Identification d'anomalies statistiques dans le trafic internet par projections aléatoires multirésolutions

Pierre Borgnat[1], Guillaume Dewaele[1], Patrice Abry[1]

[1]Laboratoire de Physique, École Normale Supérieure de Lyon, CNRS
46 allée d'Italie, 69364 Lyon Cedex 07
Pierre.Borgnat@ens-lyon.fr, Guillaume.Dewaele@ens-lyon.fr, Patrice.Abry@ens-lyon.fr

Résumé – Une méthode d'analyse des flux du trafic internet pour la détection et l'identification d'anomalies statistiques, utilise des projections aléatoires du trafic dans des *sketchs* qui sont alors agrégés à plusieurs échelles. Ces représentations permettent de formuler la détection des anomalies comme un test statistique de valeurs aberrantes, le trafic normal étant modélisé par des lois Gamma. La méthode fonctionne séquentiellement, en temps réel et permet, par inversion des *sketchs*, l'identification des adresses IP impliquées. Le résultat est illustré sur du trafic contenant des anomalies de réseau réalistes.

Abstract – An anomaly detection procedure for Internet traffic flows is proposed, using multiscale aggregated sketches (random projections). This representation enables the detection of anomalies as an outlier test, under the null hypothesis of normal traffic modeled by Gamma laws. The method works on-line, in real time, and with the inversion of the sketches, identifies the IP adresses of the anomalous flows. The result is illustrated on a realistic traffic containing anomalies.

1 Introduction

Une problématique de l'analyse du trafic internet est la détection des anomalies statistiques du trafic (par opposition aux anomalies de signature connue : virus, prise de contrôle d'un ordinateur) qu'elles soient des attaques malignes (dénis de services comme du SYN-flooding, scans de ports,...), des dysfonctionnements ou des sauts de trafic. Ces anomalies pouvant perturber la qualité de service, voire le fonctionnement même du réseau, les enjeux technologiques et économiques sous-jacents sont majeurs [1, 4].

Les méthodes classiques de détection de rupture statistique se heurtent à plusieurs difficultés dans le cas du trafic internet. Que l'on inspecte le trafic au niveau des paquets, des connexions (TCP) ou des flux (ensemble de connexions), l'espace de représentation est de grande dimension (2^{96} en considérant les adresses IP et les ports, sans prendre en compte d'autres caractéristiques – protocole, état de la connexion,...). Cela complique à la fois la détection des anomalies et l'identification des ordinateurs concernés. Une deuxième difficulté vient de la variabilité naturelle du trafic internet [8], à la fois non stationnaire (jour/nuit,...) et à longue mémoire. Choisir a priori des échelles de temps dans le trafic contraint à ne détecter que les anomalies opérant à ces échelles, alors qu'existent autant des pics intenses de trafic illégitime de courte durée que des attaques plus longues d'intensité moins visible [1].

Dans ce travail, nous combinons la représentation par *sketchs* des données internet [5], une méthode issue du *data streaming* [7], avec la modélisation des séries agrégées multirésolutions [10] pour proposer un algorithme de détection d'anomalies fonctionnant malgré ces deux écueils. Notons que des approches de réduction de dimension alternatives aux sketchs ont été proposées, utilisant une PCA [6, 11], ou une projection dans une variété non-linéaire [9]. Ici, la combinaison des sketchs et de l'approche multirésolution permet de formuler la détection comme un test statistique portant sur des valeurs aberrantes [2], la référence étant directement obtenue par moyenne sur les sketchs. La suite de l'article présente le type de trace de trafic réaliste d'intérêt pour ce travail. En section 3 nous introduisons les sketchs multirésolutions et la section 4 décrit le test de détection. La section 5 montre les résultats de détection sur la trace de validation. Nous concluons enfin.

2 Trace de validation des flux

L'étude de traces ressemblant à du trafic de réseaux commerciaux, produites par France Télécom dans le cadre du projet ANR-RNRT Oscar, a montré que les anomalies d'intérêt dans ce type de trafic sont des attaques de grande intensité, telles que des dénis de service par SYN-flooding ou des activités de scan massif. La méthode et le test proposés dans la suite sont adaptés à ce contexte. Une trace couvrant 66h, contenant $9 \cdot 10^7$ flux échantillonnés en gardant un paquet sur 500, impliquant 6,5 millions d'adresses IP source et 3 millions d'adresses IP destination sert à la validation et ici à l'illustration (fig. 1). Travailler ici sur la série des nouveaux flux permet de réduire dans un premier temps la quantité de données à traiter puisque l'on part de plus de $5 \cdot 10^{10}$ paquets.

