

Deux applications de processus ponctuels aux réseaux de communication

Olivier DOUSSE, Hung X. NGUYEN, Patrick THIRAN

Faculté Informatique et Communications
EPFL, CH-1015 Lausanne, Suisse

Olivier.Dousse@epfl.ch, Hung.Nguyen@epfl.ch
Patrick.Thiran@epfl.ch

Résumé – Cet article résume des résultats récents obtenus en utilisant des processus ponctuels (plus précisément, des bruits impulsionnels) pour deux applications dans le contexte des réseaux de communication: d’une part, la modélisation du trafic TCP/IP dans les réseaux d’épine dorsale, et d’autre part les propriétés de connectivité des réseaux ad hoc sans fil sous des contraintes de débit.

Abstract – This paper summarizes recent results obtained by using shot noise processes for two applications in communication networks: first, TCP/IP traffic models for backbone networks; and next the study of connectivity under rate constraints in wireless ad hoc networks.

1 Introduction

Les bruits impulsionnels (« shot noise ») forment une classe de processus aléatoires particulièrement utiles pour modéliser une grande catégorie de signaux rencontrés en biologie, physique ou informatique. Dans cet article, nous développons deux applications récentes de ces processus dans le contexte des réseaux de communication.

Les processus considérés dans cet articles sont des bruits impulsionnels de Poisson (voir par exemple [3]), qui s’écrivent comme des sommes

$$R(t) = \sum_{n \in \mathbb{Z}} X_n(t - T_n)$$

où $\{T_n, n \in \mathbb{Z}\}$ est une suite de variables aléatoires représentant les points d’un processus de Poisson sur la ligne \mathbb{R} (pour la première application) ou sur le plan \mathbb{R}^2 (pour la seconde application), et où $\{X_n(t), n \in \mathbb{Z}\}$ est une suite de fonctions aléatoires réelles (appelées impulsions), que nous supposons ici indépendantes et identiquement distribuées i.i.d définies sur \mathbb{R} ou \mathbb{R}^2 , et indépendantes de la suite $\{T_n\}$.

La première application, développée à la section 2, est la modélisation des flux TCP/IP dans les réseaux d’épine dorsale Internet. Dans ce cas, la fonction $X_n(\cdot)$ constituant les impulsions représente le débit instantané d’un flux TCP/IP, tandis que le processus d’arrivée des impulsions aléatoires est formé par la séquence des temps d’arrivée $\{T_n\}$ des flux. Le processus ponctuel est donc uni-dimensionnel. On s’intéresse alors aux statistiques de deuxième ordre des flux TCP/IP.

La seconde application, développée à la section 3, concerne les réseaux “ad hoc” sans fil. Contrairement aux réseaux sans fil de type cellulaires, les utilisateurs (noeuds) d’un réseau ad hoc communiquent entre eux de proche en proche, sans passer par une infrastructure d’antennes fixes. Si la destination ne peut être atteinte directement à partir de la source, la communication s’établit en plusieurs sauts, les noeuds intermédiaires placés sur le chemin entre source et destination jouant alors le rôle de relais. Le processus ponctuel est ici formé par la position

des noeuds, et est en général bidimensionnel. La condition pour l’existence d’une communication directe entre deux noeuds du réseau est qu’elle puisse s’effectuer à un débit minimal fixé. Deux facteurs limitent ce débit: d’une part, l’atténuation du signal avec la distance séparant les noeuds, et d’autre part, les interférences créées par les communications concurrentes. Le calcul exact du débit qui peut être obtenu pour un tel canal multi-utilisateurs reste un problème ouvert en théorie de l’information, ce qui nous amène à faire des hypothèses supplémentaires. La première est assez pessimiste. Elle consiste à considérer les communications interférentes comme une source de bruit, s’ajoutant au bruit de fond, et modélisée par un bruit impulsionnel. Le graphe de connectivité résultant de ce modèle sera présenté à la section 3.1. La seconde est au contraire optimiste, et consiste à traiter un modèle de canal particulier, appelé “cannal relai” (relay channel), dans lequel une seule paire source-destination est active, tous les autres noeuds pouvant alors coopérer comme relais éventuels pour acheminer dans l’information de la source vers la destination. La contribution de ces relais peut à nouveau être modélisée par un processus ponctuel, comme nous le verrons à la section 3.2.

