Mathématiques et Réseaux de Communication

Philippe Robert

MATHÉMATIQUES ET RÉSEAUX DE COMMUNICATION

PHILIPPE ROBERT

Notice Biographique. Philippe Robert est Directeur de Recherche à l'INRIA (Institut National de Recherche en Informatique et en Mathématiques Appliquées) et Professeur Chargé de Cours à l'École Polytechnique au département de Mathématiques Appliquées. Il est responsable de l'équipe de recherche INRIA "Réseaux de Communication, Algorithmes et Probabilités". Ses travaux de recherche concernent l'étude mathématique des réseaux de communication.

1. Introduction

Cet article présente brièvement l'étude des réseaux de communication sous l'angle de la modélisation mathématique. Les quarante dernières années ont vu le développement remarquable de ces réseaux, les ordres de grandeur de ces systèmes ont changé complètement : on est passé de quelques dizaines, centaines de nœuds inter-connectés à des réseaux réunissant des centaines de millions de nœuds. Parallèlement les vitesses de transmission ont elles aussi augmenté dans des proportions similaires. Les possibilités d'utilisations de ces réseaux par une large population d'utilisateurs, comme la navigation sur les pages web du réseau Internet par exemple, sont de plus en plus nombreuses, de plus en plus sophistiquées. L'ère de l'Internet et ses ordres de grandeur, l'évolution des technologies posent de nouveaux challenges aux concepteurs de réseaux. Il n'en reste pas moins que, depuis l'origine, depuis le déploiement du réseau téléphonique au début du vingtième siècle pour fixer les idées, se dégage un ensemble de questions fondamentales commun à tous ces réseaux aussi variés soient-ils :

- Comment diffuser, rechercher l'information dans un réseau ?
- Comment allouer, garantir les ressources pour l'accès à un réseau ?

C'est le propre de la démarche scientifique d'identifier les problèmes génériques, fondamentaux d'un domaine donné et ensuite d'essayer de les résoudre dans un cadre général.

Mathématiques. Les modèles mathématiques d'un réseau de communication permettent de décrire son évolution et de la quantifier précisément. L'aspect aléatoire, non prévisible, des arrivées de demandes de communication est un trait majeur de ces systèmes dont doit tenir compte le concepteur de réseau. Pour cette raison la théorie des probabilités est le cadre naturel pour l'étude mathématique de ces réseaux. Cette théorie mature et particulièrement flexible donne les outils nécessaires pour étudier non seulement les réseaux de communication mais aussi une large palette de systèmes comme les marchés financiers, les systèmes biologiques, ... Il n'y a donc pas de théorie mathématique spécifique pour les réseaux. Les réseaux de communication constituent un large domaine d'application et de développement de la théorie des probabilités. À titre d'exemple, nous présenterons deux objets mathématiques importants utilisés, entre autres, pour étudier les réseaux : les processus de Poisson et les processus de Markov.

Informatique. La modélisation mathématique n'est pas, bien sûr, la seule discipline scientifique ayant un impact sur la conception des réseaux de communication. L'informatique joue un rôle très important, notamment pour définir et concevoir les programmes ou protocoles qui gèrent les communications dans le réseau. La conception d'algorithmes pour les réseaux est un sujet d'importance majeure. La modélisation mathématique intervient aussi dans ce domaine, par exemple pour évaluer l'efficacité d'un algorithme ou pour le comparer à d'autres.

Date: 21 décembre 2009.

2. Un Bref Survol Historique

Les premiers réseaux : Le téléphone. Les premiers grands réseaux de communication ont commencé à se développer au début du vingtième siècle. C'est à cette époque que le réseau téléphonique se déploie sur une large échelle dans la plupart des pays industriels. Jusqu'aux années 1960, le réseau téléphonique sera, essentiellement, le seul "vrai" réseau au sens où nous l'entendons actuellement : un ensemble significatif de points connectés entre eux pouvant recevoir et émettre des informations via un médium (électronique, radio, optique, ...).

Au début de cette période, l'établissement des communications se fait "manuellement" : dans le cas simple où un utilisateur à Paris voulait entrer en communication avec un correspondant B géographiquement proche, à Asnières par exemple, il devait nécessairement contacter un opérateur pour que celui-ci établisse le *circuit* entre A et B. Quand le correspondant était plus distant, le circuit était établi en utilisant des relais et donc plusieurs opérateurs. Assez naturellement, les gestionnaires de réseaux de l'époque sont déjà confrontés au problème de *dimensionnement* d'un tel réseau.

> *Combien d'opérateurs faut-il pour qu'un utilisateur ait en toute circonstance une probabilité importante de pouvoir établir sa communication ?*

Cette question pose en premier lieu le problème de la description du *trafic* : comment peut-on représenter mathématiquement les demandes de communication au central pour pouvoir calculer une telle probabilité ?

