Détection d'attaques de « Déni de Services » : ruptures dans les statistiques du trafic

Pierre BORGNAT[1], Nicolas LARRIEU[2], Patrice ABRY[1], Philippe OWEZARSKI[2]

[1] Laboratoire de Physique (UMR CNRS 5672) de l'École normale supérieure de Lyon
46, allée d'Italie 69364 Lyon Cedex 07, France

[2]Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS)
7, avenue du Colonel Roche 31077 Toulouse Cedex 4, France

Pierre.Borgnat@ens-lyon.fr, nlarrieu@laas.fr, Patrice.Abry@ens-lyon.fr, owe@laas.fr

Résumé – Nous étudions les anomalies de trafic sur l'internet induites par le déclenchement d'une attaque de déni de service distribué, à partir d'une expérience où nous collectons le trafic à l'entrée du réseau local ciblé. L'étude concerne la caractérisation conjointe des statistiques d'ordre 1 (marginale) et 2 (covariance) du trafic agrégé. Les propriétés de mémoire longue du trafic ne sont pas sensiblement modifiées par l'occurrence de l'attaque, alors que les marginales des séries agrégées, correctement décrites par des lois gamma dont les paramètres évoluent continûment avec le niveau d'agrégation, portent la signature de l'attaque qui est une modification significative induite sur leur évolution à travers les échelles.

Abstract – We are studying anomalies of the Internet trafic caused by a Distributed Denial of Services (DDoS) attack, through an experiment we have done where we collect the trafic incoming on the target. The study proposes a joint characterisation of the 1st (marginal) and 2nd (covariance) order statistics of the aggregated process. The long memory, characterised by a wavelet approach, is not affected by the DoS whereas the parameters of a gamma model that fits well the marginals are: their evolution with the scale of aggregation is significantly modified by the attack and this is interpreted as a relevant statistical signature of the DoS.

1 Expérience d'analyse de trafic internet en présence d'anomalies illégitimes

1.1 Motivations

L'internet est un réseau de communication qui sert à véhiculer un grand nombre d'applications et est employé par un grand nombre d'utilisateurs. Il est soumis par conséquent à de fortes variations de trafic, certaines légitimes (variations journalières, affluence massive soudaine sur un site, de type *flash crowd* ou par effet *slashdot*, etc.) et d'autres illégitimes telles que les attaques contre un serveur. Dans le cadre du projet METROSEC nous étudions les anomalies de trafic et leurs signatures statistiques avec comme objectif de les caractériser, de savoir les détecter, de les prévenir et de défendre le réseau, contre les anomalies illégitimes en particulier. Le présent travail est consacré au premier aspect : caractériser une de ces anomalies illégitimes de trafic parmi les plus fréquentes, l'attaque de type « déni de service distribué », ou DDoS (pour *Distributed Deny of Service*) [6]. La détection d'anomalies à travers les statistiques du trafic a été abordée par le biais des caractéristiques spectrales [3, 6], de la covariance [7], ou de décompositions en ondelettes [2, 8]. Nous développons dans ce travail une description des données par une analyse multirésolution des statistiques d'ordre 1 et 2 du trafic. Nous décrivons d'abord les conditions expérimentales de réalisation de cette attaque et l'acquisition de traces de trafic à cette occasion, puis le modèle statistique employé et les paramètres multi-échelles qui apparaissent significatifs pour caractériser cette anomalie de trafic.

1.2 Attaque DDoS et collecte du trafic

Nous avons initié et contrôlé une attaque DDoS par TCP, consistant à inonder, de façon concertée à partir de plusieurs machines sources, une machine cible de paquets SYN-TCP (demande d'ouverture de connexion) — c'est le mécanisme du SYN *flooding*. Dans notre cas, 4 sites distants (université de Mont de Marsan, LIP6 de Paris, université de Coimbra au Portugal et un client situé sur une plaque ADSL parisienne) ont fourni des attaquants contre une machine située dans le réseau local de recherche du LAAS-CNRS, Toulouse. Le trafic complet (trafic de l'attaque ajouté au trafic usuel) est capturé paquet par paquet par l'intermédiaire de sondes DAG de métrologie passive [4]. Dans ce travail, nous nous concentrons en particulier sur l'analyse d'une trace enregistrée le 10 décembre 2004, mais d'autres expériences ont été réalisées.

