Détection d'attaques de Dénis de Service par un modèle non gaussien multirésolution

Pierre Borgnat, Nicolas Larrieu, Philippe Owezarski, Patrice Abry, Julien Aussibal, Laurent Gallon, Guillaume Dewaele, Karima Boudaoud, Laurent Bernaille, Antoine Scherrer, et al.

Détection d'attaques de Dénis de Service par un modèle non gaussien multirésolution

P. Borgnat[1], N. Larrieu[2], P. Owezarski[2], P. Abry[1], J. Aussibal[4], L. Gallon[4], G. Dewaele[1], K. Boudaoud[5], L. Bernaille[6], A. Scherrer[3], Y. Zhang[2], Y. Labit[2]

[1]*Laboratoire de physique, UMR CNRS-ENS Lyon ;* [2]*LAAS, UPR CNRS ;* [3]*LIP, UMR CNRS-INRIA-UCBL-ENS Lyon ;* [4]*IUT Mont-de-Marsan ;* [5]*I3S, UMR CNRS, Sophia Antipolis ;* [6]*LIP6, UMR-CNRS, Paris VI.*

RÉSUMÉ. *Nous nous intéressons à la détection d'attaques sur le réseau Internet. Nos procédures de détection reposent sur l'utilisation de lois non gaussiennes pour modéliser conjointement les distributions marginales du trafic agrégé à différents niveaux. Nous utilisons ensuite plusieurs distances statistiques pour quantifier une rupture dans cette modélisation statistique entre, d'une part, celle estimée dans une fenêtre d'observation et d'autre part celle obtenue sur une référence, constituée par un trafic normal. Les méthodes proposées ont une nature fondamentalement multirésolution : plusieurs niveaux d'agrégation sont analysés conjointement. Nous avons réalisé une série d'expériences d'attaques de Dénis de Service Distribuées (DdSD) et mesuré le trafic sur le lien d'accès de la victime. Contrôlant précisément les caractéristiques de ces DdSD, nous disposons d'une base de données documentée pour valider les performances des procédures de détection proposées. Nous traçons des courbes de performances (probabilité de détection versus probabilité de fausses alarmes) des procédures proposées et montrons qu'elles permettent la détection des attaques, à l'horizon de la minute, même lorsque celles-ci ont une faible intensité en volume.*

ABSTRACT. *In the present work, we address the issue of detecting attacks over the Internet. The detection procedures we propose are relying on a non Gaussian joint modelling of the marginal distributions of traffic, aggregated at several levels. Various statistical distances are then used to evidence a significant change between the parameters of the models estimated over a running observation window and a reference window, containing only normal trafic. The proposed approaches are fundamentally multiresolution by nature: several aggregation levels are analyzed simultaneously. We also performed a collection of Distributed Denial of Service (DDoS) attacks and collected the corresponding traffic. Because we control precisely the characteristics of these DDoS, we have at disposal a documented data base that we use to assess the statistical performances (correct detection probability versus false alarm probability) of the detection procedures. We show that they present satisfactory detection scores when the observation windows are of the order of one minute even for attacks whose intensity is low.*

MOTS-CLÉS : *Attaques, Déni de Service, modèle non gaussien, détection, divergence de Kullback.*

KEYWORDS: *Attacks, Denial of Service, non Gaussian modelling, detection, Kullback divergence.*

1. Motivation

L'Internet, parce qu'il tend à devenir le réseau universel de communications caractérisé par une offre multiservice, devient également la cible principale d'attaques et présente une sensibilité accrue à leurs impacts. Les attaques de déni de service (DdS) notamment connaissent depuis plusieurs années une forte recrudescence [MOO 01]. Elles sont fortement pénalisantes puisqu'elles se traduisent par une détérioration de la qualité de service (QdS) dont la gravité est imprévisible. Or, l'Internet est de plus en plus utilisé pour des applications nécessitant une QdS stable et garantie – par exemple, pour les applications de téléphonie sur IP. Les attaques de DdS sont ainsi susceptibles d'engendrer des pertes financières conséquentes. La détection de ces attaques constitue donc un enjeu de première importance. Aujourd'hui, la plupart des attaques de DdS sont distribuées pour être moins facilement détectables, i.e. de nombreuses sources collaborent de façon à se partager l'émission du trafic de l'attaque. Chaque source ne générant qu'un trafic d'attaque faible, il est difficile de les détecter près de leurs sources et elles n'ont d'ailleurs que peu d'impact sur la QdS des réseaux sources ou proches de la source. Dès que toutes les composantes de l'attaque se superposent près de la victime, l'effet est dramatique pour le réseau, ses équipements et ses utilisateurs. La variation de trafic devient énorme et les IDS (Intrusion Detection System) actuels détectent facilement l'attaque... Mais il est déjà trop tard ; la QdS du réseau est dégradée et l'attaque, par conséquent, un succès.

