

## **Foreword**

### **Analysis of traffic and usage traces on the Internet**

The Internet is nowadays investigated for discovering uses and practices of the different user communities. A large part of this work relies on traces and log files of Internet user activity which can be gathered from different network components (servers, routers... and users' terminals).

Researches are raised by very different aims as mathematical modeling of traffic in order to improve communication systems, applications and equipments used in telecommunication networks, for some of them; others are more interested by describing and interpreting the different kinds of uses and users from a sociological point of view.

Such researches have a wide range of applications in the Internet. On the one hand, network engineering is an obvious issue to studies aiming at analyzing and modeling traffic. On the other hand, more advanced analysis using a finer grain point of view, and then often focusing on individual network users, lead to a better understanding of uses and users requirements, and can then be used for building more suited services according to user profiles.

This special issue aims at bridging the gap between, on the one side, the uses and new sociability relations that appear on the Internet, and on the other side, technical issues for network and telecommunication engineers (for instance in terms of networks dimensioning, QoS, control, etc.). While engineers have already observed a deep evolution in the characteristics of traffic flows and launched several research projects to study this new dynamics, a deep insight of the sociological evolutions that have produced the observed phenomena has still to be achieved. Methods and tools for analyzing usage traces and logs files can then benefit from all the concerned disciplines: mathematics, graph theory, telecommunication networks engineering, computer science, networking, datamining, sociology, ergonomics...

In this special issue of "Annals of Telecommunications", we collected contributions from research teams representing the diversity of this domain.

The first paper (Beauvisage et al.) is clearly situated in the Internet uses sociology domain. In their introduction, the authors give an overview of web usage studies based on traffic analysis, divided into two categories, server-centric and user-centric; the authors' approach belonging to the latter. The authors describe a methodology and a set of tools they developed to analyze large Internet usage databases in France. They also provide some results allowing an almost historical approach (at the Internet history scale), describing what has been "Internet 1.0" in France in the 2000-2002 period.

In an effort to understand user behaviour in the context of the design of underlying networked multimedia distribution protocols, Katrinis et al report a study of turn taking in small (meetings) and large (lectures) group interactions captured in a spoken English corpus drawn from interactions in Higher Education. Building on previous studies in the CSCW literature they show, perhaps unsurprisingly that lecture time is dominated by the tutor although they note that in 15% of the cases some 50% of the attendees participated in some way. In contrast in the case of small group meetings they find that the next speaker is almost always one of the last 4 to speak irrespective of the size of the meeting group. However single speakers in meeting groups rarely achieve tutor-like dominance. These results provide useful heuristics for the design of

network protocols that can assume with a high degree of accuracy, which node will be the next main source of data although as the authors note, these kinds of interactions may be particular to the cultural context of the data corpus. In addition it is interesting to speculate on the potential intensification of these patterns that might be seen in a non-face-to-face situation where turn-taking cues may be less apparent.

Patterns of behaviour and interactions in a social context are also the subject of Dorat et al's paper which uses a corpus of 25,941 email messages sent over a period of 12 months to a French email list to demonstrate the presence of identifiable social structures and interaction networks. Taking a formal modelling approach they show that the interaction network has similar properties to previously studied 'real world' networks – its average degree is low and density small. More than 50% of the active participants had only responded to 5 messages producing an overall graph which was clearly non-random. By introducing the concept of threads and analysing the messages in a multilevel manner (messages within threads) they go on to show that it is possible to produce artificial networks matching the observed data very closely. Whilst the authors do not provide extensive sociological interpretations of their results they suggest that their modelling methods can be applied to the analysis of discussion contributions and participation in messaging systems to test the effects of various system design choices.

Fessant et al. apply sophisticated techniques (datamining, kohonen maps) to usage data of ADSL lines in order to infer qualified typologies of the customers using these lines. The paper starts by a presentation of a tool (probe) designed to gather data at the network level by analysing IP packets. This tool provides a typology of applications (VoIP, peer-to-peer, Web...) used by the ADSL customers. Then, datamining techniques, particularly Kohonen self-organising maps, allow the authors to present a segmentation of the panel of customers; each segment is described by usage variables. Thanks to a very convenient and rich visual presentation of this segmentation, the results might be of great interest for marketing purposes (typology and segmentation of customers databases).

The 5<sup>th</sup> article (Owezarski et al) serves as a transition in the thematic progression of this special issue. It opens the second part of this issue, dedicated to contributions using metrology techniques to give network engineers and researchers information they need for technologies, architectures and protocols evolution, in a way that corresponds to new applications and uses evolution, particularly in terms of Quality of Service (QoS). This article proposes a characterisation of current traffic and raises some issues with respect to implementation of the QoS mechanisms necessary for the new applications.

The 6<sup>th</sup> article (Coppens et al) deals with real-time supervision issues on high-speed networks. For this purpose, and depending on the characteristics of the traffic one wants to monitor, it is necessary to install filters, some of them being complex, for which computing time for each packet must be inferior to transit time of the same packet within the equipment. This paper presents the issues related to filtering in high-speed context and proposes solutions adapted to multi-gigabits/s networks.