Série temporelle étudiée. La trace s'étend sur $59 \cdot 10^3$s (soit 16h20) et se découpe en trois zones. L'attaque commence $8,5 \cdot 10^3$s (2h20) après le début de la mesure et dure $22,5 \cdot 10^3$s (6h15). La capture continue ensuite pendant $28 \cdot 10^3$s, mesurant alors le trafic de nuit, moins intense. La série temporelle analysée est le trafic agrégé pendant des intervalles de temps successifs de durée Δ. On étudie donc le nombre de paquets comptés durant Δ au temps $k \in \mathbb{Z}$, noté $X_\Delta(k)$. La figure 1 (en haut) présente cette trace agrégée avec $\Delta = 30$s. Elle se décompose visuellement clairement en 3 zones (avant, pendant et après l'attaque) et met en évidence l'augmentation du nombre de paquets circulant pendant l'attaque. Il est remarquable cependant de comparer sur la figure 1 la même série

agrégée à deux échelles très différentes $\Delta = 30$s et $\Delta = 1$ms. Alors que la première met clairement en évidence la non-stationnarité de la trace, du fait de l'occurrence de l'attaque ou d'augmentations spontanées du trafic, dans la seconde, agrégée à la résolution de 1ms qui correspond davantage à l'échelle de temps pertinente d'analyse et modélisation du trafic, ces non-stationnarités sont nettement moins faciles à déceler visuellement. Nous avons vérifié dans la suite que, pour chaque zone sur des périodes plus courtes (entre 1 et 30 minutes), les séries des X_Δ sont empiriquement, et raisonnablement, stationnaires, au sens où les paramètres statistiques étudiés dans la suite ne changent pas significativement.

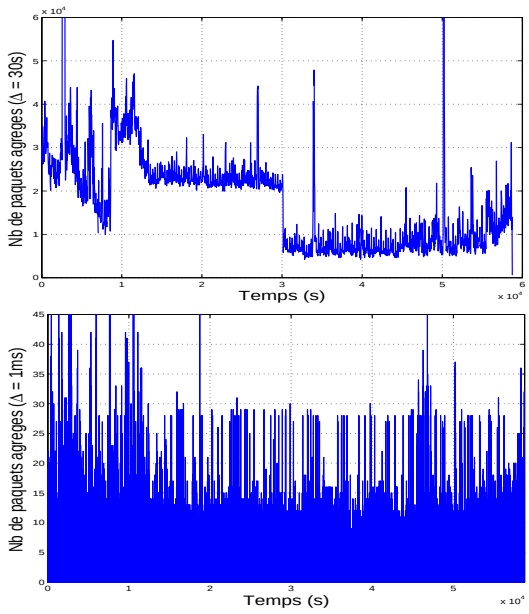

FIG. 1 – **Séries temporelles** obtenues en agrégeant le nombre de paquets dans des boîtes temporelles successives de taille $\Delta = 30$s (haut) ou $\Delta = 1$ms (bas).

Nous cherchons une signature de l'attaque dans une modification des statistiques du flux de trafic et il est naturel de s'intéresser d'abord aux propriétés statistiques aux deux premiers ordres : la distribution marginale et la covariance. Il est de plus maintenant bien établi que les traces Internet sont caractérisées par une propriété de mémoire longue, ou plus généralement par des caractéristiques multi-échelles [10]. Nous nous attachons donc à étudier les propriétés statistiques jusqu'à l'ordre 2 sur une large gamme de temps de mesure, par exemple en variant Δ.

2 Caractérisation des statistiques d'ordre 1 et 2 du trafic en présence de DDoS

2.1 Covariance et mémoire longue.

La propriété de mémoire longue des traces Internet se retrouve dans la covariance de X_Δ. Elle est particulièrement bien mise en évidence et analysée à travers une décomposition en ondelettes qui réalise une représentation sur des temps d'agrégation croissants. Notons

$$\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t-k) \text{ et } \phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t-k),$$

les dilatées et translatées sur la grille dyadique de, respectivement, une ondelette mère de référence ψ_0 et la fonction d'échelle qui lui est associée ϕ_0 [9]. On note alors $d_X(j,k) = \langle \psi_{j,k}, X_\Delta \rangle$ et $a_X(j,k) = \langle \phi_{j,k}, X_\Delta \rangle$ les coefficients d'ondelettes et d'approximations. Remarquons que la série des coefficients d'approximation $a_X(j,k)$ se lit moralement comme la série agrégée $X_{\Delta_j}(k)$ avec $\Delta_j = 2^j\Delta$. Si la paire ψ_0, ϕ_0 choisie définit l'ondelette de Haar, cette équivalence qualitative devient exacte : $X_{2^j\Delta}(k) = a_X(j,k)$. Varier le niveau d'agrégation Δ dans l'analyse revient donc essentiellement à faire une analyse multirésolution des données.