Les travaux scientifiques dans ce domaine sont essentiellement l'œuvre d'ingénieurs mathématiciens liés aux opérateurs téléphoniques de l'époque et de mathématiciens d'universités d'Europe de l'Est. On peut citer parmi eux :

- A. K. Erlang (1878 — 1929)
  Mathématicien travaillant pour la compagnie de téléphone de Copenhague ;
- T. O. Engset (1865 — 1943)
  Ingénieur de la compagnie de téléphone norvégienne Televerket ;
- C. Palm (Suède, 1907 — 1951)
  Ingénieur statisticien pour la compagnie de téléphone suédoise Ericsson ;
- A. Khinchin (1894 — 1959), Membre de l'académie des sciences de l'URSS ;
- F. Pollaczek (1892 — 1981), Ingénieur à la poste allemande.

Les réseaux informatiques. Les années 1960 voient le début de l'industrie informatique : IBM lance alors une gamme d'ordinateurs à l'échelle industrielle : la machine IBM OS/360 TSO (Time Sharing Option) est le premier d'une longue série d'ordinateurs généralistes. Le modèle générique de cette époque est celui de l'*ordinateur central* auquel sont reliés de nombreux terminaux. Un tel ordinateur traite les requêtes émanant de ces terminaux. Un exemple célèbre est HAL, l'ordinateur central du vaisseau spatial en route pour Mars dans le film de S. Kubrick "2001 : l'odyssée de l'espace" (1968). Un modèle classique d'allocation de ressources dans ce cadre est celui du mécanisme de temps partagé : Si à un instant donné N requêtes sont en cours, l'ordinateur, le programme, consacre à chacune d'elles la fraction $1/N$ sur chaque unité de temps.

Les concepteurs de ces systèmes informatiques sont amenés à résoudre des problèmes de dimensionnement et d'évaluation des performances de leurs systèmes, comme par exemple :

- *Déterminer le nombre de terminaux, la charge maximale, que peut supporter l'ordinateur.*
- *Estimer les délais moyens de traitement d'une requête, le nombre moyen de requêtes en attente, ...*

Comme dans le cas des réseaux téléphoniques, il s'agit en premier lieu de décrire le trafic qui arrive aux terminaux d'une telle architecture.

Ces problèmes sont étudiés extensivement dans le cadre de laboratoires industriels comme ceux d'IBM ou des laboratoires Bell de la compagnie AT&T ainsi que dans les universités américaines. L'architecture de ces réseaux informatiques est, en général, assez limitée. Dans la grande majorité des cas, il n'y a qu'un nœud : le serveur central qui traite les requêtes issues des points d'accès. La communication entre ordinateurs est envisagée sous la forme de tâches envoyées d'une machine à une autre plutôt que sous la forme d'une *connexion* au sens où deux machines échangent des données entre elles.

Les systèmes distribués. Pendant les années 1970 sont conçus et développés des réseaux avec une différence d'architecture fondamentale. Chacun des nœuds de ces réseaux est autonome : Si un nœud veut diffuser, rechercher une information, il ne s'adresse pas à un contrôleur au niveau du réseau, à un ordinateur central qui donne ou non la permission de transmettre. Chaque nœud utilise ou essaie d'utiliser de manière autonome les ressources communes du réseau mises à sa disposition (comme la bande passante par exemple), ce qui entraîne bien sûr des conflits d'accès à ces ressources. Ces réseaux avec des nœuds autonomes sont appelés *systèmes distribués.* L'exemple le plus connu est Internet qui est essentiellement un système distribué.

Exemples

(1) Protocoles d'accès.
Des émetteurs, appelés aussi stations, qui sont disséminés dans la nature (le désert, des îles, ...), ne peuvent émettre que sur une seule fréquence. Si au moins deux stations émettent en même temps, les signaux émis se superposent et ne peuvent donc être décodés par un éventuel récepteur. Les tentatives de transmission simultanées sont donc des échecs. La dispersion géographique interdit de plus une quelconque coordination pour que les stations puissent convenir au préalable de l'ordre dans lequel elles vont transmettre. Il n'y a pas de contrôleur central pour déterminer qui doit transmettre et quand. Le système doit donc s'auto-organiser.

Un exemple d'algorithme : ALOHA. Cette solution a été proposée par Abramson (1968). Chaque unité de temps, une station ayant un message à transmettre tire à pile ou face pour déterminer si elle essaie ou non à cet instant. Le tirage à pile ou face est en fait effectué par un générateur aléatoire. Une station avec un message à transmettre a donc une probabilité 1/2 pour faire un essai de transmission. Si N stations sont en concurrence la probabilité qu'une station donnée transmette avec succès, i.e. qu'elle émette et qu'aucune des $N-1$ autres n'essaie à ce moment, est donc donnée par

$$1/2 \times (1{-}1/2)^{N-1} = 1/2^N > 0.$$

Chaque unité de temps, il y a donc une probabilité positive de succès. On peut voir qu'en répétant ces essais, une station va finir par transmettre avec succès. Et donc, toutes les stations finiront pas transmettre avec succès. Ainsi, les communications sont acheminées dans le réseau sans l'aide d'un contrôle central.