A l'heure actuelle, les IDS restent donc peu performants face aux dénis de service distribués et sont insensibles aux attaques dont les intensités sont trop faibles. En effet, le plus souvent, leur fonctionnement repose sur l'utilisation de signatures [PAX 99] ou profils de trafic construits sur des statistiques trop peu descriptives (moyenne et écart-type). La très forte variabilité naturelle du trafic [PAR 96] produit alors une forte fluctuation de ces statistiques et ainsi de très fort taux de faux positifs (fausses alarmes) et de faux négatifs (détection manquée). C'est le reproche majeur fait aux IDS actuels [MOO 01, BRU 00, HOC 93, JAV 91, VAC 89].

Un ensemble de travaux plus récents s'efforcent de prendre en compte une forme plus riche de la structure statistique du trafic (corrélation, densité spectrale, ...) (voir e.g., [YE 00, LAK 04, HUS 03, BAR 02, JUN 02]). Le travail rapporté dans cet article s'inscrit dans cette perspective. Il repose, en effet, sur une modélisation conjointe, à plusieurs niveaux d'agrégation simultanément, des distributions marginales du trafic par des lois non gaussiennes : précisément des lois gammas, $\Gamma_{\alpha,\beta}$ (voir par exemple [EVA 00]). L'originalité de l'approche proposée réside dans sa nature *multirésolution* (plusieurs niveaux d'agrégation Δ sont analysés conjointement), qui fournit une statistique robuste (l'évolution des paramètres α et β de ces lois en fonction de Δ), prenant finement en compte la structure de corrélation (court-terme) présente dans le trafic agrégé. Cette modélisation est décrite en section 3.

Le principe de la détection proposée consiste à rechercher une rupture dans le suivi des variations au cours du temps des valeurs prises par les paramètres correspondant à cette modélisation. Pour ce faire, des distances [BAS 89] entre la statistique estimée

dans une fenêtre d'observation courante et celle calculée sur une situation de référence sont seuillées. Ces procédures sont détaillées en section 4.

La difficulté de la validation de détecteurs d'anomalies réside dans l'établissement des qualités de leurs performances statistiques. Les individus perpétrants les attaques en informent en effet rarement les victimes a priori, de sorte qu'il est difficile de disposer d'un ensemble de traces, contenant des attaques de types et caractéristiques connues, pouvant servir de base de données pour l'étalonnage des procédures de détection. Pour pallier cette difficulté, nous avons choisi de réaliser nous-mêmes un ensemble d'attaques de dénis de service distribuées dont nous faisons varier les paramètres de manière **contrôlée** et **reproductible**. La réalisation de ces attaques est décrite dans la section 2. A partir de cette base de données que nous avons constituée, nous établissons les performances statistiques (probabilité de détections correctes versus probabilité de fausses alarmes) des procédures de détection proposées. Ces résultats sont rapportés dans la dernière partie de la section 4. Bien qu'elle puisse paraître artificielle ou simplifiée, cette méthodologie de production d'une base de donnée nous semble indispensable à l'étude, la mise au point et la validation des mécanismes de détection d'attaques.

2. Attaques par dénis de service distribuées

Description. Les attaques présentées dans cet article consistent en des dénis de service (DdS) distribués, réalisés à l'aide d'UDP flooding. Nous avons pour cela, utilisé le logiciel IPERF [IPE] (sous environnement Linux standard) qui nous permet de générer des flux (UDP) de débits variables. Ces attaques ont été produites à partir de plusieurs sites répartis sur la France (l'IUT de Mont de Marsan, l'ENS Lyon, l'ESSI et le LIP6) et à destination d'une cible unique (une machine Linux standard située sur le réseau du LAAS-CNRS) via le réseau RENATER. Le LAAS est connecté à RENATER par l'intermédiaire d'un lien Ethernet 100 Mbps qui n'a pas été saturé pendant les attaques. Ces attaques ont été réalisées de façon artificielle et contrôlée, de manière à pouvoir en faire varier à dessein les caractéristiques et paramètres (durée, intensité du flux de déni de service, taille et fréquence d'émission des paquets) afin de créer différents profils d'attaques à détecter. L'ensemble des configurations que nous avons choisies est détaillé dans le tableau 1. Chaque configuration a donné lieu à une capture de trafic avant, pendant et après les attaques, de façon à encadrer convenablement la période de DdS par deux périodes de trafic « normal ». Les traces ont une durée d'environ 90 minutes, les attaques sont principalement situées dans le deuxième tiers-temps.