2 Application 1: modélisation du trafic TCP/IP

L’analyse et la modélisation de trafic dans l’Internet peut se faire à plusieurs niveaux de granularité différents, selon les données à disposition et le but poursuivi. Le niveau de granularité le plus fin est celui du paquet IP, mais il mène à des difficultés de modélisation à cause des la forte dépendance statistique entre paquets. Le niveau supérieur groupe les paquets en flux, définis comme les ensembles de paquets ayant les mêmes adresses de source et destination, mêmes numéros de port, et même numéro de protocoles. Les paquets à l’intérieur d’un flux sont fortement corrélés, mais les flux eux-mêmes sont nettement moins corrélés entre eux (en tout cas dans le coeur des

réseaux d'épine dorsale). D'autre part, pour un grand nombre d'applications, l'unité d'intérêt pour l'utilisateur est le flux plutôt que le paquet individuel. La modélisation du trafic à ce niveau de granularité a connu un intérêt important ces dernières années [2, 7, 8], et c'est celui que nous adoptons dans cet article. Enfin, signalons qu'on peut encore monter d'un niveau, en groupant les flux par session [10].

En général, les opérateurs dimensionnent les réseaux IP d'épine dorsale (« backbone IP networks ») de manière à éviter toute congestion à l'intérieur de ceux-ci, les goulots d'étranglement étant déplacés au réseau d'accès. Alors qu'un lien congestionné a pour effet de corrélérer les débits des flux TCP/IP l'empruntant, à cause du mécanisme AIMD (Additive Increase, Multiplicative Decrease) opéré par la source, un lien non congestionné n'introduit pas une telle corrélation. D'autre part, un grand nombre de flux d'origines et destinations variées sont multiplexés sur un lien du backbone, ce qui rend la séquence de leurs temps d'arrivées sur le lien proche d'une distribution de Poisson.

Désignons par $\{T_n, n \in \mathbb{Z}\}$ la séquence des temps d'arrivée des flux, et par $\{X_n(t), n \in \mathbb{Z}\}$ la suite des impulsions, représentant l'évolution du débit de chaque flux. Les observations du paragraphe précédent nous amènent à faire l'hypothèse que les variables T_n suivent un processus de Poisson et que les impulsions X_n sont i.i.d. Ces hypothèses sont assez bien vérifiées en pratique [2], et nous conduisent à décrire le débit instantané $R(t)$ de l'agrégat des flux sur le lien comme le processus ponctuel

$$R(t) = \sum_{n \in \mathbb{Z}} X_n(t - T_n). \quad (1)$$

Ce modèle, introduit dans [2] et validé sur des traces réelles du réseau IP de Sprint, nous permet de calculer les moments du premier et deuxième ordre de $R(t)$, si on dispose de la séquence des durées des flux $\{D_n, n \in \mathbb{Z}\}$, de la séquence des tailles des flux $\{S_n, n \in \mathbb{Z}\}$ et de la distribution des pertes de paquets. Notons que

$$S_n = \int_0^{D_n} X_n(t) dt.$$

Alors que d'autres méthodes [4] de calcul de ces moments sont basées sur une modélisation fine du protocole TCP/IP à l'aide de chaînes de Markov, le modèle (1) est plus simple et moins détaillé, se contentant d'une modélisation du protocole par une équation aux récurrences du premier ordre. Il donne néanmoins d'excellentes bornes, comme nous le verrons ci-dessous.

Pour un flux TCP/IP, le mécanisme AIMD permet de déterminer une borne assez étroite de la fonction d'auto-covariance de $X_n(t)$. Le protocole TCP limite le nombre maximal de paquets d'un flux TCP/IP en transit dans le réseau par une fenêtre coulissante, qui avance d'un paquet chaque fois que l'acquittement d'un paquet précédemment envoyé a été reçu. L'algorithme AIMD consiste à adapter la taille de cette fenêtre en fonction des pertes de paquets. Grosso modo, si on néglige les effets de "slow-start" et de "congestion avoidance", la fenêtre est augmentée linéairement chaque fois que l'acquittement attendu est reçu, et est divisée par deux chaque fois qu'un acquittement n'est pas reçu en dehors d'un certain intervalle de temps (time-out interval): c'est le mécanisme AIMD. En le formulant en terme d'évolution du taux instantané $X_n(t)$ du n ème flux