(2) Transmission de données dans le réseau Internet.
Pour envoyer un courrier électronique (par exemple) d'une machine A vers une machine B, le message est découpé en paquets et ceux-ci sont envoyés progressivement par A. Les paquets transitent par les nœuds du réseau, si l'un d'eux est saturé (dans le cas d'un trafic local important par exemple) alors ces paquets seront perdus. Le réseau Internet n'ayant pas de contrôle central, les paquets peuvent être donc perdus sans que A en soit informé directement. La question est de savoir comment A peut transmettre *tous* ses paquets à B, en prenant en compte le fait que le réseau n'est pas fiable et sans connaître précisément l'activité des autres machines. La procédure, l'algorithme, qui résout ce problème de l'envoi des paquets s'appelle TCP (Transmission Control Protocol).

(3) Les réseaux Pair à Pair.
Un ensemble de fichiers est réparti sur un très grand nombre de serveurs qui peuvent tomber en panne, être arrêtés, ... Par quelle méthode un utilisateur peut-il télécharger un des fichiers de cet ensemble ? En particulier comment peut-il connaître le serveur le plus proche (en terme de temps d'accès) qui possède le fichier. Une solution possible serait qu'UN serveur central A possède toutes les informations : i.e. les fichiers que possède chacun des serveurs. Le problème est que si A tombe en panne, le réseau de distribution de fichiers ainsi mis en place est inopérant. Une solution qui donnera essentiellement le même rôle à tous les serveurs présentera de meilleures garanties de robustesse. Là encore le cadre d'un système distribué s'impose donc naturellement.

Noter que les systèmes distribués sont présents dans le monde naturel, comme le déplacement des bancs de poissons, essaims, nuées d'oiseaux, ... Dans un banc de poissons, chaque individu se contente de répercuter les variations de ses voisins dans les trois directions ce qui donne à l'ensemble du banc une unité sans "un chef de banc". Ces stratégies sont en particulier bien adaptées pour désorienter (un peu) d'éventuels prédateurs.

Système centralisé vs Système distribué. Un avantage évident d'un système distribué est sa robustesse : si un des nœuds tombe en panne, le réseau continue de fonctionner. Il n'y a pas de nœud indispensable. Cette propriété de robustesse était un des sujets d'étude des programmes de recherche (financés par l'armée américaine) à l'origine de l'Internet. Le modèle du serveur central est particulièrement vulnérable de ce point de vue. Pour la même raison, un système distribué pourra intégrer facilement un grand nombre de nouveaux nœuds.

Un système centralisé, du fait du contrôle global, pourra, lui, assez simplement garantir les performances d'une communication qu'il accepte. Comme par exemple assurer que l'échange de données entre les deux nœuds de la communication se fera à un débit supérieur à une valeur donnée. Dans un système distribué, comme le réseau Internet, du fait de l'absence de contrôle de l'état des nœuds, il est beaucoup plus difficile de pouvoir offrir ce type de garantie. Une communication donnée peut éventuellement passer par des nœuds saturés qui diminueront alors son débit de façon arbitraire.

Algorithmes. Si l'idée que chaque nœud ait une activité autonome est séduisante, elle a néanmoins une contrepartie. Les différents nœuds sont en concurrence pour l'accès aux ressources du réseau comme par exemple la fréquence commune dans le cas des protocoles d'accès décrit ci-dessus. Ce problème ne se pose pas dans le cadre d'un contrôle centralisé : le contrôleur central décide qui a accès à quoi et quand. Il faut donc une procédure, un *algorithme*, pour allouer les ressources dans le cadre d'un système distribué avec la contrainte que chaque nœud utilise le même algorithme. La conception d'algorithme est un sujet clé de la recherche dans les réseaux de télécommunication.

3. Algorithmes et Modélisation Mathématique

L'Innovation dans les Réseaux : les Algorithmes. Les réseaux de communication sont identifiés assez couramment à des technologies comme les réseaux sans fil associés au GSM ou au Wifi, les réseaux optiques où le médium est une fibre optique, ... L'accroissement des performances des architectures matérielles, de la vitesse des processeurs, du nombre de canaux radio ou encore de la taille des mémoires jouent un rôle important dans le développement des réseaux de communication. Ce n'est toutefois qu'un aspect de l'innovation dans les réseaux.

La *conception d'algorithmes* est aujourd'hui un élément crucial dans toutes les avancées spectaculaires qu'a connu le monde des réseaux de communication ces quarante dernières années. Citons deux exemples représentatifs : l'algorithme TCP (Transmission Control Protocol) de transmission de données dans le réseau Internet inventé par Cerf et Kahn (1973) et Van Jacobson (1987) qui achemine près de 90% du trafic Internet et l'algorithme de classification de sites web inventé par Brin et Page (1997) qui est à la base du moteur de recherche Google.

La conception d'algorithmes est maintenant le principal facteur d'innovation pour le développement des nouvelles architectures de réseaux.

