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), *École normale supérieure de Lyon*
46, allée d'Italie 69364 Lyon Cedex 07, France
Tél : *04 72 72 80 00* – fax : *04 72 72 80 80*
Pierre.Borgnat@ens-lyon.fr, Patrice.Abry@ens-lyon.fr
[2] Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS)
7, avenue du Colonel Roche 31077 Toulouse Cedex 4, France
Tél : *05 61 33 62 00* – fax : *05 61 55 35 77* – nlarrieu@laas.fr, owe@laas.fr

Soumis à la session spéciale "Traitement du Signal et Télétrafic Internet".

Problème traité.
Nous étudions les ruptures dans les statistiques du trafic induites par le déclenchement d'une attaque de type "Déni de Service distribué".

Originalité du travail.
Nous réalisons l'attaque, collectons le trafic à l'entrée du réseau local qui accueille la machine cible. Les statistiques d'ordre 1 (marginale) et 2 (covariance) du trafic collecté sont conjointement caractérisées. Ce travail est réalisé dans le cadre du projet METROSEC de l'ACI Sécurité & Informatique.

Résultats nouveaux.
Les propriétés de mémoire longue du trafic ne sont pas sensiblement modifiées par l'occurrence de l'attaque, alors que les marginales des séries de trafic agrégé, correctement décrites par des lois gamma dont les paramètres évoluent continûment avec le niveau d'agrégation, portent la signature de l'attaque qui est ainsi une modification significative induite sur cette évolution, plutôt que la valeur des paramètres des lois à un niveau d'agrégation arbitraire.

Résumé étendu

• **Attaques et Mesures.** – Nous avons réalisé de façon contrôlée une attaque de type "Déni de Service distribué" (DDoS, Distributed Deny of Service), consistant à inonder, de façon concertée à partir de machines sources, une machine cible de paquets SYN TCP (demande d'ouverture de connexion, SYN flooding). Dans notre cas, 3 sites distants (université de Mont de Marsan, LIP6 de Paris et un client situé sur une plaque ADSL parisienne) ont attaqué 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 (métrologie passive) [2].
• **Description de la trace.** – Dans ce travail, nous nous concentrons sur l'analyse d'une trace enregistrée le 10 décembre 2004. Cette trace s'étend sur $59 \cdot 10^3$s (16h20). 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 consiste en un trafic agrégé X_Δ (nombre de paquets comptés dans des intervalles de temps successifs de durée

Δ). La figure 1, à gauche, 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 les non-stationnarités existant dans la trace, qu'elles proviennent de l'occurence 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 cherchons une signature de l'attaque dans un changement significatif 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.

