Techniques et outils de métrologie pour l'Internet et son trafic

Philippe Owezarski et Nicolas Larrieu
LAAS-CNRS

L'Internet connaît une mutation au niveau de ses usages. De réseau mono-service pour transporter des fichiers binaires ou textuels il y a vingt ans, l'Internet doit aujourd'hui être un réseau multi-services pour le transport de données diverses et variées comme des données audio et vidéo (films, vidéo à la demande, téléphonie, etc.). De fait, il faut opérer une mutation technologique du réseau de façon à le rendre capable de transporter avec des QoS adéquates et multiples les différents types d'informations proposées par toutes les applications utilisant l'Internet. Toutefois, toutes les tentatives pour garantir la qualité de service de l'Internet ont échoué, notamment à cause d'une complète méconnaissance du trafic de l'Internet et des raisons de cette complexité. Au final, tel qu'il existe aujourd'hui, personne n'a la maîtrise, ni même une connaissance, complète du réseau, ce qui va à l'encontre de la mise en œuvre de multiples services de communication à qualité garantie.

La métrologie des réseaux de l'Internet – au sens littéral « la science des mesures » appliquée à l'Internet et son trafic – doit permettre d'apporter une réponse à ces problèmes. En premier lieu, s'il faut fournir des services ayant des qualités prédéfinies, il faut être capable de mesurer ces qualités. D'autre part, la métrologie doit permettre de répondre aux questions concernant le (ou les) modèle(s) de trafic de l'Internet qui font aujourd'hui défaut. Aujourd'hui, la métrologie des réseaux – science récente s'il en est (elle est apparue au début des années 2000) – change tout le processus de recherche et d'ingénierie des réseaux de l'Internet et en devient la pierre angulaire.

La métrologie de l'Internet se décompose en deux tâches distinctes : la première consiste à mesurer des paramètres physiques de la qualité de service offerte par le réseau ou sur le trafic. Dans un réseau de la taille et de la complexité de l'Internet c'est déjà – nous le verrons – une tâche complexe. Toutefois, cette activité de mesure et d'observation ne permet de ne mettre en évidence que des phénomènes visibles. Or en réseau, ce qui est certainement encore plus important c'est d'en déduire les causes, i.e. déterminer les composants et/ou mécanismes protocolaires qui les engendrent. On se retrouve en fait confronté au même problème que Platon dans l'allégorie de la caverne [PLA avJC] : Platon, dans sa caverne, où crépitait un feu de bois, n'apercevait que l'ombre des hommes qui rodaient dans la grotte. Et les ombres projetées sur les parois de la caverne étaient immenses, pouvant laisser croire qu'elles étaient celles de géants. La métrologie réseau – mesure de la QoS ou analyse simple du trafic – nous confronte à ce problème « platonien » : elle ne nous montre que les effets de toute la mécanique des réseaux alors que ce qui nous intrigue, ce sont les causes de ces effets, les phénomènes qui les engendrent. C'est cette tâche – sans aucun doute la plus délicate et la plus importante – qui constitue le second volet de la métrologie car en mettant en évidence les causes des lacunes de l'Internet, on trace les voies de recherche pour faire évoluer les mécanismes, architectures et protocoles de l'Internet.

Cet article introduit donc les principes de base de l'Internet et de son trafic, les besoins et les métriques physiques. Il montre les différentes techniques de mesure actives et passives, leurs besoins, leurs qualités et défauts, cite un ensemble d'outils réel utilisables ainsi que leur mise en place sur le réseau RENATER dans le cadre du projet de métrologie français METROPOLIS. Puis, à partir de ces mesures ou observations sur le trafic, l'article montre une approche pour analyser le trafic qui met en évidence les causes des limitations actuelles, démontrant ainsi l'importance de la métrologie pour la recherche et l'ingénierie des réseaux.

1. De la complexité de réaliser des mesures en environnement Internet

1.1. Caractéristiques générales du trafic IP

Un réseau de type Internet doit aujourd'hui être multi-services : il a vocation à transporter un grand nombre de types de service possédant des caractéristiques de trafic différentes et des contraintes de QoS différenciées. Cependant, dans le souci d'une modélisation simplifiée autant que pour les besoins opérationnels de gestion du réseau, on recherche plutôt une classification grossière des différents types de trafic [ROB 00]. La plupart des auteurs s'accordent généralement pour distinguer deux grandes classes de trafic de télécommunications dans les réseaux à haut débit :

- Le trafic de type « **streaming** », dont la durée et le débit ont une réalité intrinsèque bien que variable éventuellement. Souvent associé à la notion de services « orientés connexion », son intégrité temporelle doit être préservée par le réseau. Le délai de transfert des données de même que sa variation, la gigue, doivent être contrôlables, tandis qu'un certain degré de perte de paquets peut être tolérable. Les flux de trafic streaming sont typiquement produits par les services téléphoniques et vidéo (vidéoconférence ou téléchargement « on-line » de séquences).
- Le trafic dit « **élastique** », ainsi nommé car son débit peut s'adapter à des contraintes extérieures (bande passante insuffisante par exemple) sans pour autant remettre en cause la viabilité du service. Cette classe de trafic est essentiellement engendrée par le transfert d'objets numériques par nature (par opposition au transfert en mode numérique d'informations analogiques à la source) tels que des pages Web (application HTTP), des messages électroniques (e-mail, application SMTP) ou des fichiers de données (application FTP). Le respect de leur intégrité sémantique est indispensable mais les contraintes de délai de transfert sont moins fortes. Cette intégrité sémantique est la plupart du temps assurée par le protocole de transport (TCP) et ne constitue donc pas un élément de performance sur lequel l'opérateur de réseau puisse agir ; en revanche, le maintien d'un certain débit effectif minimum de transfert des documents est un objectif de QoS.

Le trafic de type élastique est actuellement largement majoritaire sur les réseaux IP : on constate couramment [THO 97] des proportions supérieures à 95% en volume (octets) et à 90% en nombre de paquets pour le trafic TCP, protocole avec lequel fonctionnent la plupart des applications mentionnées ci-dessus.

L'analyse des caractéristiques du trafic Internet s'effectue commodément en se plaçant à un niveau de représentation selon trois entités de trafic, correspondant à trois échelles de temps différentes et, quoique de manière assez grossière, à trois niveaux (couches) de la pile protocolaire des réseaux de données :

- Les « **paquets** » forment l'entité de trafic la plus fine que l'on considère dans les réseaux de données, le paquet étant l'unité élémentaire traitée par la couche « réseau ». Les paquets sont *a priori* de longueur variable dans un réseau IP et leur processus d'apparition est très complexe, en raison notamment de la superposition de services de nature très diverse et de l'interaction des couches protocolaires (dispositifs de contrôle de flux et de retransmission sur perte de paquets, tels TCP [BLA 92]). Comme nous le verrons dans la partie 4, le trafic au niveau paquet possède la caractéristique unanimement reconnue

d'auto-similarité, laquelle rend très ardue l'évaluation de ses performances à ce niveau. Nous détaillerons donc des outils capables à partir de mesures du trafic d'évaluer cette caractéristiques du trafic, de la qualifier et de la quantifier. Les échelles de temps décrivant le processus des paquets sont la microseconde et la milliseconde, en fonction des ordres de grandeur du débit de transmission des liens.

- Les « **flots** » constituent une entité de trafic intermédiaire que l'on pense être la mieux adaptée pour effectuer les études d'ingénierie du trafic IP. Ils correspondent à des transferts plus ou moins continus de séries de paquets associés à une même instance d'une application donnée. Les flots de type streaming sont associés à des communications audio/vidéo (téléphonie sur IP, vidéoconférence) ou encore à des téléchargements en temps réel de séquences vidéos. Les flots de trafic élastique sont créés par le transfert d'un fichier, d'un message, d'un objet (ou document) au sein d'une page HTML, etc. Un flot correspond donc plus ou moins à la couche transport de la pile protocolaire Internet ; mais pas complètement puisque cette notion n'est pas nécessairement équivalente à celle d'une connexion TCP, p. ex., comme on le verra par la suite. On peut estimer que les flots ont une durée s'étendant de quelques secondes à quelques minutes, voire quelques heures.
- Au plus haut niveau, on peut tenter de définir la notion de « **sessions** » dans le but de se rapprocher des périodes d'activité des utilisateurs (transposition de la notion d'appels considérée en téléphonie à commutation de circuits). Pour le trafic streaming, ce niveau ne se distingue guère de celui des flots, du moins temporellement, puisque ce dernier correspond déjà à des communications ou des appels. S'agissant du trafic élastique, les sessions peuvent être associées à des connexions Telnet, FTP, ou à des envois de messages électroniques. La notion de session est (encore) plus floue au sujet des connexions de type WWW selon le protocole HTTP : on peut par exemple la définir comme étant la durée de transfert d'une page HTML dans son ensemble (comportant plusieurs objets à transférer) ou d'une suite de pages associées à une même consultation. Les sessions sont générées par la couche application des réseaux et l'ordre de grandeur de leur durée se situe entre quelques minutes et quelques heures.

Il est donc important en métrologie d'analyser le trafic Internet à ces 3 niveaux. Toutefois, en termes de mesures, les informations des 3 niveaux sont contenues dans les informations du niveau le plus bas, soit le niveau paquet. Dès lors, les mesures à proprement parler sont généralement faites au niveau paquet, et ce sont ensuite des logiciels de traitement des traces et mesures qui permettent d'en extraire les informations de niveaux flots et sessions.

1.2. Les métriques

La QoS est aujourd'hui délicate à définir, et il n'existe pas de consensus au niveau des acteurs de la communauté réseaux. En effet, beaucoup limitent la définition de la QoS au triptyque :

- Délais de transmission des paquets ;
- Débit ;
- Taux de perte au niveau des paquets.

Cette définition se limite à considérer des paramètres physiques de la transmission de paquets de données. Toutefois, de nombreux acteurs du domaine des réseaux considèrent également que la QoS intègre des paramètres plus complexes (et moins physiques) comme :

- la disponibilité des services réseau, i.e. la possibilité pour un utilisateur d'envoyer une information avec une QoS donnée, à tout moment ;
- la sécurité des transmissions, i.e. la capacité du réseau à assurer que les messages ne seront pas lus ou altérés par un pirate informatique ;

- la robustesse du réseau, i.e. sa capacité à continuer à fonctionner correctement et avec le niveau de performance et de QoS requis même en cas de panne de certains de ses composants, ou en cas d'attaque ;
- etc.

En ce qui concerne la métrologie, et sa première étape qui est la mesure ou la collecte d'informations, il est évident que ce sont des paramètres physiques qui vont pouvoir être mesurés. Les paramètres plus complexes, dont la définition et les métriques pourront être fixées par chacun, seront en fait déterminés à partir de la mesure de ces paramètres de base. Dans cette première partie qui se focalise sur la mesure et l'observation des effets du réseau sur le trafic et sa QoS, nous allons donc nous limiter à la mesure et l'observation du triptyque (délai, débit, perte).

Toutefois, le terme QoS peut avoir différentes connotations selon la personne qui l'emploie. Pour simplifier un peu la problématique, on peut considérer qu'il existe deux points de vue principaux :

- **Le point de vue des applications** (ou des utilisateurs) : chaque application a des besoins en termes de débit, délai, gigue, fiabilité, etc. Ces besoins sont naturellement différents d'une application à l'autre, et chaque application souhaiterait pouvoir bénéficier d'un service de communication spécifiquement adapté à ses besoins.
- **Le point de vue des opérateurs** dont les objectifs sont d'optimiser l'utilisation des ressources de l'infrastructure de communication (et ainsi de maximiser leurs gains), de limiter les pertes et les délais, et de pouvoir facturer de façon juste et cohérente les services rendus aux utilisateurs.

Enfin, il existe une dernière complexité à la mesure de ces paramètres de QoS physiques (aussi qualifiés de simples) et qui est liée à la dimension de l'Internet, composé d'une multitude de réseaux interconnectés, eux mêmes composés de nombreux nœuds et de points de présence dans les différentes villes du monde desservies. Ainsi, il est important de décomposer les mesures en deux nouvelles classes :

- les mesures de bout en bout qui concernent le triptyque {délai, débit, perte} entre deux utilisateurs terminaux. Ce sont des mesures qui rejoignent le point de vue des applications précédemment cité ;
- les mesures de proche en proche qui concernent le triptyque {délai, débit, perte} entre des nœuds intermédiaires du réseau (routeurs, commutateurs, passerelles, etc.). Ce sont des mesures qui rejoignent le point de vue des opérateurs.

La suite donne une définition des paramètres du triptyque physique {délai, débit, perte} tout en en donnant les besoins liés aux différentes applications qui existent dans l'Internet. Cela permettra au lecteur d'appréhender la grande diversité des besoins de ces applications, et de toucher du doigt la complexité de l'analyse du trafic Internet composé de toute cette variété de classes de trafic différentes.

1.2.1 Besoins en terme de fiabilité

Les multiples applications de l'Internet, qui utilisent de nombreux médias, ont donc des besoins très variables. Ainsi, les medias continus (audio et vidéo) ont comme caractéristiques d'être plus ou moins redondants. Ainsi deux images successives d'une transmission vidéo comportent généralement peu de différences. De cette redondance résulte la possibilité que des pertes d'information (image) soient acceptables du point de vue de l'utilisateur final. Il

apparait donc ici pour les medias les plus importants d'une application multimédia (audio et vidéo) une contrainte de fiabilité du transfert des données non plus totale mais partielle, la perte de certaines informations pouvant être acceptable. Notons cependant que l'expression d'une contrainte de fiabilité partielle est à coupler avec la façon dont sont codées les données audio et vidéo. En effet, certains codages (MPEG par exemple) introduisent une dépendance entre les images qui peut rendre indécodables plusieurs images consécutives en cas de perte de l'une d'entre elles, plus importante que les autres. A l'inverse, un codage de type M-JPEG n'introduisant aucune dépendance entre les images, une contrainte de fiabilité exprimée en termes d'un pourcentage maximum de pertes admissibles et d'un nombre maximum de pertes consécutives s'avère alors valide.

1.2.2 Besoins temporels

Les applications de diffusion différée de médias continus traitent leurs données de la même façon que s'il s'agissait de medias discrets. Deux types d'applications multimédias présentent des contraintes temporelles : les applications de diffusion en temps réel de médias continus et les applications multimédias interactives. Les contraintes temporelles s'expriment généralement par le biais de deux paramètres : le délai de transit des données et la gigue.

- Délai
  Pour les applications interactives (telle que la visioconférence), et à un degré moindre pour les applications de diffusion en temps réel (telles que le streaming audio ou vidéo), afin que la communication se déroule comme si elle avait lieu localement, il faut que les données soient transmises en un temps inférieur au seuil de perception humain lié au media considéré. Il apparaît ainsi une contrainte sur le délai de bout en bout du transfert des données.
- Gigue
  Les medias continus (tels que l'audio et la vidéo) présentent des contraintes temporelles dues à leur caractère isochrone. Ces contraintes s'expriment en terme de régularité dans l'arrivée des données (on suppose que la source des données émet à un débit correspondant au débit idéal de présentation). Cette régularité s'exprime par une contrainte sur le temps inter-arrivées des données, c'est-a-dire sur la différence entre les dates d'arrivée de deux données successives. Cette contrainte est appelée la « gigue ». La date d'arrivée d'une donnée étant calculée par :

  $$t_{réception} = t_{émission} + dt_{min} + \delta dt$$

  où :
  - dt_{min} désigne le temps de transmission optimal (sans attente dans le réseau) ;
  - δdt désigne le temps d'attente dans le réseau.

  On peut alors exprimer la gigue par :

  $$t_{inter\ réception} = t_{inter\ émission} + \delta dt_2 - \delta dt_1$$

  où :
  - δdt_1 et δdt_2 représentent les temps d'attente dans le réseau pour deux données successives.

1.2.3 Besoins en terme de débit

Les besoins des applications en terme de débit sont très variables. Certaines, comme les applications Web ou mail ne requièrent que quelques kiloOctets de bande passante pour les flux qu'elles échangent. D'autres, au contraire, sont beaucoup plus exigeantes. Bien sûr, c'est cette dernière catégorie qui tend à se généraliser car de plus en plus d'applications récentes nécessitent une bande passante importante. On peut situer les applications de diffusion en temps réel comme par exemple les chaînes de télévision sur Internet. Dans ce dernier cas, il

est d'ailleurs primordial de pouvoir fournir un service le plus stable et le plus régulier possible de façon à ce que l'utilisateur à l'extrémité du réseau reçoive son flux multimédia avec un niveau de qualité le plus régulier possible.

1.3. Points durs techniques

Toutefois, la mesure de paramètres « physiques » simple – sans parler de l'estimation de paramètres complexes à partir de ces mesures qui sera présentée ultérieurement (partie 4) – reste un problème complexe dans l'Internet. Cette partie dresse une liste la plus exhaustive possible des problèmes techniques qui se présentent pour réaliser des mesures sur le réseau planétaire qu'est l'Internet.

1.3.1 Dimension géographique et administrative

De façon évidente, les grandes difficultés inhérentes à la réalisation de mesures de trafic ou de QoS sont liées à la taille du réseau et à son étendue géographique. Ainsi, il est indispensable d'avoir de très nombreux points de mesure dans le réseau. Regarder le trafic en un seul point à un moment donné est finalement assez peu significatif. Les communications sur l'Internet se font souvent sur de très longues distances, et ne regarder ces communications qu'en un seul point est très réducteur. Toutefois, l'Internet est une interconnexion de réseaux, et à ce titre de multiples opérateurs sont propriétaires et en charge de la gestion et de l'opération d'un de ces réseaux qui composent l'Internet. Aussi, il est impossible aujourd'hui de positionner des outils de mesure en certains points du réseau contre la volonté de l'opérateur qui possède et gère ce point. Pour pouvoir faire de la métrologie du réseau, il faudrait pouvoir disposer de points de mesures au niveau de tous les nœuds du réseau, mais la construction de l'Internet l'interdit par défaut. De plus, monitorer tous les nœuds de l'Internet est une tâche titanesque. Pour contourner cette difficulté, de nombreux travaux sont en cours depuis 3 à 4 ans sur les problèmes d'échantillonnage à la fois spatial et temporel. L'idée serait de trouver des techniques qui à partir de quelques points de mesure, qui ne fonctionneraient que pendant des périodes de temps courtes, permettraient d'estimer de façon précise le trafic sur l'ensemble du réseau 24 heures sur 24. Mais aujourd'hui, malgré les efforts consentis, les résultats en matière d'échantillonnage sont au point mort, et d'ailleurs, nous n'en parlerons pas dans le reste de cet article par manque de résultats positifs significatifs.