Impacts. L'impact des attaques sur le débit global du lien analysé est très variable en fonction de leurs paramètres. La figure 1 présente les débits en octets et paquets de plusieurs attaques (attaques I, III et G) pendant toute la durée de la capture. Il est intéressant de noter que certaines attaques (attaques III et G) ont un impact très important sur le profil du trafic global, alors que d'autres (attaque I par exemple), au contraire, sont totalement noyées dans le trafic global et donc quasiment invisibles.

Label	Début trace	T (s)	Début attaque	T_A (s)	D (Mbps)	V (octets)	Int (%)
I	06-15(09 :54)	5400	10 :22	1800	0.25	1500	17.06
II	06-15(14 :00)	5400	14 :29	1800	0.5	1500	14.83
III	06-15(16 :00)	5400	16 :29	1800	0.75	1500	21.51
IV	06-16(10 :09)	5400	10 :16	2500	1.0	1500	33.29
V	06-17(10 :00)	5400	10 :28	1800	1.25	1500	39.26
A	06-17(14 :00)	5400	14 :28	1800	1	1000	34.94
B	06-17(16 :00)	5400	16 :28	1800	1	500	40.39
C	06-20(10 :03)	5400	10 :28	1800	1	250	36.93
G	06-21(14 :00)	5400	14 :28	1800	5	1500	58.02

Tableau 1. Description des attaques. *Paramètres des attaques de DdS distribuées réalisées (en 2005). T et T_A indiquent les durées des traces et des attaques, en secondes, D le débit (en Mbps) de chacune des sources (contrôlé en fixant la durée entre paquets) et V la taille (en octets) de chaque paquet de l'attaque, Int représente l'intensité relative de l'attaque (i.e. le ratio entre la somme des débits des flux d'attaques reçus et le débit moyen mesuré sur le lien du LAAS pendant l'attaque ; Débit (*flux attaquants*) / Débit (*global*)*100 pendant l'attaque).*

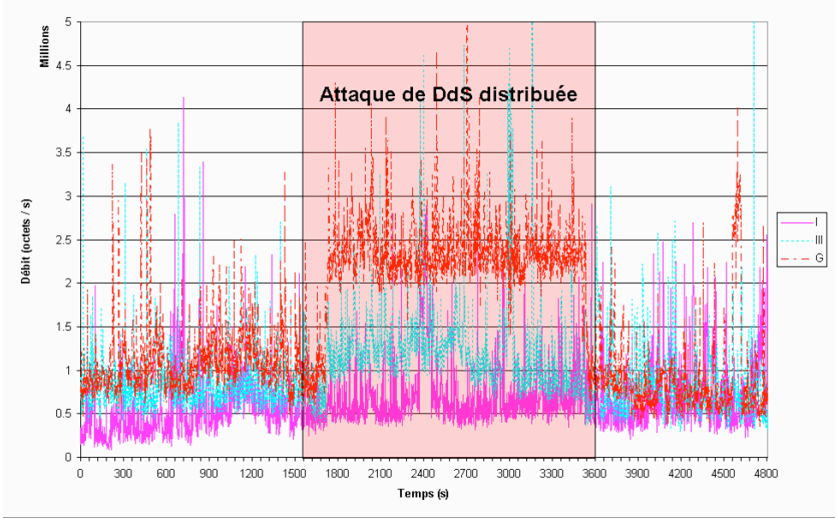

Figure 1. Trafic collecté pendant les attaques de DdS (débit en octet/s). *Attaques I, III et G. On montre ici 1h20 de chaque trace afin de recaler les périodes d'attaque dans la même zone. Les attaques se produisent pendant le second tiers-temps.*

L'objectif des procédures de détection est précisément de pouvoir détecter les attaques, y compris celles dont l'intensité est la plus faible, avant qu'elles n'aient un impact négatif voire dramatique sur la QdS du réseau.

