

De l'efficacité de la modélisation dynamique du trafic des réseaux multiservices, à l'hypothèse de la réaction à la congestion par la compression adaptative du trafic

[From the effectiveness of dynamic traffic modeling of multiservice networks, to the congestion reaction hypothesis by adaptive traffic compression]

Ilunga Mbuyamba Elisée and Tshibengabu Kalonji Simon

Institut Supérieur de Techniques Appliquées (ISTA), Kinshasa, RD Congo

Copyright © 2023 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: In this paper, we first describe the general principle of dynamic modelling, the effectiveness of which in analyzing the traffic of multiservice networks is one of the main objects of this study. By showing how this technique can be applied to obtain the elementary Markovian system: of the queue $M/M/1/\infty$; we partly answer the big problem on the management of the QoS in the case of an isolated node then of a network. We then focus on the principles of traffic control of multiservice networks, drawing inspiration from existing work in the literature and from policies for managing congestion phenomena defined in recommendations I.371 of the ITU-T and ATM-Forum UNI specification V3.1. As a possible solution to this problem, we have evoked the hypothesis of the reaction to congestion by dynamic traffic compression, in the event of the appearance of a phenomenon of congestion exceeding any prediction. The progress of this hypothesis will have the merit of resurfacing and increasing the interest of resorting to compression as a solution to the problem of congestion, in particular for wireless networks; since radio resources are scarce and shared.

KEYWORDS: Dynamic Traffic Modeling, Multiservice Networks, Congestion, Compression, QoS.

RÉSUMÉ: Dans cet article, nous décrivons tout d'abord le principe général de la modélisation dynamique dont l'examen de l'efficacité dans l'analyse du trafic des réseaux multiservices constitue l'un des objets principaux de cette étude. En montrant comment cette technique peut être appliquée pour obtenir le système élémentaire markovien: de la file d'attente $M/M/1/\infty$; nous répondons en partie à la grande problématique sur la gestion de la QoS dans le cas d'un nœud isolé puis d'un réseau. Nous nous intéressons ensuite aux principes du contrôle de trafic des réseaux multiservices, en nous inspirant de travaux existants dans la littérature et de politiques de gestion des phénomènes de congestion défini dans les recommandations I.371 de l'UIT-T et ATM-Forum UNI spécification V3.1. Comme solution possible à cette problématique, nous proposons l'hypothèse de la réaction à la congestion par compression adaptative du trafic, en cas d'apparition d'un phénomène de congestion dépassant toute prédiction. Le progrès de cette hypothèse aura le mérite de resurgir et accroître l'intérêt du recours à la compression comme solution à la problématique de congestion, en particulier pour les réseaux sans fil; étant donné que les ressources radio sont rares et partagées.

MOTS -CLEFS: Modélisation dynamique du trafic, Réseaux multiservices, Congestion, Compression, QoS.

1 INTRODUCTION

Le progrès des services que doivent fournir les réseaux multiservices, les tailles de ces réseaux, l'augmentation du trafic qui s'y rapporte et sa nature de plus en plus complexe appellent à des besoins importants en bande passante et à des

exigences très variés. C'est pour cette raison que nous proposons, une réflexion tendant à cerner les contours de la question suivante: "est-il possible de contrôler la qualité de service autrement qu'en surdimensionnant les réseaux multiservices " ?

En effet, si un changement de trafic se produit dépassant les prédictions, une adaptation simultanée de la topologie du réseau à cette variation brusque de trafic parait en pratique irréalisable. Par contre, l'ingénierie de trafic permet d'adapter le trafic à la topologie du réseau. Dans la thèse de K. Abboud [1], l'expression "Ingénierie de trafic" désigne l'ensemble des mécanismes de contrôle de l'acheminement du trafic dans le réseau afin d'optimiser l'utilisation des ressources et de limiter les risques de congestion. L'objectif de l'ingénierie de trafic est de maximiser la quantité de trafic pouvant transiter dans le réseau, tout en maintenant la qualité de service (bande passante, délai...) offerte aux différents flux de données.

Aussi, nous nous proposons d'examiner l'efficacité de la modélisation dynamique du trafic de réseau multiservice en vue de prévenir les congestions; puis nous étendons la réflexion à l'hypothèse de la réaction à la congestion par la compression adaptative du trafic de manière à limiter les effets de la congestion, en l'occurrence, minimiser l'intensité du volume du flux transitoire d'entrée en l'adaptant à la bande passante des ressources du réseau.