De plus, l'étendue planétaire de l'Internet entraîne des variations journalières de l'activité. Cela n'a donc que peu de sens de ne regarder qu'un seul point du réseau, car une connexion inter-continentale peut être confrontée sur un des continents où la nuit est tombée à un trafic potentiellement plus réduit (et qui posera donc peu de problèmes par rapport au respect de contraintes de QoS), puis traverser quelques millisecondes plus tard un continent aux heures pleines de la journée, et se heurter ainsi à un trafic conséquent et à de possibles phénomènes de congestion très problématiques pour la QoS de la communication. Dans un tel contexte, il est donc indispensable de disposer de points de mesure sur ces différents continents, car le trafic local peut avoir des conséquences importantes sur la connexion inter-continentale que nous considérions. De la même façon, on retrouve le problème évoqué dans la partie 1.2. au sujet des mesures de bout en bout ou de proche en proche. On voit clairement sur cet exemple de la connexion inter-continentale que la mesure de bout en bout ne donne qu'un résultat global de la QoS effective pour cette connexion, alors que les informations de proche en proche permettent de localiser le problème entrainant les limitations de service.

1.3.2 Problèmes liés à la mesure dans des systèmes distribués

Une autre classe de problèmes à laquelle on se retrouve confronté lorsque l'on souhaite réaliser des mesures sur l'Internet est celle liée aux systèmes distribués – problèmes d'autant plus délicats à résoudre que l'Internet est vaste, et surtout non maîtrisé dans son ensemble par

une entité unique. Ainsi, un des principaux problèmes à régler concerne la possibilité d'avoir recours à une référence temporelle unique. En effet, si les horloges des machines sonde impliquées dans une mesure de délai par exemple sont désynchronisées, le résultat n'aura aucune valeur et ne sera pas exploitable. Il existe bien des protocoles de synchronisation d'horloges en réseau, dont le plus connu et utilisé est NTP [MIL 96], mais les évaluations de performance de ce protocole ont montré que même s'il parvenait à des résultats assez satisfaisants pour les réseaux locaux, il est totalement insuffisant sur les réseaux longues distances étendus comme l'Internet. En l'absence de protocoles de synchronisation d'horloges sur des réseaux longues distances plus performants que NTP, c'est tout un pan de la recherche sur les systèmes distribués à larges échelles et étendus géographiquement qui doit être inventé. Ces recherches ne peuvent a priori explorer que deux axes :

- Soit trouver des mécanismes qui soient capables de déterminer le décalage et la dérive qui existe entre les horloges impliquées dans une mesure entre deux sondes distantes, de façon à pouvoir ensuite corriger les valeurs retournées par les sondes. C'est en fait la stratégie qui est employée par NTP, mais avec un algorithme inadapté à l'Internet.
- Soit trouver un moyen de synchroniser physiquement et très précisément les horloges de toutes les sondes de mesure et métrologie sur une référence temporelle unique.

Nous verrons par la suite quelle solution nous proposons pour résoudre ce problème de synchronisation des horloges.

Un autre problème que l'on rencontre dans les systèmes distribués – et encore accentué dans l'Internet à cause de sa taille – est lié à la localisation des mesures qui ne sont pas nécessairement faites sur le point où elles vont être exploitées. Ces mesures devront donc être rapatriées sur le lieu de leur exploitation. Ce rapatriement engendre au moins deux problèmes :

- Le premier est lié à la quantité de trafic que cela peut engendrer sur le réseau. En effet, même si parfois, par exemple sur la mesure d'un délai, seule une valeur scalaire est transmise par les sondes de mesure et n'engendre donc pas un surplus de trafic important, il se peut que l'information à rapatrier soit une trace de paquets complète qui peut elle représenter plusieurs MégaOctets, voir plusieurs GigaOctets. Dans un tel cas, la signalisation des informations de mesure est loin d'être transparente pour le réseau et son trafic.
- Le second se présente lorsque l'on souhaite exploiter les mesures effectuées en temps réel. Naturellement, le rapatriement des données sur leur lieu d'exploitation prend du temps… Plus ou moins selon les cas. Mais cependant assez pour se poser la question de la validité temporelle de cette mesure. Dans un réseau à haut débit – ce qui est le cas dans l'Internet aujourd'hui – et dont le trafic varie beaucoup et très rapidement, il est légitime de se demander si la valeur de la mesure reçue est encore valide. C'est en fait un problème que l'on rencontre dans les réseaux et les systèmes distribués du fait que les messages de contrôle ou de signalisation dans le réseau utilisent le même support de transmission que les données de communications des utilisateurs… Dans le cas de l'Internet, messages de signalisation et données utilisateurs sont transportés dans les paquets qui empruntent les mêmes liens, passent par les mêmes routeurs, etc. En faisant une analogie avec le trafic routier, cela signifierait par exemple que pour déterminer le temps que l'on va mettre un samedi pour rallier Paris à Marseille, on envoie un véhicule effectuer ce même trajet dans la nuit précédent le jour du départ. Naturellement, la valeur obtenue le vendredi soir peut ne pas être très différente de ce qu'elle sera le samedi. Mais si le samedi en question est le premier samedi du mois d'août avec le tristement célèbre chassé-croisé des vacances d'été, il est évident que la valeur obtenue le vendredi soir est sans rapport avec celle que mettra le véhicule le samedi. La brusque augmentation du trafic automobile entre Paris et

Marseille le premier samedi du mois d'août, avec la création de bouchons (assimilables aux congestions du réseau), fait que la mesure effectuée la nuit auparavant est sans valeur. Or en réseau, il n'existe pas de médium parallèle pour transporter les messages de contrôle ou de signalisation urgents, ce qui va rendre les performances des mécanismes de gestion du réseau ou de signalisation des mesures sensibles au trafic de données existant sur le réseau. C'est là aussi un problème majeur qu'il faudra régler pour pouvoir mettre en œuvre et déployer des systèmes de mesure dans l'Internet, et en exploiter efficacement les résultats.

1.3.3 Mesures, estimations et analyses

Ce problème nous amène tout naturellement au problème suivant pour les mesures qui concerne la dépendance des techniques de mesure ou leur calibration à la nature du trafic. Ainsi, il paraît évident que l'on ne pourra pas du tout utiliser les mêmes solutions pour mesurer du trafic ADSL à 2 Mbps et le trafic d'un opérateur de cœur de réseau à plusieurs dizaines de Gbps. De la même façon, la granularité d'observation ne pourra pas être la même dans les deux cas. Egalement, cette granularité devra aussi être adaptée à d'autres caractéristiques du trafic : par exemple, si la taille moyenne des flux transmis sur différents liens varie, il sera certainement judicieux d'adapter la granularité d'observation à la taille des flux transportés.

Finalement, et c'est certainement là le problème principal de la mesure et de la métrologie dans les réseaux, il est quasiment impossible de faire des mesures sur la multitude de mécanismes et protocoles que comporte l'architecture appelée TCP/IP de l'Internet. Ce problème général évoqué dans l'introduction de l'article fait que la mesure seule ne permet pas d'obtenir des informations sur le réseau et son comportement. Elle ne permet de ne voir que les effets des mécanismes et protocoles, ce qui est loin d'être satisfaisant. La mesure doit donc être suivie d'une phase au cours de laquelle des méthodes et des techniques de métrologie, par exemple basées sur la caractérisation et l'analyse du trafic et/ou des mesures de QoS, vont permettre d'expliquer les causes des phénomènes observés. Et c'est cette impossibilité avec des mesures physiques simples qui justifie le plan de cet article qui présente d'abord les techniques de mesures physiques simples (parties 2 et 3) mais surtout – et c'est la contribution la plus importante de cet article – les techniques de métrologie (caractérisation et analyse) qui vont permettre d'estimer les causes des phénomènes observés sur le trafic ou la QoS du réseau (partie 4).

1.4. Points durs juridiques

Toutefois, les difficultés avec les mesures et la métrologie ne s'arrêtent pas aux problèmes techniques. De nombreux problèmes juridiques se posent. Le premier a été imposé par le législateur français avec la création de la CNIL, laquelle interdit de créer des fichiers contenant des données sur les individus et leurs usages. Typiquement, la collecte de traces de trafic entre sous le coup de cette loi. Il faut donc, pour être en toute légalité, demander une autorisation à la CNIL pour constituer de tels fichiers de traces. En effet, une trace contient les adresses IP des machines source et destination, et à ce titre, les ordinateurs étant de plus en plus individuels, cela identifie une personne physique, dont on pourra analyser les faits et gestes sur le réseau. Cependant, le débat est ouvert sur cette question car avec les systèmes multi-utilisateurs actuels, il faut clairement définir si juridiquement une adresse IP identifie vraiment une personne physique ou non ? La question n'a pas encore été tranchée, et comme souvent avec les nouvelles technologies, le flou juridique le plus total est de rigueur. Dans les faits, et à notre connaissance, peu de chercheurs ou d'ingénieurs en réseau qui collectent des traces de trafic n'ont demandé d'autorisation à la CNIL. D'ailleurs, ceux qui ont effectué une

demande attendent encore la réponse, la CNIL ne se pressant pas pour légiférer sur cette question. Légalement, la collecte de trace est donc a priori interdite. Il convient donc à toute personne souhaitant collecter du trafic d'être très prudente avec l'utilisation qu'elle en fera. A priori, dans notre cas qui consiste à n'analyser les traces que par rapport aux aspects techniques relatifs aux protocoles et mécanismes réseaux à des fins d'amélioration, nous n'attentons pas au respect de la vie privée[1]. Il y a donc peu de chances qu'une plainte soit déposée, ce qui semble laisser penser que notre activité en métrologie est légale. C'est du moins l'interprétation que nous avons faite de la loi avant de démarrer nos travaux de recherche en métrologie. Cet article ne contient donc a priori que des informations dont la divulgation ne porte atteinte à personne. Par contre pour quelqu'un qui utiliserait des traces de trafic pour connaître les usages d'un individu ou d'une famille comme les sites web consultés ou les fichiers téléchargés, il serait clairement en violation du droit d'autrui à la vie privée. Toutefois, le flou existe à ce niveau là, car il semble que l'on se dirige vers le devoir pour les FAI de dénoncer leurs utilisateurs qui téléchargeraient des films ou des musiques non libres de droit[2]. Cela semble ouvrir une brèche dans la surveillance des internautes si cette loi est votée un jour. Mais le débat fait rage à l'heure actuelle, et bien malin est celui qui pourrait en deviner l'issue.

Toutefois, même en restant sur des aspects d'analyse technique des réseaux, il est très difficile de convaincre les opérateurs de donner accès à des traces de trafic provenant de leur réseau. En effet, ces traces contiennent, pour un métrologue averti de très nombreuses informations sur l'architecture du réseau, les mécanismes protocolaires utilisés et les politiques de gestion mises en place. De la même façon, cela donne des informations sur la QoS que ce réseau peut offrir aux utilisateurs, sur sa robustesse face aux attaques, etc. Or la fourniture d'accès et de services Internet est un domaine économique très concurrentiel. Révéler les secrets de conception et d'administration d'un réseau revient donc à révéler un secret industriel et les concurrents pourraient ainsi en profiter en copiant cet opérateur. Egalement, révéler la qualité de service que peut fournir un réseau – surtout si elle est basse – ou les failles de sécurité peut aussi porter préjudice à un opérateur qui pourrait être boudé par les utilisateurs potentiels qui lui préfèreront un concurrent. Révéler les secrets de conception d'un réseau pourrait également servir à des pirates informatiques pour l'attaquer de façon imparable. Finalement, ces informations de métrologie sont très stratégiques pour les opérateurs réseau et FAI. Ils les gardent donc jalousement, et contrôlent très fortement les informations qu'ils rendent publiques. C'est également le cas pour les réseaux de la recherche, qui sont pourtant des réseaux publics. Leurs administrateurs veillent à ce que les informations que nous publions ne soient pas de nature à porter préjudice aux chercheurs, enseignants-chercheurs et étudiants de nos universités et écoles d'ingénieurs, par exemple en révélant des usages ne respectant pas la charte de ces réseaux pour l'enseignement et la recherche.

---

[1] De plus, pour encore plus se protéger, il est possible pour les chercheurs ou ingénieurs qui collectent des traces de trafic de les anonymiser, i.e. de transformer les adresses IP contenues dans les traces par un processus irréversible, mais qui a la propriété de conserver la structure hiérarchique de l'adressage Internet. Ainsi, les ingénieurs et chercheurs pourront continuer à mener leurs recherches sur le routage ou les matrices de trafic par exemple, tout en se prémunissant contre toute possibilité d'être accusé de violer le droit à la vie privée des individus qui ont généré le trafic capturé (puisque le mécanisme d'anonymisation irréversible interdit de revenir aux adresses initiales et de pouvoir donc identifier la personne physique derrière sa machine). Nous reviendrons sur ces mécanismes d'anonymisation dans la partie 3.2.2.

[2] Déjà, un ingénieur ou un chercheur en réseau qui dans ses traces découvrirait un individu qui se livre à des activités pédophiles pourrait le dénoncer sans risque d'être accusé d'avoir violé son droit à la vie privée. Mais le cas de la pédophilie est aujourd'hui à notre connaissance la seule exception pour laquelle la faute de l'individu est de très loin supérieure à celle de l'ingénieur. Ce n'est par exemple pas le cas pour la dénonciation d'activités terroristes pour lesquelles seuls les services secrets et forces de l'ordre sont habilités.

Toutes ces limitations sont problématiques dans notre travail quotidien. Toutefois, elles ne le sont pas dans cet article qui ne traite que des techniques de mesure et des méthodes de métrologie et d'analyse. Les exemples que nous donnons pour illustrer ces propos utilisent des traces de plus de 12 mois qui ont de fait perdu de leur intérêt stratégique vue la vitesse d'évolution des réseaux de l'Internet. Toutefois, elles restent très intéressantes pour valider nos méthodes d'analyse et de métrologie.

2. Techniques et outils de mesure et métrologie

Après avoir posé les problèmes techniques et juridiques, cette partie va maintenant montrer comment on peut y apporter une solution. Nous nous focaliserons sur les seuls problèmes techniques, les problèmes juridiques n'étant pas de notre domaine de compétence. Cette partie va donc montrer les différentes techniques qui ont été conçues et développées au cours de ces dernières années pour mesurer les paramètres physiques de base dans les réseaux, ou collecter des traces de trafic. Le chapitre 3 illustrera la mise en place d'équipements de mesure sur Renater dans le cadre du projet METROPOLIS. Ensuite, le chapitre 4 et ses suivants traiteront des méthodes d'analyse et de métrologie qui permettront d'aller plus loin dans la compréhension des phénomènes qui sont observés dans le réseau.

2.1. Mesures actives

2.1.1. Définition

Le principe des mesures actives consiste à générer du trafic dans le réseau à étudier et à observer les effets des composants et protocoles – réseaux et transport – sur le trafic : taux de perte, délai, RTT, etc. Cette première approche possède l'avantage de prendre un positionnement orienté utilisateur. Les mesures actives restent le seul moyen pour un utilisateur de mesurer les paramètres du service dont il pourra bénéficier.

2.1.2. Problématique associée aux mesures actives

L'un des inconvénients majeurs pour le réseau avec les mesures actives est la perturbation introduite par le trafic de mesure qui peut faire évoluer l'état du réseau et ainsi fausser la mesure. En effet, le résultat de la mesure donne une information sur l'état du réseau transportant à la fois les données normales des utilisateurs et de signalisation du plan de contrôle du réseau, mais également l'ensemble des paquets sondes. Or on souhaiterait avoir une information qui correspond au trafic normal uniquement, sans les paquets sondes lesquels ont forcément un impact sur les performances du réseau. Il faut donc soit être capable d'estimer l'impact des paquets sondes sur les performances du réseau, soit être sûr que ces paquets sondes auront un impact minimal, si possible quasi nul. C'est cette dernière proposition, a priori plus simple, qui suscite le plus d'efforts de recherche. On parle de trafic de mesure non intrusif. Ainsi, de nombreux travaux menés actuellement abordent ce problème en essayant de trouver les profils de trafic de mesure qui minimisent les effets du trafic supplémentaire sur l'état du réseau. C'est par exemple le travail en cours au sein du groupe IPPM de l'IETF [PAX 98] [ALM 99a] [ALM 99b] [ALM 99c].

Un autre élément qui se pose dans ce genre de mesures est lié à la vitesse de convergence des mesures vers un résultat dont la fiabilité est bonne. En effet, pour mesurer certains paramètres, il faut parfois mettre en œuvre tout un processus complexe pour approcher de la solution. Par exemple, pour mesurer le débit disponible sur un chemin entre une source et une destination, certains outils transmettent des débits de paquets sondes qui augmentent à chaque tentative jusqu'à ce que des pertes apparaissent, ces pertes étant

assimilées à des phénomènes de congestion. La valeur du débit généré est ainsi la valeur retournée par l'outil de mesure comme étant le débit disponible sur le chemin. Toutefois, le processus peut être long, et dans le cas où le trafic transporté sur le chemin est très variable, le résultat est peu fiable. Parfois même il est possible que l'outil ne converge pas vers un résultat. Converger rapidement est donc essentiel pour pouvoir connaître précisément les variations du trafic sur un chemin.

Enfin, la précision est un élément déterminant. Si les mesures ont des intervalles de confiance importants, les résultats sont sans intérêt pour les chercheurs, ingénieurs ou administrateurs réseaux. Par exemple, dans le cas de l'outil de mesure du débit disponible sur un lien, la précision est intimement liée à la vitesse de convergence du processus de mesure, et au pas d'augmentation du trafic sonde lors de chaque changement d'itération. Mais dans d'autres cas, comme par exemple la mesure de délais, la précision peut seulement être liée à la qualité de la synchronisation temporelle entre la source et la destination des paquets sondes.