Exemples

- Réseaux Mobiles.
  Déterminer une politique d'allocation de bande passante pour des utilisateurs ayant des demandes variées : pour le téléphone ou la transmission vidéo par exemple. Faut-il privilégier les applications gourmandes en bande passante (comme la vidéo) ou plutôt les communications téléphoniques ? ou faire un partage entre ces deux types de trafic ?
- Réseaux Pair à Pair.
  Comment répartir des fichiers vidéo sur une fraction des nœuds d'un réseau de telle sorte que chaque utilisateur accède en un temps minimal à un ensemble de films.
- Réseaux de capteurs.
  Un ensemble de petits modules, des capteurs, sont disséminés dans un environnement (pour recueillir des données météorologiques par exemple). Ils recueillent chacun localement des données et ils peuvent communiquer entre eux pour se transmettre leurs données pour les acheminer à un lieu de collecte. Chaque transmission entame les capacités de la batterie d'un capteur et deux capteurs proches ne peuvent émettre en même temps. Comment rassembler les informations recueillies par chaque capteur tout en consommant le moins possible d'énergie pour préserver la longévité du réseau ?

L'Innovation dans les Réseaux : Modélisation Mathématique. Les modèles mathématiques ont joué un rôle important au tout début de l'essor des réseaux téléphoniques. La complexité des nouvelles architectures a renforcé la nécessité de pouvoir représenter mathématiquement l'état d'un réseau pour pouvoir répondre à des questions élémentaires comme

– pour quelles configurations de trafic la charge de chaque nœud de celui-ci reste "raisonnable" au cours du temps. On dira dans ce cas que le réseau est stable.
– Quel est le pourcentage des requêtes qui ne peuvent pas accéder aux ressources du réseau et qui sont donc rejetées par celui-ci ?
– Quel est le temps de traitement d'une demande de transmission ?

L'ordre de grandeur du nombre de nœuds d'un réseau téléphonique est de l'ordre de quelques centaines (au plus), en principe il est possible de concevoir un programme informatique qui simulera le comportement de celui-ci dans une configuration de charge donnée. Les échelles des réseaux actuels sont bien au-delà de ces nombres. La simulation complète de tels systèmes par un programme informatique est difficilement envisageable malgré la puissance de calcul des ordinateurs actuels.

Quand la modélisation mathématique est possible, elle permet d'éviter ces lourds traitements informatiques. Elle a en outre l'avantage de pouvoir garantir des propriétés de stabilité du réseau, comme par exemple assurer, sous certaines hypothèses, que le nombre moyen de communications simultanées dans tout le réseau sera toujours majoré par une constante fixée. Un concepteur de système n'utilisant que les simulations ne pourra guère faire mieux que montrer que le réseau fonctionne correctement sur une certaine plage de temps et pour un ensemble de paramètres fixés numériquement : taux d'arrivée, capacité, ...

Modélisation Mathématique : La Théorie des Probabilités. Un réseau de communication reçoit et traite des flots de requêtes qui sont définis par plusieurs caractéristiques parmi lesquelles :
– les instants d'arrivées des demandes de communication ;
– la durée de chacune de ces requêtes ;
– l'ensemble des nœuds qui doivent être utilisés par celles-ci.

Tous ces paramètres qui correspondent à des demandes de ressources du réseau qui ne peuvent être bien sûr prévues à l'avance. Le réseau se doit de réagir au fur et à mesure des arrivées.

Une solution (irréaliste) serait de concevoir un réseau pouvant fonctionner dans tous les scénarios d'arrivées de requêtes, de demandes, ... Par exemple, d'envisager sur un réseau que tous les abonnés puissent appeler un service donné sur une plage de temps très courte. La conséquence serait de sur-dimensionner tout le réseau pour que l'accès aux services se fasse à un débit gigantesque pour pouvoir accepter ces pointes de trafic. Ce serait comme remplacer toutes les routes départementales du réseau routier par des autoroutes pour éviter absolument tous les embouteillages possibles.

Si cette éventualité n'est clairement pas réaliste dans le cas des réseaux de communication, elle l'est dans d'autres domaines comme les systèmes dits critiques. Par exemple, le centre de contrôle d'un avion de ligne doit pouvoir répondre à tous les scénarios possibles et imaginables entre tous les paramètres clés de l'avion : moteurs, capteurs de vitesse, paramètres de vol, défaillance d'éléments, ... Il n'est absolument pas possible d'accepter que l'avion ait ne serait-ce qu'une probabilité de 1/1000, 1/10000, ... de s'écraser dans certaines situations "malheureuses"...

En revenant au cas des réseaux, mathématiquement, cela se traduit par l'introduction de modèles probabilistes pour les arrivées, les demandes, ... Essentiellement cela consiste à quantifier l'aléatoire auquel est soumis un réseau. Un exemple de scénario de trafic : les instants d'arrivées des requêtes sont espacées entre 10 et 20 millisecondes mais avec une probabilité de 1/1000 il peut arriver une centaine de requêtes en moins de 5 millisecondes. C'est le rôle du concepteur de réseau de proposer une architecture qui puisse prendre en charge un ensemble de scénarios de trafic spécifiés. La partie 4 donne un aperçu de la modélisation pour décrire les instants d'arrivées des requêtes.