Un tel modèle doit représenter la variation stochastique du trafic des réseaux multiservices et les phénomènes qui en résultent, spécialement l'état de charge au niveau de ses différents composants (routeurs, liens, chemins...) et réagir le cas échéant à la congestion, en adaptant le trafic à l'état du réseau.

En d'autres termes, limiter les pics des débits instantanés d'entrée à la capacité de la bande passante disponible du réseau, au moyen d'une compression adaptative. Le taux de compression doit suivre dynamiquement l'importance de la congestion survenue et, plus ce taux de compression est faible, plus la probabilité de congestion est élevée.

Pour y parvenir, nous présenterons d'abord le problème lié à l'efficacité de la modélisation dynamique de la file M/M/1/∞, dans l'analyse du trafic des réseaux multiservices. Telle qu'étudiée dans [2], l'analyse du processus markovien permet d'obtenir l'expression exacte de l'équation différentielle gouvernant le trafic des réseaux à file d'attente. Malheureusement, cette équation fait intervenir la probabilité d'état transitoire dont l'expression analytique complexe ne peut être calculée. La solution proposée consiste à transformer l'équation différentielle associée à X (t) en une équation différentielle autonome en utilisant une approximation basée sur une relation implicite entre le trafic moyen X (t) et la probabilité d'état transitoire P₀(t).

Et comme contribution, nous enrichissons la discussion en proposant: « l'hypothèse de la réaction à la congestion par une compression adaptative du trafic ». Cette hypothèse présage un pouvoir d'expression important, a priori pour des réseaux de grande taille.

2 THEORIE DIFFERENTIELLE DU TRAFIC

Dans cette partie, nous présentons le principe général de la théorie différentielle du trafic tel que proposé dans [2]. Considérons le système représenté sur la Figure 1. Soit x (t) la variable aléatoire associée au nombre de paquets dans ce système. Notons TE (t) et TS (t) respectivement les débits instantanés d'entrée et de sortie du système.

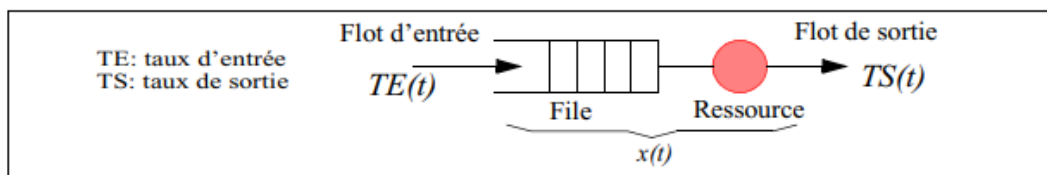


Fig. 1. Modèle de file d'attente d'une ressource réseau [2]

L'idée fondamentale de la théorie différentielle du trafic est d'établir l'expression exacte de l'équation différentielle gouvernant le trafic moyen transitoire X (t) = E [x (t)]. Cette équation est obtenue par une analyse du processus markovien ou semi-markovien associé à x (t). Sa forme générale est la suivante [2], (où P [x(t)] est la probabilité de l'état x (t)):

$$\frac{d}{dt} X(t) = TE(t) - TS(t) = \sum_i x_i(t) \cdot \frac{d}{dt} P[x_i(t)] \tag{1}$$

X(t₀) donné

Cette expression est très générale: elle indique tout simplement que la variation du nombre moyen de paquets dans le système est donnée par la différence entre le débit instantané d'entrée TE(t) et celui de sortie TS(t).

3 MODELE DYNAMIQUE DE LA FILE M/M/1/∞

Selon [2], la file M/M/1/∞ est un modèle très simple souvent utilisé pour représenter un lien de transmission. Il s'agit d'un modèle à capacité infinie. Les clients arrivent suivant un processus de Poisson de paramètre λ. Ils sont servis suivant la discipline PAPS (premier arrivé premier servi) et la durée de service de chaque client est une variable aléatoire exponentiellement distribuée de moyenne 1/μ.