3. Modélisation du trafic

Marginales non gaussiennes. Dans ce travail, nous étudions les séries temporelles $X_\Delta(k), k \in \mathbb{Z}$, constituées par le nombre de paquets agrégés dans des boîtes de taille Δ autour du temps $k\Delta$. Un travail équivalent pourrait être conduit sur les volumes

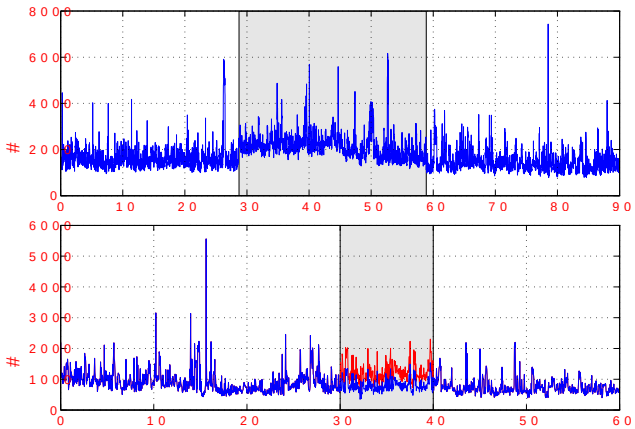

Figure 2. Trafics agrégés. *Séries temporelles de trafic agrégé, $\Delta_0 = 1$ms, temps en minutes. En haut : attaque III, se produisant entre les instants $t = 29{,}15$ min et $t = 59{,}39$ min. En bas : trafic normal ; on représente aussi ce trafic sur lequel on ajoute une anomalie par multiplication par 2 entre les instants $t = 30$ min et $t = 40$ min.*

agrégés. Des exemples de série temporelle de trafics agrégés contenant ou non une attaque ou une anomalie sont présentés sur la figure 2.

Nous proposons de modéliser la densité de probabilité (c'est à dire sa loi marginale) de X_Δ (pour chaque niveau d'agrégation indépendamment) par une distribution non gaussienne : la loi $\Gamma_{\alpha,\beta}$. Celle-ci est définie pour des variables aléatoires positives X par (où $\Gamma(u)$ est la fonction Gamma standard (voir [EVA 00])) :

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

Les lois $\Gamma_{\alpha,\beta}$ sont caractérisées par deux facteurs : la forme α et l'échelle β. Elles sont stables par multiplication, i.e. si X est $\Gamma_{\alpha,\beta}$, alors λX est $\Gamma_{\alpha,\lambda\beta}$, β consiste donc essentiellement en un facteur multiplicatif. Elles sont également stables sous addition, si X et X' sont $\Gamma_{\alpha,\beta}$ et $\Gamma_{\alpha',\beta}$ et indépendantes l'une de l'autre, alors $X + X'$ est $\Gamma_{\alpha+\alpha',\beta}$. Le facteur α rend compte de la forme de la distribution ; quand α est proche de 1, la distribution est exponentielle. Quand α est grand, elle s'approche d'une loi gaussienne. Le paramètre $1/\alpha$ peut donc être envisagé comme une mesure de l'écart à la gaussienne.

Cette stabilité sous addition s'avère particulièrement intéressante vis-à-vis du processus d'agrégation. En effet, puisque $X_{2\Delta}(k) = X_\Delta(2k) + X_\Delta(2k+1)$, elle nous indique que les lois $\Gamma_{\alpha,\beta}$ constituent une famille permettant de décrire les marginales de X_Δ pour les petits comme pour les grands niveaux d'agrégation Δ. La pertinence des lois $\Gamma_{\alpha,\beta}$ pour décrire le trafic agrégé sur une large gamme de niveaux d'agrégation, 2 ms $\leq \Delta \leq$ 500 ms, est illustrée sur la figure 3. Cette figure montre de plus que ce modèle reste pertinent non seulement pour du trafic normal, mais également pour du trafic *anormal*, dans ce cas subissant une attaque de DdSD.

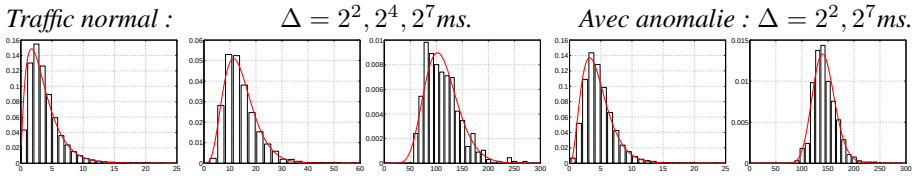

Figure 3. Marginales du trafic et Lois Gamma. *Distributions marginales du trafic agrégés à différents niveaux différents et ajustées par des lois $\Gamma_{\alpha,\beta}$. À gauche : trafic normal ($\Delta = 2^2, 2^4, 2^7 ms$). À droite : trafic pendant l'attaque III ($\Delta = 2^2, 2^7 ms$).*