Algorithmes Probabilistes. Les arrivées de requêtes de communications constituent la majeure partie de l'incertitude, l'aléatoire auquel est soumis un réseau, ce n'est cependant pas la seule source d'aléatoire. Certains des algorithmes qui contrôlent les communications utilisent aussi des composantes aléatoires, on parle d'*algorithmes probabilistes*. C'est le cas par exemple de l'algorithme ALOHA décrit plus haut où un émetteur utilise un tirage à pile ou face (une variable aléatoire donc) pour décider d'un essai de transmission ou non. Ces algorithmes utilisant des tirages aléatoires ont connu un remarquable développement dans les réseaux de communication depuis la fin des années 1960, ils permettent en particulier d'éviter des conflits répétés pour l'accès aux ressources. Leur domaine d'application est en fait considérablement plus grand maintenant : reconstruction d'images, géométrie algorithmique, cryptographie, ... Dans ce contexte, l'aléatoire n'est plus seulement un élément extérieur subi (comme dans le cas des arrivées) mais aussi une composante introduite pour résoudre un problème.

4. Représentations mathématiques du trafic

Les instants d'arrivées des demandes de communications sont, par nature, aléatoires et par conséquent un concepteur de réseau doit intégrer que le réseau devra ponctuellement supporter des pointes de trafic. Il est donc impossible de garantir, par exemple, que le temps d'attente de chaque demande soit toujours inférieur à x millisecondes. On pourra éventuellement assurer, via un modèle mathématique, que la probabilité que le temps d'attente excède x millisecondes soit plus petite que 0.00001. Les critères de performances s'exprimeront par conséquent en terme de moyenne : délai moyen d'attente, nombre moyen de requêtes en attente, etc...

Dans le cas du trafic téléphonique, on pourrait penser caractériser le trafic par le nombre moyen d'appels par heure, qui peut être assez facilement estimé. Mais si ce nombre est de 60 par exemple, cela peut recouvrir des situations assez différentes, prenons les plus extrêmes :

(1) Ils ont tous lieu au même moment à chaque heure ;

(2) Il y a un appel chaque minute.

Si un opérateur peut traiter un appel par minute, dans le cas (1) il y a besoin de 60 opérateurs alors que dans le cas (2) un seul suffit pour que toutes les requêtes soient satisfaites. L'affectation des ressources ne sera donc pas la même suivant les scénarios. En pratique, le scénario probable est entre ces deux cas : les arrivées sont espacées mais ponctuellement il peut y avoir des arrivées très rapprochées.

Les processus de Poisson. Les instants d'arrivées des requêtes peuvent être décrits comme des points dispersés au hasard sur la droite temporelle. Rigoureusement, cela n'est pas possible, on ne peut lancer un point au hasard, uniformément, entre 0 et l'infini. Comme souvent en mathématiques, on peut toutefois donner un sens à cela en considérant des intervalles de temps dont la taille T est de plus en plus grande. Cela se fait de la manière suivante.

Pour $\lambda > 0$, on suppose que les instants d'arrivées de λT communications ont lieu au hasard entre 0 et T (en supposant que λT soit un entier pour simplifier). En particulier la probabilité qu'une demande de communication donnée arrive pendant l'intervalle de temps $[a, b]$ est donnée par $(b - a)/T$. Dans cette configuration, le taux d'arrivée des demandes de communication par unité de temps est donné par $\lambda T/T = \lambda$. On a ainsi une suite ordonnée de points sur l'intervalle $[0, T]$ qui correspondent aux arrivées de λT communications.

Un résultat mathématique montre que lorsque T tend vers l'infini, la répartition de cette suite de points aléatoires se rapproche de celle d'une suite aléatoire infinie (t_n) croissante de points que l'on appelle *un processus de Poisson* d'intensité λ. Les instants t_n sont les instants d'arrivées associés au processus de Poisson. Un processus de Poisson est donc, d'une certaine façon, le lancer au hasard d'un nombre infini de points sur une demi-droite.

On peut aussi représenter un processus de Poisson de la façon suivante. Considérons un ensemble de N individus, $N = 10000$ par exemple, qui peuvent demander l'accès à un réseau ou à un service donné. Supposons que pour chaque unité de temps chacun d'eux essaie d'y accéder avec une petite probabilité, de l'ordre de λ/N, pour un $\lambda > 0$. On peut montrer mathématiquement que l'ensemble de tous les instants où des utilisateurs accèdent au service est aussi proche d'un processus de Poisson de paramètre λ. Supposer que l'ensemble des instants d'arrivée de demandes de communications soit un processus de Poisson est donc une hypothèse tout à fait réaliste pour de larges classes de réseaux. Comme on va le voir, il y a toutefois des situations, le réseau Internet par exemple, où il convient d'être prudent avec ce type de représentation.