Le spectre S_X, ou la covariance de X_Δ, sont reliés aux coefficients d'ondelettes par [1] :

$$\mathbb{E}\{d_X(j,k)^2\} = \int S_X(\nu) 2^j |\Psi_0(2^j\nu)|^2 d\nu,$$

où Ψ_0 désigne la transformée de Fourier de ψ_0 et $\mathbb{E}$ l'espérance mathématique. La moyenne temporelle $1/n_j \sum_{k=1}^{n_j} |d_X(j,k)|^2$, notée S_j, estime la moyenne d'ensemble $\mathbb{E}\{d_X(j,k)^2\}$. On trace ensuite le diagramme log-échelle : $\log_2 S_j$ en fonction de $\log_2 2^j = j$. Dans ce diagramme, la longue mémoire se matérialise par l'apparition d'un segment de droite dans la limite des grandes échelles (j grand). Les performances attendues pour l'estimation par cette méthode ont déjà été discutées ailleurs [1].

La figure 2 (en haut) compare les diagrammes log-échelles obtenus, pour $\Delta = 1$ms, sur des portions de trace (choisies parce que stationnaires) d'une heure environ, avant, pendant et après l'attaque. Ces diagrammes se superposent quasiment, indiquant ainsi que la structure de dépendance statistique n'est pas notablement modifiée par l'occurence de l'attaque. La longue mémoire notamment, patente par l'alignement des diagrammes des droites pour les grands j, ne disparaît pas et n'est pas non plus modifiée sensiblement ni dans le sens d'une augmentation ni dans celui d'une diminution de son paramètre.

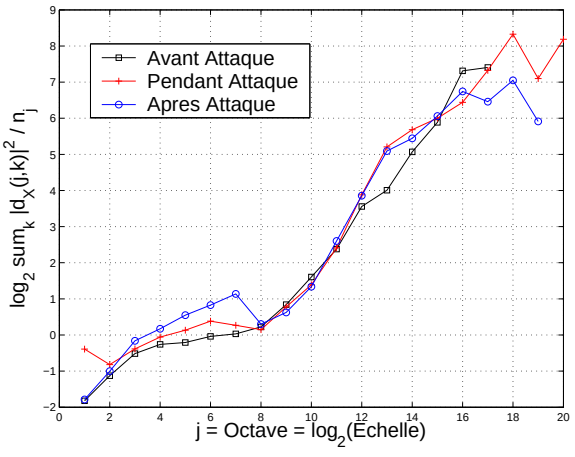

FIG. 2 – **Diagramme log-échelles.** L'occurrence de l'attaque ne modifie pas sensiblement les diagrammes donc la structure de covariance des données. La longue mémoire notamment n'est pas affectée.

2.2 Lois marginales.

Du fait de l'arrivée ponctuelle (paquet par paquet) du trafic, on s'attend à ce que la loi marginale passe d'un comportement exponentiel ou Poisson à petit temps d'agrégation Δ, à

un comportement proche du gaussien à grand Δ. Les lois marginales des coefficients d'approximations $a_X(j,k)$ en fonction de l'échelle j, donc, comme déjà dit, en fonction du niveau d'agrégation Δ_j, révèlent cet aspect. Une étude parallèle nous indique que les lois des coefficients d'ondelettes $d_X(j,k)$ sont moins pertinentes pour cette étude. Empiriquement, la famille des lois gamma fournit un modèle raisonnable des marginales des $a_X(j,k)$, pour les différents niveaux d'agrégation j. Rappelons qu'une loi $\Gamma_{\alpha,\beta}$ est définie par [5]

$$\Gamma_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right),$$

où $\Gamma(u)$ est la fonction gamma. Elle dépend d'un paramètre de forme α et d'un paramètre d'échelle β. Sa moyenne est $\mu = \alpha\beta$ et sa variance $\sigma^2 = \alpha\beta^2$. L'inverse du paramètre de forme, $1/\alpha$, indique la distance entre cette loi et la gaussienne. Par exemple, l'asymétrie et la kurtosis (moments relatifs d'ordre 3 et 4) se comportent respectivement en $2/\sqrt{\alpha}$ et $3 + 6/\alpha$.

