Estimation de la Matrice de Trafic en temps réel par filtres de Kalman

Augustin Soule, Kavé Salamatian
LIP6-UPMC

***Abstract*—**
Dans ce travails nous développons une nouvelle approche pour l'estimation de matrices de trafic. Nous construisons un modèle d'états pour représenter les flots de paires Origine Destination d'un grand réseau d'opérateur. Ce modèle est suffisamment riche pour capturer les corrélations spatio-temporelles des flots. Nous appliquons ensuite un filtre de Kalman a ce système linéaire dynamique. Cette approche diffère des approches précédentes car tandis que les approches précédentes étaient focalisées uniquement sur l'ingénierie de trafic, nous nous concentrerons sur la détection d'anomalies. Notre modèle est suffisamment riche pour définir le comportement normal d'un flot et ainsi détecter les comportements anormaux. De plus la simplicité du modèle permet d'effectuer les calculs en temps réel. Nous avons ensuite validé cette méthode à l'aide de traces d'un grand fournisseur d'accès américain.

***Index Terms*—Traitement du signal et Télétrafic Internet, Métrologie des réseaux, statistiques, Filtre de Kalman, Matrices de Trafic, Système linéaires dynamiques,...**

I. Introduction

La gestion des réseaux d'opérateurs est une tache de plus en plus complexe. Le nombre et la complexité des équipements actifs a incroyablement augmenté ces dernières années. Dans le même temps le trafic s'est complexifié allant jusqu'à contenir une part non négligeable de trafic hostile. Les opérateurs se reposent sur des outils de surveillance pour les aider à gérer ce gigantesque système que constitue l'Internet.

Une vue d'ensemble de la demande de trafic dans un domaine est souvent appelé Matrice de Trafic. La matrice de trafic est la représentation sous forme matricielle du volume de trafic entre une paire de noeud Origine et Destination (OD). Il est possible de définir la demande de trafic entre les liens, les routeurs ou encore les POP (point of presence) [1]. Dans ce travail nous nous concentrons sur la demande de trafic entre POP. C'est en effet cette granularité qui permet une vue globale du réseau de l'opérateur. La méthode est néanmoins applicable a tout les niveaux d'aggégation souhaité.

La mesure directe de la matrice de trafic s'avère complexe car les routeurs ne supportent pas tous cette fonctionnalité. De plus la surcharge au niveau du réseau pour le rapatriement de telles statistiques est coûteuse et nécessite l'envoi de volume important de données. Il convient donc d'inférer les matrices de trafic à partir des données que l'opérateur collecte déjà. L'opérateur collecte dans son centre de contrôle beaucoup de données à l'aide du protocole SNMP (Simple Network Managment Protocol). Par exemple il collecte à des intervalles de 5 minutes le volume de trafic traversant chaque lien du réseau. Mais le trafic traversant un lien est la somme de plusieurs composantes de la matrice de trafic. L'inférence des matrices de trafic consiste à utiliser ce genre de données de volume sur un lien afin d'estimer les éléments composants chaque terme. Ce domaine de recherche à été très actif ces dernières années[1], [2], [3]. Les approches précédentes d'estimation de la matrice de trafic ne sont pas forcément capables de donner une estimation en temps réel de la matrice de trafic, que ce soit pour des problèmes de modèle ou encore de complexité. Le travail présenté dans cet article présente une méthode permettant une estimation temps réels de la matrice de trafic.

Nous allons utilisons les mesures effectuées sur la partie européenne du réseau de Sprint pour développer un modèle linéaire dynamiques d'espace d'état. Ce modèle décrit l'évolution temporelle de la matrice de trafic ainsi que la corrélation spatiale de celle-ci . Mais les valeurs de la matrice de trafic n'étant pas directement observable, nous utilisons un filtre de Kalman pour estimer et prédire ces valeurs en utilisant les volumes de trafic observé sur les liens. Cette approche permet un suivi, même à des échelles de temps très petites, en temps réel de la matrice de trafic .