On présente quelques unes des propriétés mathématiques importantes d'un processus de Poisson.

- Instants d'arrivées.

  La distribution du premier instant d'arrivée t_1 est donnée par

  $$\text{Proba}(t_1 \geq y) = e^{-\lambda y}, \text{ pour } y \geq 0.$$

  C'est une *loi exponentielle* de paramètre λ. La durée $t_{n+1} - t_n$ entre le n-ième point et le $(n + 1)$-ième suit aussi une loi exponentielle de paramètre λ. Un processus de Poisson est donc un ensemble d'instants espacés par des variables aléatoires de loi exponentielle.

– Nombre d'arrivées.
  Le nombre d'arrivées N_{ab} entre les instants a et b est le nombre de points t_n, tels que $a \leq t_n \leq b$. La variable entière N_{ab} suit une *loi de Poisson* : la probabilité qu'il y ait n points dans l'intervalle de temps $[a, b]$ vaut
$$\text{Proba}(N_{ab} = n) = \frac{(\lambda(b-a))^n}{n!} e^{-\lambda(b-a)}, \text{ pour } n \geq 1.$$
  Le nombre moyen d'arrivées sur $[a, b]$ vaut $\lambda(b-a)$, λ est donc le nombre moyen de demandes par unité de temps.
  Une propriété cruciale des processus de Poisson est l'*indépendance suivante* : elle s'exprime par le fait que savoir qu'il y a dix millions d'arrivées entre les instants 0 et 1 (par exemple) n'apporte aucune information sur le nombre d'arrivées qu'il y a dans l'intervalle suivant, entre 1 et 2. Plus globalement, les nombres d'arrivées dans deux intervalles de temps disjoints sont des *variables aléatoires indépendantes.*

Si le processus de Poisson est un modèle naturel pour représenter des instants d'arrivées, sur le plan mathématique, il permet en outre le calcul de nombreuses caractéristiques associées aux arrivées de requêtes.

Processus de Poisson et Mouvement Brownien. Le processus de Poisson et le mouvement brownien sont les deux objets mathématiques fondamentaux de la théorie des probabilités. Le mouvement Brownien décrit les perturbations aléatoires continues autour de la trajectoire d'un point dues à l'interaction avec le milieu. Comme par exemple la perturbation d'un signal radar par les variations de l'atmosphère. Les processus de Poisson sont eux des objets naturels pour décrire des instants isolés où se passent des événements aléatoires.

Du point de vue de la modélisation, le mouvement brownien est l'objet de base pour décrire toutes les évolutions aléatoires *continues*, par exemple le mouvement d'un grain de pollen à la surface de l'eau pour reprendre le modèle originel du botaniste Robert Brown, ou encore les évolutions des cours de portefeuilles boursiers. Le processus de Poisson est lui utilisé pour décrire les évolutions *discontinues*, c'est-à-dire procédant par sauts : instant d'arrivée d'une requête, de mutation d'un gène, d'émission d'une particule radioactive, de déclenchement d'une avalanche, etc...

Les phénomènes de dépendance longue dans l'Internet. Les processus de Poisson sont largement utilisés pour décrire mathématiquement les arrivées de requêtes aux réseaux de communication. Avec ces modèles de trafic, de nombreuses quantités comme les charges moyennes des nœuds, le débit moyen d'une connexion ou encore la probabilité de rejet d'une requête peuvent être exprimées par des formules mathématiques. Ces formules permettent au concepteur de réseau de calibrer la capacité de celui-ci.

Par exemple, considérons une mémoire pouvant accueillir au maximum C requêtes en attente de traitement à un nœud de communication. Si le processus d'arrivées est un processus de Poisson, un résultat classique donnera que la probabilité P_0 de rejet d'une nouvelle requête, i.e. que celle-ci trouve la mémoire complètement occupée (C requêtes sont déjà en attente de traitement), est majorée par

$$P_0 \leq a \exp(-bC), \tag{1}$$

pour des constantes a et b qui dépendent des paramètres du système. Si le réseau doit satisfaire la contrainte que la probabilité de rejet soit plus petite que ε, $P_0 \leq \varepsilon$, ($\varepsilon = 0.0001$ par exemple) avec la majoration précédente il suffit de choisir, de dimensionner, la capacité C telle que $C \geq -\log(\varepsilon/a)/b$. Cette procédure est utilisée dans de nombreux exemples de réseaux.

Si cette mémoire est un des nœuds du réseau Internet, il a été montré qu'une majoration similaire à (1) n'est plus valable. En fait l'analogue d'une telle relation serait dans ce cas
$$P_0 \leq \frac{a}{C^h},$$
pour des constantes $a > 0$ et $h > 1$. La conséquence est que pour le même critère sur P_0, la capacité C devra être de l'ordre $1/\varepsilon^{1/h}$ soit nettement plus grande que l'ordre de grandeur $-\log \varepsilon$ précédent. Cette relation a bien entendu un impact très important sur le dimensionnement.