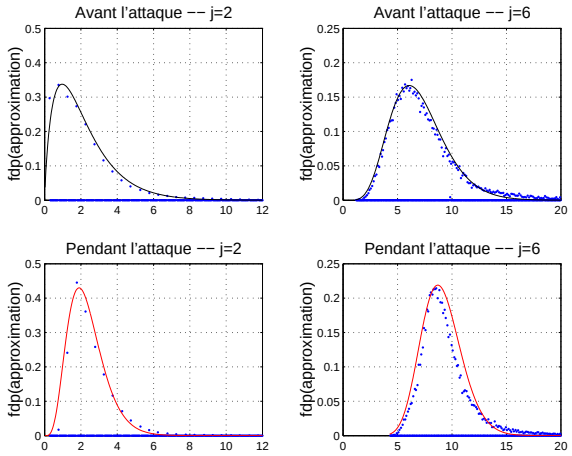

FIG. 3 – **Estimation et modèles des lois des $a_X(j,k)$.** Les lois marginales de $a_X(j,k)$, estimées sur une tranche de 30 minutes, sont tracées avant et pendant l'attaque, à $j = 2$ et $j = 6$. L'histogramme expérimental est superposé au modèle en loi gamma utilisé ici.

Estimation des paramètres à travers les échelles. Sur des périodes de temps assez courte pour que la série reste stationnaire (pour des critères empiriques), nous estimons les paramètres $\hat{\alpha}$ et $\hat{\beta}$ qui permettent de modéliser la loi de X_Δ par une loi gamma. L'analyse a été conduite sur des périodes de temps allant de 1 à 30 minutes ; ici on montre les résultats pour des tranches de 15 minutes. La méthode classique du maximum de vraisemblance est utilisée pour chaque niveau d'agrégation j, soit un temps Δ_j. L'accord du modèle avec les lois marginales expérimentales des traces est illustré sur la figure 3. Ce modèle est valable avec et sans anomalie de trafic : pendant l'attaque ou avant et après. Les mesures livrent donc une évolution à travers les échelles $\hat{\alpha}(j)$ et $\hat{\beta}(j)$ et c'est l'évolution qui est intéressante. La zone pertinente se développe sur les échelles $1 \leq j \leq 8$, c'est-à-dire pour des Δ_j de 1 ms à 0,3 s. Ce sont des temps plus courts que la zone dans laquelle la longue mémoire existe (au-delà de $j = 9$, soit à partir de 0,5 s environ).

La fiabilité de l'estimation de $\hat{\alpha}$ et $\hat{\beta}$ a été testée à l'aide de tests statistiques usuels. En fait, comme on dispose de générateurs de séries temporelles ayant des lois gamma comme marginales et une covariance avec de la longue mémoire [11], il a été possible de vérifier que la méthode classique d'estimation reste valable avec de telles corrélations à temps long.

Constations empiriques. Pour caractériser l'occurrence de l'anomalie, nous ne cherchons pas un changement dans les valeurs prises par ces paramètres à un niveau d'agrégation j fixé *a priori*, mais dans leur évolution en fonction de l'échelle j (ou le temps d'agrégation Δ_j). Notons que les lois gamma sont stables par addition : soient X_i and $i = 1, 2$ deux variables aléatoires indépendantes de loi $\Gamma_{\alpha_i,\beta}$, leur somme $X = X_1 + X_2$ suit une loi $\Gamma_{\alpha_1+\alpha_2,\beta}$. L'agrégation réalise

$$X_{2\Delta}(k) = X_\Delta(2k) + X_\Delta(2k+1).$$

Ainsi, si il n'y avait pas de dépendance d'un temps à l'autre, on s'attendrait à ce que α augmente linéairement avec Δ et que β reste constant. Nous montrons sur la figure 4 que $\hat{\alpha}(j)$ et $\hat{\beta}(j)$ ne suivent pas ces comportements : cela caractérise les dépendances temporelles fortes dans le trafic.

