entre les noeuds) selon l'état des noeuds. Il existe de nombreux outils pour sonder et faciliter les tâches d'administration de ces réseaux. Ces outils permettent aux administrateurs réseau de manager efficacement et de découvrir le réseau. Nous citons HPOenView [?] capable de localiser une erreur et d'envoyer des notifications de plusieurs événements incluant principalement des pertes dans le réseau et l'utilisation des liens. L'inconvénient de ces outils est leur coût et ils ne sont pas abordables pour les petites et moyennes organisations. La découverte de réseaux informatiques s'effectue aussi avec des protocoles réseaux dont les plus connus sont ICMP [?] et SNMP [?]. Divers algorithmes ont été proposés. Nous pouvons citer l'algorithme de *Narayan et al.* [?] qui réalise la découverte de topologie et de services pour des réseaux hétérogènes en se servant du système netInventory [?]. Cet algorithme est basé sur deux hypothèses : (i) chaque domaine doit avoir un seul sous-réseau et (2i) les tables de routage sont complètes. Pour la découverte de topologie, le système *netInventory* énumère la liste des adresse IP, envoie des messages ECHO or ICMP pour déterminer si un noeud est actif. Dans le cas où PING est désactivé, netInventory se sert de ipRouteTable et ipNetToMediaTable dans les routeurs pour connaître l'état d'un noeud.

De même, l'algorithme proposé par *Kuangyu Qin* [?] est basé sur *SNMP* et suppose que l'administrateur réseau est connecté à l'interface de routage et envoie des paquets de découverte aux routeurs. La station de gestion débute la découverte par la lecture des tables de routage. En utilisant les MIB (Management Information Base) et les agents SNMP, cet algorithme élimine les équipements redondants et génère une topologie efficiente même quand il existe des VLAN dans le réseau.

Un autre algorithme, proposé par *Bilal Saeed, TarekSheltami et Elhadi Shakshuki* [?], est basé sur netInventory et accélère la découverte de la topologie sans générer du trafic supplémentaire. Cet algorithme, nommé TDA (Topology Discovery Algorithm), se sert du multi-threading pour gérer toutes les requêtes de découverte réseau sur la plateforme Android. Une étude bibliographique, effectuée par *Ahmed et al.* [?], fournit les différentes techniques et algorithmes pour la découverte de topologie de réseaux informatiques. Les auteurs déclarent que la plupart des algorithmes de

découvertes physiques de réseau sont basés sur le protocole SNMP. En d'autres termes, ils supposent que tous les noeuds de ce réseau sont repertoriés et actifs au moment de la découverte.

**Conclusion :** cette méthode est contraire à notre sujet de recherche dans lequel nous ignorions les noeuds de notre réseau.

### 1.3.2 Tomographie des réseaux

La tomographie de réseaux est une méthode qui étudie les caractéristiques internes d'un réseau (c'est-à-dire les liens et noeuds ON/OFF, la bande passante, la congestion du réseau) en utilisant les mesures point-à-point (entre noeuds) obtenues à partir des sondes placées dans ce réseau et en supposant que le réseau est modélisable (on peut définir un modèle avec l'estimateur du maximum de vraisemblance ou l'inférence bayésienne). Il fait aussi la prédiction de la topologie du réseau. Cette méthode est utilisée en l'absence de système de supervision fiable pour identifier les caractéristiques des liens et faire un diagnostic de ce dernier (quel noeud/liens est indisponible, congestionné ou ajouté) car il est impossible de contrôler les flux et l'état des équipements. Les problèmes, résolus par la tomographie des réseaux, sont regroupés en 3 catégories :

#### 1.3.2.0.0.1 l'estimation des paramètres (caractéristiques) d'un lien :

L'article de *Ghita et al.* [?] se propose de découvrir la congestion des liens dits "correlés" dans un réseau informatique. Un lien entre deux noeuds du réseau est une connexion logique au niveau de la couche 3 du protocole TCP/IP et deux liens sont corrélés s'ils appartiennent au même domaine ou sous-réseau. Pour réaliser cet algorithme, il considère que la topologie du réseau et le degré de corrélation entre les liens sont connus et qu'il existe un trafic unicast entre les noeuds dans le réseau (ce trafic est désigné par chemin). Il énonce quatre hypothèses pour l'expérimentation de l'algorithme : (i) l'ensemble des chemins reste inchangé durant chaque simulation; (2i) chaque chemin est congestionné si au moins un lien du chemin est congestionné; (3i) le comportement de congestion de chaque lien pendant chaque simulation est modélisée par un processus aléatoire stationnaire; (4i) deux ensembles de corrélations disjoints ne doivent pas être traversés par les mêmes chemins.

Différentes méthodes ont été proposées et validées mais elles diffèrent de l'algorithme *Ghita et al.* [?] par les caractéristiques des liens fournis. En effet, les méthodes initiales se basent sur les corrélations temporelles, chacune parfaitement corrélée, envoyée par des packets multicast [?, ?, ?, ?]. Tous les taux de pertes des liens sont statistiquement identifiables dans une topologie en arbre [?]. Cependant le multicast n'est pas largement déployé et les groupes de paquets unicast exigent un développement substantiel et un coût d'administration élevé. D'où il est moins aisé de sélectionner les corrélations temporelles.

