

DES CHIFFRES ET DES LETTRES, DEUXIÈME SAISON



MARTIN.OUWELAND@epfl.ch, DOMAINE IT

Un article du FI4/05 présentait il y a trois ans quelques statistiques concernant le courrier électronique. Dans le présent article, nous allons présenter les données analogues pour la situation présente, permettant ainsi de se faire une idée de l'évolution durant cette période.

ADRESSES E-MAIL

Au moment où je rédige ces lignes, le 5 mai 2008, il y a 11524 adresses personnelles au format *prénom.nom@epfl.ch* en usage (4576 étudiants, 1920 assistants/doctorants, 3904 employés/accrédités et 1124 ex-membres qui verront leur adresse supprimée sous peu), 1020 adresses abstraites couvrant des fonctions, des projets, etc. (p.ex. *fi@epfl.ch* ou *support-it@epfl.ch*), et enfin 14803 adresses de listes de distribution telles que *personnel.epfl@epfl.ch*, automatiquement générées à partir des unités, classes et rôles figurant dans les données de l'accréditation (voir dinfo.epfl.ch/cgi-bin/listes.pl), pour un total de 27347 adresses.

adresses personnelles	11524
étudiants/doctorants/postgrade	4576
assistants-doctorants	1920
employés et autres accrédités	3904
<i>en partance</i>	1124
adresses de fonctions (aliases)	1020
adresses listes (1708 unités, 104 fonctions)	14803
adresses @epfl.ch	27347

MESSAGES

Passons au nombre de messages traités par les serveurs de messageries de l'EPFL. La figure ci-contre résume la situation, indiquant diverses moyennes par jour sur tout le mois d'avril 2008. Ainsi, sur la partie gauche on voit que nos serveurs reçoivent et acceptent de traiter chaque jour 515'000 messages, (465'000 externes + 50'000 internes) mais en refusent près de 200'000, essentiellement parce qu'ils s'adressent à des destinataires inexistantes (il y a énormément d'adresses périmées dans les listes d'adresses utilisées par les spammeurs.)

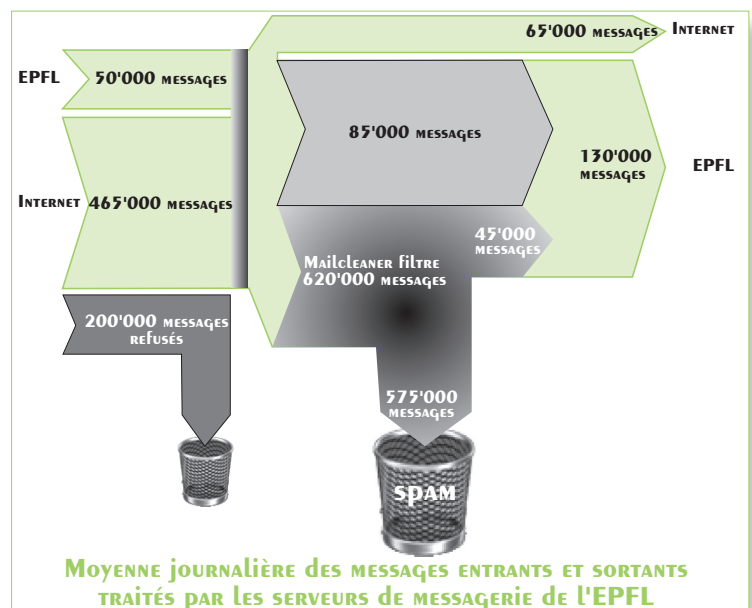
Sur la partie à droite figurent les messages envoyés par les serveurs, avec un effet de multiplication (les 770'000 messages en sortie représentant 1,5 fois le nombre de messages entrants) qui n'est que partiellement représenté avec les échelles choisies. L'explication en est simplement qu'un

message peut avoir plusieurs destinataires, un cas extrême étant représenté par un envoi à *personnel.epfl@epfl.ch* qui générera 5163 messages en sortie.

Sur les 705'000 messages destinés à l'EPFL, 620'000 sont d'abord soumis au filtre anti-spam MailCleaner (voir FI8/03) qui en retient 575'000 comme spam (chacun peut constater que son taux d'erreur ne dépasse pas quelques pour cent). Les autres 85'000 messages ne sont pas filtrés soit parce qu'ils proviennent de l'EPFL soit parce qu'ils sont destinés à des adresses qui ne sont pas abonnées à MailCleaner. Avec les 45'000 messages que MailCleaner laisse passer, nous obtenons 130'000 messages parvenant finalement dans les boîtes aux lettres des membres de l'EPFL. On peut donc dire que le courrier électronique est à l'EPFL plutôt un outil de travail interne, bien que les 65'000 messages envoyés depuis l'EPFL vers l'extérieur ne soient pas négligeables.

TAILLE DES MESSAGES

Passons à quelques chiffres sur la quantité de données que représentent tous ces messages, toujours en moyenne journalière sur le mois d'avril 2008. Tout d'abord en sortie, tous les messages à destination interne représentent 13 Gigaoctets, et les messages de l'EPFL vers l'extérieur 4,5 Gigaoctets. Si on sait que les 620'000 messages soumis à MailCleaner ne font que 7.6 Gigaoctets (comparés aux 5.4 Gigaoctets des 85'000 autres messages internes), on voit que le spam a au moins cette vertu d'être de taille plus modeste (environ 12 kilooctets en moyenne) que le reste des messages (64 kilooctets). Ceci explique aussi pourquoi, en entrée, les messages provenant de l'EPFL représentent au total 4,4 Gigaoctets et les messages externes 7,5 Gigaoctets, alors que ces derniers sont presque



10 fois plus nombreux: les messages internes sont donc en moyenne près de 6 fois plus gros que les messages externes. Un autre facteur contribuant à cette différence est le type d'utilisation de la messagerie électronique: en interne, on hésitera moins à se communiquer des versions provisoires de gros documents en ébauche.

Ces chiffres, de l'ordre de la dizaine de Gigaoctets, sont à comparer avec le trafic journalier global de l'EPFL, qui se comptera bientôt en milliers de Gigaoctets (voir *mathe.epfl.ch*). On peut donc mentionner, si on considère sa part du trafic global, que le spam n'entraîne sans doute pas comme on l'entend parfois dire des coûts importants, qui se situent plutôt dans son filtrage, automatique (MailCleaner) ou manuel (le temps consacré par chacun d'entre nous pour s'en débarrasser).

CONCLUSIONS

En comparant ces chiffres avec ceux d'il y a trois ans, on peut dégager deux tendances bien marquées. En premier lieu, et ce n'est pas une surprise, l'accroissement très important du spam: alors que le nombre de messages finaux et nettoyés du spam arrivant dans les boîtes aux lettres des membres de l'EPFL n'a que peu augmenté (8 messages par personne et par jour en moyenne en 2005, 11 en 2008) le nombre des messages externes entrants a été multiplié par 7 et le nombre de messages rejetés comme spam par MailCleaner a quant à lui été multiplié par 9 ! D'autre part, on constate aussi une forte croissance du nombre de messages envoyés par l'EPFL vers des adresses externes, puisqu'il a été multiplié par 4. ■