Si nous prenons pour variable d'état x (t) le nombre de clients dans le système, alors la chaîne de Markov associée est un processus de naissance et de mort ergodique de paramètres λ et μ à condition que le taux d'utilisation ρ = λ/μ < 1. Notons P_i(t) la probabilité que x (t) = i. Il est bien connu qu'à l'équilibre, nous avons l'équation I et l'équation différentielle exacte du trafic moyen transitoire X (t) de la file M/M/1/∞ s'écrit:

$$\frac{d}{dt} X(t) = \sum_i \frac{d}{dt} P_i(t) = \lambda - \mu (1 - P_0(t)) \quad (II)$$

Malheureusement, cette équation fait intervenir la probabilité transitoire P₀(t) qui a l'expression complexe suivante (où i représente le nombre initial de clients) [2]:

$$P_k(t) = e^{-(\lambda + \mu)t} \left[\rho^{\frac{k-i}{2}} I_{k-i}(at) + \rho^{\frac{k-i-1}{2}} I_{k+i+1}(at) + (1-\rho)\rho^k \sum_{j=k+i+2}^{\infty} \rho^{\frac{j}{2}} I_j(at) \right]$$

Avec $a = 2\mu\sqrt{\rho}$ et $(\forall k \geq -1), I_k(x) = \sum_{m=0}^{\infty} \frac{x^{k+2m}}{(k+m)! m!}$ (III)

En pratique, cette expression analytique ne peut être utilisée pour calculer la valeur exacte de P₀(t) car elle fait intervenir des sommes infinies. Une autre possibilité serait de calculer le vecteur de probabilités par intégration numérique de la chaîne de Markov. Mais là encore cette solution est limitée par la dimension infinie de l'espace d'états. Bien que dans les deux cas des approximations soient possibles en tronquant les calculs, les approximations résultantes ne sont pas très efficaces d'un point de vue numérique (ce qui est essentiel pour de grands réseaux).

La solution proposée dans le cadre de la théorie différentielle du trafic consiste à transformer l'équation différentielle associée à X (t) en une équation différentielle autonome en utilisant une approximation basée sur une relation implicite entre le trafic moyen X (t) et P₀(t). Plus précisément, on observe qu'à l'équilibre, la relation suivante est vérifiée:

$$\rho = 1 - P_0(\infty) = \frac{X(\infty)}{1 + X(\infty)} \quad (IV)$$

L'idée est alors d'étendre la validité de cette relation au régime transitoire:

$$1 - P_0(t) = \frac{X(t)}{1 + X(t)} \quad (V)$$

On obtient ainsi l'approximation dynamique suivante pour la file M/M/1/∞ :

$$\frac{d}{dt} X(t) \approx \lambda - \mu \frac{X(t)}{1 + X(t)} \quad (VI)$$

X(t₀) donné

4 RESULTATS EN REGIME STATIONNAIRE ET TRANSITOIRE

Il s'agit d'une équation différentielle autonome qui permet de calculer efficacement X (t) par intégration numérique. Il est facile de montrer que cette approximation différentielle converge vers la valeur stationnaire exacte X(∞) = ρ(1-ρ).

Remarquons également que bien que l'équation (VII) permette d'obtenir une expression explicite de $X(t)$, il est plus intéressant en pratique d'utiliser la forme « entrée/sortie » ci-dessous pour de grands réseaux. Les résultats numériques ainsi que les graphiques des figures 2 et 3 sont ceux présentés dans [2].

Tableau 1. Comparaison des valeurs de $X(\infty)$ obtenues avec la formule exacte et avec le modèle différentiel [2]

| | Solution exacte | Modèle différentiel |
|--------------|-----------------|---------------------|
| $\rho = 0.1$ | 0.1111 | 0.1111 |
| $\rho = 0.5$ | 1.0000 | 0.9999 |

Le tableau 1 présente une comparaison des valeurs stationnaires de $X(t)$ obtenues en utilisant la formule exacte et en intégrant le modèle différentiel pour $\rho=0.1$ et $\rho=0.5$. On peut constater que le modèle différentiel converge bien vers les valeurs stationnaires exactes. Pour les mêmes valeurs de ρ , la figure 2 compare le comportement transitoire de $X(t)$ obtenu avec le modèle différentiel à celui obtenu avec 10 000 simulations de Monte-Carlo. On peut voir que le modèle différentiel se révèle assez précis: pour $\rho=0.1$, les erreurs moyennes et maximales sur l'intervalle $[0,10]$ sont respectivement 3.3% et 10%, tandis que pour $\rho=0.5$, ces erreurs sur l'intervalle $[0,30]$ sont respectivement de 4.4% et 8.1%. Mais l'avantage principal du modèle différentiel est qu'il permet de déterminer la réponse dynamique du système à des variations de trafic, comme illustré sur la figure 3 (dans ce cas, l'erreur moyenne est de 4.4% est l'erreur maximale de 13.4%).