In the last paper, JJ Pansiot presents a measure tool based on the *mrinfo* utility. This tool allows a local approach of network topology, thanks to fine-grained information

provided on a router neighbourhood. Moreover, thanks to its low cost, this measure can be performed at a relatively high frequency, which allows a precise approach of the network's topology dynamics. The author presents results obtained from a daily analysis over a long period (8 months). This approach is complementary with more classical methods mainly based on the usage of *traceroute*. Even if these approaches remain necessary for an overall approach of network topology, the tool proposed by Pansiot allows making a more precise zoom on local dynamics.

We are grateful to authors who positively replied to our solicitation, while the "bet" of this special issue was risky, given the broad scope of the subject and the difficulty to put together such various scientific disciplines and methodological approaches. We also warmly thank all experts who gave from their precious time to review the papers, their contribution was determining for the quality of this issue.

## **Editorial**

### **Analyse de trafic et de traces d'usage sur Internet**

Internet fait aujourd'hui l'objet de nombreuses investigations et recherches sous l'angle des usages et des pratiques qu'y développent les différentes populations d'internautes. Une part importante de ces travaux s'appuie sur les traces que laissent les internautes sur leur passage et qui peuvent être relevées à différents endroits du réseau (serveurs, routeurs, postes clients...).

Les chercheurs sont à la poursuite d'objectifs aussi divers que la modélisation mathématique du trafic dans le but d'améliorer les systèmes de communication, les applications et les équipements utilisés sur les réseaux pour les uns ; la description et l'interprétation d'un point de vue sociologique des différentes catégories d'usages et d'utilisateurs pour les autres.

Les applications de telles recherches sont nombreuses. D'une part, l'ingénierie des réseaux constitue un débouché évident aux études visant à analyser et à modéliser le trafic. D'autre part, des analyses plus fines, souvent centrées sur des utilisateurs du réseau, aboutissent à une meilleure compréhension des situations d'usage et des pratiques des internautes et peuvent être exploitées pour la conception de services plus adaptés aux profils des utilisateurs.

Ce numéro a pour ambition de faire le lien entre les questions d'usage et des nouvelles formes de sociabilité qui se développent sur Internet et des problématiques concrètes pour les ingénieurs en télécommunications et réseaux informatiques (en termes de dimensionnement des réseaux, de maîtrise de la QoS, etc.). En effet, si les ingénieurs ont déjà largement fait le constat d'une mutation profonde des caractéristiques des flux transitant sur les réseaux et entreprennent actuellement des programmes de recherche pour mieux modéliser cette nouvelle dynamique et en déduire des architectures et mécanismes protocolaires adaptés, une compréhension profonde des évolutions d'usage ayant engendré cette mutation demeure nécessaire. Les méthodes et outils d'analyse des traces d'usages constituent un terrain de rencontre possible entre les différentes disciplines concernées : mathématiques et théorie des graphes, ingénierie des réseaux télécoms, informatique, *datamining*, sociologie, ergonomie... Dans ce numéro spécial des Annales des Télécommunications, nous avons réuni des contributions d'équipes représentant la diversité de ce domaine.

Le premier article (Beauvisage et al.) est une contribution clairement située dans le domaine de la sociologie des usages d'Internet. Dans leur introduction, les auteurs situent leur travail dans le panorama plus large des études d'usage d'Internet fondées sur l'analyse de données de trafic, en distinguant les approches centrées serveur de celles centrées utilisateur, leur travail se situant dans cette dernière catégorie. Les auteurs présentent la méthodologie et les techniques qu'ils ont développées pour analyser des bases de données volumineuses d'usage d'Internet en France. Ils donnent également quelques résultats significatifs permettant un recul quasi-historique (à l'échelle de l'histoire d'Internet) de ce qu'a été "Internet 1.0" en France dans les années 2000 à 2002.

Dans un effort destiné à comprendre le comportement de l'utilisateur dans le contexte de conception des protocoles de distribution multimédia, Katrinis et al. présentent une étude de corpus d'interactions en langue anglaise au sein de groupes de tailles diverses (conférences avec une large audience, réunions en petits comités), les locuteurs appartenant à l'enseignement supérieur. En s'appuyant sur l'état de l'art dans

le domaine du CSCW, ils montrent que le temps des conférences est dominé par le tuteur ou enseignant, mais ils montrent également que dans 15% des cas, la moitié des participants contribuent d'une manière ou d'une autre. Par contraste, dans le cas de réunions en petit comité, ils montrent que le prochain locuteur est presque toujours parmi les 4 derniers locuteurs, quelque soit la taille du groupe. Ces résultats fournissent des heuristiques utiles pour la conception des protocoles réseau qui peuvent déduire, avec une précision élevée, quel nœud sera la prochaine source principale de données. Les auteurs notent toutefois que ces résultats sont sans doute spécifiques au contexte culturel particulier au sein duquel se sont déroulées les interactions observées.