2.1.3. Exemples d'outils

Les mesures actives simples restent tout de même monnaie courante dans l'Internet pour lequel de nombreux outils de test, validation et / ou mesure sont disponibles. Parmi eux, on peut citer les très célèbres *ping* et *traceroute*. *Ping* permet de vérifier qu'un chemin est valide entre deux stations et de mesurer certains paramètres comme le RTT ou le taux de perte. *Traceroute* permet de voir apparaître l'ensemble des routeurs traversés par les paquets émis jusqu'à leur destination et donne une indication sur les temps de passage en chacun de ces nœuds. On peut également citer MGEN qui présente la particularité d'émettre des paquets en multicast, c'est-à-dire à destination de plusieurs récepteurs à la fois, et permet donc, grâce à un mécanisme de duplication des paquets au dernier moment dans le réseau, de minimiser le trafic des paquets sondes.

Il existe également toute une batterie d'outils pour l'estimation de la capacité d'un lien ou d'un chemin, comme clink [DOW 99], pchar [MAH 00], pathchar [JAC 97], etc. De la même façon des outils, souvent assez similaires dans leurs processus estiment le débit disponible sur un chemin du réseau. On peut citer Abing [NAV 03], Spruce [STR 03], pipechar [GUO 05], pathchirp [RIB 03], IGI [HU 03], etc. Cet article, dont l'objectif est de faire un tour d'horizon des différentes techniques de mesure, n'a pas pour objectif de rentrer dans les détails des algorithmes de chacun de ces outils. Le lecteur pourra se référer aux publications originales qui décrivent ces outils. A noter toutefois, et c'est aussi une des raisons pour laquelle nous ne nous attardons pas sur la description de ces outils, qu'ils restent assez peu efficaces : ils sont très long à converger et de fait inadaptés pour un réseau dont le trafic varie beaucoup. Ils sont également imprécis ne serait-ce que pour estimer le débit moyen disponible [LAB 05].

Il existe également des outils matériels pour les mesures actives, comme les sondes RIPE TTM[3] par exemple. Ces sondes sont en fait des PC utilisant le système FreeBSD et équipées de cartes GPS reliées à une antenne satellite. Avec le logiciel de génération et d'analyse des paquets sondes déployés sur ces sondes, il est possible de connaître les délais, les taux de perte et la topologie du réseau entre les différentes sondes RIPE déployées dans l'Internet. L'élément le plus intéressant dans les équipements RIPE reste l'existence d'un service commercial (offert par RIPE) pour gérer ces sondes et leurs mesures en continu. Il existe aujourd'hui plus de 250 sondes de ce type dans le monde, ce qui en fait une des plus grandes plates-formes de mesures actives. D'autres plates-formes de ce type existent comme NIMI [PAX 00], initiée et maintenue par des universités, américaines à la base, de tous pays aujourd'hui. Toutefois, dans la partie 3.1, cet article décrira en détail une de ces plates-formes

[3] http://www.ripe.net/projects/ttm/index.html

que nous connaissons bien car étant impliqués dans ce projet. Il s'agit de la plate-forme MetroMI du projet METROPOLIS déployée sur le réseau Renater.

2.2.Mesures passives

2.2.1. Définition

Les projets de mesures passives sont apparus beaucoup plus récemment que les projets de mesures actives car ils nécessitent des systèmes de capture ou d'analyse du trafic en transit relativement avancés, et développés très récemment – même si quelques logiciels simples, mais aux capacités limitées, existaient déjà comme TSTAT, NTOP, LIBCAP, TCPdump, TCPtrace, etc. sur lesquels nous reviendrons dans la partie 2.2.3. Ils ont néanmoins mis en évidence que des outils de supervision, fonctionnant avec un approche passive étaient de nature à résoudre bon nombres de problèmes qui se posent pour la conception, l'ingénierie, l'opération et la gestion des réseaux de l'Internet, et c'est là un des objectifs de cet article de le démontrer.

Des équipements matériels plus performants sont en fait à la base de l'essor actuel en matière de métrologie passive, et ont notamment ouvert la voie à la métrologie passive microscopique (définie plus loin). Le principe des mesures passives est de regarder le trafic et d'étudier ses propriétés en un ou plusieurs points du réseau. L'avantage des mesures passives est qu'elles ne sont absolument pas intrusives et ne changent rien à l'état du réseau lorsqu'on utilise des solutions matérielles dédiées (par exemple sur la base de cartes DAG [DAG 01] présentées dans la partie 3.2.1). Elles permettent des analyses très avancées. En revanche, il est très difficile de déterminer le service qui pourra être offert à un client en fonction des informations obtenues en métrologie passive. Il vaut mieux pour cela utiliser des techniques actives.

Les systèmes de métrologie passive, peuvent également se différencier en fonction du mode d'analyse des traces. Ainsi, le système peut faire une analyse en-ligne ou hors-ligne. Dans le cadre d'une analyse en-ligne, toute l'analyse doit être effectuée dans le laps de temps correspondant au passage du paquet dans la sonde de mesure. Une telle approche, temps-réel, permet de faire des analyses sur de très longues périodes et donc d'avoir des statistiques significatives. Par contre, la complexité maximale pour ces analyses reste très limitée à cause du faible temps de calcul autorisé (d'autant plus faible que la capacité du réseau est importante). A l'inverse, une analyse hors-ligne oblige la sonde à sauvegarder une trace du trafic pour analyse ultérieure. Une telle approche demande ainsi d'énormes ressources ce qui représente une limitation pour des traces de très longue durée. Par contre, une analyse hors-ligne permet des analyses extrêmement complètes et difficiles, capables d'étudier des propriétés non triviales du trafic. De plus, comme les traces sont sauvegardées, il est possible de faire plusieurs analyses différentes sur les traces, et de corréler les différents résultats obtenus sur la trace, ou obtenus sur des traces différentes, pour une meilleure compréhension des mécanismes complexes du réseau.

2.2.2. Problématique associée aux mesures passives

La principale contrainte qui se pose pour l'installation de sondes de mesure est due au fait que le réseau dont nous souhaitons analyser le trafic est presque toujours un réseau opérationnel (à l'exception de quelques réseaux d'expérimentation dans les laboratoires de recherche), et que malgré la présence de la sonde, ce réseau doit continuer à fonctionner sans aucune dégradation du service qu'il offre.

- Le premier besoin pour le système de mesure à mettre en place est donc une transparence totale pour le réseau et son trafic. Cela signifie que pour être non intrusif,

cet équipement ne devra pas provoquer de pannes, d'erreurs de transmission et ne pas introduire de délais pour ne pas modifier le profil du trafic et les performances du réseau.

- Le second besoin lors du choix des sondes de mesure passive concerne sa précision et la validité des traces qu'elle produira. Ainsi, il est essentiel de ne pas « manquer » de paquets transitant sur le réseau, et d'avoir des informations précises sur le passage de ces paquets, notamment au niveau temporel, qui représente aujourd'hui une des difficultés majeures avec les systèmes actuels. Le système devra donc être bien dimensionné et offrir une horloge précise qui ne dérive pas.
- Le troisième besoin qui apparaît concerne la possibilité de corréler des événements de plusieurs traces, par exemple de suivre un paquet en plusieurs points du réseau, ou d'analyser de façon croisée le passage des paquets et de leurs acquittements, etc. Pour pouvoir finement analyser de tels événements se produisant en des points géographiquement distants et à des instants distincts mais faiblement éloignés temporellement, il est nécessaire de disposer d'une base temporelle commune et universelle pour toutes les sondes.

Enfin, il existe également d'autres besoins qui peuvent apparaître de seconde importance lors de la conception d'un outil de mesure mais qu'il ne faut pas négliger. Le premier concerne le problème de la métrologie d'un réseau complet et du rapatriement des données collectées au niveau des différents points de métrologie. Il est important de trouver un moyen de ramener ces données vers les machines d'analyse de façon efficace et sans influer sur le réseau et sa charge. Le second concerne la nature des informations à collecter, et notamment la taille des enregistrements faits sur chaque paquet. En effet, collecter la totalité des données du paquet, avec donc toutes les informations applicatives, est a priori contraire aux règles fixées par la CNIL sur le droit à la vie privée. Il faut donc bien réfléchir aux données à collecter sur chaque paquet par rapport à l'analyse que l'on souhaite faire.

2.2.3. Exemples d'outils

Parmi les outils de métrologie passive à notre disposition, il existe tout un ensemble d'outils logiciels, dont il serait illusoire de vouloir faire une liste exhaustive. Nous nous limiterons donc aux plus connus et utilisés.

2.2.3.1. Solutions logicielles

La plus célèbre de ces familles d'outils logiciels est celle composée de TCPdump[4], TCPtrace[5], Ethereal[6], etc. qui sont tous des outils basés sur l'utilisation de la librairie LIBPCAP. Cette librairie permet d'aller lire sur une interface réseau les paquets qui transitent, d'en récupérer une copie, et de la stocker et/ou analyser. Naturellement, ces outils possèdent les avantages et les inconvénients de leur nature logicielle. Les avantages font qu'ils sont faciles à utiliser, largement disponibles (les 3 outils cités sont d'ailleurs des logiciels libres), et ne nécessitent pas d'équipements matériels coûteux pour fonctionner. A l'opposé, ces outils souffrent de problèmes d'imprécisions, notamment temporelles, car le système opératoire effectue de nombreuses tâches – et peut même être interrompu – entre la capture du paquet et son estampillage par exemple. De même, des problèmes de performance limitent les capacités de traitement et d'analyse des ces outils logiciels, et parfois même de capture. Dès lors que le débit du trafic sur le lien monitoré augmente, ces outils ne donnent plus satisfaction ; ils ne sont guère efficaces avec des ordinateurs de bureau classiques dès que le débit excède une

[4] http://www.tcpdump.org/
[5] http://www.tcptrace.org/
[6] http://www.ethereal.com/

dizaine de Mbps. Pour augmenter ces capacités, il faut utiliser des machines dont les capacités de traitement et de communication externes et internes sont largement supérieures, mais sans que les problèmes de précision, ni même de possibilités d'analyse ne soient largement améliorés. En fait, pour monitorer du trafic sur des liens à très hauts débits, nous recommandons de n'utiliser ces outils logiciels que pour l'analyse du trafic en temps différé ; en aucun cas, ils ne donnent satisfaction pour des analyses en temps réel. L'approche consiste donc à utiliser des outils matériels – plus avancés et plus récents – de capture du trafic sur des liens à hauts débits, de stocker la trace de ce trafic sur un disque dur, puis de faire l'analyse de cette trace en temps différé avec les outils logiciels cités. Naturellement, cela interdit un certain nombre d'applications au réseau qui nécessitent des résultats d'analyse en temps réel. Mais cela permet néanmoins de travailler à la caractérisation et modélisation du trafic, à l'analyse de phénomènes particuliers sur les protocoles, l'analyse des attaques, etc. Il est à noter cependant que de part le faible investissement que nécessite la prise en main de ces outils, couplé au fait qu'il n'est pas nécessaire de faire des investissements matériels, ils sont massivement utilisés dans la communauté réseau, et seuls les spécialistes en métrologie investissent dans des solutions matérielles plus performantes et plus coûteuses. A noter également dans la même famille, des outils comme TSTAT[7] ou NTOP[8] qui permettent respectivement d'analyser des connexions TCP et de monitorer le nombre de connexions sur une machine. Ils présentent évidemment les mêmes avantages et défauts que TCPdump ou Ethereal. TSTAT est utilisé en général par les chercheurs et ingénieurs qui cherchent les limitations de TCP sur une infrastructure de communication particulière. NTOP peut par exemple être utilisé pour détecter certaines attaques de déni de service.

2.2.3.2. Traffic Designer

S'inspirant de ces outils logiciels, la société QoSMOS (une jeune pousse issue du LIP6 à Paris) a conçu, développé et commercialise un outil de supervision et de mesure du trafic – Traffic Designer[9] – mais qui en plus exploite ces résultats pour adapter sa gestion du trafic aux besoins des utilisateurs en termes de QoS. L'objectif est donc d'exploiter au mieux les ressources du réseau pour maximiser la satisfaction des utilisateurs. L'outil Traffic Designer est complètement « packagé ». Il arrive chez l'utilisateur sous la forme d'un PC spécifiquement adapté à cette tâche, et avec tous les logiciels d'analyse et de gestion du trafic. Notre expérience avec cette solution a montré que les problèmes de précisions et de performances sont rémanents. Avec les trafics actuels de l'Internet, Traffic Designer n'est guère capable aujourd'hui de traiter des débits supérieurs à une dizaine de Mbps. Toutefois cet outil reste très intéressant par la quantité des logiciels d'analyse fournis et par la qualité de son interface graphique, et ce autant pour la facilité de définition des analyses particulières que pour la présentation des résultats. En particulier, un des points forts de la solution QoSMOS est son logiciel de classification applicative basé sur une reconnaissance des protocoles applicatifs, et non pas sur l'utilisation des numéros de port IANA comme cela est fait dans tous les autres outils d'analyse. En effet, aujourd'hui, de plus en plus d'applications utilisent des numéros de port dynamiques (comme le P2P par exemple) ou encapsulent leur connexions dans des connexions d'autres protocoles applicatifs (les applications de video streaming par exemple encapsulent souvent leurs connexions dans des connexions web pour pouvoir passer les pare-feux ou firewalls). Les numéros de port fixés par l'IANA pour les différentes applications n'ont donc plus grande signification. Le mode de reconnaissance du protocole applicatif par « pattern matching » sur les premiers paquets de la connexion applicative est donc d'une précision remarquable qui limite l'introduction de biais dans l'analyse. De plus, la

[7] http://tstat.tlc.polito.it/
[8] http://www.ntop.org/
[9] http://www.qosmos.net/Produits_et_services/Traffic_Designer/

société QoSMOS, pour passer outre les problèmes de performance propose pour les spécialistes de la métrologie réseau un logiciel d'injection de traces de trafic (capturées par ailleurs) dans l'outil Traffic Designer : l'outil TDplayer. Comme pour les outils précédemment cité, cela interdit des analyses en temps réel, mais cela permet d'un autre côté d'utiliser les capacités des logiciels d'analyse Traffic Designer, qui restent intéressantes même en temps différé.

2.2.3.3. IPANEMA

En allant encore plus loin dans les solutions de métrologie dédiées, on peut citer IPANEMA[10] qui est une sonde qui combine mesures actives et passives. Ces sondes sont constituées autour d'un hardware dédié, et possédant un logiciel pour la génération de paquets sondes et un logiciel d'analyse, l'analyse portant à la fois sur les paquets sondes et sur le trafic normal du lien monitoré. Il faut donc pour que la solution fonctionne disposer de plusieurs sondes dans le réseau pour pouvoir exploiter à la fois les mesures actives et passives. Ainsi, les sondes qui émettent le trafic et celles qui le voient passer peuvent s'échanger les mesures qu'elles ont réalisées pour une analyse complète et précise. Toutefois, l'inconvénient majeur de ces sondes IPANEMA est leur coût quasiment prohibitif. De plus, pour que la solution puisse être utile, il faut en déployer un assez grand nombre dans le réseau, ce qui amène à des tarifs indécents. C'est d'ailleurs certainement la raison pour laquelle les sondes IPANEMA restent marginales aujourd'hui, et que les spécialistes de la métrologie leur ont préféré des solutions tout aussi performantes mais moins onéreuses comme les systèmes DAG sur lesquels nous reviendront au paragraphe 3.2.

2.2.3.4. Netflow

Toutefois, en pensant à la mise en œuvre opérationnelle des solutions de métrologie dans un réseau, il apparaît que l'endroit idéal pour positionner des sondes de mesures passives est indéniablement dans les routeurs. CISCO a ainsi développé le module *Netflow* [CIS 01] dans ses routeurs, qui scrute le trafic en transit, et génère régulièrement des informations statistiques sur ce trafic. *Netflow* a ainsi été utilisé dans de nombreux projets de recherche, mais tend aussi à remplacer les solutions de gestion de réseaux chez les ingénieurs réseaux d'exploitation. Netflow est à la base un mécanisme pour améliorer les performances des routeurs CISCO avec une gestion interne du routage par flux qui ne nécessitait de ne router que le premier paquet de chaque flux, les paquets suivants étant ensuite commutés vers le même port que le premier paquet du flux. Ainsi, même si la commutation / routage Netflow a disparu des nouvelles générations de routeurs CISCO, le composant logiciels de mesures statistiques est resté. L'expérience montre toutefois que les performances de Netflow restent limitées (code écrit en Java et interprété), et que l'influence sur les performances du routeur est non négligeable. Mais il s'agit certainement là d'une voie à explorer car ce sera certainement la plus viable et la plus efficace dès lors que des solutions de métrologie devront être mises en exploitation sur un réseau opérationnel. Elles prendront en fait la place et étendront les solutions actuelles de gestion du réseau basées sur le protocole SNMP et ses MIB (qui contiennent les informations mesurées aux différents points monitorés). A noter d'ailleurs qu'il existe des travaux au niveau du groupe MRTG[11] de l'IETF qui essaient de normaliser la représentation de ces résultats de monitoring. Pour l'instant ils portent sur SNMP et ses MIB, mais il faudra qu'ils puissent prendre en compte d'autres solutions de métrologie et d'autres formats qui vont apparaître prochainement.

[10] http://www.ipanematech.com/New/FR/index.php
[11] http://people.ee.ethz.ch/~oetiker/webtools/mrtg/

2.2.3.5. OCxMON

Enfin, des cartes dédiées à la capture du trafic sont apparues. Les premières dont l'intérêt a été significatif sont les cartes OCxMON [APS 97] mais qui ne fonctionnent que pour des réseaux ATM (OC3MON étant pour les liens à 155 Mbps et OC12MON pour les liens à 622 Mbps). OCxMON est donc une carte de capture des entêtes de chaque paquet qui transitent sur un lien. Ces paquets sont ensuite stockés dans un fichier sur un disque dur. Il est ainsi possible de passer tous les logiciels d'analyse que l'on souhaite sur cette trace. A noter toutefois, qu'il est possible aussi de récupérer en direct la sortie de la carte pour la rediriger vers les logiciels d'analyse, si ceux-ci sont assez rapides pour fonctionner à la vitesse du lien.