Au-delà de cette réduction de données, il est plus facile pour des réseaux commerciaux d'obtenir des traces au niveau des flux que des traces avec des informations sur les paquets. Les routeurs peuvent en effet fournir les flux au cours du temps grâce aux logiciels netflow (Cisco) ou sflow (Juniper) par exemple, tandis que la mesure au niveau des paquets demande des instruments dédiés (cartes DAG,...).

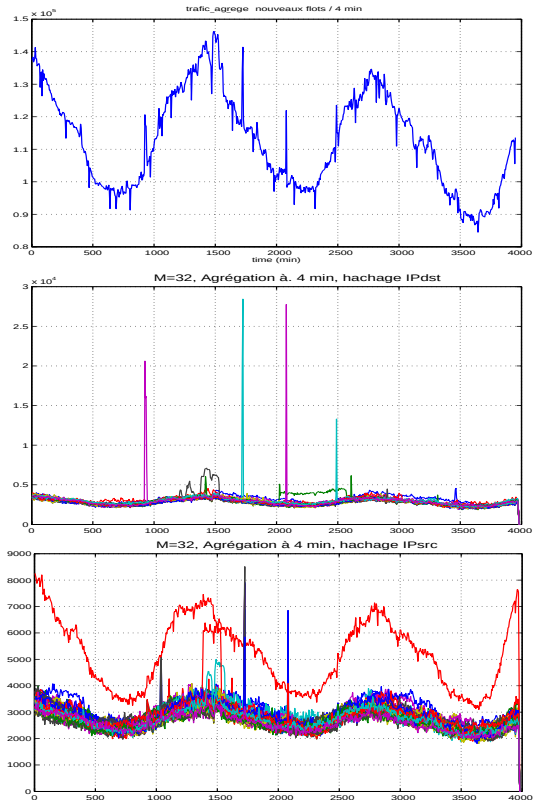

FIG. 1 – Trace étudiée (temps en min) — **Haut :** nombre de nouveaux flux agrégé à $\Delta_j = 4$ min. **Milieu :** séries des sketchs (#flux/Δ_j) agrégés à 4 min (une couleur par sortie m) pour une table de hachage sur IPdst. **Bas :** même chose avec hachage sur IPsrc. Les anomalies sont plus apparentes car le contraste est renforcé entre le trafic d'une anomalie (en général dans une seule boîte IPsrc ou IPdst) et le reste du trafic dans les autres boîtes.

Pour des anomalies de très faible intensité, des méthodes de détection basées sur des mesures de paquets, employant des sketches et les statistiques à plusieurs échelles (typiquement entre 1ms et 1s) des marginales des séries agrégées fonctionnent de manière satisfaisante [3]. Dans le présent travail, la contrainte est de se restreindre aux informations portant sur les flux, avec une granularité temporelle plus élevée (au minimum la seconde) car les statistiques des flux échantillonnés sont peu fiables en dessous.

3 Sketchs multirésolution

Le trafic internet est constitué de paquets ayant chacun des attributs tels les adresses source ou destination (notées IPsrc et IPdst), ses ports, son protocole, sa longueur, etc. On appelle *sketch* de la trace une représentation du trafic obtenu en appliquant N tables de hachage aléatoires différentes à l'un des attributs surveillés : ici IPsrc puis IPdst [5]. Chaque flux est ainsi attribué à une des M sorties de la table de hachage qui classe en mélangeant (mais de façon fixée et connue) l'espace des attributs et agit donc comme une projection aléatoire dans les M boîtes. La puissance de la méthode est que ce classement s'opère en temps réel grâce aux techniques de hachage rapide [12]. Si une anomalie existe, elle provoquera une rupture statistique dans une boîte donnée de chacune des N tables de hachage.