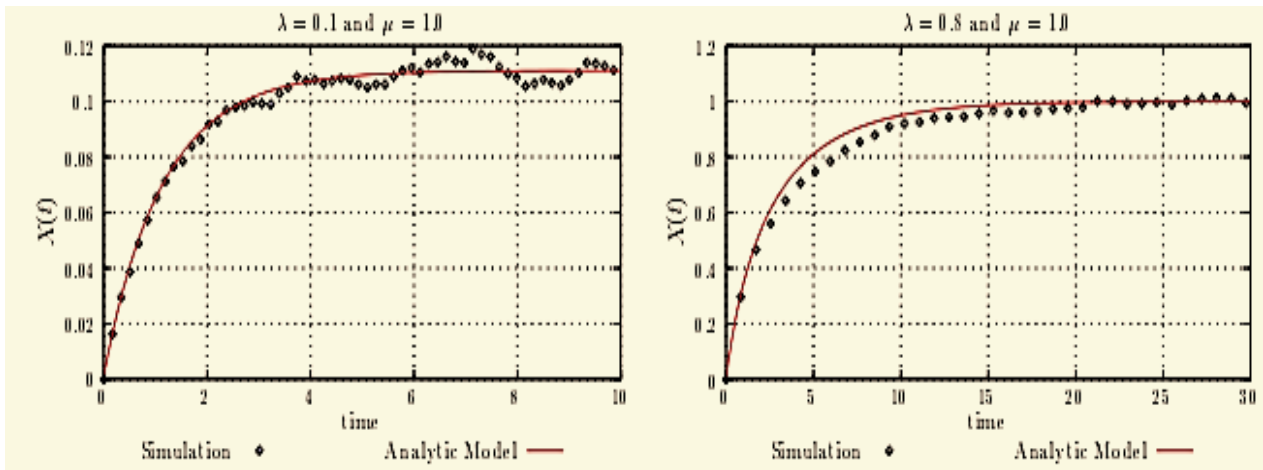


Fig. 2. $M/M/1/\infty$: comparaison entre le modèle analytique et des simulations [2]

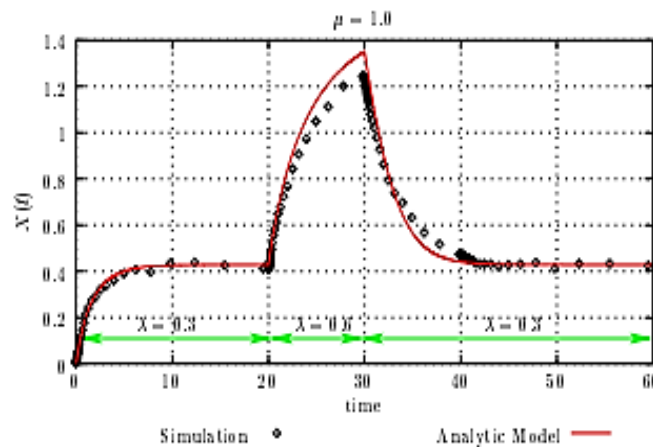


Fig. 3. $M/M/1/\infty$: comparaison entre le modèle analytique et des simulations dans le cas de variation de trafic [2]

4.1 MODÈLES DYNAMIQUES DE RÉSEAUX DE FILES D'ATTENTE

Une hypothèse classique pour de grands réseaux est de supposer des arrivées suivant le processus de Poisson et des durées de service exponentiellement distribuées en chaque nœud. Les files d'attente peuvent être à capacités finies ou infinies.

Sous ces hypothèses, considérons un réseau de n nœuds. Chaque nœud reçoit du trafic direct (exogène) et du trafic sortant des autres nœuds. Notons:

- λ_i le taux d'arrivée du trafic direct au nœud i . On suppose que les paquets arrivent suivant un processus de Poisson,
- μ_i le taux de service au nœud i . On suppose que les durées de service sont des variables aléatoires exponentiellement distribuées,
- $r_{i,j}$ le paramètre de routage donnant la proportion de trafic sortant du nœud i qui est envoyée vers le nœud j .