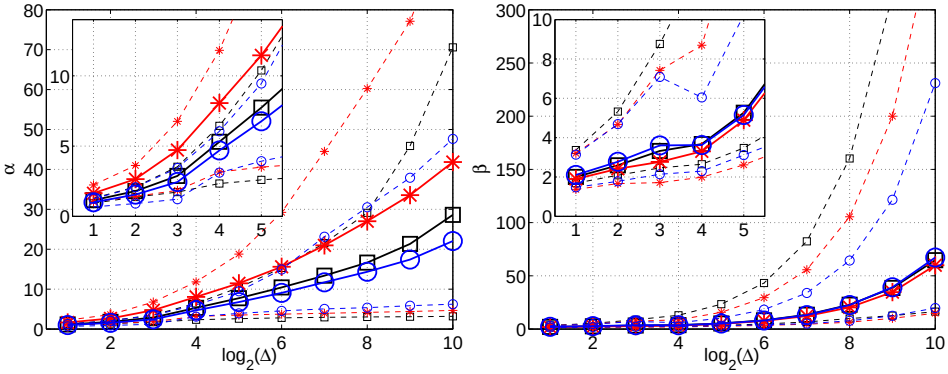

Figure 4. Evolution des paramètres α et β en fonction du niveau d'agrégation Δ. *A gauche : α_Δ ; à droite : β_Δ, estimés sur des tranches d'une minute, avant (cercles bleus), pendant (astérisques rouges) et après (carrés noirs) l'attaque. En trait plein : la moyenne ; en trait pointillé : les valeurs extrêmes de chaque période. L'évolution de α_Δ diffère notablement pendant l'attaque de celles avant et après l'attaque. C'est cette différence d'évolution qui donne le principe de la détection.*

Une version plus complète de ce modèle qui inclut conjointement les modélisations des statistiques des premier (marginale non gaussienne $\Gamma_{\alpha,\beta}$) et second ordres (covariance à longue mémoire type FARIMA) a été proposé dans [SCH 06] et détaillée mathématiquement dans [SCH 05].

Evolution des paramètres en fonction du niveau d'agrégation. Nous nous intéressons aux valeurs prises par les paramètres α et β en fonction du niveau d'agrégation Δ. En effet, si les $X_\Delta(k)$ consistaient en variables aléatoires indépendantes, la stabilité sous addition induirait simplement les évolutions $\alpha_\Delta = \alpha_0 \Delta$ et $\beta_\Delta = \beta$. Tout écart à ces évolutions rend donc compte d'une structure de dépendance (à court-terme) entre les $X_\Delta(k)$. Nous séparons les données en fenêtres adjacentes disjointes de taille T. Pour chacune, et pour différents niveaux d'agrégation Δ, nous estimons indépendamment les paramètres α et β. Cette estimation est réalisée à partir d'estimateurs dits *à maximum de vraisemblance*, décrits dans [EVA 00]. Nous obtenons donc des courbes $\hat{\alpha}_\Delta(l)$ et $\hat{\beta}_\Delta(l)$, où l'indice l rend compte de la position en temps, lT, de la $l-$ème fenêtre. Celles-ci sont représentées sur la figure 4 avec les choix $T = 1$min, $\Delta = 2^1, \ldots, 2^{10}$ ms ; nous traçons principalement la moyenne sur les différents l des

courbes, $\hat{\alpha}_\Delta$ et $\hat{\beta}_\Delta$. Sur ces courbes, nous observons d'abord que les courbes $\hat{\alpha}_\Delta$ et $\hat{\beta}_\Delta$ différent notablement de celles prévues par la situation d'indépendance indiquant ainsi la présence d'une forte structure de corrélation à temps court dans les $X_\Delta(k)$. Nous notons ensuite que ces évolutions en fonction de Δ sont significativement différentes pour des trafic normaux (carrés noirs et cercles bleus) et sous attaque (astérisques rouges) (les traits pointillés rendent compte de la dispersion des estimées obtenues à partir de blocs de durée $T = 1$min). On observe que l'occurrence de l'attaque induit une augmentation nette dans la croissance de α en fonction de Δ (cf. figure 4, à gauche). Ce changement indique que l'évolution de la loi $\Gamma_{\alpha,\beta}$ vers une loi normale sous l'effet de l'agrégation (c'est-à-dire de la loi des grands nombres) se trouve considérablement accélérée par une attaque de DdSD, qui donc change fortement la structure de corrélation à court terme d'arrivée des paquets (cette interprétation est complétée dans [SCH 06]). C'est dans l'exploitation de ces différences d'évolution de α et β en fonction du niveau d'agrégation que vont résider les détecteurs de DdSD présentés dans la section suivante.