sur un interval infinitésimal $[t, t + dt]$ au lieu de fenêtre coulissante, on trouve

$$X_n(t + dt) - X_n(t) = -\frac{X_n(t)}{2} p_l(t) + \alpha dt, \quad (2)$$

où α désigne un facteur inverse au temps aller-retour d'un paquet, et $p_l(t)$ est la probabilité de perte d'un paquet dans l'intervalle $[t, t + dt]$. Soit $r_d(\tau) = \mathbb{E}[X_n(t)X_n(t + \tau) \mid D_n = d]$ la fonction d'autocorrélation des flux conditionnée sur leur durée. En supposant que les pertes suivent un processus de Poisson, d'intensité constante λ_l , on déduit de (2) l'équation différentielle

$$\frac{dr_d}{d\tau}(\tau) = -\frac{\lambda_l r_d(\tau)}{2} + \alpha \mathbb{E}[X_n(t) \mid D_n = d]. \quad (3)$$

pour $\tau > d$. On calcule [1, 2] que $\mathbb{E}[X_n(t) \mid D_n = d] = 2\alpha/\lambda_l$ et $r_d(0) = 4\mathbb{E}^2[S_n \mid D_n = d]/3d^2$, dont on déduit la solution de (3) pour $|\tau| > d$:

$$r_d(\tau) = \frac{\mathbb{E}^2[S_n \mid D_n = d]}{d^2} \left(1 + \frac{1}{3} \exp\left(\frac{-\alpha|\tau|d}{\mathbb{E}[S_n \mid D_n = d]}\right) \right).$$

A partir de ce résultat, on obtient une borne la fonction d'auto-covariance $C_R(\tau)$ sur l'entière distribution des durées des flux. En notant $1/\lambda = \mathbb{E}[T_n]$ la temps moyen entre arrivées de flux, on trouve [9]

$$C_R(\tau) \leq \lambda \mathbb{P}(D_n > |\tau|) \left(\frac{1}{3} \mathbb{E} \left[\frac{S_n^2(D_n - |\tau|)}{D_n^2} e^{-\alpha|\tau|\frac{D_n}{S_n}} \mid D_n > |\tau| \right] + \mathbb{E} \left[\frac{S_n^2(D_n - |\tau|)}{D_n^2} \mid D_n > |\tau| \right] \right). \quad (4)$$

Cette borne comporte deux termes. Le premier décroît exponentiellement avec τ , le second seulement linéairement. Il existe donc deux régimes différents: pour de faibles valeurs de τ , le premier terme exponentiel domine la somme, tandis que le second terme linéaire l'emporte pour de grandes valeurs de τ . C'est en effet le comportement observé à la figure 1, qui compare la borne avec la fonction $C_R(\tau)$ obtenue par simulation (Simulateur ns, les paramètres de simulation sont décrits dans [9]). On remarque l'excellente approximation du résultat de simulation par l'expression (4), dans laquelle par ailleurs le taux de pertes λ_l n'apparaît pas.

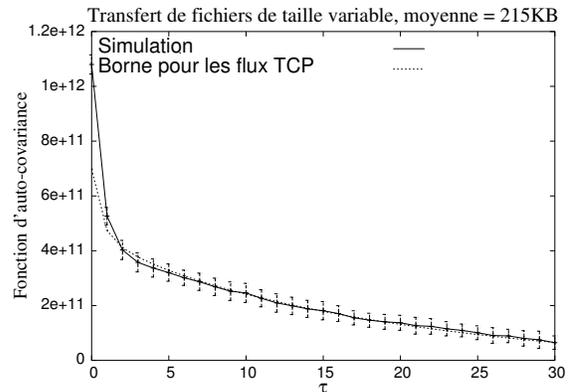


FIG. 1: Résultats de simulation montrant la fonction d'auto-covariance $C_R(\tau)$ en fonction de l'écart temporel τ .

3 Application 2: connectivité et débit dans les réseaux ad hoc

Dans cette seconde application, nous considérons des réseaux sans fil ad hoc. Nous supposons que les positions des noeuds du réseau ad hoc suivent un processus de Poisson \mathcal{N} sur \mathbb{R}^2 d'intensité $\lambda > 0$.