Nous construisons alors $N \times M$ séries temporelles agrégées à une première échelle de temps fine de surveillance Δ_0 (typiquement 1s) : $X_0^{n,m}(t)$ est alors le nombre de nouveaux flux pendant Δ_0 au temps t, pour la sortie m de la table de hachage n. Ne sachant pas sur quelle durée une anomalie sera visible, les séries sont (séquentiellement) agrégées sur des échelles de temps $\Delta_j = \Delta_0 2^j$. On note $X_j^{n,m}(t)$ ces séries agrégées pour $j \in \{0, ..., J\}$,. La figure 1 montre les M trafics agrégés, ici à $\Delta_j = 4$ min., après hachage sur IPdst ou IPsrc. Ces séries sont calculables en temps réel et stockées sur une fenêtre d'analyse $T > 2^J$, ce qui donne au final $N \times M \times T/\Delta_0$ données à manipuler. Typiquement on prend $N = 8$, $M = 32$, $T = 900$s soit une dimension acceptable de $2,3 \cdot 10^5$.

L'objectif est alors de déterminer, si I anomalies ont lieu au temps t, les classes $m_i^n(t)$, avec $i = 1, ...I < M$ pour $n \in \{1, ...N\}$, où une anomalie est présente. Ceci est l'objet du test de la partie suivante. Supposons dès maintenant les sorties avec alerte $\{m_i^n(t)\}_{i \in I_n, n=1,...N}$ obtenues, il suffit d'inverser la table de hachage pour identifier l'attribut associé aux anomalies repérées : IPdst ou IPsrc (selon la clef de hachage initiale). Ceci est réalisé par recherche exhaustive parmi les adresses IP observées, ce qui est réalisable puisque le hachage direct est très peu coûteux. On peut montrer de plus que le nombre de collisions[1] attendues est proportionnel à M^{-2N} [3, 12]. Il devient inférieur à 1 dès que $N \geq 6$ pour $M = 32$ pour un espace de départ de 2^{32} adresses. En utilisant ici $N = 8$, on peut identifier des anomalies qui arrivent aux mêmes instants tant que I reste petit devant N.

On identifie ainsi les ensembles de flux anormaux dans le trafic, en disposant de leur IPsrc ou IPdst, éventuellement les deux si un flux entre 2 ordinateurs seulement produit une anomalie de trafic. Le classement par IPsrc et IPdst permet de distinguer des anomalies de type différents : dénis de service (éventuellement distribués) pour IPdst et scans pour IPsrc.

4 Procédure de détection

La détection de la présence d'une anomalie dans une sortie de sketch s'appuie sur une modélisation non-gaussienne des statistiques du trafic. Les séries $X_j^{n,m}(t)$ suivent des lois Gamma comme montré en général dans [10]. Rappelons qu'une loi $\Gamma_{\alpha,\beta}$ est définie par

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

où α est le paramètre de forme et β le paramètre d'échelle.

Nous estimons d'abord les paramètres de la loi Gamma à partir du trafic (à l'aide de moyennes temporelles et à travers les sketchs) pour chaque fenêtre d'observation de durée T, finissant au temps kT. Puis nous détectons les anomalies par un test de valeur aberrante [2] contre ce modèle, en testant l'ensemble des séries agrégées $X_j^{n,m}(t)$

[1] Une collision se produit quand 2 attributs différents sont envoyés dans les mêmes sorties pour N fonctions de hachage différentes.

sur la fenêtre de temps $t \in [(k-1)T, kT]$. Les anomalies d'intérêt sont de deux types :

1. pic soudain d'activité dans une sortie (t, j et n fixés),
2. rupture statistique (potentiellement plus douce) d'un $X_j^{n,m}(t)$ au cours du temps, (j, n et m fixés).

Dans les deux cas, l'anomalie peut être vue comme une valeur aberrante pour la (ou les) sortie(s) contenant le trafic anormal, en comparaison du trafic normal dans l'ensemble de boîtes $m \in \{1, ...M\}$. En particulier dans le 2e cas, la surveillance des différentes échelles Δ_j en parallèle implique qu'on verra la rupture elle aussi comme une valeur aberrante, à une échelle j assez grande, plutôt qu'à l'aide d'un test de rupture statistique pour lequel on ne dispose pas de modèle d'évolution temporelle (potentiellement non-stationnaire) du trafic. La détection des anomalies revient ainsi à déterminer, pour chaque couple (n, m), si l'approximations à chaque échelle $X_j^{n,m}(t)$ suit la loi du modèle sans anomalie $\Gamma_{\tilde{\alpha}_j^{n,m}(k), \tilde{\beta}_j^{n,m}(k)}$.