4. Détection d'attaques de DdSD

Principes. La caractéristique fondamentale des détecteurs que nous proposons réside dans leur nature multirésolution : plusieurs niveaux d'agrégation Δ sont utilisés conjointement. La détection repose sur le principe suivant : on repère les dépassements de seuil d'une distance calculée entre des fenêtres adjacentes disjointes de durée T et indicée par l et une fenêtre de référence, choisie a priori, de taille T_{Ref}, et correspondant à une portion de trafic pouvant être considérée comme normale. Ici, $T_{Ref} = 10$ min, $T = 1$ min. L'élaboration de cette procédure requiert donc plusieurs choix : celui de la référence, celui de la distance et celui du seuil. Dans tout le travail présenté ici, cette référence est constituée par les T_{Ref} premières minutes de chaque trace. Il existe une large gamme de distances ou divergences pouvant être utilisées, dont une revue particulièrement exhaustive peut être consultée, par exemple, dans [BAS 89]. Trois distances sont considérées ici : les distances quadratiques pour les fonctions α_Δ et β_Δ, les divergences de Kullback des distributions marginales monodimensionnelles de X_Δ pour différents Δ et les divergences de Kullback des distributions bidimensionnelles de $(X_\Delta, X_{\Delta'})$ pour différents couples $(\Delta, \Delta' \neq \Delta)$. Ici, $\Delta = 2^1, 2^2, \ldots, 2^9, 2^{10}$ ms. Plusieurs valeurs de seuils seront examinées.

Distance quadratique moyenne. A partir des estimées $\hat{\alpha}(\Delta)$ et $\hat{\beta}(\Delta)$, nous définissons les distances quadratiques moyennes (DQM) :

$$D_\alpha(l) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\alpha}_{2^j}(l) - \hat{\alpha}_{2^j}(ref)\right)^2, \quad D_\beta(l) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\beta}_{2^j}(l) - \hat{\beta}_{2^j}(ref)\right)^2 \quad (2)$$

Les distances estimées correspondantes sont illustrées sur les figures 5, pour une trace avec attaque (Attaque III), et 6 pour un trafic normal. Ces figures confirment que l'évolution en fonction des niveaux d'agrégation du paramètre de forme, α_Δ, est notablement modifiée par l'occurrence de l'attaque, conduisant ainsi à des distances anor-

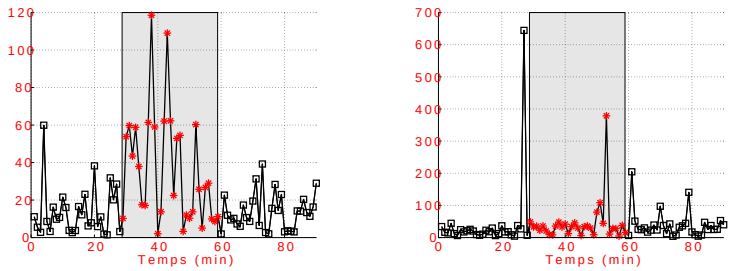

Figure 5. Distance quadratique moyenne (DQM). *DQM calculée sur α_Δ à droite, et β_Δ à gauche, sur des fenêtres successives (d'une minute). Les fenêtres correspondant à l'attaque (III) sont marquées par des astérisques (rouges), celles correspondants au trafic normal par des carrés (noirs).*

malement élevées, qui permettent sa détection (figure 5, à gauche). A l'inverse, le paramètre β est principalement sensible à une multiplication en volume du trafic. Or, on observe que $D_\beta(l)$ évolue peu pendant l'attaque (figure 5, à droite), indiquant que celle-ci ne se matérialise pas par une forte augmentation du trafic mais plutôt par une modification de sa structure statistique. Les fortes valeurs de $D_\beta(l)$ visibles sur la figure 5, à droite, correspondent à des fenêtres d'observation contenant la phase de transition (le démarrage de l'attaque - on note, en effet, que ces grandes valeurs se produisent en début et en fin d'attaque). Le mélange des statistiques des trafics normaux et sous attaque conduit à une distribution bimodale dont les paramètres ne peuvent être correctement estimés, d'où des valeurs aberrantes. A contrario, la figure 6 illustre les mêmes distances mesurées sur un trafic normal. On constate, comme attendu, que ces distances restent faibles et ne conduiraient donc pas à la production de fausses alarmes. Cette même figure 6 illustre les distances obtenues à partir d'un trafic normal sur lequel a été pratiquée une multiplication du nombre de paquets, artificielle et a posteriori (ici entre les minutes 30 et 40). Par construction, $\hat{\beta}_\Delta(l)$ voit parfaitement l'anomalie due à la multiplication artificielle du trafic, alors qu'elle est transparente pour $\hat{\alpha}_\Delta(l)$ (voir la figure). Cependant, on observe alors que les distances $D_\alpha(l)$ et $D_\beta(l)$ restent quasiment stables car les dépendances en Δ des courbes α_Δ et β_Δ n'ont pas changées. Les DQM $D_\alpha(l)$ et $D_\beta(l)$ nous permettent donc de détecter des changements dans les statistiques du trafic qui correspondent à un changement de sa structure de corrélation (donc de sa dynamique) tout en restant insensibles à ceux dus à de simples effets multiplicatifs. L'ensemble d'outils, $\hat{\alpha}_\Delta(l)$, $\hat{\beta}_\Delta(l)$, $D_\alpha(l)$ et $D_\beta(l)$ permet donc de détecter toutes les anomalies des statistiques du trafic et de discriminer entre légitimes (augmentation en volume d'un même trafic) et illégitimes (attaques induisant un changement dans la structure de corrélation du trafic).