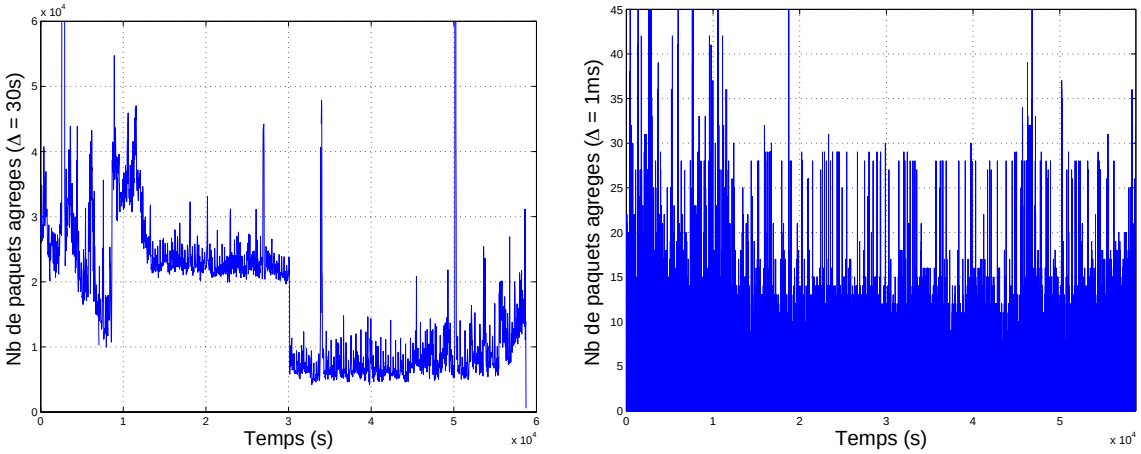

Fig. 1: **Séries temporelles :** obtenues en agrégeant le nombre de paquets dans des boîtes temporelles successives de taille $\Delta = 30$s (gauche) ou $\Delta = 1$ms (droite).

• **Covariance et mémoire longue.** – Il est maintenant bien établi que les traces Internet sont caractérisées par une propriété de mémoire longue, visible, par exemple, dans la covariance de X_Δ. Cette propriété est particulièrement bien mise en évidence et analysée à travers une décomposition en ondelettes. Notons $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ et $\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t - k)$ les dilatées et translatées sur la grille dyadique d'une ondelette mère de référence ψ_0 et de la fonction d'échelle qui lui est associée ϕ_0 [3]. On note alors respectivement $d_X(j,k) = \langle \psi_{j,k}, X_\Delta \rangle$ et $a_X(j,k) = \langle \phi_{j,k}, X_\Delta \rangle$ les coefficients d'ondelettes et d'approximations. Notons 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. Spectre Γ_X ou covariance de X_Δ sont reliés aux coefficients d'ondelettes par [1] :

$$\mathbb{E}\{d_X(j,k)^2\} = \int \Gamma_X(\nu) 2^j |\Psi_0(2^j\nu)|^2 d\nu, \tag{1}$$

où Ψ_0 désigne la transformée de Fourier de ψ_0 et $\mathbb{E}$ l'espérance mathématique. La moyenne temporelle $1/n_j \sum_{k=1}^{n_j} |d_X(j,k)|^2$ 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). La figure 2 (gauche) 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.

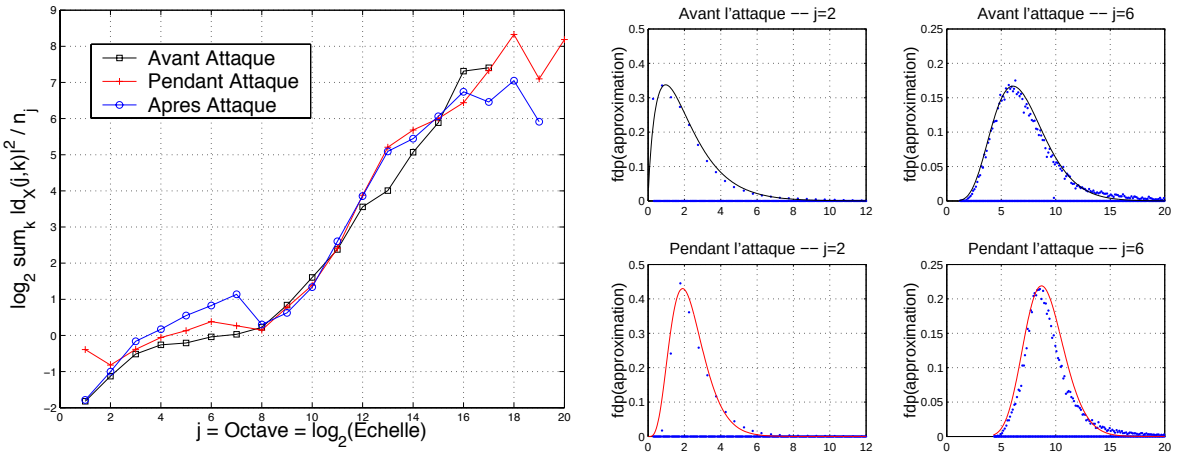

Fig. 2: **Diagramme Log-Echelles (à gauche) :** l'occurrence de l'attaque ne modifie pas sensiblement les diagrammes donc la structure de covariance des données. La longue mémoire notamment n'est pas affectée. **Estimation et modèles des lois des $a_X(j,k)$ (à droite) :** les lois marginales de $a_X(j,k)$, estimées sur une tranche de 30 minutes, sont tracées avant et pendant l'attaque, à $j = 2$ et $j = 6$. On a superposé aux points expérimentaux le modèle en loi *gamma* utilisé dans la suite.