Paramètre d'échelle $\tilde{\beta}_j^{n,m}(k)$: estimation et 1er test d'adéquation du modèle sur β. Pour du trafic normal, le paramètre de forme suit l'évolution du niveau de trafic alors que le paramètre d'échelle reste constant tant que le trafic ne change pas de nature. On utilise l'estimateur des moments (variance divisée par la moyenne) $\hat{\beta}_j^{n,m}(k)$ sur la k-ème fenêtre de durée T. Un premier test de valeur aberrante détecte les sorties où $\hat{\beta}_j^{n,m}(k)$ dévie significativement de l'ensemble des $\{\hat{\beta}_j^{n,m}(k), m = 1, ...M\}$. Le test est construit sur la distance de Mahalanobis sommée sur les échelles (en fixant le seuil $\zeta = 0,5$) [3] :

$$\left(\frac{1}{J} \sum_{j=1}^{J} \frac{(\hat{\beta}_j^{n,m}(k) - < \hat{\beta}_j^{n,m}(k) >_m)^2}{\sigma^2_{\hat{\beta}_j^n}} \right)^{1/2} > \zeta : \text{anomalie}$$

Les valeurs aberrantes ainsi détectées correspondent aux anomalies de type 1 de durée plus petite qu'au moins un des Δ_j, ou aux sauts brusques, conduisant à une distribution bimodale en général (pour laquelle le modèle ne tient en fait plus). Pour les sorties non aberrantes retenues, on lisse cette estimation par filtrage exponentiel récursif :

$$\tilde{\beta}_j^{n,m}(k) = (1-\gamma)\tilde{\beta}_j^{n,m}(k-1) + \gamma \hat{\beta}_j^{n,m}(k).$$

Le paramètre de mémoire γ est choisi pour fixer un temps caractéristique de une à deux heures (temps sur lequel le trafic reste stationnaire). On passe à la 2e étape pour les autres sorties de sketchs ; les anomalies déjà détectées sur $\hat{\beta}$ sont elles retenues en tant qu'alarmes dans les $\{m_i^n\}$.

Paramètre de forme $\tilde{\alpha}_j^{n,m}(k)$. Pour s'adapter à l'évolution temporelle du trafic, le paramètre de forme est estimé par la relation $\alpha\beta = \mu$ (valable pour une loi $\Gamma_{\alpha,\beta}$) à l'instant courant en utilisant la moyenne du trafic prise à la fois sur le temps T et **à travers les sketchs** :

$$\tilde{\alpha}_j^{n,m}(k) = \frac{< \mu_j^{n,m}(k) >_m}{\tilde{\beta}_j^{n,m}(k)},$$

où $\mu_j^{n,m}(k)$ est le trafic moyen. L'idée de cette estimation est qu'en situation de trafic normal, le hachage répartit de manière homogène le trafic dans les sorties. La moyenne prise sur les sorties de sketchs est donc censée être représentative de ce qui se passe dans chaque sketch.

2e test de détection des anomalies. Un test unilatéral (sur les grandes valeurs) est ensuite mis en place pour chaque série $X_j^{n,m}(t)$, contre la statistique sous hypothèse nulle (trafic normal) modélisée par $\Gamma_{\tilde{\alpha}_j^{n,m}(k), \tilde{\beta}_j^{n,m}(k)}$. Ce test est simplement le calcul de la probabilité P sous cette loi d'avoir une v.a. supérieure ou égale à $X_j^{n,m}(t)$. On fixe un seuil λ de fausse alerte, et on seuille la p-valeur :

$$\begin{cases} P_{\tilde{\alpha}_j^{n,m}(k), \tilde{\beta}_j^{n,m}(k)}(X > X_j^{n,m}(t)) \leq \lambda & : \text{anomalie,} \\ P_{\tilde{\alpha}_j^{n,m}(k), \tilde{\beta}_j^{n,m}(k)}(X > X_j^{n,m}(t)) > \lambda & : \text{normal.} \end{cases}$$

Basé sur le comportement empirique de cette statistique, on fixe dans la suite un seuil empirique à $\lambda = 10^{-5}$ pour décider entre ce qui est une anomalie et ce qui est normal, ce qui laisse une probabilité de fausse alarme inférieure à 1% pour un test sur $T/\Delta_0 \leq 10^3$ valeurs.