La raison principale est que les processus de Poisson ne sont pas des modèles adaptés pour décrire les arrivées de paquets du trafic Internet et donc la relation (1) n'est pas vraie dans ce cas. On a vu que pour le processus de Poisson, les arrivées sur des intervalles de temps différents sont indépendantes. Sur le réseau Internet, ce n'est pas le cas, entre deux plages de temps même très éloignées, il y aura toujours un nombre significatif de connexions qui transmettront continuellement des paquets entre ces deux périodes et donc cette situation crée une dépendance pour les arrivées de paquets sur ces deux intervalles de temps. Les transferts de très gros fichiers, par les réseaux pair à pair par exemple, sont à l'origine de l'existence de ces très longues connexions

Cette difficulté peut être résolue en décrivant une connexion, non comme une succession de paquets, mais comme un fluide s'écoulant avec un débit variable. Les instants de début de connexion peuvent alors être raisonnablement représentés par un processus de Poisson.

L'expérience montre que, pour que les modèles mathématiques décrivant l'état d'un réseau soient analysables, ils doivent d'une façon ou d'une autre prendre en compte un modèle d'arrivées de requêtes du type processus de Poisson. Le processus de Poisson est un objet incontournable pour obtenir des résultats utilisables en pratique.

5. État d'un Réseau : Les processus de Markov

Le cahier des charges d'une architecture de réseau est défini par un ensemble de fonctionnalités attendues par l'utilisateur et aussi par un ensemble de contraintes sur les valeurs de plusieurs des caractéristiques du réseau comme la charge de trafic sur chacun des nœuds, le débit obtenu par chacune des connexions, ou encore le taux de rejet des connexions. Par exemple, on voudra que le débit moyen d'une communication acceptée soit supérieur à 10 Mega-bit par seconde ou encore que le nombre moyen de requêtes en attente soit plus petit que 100, ... Le concepteur de réseau attendra donc d'un modèle mathématique qu'il puisse lui permettre d'estimer ces quantités pour différents scénarios de trafic. Nous venons de voir que les arrivées de requêtes, le trafic, peuvent être représentées assez souvent par des processus de Poisson.

Pour ce faire, le modèle mathématique de l'état du réseau devra donner le nombre de connexions en cours sur chacun des liens du réseau, leurs débits respectifs, le nombre de requêtes en attente de traitement sur chaque nœud.... Plus précisément, l'état $X(t)$ à l'instant t du réseau est un vecteur $X(t) = (x_1(t), \dots, x_n(t))$ où $x_i(t)$ est le nombre de communications en attente, en cours, au nœud i d'un réseau de n nœuds. Du fait des arrivées et départs aléatoires, les quantités $x_i(t)$, $i = 1, \dots, n$, sont des variables aléatoires. La fonction $t \to X(t)$ est donc aléatoire. Dans le langage de la théorie des probabilités, on parle de *processus aléatoire* pour marquer que la variable t a une interprétation temporelle. Cette modélisation est une partie importante de l'analyse mathématique d'un réseau. Elle a deux objectifs qui peuvent être antagonistes :

a) La représentation doit être suffisamment détaillée pour pouvoir obtenir des caractéristiques importantes du réseaux, par exemple, les probabilités de perte, les débits obtenus par les communications,... Ainsi, si C_i est le nombre maximum que peut accepter le nœud i, la probabilité que ce nœud soit saturé à l'instant t est donnée par
$$\text{Proba}(x_i(t) = C_i).$$
b) La représentation doit être aussi assez simple pour que des formules mathématiques utilisables puissent être obtenues.

Propriété de Markov. Le processus $t \to X(t)$ est dite posséder la *propriété de Markov*, si, pour un instant $t_0 > 0$, sachant que l'état $X(t_0)$ du réseau à cet instant est un vecteur donné x_0, l'évolution de l'état du réseau après t_0 est indépendante du comportement du réseau avant l'instant t_0. Autrement dit : connaissant le présent (l'instant t_0), les évolutions du réseau avant l'instant t_0 et après l'instant t_0 sont indépendantes. La conséquence d'une telle propriété est que connaissant $X(t_0)$, on peut déterminer l'évolution ultérieure de $(X(t))$ après l'instant t_0 sans connaître ce qui s'est passé avant cet instant.

Cette notion naturelle simplifie beaucoup l'analyse mathématique. Quand celle-ci est satisfaite, elle permet d'écrire un ensemble d'équations différentielles, *les équations de Kolmogorov* qui décrivent l'évolution de la distribution de l'état du réseau au cours du temps. Pour illustrer ces idées, on va considérer le cas simple d'un seul lien de communication.

Exemple d'un lien de communication. On suppose que le lien peut accepter au maximum C communications simultanées et que les arrivées forment un processus de Poisson de paramètre λ et que la durée d'une communication est de moyenne m. Si $X(t) \in \{0, 1, \dots, C\}$ est le nombre de communications en cours à l'instant t, sous certaines hypothèses, le processus $(X(t))$ a la propriété de Markov.