Divergence de Kullback. Une autre mesure de proximité est donnée par la divergence de Kullback. Elle ne s'applique plus aux fonctions α_Δ ou β_Δ, mais aux densités de probabilité du trafic agrégé. Soient p_1 et p_2 deux densités de probabilités, la divergence de Kullback (DK) [BAS 89] se définit par :

$$DK(p1, p2) = \int (p_1(x) - p_2(x))(\ln p_1(x) - \ln p_2(x))dx. \quad (3)$$

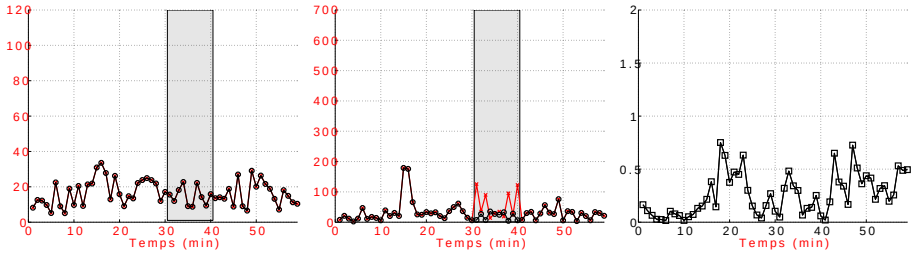

Figure 6. Trafic normal et trafic contenant une multiplication arbitraire. *De gauche à droite : DQM $D_\alpha(l)$, DQM $D_\beta(l)$ et divergence de Kullback, $K^{(1D)}_{2^4}(l)$, estimée sur des tranches successives d'une minute, pour la trace de référence (cercles noirs). On observe ici les fluctuations statistiques normales des distances, qui sont donc d'un niveau inférieur à celui que prennent les distances en présence d'une attaque. Pour les DQM, on a superposé (astérisques rouges) aux graphes, la DQM obtenue pour la trace contenant la multiplication arbitraire (astérisques quasi superposées aux cercles noirs, dans la zone grisée), entre les minutes 30 et 40. Ces deux distances ne sont presque pas sensibles aux simples augmentations en volume.*

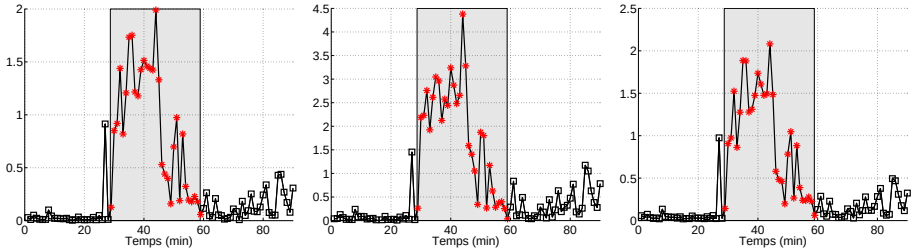

Figure 7. Divergences de Kullback. *A gauche, $K^{(1D)}_{2^4}(l)$, au milieu $K^{(1D)}_{2^7}(l)$, à droite, $K^{(2D)}_{2^4,2^7}(l)$, estimées sur des tranches successives d'une minute. On constate que l'attaque est très correctement mise en évidence.*

Suivant cette définition, nous calculons pour chaque niveau d'agrégation Δ, d'une part la divergence de Kullback entre les densités de probabilité de X_Δ (notées $p_{\Delta,l}$) de la fenêtre courante, et $p_{\Delta,Ref}$ de la fenêtre de référence ; d'autre part, pour chaque couple de niveaux d'agrégation $(\Delta, \Delta' \neq \Delta)$, nous calculons la divergence de Kullback entre les densités de probabilité conjointes de $(X_\Delta, X_{\Delta'})$, $p_{\Delta,\Delta',l}$, de la fenêtre courante, et $p_{\Delta,\Delta',Ref}$ de la fenêtre de référence :

$$K^{(1D)}_{\Delta}(l) = DK(p_{\Delta,l}, p_{\Delta,Ref}), \quad K^{(2D)}_{\Delta,\Delta'}(l) = DK(p_{\Delta,\Delta',l}, p_{\Delta,\Delta',Ref}). \qquad (4)$$

Les divergences de Kullback mesurées sur l'attaque III sont représentées sur la figure 7. Ces graphiques montrent que ces distances *voient* parfaitement les attaques et permettent donc de les détecter.