2.2.3.6. DAG

Finalement, c'est cette approche avec cartes de capture du trafic qui a aujourd'hui le plus de succès auprès des spécialistes de la métrologie. La solution OCxMON n'étant valide que pour des liens ATM assez rares aujourd'hui, c'est vers la solution DAG que se sont le plus souvent tournés les chercheurs et ingénieurs en métrologie réseau, la carte DAG existant pour de très nombreuses technologies de réseaux et couvrant (presque ?) toutes les capacités de liens. Cette solution sera décrite précisément dans la section 3.2.

3. Exemple de la plate-forme de mesure et métrologie sur Renater

Une fois effectué ce bref tour d'horizon de la problématique associée aux mesures dans les réseaux de communication – et notamment dans l'Internet – et des différentes techniques de mesures existantes, nous allons maintenant donner plus de détails sur un exemple que nous connaissons particulièrement bien pour en être des contributeurs : l'exemple de la plate-forme de métrologie déployée autour du réseau Renater dans le cadre du projet METROPOLIS. Nous allons donc donner tous les détails de conception, mise en œuvre et déploiement des outils de mesures actives (partie 3.1) et de mesures passives (partie 3.2).

3.1. Plate-forme de mesure active

Comme nous l'avons vu dans la partie 2.2.3, il existe aujourd'hui deux plates-formes de métrologie active publiques, mondiales et de tailles significatives : les plates-formes RIPE-TTP et NIMI. Naturellement, et c'est d'autant plus vrai pour les mesures actives, plus une plate-forme de mesure comporte de sondes plus elle est à même de fournir des résultats intéressants à l'échelle d'un réseau ou d'une interconnexion de réseaux. Aussi, notre choix en termes d'équipements de mesures actives pour notre plate-forme MetroMI (METROPOLIS Measurement Infrastructure) a été conduit par la volonté d'être compatible avec le petit millier de machines existantes et composant ces deux plates-formes. Ainsi, les sondes de mesures actives sont des sondes RIPE-TTM, i.e. des PC standards, avec le système FreeBSD et une carte GPS connectée à des antennes GPS. Grâce à ce GPS, l'ensemble de la plate MetroMI dispose d'une référence temporelle universelle et toutes les sondes sont donc synchronisées. De plus, le GPS fonctionne par l'émission toutes les secondes d'un signal de synchronisation sur lequel se calent les horloges des différentes machines. Cela assure aussi une grande précision des horloges dont la dérive est annulée toutes les secondes – et la dérive des horloges actuelles sur une seconde est quasiment inexistante (dans tous les cas non mesurable avec les outils classiques – non atomiques – dont nous disposons). D'autre part, nos sondes RIPE-TTM hébergent une version du logiciel NIMI portée par les chercheurs du projet METROPOLIS pour le système FreeBSD. Nos sondes MetroMI deviennent donc compatibles avec les sondes NIMI (par nature elles étaient déjà compatibles avec les machines de la plate-forme RIPE-TTM).

En ce qui concerne la non intrusivité des mécanismes de mesure qui injectent des paquets sondes et leur vitesse de convergence, c'est naturellement le soucis des concepteurs et développeurs d'outils de mesure active, et les solutions vont dépendre des paramètres qu'il faut mesurer et de la manière employée pour les mesurer. Ce n'est en tout cas pas un problème de nature à impacter la conception matérielle des machines de la plate-forme de mesure active.

Enfin, il faut également noter que la plate-forme MetroMI intègre ses propres logiciels de mesure, développés dans le cadre de METROPOLIS. D'autre part, les machines de la plate-forme MetroMI peuvent être isolées des plates-formes RIPE-TTM et NIMI dans le cas où nous souhaiterions conduire des expérimentations pouvant être incertaines pour l'Internet et la vivacité des machines des plates-formes RIPE-TTM et NIMI.

Le déploiement des sondes MetroMI est représenté sur la figure 1.

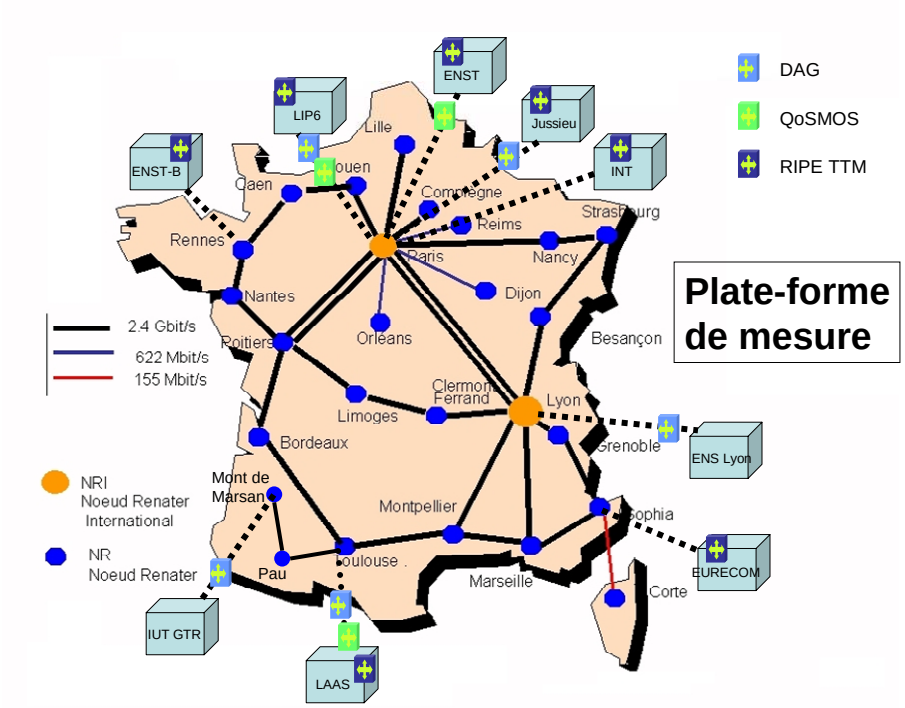

Figure 1. La plate-forme de mesure (active et passive) déployée dans le cadre de METROPOLIS

3.2. Plate-forme de mesure passive

Par rapport à la plate-forme passive, comme nous l'avons vu dans la partie 2.2.3, il n'existe pas de solution satisfaisante pour des analyses avancées qui fonctionne en temps réel. De même, nous avons vu qu'elles étaient pour la plupart inadaptées aux réseaux à très hauts débits car pas capables de capturer les paquets qui arrivent à la vitesse du lien. De fait, nous avons du nous orienter vers des solutions à partir de cartes de capture dédiées, et notamment les cartes DAG. Ce choix présente également l'avantage, étant donné l'état du marché actuel autour de ce genre d'équipements, d'être presque un standard de fait. Tous les logiciels que nous développons pourront donc être utilisés par la presque totalité de la communauté métrologie, et nous pourront de même bénéficier des développements faits par d'autres.

3.2.1. La solution DAG

Pour répondre aux besoins (transparence, précision temporelle, temps universel) énoncés dans la partie 2.2.2, la solution existante la mieux adaptée est indéniablement une

solution basée sur les cartes DAG conçues et développées à l'université de Waikato en Nouvelle-Zélande et, à l'heure actuelle, commercialisées, maintenues et améliorées par la société ENDACE. Le principe de fonctionnement des sondes DAG est décrit sur la figure 2. Le premier avantage de cette carte est de pouvoir travailler en dérivation du lien à analyser. Ainsi, dans le cadre de réseaux sur fibres optiques, le principe de branchement de la sonde consiste à insérer un « splitter » optique qui laisse passer 80 % de la puissance optique sur la fibre originelle (chemin normal), et récupère 20 % de cette puissance à destination de la sonde DAG. Ainsi, le trafic n'est absolument pas perturbé, aucun délai n'est introduit au niveau du « splitter »[12] et le trafic conserve donc les mêmes caractéristiques et profils. Le système de mesure est ainsi totalement transparent.

De son côté, la carte DAG est une carte dédiée qui réalise, en temps-réel, l'extraction des entêtes de tous les paquets qui passent sur le lien. La taille de l'entête est précisée au moment de la configuration de la carte pour la capture. Dans notre cas, nous souhaitons pouvoir capturer les entêtes IP et TCP (ce qui nous éloigne a priori de certains problèmes juridiques qui peuvent se poser lorsque l'on capture un échantillon plus grand de chaque paquet). Enfin, pour chaque paquet capturé, la carte ajoute une estampille codée sur 64 bits à l'entête capturée. Le tout est ensuite stocké sur disque. Il est à noter que les traces ainsi constituées deviennent rapidement très volumineuses, surtout sur les réseaux à hauts débits, et nécessitent donc d'utiliser des disques de grandes capacités et en nombres suffisants. Toutes les machines sont donc équipées de 3 disques durs de 73 Go.

Pour la même raison, le trafic qui transite entre la carte DAG et le disque dur de la station hôte est très élevé, et pour les réseaux aux capacités les plus fortes, les bus PCI classiques des ordinateurs habituels ne suffisent pas. Il est nécessaire dans ce cas d'utiliser la dernière génération de bus PCI, à savoir les bus 64 bits à 66 MHz qui offrent des bandes passantes bien supérieures aux bus PCI traditionnels de 32 bits cadencés à 33 MHz. Tous ces éléments (carte dédiée temps-réel, bus haute capacité, mémoire importante et disques durs de grandes capacités) sont les éléments indispensables pour garantir un système bien dimensionné capable de capturer une trace de tous les paquets ayant transité sur le lien mesuré.

En ce qui concerne l'estampille de passage de chaque paquet, stockée avec l'entête du paquet, une référence GPS est utilisée. La carte est en effet directement reliée à une antenne GPS. Ainsi, l'horloge de la station qui héberge la carte DAG est resynchronisée chaque seconde sur un signal GPS qui transporte le temps universel venant des horloges atomiques de référence. Ainsi, la dérive de l'horloge est quasiment inexistante, garantissant une grande précision des mesures temporelles, ainsi que le temps universel, car les sondes seront effectivement synchronisées sur le temps de référence universel.

[12] D'ailleurs, le « splitter » est un élément complètement passif basé sur des jeux de miroirs qui ne sont même pas alimentés électriquement, garantissant ainsi un fonctionnement normal même en cas de panne d'électricité.

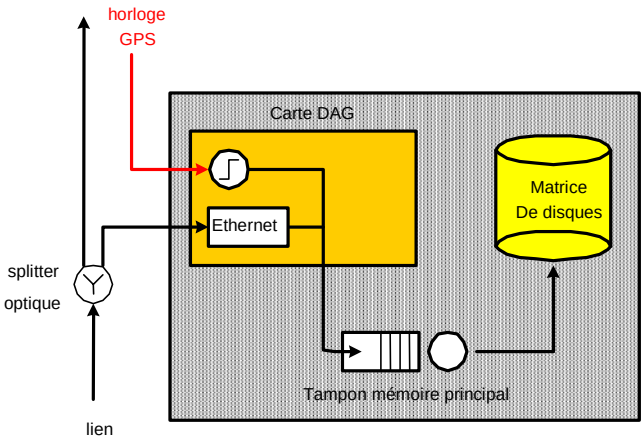

Figure 2. Principe opératoire des cartes DAG

Respectant ces principes de conception, la sonde conçue pour capturer le trafic sur un lien Giga-Ethernet est détaillée sur la figure 3.

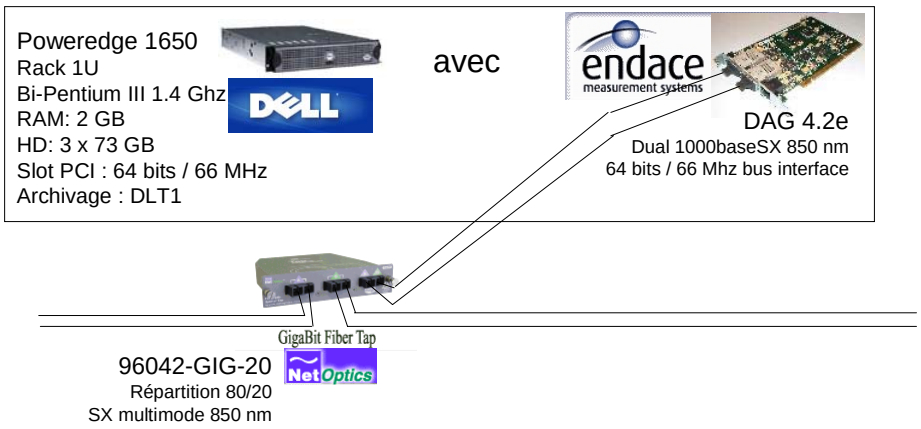

Figure 3. Conception de la sonde de métrologie passive pour un lien Giga-Ethernet (le modèle de la station de travail qui accueille la carte DAG est naturellement à la discrétion de l'utilisateur. Nous indiquons le modèle exact que nous utilisons pour situer le niveau de performance requis, mais n'importe quelle autre machine de puissance et de qualité de communication interne équivalentes ou supérieures sera satisfaisante)

De la même façon, la sonde pour capturer le trafic sur des réseaux Fast-Ethernet est détaillée sur la figure 4. A noter que dans ce cas, le support physique est de la paire torsadée. Il n'y a donc pas de « splitter », inutile avec les propriétés naturelles de propagation de l'électricité, et un système de « bypass » le remplace pour garantir un bon fonctionnement du réseau même si la sonde s'arrête.

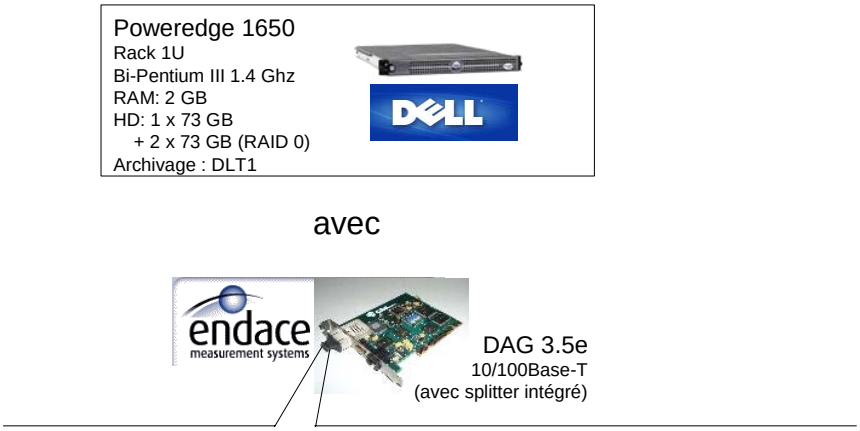

Figure 4. Conception de la sonde de métrologie passive pour un lien Fast-Ethernet (le modèle de la station de travail qui accueille la carte DAG est naturellement à la discrétion de l'utilisateur. Nous indiquons le modèle exact que nous utilisons pour situer le niveau de performance requis, mais n'importe quelle autre machine de puissance et de qualité de communication interne équivalentes ou supérieures sera satisfaisante)

Enfin, pour analyser les traces capturées par les sondes, une plate-forme de stockage des traces et d'analyse a été conçue et mise en place. Cette plate-forme est hébergée au LAAS à Toulouse et ouverte aux partenaires de METROPOLIS. Les besoins de cette plate-forme sont donc essentiellement une grande capacité de stockage, et une grande capacité de traitement (processeurs et mémoire essentiellement). Cette plate-forme est donc composée de 2 PowerEdge 4600 dont les caractéristiques sont :

Poweredge 4600
Rack 6U
Bi-Pentium Xeon 2.4 Ghz
RAM: 8 GB
HD: 8 x 73 GB
Archivage : DLT1

D'autre part, il a fallu penser à un système pour rapatrier les fichiers de traces des sondes de mesures vers la plate-forme d'analyse. La solution idéale aurait été de pouvoir utiliser les réseaux académiques, mais face à la charge supplémentaire qu'auraient représentée ces transferts, certains réseaux académiques auraient eu du mal à tenir. Il a donc été décidé d'équiper toutes ces machines avec des lecteurs de bandes au format DLT1, et nous faisons ces transferts en envoyant les bandes par les services postaux ou les transporteurs rapides.

Cette solution a aussi l'avantage de pouvoir utiliser les bandes comme système d'archivage des traces que nous n'utilisons plus pendant quelques temps, ce qui a permis de réduire la capacité des disques de la plate-forme d'analyse, et de réduire sensiblement son prix.

3.2.2. Carte de déploiement des sondes DAG

Nous avons demandé à plusieurs responsables des réseaux académiques français l'autorisation de déployer nos équipements de mesure sur le réseau qu'ils opèrent. Notre objectif était de pouvoir déployer ces sondes sur des réseaux de natures différentes, soit sur le réseau de cœur, sur les réseaux régionaux et à la sortie des laboratoires. Toutes les autorisations ne nous ont pas été accordées pour des raisons sur lesquelles nous reviendront plus loin. Toutefois, aujourd'hui, 5 sondes DAG ont été déployées. Ainsi, comme le montre la figure 1, deux sondes en Fast-Ethernet ont été positionnées à la sortie de deux grands laboratoires que sont le LAAS (figure 5) et le LIP6 (figure 6). La troisième sonde en Giga-Ethernet a, quant à elle, été positionnée à la sortie du réseau de Jussieu sur RAP (figure 7), nous permettant ainsi d'avoir accès aux traces du trafic d'un réseau à très haut débit. Une autre sonde Giga-Ethernet a plus récemment été installée à la sortie du réseau de l'ENS Lyon, et une autre en Fast-Ethernet à la sortie du lien d'accès de l'IUT GTR de Mont-de-Marsan à Renater (via l'université de Pau).

