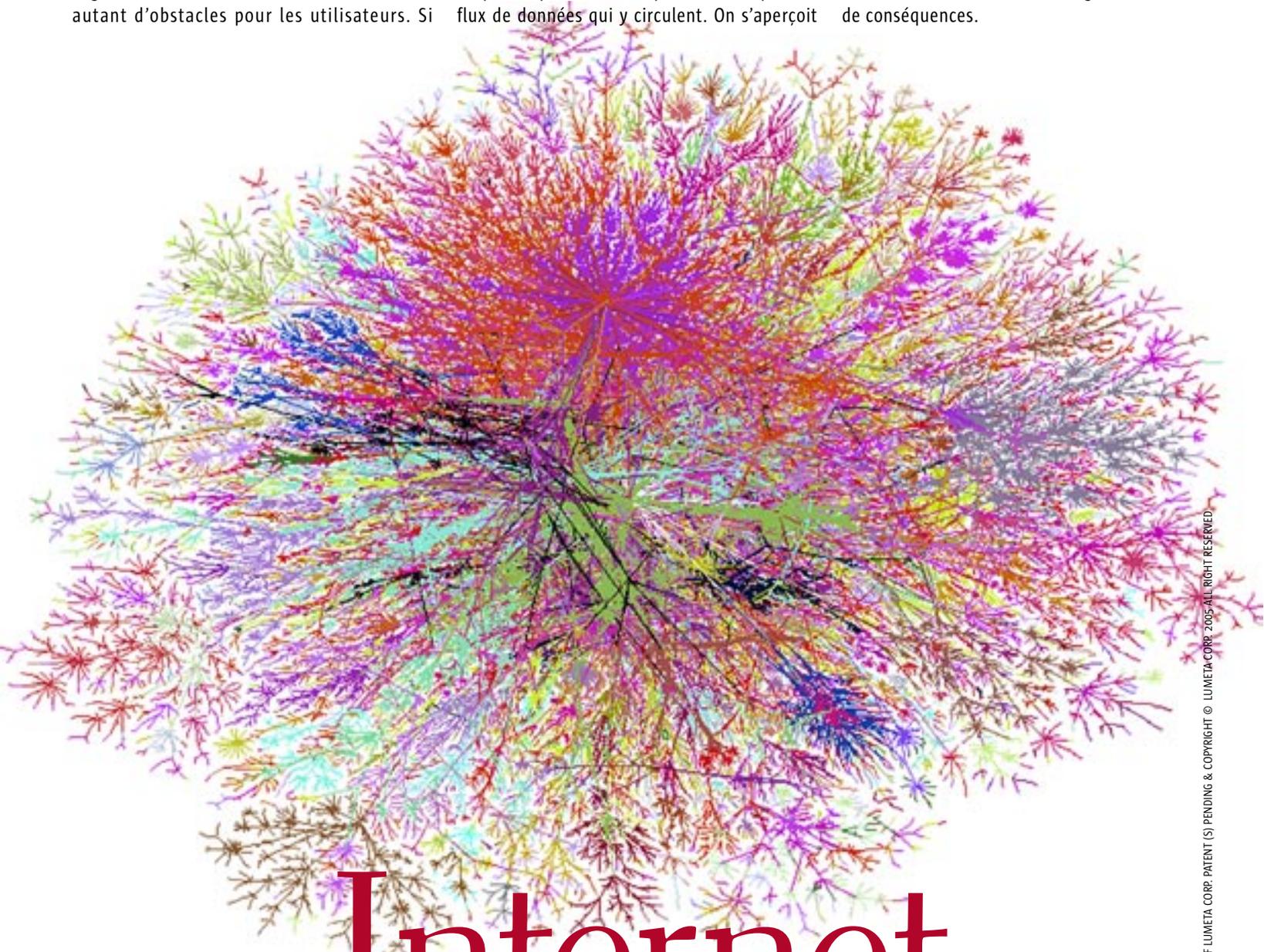


■ **EN DEUX MOTS** ■ Revers de la médaille Internet, la lenteur des connexions, les brouillages ou les transmissions défectueuses sont autant d'obstacles pour les utilisateurs. Si

l'on souhaite accroître la fluidité du trafic et prévenir les pannes du super-réseau, il faut en passer par une description statistique des flux de données qui y circulent. On s'aperçoit

alors que, à l'instar d'autres systèmes complexes, un phénomène caractérise Internet : l'invariance d'échelle. Une singularité lourde de conséquences.