Plus précisément, en travaillant sur des fenêtres de durée T, on dit qu'une sortie de sketch m^n est anormale au temps k si le 1er test portant sur $\hat{\beta}$ lève une alarme ou s'il existe (au moins) une échelle j et un temps t dans la fenêtre tels que le 2e test révèle $X_j^{n,m}(t)$ comme étant une valeur aberrante. Il reste à inverser les sketchs $\{m_i^n(kT)\}_{i,n}$ pour retrouver les anomalies et identifier les ordinateurs, ainsi qu'expliqué à la fin de la section précédente.

5 Illustration et conclusion

La méthode a été testée sur la trace décrite plus haut. Les paramètres ont été fixés à $\Delta_0 = 1s$, $T = 15$min, $J = 8$ pour la représentation et les seuils sont $\zeta = 0,5$ et $\lambda = 10^{-5}$. La figure 2 montre la p-valeur du 2e test dans 2 sorties contenant une anomalie repérée ainsi. Les fluctuations normales des p-valeurs montre que le choix du seuil permet de garder un niveau de fausses alertes acceptable, tout en détectant les zones de ruptures statistiques de trafic. Notons en particulier l'intérêt d'une détection intrinsèquement multi-échelle (plusieurs j) : certaines anomalies ne se voient qu'à certaines échelles de temps.

Grâce à l'inversion des sketchs, on identifie les flux responsables des anomalies statistiques. On a ainsi reporté en figure 2 les trafics extraits pour les ordinateurs impliqués (ici des floodings). L'illustration porte ici sur la détection d'attaque de déni de service (anomalie impliquant une seule IPdst) mais le même travail a été fait pour détecter et identifier les anomalies à partir des IPsrc. Par exemple, le saut depuis une seule machine que l'on peut voir sur la figure 1 (en bas à droite, autour du temps 1500 et menant l'intensité dans sa boîte à 6000 flux/Δ) se révèle être ici un port de scan massif, distribué en terme de destination (donc invisible par hachage selon IPdst).

La méthode de détection proposée, basée sur deux tests de valeurs aberrantes pour des sketchs agrégés multirésolution, est adaptée au temps réel et à la surveillance pendant un long temps, en inspectant le comportement du trafic au niveau des flux, plutôt qu'à l'échelle des paquets. Malgré cela, nous avons illustré sur une trace que la méthode