II. Méthode

Un modèle réaliste, *i.e.* capable de prédire la dynamique du réseau, doit intégrer la corrélation des données observées, à l'aide de SNMP, avec chacune des paires OD qui la constitue. Ce modèle doit capturer l'évolution temporelle de chaque pair OD qui traverse le réseau. Soit X_t un vecteur contenant les valeurs de la matrice de trafic. Nous supposons que la dynamique temporelle des éléments de la matrice de trafic est capturée par un modèle Linéaire et invariant dans le temps de la forme $X_{t+1} = CX_t + W_t$, où la matrice C représente la corrélation spatiale et temporelle de la matrice de trafic, W_t représente un bruit contenant l'imprécision du modèle. D'autre part le volume de trafic sur chaque lien (les données SNMP) que nous représentons par un vecteur Y_t est lié au vecteur X_t par une relation linéaire de la forme $Y_t = AX_t + V_t$, où A représente la matrice de routage et V_t représente un bruit de mesure.

$$\begin{cases} X_{t+1} & = CX_t + W_t \\ Y_t & = AX_t + V_t \end{cases} \quad (1)$$

On obtient ainsi un modèle d'état classique sur lequel un filtre de Kalman peut être appliqué. Le filtre de Kalman est le filtre linéaire optimal au sens du Minimum de Variance de l'Erreur d'Estimation de l'état X_t en utilisant l'observation Y_t.

En conséquent l'estimation de la matrice de trafic ce fait en deux étapes :

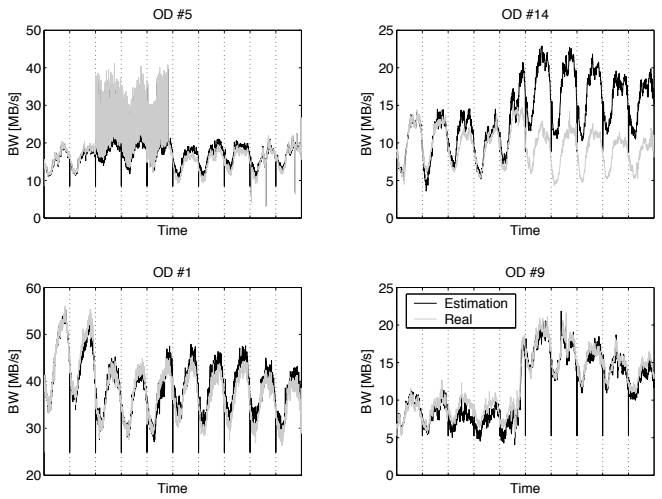

Fig. 1. Paire Origine Destinatioin Réelle (Gris clair) et inférée (Gris foncé) à l'aide du filtre de Kalman. Une seule calibration est utilisé lors du premier jour.

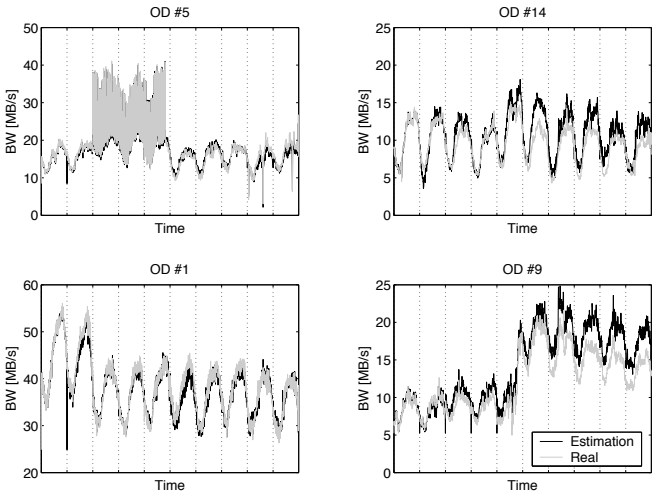

Fig. 2. Paire Origine Destinatioin Réelle (Gris clair) et inférée (Gris foncé) à l'aide du filtre de Kalman. Plusieurs calibrations sont utilisés au cours des dix jours.