• **Lois marginales.** – 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 elles une plus grande sensibilité. Une étude préliminaire nous indique que les lois des coefficients d'ondelettes $d_X(j,k)$ sont moins pertinentes pour notre étude. La zone intéressante se développe sur les échelles $1 \leq j \leq 8$, qui n'est pas celle dans laquelle la longue mémoire se développe (au-delà de $j = 9$).

Nous observons (cf. figure 2, à droite) que les lois marginales des $a_X(j,k)$ sont raisonnablement bien modélisées par des lois $\Gamma_{\alpha,\beta}$, pour les différents niveaux d'agrégation j, pendant l'attaque ou avant (après l'attaque a les mêmes propriétés). Pour chaque niveau d'agrégation j, nous estimons, dans des fenêtres d'observations suuccessives de durée 30 min, les moyennes $\hat{\mu}(j)$, variances $\hat{\sigma^2}(j)$, $\alpha(j)$ et $\beta(j)$ (α et β sont estimés par la méthode du maximum de vraisemblance).

Pour détecter l'occurrence d'attaques, nous ne cherchons pas un changement dans les valeurs prises par ces paramètres à un niveau d'agrégation j fixé *a priori*, mais dans l'évolution de ces paramètres en fonction de j. Nous mettons en évidence (figure 3, à droite) que de $\hat{\mu}(j)$ et $\hat{\sigma^2}(j)$ ne sont pas des caractéristiques significatives de l'attaque, car la variabilité intrinsèque du trafic suffit à provoquer de fortes variations des paramètres. Au contraire, nous observons, et cela constitue le résultat principal de ce travail, que les évolutions selon j des $\alpha(j)$ et $\beta(j)$ diffèrent notablement entre les portions avec et sans attaque. Pendant l'attaque, le paramètre d'échelle β diminue jusqu'à $j = 5$ alors qu'il augmente d'abord rapidement dans les autres zones, que le trafic soit chargé (le jour) ou non (après l'attaque, de nuit). Le paramètre de forme α sépare mieux encore les zones : sans attaque, il diminue jusqu'à un minimum entre $j = 3$ et 5 pour augmenter linéairement ensuite. Pendant l'attaque, α croît rapidement avec j. Pour $j = 2$, $\alpha = 1,9 \pm 0,3$ sans attaque et $\alpha = 4,3 \pm 0,2$ pendant l'attaque, pour atteindre des valeurs à $j = 8$ de $\alpha = 4,9 \pm 1,9$ et $\alpha = 31 \pm 6$ respectivement (moyennes et écarts type ont été calculées à partir des estimations sur chaque tranche). La même étude a été faite pour des fenêtres successives de 5 min. Les graphiques correspondants, qui seront montrés dans la version complète, conduisent aux mêmes conclusions.

• **Discussion.** – Guidé par les observations empiriques, le choix d'une description des marginales par des lois $\Gamma_{\alpha,\beta}$ est particulièrement pertinent. Par définition stable par addition (donc par agrégation), ces lois fournissent une famille accommodant *naturellement* une évolution avec la variation du niveau