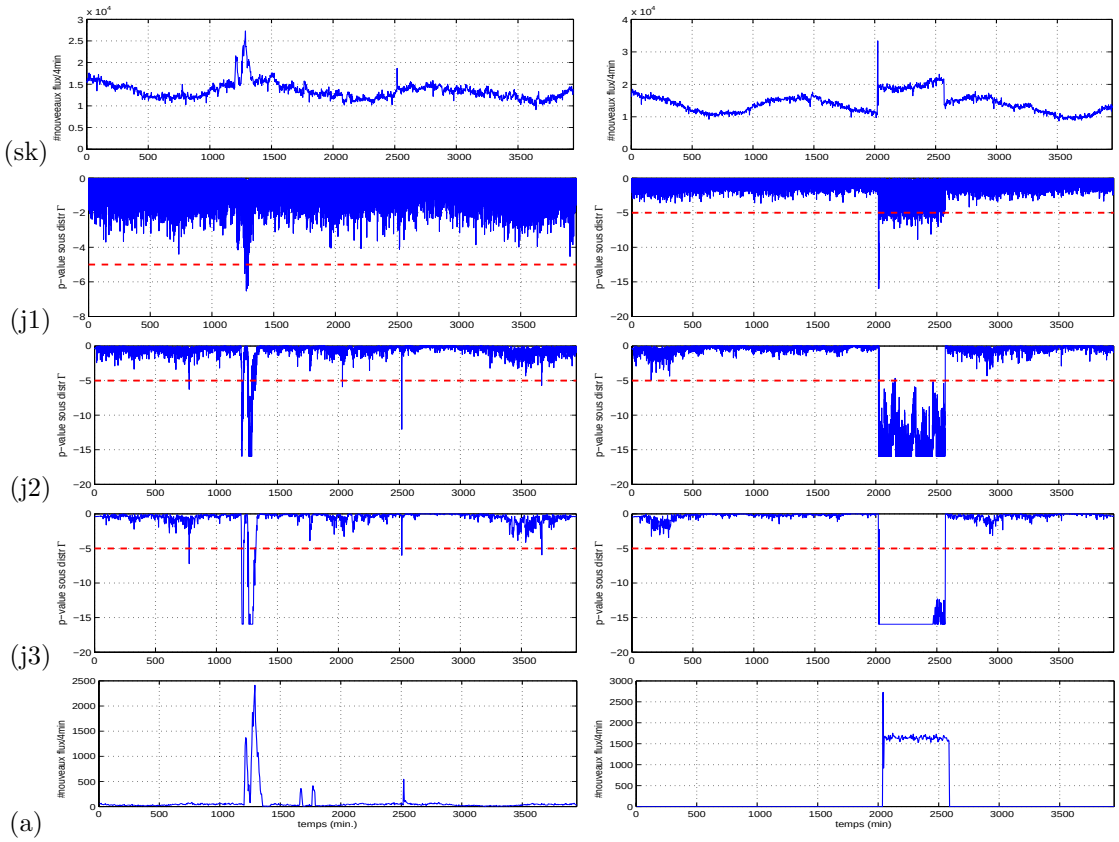

FIG. 2 – Résultat du test — Pour deux sorties m_1^n et m_2^n (gauche et droite) d'un sketch n (hachage sur IPdst), où apparaissent des anomalies, la première ligne (sk) montre les séries agrégées à $\Delta_j = 4$ min (comme en fig. 1, abscisses en min). En dessous : log(p-valeurs) pour le 2^e test, pour j1=1, j2=6 et j3=8 correspondant à 2s, 1 min et $\sim$4 min. Le seuil du test est $\lambda = 10^{-5}$ (ligne pointillée). Certaines anomalies ne se voient qu'à certaines échelles (comme le pic à $t = 500$ (à gauche), visible pour (j2) seulement). La ligne (a) correspond au trafic (agrégé à 4 min.) à destination de l'ordinateur identifié ainsi et révèle bien un trafic anormal aux instants détectés par le test.

est opérationnelle. La comparaison avec les méthodes de détection et d'identification d'anomalies (toujours à l'aide de sketchs) au niveau paquet [3] est une perspective de ce travail, de manière à mieux quantifier ce que l'on perd en opérant avec des données de flux seulement.

Remerciements. Ce travail a été mené dans le cadre de l'ANR-RNRT OSCAR. Nous remercions F. Guillemin (RD-CORE-LAN, France Telecom-Orange) pour la mise à disposition de la trace.

Références

[1] P. Barford, J. Kline, D. Plonka, A. Ron. "A signal analysis of network traffic anomalies." *ACM/SIGCOMM IMW*, 2002.

[2] V. Barnett, T. Lewis. *Outliers in Statistical Data.* John Wiley & Sons, 1994.

[3] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, K. Cho. "Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures." *ACM/SIGCOMM Workshop LSAD*, 2007.

[4] J. Jung, B. Krishnamurthy, M. Rabinovich. "Flash Crowds and Denial of Service Attacks : Characterization and Implications for CDNs and Web Sites". *Int. WWW Conference*, 2002.

[5] B. Krishnamurty, S. Sen, Y. Zhang, Y. Chen "Sketch-based change detection : Methods, evaluation, and applications." *ACM IMC*, 2003.

[6] A. Lakhina, M. Crovella, C. Diot. "Characterization of network-wide anomalies in traffic flows." *ACM IMC*, 2004.

[7] S. Muthukrishnan. "Data streams : Algorithms and applications." *Proc. ACM-SIAM SODA*, 2003.

[8] K. Park, W. Willinger (ed.) *Self-Similar Network Traffic and Performance Evaluation*, Wiley (Interscience Division), 2000.

[9] N.Patwari, A. Hero : "Manifold learning visualization of network traffic data." *SIGCOMM Workshop on Mining Network Data*, 2005.

[10] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, P. Abry. "Non gaussian and long memory statistical characterisations for internet traffic with anomalies." *IEEE TDSC*, 4(1), p. 56–70, 2007.

[11] M.L. Shyu, S. C. Chen, K. Sarinnapakorn, L. Chang. "A novel anomaly detection scheme based on principal component classifier." *ICDM'03 Workshop*, 2003.

[12] M. Thorup, Y. Zhang. "Tabulation based 4-universal hashing with applications to second moment estimation." *Proc. ACM-SIAM SODA*, 2004.