La figure 4 met en évidence que les estimées des moyennes $\hat{\mu}(j)$ et variances $\hat{\sigma^2}(j)$ des traces en fonction de l'échelle ne sont pas des caractéristiques significatives de l'attaque. La variabilité intrinsèque du trafic suffit à provoquer de fortes variations des paramètres et l'augmentation suit des lois similaires qu'il y ait une anomalie ou non ; le trafic de jour avant l'attaque et celui pendant l'attaque suivent des comportements analogues pour ces paramètres qui ne décrivent qu'un trafic chargé (alors qu'il est plus faible la nuit, après la DDoS). Au contraire, nous observons que les évolutions selon j des $\hat{\alpha}(j)$ et $\hat{\beta}(j)$ diffèrent notablement entre les portions avec et sans attaque. En condition de trafic normal, le paramètre de forme $\hat{\alpha}$ n'augmente pas à petite échelle mais reste constant jusqu'à $\Delta \simeq 20$ ms ; puis il croît comme $\log_2 \Delta$. Le paramètre d'échelle $\hat{\beta}$ évolue comme une loi de puissance sur toute la gamme d'échelle. Lors de l'anomalie, $\hat{\alpha}(j)$ subit une forte augmentation dès les petites échelles sans rester jamais constant. Parallèlement, $\hat{\beta}(j)$ suit un changement inverse : il diminue entre $\Delta \simeq 1$ ms et $\Delta \simeq 30$ ms, là où il devrait être croissant en trafic normal.

3 Discussion

3.1 Interprétation

La rupture de trafic causée par la DDoS trouve ici sa signature dans le changement des lois marginales mais pas dans la longue mémoire de la corrélation. Les noms de *forme* et d'*échelle* pour les paramètres α et β conviennent en fait à la description de ces changements quand l'attaque survient. Une première interprétation s'appuie sur $1/\alpha$ qui, on le rappelle, donne la distance entre la loi $\Gamma_{\alpha,\beta}$ et la loi gaussienne. Par agrégation, α augmente en cas de trafic normal. Ici l'attaque accélère la convergence vers une distribution gaussienne des traces et réduit l'*échelle* des fluctuations autour du trafic moyen.

Pour une interprétation plus poussée, on remarque que les cas où l'histogramme des a_X n'est pas nul en zéro (ou tend vers 0 en 0) sont typiques du trafic sans attaque et conduit à un facteur de forme α peu élevé mais une variance grande, et donc un β qui peut être grand (selon les fluctuations naturelles du trafic). D'un autre côté, les cas où l'histogramme s'annule

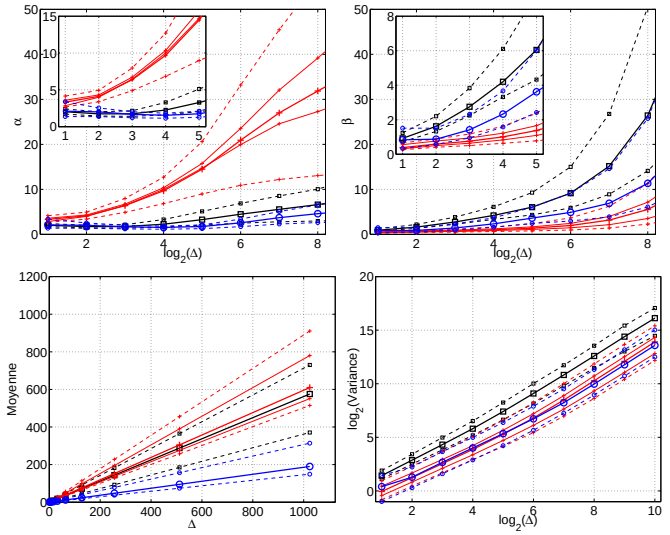

FIG. 4 – **Paramètres des lois marginales.** En haut : évolution en fonction de l'échelle de $\hat{\alpha}(j)$ et $\hat{\beta}(j)$ sur des tranches de 15 min, avant (carrés noirs), pendant (croix rouges) et après (cercles bleus) l'attaque. On montre l'évolution moyenne (traits épais) sur ces tranches et les valeurs extrêmes de l'évolution (pointillés) ainsi que deux exemples typiques d'évolution pendant la DDoS (traits fins) estimée sur 15 min. En bas : évolution de la moyenne (axes linéaires, Δ en s) et de la variance (axes log-log), représentées avec les mêmes conventions.