En posant $f_k(t) = \text{Proba}(X(t) = k)$, $0 \le k \le C$, les équations de Kolmogorov sont, dans ce cas,

$$\frac{d}{dt} f_k(t) = \lambda f_{k-1}(t) + \frac{k+1}{m} f_{k+1}(t) - \left(\lambda + \frac{k}{m}\right) f_k(t), \quad 0 < k < C. \tag{2}$$

Si $F(t)$ est le vecteur $(f_k(t), 0 \le k \le C)$, les équations précédentes peuvent s'exprimer sous la forme matricielle compacte

$$\frac{d}{dt} F(t) = Q \cdot F(t),$$

où Q est une matrice fixée.

C'est l'analogue matriciel de l'équation différentielle classique en dimension 1, $f'(x) = \alpha f(x)$ dont la solution est bien sûr $f(x) = f(0) \exp(\alpha x)$. Ici la solution de ce système est $F(t) = \exp(tQ) \cdot F(0)$, mais le calcul de $\exp(tQ)$, l'exponentielle de la matrice Q, n'est pas simple. Les solutions de ces équations ne peuvent donc pas, en général, être exprimées avec des expressions mathématiques utilisables.

Ce système permet cependant de décrire le comportement asymptotique de ce lien, c'est-à-dire la distribution de $X(t)$ quand t est très grand. Un résultat de probabilité montre en effet que le réseau converge assez rapidement vers un état d'équilibre où l'état du réseau reste stable : c'est-à-dire que pour t assez grand la distribution de $X(t)$ ne dépend plus de t. En particulier les limites

$$\pi(k) = \lim_{t \to +\infty} \text{Proba}(X(t) = k), \quad 0 < k < C,$$

existent. La quantité $\pi(k)$ s'interprète comme la probabilité qu'à l'équilibre, il y ait k communications en cours. Si le membre de droite de (2) tend vers 0 quand t devient grand (ce qui est intuitivement plausible puisque les fonctions convergent), on obtient donc la relation satisfaite par $(\pi(k), 0 \le k \le C)$,

$$0 = \lambda \pi(k-1) + \frac{k+1}{m} \pi(k+1) - \left(\lambda + \frac{k}{m}\right) \pi(k), \quad 0 < k < C. \tag{3}$$

Ces équations sont appelées *équations d'équilibre* du processus $(X(t))$.

Il est facile de vérifier que si $\rho = \lambda m$, alors

$$\pi(k) = \frac{1}{Z} \frac{\rho^k}{k!}, \quad 0 \le k \le C,$$

est solution avec Z dite constante de normalisation égale à $Z = 1 + \rho + \rho^2/2 + \cdots + \rho^n/n!$, pour que la somme des $\pi(k)$ vaille 1 (π est une probabilité !).

La probabilité qu'à l'équilibre C communications soient en cours, ou encore la probabilité qu'une demande de communication qui arrive soit rejetée, est donnée par

$$\pi(C) = \frac{\rho^C / C!}{1 + \rho + \rho^2/2 + \cdots + \rho^n/n!}.$$

C'est la célèbre formule d'Erlang. Le trafic entrant est représenté par la quantité $\rho = \lambda m$, la relation précédente permet de choisir C pour que la probabilité $\pi(C)$ de rejet soit plus petite qu'une quantité ε fixée à l'avance.

La formule d'Erlang et ses généralisations sont à la base de beaucoup de méthodes de dimensionnement des réseaux. Des équations similaires à (3) sont aussi valides pour des classes assez générales de réseaux. C'est un des avantages de la représentation de l'état d'un réseau par un processus de Markov, le comportement à l'équilibre du réseau est alors décrit par un ensemble d'équations linéaires.

Conclusion

Nous n'avons donné qu'un aperçu de la richesse des problèmes que posent les réseaux de communication. Les réseaux pair à pair, les réseaux sans fil, les réseaux optiques, les recherches d'information sur le web, chacun dans sa catégorie est à l'origine d'une large gamme de problèmes novateurs qui motivent l'étude de nouveaux modèles mathématiques et quelquefois le développement de nouvelles méthodes pour comprendre le fonctionnement global de ces systèmes complexes. On observe de plus un rapprochement récent avec l'étude des réseaux biologiques. Les réseaux du vivant (cellules, neurones, sociétés d'individus auto-organisées) sont proches en terme de modèles mathématiques mais sont aussi une possible source d'inspiration pour la conception d'algorithmes dans les réseaux de communication. Il est donc clair que, malgré les résultats déjà obtenus, l'étude scientifique des réseaux n'en est qu'à ses débuts.

(Ph. Robert) INRIA Paris — Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay

École Polytechnique, Département de Mathématiques Appliquées, 91128 Palaiseau Cedex
E-mail address: Philippe.Robert@inria.fr
URL: http://www-rocq.inria.fr/~robert