L'ensemble des méthodes qui suivent [?, ?, ?, ?] utilisent seulement des mesures unicast point-à-point (c'est-à-dire des mesures sur les liens) dans le simple but d'identifier les congestions de liens. Les méthodes booléennes de tomographie de réseaux considèrent des hypothèses supplémentaires [?, ?] pour identifier les liens congestionnés en trouvant le plus petit ensemble de liens qui peut être expliqué par ces mesures. Ces hypothèses sont : (i) les liens sont indépendants; (2i) les liens sont congestionnés équiprobablement; (3i) le nombre de liens congestionnés est faible. Toutes les précédentes méthodes [?, ?, ?, ?, ?] se basent sur l'indépendance des liens c'est-à-dire l'hypothèse (i).

#### 1.3.2.0.0.2 La prédiction de topologie :

L'objectif de cette méthode est d'identifier l'arbre de la topologie connectant un serveur aux autres machines du réseau. L'idée est d'utiliser une fonction croissante (ou monotone) du nombre de liens partagés entre deux noeuds ou le maximum de vraisemblance pour trouver l'arbre.

L'article de *Coates and al.* [?] suppose que les mesures de la couche 2 (protocole TCP/IP) des machines clientes sont assez fiables pour découvrir le réseau. Les mesures sont collectées à l'aide de l'outil Traceroute. Ensuite l'auteur définit un modèle basé sur les chaînes de Markov de Monte Carlo pour déterminer les caractéristiques des liens du réseau et déduire les topologies probables. Enfin il propose un critère global du seuil maximum pour l'identification de topologie contrairement aux autres travaux [?, ?] qui emploient des stratégies semi-optimales de fusion des liens.

#### 1.3.2.0.0.3 La densité du trafic entre émetteur/recepteur :

Un problème de tomographie de réseaux qui retient notre attention est l'estimation de la matrice de trafic qui prévoit le volume de flots entre les noeuds point-à-point à partir des mesures [?, ?]. En effet, les variables inconnues sont les volumes de flots et nous suivons une certaine loi de probabilité.

Les corrélations de flots ont été étudiées par *Singal et Michailidis* [?] et ils montrent que, sous certaines classes de dépendances, les moments d'ordre  $n$  de ces volumes de flots sont identifiables à partir des mesures de liens pour  $n \geq 2$ . Cette approche diffère de 2 aspects par rapport l'estimation des caractéristiques de liens. Premièrement, l'estimation des caractéristiques s'intéresse aux mesures point-à-point c'est-à-dire sur les liens tandis que le calcul du trafic point-à-point doit être estimé dans la densité de trafic [?, ?, ?]. Deuxièmement, l'estimation des caractéristiques utilise les variables booléennes et la connaissance des valeurs des variables des lois de distributions n'est pas nécessaire comme cela se fait dans la densité de trafic. La connaissance des valeurs des paramètres engendre la restriction de l'extension des résultats théoriques [?].

**Conclusion :** la tomographie de réseaux n'est pas appropriée pour notre sujet de recherche car nos mesures d'arcs déterminent les caractéristiques de ces arcs et les noeuds sont inconnus empêchant la découverte de réseau qui est l'étape nécessaire pour débiter la tomographie de réseaux.

### 1.3.3 Reconstruction de la topologie par les séries temporelles

Le brevet de *Chaudhary et al.* [?] décrit comment trouver le graphe induit par un réseau de canalisations de fluides en se servant des mesures de capteurs de différentes stations qui émettent des fluides. Il considère le réseau comme un arbre dans lequel la racine est une station comprenant un compresseur. La présence du compresseur induit un délai de livraison entre la station source et les stations de livraison (délai d'acheminement du fluide entre noeuds du réseau). Le délai est dû au temps mis par le fluide pour atteindre la station de livraison.

Le traitement des mesures [?, ?, ?] porte sur la recherche de pics, la suppression d'anomalies (observations non conformes à un motif dans le dataset de données) et le lissage des données. Les données obtenues identifient la relation d'adjacences entre les noeuds. Le traitement des retards temporels déterminent la distance entre des noeuds du réseau. L'analyse des séries temporelles par paires détermine la causalité entre les capteurs associés. La causalité est la recherche d'évènements identiques observés dans deux séries de mesures. En d'autres termes, la causalité consiste à savoir si les évènements, observés dans une série de mesures, se reproduisent dans une autre série. Un modèle de causalité basé sur les séries temporelles des stations est défini comme une régression

multiple avec un modèle de Granger. À partir de ce modèle, on calcule les différents retards  $X(t)$  et leur coefficients. On définit également un modèle de pénalisation basé sur une régression LASSO afin de filtrer les relations de causalité et obtenir un graphe creux (sparse). Le graphe de causalité est alors un arbre.