4.2 FORME PRODUIT [3]

Soit a_j le taux d'arrivée au nœud j (trafic direct et trafic provenant des autres nœuds):

$$a_j = \lambda_j + \sum_{i=1}^n a_i r_{i,j} \quad (\text{VII})$$

La probabilité d'un état global du réseau est le produit des probabilités des états des différents nœuds pris isolément:

$$P = \prod_{i=1}^n (1 - \rho_i) \rho_i^{a_i} \text{ ou } \rho_i = \frac{a_i}{\mu_i} \quad (\text{VIII})$$

Une conséquence importante de la forme produit est que le comportement de chaque nœud du réseau est similaire à celui d'une file $M/M/1/\infty$. Le trafic moyen stationnaire au nœud i a l'expression suivante:

$$X_i(\infty) = \frac{\rho_i}{1 - \rho_i} \quad i=1 \dots n \quad (\text{IX})$$

4.3 MODÈLE DIFFÉRENTIEL DU RÉSEAU À FIL D'ATTENTE

Soit $X_i(t)$ le nombre moyen de paquets au nœud i à l'instant t . Puisque chaque nœud se comporte comme une file $M/M/1/\infty$ isolée, le modèle différentiel développé au point (2) peut être appliqué. Evidemment, il est nécessaire d'introduire le couplage entre les différents nœuds [2].

$$\frac{d}{dt} X_i(t) = \lambda_i + \sum_{j=1}^n r_{j,i} \mu_j [1 - P_j^0(t)] - \mu_i (1 - r_{i,i}) [1 - P_i^0(t)] \quad (\text{X})$$

Où $P_i^0(t)$ est la probabilité que le nœud i soit vide à l'instant t . Ces probabilités peuvent être approximées de la même façon qu'au point (2). On obtient alors l'approximation différentielle du trafic transitoire $X_i(t)$ au nœud i :

En substituant $1 - P_j^0(t)$ et $1 - P_i^0(t)$ par leurs expressions, tel que dans la relation V, on obtient:

$$\frac{d}{dt} X(t) \approx \lambda_i + \sum_{j=1}^n r_{j,i} \mu_j \frac{X_j(t)}{1 + X_j(t)} - \mu_i (1 - r_{i,i}) \frac{X_i(t)}{1 + X_i(t)} \quad (\text{XII})$$

Ce modèle différentiel est dynamique: à chaque instant t , on obtient l'évolution de l'espérance $X_i(t)$ du nombre de clients en chaque nœud. Si les taux d'arrivées et les paramètres de routage sont stationnaires, alors le système d'équations différentielles couplées converge vers l'état moyen du système à l'équilibre.

Deux approches sont possibles pour résoudre ce système d'équations différentielles du premier ordre:

- via des techniques classiques d'intégration numérique si l'on s'intéresse à la solution transitoire. En général, la méthode d'Euler est suffisamment précise,
- via un algorithme d'approximation successive si l'on s'intéresse uniquement à la solution stationnaire.

5 LIMITATION DES MODELES EXISTANTS

5.1 LIMITATION DES MODÈLES STATISTIQUES CLASSIQUES

La piètre efficacité des modèles statistiques classiques (processus de Poisson, loi de Gauss, processus de Markov) est patente et les outils actuellement disponibles pour l'évaluation des performances et le dimensionnement des réseaux à file d'attente répondent peu, ou mal aux problèmes posés:

- Le modèle de Poisson est caractérisé par une variabilité faible, et ne permet qu'un dimensionnement simple du réseau;
 - ✓ Echelle de variabilité $\sim 1/\lambda$
- Le trafic d'un réseau multiservices tel que l'internet n'est pas amnésique mais fortement corrélé, son comportement fractal rend difficile le dimensionnement du réseau;
 - ✓ Le contrôle de congestion TCP représente une mémoire des flux,
 - ✓ Un lien peut devenir saturé quel que soit sa capacité,
 - ✓ Les files d'attente peuvent déborder quel que soit leur longueur
- Le trafic d'un réseau multiservices présente une variabilité extrême;
 - ✓ Durée des connexions: de quelques ms à des jours, semaines, mois,
 - ✓ Volume de données par connexion: de quelques bytes à des téraoctets
 - ✓ De multiples échelles de variabilité allant de ms à des minutes

5.2 LIMITATION DE LA FORME PRODUIT

Dans le cas d'un seul faisceau emprunté par plusieurs flux multiservices, le problème d'évaluation de performances a une solution exacte donnée par la forme produit des probabilités.

Dans le cas d'un réseau multiservices, si la forme produit des probabilités d'état est encore vérifiée, son utilisation est complètement inenvisageable du fait de l'explosion combinatoire de l'espace d'état.