pour des valeurs inférieures à un seuil non nul est typique de l'attaque car le temps maximal d'inter-arrivée entre paquets est alors imposé par le SYN *flooding*. Cela conduit à une croissance constante et rapide de $\alpha(j)$ en fonction de j mais aussi à une dispersion (et donc un $\beta(j)$) faible et bloquée à court temps d'agrégation. L'attaque impose donc son rythme et supprime les fluctuations statistiques du trafic normal qui admettait des longs temps d'inter-connections, c'est-à-dire la possibilité de durées éventuellement longues durant lesquelles le nombre moyen de paquets reste très bas.

3.2 Perspectives

Nous avons analysé une signature de l'attaque par DDoS dans les statistiques du trafic. Plusieurs questions sont ouvertes par ce travail. La première est de savoir si cette signature, qui discrimine le trafic avec anomalie du trafic normal, existe aussi quand la rupture est légitime du fait d'un afflux important de trafic. Des premiers tests sur des expériences de *flash crowd*, où une centaine d'utilisateurs se sont connectés en même temps sur un serveur `WWW`, montrent que ce n'est pas le cas. Pour une telle anomalie légitime, c'est le diagramme log-échelle qui est modifié (pour des temps caractéristiques de l'utilisateur, autour de 1s) alors que l'évolution des paramètres de la loi marginale ne sont pas sensiblement affectés. Nous prévoyons de valider la pertinence de nos observations sur d'autres scénarios d'attaque, en classifiant différentes conditions de trafic normaux puis plusieurs types d'anomalies par cette démarche d'analyse des statistiques d'ordre 1 et 2.

Une deuxième perspective est de développer une stratégie de détection d'attaque, susceptible de déclencher une alerte dans un délai aussi réduit que possible. Pour cela nous étudions comment formuler un test statistique de changement des paramètres de la loi marginale à partir des constatations reportées ici. Un premier élément en faveur de cette extension est que les caractéristiques restent similaires si on estime les paramètres sur des tranches d'une minute seulement. Dans ces conditions, on peut envisager de mettre en œuvre une détection sur un horizon de quelques minutes, ce qui est un temps raisonnable pour des protocoles de défense du réseau.

Enfin nous travaillons à intégrer la non-stationnarité normale du trafic, due aux variations légitimes et attendues, dans la caractérisation statistique en allant au-delà d'une vérification empirique de la stationnarité sur des périodes courtes.

Remerciements. Ce travail a été mené dans le cadre du projet METROSEC de l'ACI Sécurité & Informatique. Nous remercions L. Gallon (LIUPPA, Université de Mont de Marsan) et L. Bernaille (LIP6, Paris) pour l'aide apportée dans la mise en place de l'attaque.

Références

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

[2] P. Barford, J. Kline, D. Plonka, and A. Ron. A signal analysis of network traffic anomalies. In *ACM/SIGCOMM Internet Measurement Workshop*, Marseille, France, 2002.

[3] C-M. Cheng, H.T. Kung, and K-S. Tan. Use of spectral analysis in defense against DoS attacks. In *IEEE Globecom*, Taipei, Taiwan, 2002.

[4] J. Cleary, S. Donnely, I. Graham, A. McGregor, and M. Pearson. Design principles for accurate passive measurement. In *PAM (Passive and Active Measurements) Workshop*, Hamilton, New Zealand, April 2000.

[5] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. Wiley (Interscience Division), June 2000.

[6] A. Hussain, J. Heideman, and C. Papadopoulos. A framework for classifying Denial of Service attacks. In *Proc. of SIGCOMM'03*, Karlsruhe, Germany, 2003.

[7] S. Jin and D. Yeung. A covariance analysis model for DDoS attack detection. In *IEEE International Conference on Communications*, Paris, France, June 2004.

[8] L. Li and G. Lee. DDoS attack detection and wavelets. In *International Conference on computer communications and networks*, August 2003.

[9] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Boston, 1998.

[10] K. Park and W. Willinger. Self-similar network traffic : An overview. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 1–38. Wiley (Interscience Division), 2000.

[11] A. Scherrer and P. Abry. Marginales non gaussiennes et longue mémoire : analyse et synthèse de trafic internet. In *Colloque GRETSI-05*, Louvain-la-Neuve, Belgique, 2005.