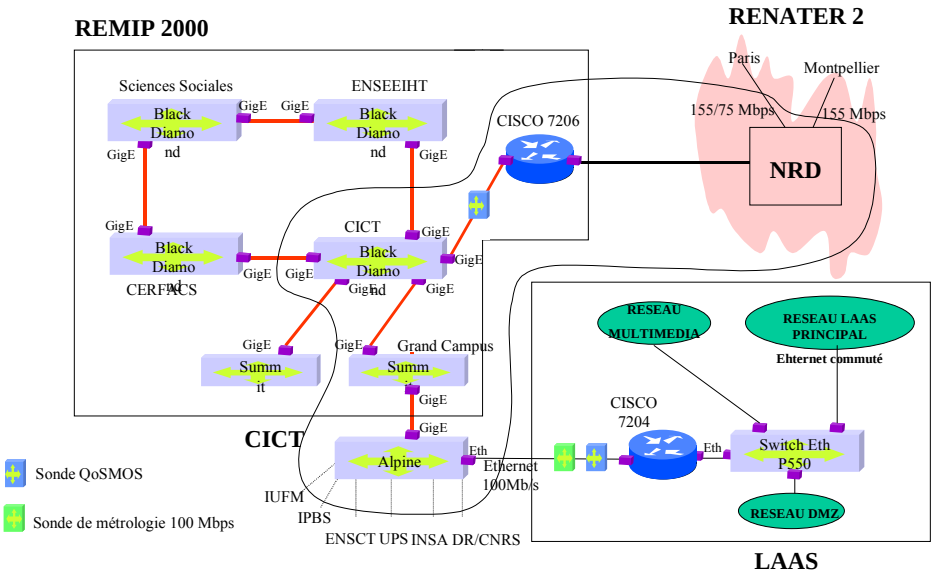

Figure 5. Schéma de déploiement sur la plate-forme toulousaine

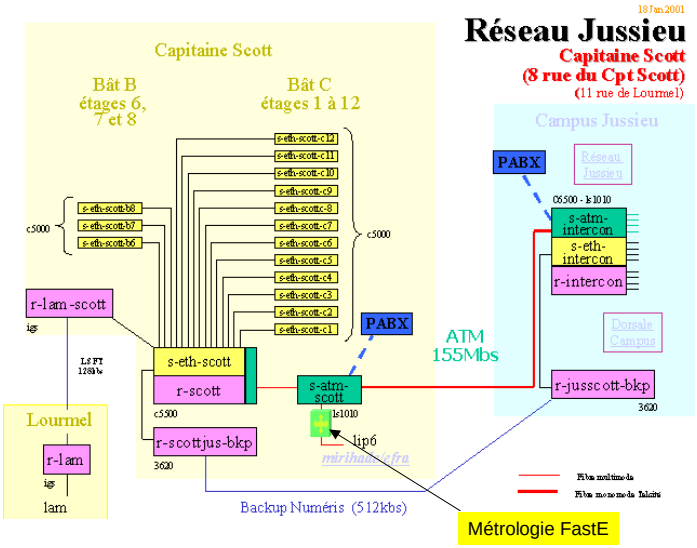

Figure 6. Schéma de déploiement au LIP6

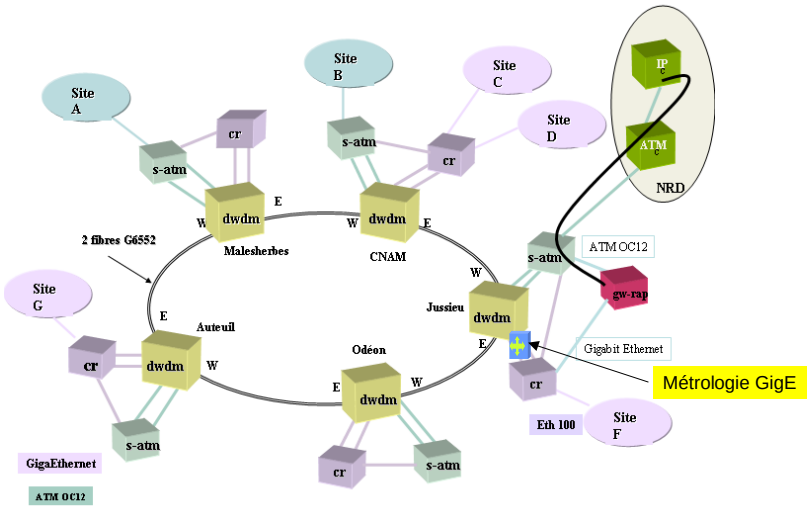

Figure 7. Schéma de déploiement à Jussieu

3 sondes QoSMOS Traffic Designer ont également été installées au LAAS, LIP6 et à l'ENST. Le choix de positionnement des sondes de métrologie passives microscopiques (système DAG) est aussi stratégique par rapport au positionnement des sondes de métrologie passive macroscopique et de sondes de métrologie active. Ainsi, il sera possible, pour un même trafic de corréler les analyses micro- et macroscopiques, ainsi que les mesures actives

et passives, et ce en plusieurs points du réseau et ce sur leur chemin entre Toulouse et Paris. Enfin, nous rappelons que la plate-forme d'analyse est hébergée par le LAAS à Toulouse.

Au delà de ces problèmes techniques, une nouvelle contrainte venant des administrateurs des réseaux qui hébergeaient nos sondes est apparue. Au delà de la non intrusivité de nos sondes qui a rapidement été avérée et donc garantissant que les performances des réseaux dont le trafic était analysé ne seraient pas perturbées, les administrateurs ont demandé à ce que les traces soient anonymes. D'ailleurs, les refus que nous avons essuyés parfois lors de nos demandes d'installation de sondes sur certains réseaux ont été motivés par l'aspect confidentiel des communications. Rendre les traces anonymes signifie que les adresses des sources et destinations des paquets dont nous capturons les entêtes soient anonymisées, de façon à ne pas pouvoir savoir quels en étaient les auteurs et destinataires. De plus, l'anonymisation nécessite la suppression des données utilisateurs contenues dans la partie applicative de chaque paquet de données prélevés sur le réseau. Naturellement, ces informations portent un certain nombre d'informations qui peuvent être utiles pour découvrir les topologies des réseaux, les usages que font certains du réseau, les protocoles de routage, etc. De tels aspects ne pourront donc pas être traités. Toutefois, pour pouvoir tout de même ne pas trop limiter les analyses et études que l'on peut faire à partir des traces capturées, il nous a fallu mettre en place un mécanisme d'anonymisation des adresses déterministe (pour pouvoir continuer à reconnaître les flux de l'Internet par exemple) et surtout qui ne casse pas la structure des adresses, dont les différents octets ne sont pas attribués de façon aléatoire. Un algorithme décrit dans [XU 01] [XU 02] a ainsi été mis en place non sans difficulté, en particulier à cause du nouveau format de stockage des traces DAG, qui est dans cette nouvelle version des cartes DAG un format dynamique. Il faut toutefois noter que l'anonymisation ne satisfait pas tous les administrateurs de réseaux académiques car malgré tout on connaît la provenance des traces, et les résultats obtenus pourraient mettre en évidence les comportements d'une communauté assez réduite et identifiable.

4. **Analyse du trafic Internet**

Nous disposons maintenant d'une plate-forme de métrologie qui nous permet de réaliser des mesures de paramètres simples comme le débit, les délais, les taux de perte, etc. et de capturer des traces de trafic dont chaque échantillon aura une taille plus ou moins importante selon le type d'information recherchée. Toutefois, comme cela a été présenté dans l'introduction, tout ceci ne donne qu'une vision superficielle du réseau, de son comportement et des services qu'il offre. En aucun cas, cela ne donne des informations sur le fonctionnement des différents composants du réseau comme les routeurs, les protocoles, etc. Or pour les ingénieurs, administrateurs et chercheurs en réseau, c'est bien la mécanique du réseau et de ses composants qui est importante et qu'il est nécessaire de mettre en évidence. Les mesures et captures de trafic effectuées ne sont que le résultat des effets des mécanismes et composants du réseau. En fait, le problème le plus important et le plus complexe à résoudre en métrologie des réseaux de communication consiste à déterminer à partir de l'observation des effets externes les causes internes (les mécanismes protocolaires) qui en sont responsables.

C'est donc au travers de la caractérisation et l'analyse des résultats de mesure et des traces de trafic qu'une telle tâche pourra être réalisée. Nous insistons d'ailleurs sur le fait que c'est là aujourd'hui le principal point dur de la métrologie et de l'étude du trafic. Certes, tous les problèmes au niveau de la mesure physique des paramètres simples n'ont pas encore été résolus, mais c'est au travers de l'analyse que la métrologie pourra vraiment atteindre les objectifs que nous avons fixés pour elle. Cette partie de l'article va donc se focaliser sur ces problèmes de caractérisation et d'analyse du trafic, même si dans ce domaine les chercheurs

n'en sont qu'au tout début de leurs travaux et tâtonnent encore. Nous allons toutefois présenter les méthodes d'analyse qui ont aujourd'hui donné le plus de résultats significatifs et qui ont permis de progresser dans la connaissance et la compréhension de l'Internet et de ses mécanismes protocolaires et architecturaux.

En particulier, un des éléments clés, et qui reste aujourd'hui un des seuls éléments de conclusion quant à l'analyse du trafic, c'est l'aspect multi-échelles qui semble déterminant – mais aussi des plus complexes à mettre en œuvre – pour l'analyse du trafic. Comme cela a été présenté dans la partie 1.1, il existe différents niveaux d'études du trafic, des bits ou octets jusqu'aux sessions, en passant par les flux. Or chaque niveau correspond à des granularités différentes et le traitement de ces différents éléments du réseau a donc des impacts sur des échelles temporelles différentes. De plus, les flux (ou les sessions) ont des tailles variables, et ont donc des conséquences sur des échelles de temps variables. C'est là la principale difficulté pour l'analyse du trafic qui doit être multi-échelles. C'est là aussi une des caractéristiques qui différencie la métrologie des réseaux de communication des mesures physiques en général. En physique, une granularité d'étude et un point de vue, cohérents par rapport à l'échelle du mécanisme que l'on veut observer, est choisie une bonne fois pour toute, et l'analyse est faite avec cette granularité. En métrologie des réseaux, toutes les échelles de temps sont indissociables, et toutes les granularités d'analyse doivent être considérées en même temps. Cela rajoute naturellement une dimension à la complexité des analyses, et nécessite de fait des outils mathématiques complexes que nous allons maintenant introduire.

4.1. Trafic Internet et notions associées

Les premières études métrologiques sur le trafic Internet menées partout dans le monde ont globalement montré que ce dernier est particulièrement instable, à cause des propriétés d'auto-similarité et de dépendance à long terme appelée aussi (LRD) [LEL 93]. Il est aussi montré que la distribution à queue lourde est très impliquée dans ces propriétés [WIL 98]. Avant de détailler dans la suite de ce chapitre toutes ces caractéristiques du trafic Internet et d'en analyser les causes, il est nécessaire d'introduire les notions mathématiques associées aux différents comportements observés dans le réseau.

4.1.1. Fonction d'auto-corrélation

Avant de présenter cette fonction mathématique, il faut définir les notions d'indépendance et de corrélation :

- X, Y sont deux v.a. indépendantes ssi

$$P(X<x \cap Y<y) = P(X<x).P(Y<y)$$

- X, Y sont deux v.a. décorrélées ssi

$$E(XY)=E(X).E(Y)$$

En pratique, on dispose de N mesures. La fonction d'auto-covariance se calcule comme la fonction de covariance entre deux séries. La seconde série est ici la même que la première mais décalée d'un nombre K d'éléments. La fonction d'auto-covariance C_K s'écrit alors :

$$C_K = \sum_{k=0}^{K} \left(\frac{1}{N-k} \sum_{t=0}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \right)$$

où $\bar{x}$ représente la moyenne de la série de points.

- Interprétation des résultats

Une caractéristique des lois et des distributions que l'on peut calculer à partir des traces réseaux est d'avoir une fonction d'auto-corrélation spécifique. En effet, elle traduit une corrélation persistante dans le temps[13] ainsi que la présence de dépendance à long terme entre les objets analysés (les paquets TCP la plupart du temps). Ainsi, il est nécessaire de pouvoir calculer de façon systématique la fonction d'auto-covariance d'une série représentant ses caractéristiques d'auto-corrélation. Pour ce qui est de l'interprétation de l'auto-corrélation d'une série, on considère que la covariance est nulle lorsque la corrélation empirique (donné par le graphique de $\frac{Auto\,\mathrm{cov}(K)}{Auto\,\mathrm{cov}(0)}$) est contenue entre $\frac{2}{\sqrt{n}}$ et $-\frac{2}{\sqrt{n}}$ (n étant le nombre de points sur lesquels on calcule la fonction d'auto-corrélation). Il s'agit de l'application du théorème de la limite centrale avec des hypothèses gaussiennes. Cela correspond à un intervalle de confiance de 95 % (cf. [DAC 94]). La courbe représentant l'auto-corrélation s'analyse en ayant au préalable calculé l'intervalle de confiance dans lequel la fonction doit se situer. Si la courbe dépasse cet intervalle, il existe de la corrélation. C'est une première information pour mettre en évidence dans un deuxième temps la présence de LRD dans la série analysée. En effet, on parle de LRD dans la série quand la corrélation reste présente pour une valeur de K grande.

4.1.2. Processus à dépendance longue (LRD)

Un processus à dépendance longue ou à mémoire longue signifie que la dépendance entre deux variables du processus ne diminue pas trop rapidement avec l'éloignement temporel. La définition mathématique introduite par [COX 84] est présentée ci-dessous :

Soit $X=X_t$ un processus stochastique (à covariance) stationnaire à temps discret, on dit que X_t est à mémoire longue s'il satisfait les propriétés suivantes :

- $\sum_{t=0}^{\infty} \rho(t) = \infty$ (ρ est la fonction d'auto-corrélation),
- La densité spectrale S est singulière[14] à l'origine,
- $m.\mathrm{var}X^m \rightarrow \infty$ quand $m \rightarrow \infty$

Un processus à dépendance longue possède la propriété suivante :

$\rho(t) \xrightarrow{t\to\infty} ct^{-\beta}$ $0<\beta<1$ (ρ est la fonction d'auto-correlation).

Ainsi la fonction d'auto-corrélation décroît hyperboliquement.

La dépendance à long terme a été découverte en premier par Hurst qui la définit comme un processus ayant une fonction d'auto-corrélation non sommable (première propriété) et

[13] La fonction d'auto-corrélation ne se situe pas dans l'intervalle de confiance pour K grand (supérieur à 5000 : cette borne a été définie de manière empirique grâce aux calculs qui sont détaillés dans [LAR 02]).

[14] Une fonction *f* est dite singulière en un point *a* si elle n'est pas explicitement définie en ce point (à cause par exemple d'une division par zéro si *x = a* ou dans le cas d'une fonction définie sur un ensemble topologiquement ouvert, d'un point a qui est à la frontière de l'ensemble de définition de la fonction - C'est le cas de la fonction *ln(x)* lorsque *x=0*).

caractérisée par un paramètre *H*, défini par la formule $H = 1 - \frac{\beta}{2}$ [15], et appelé paramètre de Hurst. En 1993, Leland, Taqqu, Willinger et Wilson ont mis en évidence la LRD pour des séries temporelles de paquets Ethernet. Depuis, une multitude de travaux et d'articles ont traité de la dépendance à long terme du trafic Internet (voir [WIL 98], [VER 00b], [PAR 96] et [DOW 01]). Nous y reviendrons dans les sections suivantes.

4.1.3. Distribution à décroissance lente

Plusieurs travaux de recherches (voir [WIL 98], [VER 00b], [PAR 96] et [DOW 01]) ont démontré que la distribution à queue lourde pour certaines caractéristiques du trafic (distribution des tailles de fichiers, des durées de transfert…) est l'une des principales causes de LRD du trafic Internet.

Une distribution est à queue lourde si sa fonction de distribution a la propriété suivante :

$$P[X > x] \xrightarrow{t \to \infty} x^{-\alpha}, \alpha \in]0,2[$$

En d'autres termes, la forme asymptotique de la distribution à queue lourde suit une loi exponentielle avec α inférieur à 2.

On l'appelle « à queue lourde » car, comparée à la distribution exponentielle et la distribution normale, une variable aléatoire qui suit une distribution à queue lourde peut montrer pour des très grandes valeurs de *X* une probabilité *P[X]* supérieure à celle obtenue pour une distribution exponentielle équivalente. Cette variable a une variance infinie si $\alpha \in]0, 2[$ et une moyenne infinie si $\alpha \in]0, 1[$.

4.1.4. Processus auto-similaire

L'auto-similarité est une notion très importante dans la caractérisation du trafic Internet. En effet, la nature du trafic de données en général, celui d'Internet plus particulièrement, présente un aspect auto-similaire. Il s'agit de la manifestation du phénomène suivant : la structure des variations d'amplitude du signal analysé (par exemple le nombre d'octets transférés par unité de temps) se reproduit de manière similaire quelle que soit la finesse temporelle avec laquelle il est représenté. Ainsi, le comportement d'un trafic auto-similaire est à l'opposé de celui d'un trafic poissonnien, dont les variations d'amplitude sont filtrées au fur et à mesure que l'on augmente la taille de la fenêtre d'observation [PAX 95]. Il existe différentes définitions mathématiques de l'auto-similarité. La suivante concerne les processus à temps continu :

Un processus *X(t)* est dit auto-similaire de paramètre $H \in R$, si et seulement si pour tout $c > 0$, $c^H X(t)$ et *X(ct)* possèdent les mêmes distributions jointes à tous les ordres. Ainsi pour tout entier $n, t_1, \ldots, t_n, x_1, \ldots, x_n$:

$$P(X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n) = P(X(ct_1) \leq c^H x_1, \ldots, X(ct_n) \leq c^H x_n)$$

Cette définition signifie que si l'on modifie l'échelle sur laquelle on observe le processus par un facteur positif *c* et que l'on « zoome » le même processus par ce facteur élevé à la puissance *H*, alors l'allure des deux processus obtenus est la même. Par conséquent, il n'y a pas une stabilisation vers une moyenne comme dans le cas du processus de Poisson.

[15] β étant le coefficient de la fonction d'auto-corrélation [BER 94].

4.2. Analyse par décomposition en ondelettes du trafic

Ainsi, ces premiers résultats de caractérisation du trafic Internet confirment la nature de nombreux phénomènes se produisant avec des granularités différentes (e.g. l'auto-similarité), et nécessitent donc de mettre en œuvre une méthode d'analyse multi-échelles. La technique qui est aujourd'hui la plus utilisée pour une telle analyse multi-échelle nous vient du domaine du traitement du signal. Il s'agit d'analyse à base d'ondelettes que nous allons décrire et illustrer dans la suite.