En pratique, la plupart des méthodes proposées dans la littérature sont basées sur l'hypothèse d'indépendance des probabilités de blocage et sur l'approximation du trafic réduit. L'idée est alors de supposer que les phénomènes de blocage apparaissent de manière indépendante sur les différents liens de communications empruntés par un flux au cours de son acheminement, et de considérer que le trafic offert à chaque lien est d'autant plus réduit que les taux de blocage du flux sur les autres liens sont élevés. Sous ces hypothèses, on obtient un système d'équation de type point fixe, dont la solution fournit les valeurs approchées des probabilités de blocage.

6 LA GESTION REACTIVE DES PHENOMENES DE CONGESTION

Rappelons que dans [1] la philosophie du contrôle réactif est d'accepter un maximum de connexions. Lors de la congestion d'un équipement réseau (mesurée en termes de pertes, de délais et de remplissage des tampons), les sources réduisent leurs débits. Ainsi, dans le modèle Best-Effort (« ou au mieux ») du réseau Internet actuel, la gestion du phénomène de congestion est faite de manière réactive par le mécanisme de fenêtre glissante (adaptation du débit) du protocole TCP. Toutefois, ce contrôle n'est pas adapté pour pouvoir offrir des garanties de QoS par exemple à des flux audio et vidéo temps réel: ils ne peuvent pas adapter leur débit.

7 HYPOTHESE DE LA COMPRESSION ADAPTATIVE

7.1 CONTEXTE GÉNÉRAL

Le travail de Emmanuel Besson [7] montre que dans un réseau de transmission global, on attend d'un routeur qu'il dirige les paquets avec le minimum de délai et dans les meilleures conditions de sécurité (c'est-à-dire avec une probabilité de perte la plus faible possible). Cependant, on ne peut pas concevoir d'équipements capables de soutenir n'importe quel débit en entrée. Les limitations technologiques des processeurs, et les contraintes au niveau de la place mémoire sont telles qu'un routeur ne peut soutenir, sans saturer, qu'un débit maximal à déterminer. En conséquence, les trois critères retenus sont:

- Le délai de transfert d'un paquet à travers la structure de routage (depuis la file de réception jusqu'à la file de transmission);
- Le débit maximal autorisé par l'implémentation sans qu'il y ait saturation;
- La probabilité de perte d'un paquet à la suite d'un dépassement de capacité mémoire.

Ainsi, nous proposons une hypothèse qui expose l'idée d'un mécanisme de contrôle réactif qui agit en fonction de l'évolution du trafic c.à.d., qui réagit à la congestion. Dans ce cas, tant qu'il n'y a pas de congestion, le réseau établit toutes les connexions sans compresser le débit. La mesure de congestion peut se faire de plusieurs manières (e.g.: en mesurant les pertes, les délais, le remplissage des buffers...) par le réseau ou par les équipements terminaux. Si une congestion se déclare, le réseau compresse le débit à l'entrée en l'adaptant à la taille de ses ressources disponibles, idéalement, de telle sorte que la stabilité du système et le temps de traitement restent constants.

7.2 LA COMPRESSION ADAPTÉE AUX FLUX DU TRAFIC

Comme nous venons de le rappeler, au point 6 ci-dessus, le mécanisme de fenêtre glissante (adaptation du débit) du protocole TCP n'est pas adapté pour pouvoir offrir des garanties de QoS par exemple à des flux audio et vidéo temps réel.

Quand le trafic est composé de plusieurs types des flux, cas des réseaux multiservices, l'idée d'une technique universelle de compression donnant des meilleures performances pour toutes types de données n'est pas bénéfique. Dans [14], S. Bannour et al expliquent que les diverses techniques de compression offrent différentes performances en termes d'espace, de temps et de taux de compression selon les caractéristiques du réseau et celles du trafic à transmettre. D'où l'idée d'un algorithme de sélection de techniques de compression qui permet de choisir dynamiquement la technique de compression la plus adéquate au flux à transmettre.

L'hypothèse que nous proposons obéit aux normes de la compression **Stacker [4]** et tiendra donc compte de l'évidence suivante: qu'il n'existe pas de technique de compression universelle qui puisse réduire la taille de toutes les données possibles. Si cela existait, on commencerait par réduire les données originales, puis on réapplique la compression aux données compressées, et on réapplique, jusqu'à aboutir au résultat absurde de 0 octet. Cette succession de compressions aboutit en fait à une perte d'information à chaque étape.

La compression a pour but de réduire la longueur d'une chaîne sans affecter son contenu informatif. Cela permet à la fois de réduire les exigences en mémoire et d'augmenter la capacité d'un canal de transmission (théorie de Shannon) [12].