COURTESY OF LUMETA CORP. PATENT (S) PENDING & COPYRIGHT © LUMETA CORP. 2005 ALL RIGHT RESERVED.

Internet

Comment réguler le trafic

Proposée par une équipe franco-australienne, une modélisation statistique particulière au réseau Internet peut permettre, à terme, d'en améliorer le fonctionnement. Reposant sur l'analyse mathématique en ondelettes, elle explore des phénomènes comme l'invariance d'échelle et la mémoire longue, propres aux grands systèmes complexes.

Délais de transmission, fichiers perdus, congestions, voire écroulements partiels : qui n'a jamais expérimenté, impuissant, les failles d'Internet ? L'amélioration du fonctionnement du super-réseau constitue aujourd'hui une priorité, alors que la perspective de raccordement de tous les réseaux de communication, y compris les mobiles, va encore accroître la complexité de ce grand système. Mais, pour cela, il faut savoir décrire Internet le plus précisément possible. C'est-à-dire esquisser une description statistique des échanges d'informations qui s'y déroulent.

Prévoir les performances du réseau

Au début des années quatre-vingt, des équipes réunissant statisticiens, traiteurs de signaux et spécialistes des réseaux se sont attelés à cette tâche. Ils se sont appuyés – c'était logique – sur les concepts et les outils statistiques qui avaient été utilisés avec succès sur le premier grand réseau de communication du monde moderne : le réseau téléphonique. Quelques années plus tard, on s'est malheureusement aperçu qu'ils n'étaient pas en mesure de prendre en compte la complexité d'Internet. La piètre efficacité des modèles statistiques mis en œuvre (processus de Poisson, loi de Gauss, processus de Markov), est patente, par exemple pour cette tâche cruciale de prévision des performances du réseau. On a observé à ce propos une sous-évaluation des délais dans les temps de transmission ou des taux de paquets IP* perdus.

Explication de ces difficultés : Internet, ça n'est pas simple comme un coup de fil ! Au téléphone, le réseau fait circuler un signal de parole au moyen d'un canal unique, fixé par l'opérateur au moment de l'appel et figé pour toute la durée de la communication. Sur Internet, l'envoi d'un mail, la consultation d'une page Web, le téléchargement d'un morceau de musique – toutes activités de nature et de contraintes très différentes – revient à faire circuler un très grand nombre de paquets IP entre machines. Tous ces paquets voyagent de relais en relais, *via* des « routeurs », des « serveurs », des « files d'attente ». Ils sont tour

à tour mélangés avec un grand nombre d'autres paquets, issus de communications sans relation les unes avec les autres, ensuite ils sont dirigés, triés, vers des routes secondaires, des rues – voire des ruelles – jusqu'à leurs destinataires (voir ci-contre). Pour garantir, *in fine*, la reconstitution d'une information sensée, ces mélanges et ces tris de paquets sont, par ailleurs, réalisés *via* des protocoles* qui assurent une bonne qualité du service.

Rien n'est homogène sur le super-réseau : ni sa topologie, ni sa géographie, ni les technologies que l'on y croise, ni le comportement des internautes... L'hétérogénéité est bien le maître mot. C'est cette donnée fondamentale qui induit la complexité d'Internet. Comment, alors, pour

Rien n'est homogène sur le super-réseau : ni sa topographie, ni sa géographie, ni ses technologies



SUR LE WEB, TOUTE ACTION SE RÉSUME EN LA CIRCULATION d'un très grand nombre de « paquets » entre ordinateurs « serveurs » et ordinateurs « clients ». © GETTY IMAGES/BOB STELKO/IMAGE BANK

plus de performances et une meilleure gestion, prendre en compte cette diversité ? Comment modéliser les flux pour mieux utiliser la bande passante*, éviter les pertes d'informations, réduire les délais ?

Parmi d'autres, notre équipe de recherche franco-australienne y a travaillé ces dernières années. Constat de départ : la singularité du trafic Internet par rapport à d'autres systèmes [1] est perceptible lorsque l'on examine les échelles de temps des échanges qui s'y produisent.

Reprenons l'exemple du téléphone et imaginons que l'on effectue la moyenne des flux de données dans des « boîtes temporelles » de taille variable, par exemple, 3, 6, 12, 24, 48 et 96 secondes. On obtient alors, pour les boîtes les plus grandes, un flux moyen lissé quasi uniforme : le trafic est régulier et vous téléphonez sans problème. Mais si vous tentez la même opération sur les flux Internet, l'irrégularité du trafic persistera même pour les boîtes de plus grande taille [fig. 1] : problèmes de transmission en perspective... Cette particularité du réseau Internet est appelée « invariance d'échelle ».