La puissance du signal émis par le noeud i et reçue par le noeud j est $PL(\mathbf{x}_i - \mathbf{x}_j)$, où \mathbf{x}_i et \mathbf{x}_j sont les positions respectives des noeuds i et j dans le plan, $L(\cdot)$ est la fonction d'atténuation des ondes radio dans le milieu, et P est la puissance émettrice (supposée identique pour tous les noeuds).

3.1 Modèle basé sur le rapport signal sur bruit et interférences

Nous supposons que le noeud i peut transmettre des données au noeud j si le signal reçu par j est assez puissant, en comparaison au bruit. Ce dernier est la somme d'une composante constante N_0 , représentant le bruit thermique de fond, et d'une composante représentant les interférences générées par les transmissions correspondantes. Formellement, cette condition s'écrit

$$\frac{PL(\mathbf{x}_i - \mathbf{x}_j)}{N_0 + \gamma P \sum_{k \neq i, j} L(\mathbf{x}_k - \mathbf{x}_j)} \geq \beta, \quad (5)$$

où β est le rapport signal sur bruit requis pour pouvoir décoder l'information transmise. Le coefficient γ est l'inverse du gain du système, il pondère l'effet des interférences et dépend de l'orthogonalité des codes utilisés pendant les transmissions simultanées. Un lien, représentant une communication directe entre les noeuds i et j , existe si (5) est valable dans les deux directions i vers j et j vers i . Cette condition définit un graphe $\mathcal{G}(\gamma, \lambda)$, dont certaines propriétés sont analysées dans [5]. En particulier, l'existence d'une transition de phase est démontrée: lorsque la densité spatiale des noeuds λ est suffisamment élevée, il existe une valeur $\gamma^*(\lambda)$, appelée seuil de percolation, telle que si $\gamma < \gamma^*$ alors le graphe comporte une composante de noeuds connectés de taille et d'étendue infinie (on parle de phase sur-critique); tandis que si $\gamma > \gamma^*$, alors toutes les composantes de noeuds connectés sont de taille presque sûrement finie (phase sous-critique). Les figures 2 et 3 montrent le graphe $\mathcal{G}(\gamma, \lambda)$, pour la même réalisation du processus \mathcal{N} , et deux valeurs différentes de γ . La première, $\gamma = 0$, revient à négliger les interférences, et donne un graphe sur-critique, tandis que la seconde $\gamma = 0.02 > \gamma^*(\lambda)$ tient compte des interférences, et rend le graphe sous-critique. Néanmoins, en augmentant la densité λ de noeuds, on rend à nouveau le graphe sur-critique (fig 4).

Nous remarquons que le terme d'interférence est un bruit impulsif sur \mathbb{R}^2 , qui s'écrit

$$I(\mathbf{x}) = P \sum_{i \in \mathcal{N}, \mathbf{x}_i \neq \mathbf{x}} L(\mathbf{x} - \mathbf{x}_i), \quad (6)$$

et dont la valeur influence fortement le seuil de percolation $\gamma^*(\lambda)$. En particulier, il est nécessaire que l'espérance de $I(\mathbf{x})$ soit finie, ce qui entraîne que pour des fonctions d'atténuation isotropiques $L(\mathbf{x}) = l(\|\mathbf{x}\|)$,

$$\int_y^\infty l(t) dt < \infty, \text{ pour } y \text{ suffisamment grand.}$$

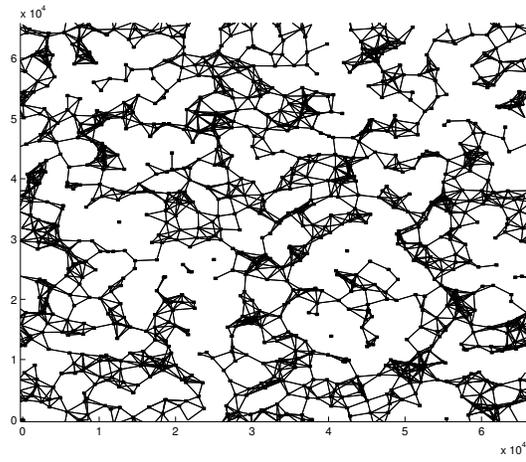


FIG. 2: Un exemple de graphe $\mathcal{G}(0, \lambda)$ avec $\gamma = 0 < \gamma^*(\lambda)$.