7.2.1 COMPRESSION DE PILE [4]

La compression Stacker est basée sur l'algorithme de compression Lempel-Ziv. L'algorithme Stacker utilise un dictionnaire codé qui remplace un flux continu de caractères par des codes. Ceci stocke les symboles représentés par les codes en mémoire dans une liste de style dictionnaire. Comme la relation entre un code et le symbole d'origine varie selon les données, cette approche est plus réactive aux variations des données. Cette flexibilité est particulièrement importante pour les données LAN, car de nombreuses applications différentes peuvent transmettre sur le WAN à la fois. En outre, à mesure que les données varient, le dictionnaire change pour s'adapter aux besoins variables du trafic. La compression de pile est dynamique mais plus gourmande en CPU et moins gourmande en mémoire.

7.2.2 COMPRESSION DU PRÉDICTEUR [4]

L'algorithme de compression du Prédicteur tente de prédire la séquence suivante de caractères dans un flux de données en utilisant un index pour rechercher une séquence dans le dictionnaire de compression. Il examine ensuite la séquence suivante dans le flux de données pour voir si elle correspond. Si c'est le cas, cette séquence remplace la séquence recherchée dans le dictionnaire. S'il n'y a pas de correspondance, l'algorithme localise la séquence de caractères suivante dans l'index et le processus recommence.

L'index se met à jour en hachant quelques-unes des séquences de caractères les plus récentes du flux d'entrée. Aucun temps n'est consacré à la compression de données déjà compressées. Le **taux de compression** obtenu à l'aide du prédicteur n'est pas aussi bon que d'autres algorithmes de compression, mais il reste l'un des algorithmes les plus rapides disponibles. Predictor consomme plus de mémoire et moins de CPU.

- **Utilisation du processeur:** Les solutions de compression logicielle consomment de précieux cycles de processeur dans le routeur. Les routeurs doivent également prendre en charge d'autres fonctions telles que la gestion, la sécurité et les traductions de protocole; la compression de grandes quantités de données peut ralentir les performances du routeur et provoquer une latence du réseau
- **Taux de compression:** le taux le plus élevé est généralement atteint avec des fichiers texte très compressibles. La compression de données peut entraîner une dégradation des performances, car il s'agit d'une compression logicielle et non matérielle

8 DISCUSSION

Après avoir introduit un des principes de l'ingénierie du trafic qui entrent en jeu dans les solutions de régulation, nous exposons des savoirs actuels relatifs aux mécanismes d'efficacité de ces modèles.

Ces connaissances ainsi constituées permettent d'agir en amont des problèmes en mettant en place des stratégies d'exploitation innovantes et pertinentes. Elles permettent également, par une meilleure compréhension des phénomènes, d'étudier un mécanisme susceptible de réagir en temps réel face à des situations de trafic générant de la congestion.

Le contrôle de trafic désigne l'ensemble des actions du réseau qui permettent d'éviter les congestions.

De même, le contrôle de congestion désigne l'ensemble des actions du réseau qui permettent de minimiser l'intensité et les effets d'un éventuel état de congestion.

Dans cette première étape de notre réflexion, la modélisation du flux de trafic proposée dans la méthode développée, correspond à une solution au problème de contrôle préventif de trafic, particulièrement dans la détermination plus ou moins exacte de la probabilité de blocage, pour un bon dimensionnement des ressources du réseau; tandis que l'hypothèse de la réaction à la congestion par la compression adaptative du trafic y est annoncée comme argument parmi les solutions envisageables dans la problématique de contrôle d'une congestion survenant malgré la planification; dépassant ainsi toute prédiction.

AVANTAGES ET INCONVÉNIENTS DE LA COMPRESSION

Les principaux avantages de la compression sont la réduction du matériel de stockage, des temps de transmission de données et de bande passante, et les économies qui en résultent. Un fichier compressé nécessite une capacité de stockage moindre comparé à un fichier non compressé. L'utilisation de la compression peut faire chuter les dépenses en disque et/ou en disques SSD (Solid-State Drives). Un fichier compressé limite également les temps de transfert et consomme moins de bande passante réseau qu'un fichier non compressé.

Le principal inconvénient de la compression est son impact sur les performances, car les ressources de l'unité centrale et de la mémoire sont sollicitées pour compresser et décompresser les données. De nombreux fournisseurs ont conçu leurs systèmes pour minimiser l'impact des calculs gourmands en processeur associés à la compression. Si celle-ci est effectuée au fil de l'eau (*in-line*), avant que les données soient inscrites sur le disque, le système peut la décharger pour préserver ses ressources. Par exemple, IBM utilise une carte d'accélération matérielle distincte pour gérer la compression avec certains de ses systèmes de stockage d'entreprise.