Rappelons d'abord le principe des ondelettes (voir [MAL 99] pour une introduction complète). ψ_0 représente l'ondelette-mère. $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ représente sa forme dilatée et translatée et $d_X(j, k) = \langle \psi_{j,k}, X_0 \rangle$ les coefficients d'ondelette correspondants. L'ondelette-mère ψ_0 est aussi caractérisée par un entier $N \geq 1$, le nombre de moments évanescents qui joue un rôle clé dans l'analyse pratique et théorique de la longue mémoire. Pour tous les processus X stationnaires au second ordre, son spectre $f_X(\nu)$ peut être exprimé à l'aide de ses coefficients d'ondelette par l'équation :

$$E(d_X(j,k)^2) = \int f_X(\upsilon) 2^j \left| \Psi_0 2^j \upsilon \right|^2 d\upsilon$$

où Ψ_0 est la transformée de Fourier de ψ_0 et E l'espérance mathématique. Si X est un processus dépendant à long terme de paramètre d, cela implique que :

$$E(d_X(j,k)^2) \propto C 2^{j(2d+1)}, \text{ si } 2^j \rightarrow +\infty$$

De plus, il a été prouvé que $\{d_X(j, k), k \in Z\}$ forment une séquence dépendante à court terme si $N > d + 1/2$. Cela signifie qu'ils ne souffrent pas des difficultés statistiques dues aux propriétés de longue mémoire. En particulier, les moyennes temporelles $S_j = 1/n_j \sum_{k=1}^{n_j} \left| d_X(j,k) \right|^2$ peuvent être utilisées comme des estimateurs efficaces et robustes pour $E(d_X(j,k)^2)$. Ceci conduit à la procédure d'estimation suivante : une régression linéaire pondérée de $log_2 S_j$ par rapport à $log_2 2^j = j$, réalisée à la limite de la granularité d'étude la plus grande, fournit une estimation de $2d+1$, et par conséquent de d. Les représentations graphiques de $log_2 S_j$ en fonction de $log_2 2^j = j$ sont communément qualifiées de diagrammes logarithmiques (LD : logscale diagrams) d'estimation de la LRD, comme par exemple le diagramme de la figure 9. La possibilité de faire varier N apporte de la robustesse à ces procédures d'analyse et d'estimation. La définition complète ainsi que les performances de cette procédure d'estimation sont détaillées dans [ABR 00] [ABR 98] [VEI 99].

Pour ceux que les équations rebutent, et pour visualiser le principe de l'analyse du trafic par décomposition en ondelettes, nous l'illustrons sur la figure 8. Ainsi, la propriété d'auto-similarité du trafic signifie que le schéma oscillant du trafic Internet se produit à toutes les échelles de temps. Il est donc important pour analyser le trafic de disposer d'un outil capable d'analyser le comportement du trafic, et notamment ses variations, à toutes les échelles de temps, i.e. pour toutes les granularités. Pour cela, nous avons utilisé une analyse en ondelettes [ABR 98]. La fonction en ondelettes a été sélectionnée car elle représente bien les variations du trafic, et ce quelles que soient leurs durée dans le temps (la preuve est visuelle sur la figure 8). Le principe de cette analyse consiste à extraire du trafic toutes les ondelettes possibles. Pour cela, nous utilisons plusieurs fonctions en ondelettes chacune de fréquence différente afin d'obtenir les différentes granularités temporelles d'observation. Les fonctions avec les périodes les plus larges représentent les plus longues vagues, c'est à dire celles générées par les flux éléphants.

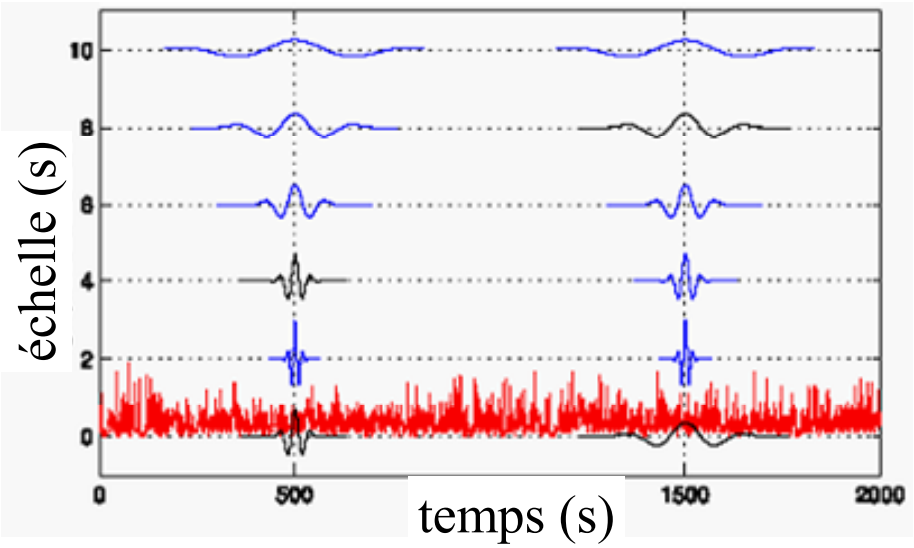

Figure 8 : Analyse en ondelettes de la LRD du trafic Internet

La courbe de la figure 9 a été obtenue en utilisant l'outil LDEstimate [ABR 98] qui estime la LRD qui se manifeste dans le trafic à toutes les échelles temporelles. Le résultat produit par cet outil est une représentation graphique en échelle logarithmique des lois qui régissent le niveau de dépendance du trafic à différentes échelles temporelles. Il est obtenu grâce à l'analyse en ondelettes du trafic Internet que nous venons de présenter et représente le niveau de variabilité des oscillations en fonction de la granularité d'observation. Le facteur de Hurst (caractéristique des processus auto-similaires qui se retrouvent dans le trafic Internet, cf. [PAR 00]) est obtenu directement sur la courbe de LRD en mesurant sa pente. La figure 9 montre un comportement différent pour deux échelles temporelles (appelé phénomène de « bi-scaling »). La frontière entre ces deux niveaux de LRD se trouve autour de l'octave 8 (ce qui correspond à 2^8 unités de temps, soit ici environ 250 ms) et met en évidence des niveaux de LRD différents pour les échelles de temps courtes et longues, ceci se traduisant par différentes lois de puissance. Pour les échelles petites (octave < 8), c'est à dire les paquets proches les uns des autres, la dépendance est peu marquée. Par contre, pour les échelles plus grandes octave > 8), c'est à dire des paquets plus éloignés, la dépendance est beaucoup plus importante. Il est important de noter ici que toutes les analyses de traces de trafic qui ont été faites de par le monde, et ce que les traces proviennent de n'importe quel type de réseau, indépendamment de la période de la journée, ont montré le même phénomène de bi-scaling. Il semble donc que l'on ait affaire à une signature significative et robuste du trafic qu'il faudra analyser plus en détail. Nous approfondissons maintenant ce phénomène et essayons d'en découvrir les causes avec des analyses complémentaires, et notamment une analyse fine de la répartition du trafic par application (répartition faite par l'outil QoSMOS Traffic Designer) et une étude de la taille des flux contenus dans le trafic (puisqu'on a vu qu'elle pouvait influer sur l'auto-similarité et la dépendance longue du trafic).

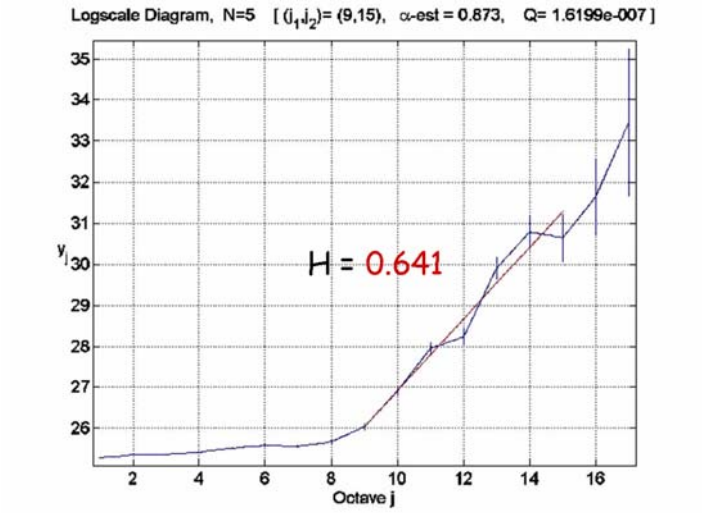

Figure 9 : Evaluation de la LRD dans le trafic Internet

4.3. Analyse des phénomènes de dépendance longue dans le trafic

4.3.1. Tendance d'évolution du trafic

Commençons cette partie par décrire l'évolution de la distribution du trafic par application mesurée dans l'Internet ces dernières années. La figure 10 illustre cette distribution mesurée en mai 2000 sur le réseau SPRINT. La grande proportion représentée par le trafic HTTP (plus de 75 % du trafic Internet) est remarquable. On note aussi que les principales applications standards sont représentées : web, web sécurisé, courriel, ftp ou news. Cependant, de nouvelles applications émergentes (à cette époque) sont présentes : flux de streaming multimédia (comme MediaPlayer ou RealAudio) ou jeux distribués en réseau (comme Quake). Néanmoins, la caractéristique la plus importante de ce trafic reste son élasticité et ses contraintes temporelles de QdS qui ne sont pas importantes (pour la grande majorité de ses applications).

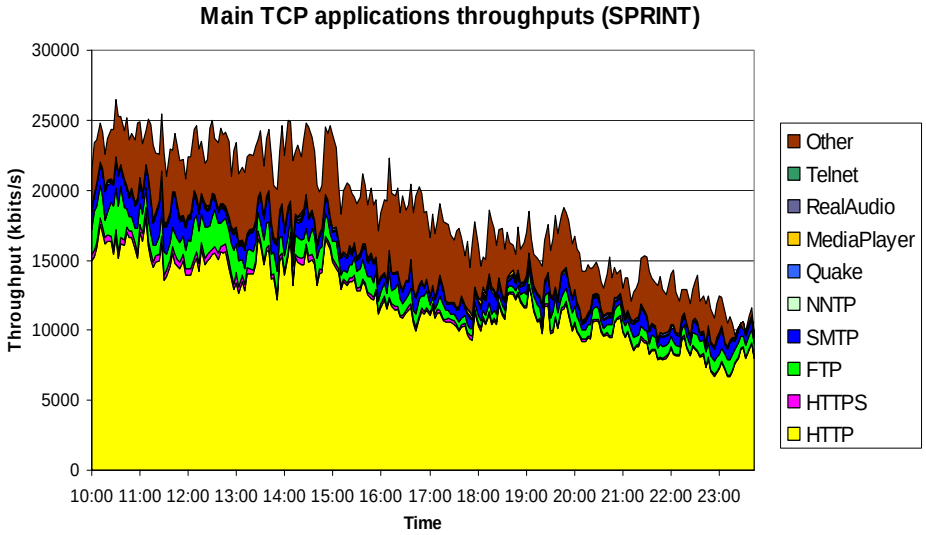

Figure 10. Répartition du trafic sur le réseau SPRINT (mai 2000) – les applications sont classées dans le même ordre sur la légende et dans le graphique}

Trois mois plus tard, la distribution était à peu près la même à l'exception d'une nouvelle application, la deuxième en partant du haut sur le graphique de la figure 11, qui est devenue, en quelques semaines, l'une des applications majeures de l'Internet. Il s'agissait de Napster, affublée du terme « application tueuse » car, en l'espace de trois mois, elle représenta entre 20 et 30 % du trafic. Ce type d'application P2P connut, au fil du temps, un succès de plus en plus important auprès des utilisateurs, représentant ainsi une part de plus en plus importante du trafic au sein de l'Internet. Bien que Napster eût quelques déboires avec la justice américaine, elle a ouvert la voie à tout un ensemble d'applications P2P comme Gnutella, E-donkey, Morpheus et d'autres.

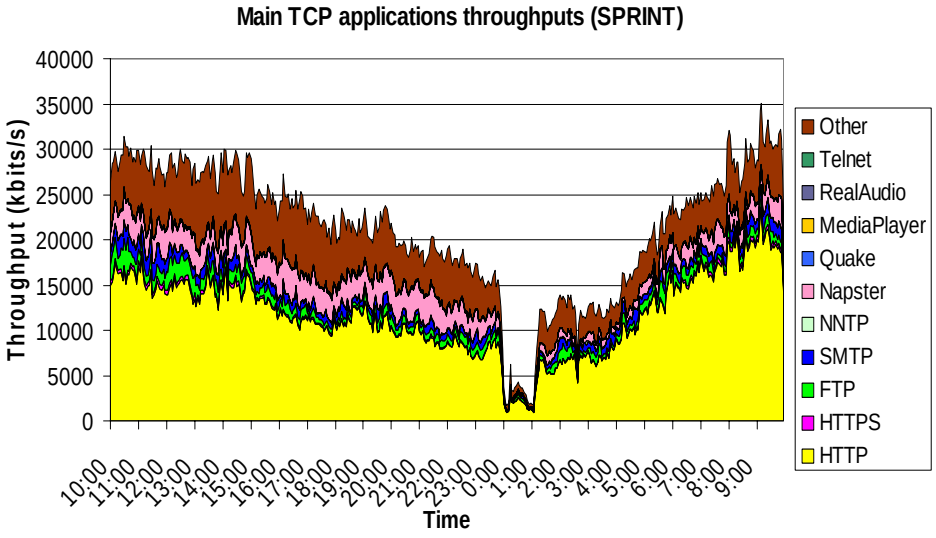

Figure 11. Répartition du trafic sur le réseau SPRINT (août 2000) – les applications sont classées dans le même ordre sur la légende et dans le graphique}

En effet, trois ans plus tard, le trafic P2P n'a cessé d'augmenter et à l'heure actuelle, sur certains liens du réseau Renater, il peut représenter la même proportion que le trafic HTTP (cf. figure 12). Bien sûr, Napster a été remplacé par Kazaa ou E-donkey. Une telle augmentation du trafic P2P a irrémédiablement eu un impact sur les caractéristiques du trafic global,en particulier, à cause de la nature des fichiers échangés (la plupart du temps de la musique ou des films) qui sont comparativement beaucoup plus longs que les flux du trafic web, le trafic majoritaire quelques années auparavant.

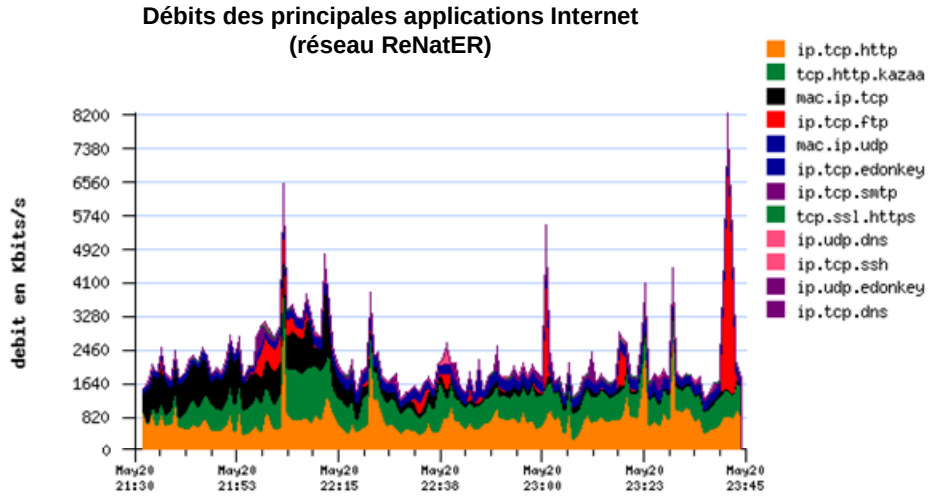

Figure 12. Répartition du trafic sur le réseau RENATER (mai 2003) - les applications sont classées dans l'ordre inverse sur la légende et dans le graphique}

En fait, cette augmentation du trafic P2P couplée à la présence du trafic classique induit les caractéristiques suivantes :

- Le trafic Internet est toujours composé de milliers de petits flux appelés souris (imputables principalement au trafic web ainsi qu'au trafic de contrôle P2P),
- Un nombre de flux éléphants qui ne cesse d'augmenter.

A tel point que la distribution de la taille des flux dans l'Internet change de façon importante. Ce phénomène a été analysé depuis le début des années 2000 et les résultats sont présentés dans la figure 13. La distribution exponentielle (celle possédant la queue la moins lourde), sert de référence car elle est proche du modèle de Poisson[16]. Nous pouvons voir sur cette figure que la proportion de très longs flux a augmenté de façon très importante depuis l'an 2000. Si en 2000, cette distribution n'était pas très éloignée d'une exponentielle, ce n'est plus du tout le cas à l'heure actuelle. Au contraire, elle dispose d'une queue très lourde et est très éloignée de la distribution exponentielle.

[16] Ce modèle est utilisé comme référence dans la majorité des cas lorsqu'il s'agit de réaliser des simulations ou des évaluations de performance en réseau.

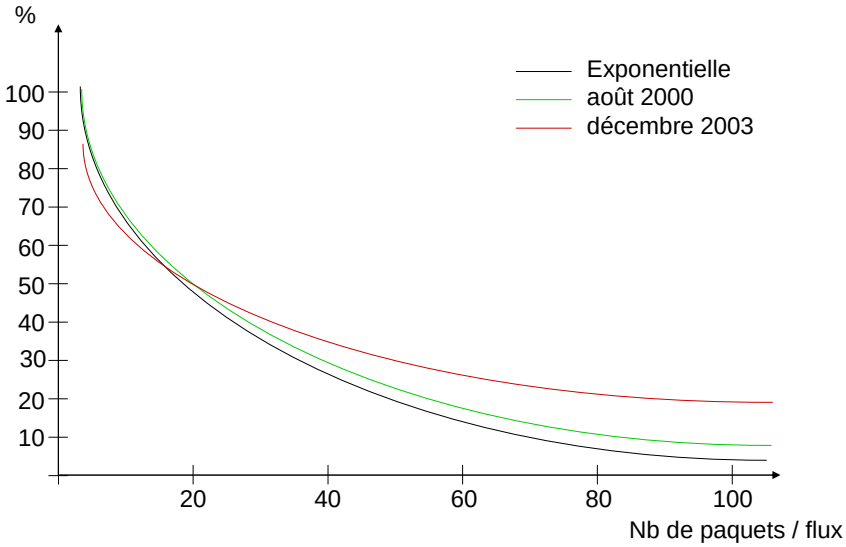

Figure 13. Evolution de la distribution des tailles de flux dans l'Internet entre 2000 et 2003

4.3.2. Mise en évidence de la dépendance longue dans le trafic

En revenant à l'évolution majeure du trafic Internet qui consiste en un nombre de plus en plus grand de flux longs, la figure 14 illustre les modifications que nous pouvons observer. Pour cela, elle compare le trafic Internet actuel avec un trafic qui suit un modèle de Poisson. Ces deux trafics sont observés à différentes granularités (0,01 s, 0,1 s et 1 s) et il est facile de remarquer que le trafic Internet ne se lisse pas aussi vite que le trafic Poissonnien.