Les auteurs *Chaudhary et al.* proposent un script s'exécutant récursivement sur les noeuds du réseau en choisissant, à chaque étape, les noeuds qui ne sont pas des puits. En se servant du graphe de causalité (corrélation) et des retards de propagation, il détermine les voisins du noeud  $X_i$  et supprime les noeuds voisins de  $X_i$  ayant une causalité entre eux car le réseau est un arbre. La méthode décrite ici est la construction du graphe itérativement à partir des sous-arbres du graphe. En effet, le nombre de sous-arbres est l'ordre dans lequel on a supprimé les noeuds puits. Cette méthode est réalisable en  $O(n^2)$  car chaque sommet est traité une fois et la recherche de son voisinage est en  $O(n)$ .

Cette méthode est inadaptée pour un DAG car sa complexité est en  $O(n^n)$ . Elle ne s'exécute que sur des arbres.

**Conclusion :** l'étude bibliographique effectuée sur la découverte de topologie montre que la recherche sur ce sujet est très active dans le domaine des réseaux informatiques, précisément dans la détection de noeuds/liens congestionnés. Toutefois, dans le domaine énergétique, les rares travaux réalisés sur ce sujet s'accroissent autour de la découverte de topologie par la reconstruction des sous-graphes en supposant que les noeuds du réseau sont connus, que certains liens sont absents, que les mesures sont influencées par les incidents et que des erreurs peuvent être présentes sur nos données. Ces travaux s'avèrent moins pertinents pour notre problématique parce que les résolutions proposées s'appuient sur des hypothèses qui sont différentes des nôtres. En effet, nous connaissons que les liens et les mesures qui circulent sur ces liens. Mais nous ignorons exactement les extrémités de ces liens. Par ailleurs, ces mesures suivent des lois physiques qui impliquent la propagation d'événements dans ces réseaux. Notre problématique est alors de déterminer les extrémités partagées entre les liens grâce aux lois et aux mesures physiques et aussi à la théorie des graphes.

## 1.4 Problème de découverte de topologie électrique

Notre problème est de découvrir la topologie d'un réseau électrique dont on ignore les sommets et on ne connaît que les flots dans chaque arc.

### 1. Données :

Nous avons donc

- Un ensemble d'arcs distincts 2 à 2 du graphe  $G$  dont les extrémités **initiales** et **finales** sont inconnues.  

$$A = \{a_1, \dots, a_m\}$$
- À chaque arc sont associées des séries de mesures  $M(a_i)$ .  

$$\forall a \in A, M(a) = (gp_a^x)_{x \in GP^a}$$
- Chaque série de mesures  $gp_a^x$  est de norme  $T^{a,x}$  vérifiant la remarque de la section 1,  
 $x \in GP^a \subset GP$
- $\forall t < T^{a,x}, gp_a^x[t]$  respecte les règles dans  $R$ .

**2. Objectif :**

Déterminer la topologie du graphe  $G$  à partir des flots des arcs  $M(a_i)$  et des règles  $R$ .

**3. Approche :**

Notre objectif est de déduire la topologie du réseau électrique représentée sous la forme d'un graphe de flots. Notre approche se subdivise en 2 étapes :

- (a) La première étape consiste à la recherche d'arcs ayant des extrémités communes. Pour ce faire, nous calculons la similarité entre les paires de mesures d'arcs pour chaque grandeur dans le but de déterminer les arcs corrélées puis nous construirons la matrice de corrélation.
- (b) La seconde étape est la construction du graphe à partir de la matrice de corrélation.

## 1.5 Conclusion du chapitre

Dans ce chapitre, nous avons montré que le réseau électrique se modélise par un graphe de flots dont les sommets sont des équipements, les arcs sont les câbles électriques unidirectionnels et les flots par les mesures des équipements. Par convention, ces mesures sont portées par les arcs incidents entrants dans chaque sommet du graphe. Chaque mesure est associée à une grandeur physique et à chaque grandeur est définie une capacité sur l'arc. Nous avons redéfini la loi de conservation  $R$  en fonction de nos mesures, de nos grandeurs et nos capacités.

Dans ce graphe, nous connaissons les liens et les mesures qui circulent sur ces liens. Mais nous ignorons exactement les extrémités de ces liens. Par ailleurs, ces mesures suivent des lois physiques qui impliquent la propagation d'évènements dans ces réseaux. Notre problème est alors de déterminer les extrémités partagées entre les liens grâce aux lois  $R$  et aux mesures physiques et aussi à la théorie des graphes.

Une étude bibliographique a été réalisée sur la découverte de topologie. Les travaux concernent principalement la reconstruction de topologie dans le domaine informatique à partir de sondes et de protocoles de réseau. Ces travaux s'accroissent sur la supervision et l'administration du réseau et aussi sur la recherche des caractéristiques des éléments du réseau. Dans le domaine énergétique, nous avons trouvé un brevet qui reconstruit des sous-graphes du réseau à partir de la propagation des incidents. Tous ces travaux supposent que nous connaissons l'état de tous les éléments du réseau. Ce qui est contraire à notre problématique.

Nous avons proposé deux approches pour résoudre notre problème. La première approche consiste à déterminer la corrélation entre les arcs à partir des mesures et des règles de flots puis à construire une matrice de corrélation. La seconde approche découvre le réseau en effectuant certaines transformations sur la matrice de corrélation.