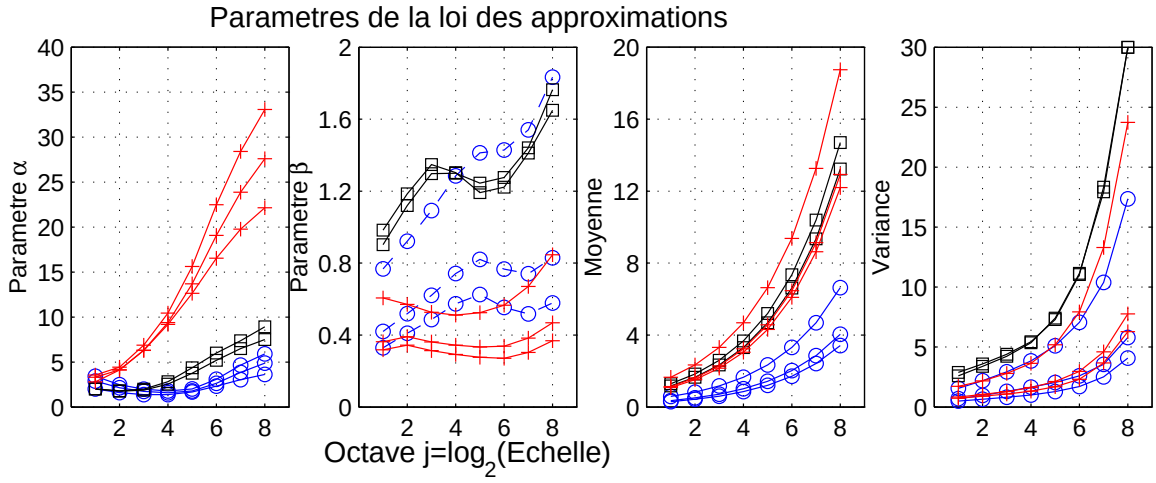

Fig. 3: **Paramètres des lois marginales :** Évolution en j des estimées (de gauche à droite) de estimations de $\alpha(j)$, $\beta(j)$ $\hat{\mu}(j)$ et $\hat{\sigma^2}(j)$ sur des tranches successives de 30 minutes. Chaque courbe correspond à une tranche : le jour avant attaque (ronds bleus), la nuit après l'attaque (carrés noirs), pendant l'attaque (croix rouges).

d'agrégation. De plus, les interprétations de *forme* et *échelle* des paramètres α et β semblent convenir à la description des changements de ces lois quand l'attaque survient. En effet, les cas où l'histogramme n'est pas nul en zéro (ou tend vers 0 en 0) est typique des zones sans attaque et conduit à un facteur de forme α peu élevé mais une variance grande, et donc un β qui peut être grand (fluctuations naturelles intermittentes du trafic) ; les cas où l'histogramme s'annule pour des valeurs inférieures à un seuil non nul est typique de l'attaque (car le temps maximal d'inter-arrivée est alors imposé par le SYN flooding) et conduit à un $\alpha(j)$ élevé croissant rapidement en fonction de j mais aussi à une dispersion (et donc un $\beta(j)$) faible. L'attaque impose son rythme et supprime les fluctuations statistiques du trafic normal qui autorise les 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. Le fait que les niveaux d'agrégation utilisés, $1 \leq j \leq 8$, restent nettement en deça de la zone où la longue mémoire se développe, $j \geq 10$, indique que l'évolution en j des $\alpha(j)$ et $\beta(j)$ rend compte des corrélations court-termes de X_Δ. Ce sont donc ces corrélations court-termes, beaucoup plus que la longue mémoire, qui sont modifiées par l'attaque. Cela confirme la pertinence de la concentration d'énergie visible pour $5 \leq j \leq 7$ dans les digrammes log-echelle de la figure 2 pendant l'attaque.
Nous envisageons, dans la suite de travail, d'une part, de valider la pertinence de nos observations sur d'autres scénarios d'attaque et, d'autre part, d'exploiter celles-ci pour développer une stratégie de détection d'attaque, susceptible de déclencher une alerte dans un délai aussi réduit que possible.

Remerciements. Ce travail a été mené dans le cadre du projet Metrosec sponsorisé par l'ACI Sécurité et Informatique. Nous remercions en particulier 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, D. Veitch, and P. Flandrin. Long-range dependence : revisiting aggregation with wavelets. *Journal of Time Series Analysis*, 19(3) :253–266, 1998. 2

[2] 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. 1

[3] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Boston, 1998. 2