L'analyse a montré que ce résultat est totalement dû aux éléphants présents dans le trafic Internet. En effet, la transmission d'éléphants crée dans le trafic l'arrivée d'une grande vague de données qui a la particularité de durer un temps relativement long (plus d'une seconde[17]). C'est pour cela que l'on observe cette différence entre les deux types de trafic : la transmission des éléphants induit des oscillations persistantes dans le trafic actuel.

[17] Les flux web sont traditionnellement transmis en moins d'une seconde dans l'Internet actuel.

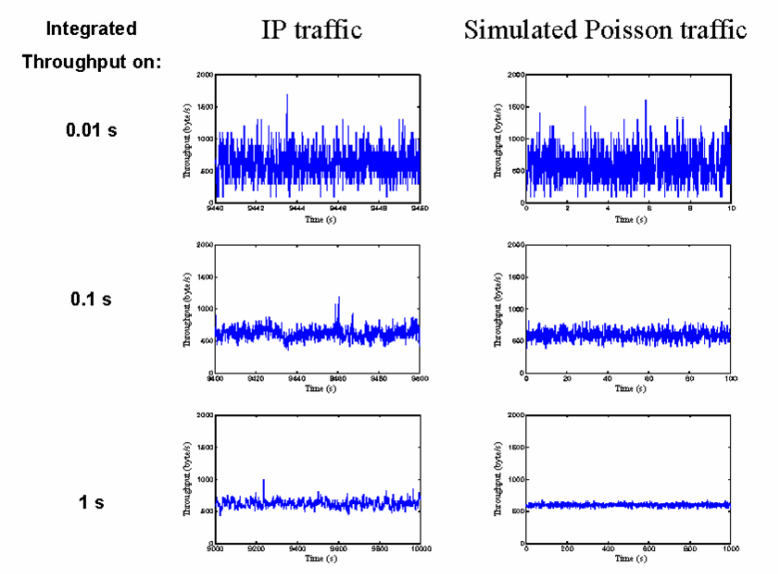

Figure 14 : Comparaison entre les oscillations observables dans un trafic Internet et un trafic Poissonnien. Cette étude réalisée dans le cadre de METROPOLIS se base sur du trafic d'une plaque ADSL de France Télécom.

De plus, les connexions TCP utilisées pour transmettre les flux éléphants plus volumineux durent plus longtemps et la dépendance qui existe entre les paquets d'une même connexion se propage ainsi sur des échelles de temps plus longues. C'est ce phénomène que l'on nomme traditionnellement LRD. On lui attribue plusieurs causes dont la principale est imputable aux mécanismes de contrôle de congestion de TCP (le protocole dominant de l'Internet). Parmi tous les mécanismes de TCP, il est évident que celui basé sur un contrôle en boucle fermée introduit de la dépendance à court terme, étant donné que les acquittements dépendent de l'arrivée d'un paquet, et que l'émission de tous les paquets suivants de la connexion est conditionnée par cet acquittement. De la même façon, les deux mécanismes de TCP (« slow-start » et « congestion avoidance ») introduisent de la dépendance à plus long terme entre les paquets de différentes fenêtres de congestion. Ainsi, en généralisant ces observations, il est évident que tous les paquets TCP d'une connexion sont dépendants les uns des autres. En plus, l'augmentation des capacités des liens de l'Internet en permettant la transmission de flux de plus en plus longs, augmente le phénomène de LRD. C'est pourquoi on observe sur la figure 14, la persistance d'un comportement oscillatoire dans le trafic Internet qui reste très marqué même avec une granularité d'observation importante (1 s).

Etant donné que le phénomène de dépendance de TCP se propage dans le trafic par l'intermédiaire des flux (i.e. les connexions TCP) [VER 00], l'augmentation de la taille des flux induit une augmentation de la portée de la dépendance qui peut atteindre des échelles très importantes. Ainsi, une oscillation au temps t induit alors d'autres oscillations à d'autres instants qui peuvent être potentiellement très éloignés de t. D'autre part, il est évident que les éléphants, en raison de leur durée de vie très importante dans le réseau et des grandes capacités de ce dernier (la plupart du temps les liens étant sur-dimensionnés), ont le temps d'atteindre de grandes valeurs pour leur fenêtre de contrôle de congestion. Ainsi, une perte induit pour le flux qui la subit une importante diminution, suivie par une importante augmentation de son débit. L'augmentation de la taille des flux favorise donc les oscillations avec une forte amplitude et un phénomène de dépendance à long terme.

Bien sûr, les oscillations sont très néfastes pour une utilisation optimale des ressources globales du réseau étant donné que la capacité libérée par un flux subissant une perte ne peut pas être immédiatement utilisée par un autre (en raison de la phase de slow-start notamment).

Ceci se traduit par un gaspillage de ressources et induit, une diminution de la QoS globale du réseau. En fait, plus le trafic oscille, moins les performances sont importantes [PAR 97].

En revenant à la courbe de la figure 9, obtenue en utilisant l'outil LDEstimate, et qui estime la LRD qui se manifeste dans le trafic à toutes les échelles temporelles, il est maintenant possible d'interpréter le résultat de bi-scaling observé. Ainsi, pour les échelles petites (octave < 8), c'est à dire les paquets proches les uns des autres, la dépendance est peu marquée. Par contre, pour les échelles plus grandes octave > 8), c'est à dire des paquets appartenant à des fenêtres de congestion consécutives, la dépendance est beaucoup plus importante. Evidement, ce phénomène existe pour l'ensemble des fenêtres de congestion d'un même flux. Ainsi, la présence dans le trafic de très long flux introduit un phénomène de dépendance à très long terme qui est visible sur la figure 12 pour les octaves très grands (> à 12). Ce niveau de LRD dans le trafic devient un problème majeur étant donné que chaque oscillation se produisant à un temps t peut se reproduire à n'importe quel temps t' qui est dépendant de t (en raison de la LRD qui existe entre les paquets échangés par le biais des protocoles traditionnels : ici TCP sur les longs flux). Il est intéressant de noter que nos expériences ont montré que le coude présent sur la courbe de LRD correspondait à la taille moyenne des flux, la partie droite de la courbe correspondant donc à l'impact des flux éléphants.

4.3.3. Démonstration du rôle de TCP sur la LRD

L'analyse du trafic qui vient d'être décrite, à laquelle vient s'ajouter notre connaissance des architectures et protocoles de l'Internet – sans oublier une part non négligeable d'intuition de la part de chercheurs – nous a permis de montrer l'effet négatif de TCP sur le trafic, lequel TCP apparaît comme inadapté pour transmettre des gros flux sur des liens à très hauts débits. Toutefois, en l'état actuel, ce résultat, même s'il apparaît cohérent, n'est pas prouvé. Mais maintenant que l'analyse métrologique nous a fait toucher du doigt les problèmes que TCP engendre avec le trafic actuel, il est facile de monter une expérimentation pour démontrer ce résultat. C'est ce qui va être fait dans la suite.

Evaluation de l'impact de TFRC sur la QoS

Ainsi, l'analyse précédente nous amène à penser que la LRD est un bon moyen de caractériser la variabilité du trafic, en particulier dans sa persistance. Aussi, l'expérience qui va être décrite dans la suite a pour objectif, sur un exemple, de montrer l'existence de ce lien entre les deux aspects variabilité et LRD. Pour ce faire, l'expérience menée s'est proposée de comparer au travers de simulations NS le trafic réel avec le même trafic re-simulé (le principe de la technique de rejeu est présentée dans [OWE 04a]), mais pour lequel le mécanisme de contrôle de congestion du protocole de transmission TCP a été remplacé par TFRC [OWE 04b] [FLO 01]. L'objectif de TFRC par rapport à TCP est de fournir des sources de trafic beaucoup plus lisses et régulières, c'est-à-dire des sources qui ne présentent pas ou peu d'oscillations. Ce mécanisme a été aussi défini pour permettre un meilleur transfert du trafic généré par les applications de streaming dans l'Internet qui nécessitent naturellement un maximum de régularité pour leurs débits d'émission et de réception. On montre ainsi que lorsque l'on emploie TFRC, et donc quand on génère un trafic régulier et lisse, la LRD qui apparaît dans le réseau est très réduite par rapport au cas où TCP est utilisé.

Cette expérience décrite plus précisément dans [OWE 04b] a pour objectif de fournir une étude comparative des caractéristiques globales du trafic suivant que les éléphants sont transmis en utilisant TCP ou TFRC. Cette expérience vise aussi à fournir des résultats dans un

environnement réaliste. L'étude comparative porte sur la trace originale d'une part et sur la trace simulée d'autre part dans laquelle les flux éléphants sont transmis en utilisant TFRC.

Par rapport au thème de cette étude comparative qui vise à étudier les effets de TFRC sur le caractère oscillant du trafic, les paramètres qui vont être évalués sont les paramètres traditionnels de débit, mais aussi des paramètres statistiques du trafic comme la LRD, et quelques paramètres mesurant le niveau de variabilité du trafic. Pour cela, nous utilisons un coefficient de stabilité (SC) qui est défini par le quotient :

$$Coefficient\_de\_stabilité(SC) = \frac{trafic\_moyen\_échangé}{écart\_type\_du\_trafic\_échangé(\sigma)}$$

La figure 15 présente le trafic dans les deux cas d'étude soit le cas réel et le cas simulé (avec TFRC). Visuellement, il apparaît clairement qu'en utilisant TFRC pour transmettre les éléphants à la place de TCP, le trafic global est bien plus lisse et régulier, et que tous les grands pics de trafic que l'on peut voir sur le trafic réel ont disparu du trafic simulé avec TFRC.

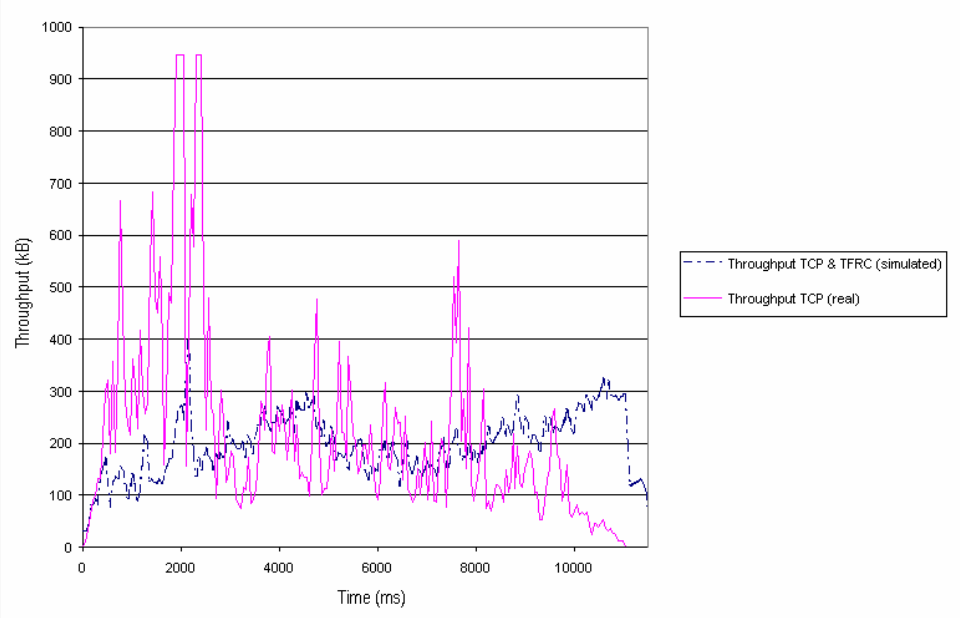

Figure 15 : Evolution du débit au cours du temps

Les résultats quantitatifs sont présentés dans le tableau 1. Ils confirment que la variabilité du trafic dans le cas du trafic réel (utilisant TCP pour transmettre les éléphants) est bien plus importante par rapport au cas simulé dans lequel les éléphants sont générés avec le protocole TFRC (voir les différences sur les écarts types et les coefficients de stabilité).

En ce qui concerne le débit global, nous avons mesuré des débits assez proches dans les deux cas. Ce résultat est excellent pour TFRC qui est par définition borné par le débit théorique de TCP. Il n'est de plus pas conçu pour consommer rapidement une grande quantité de ressources [OWE 03a], et même si TFRC est donc moins agressif que TCP, il est capable d'atteindre quasiment le même niveau de performance que TCP. Ceci confirme l'importance de la stabilité du trafic pour obtenir des performances de haut niveau et optimisées pour les réseaux de communication [PAR 97].

Protocole	Débit moyen (ko)	σ (ko)	CS

TCP New Reno (NR) : cas réel	82,335	157,959	0,521
TCP NR & TFRC : cas simulé	77,707	102,176	0,761

Tableau 1 : Caractérisation du débit pour les protocoles TCP et TFRC

En ce qui concerne la LRD, la figure 16 montre que dans le cas simulé la propriété de bi-scaling est sensiblement réduite et la courbe, même pour les grandes octaves a une pente peu marquée. Cela signifie que toutes les formes de dépendance, et en particulier celles à long terme ont été réduites de façon drastique. Les valeurs pour la LRD qui s'exprime à l'aide du facteur de Hurst sont: H(trafic réel) = 0.641 et H(trafic simulé) = 0.194 (ce qui dans ce dernier cas est non significatif, mais marque bien l'absence de LRD).

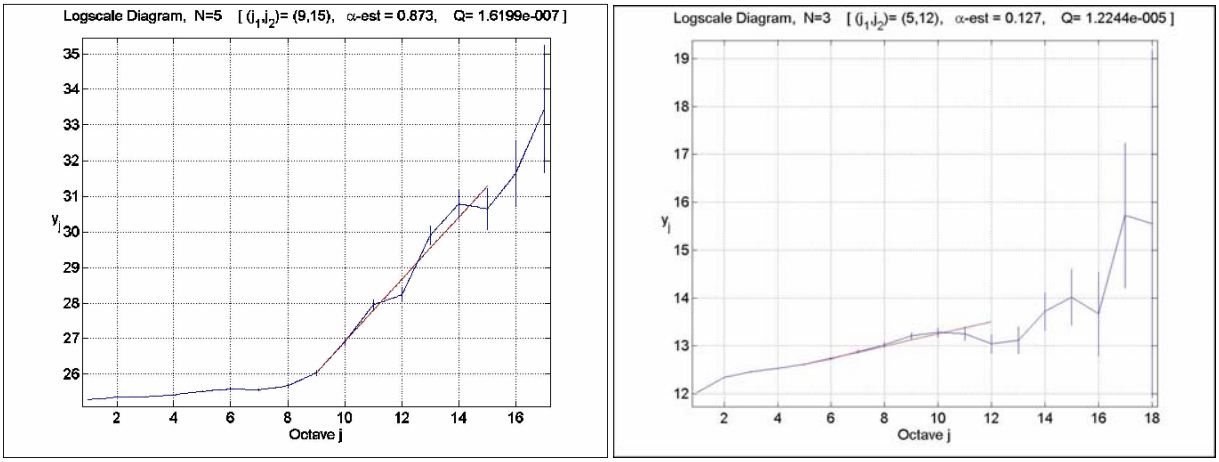

Figure 16 : Evaluation de la LRD pour le trafic simulé incluant des éléphants TFRC

La LRD : une métrique caractéristique de la QoS

Cette expérience a permis de mettre en évidence le lien étroit qui existe entre la caractéristique oscillante du trafic et la LRD. En effet, à partir du moment où on utilise pour transmettre les flux éléphants un protocole qui ne crée pas d'oscillations (TFRC) et qui brise le modèle de dépendance lors de la récupération des pertes, la LRD disparaît quasiment du trafic.

Ce résultat d'analyse est important car il donne un outil pour caractériser qualitativement et quantitativement un des phénomènes caractéristiques du trafic Internet, qui est de plus un élément dégradant de la performance du réseau. Surtout, il permet de donner des directions de recherche pour trouver des parades à ce phénomène, en particulier concernant les protocoles de transport et leurs mécanismes de contrôle de congestion.

Ce travail de caractérisation et d'analyse – par rapport à ce qui se fait dans le domaine et qui a juste pour vocation de trouver un modèle mathématique décrivant le trafic Internet – a donc bien permis d'analyser tous les phénomènes de variabilité du trafic de l'Internet et de les expliquer, mettant en cause notamment le comportement de TCP lorsqu'il est utilisé pour transmettre des flux éléphants sur des réseaux à hauts débits. En connaissant maintenant les mécanismes de TCP qui engendrent cette dynamicité dans son débit d'émission (conduisant à une inefficacité et une instabilité dommageable), nous avons maintenant les cartes en main pour proposer des solutions pour éviter la variabilité du trafic.