Les schémas d'interaction dans un réseau social constituent le sujet central du travail présenté par Dorat et al. Les auteurs utilisent un corpus de 25941 courriels envoyés sur une période de 12 mois à une liste de diffusion française pour démontrer la présence de structures sociales identifiables et de réseaux d'interaction. En adoptant une approche formelle, les auteurs montrent que les réseaux d'interaction ont des propriétés similaires à d'autres réseaux du "monde réel" étudiés dans la littérature, à savoir un degré moyen et une densité faibles. Plus de 50% des participants actifs ont seulement répondu à 5 messages produisant ainsi un graphe qui était clairement non aléatoire. En introduisant le concept de fil de discussion et en effectuant une analyse "multi-niveaux" (messages à l'intérieur des fils de discussion), les auteurs montrent qu'il est possible d'engendrer des graphes qui collent assez fortement au graphe réel. Sans fournir une analyse sociologique très poussée, les auteurs proposent que leurs méthodes de modélisation soient appliquées à l'analyse de discussions dans des systèmes de messagerie afin de tester les effets des différents choix de conception. Fessant et al appliquent des techniques pointues (*datamining*, cartes de Kohonen) à des données d'usage de lignes ADSL pour en déduire des typologies qualifiées des clients utilisant ces lignes. Les auteurs présentent pour commencer un outil (sonde) de recueil de données d'usage situé au niveau du réseau et opérant par analyse des paquets IP. Ces données permettent d'avoir une typologie des applications (VoIP, peer-to-peer, Web...) utilisées par les clients ADSL. Par la suite, les techniques *datamining*, et en particulier les cartes auto-organisatrices de Kohonen permettent aux auteurs de présenter des segmentations des clients ADSL constituant leur échantillon, chaque segment étant décrit par des variables d'usage. Grâce à une restitution visuelle intuitive et riche, de tels résultats pourraient par exemple avoir des applications en marketing (typologies et segmentations de bases de clients)

Le 5<sup>ème</sup> article (Owezarski et al.) sert d'articulation dans la progression thématique de ce numéro spécial. Il ouvre la partie du numéro regroupant les contributions qui utilisent la métrologie afin de fournir aux chercheurs et ingénieurs en réseaux les informations dont ils ont besoin pour faire évoluer les technologies, architectures et mécanismes protocolaires de façon à leur faire mettre en œuvre les besoins liés aux nouvelles applications et nouveaux usages, notamment en termes de qualité de service (QoS). Ainsi, cet article propose une caractérisation du trafic actuel et montre les problèmes qu'ils posent par rapport à la mise en œuvre de mécanismes garantissant la QoS requise pour de nouvelles applications.

Le 6<sup>ème</sup> article (Coppens et al.) se focalise quant à lui sur les problèmes de supervision en temps-réel sur des réseaux à hauts débits. Pour cela, selon les caractéristiques du trafic que l'on souhaite surveiller, il faut installer des filtres, parfois complexes, pour lesquels la vitesse de traitement pour chaque paquet ne dépasse pas le temps de transit

du paquet au sein de l'équipement. Cet article présente les problèmes liés au filtrage à haut débit, et propose des solutions adaptées à des réseaux multi gigabits/s.

JJ Pansiot présente dans sa contribution un outil de mesure fondé sur l'utilitaire *mrinfo*. Cet outil permet une approche locale de la topologie du réseau, grâce aux informations très fines obtenues sur le voisinage d'un routeur. Par ailleurs, grâce à son faible coût, cette mesure peut être effectuée à une fréquence élevée, ce qui permet une approche précise de la dynamique de la topologie du réseau. Pour illustrer cela, l'auteur présente des résultats issus d'une analyse quotidienne sur une longue période (environ 8 mois). Cette approche présente un outil complémentaire aux méthodes plus classiques notamment celles fondées sur l'utilitaire *traceroute*. Même si ces dernières demeurent indispensables pour une analyse globale de la topologie du réseau, la méthode proposée par Pansiot permet de faire un zoom plus précis sur les dynamiques locales.

Nous tenons à remercier l'ensemble des auteurs pour avoir répondu positivement à notre sollicitation, alors que le pari de ce numéro spécial était risqué, étant donné la "largeur" du sujet et les difficultés à réunir dans le même numéro autant de disciplines et de méthodologies de recherche aux approches très différentes. Ce numéro n'aurait pas pu exister sans la participation active des relecteurs, qu'ils soient ici chaleureusement remerciés.

Housseem ASSADI

France Telecom, Division Recherche et Développement

42, rue des coutures, 14000 Caen, France

[housseem.assadi@orange-ftgroup.com](mailto:housseem.assadi@orange-ftgroup.com)

Philippe OWEZARSKI

LAAS-CNRS

7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, France

[owe@laas.fr](mailto:owe@laas.fr)

Ben ANDERSON

University of Essex – Chimera

Adastral Park, Ipswich, Suffolk, IP5 3RE, UK

[benander@essex.ac.uk](mailto:benander@essex.ac.uk)