Objets fractals

L'invariance d'échelle se définit précisément par l'impossibilité d'identifier, pour un phénomène donné, une quelconque échelle qui, parce qu'elle jouerait un rôle ⇒

Patrice Abry
est chargé de recherche au CNRS.
Patrice.Abry@ens-lyon.fr

Patrick Flandrin
est directeur de recherche au CNRS. Ils travaillent tous deux au laboratoire de physique de l'école normale supérieure de Lyon.
flandrin@ens-lyon.fr

Darryl Veitch
est senior research fellow au département Darryl Veitch.
dveitch@animelb.edu.au

***Paquet IP :** cellule élémentaire d'information écrite au format spécifique « Internet Protocol ».

***Protocole :** ensemble de règles qui définissent les modalités de fonctionnement d'une communication entre deux ordinateurs.

* **Bande passante :** nombre maximal par seconde d'octets qui peuvent circuler sur un lien Internet.

[1] W.E. Leland *et al.*, « On the self-similar nature of Ethernet traffic (extended version) », *IEEE/ACM Trans. on Networking*, 2, 1, 1994.

[2] K. Park et W. Willinger (dir.), *Self Similar Traffic Analysis and Performance Evaluation*, Wiley, 2000.

[3] *IEEE Signal Processing Magazine*, numéro spécial « Network Traffic : Scaling and Complexity », 19, mai 2002.

[4] S. Mallat, *Une exploration des signaux en ondelettes*, Éditions de l'École polytechnique, 2002.

[5] J.-M. Ghidaglia, « Les ondelettes font des vagues », *La Recherche*, juin 2003, p. 90 ; M. Nowak et Y. Meyer « La surprenante ascension des ondelettes » *La Recherche*, février 2005, p. 56.

[6] D. Veitch, « Lois d'échelle en télétrafic informatique », in P. Abry et al., *Lois d'échelle, fractales et ondelettes* Hermes-Lavoisier, 2, 207, 2002.

[7] P. Abry et al., « Wavelets for the analysis, estimation and synthesis of scaling data », in 1, p. 39-88.

[8] P. Doukhan et al., *Theory and Applications of Long-Range Dependence*, Birkhäuser, 2003.

[9] <http://perso.ens-lyon.fr/patrick.flandrin/Lr05.html>

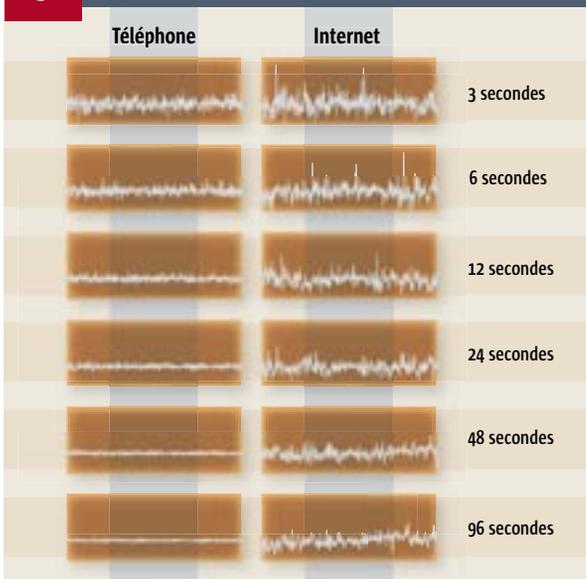
⇒ particulier, serait caractéristique. Autre façon d'exprimer cette idée : toutes les échelles jouent des rôles également équivalents. C'est l'une des raisons de l'inefficacité relative des modèles statistiques cités plus haut qui reposent sur l'existence d'une échelle de temps caractéristique.

D'un point de vue mathématique, l'invariance d'échelle renvoie aux objets fractals. Ces figures géométriques sont, on le sait, obtenues par répétitions d'une même procédure de construction élémentaire sur chaque sous-partie de l'objet. Ils présentent donc eux aussi cette particularité de conserver une même apparence lorsque l'on « zoome », quelle que soit l'échelle à laquelle ils sont observés. Voilà pour le modèle. Quant à l'outil pour l'étudier, l'analyse en ondelettes s'est révélée particulièrement bien adaptée à l'invariance d'échelle [2] [3]. Ce procédé mathématique [4] [5] permet de mettre en évidence et de caractériser de façon fiable une invariance d'échelle en la matérialisant, sur un diagramme, par un segment de droite. Si, comme dans l'exemple du téléphone, le système étudié ne connaît pas d'invariance d'échelle, c'est une courbe qui apparaîtra à la place de la droite sur le diagramme.