5. Conclusion

Cet article vient donc de dresser un panorama des activités de métrologie dans les réseaux de communication en général, et l'Internet en particulier. En premier lieu, nous y avons défini ce

qu'était la métrologie et avons situé son besoin au niveau de l'ingénierie et de la recherche en réseau. Un des éléments de la contribution de cet article – et qui est un élément clé de la métrologie de l'Internet – a été de décomposer les activités de métrologie en deux parties : une première partie concerne la mesure et l'observation à proprement parler des phénomènes se produisant sur le réseau, que ce soit en termes de mesures de la QoS ou d'observation du trafic transitant sur les liens ou dans le nœuds du réseau. La seconde partie concerne la mise en évidence au travers d'analyses poussées de phénomènes invisibles que l'on doit déduire des phénomènes observés. Ces phénomènes sont naturellement assez peu explicites puisqu'ils résultent de la superposition de très nombreux mécanismes et protocoles, de multiples applications, de comportements d'utilisateurs divers et variés et de topologies de réseaux très différentes d'un domaine (ou AS) à l'autre. En isoler les différentes composantes est donc particulièrement ardu et est aujourd'hui l'obstacle le plus important rencontré en métrologie.

L'article a donc présenté des outils de métrologie actifs et passifs qui permettent de répondre aux besoins de mesure ou d'observation du trafic et de la QoS du réseau. Après avoir détaillé ces besoins, et naturellement les points durs qui se présentent, l'article a mis en évidence les points de vue différents de chaque technique, et pour chacune d'elles, leurs avantages et inconvénients. Ces outils de mesure, dont le descriptif pouvait paraître vague dans leur présentation générale, ont été illustrés sur l'étude de cas que fut la mise en place d'une plate-forme de métrologie en bordure du réseau Renater dans le cadre du projet METROPOLIS. Il a ainsi été montré comment concevoir cette plate-forme pour répondre point par point aux besoins et résoudre les points durs.

Ensuite, l'article s'est focalisé sur les problèmes d'analyse du trafic permettant d'isoler, à partir de l'observation d'effets, les phénomènes et comportements invisibles du réseau. Comme nous l'avons déjà dit à plusieurs reprises, il s'agit là du problème majeur qu'il reste à résoudre, et l'état de l'art est encore loin des objectifs. L'article présente toutefois une méthode d'analyse qui emprunte ses outils à nos collègues du traitement du signal. L'article a ainsi montré que l'on pouvait mettre en évidence la responsabilité de certains protocoles (ici TCP) dans la limitation des performances et de la qualité des réseaux actuels. De plus, en mettant ce problème en évidence, et grâce à l'analyse précise qui a été faite des effets négatifs du protocole, la route à suivre pour corriger ce mécanisme protocolaire est tracée.

C'est donc bien le processus de recherche et d'ingénierie des réseaux qui est bouleversé par l'arrivée récente des activités de métrologie. Comme nous venons de le voir, une analyse hors ligne du trafic va permettre de modifier un des principaux protocoles de transport de l'Internet – le protocole TCP – afin de l'adapter aux besoins actuels (qui n'étaient pas ceux qu'on lui demandait de remplir il y a plus de 20 ans, au tout début de l'Internet). Ce travail est en cours, notamment à l'IETF avec la conception du protocole DCCP [HAN 03]. D'ailleurs, dans ce processus de recherche et d'ingénierie des réseaux, la connaissance accrue que l'on gagne sur le trafic ou la topologie des réseaux permet de modifier notre façon de faire des simulations ou des émulations. Ces simulations qui servent à tester des idées de mécanismes, de protocoles ou d'architectures que l'on peut avoir vont donc pouvoir bénéficier de cette connaissance, car nous pourrons simuler des topologies et des trafics réalistes. Nos mécanismes, protocoles ou architectures seront donc confrontés à des conditions expérimentales réalistes, et nous ne connaîtrons a priori plus les écarts incroyables entres les performances obtenues par nos propositions en simulation et dans le cas réel.

Enfin, le grand dessein de la métrologie est bien évidemment de passer du stade hors ligne au stade en-ligne (ou temps réel). En effet, l'objectif qu'il faut atteindre est de pouvoir bénéficier de tous les résultats de caractérisation et d'analyse de la QoS et/ou du trafic en temps réel, et ce en tous les points du réseau. Ainsi, on pourrait concevoir des mécanismes capables de réagir instantanément à l'apparition d'un phénomène dans le réseau. Toutefois,

avant qu'une telle approche ne puisse être déployée, il faut résoudre un certain nombre de problèmes : d'abord, avoir des outils de mesure active qui convergent très rapidement, ce qui n'est pas le cas aujourd'hui pour tous les paramètres de QoS. Il faut aussi que les sondes de capture puissent passer à la volée les données aux outils d'analyse, en évitant ainsi le délai du à l'utilisation d'un fichier intermédiaire. Naturellement, ceci ne peut être possible que si les outils d'analyse peuvent travailler à la vitesse du lien, ce qui est aujourd'hui loin d'être le cas. Cela situe l'ampleur de la tâche car nos analyses restent encore aujourd'hui très basiques par rapport à tous les phénomènes qu'il pourrait être nécessaire d'analyser pour concevoir un réseau dont les mécanismes se baseraient sur ces résultats d'analyse. Enfin, il faudra aussi concevoir un protocole de diffusion des résultats d'analyse dans tous le réseau qui soit à la fois performant face à un facteur d'échelle important et très rapide. Les premiers travaux de recherche autour d'une architecture réseau orientée mesures ont débuté, notamment au LAAS. Une de nos première contribution significative dans ce domaine s'appelle MBN (Masurement Based Networking), et repose sur deux éléments essentiels : une architecture de mesure MBA (Measurement Based Architecture), et un protocole de Reporting des résultats de mesure et d'analyse MRP (Measurement reporting Proctocol). Les premiers résultats d'application de cette architecture à un exemple de contrôle de congestion corrigeant les défauts de TCP sont présentés dans [LAR 02].

6. **Remerciements**

Les résultats présentés dans cet article sont le fruit d'un travail collectif qui a été mené pendant un petit peu plus de 3 ans dans le cadre du projet RNRT (Réseau National de la recherche en Télécommunications) METROPOLIS. Aussi, nous souhaitons remercier tous nos partenaires du projet METROPOLIS qui ont contribué à la mise en place des équipements de métrologie active et passive, et aux travaux d'analyse des mesures et traces collectées. Enfin, nous tenons à remercier Patrice Abry de l'ENS de Lyon et directeur de recherche au CNRS pour avoir contribué à l'analyse des traces de trafic en nous formant à l'utilisation des analyses à partir d'ondelettes et à l'outil LDestimate.

7. **Références**

[ABR 98] P. Abry and D. Veitch, 'Wavelet analysis of long-range dependent traffic', IEEE Trans. Information Theory, Vol. 44, 1998

[ABR 00] P. Abry., P. Flandrin., M. Taqqu, D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data", In Self-Similar Network Traffic and Performance Evaluation, K. Park and W. Willinger, Eds., Wiley, 2000

[ALM 99a] G. Almes, S. Kalidindi, M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999

[ALM 99b] G. Almes, S. Kalidindi, M. Zekauskas, "A One-way Packet Loss Metric for IPPM", RFC 2680, September 1999

[ALM 99c] G. Almes, S. Kalidindi, M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, September 1999

[APS 97] J. Apsidorf, "OC3MON: Flexible, affordable, high performance statistics collection", Proceedings of INET, June 1997

[BER 94] J. Beran, "Statistics for Long-Memory Processes", Monographs on Statistics and Applied Probability, Chapman and Hall, New York, NY, 1994

[BLA 92] U. Black, 'TCP/IP and related protocols', McGraw-Hill, 1992

[CIS 01] "NetFlow Services Solutions Guide", http://www.cisco.com/univercd/cc/td/doc/-cisintwk/intsolns/netflsol/

[COX 84] D. R. Cox, “Long-Range Dependance: A Review”, The Iowa State University Press, 1984

[DAC 94] D. Dacunha-Castelle and M. Duflo, « Probabilités et statistiques Tome 1 Problèmes à temps fixe », Editions Masson, Collection mathématiques appliquées pour la maîtrise, 2ème édition, pages 47-48, 1994

[DAG 01] „Dag 4 SONET network interface“, http://dag.cs.waikato.ac.nz/dag/dag4-arch.html

[DOW 99] A. B. Downey, “Using Pathchar to Estimate Internet Link Characteristics”, ACM SIGCOMM, 1999, pp. 222-23

[DOW 01] A. B. Downey, “Evidence for long tailed distributions in the Internet”, ACM SIGCOMM Internet Measurement Workshop, November 2001

[FLO 01] S. Floyd and V. Paxson, ‘Difficulties in simulating the Internet’, IEEE/ACM Trans. on Networking, Vol. 9, n° 4, Aug. 2001.

[GUO 05] J. Guojon. “Pipechar”, http://www-didc.lbl.gov/Pipechar

[HAN 03] M. Handley, S. Floyd, J. Pahdye and J. Widmer, ‘TCP Friendly Rate Control (TFRC): Protocol Specification’, RFC 3448, Proposed Standard, January 2003

[HU 03] N. Hu, P. Steenkiste, “Evaluation and Characterization of Available Bandwidth Probing Techniques”, IEEE Journal on Selected Areas in Communication, No 21, 2003

[JAC 97] V. Jacobson. “Pathchar -- a tool to infer characteristics of internet paths”. April 1997

[LAB 05] Y. Labit, P. Owezarski, N. Larrieu, "Evaluation of active measurement tools for bandwidth estimation in real environment", 3rd IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services (E2EMON'05), Nice (France), 15 Mai 2005

[LAR 02] N. Larrieu, « Métrologie des réseaux IP : développement de nouveaux outils pour caractériser, analyser et rejouer le trafic réseau », rapport de diplôme ingénieur INSA, juin 2002.

[LAR 05] N. Larrieu, P. Owezarski, "Measurement based networking approach applied to congestion control in the multi-domain Internet", 9th IFIP/IEEE International Symposium on Integrated Network Management (IM'2005), Nice, France, 15-19 May 2005

[LEL 93] W. Leland, M. Taqqu, W. Willinger and D. Wilson, “On the self-similar nature of Ethernet traffic”, ACM SIGCOM, September 1993

[MAH 00] B. A. Mah. Pchar. http://www.ca.sandia.gov/ bmah/Software/ Pchar/, 2000

[MAL 99] S. MALLAT, “A Wavelet tour of signal processing”, Academic Press, 1999

[MIL 96] D. Mills, “Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI”, Request for Comments 2030, October 1996

[NAV 03] J. Navratil. “ABwE: A Practical Approach to Available Bandwidth Estimation”, PAM 2003, La Jolla, April 2003

[OWE 03a] P. Owezarski, N. Larrieu, “Coherent charging of differentiated services in the Internet depending on congestion control aggressiveness”, Computer Communications Journal, Issue 13, Vol.26, August 2003

[OWE 04a] P. Owezarski, N. Larrieu, “A trace based method for realistic simulations”, IEEE International Conference on Communications (ICC'2004), Paris, France, June 20th-24th, 2004

[OWE 04b] P. Owezarski, N. Larrieu, “Internet traffic characterization - An analysis of traffic oscillations”, 7th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC'04), Toulouse (France), June 30 - July 2, 2004

[PAR 96] K. Park, G. Kim and M. Crovella, “On the relationship between file sizes, transport protocols, and self-similar network traffic”, IEEE ICNP, 1996

[PAR 97] K. Park, G. Kim, M. Crovella, “On the Effect of Traffic Self-similarity on Network Performance”, SPIE International Conference on Performance and Control of Network Systems, November, 1997

[PAR 00] K. Park and W. Willinger, ‘Self-similar network traffic : an overview’, In ‘Self-similar network traffic and performance evaluation’, edited by K. Park and W. Willinger, J. Wiley & Sons, 2000

[PAX 95] V. Paxson and S. Floyd, “Wide area traffic: The failure of Poisson modeling”, IEEE/ACM Trans. on Networking, Vol. 3, pp. 226-244, 1995

[PAX 98] V. Paxson, G. Almes, J. Mahdavi, M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998

[PAX 00] V. Paxson, A. Adams, M. Mathis, "Experiences with NIMI", PAM (Passive and Active Measurements) Workshop, 2000

[PLA avJC] Platon, « l’allégorie de la caverne », La République, Livre VII, IVème siècle av JC

[RIB 03] V.J. Ribeiro, R.H. Riedi, R.G. Baraniuk, J. Navratil, L. Cottrell, “PathChirp: Efficient Available Bandwidth Estimation for Network Paths”, *Passive and Active Measurement Workshop*, 2003

[ROB 00] J. Roberts, “Engineering for Quality of Service”, In ‘Self-similar network traffic and performance evaluation’, edited by K. Park and W. Willinger, J. Wiley & Sons, 2000

[STR 03] J. Strauss, D. Katabi, F. Kaashoek, “A Measurement Study of Available Bandwidth Estimation Tools”, ACM SIGCOMM Internet Measurement Workshop, 2003

[THO 97] K. Thompson, G. Miller and M. Wilder, ‘Wide-area internet traffic patterns and characteristics’, IEEE Network, Vol. 11, n° 6, Nov./Dec. 1997

[VEI 99] D. Veitch, P. Abry, “A wavelet based joint estimator of the parameters of long-range dependence”, IEEE Trans. on Info. Theory special issue on Multiscale Statistical Signal Analysis and its Applications 45, 3, Apr. 1999

[VER 00] A. Veres, Z. Kenesi, S. Molnar, G. Vattay, ‘On the propagation of long-range dependence in the Internet’, SIGCOMM'2000, Stockholm, Sweden, September 2000

[VER 00b] A. Veres and M. Boda, “On the Impact of Short Files and Random Losses on Chaotic TCP Systems”, in Proc.IFIP ATM & IP 2000 Workshop, Ilkley, UK, July 2000

[WIL 98] W. Willinger, V. Paxson and M. Taqqu, “Self-Similarity and Heavy Tails: Structural Modeling of Network traffic”, In A Practical Guide To Heavy Tails: Statistical Techniques and Applications, ISBN 0-8176-3951-9, 1998

[XU 01] J. Xu, J. Fan, M. Ammar, S.B. Moon, "On the Design and Performance of Prefix-Preserving IP Traffic Trace Anonymization", ACM SIGCOMM Internet Measurement Workshop, San Francisco, CA, November, 2001

[XU 02] J. Xu, J. Fan, M. Ammar, S.B. Moon, "Prefix-Preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme", IEEE International Conference on Network Protocols, Paris, 2002

8. Glossaire

ADSL	*Asymetric digital Subscriber Line*. ADSL est une technique de multiplexage de données numériques sur une ligne téléphonique analogique utilisée pour interconnecter des utilisateurs finaux au réseaux Internet.
AS	*Autonomous System*. Il s’agit d’un réseau opéré par une seule et même autorité.
ATM	*Asynchronous Transfer Mode*. C’est une technique de commutation de cellules qui a été pendant longtemps été utilisée comme une alternative à la commutation de paquets pour les réseaux numériques longues distances.
CNIL	Commission Nationale Informatique et Libertés.
DCCP	*Datagram Congestion Control Protocol*. Il s’agit d’un nouveau protocole de transport en cours de discussion à l’IETF qui pourrait remplacer TCP dans le futur.
ENS	Ecole Normale Supérieure.
ENST	Ecole Normale Supérieure des Télécommunications.
FAI	Fournisseur d’Accès Internet.
FTP	*File Transfer Protocol*. C’est le protocole applicatif de téléchargement des fichiers depuis un serveur.
GPS	*Global Positionning System*.
GTR	Génie Telecom et Réseaux.
HD	*Hard Drive* ou disque dur.
HTML	*Hyper Text Markup Language*. Langage de description de pages web.
HTTP	*Hyper Text Transfer Protocol*. C’est le protocole utilisé par le web pour le téléchargement et la navigation sur le web.
IANA	*Internet Assigned Numbers Authority*. C’est l’autorité qui attribute les numéros de ports aux applications references dans l’Internet.
IETF	*Internet Engineering Task Force*. Il s’agit d’un organisme qui développe et « normalise » les nouveaux protocoles et architectures pour l’Internet.
IP	*Internet Protocol*. C’est le protocole de base des communications Internet.

IPPM	*IP Performance Metrics*. Groupe de travail de l'IETF qui réfléchit aux problèmes de la mesure dans l'Internet.
IUT	Institut Universitaire de Technologie.
LAAS	Laboratoire D'analyse et d'Architecture des Systèmes.
LIP6	Laboratoire d'Informatique de Paris 6.
LRD	*Long Range Dependence*.
MetroMI	*Metropolis Measurement Infrastructure*.
METROPOLIS	Metrologie pour l'Internet et ses Services.
MGEN	*Multicast Generator*. Outil de génération de trafic multicast largement utilisé.
MIB	*Management Information Base*. C'est une base de données normalisée pour la gestion des réseaux à l'aide du protocole SNMP.
M-JPEG	*Moving-Joint Picture Expert Group*. C'est une norme de compression de séquences d'images vidéos pour laquelle chaque image est compressée et codée individuellement.
MPEG	*Moving Picture Expert Group*. C'est une norme de compression de sequences videos basée sur un codage différentiel.
MRTG	*Multi Router Traffic Grapher*. C'est un outil de supervision de la charge des liens d'un réseau.
NIMI	*National Internet Measurement Infrastructure*.
NS	*Network Simulator*.
NTP	*Network Time Protocol*. C'est un protocole de synchronisation des horloges de machines interconnectées par un réseau de communication.
P2P	*Peer-to-Peer* ou Pair-à-Pair. Protocole d'échange de fichier où tous les participants sont à la fois clients et serveurs.
PC	*Personal Computer* ou Ordinateur personnel.
PCI	*Peripheral Component Interconnect*. Standard pour les bus de communication des ordinateurs, notamment les PC.
QoS	Qualité de Service
RAM	*Random Access Memory*. Mémoire vive d'un ordinateur.
RAP	Réseau Académique Parisien.
RENATER	REseau NATional pour l'Enseignement et la Recherche
RIPE	Réseau IP Européens.
RTT	*Round Trip Time* ou temps d'aller-retour.
SMTP	*Simple Mail Transfer Protocol*. Protocole d'envoi d'e-mails.
SNMP	*Simple Network Management Protocol*.
TCP	*Transfer Control Protocol*. C'est le protocole de transport le plus utilisé dans l'Internet.
TFRC	*TCP Friendly rate Control*. C'est un nouveau mécanisme de contrôle de congestion proposé pour les applications orientées flux dans l'Internet.
TTM	*Test Traffic Measurement*. C'est le nom des sondes de mesure active de la société RIPE.
WWW	*World Wide Web* ou la toile.