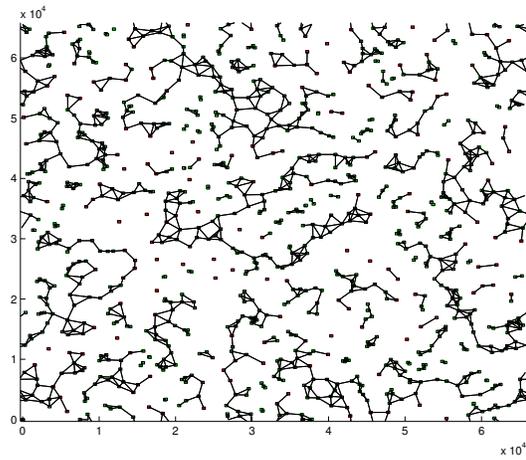


FIG. 3: Un exemple de graphe $\mathcal{G}(0.2, \lambda)$ avec $\gamma = 0.02 > \gamma^*(\lambda)$. Les interférences fragmentent le graphe en un grand nombre de petites cliques.

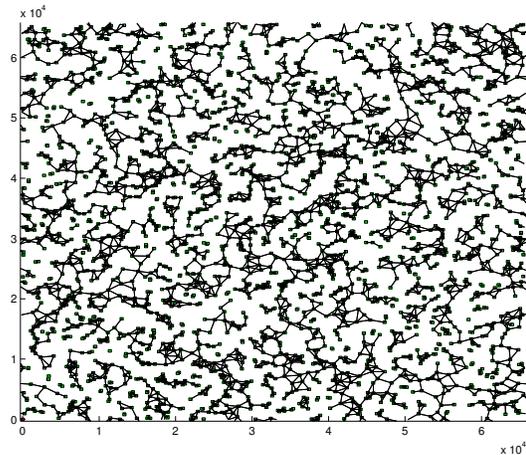


FIG. 4: Un exemple de graphe $\mathcal{G}(0.2, \lambda)$ avec $\gamma = 0.02 < \gamma^*(\lambda)$.

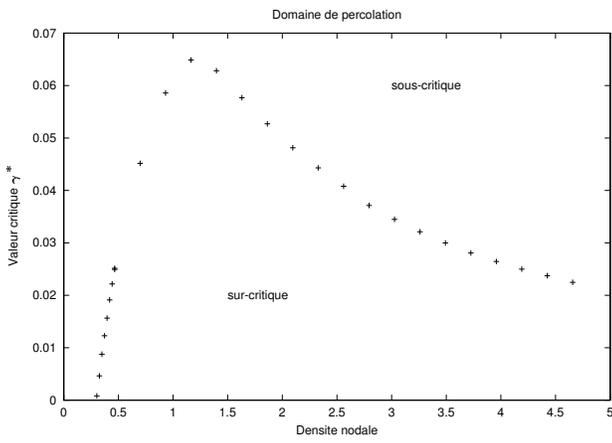


FIG. 5: *Seuil critique $\gamma^*(\lambda)$ en fonction de la densité nodale λ . [Les paramètres de simulation sont $\beta = 1$, $N_0 = 10^4$ et $P = 10^5$.]*

Nous avons montré que cette condition est non seulement nécessaire, mais pratiquement suffisante (si on ajoute les hypothèses naturelles d’avoir une fonction $l(\cdot)$ continue, strictement décroissante, et telle que $\beta N_0/P < l(0) \leq 1$) pour démontrer l’existence d’un seuil de percolation $\gamma^*(\lambda) > 0$ pour des valeurs de λ suffisamment grandes. La forme de la fonction $\gamma^*(\lambda)$ obtenue par simulation est représentée à la figure 5.

3.2 Modèle basé sur le canal-relai

La définition de connectivité basée sur le rapport (5) ne permet pas d’établir directement des résultats validés par la théorie de l’information, à cause de l’hypothèse sur les noeuds traitant les interférences comme du bruit. Pour obtenir un résultat de nature fondamentale en théorie de l’information, on prend la définition suivante: on suppose que deux noeuds sont connectés si et seulement si ils peuvent échanger des données à un débit au moins égal à R , les autres noeuds pouvant servir de relais.