Si les données sont compressées après l'inscription sur le disque ou après traitement, le processus peut être exécuté en arrière-plan pour limiter l'impact sur les performances. La compression post-traitement peut réduire le temps de réponse pour chaque entrée-sortie (E/S), mais elle consomme toujours de la mémoire et des cycles de processeur, ce qui peut affecter le nombre total d'E/S gérable par un système de stockage. De plus, comme les données doivent être initialement inscrites sur le disque ou les lecteurs de stockage flash sans compression, les économies en stockage physique ne sont pas aussi importantes qu'avec la compression au fil de l'eau.

9 CONCLUSION

Dans cet article, nous avons montré que le modèle dynamique de la file $M/M/1/\infty$ permet d'obtenir des expressions analytiques (parfois exactes) de critères de performances sous des hypothèses relativement fortes pour certains systèmes d'attente.

Toutefois, la simulation a un pouvoir d'expression beaucoup plus important (a priori sans limite) mais se heurte à des temps de calcul prohibitifs pour des réseaux de tailles réelles.

Par ailleurs, une modélisation analytique exacte ayant toutes les propriétés semble aujourd'hui complètement inenvisageable compte tenu des connaissances théoriques disponibles. En particulier, il est bien connu qu'une modélisation analytique des phénomènes transitoires dans les réseaux de files d'attente est très complexe et peu de résultats généraux existent dans ce domaine.

En outre, nous ouvrons une discussion sur l'hypothèse de la réaction à la congestion par la compression adaptative du trafic qui présage un pouvoir d'expression important, a priori pour des réseaux de grande taille. Le résultat final que nous souhaitons présenter dans nos prochains travaux, est la mise en œuvre d'un algorithme qui déterminera toutes les étapes du traitement et donc les instructions à donner pour une exécution automatique du système de compression adaptative de débit d'entrée des réseaux de taille réelle.

REFERENCES

- [1] Khodor A., Conception et évaluation d'un modèle adaptatif pour la qualité de service dans les réseaux MPLS, ÉCOLE CENTRALE DE LILLE, décembre 2010, p.28.
- [2] David GAUCHARD, Simulation hybride des réseaux IP-DiffServ-MPLS multi-services sur environnement d'exécution distribuée, HAL Id: tel-00011034, novembre 2005, p.53-54.
- [3] O. Brun et J.M Garcia, Modélisation Dynamique du Trafic dans les Réseaux multiservice, LAAS-CNR, 2000, p.2&4
- [4] <https://www.cisco.com> › 14156-compress-overview, Présentation de la compression de données - Cisco.
- [5] UIT-T, Gestion du trafic et des encombrements dans le RNIS-LB, Recommandation I.371 élaborée par la Commission d'études 13 (1997-2000) de l'UIT-T et approuvée le 10 mars 2000.
- [6] Garcia J.M., 1980, p.17.
- [7] TERRE M., Introduction aux télécommunications, édition Cnam, 2005.
- [8] E. Besson, Modèles analytiques de routeurs, RAIRO. Recherche opérationnelle, tome 34, n°2, 2000, page 234.
- [9] Alexandre DUGAS: Architecture de transformée de cosinus discrète sur deux dimensions sans multiplication et mémoire de transposition, Sherbrooke (Québec) canada, novembre 2012].
- [10] F. Oueslati: Un algorithme d'adaptation dynamique de débit pour les communications vidéo interactives sur internet, Université du Québec, 2014 page 63-65.
- [11] M. Boukadoum et al., MIC 4220, Traitement numérique des signaux, Compression de signaux numériques, UQAM, Automne 2014.
- [12] C. VALADE, Compression d'images complexes avec pertes: application à l'imagerie Radar, HAL Id: pastel-00002064, Thèse, décembre 2006.
- [13] D. Renaud, Support, Compression et Stockage. Page 18., Structure de l'information, Université de liège, belgique, 2000.
- [14] Frédéric SUR, Initiation au traitement du signal et applications, École des Mines de Nancy, 2009-2012] page 18 & 72.
- [15] S. Bannour et al., Etude de paramètres de techniques de compression pour la conception d'un algorithme de sélection, SETTIC, Tunis, Mars 2007.