Mémoire longue

L'analyse en ondelettes a été appliquée aux données Internet les plus récentes et modernes collectées en différents points du monde et correspondant à des types de trafics très différents [6] [7]. Elle a confirmé l'existence d'invariance d'échelle. Sur le diagramme que nous reproduisons ici [fig. 2], on observe non pas un, mais deux segments de droite. Car il y a non pas un, mais deux phénomènes d'invariance d'échelle qui se développent dans deux gammes d'échelles de temps distinctes. La première, dite des petites échelles, se situe en deçà de la seconde : elle correspond au temps de fonctionnement technologique du réseau (il faut compter par exemple une dizaine de microsecondes pour une transmission de quelques « paquets » d'Internet vers un

Fig.2 Représentation de l'invariance d'échelle



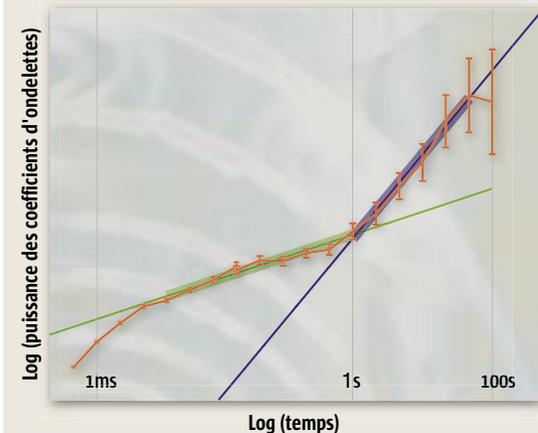
COMME À LA FIGURE 1, ON FAIT LA MOYENNE D'UN FLUX DE TRAFIC, mais les boîtes rectangulaires sont remplacées par des « boîtes oscillantes », les ondelettes. Pour chaque taille de boîte, on mesure la variance (fluctuations autour de la moyenne). Dans un diagramme logarithmique représentant ces variances en fonction de la taille de la boîte, l'invariance d'échelle est mise en évidence par une droite. Pour Internet, on observe deux segments de droite correspondant à deux régimes distincts d'invariance d'échelle, grandes et petites échelles.

réseau local haut débit). La seconde, dite des grandes échelles, au-delà de la seconde, correspond au comportement de l'internaute (qui peut passer quelques heures à surfer), ainsi qu'à la diversité des quantités de données transmises (temps de téléchargement des pages Web). De la plus petite à la plus grande échelle, les temps s'échelonnent sur six ordres de grandeur : une situation exceptionnelle en science expérimentale.

Grandes et petites échelles doivent absolument être distinguées car elles ne peuvent s'interpréter de la même façon. Dans la gamme des grandes échelles, en particulier, on observe une correspondance entre l'invariance et un phénomène dit de « mémoire longue » (ou longue dépendance) [8]. Ce concept signifie que des liens statistiques significatifs peuvent se maintenir sur des intervalles de temps arbitrairement longs.

Pour illustrer cette idée, imaginons un réservoir destiné à réguler les fluctuations d'un fleuve [9]. Si l'on connaît son débit moyen et la taille caractéristique de ses vagues, il suffira de construire un bassin adapté à la durée la plus longue durant laquelle le débit effectif reste au-dessus de la valeur du débit moyen. Mais s'il n'existe pas d'échelle caractéristique de fluctuations, c'est-à-dire si des vagues de durée arbitrairement grande peuvent déferler, il devient impossible de construire un réservoir à la bonne taille : des

Fig.1 Lissage des données, cas du téléphone et cas d'Internet



À GAUCHE, UN FLUX DE DONNÉES DE TYPE « TÉLÉPHONE ». On fait la moyenne du trafic dans des « boîtes » de tailles successives : 3, 6, 12, 24, 48 et 96 secondes. Le flux tend à devenir uniforme pour les boîtes de grandes tailles. À droite, pour le flux « Internet », il y a persistance de l'irrégularité : c'est la signature de l'« invariance d'échelle ».

© INFOGRAPHIES/GRÉGOIRE CIRADE

débordements sont inévitables. Dans le cas d'Internet, ils correspondront à des pertes d'informations dommageables au bon fonctionnement du réseau.