Nous montrons alors dans [6] que la proportion de noeuds pouvant communiquer dans ce modèle est au plus $\hat{\theta}$ avec grande probabilité (i.e., une probabilité tendant vers 1 quand le nombre total de noeuds du réseau $n \rightarrow \infty$), où

$$\hat{\theta} = \mathbb{P}[I(\mathbf{0}) \geq N_0 (e^{2R} - 1)],$$

avec $I(\mathbf{0})$ désignant le même processus ponctuel que (6) à l’origine:

$$I(\mathbf{0}) = P \sum_{i \in \mathcal{N}} L(\mathbf{x}_i).$$

En d’autres termes, même dans le cas le plus favorable où une seule paire de noeuds communique entre eux et où tous les autres noeuds ne peuvent qu’aider à cet échange d’information (au contraire du modèle de la section 3.1), une connectivité à un taux $R > 0$ donné ne peut être garantie à toutes les paires possibles de noeuds du réseau. La démonstration de ce résultat fait appel à des résultats connus sur la capacité des réseaux multi-utilisateurs en théorie de l’information, et à des propriétés de stationnarité et d’ergodisme du processus ponctuel I .

Par contre, si on s’impose de connecter au moins une fraction $0 < \theta < 1$ de noeuds du réseau, il existe un taux $R > 0$, indépendant du nombre de noeuds n dans le réseau, tel que n’importe quel noeud pris dans un sous-ensemble de $n\theta$ noeuds

du réseau puisse communiquer au taux R et avec grande probabilité avec n’importe quel autre noeud de ce sous-ensemble [6]. En d’autres termes, si on laisse tomber une fraction $1 - \theta$ de noeuds particulièrement mal placés, on peut assurer un taux R à tous les autres noeuds. Cette fraction $1 - \theta$ de noeuds éliminés peut être aussi petite que l’on veut, mais non nulle.

4 Conclusion

Cet article a montré l’importance et l’utilité des processus ponctuels de Poisson pour la modélisation de réseaux de communication, que ce soit dans sur support filaire (Application 1) ou sans fil (Application 2).

Remerciements Ce travail est financé en partie par le fonds DICS 1830 de la Fondation Hasler, Berne, et par le Pôle de Recherche National MICS du Fonds National Suisse.

Références

- [1] E. Altman, K. Avrachenkov and C. Barakat, “A stochastic model for TCP/IP with stationary random losses”, *Proc ACM SIGCOMM*, Sep. 2000.
- [2] C. Barakat, P. Thiran, G. Iannaccone, C. Diot et P. Owezarski, “Modeling Internet Backbone Traffic at the Flow Level”, *IEEE Transactions on Signal processing*, vol. 51(8), pp. 2111-2124, Août 2003.
- [3] D. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer-Verlag, New York, 1988.
- [4] D. R. Figueiredo, B. Liu, V. Misra and D. Towsley, “On the Autocorrelation Structure of TCP Traffic”, *Computer Networks Journal*, ‘Special Issue on ‘Advances in Modeling and Engineering of Long-Range Dependent Traffic’, 2002.
- [5] O. Dousse, F. Baccelli et P. Thiran, “Impact of Interferences on Connectivity of Ad Hoc Networks”, *IEEE/ACM Transactions on Networking*, vol. 13(2), pp. 425-436, Avril 2005.
- [6] O. Dousse, M. Franceschetti et P. Thiran, “Information theoretic bounds on the throughput scaling of wireless relay networks”, in *Proc. Infocom’05*, Miami, Mars 2005.
- [7] N. Hohn, D. Veitch and P. Abry, “Cluster processes: a natural language for network traffic”, *IEEE Transactions on Signal processing*, *IEEE Transactions on Signal processing*, vol. 51(8), Août 2003.
- [8] M. Krunz and A. Makowski, “Modelling video traffic using M/G/Infinity input processes: A compromise between Markovian and LRD models”, *IEEE Journal on Selected Areas in Communication*, vol. 16, pp. 733-748, Juin 1998.
- [9] H. X. Nguyen, P. Thiran, G. Iannaccone, C. Diot et C. Barakat, “On the Correlation of TCP Traffic in Backbone Networks”, in *Proc. ISCAS’04*, Mai 2004.
- [10] C. Nuzman, I. Saniee, W. Sweldens and A. Weiss, “A compound model for TCP connection arrivals”, in *Proc. ITC workshop*, Monterey, Sept. 2000.