– Dans une première étape il faut passer par une estimation de la matrice C décrivant l'évolution temporelle de la matrice de traffic ainsi que les matrices de covariance des bruits W_t et V_t. Cette calibration s'effectue sur une mesure partielle de la matrice de trafic. Pour calibrer ces paramètres, nous avons implémenté la méthode décrite dans [4] qui consiste à utiliser une méthode EM pour touver les paramètres les plus probable.
– Dans la seconde étape, le modèle obtenu dans la première étape est utilisé et un suivi de l'état du système, *i.e.* de la matrice de trafic, est effectué en appliquant un filtre de Kalman prenant en entrée les volumes de trafic observés sur les liens du réseau. Ce suivi permet au filtre de Kalman de faire évoluer les variables d'état en fonction des observations et donc d'être capable d'estimer la dynamique des paires OD. De plus il est possible de calculer l'innovation de la demande de trafic. Il est alors possible en regardant l'innovation de détecter que le modèle n'est plus valable et qu'une nouvelle étape de calibration est nécessaire.

III. Validation de la méthode

Dans cette section, nous allons présenter les résultats numériques obtenus en appliquant notre méthode sur des données réelles. Les données ont été collectés en utilisant NetFlow [5] sur tout les routeurs connectant les clients au coeur du réseau européen de Sprint. 1 paquet sur 250 à été collecté. Nous avons ensuite regroupé ces valeurs toutes les 1à minutes. De plus nous avons conjointement enregistré les données de routage pour reproduire la matrice de routage. Nous avons ainsi reconstruit la matrice de trafic entre les 11 POP soit 121 paires Origine Destination. Nous allons présenter brièvement les résultats obtenus à l'aide du filtre de Kalman.

Dans la figure 1 nous montrons un exemple de la qualité de notre approche. Nous présentons ainsi 4 paires Origine Destination différentes. Pour les paires OD 1, 5, 9, l'estimation est correcte tout au long des 10 jours. Tandis que pour la paire OD 14, le 5^{e}jour elle dévie complètement. En utilisant l'innovation, nous pouvons détecter automatiquement ce type de comportement et calibrer le modèle à chaque fois que les paires OD changent de comportement rendant le modèle invalide. Nous présentons dans la figure 2 le résultat de l'estimation pour les même paires OD après avoir recalibré le modèle.

IV. Conclusions

Dans cette article nous avons présenté l'application des filtres de Kalman au suivi et à l'estimation en temps réel de la matrice de trafic dans un réseau d'opérateur. La validation de cette méthode a été effectuée sur un des mesures obtenues dans une situation réaliste sur un réseau opérationel. Les résultats obtenus montrent que cette application semble prometteuse. En particulier l'estimation par filtre de Kalman semble particulièrement adaptée à la détection d'anomalies. C'est vers cette direction que nos recherches actuelles nous mènent.

References

[1] Alberto Medina, Nina Taft, Kavé Salamatian, Supratik Bhattacharyya, and Christophe Diot, "Traffic matrix estimation : existing techniques and new directions," in *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*. 2002, pp. 161–174, ACM Press.

[2] Yin Zhang, Matthew Roughan, Carsten Lund, and David Donoho, "An information-theoretic approach to traffic matrix estimation," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. 2003, pp. 301–312, ACM Press.

[3] Augustin Soule, Antonio Nucci, Rene Cruz, Emilio Leonardi, and Nina Taft, "How to identify and estimate the largest traffic matrix elements in a dynamic environment," in *ACM Sigmetrics*, New York, 2004.

[4] Zoubin Ghahramani and Geoffrey E. Hinton, "Parameter estimation for linear dynamical systems," Tech. Rep. CRG-TR-96-2, University of Toronto, 22˜ February 1996.

[5] CISCO, "Netflow services and applications," 2002.