Pour rendre compte de ces phénomènes d'invariance d'échelle et de mémoire longue, de nouvelles approches mathématiques ont été développées. Appliqué au trafic Internet, un premier modèle a été proposé, en 1997, par M.S. Taqqu et ses collaborateurs [10]. Pour caractériser la mémoire longue, il décrit le trafic comme une superposition d'un grand nombre de connexions possédant la particularité suivante: la probabilité qu'elles soient de longue durée est notablement plus élevée que celle que l'on pourrait prévoir avec un modèle classique utilisé dans le cadre du réseau téléphonique. Il s'agit là d'un mécanisme simple, mais fondé sur des données réalistes concernant les durées de connexions, les tailles de documents, etc. [11].

Petites et grandes échelles

Grâce à lui, on peut décrire correctement le fonctionnement du réseau aux grandes échelles de temps, et obtenir des prévisions de taux de remplissage des files d'attente, de pertes ou de délais dans les transmissions beaucoup plus précises que ne le permettent les modèles de type téléphone. Mais, il ne permet pas, en revanche, de faire la lumière sur les comportements aux petites échelles de temps. En deçà de la seconde, le phénomène d'invariance d'échelle ne correspond plus à un phénomène de mémoire longue. Il est encore mal compris. Dans cette gamme, on peut poser l'hypothèse que l'invariance observée est essentiellement d'origine technologique et non humaine, tout simplement parce que, en général, nous ne sommes pas capables d'agir en quelques millisecondes.

Tant les chercheurs qui les étudient que les opérateurs de réseaux qui doivent « faire avec » s'accordent aujourd'hui quant à l'existence de phénomènes d'invariance d'échelle et de mémoire longue dans la gamme des grandes échelles. Il apparaît nécessaire de les caractériser et de les modéliser précisément pour prendre en compte leur nature, leur origine et l'importance de leurs impacts sur le bon fonctionnement du réseau. *A contrario*, les observations et les interprétations de ce qui se passe dans la gamme des petites échelles restent discutées et sont actuellement l'objet d'études plus avancées [12] [13]. Les réponses à apporter dans ce domaine restent encore nombreuses. Quels sont les mécanismes de l'invariance d'échelle dans la gamme des petites échelles? Quelle est l'ampleur de ses impacts et conséquences pratiques pour les opérateurs de réseaux? La séparation des deux gammes a-t-elle vraiment une pertinence? Et finalement l'invariance d'échelle dans la gamme des petites échelles n'offre-t-elle pas seulement une description approximative et partielle d'une réalité plus complexe?

Il semble que l'invariance d'échelle puisse être repérée dans d'autres systèmes complexes



UN SALON INTERNET ET TÉLÉCOMS. Les opérateurs de réseau attendent des scientifiques une description fine de la structure du Web afin de pallier au mieux ses blocages © IAN HANNING/REA

Répondre à toutes ces questions devrait permettre à l'avenir une meilleure compréhension d'Internet et, partant, faciliter sa gestion. Une autre piste, plus incertaine, va plus loin: il semble que le phénomène d'invariance d'échelle puisse être repéré dans d'autres systèmes complexes, physiques (turbulence hydrodynamique), biologiques (variabilité des rythmes du corps humain), géophysiques (répartition des failles géologiques), démographiques (distribution des populations), et même financiers (évolution du cours des actions et des taux de change sur les marchés internationaux). Troublant... On peut à ce propos se poser la question de savoir s'il existe des dénominateurs communs à tous ces systèmes apparemment si différents et si le fait relève ou non du hasard. On pourrait peut-être alors, par analogie, en déduire des solutions applicables à Internet. ■ P. A., P. F., D. V.

- [10] M.S. Taqqu *et al.*, « Proof of a fundamental result in self similar traffic modeling », *Computer Comm. Rev.*, 27, 5, 1997.
- [11] M.E. Crovella *et al.*, « Self-similarity in world wide web traffic - evidence and possible causes », *IEEE/ACM Trans. on Networking*, 5, 835, 1997.
- [12] F. Baccelli, « Internet : modéliser le trafic pour mieux le gérer », in *L'Explosion des Mathématiques*, pp.75-79, SMF-SMAI, 2002.
- [13] N. Hohn *et al.*, « Cluster processes, a natural language for network traffic », *IEEE Trans. On Signal Processing*, numéro spécial *Signal Processing in Networking*, 51(6): 2229-2244, 2003.