Courbes de Performances. Pour qualifier les performances statistiques des procédures de détection proposées, nous traçons les courbes dites de Caractéristiques Opérationnelles des Récepteurs (COR), qui représentent les probabilités de détection correcte en fonction des probabilités de fausses alarmes, $P_D = f(P_F)$, ou encore

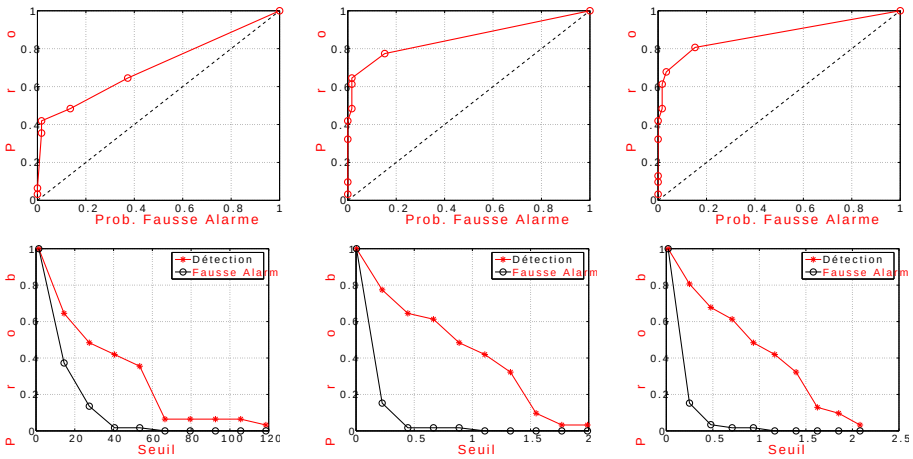

Figure 8. Courbes de performances. *En haut, probabilité de détections vs probabilité de fausses alarmes, $P_D = f(P_F)$, en bas, $P_D = f(\lambda)$ et $P_F = f(\lambda)$, pour $D_\alpha(l)$ (gauche), $K_{2^4}^{(1D)}(l)$ (milieu), $K_{2^4,2^7}^{(2D)}(l)$ (droite).*

$P_D = f(\lambda)$ et $P_F = f(\lambda)$, où λ est le seuil de détection sur la distance choisie. Pour obtenir ces courbes, nous procédons de la façon suivante. Nous découpons les traces de trafic collectées pendant les attaques entre fenêtres successives adjacentes de durée $T = 1$ min. Connaissant les caractéristiques de l'attaque, nous savons quelles fenêtres correspondent à l'attaque et au trafic normal. En faisant varier la valeur d'un seuil sur les fonctions distances $D_\alpha(l), D_\beta(l), K_\Delta^{(1D)}(l), K_{\Delta,\Delta'}^{(2D)}(l)$, nous comptons le nombre de fenêtres correspondants à l'attaque dont la distance dépasse le seuil et en déduisons la probabilité de détection ; inversement, on compte le nombre de dépassements de seuil pour les fenêtres correspondant au trafic normal, on obtient la probabilité de fausse alarme.

Les courbes COR, obtenues avec les paramètres : $T_{Ref} = 10$ min, $T = 1$ min, $\Delta = 2^1, 2^2, \ldots, 2^9, 2^{10}$ ms, sont illustrées sur la figure 8 pour l'attaque III et les trois distances choisies. Ces graphiques montrent que les courbes $P_D = f(P_F)$ sont localisées clairement près du coin supérieur gauche (le point de fonctionnement idéal aurait pour coordonnées $(0, 1)$) indiquant leurs bonnes performances. On note que $K^{(2D)}$ qui fait explicitement appel à deux niveaux d'agrégation conjointement présente des performances supérieures à celles obtenues avec $K^{(1D)}$. Sur cet exemple, les performances obtenues avec $D_\alpha(l)$ sont moins bonnes. Cependant, cette dernière distance présente des propriétés de robustesse et un intéressant pouvoir de classification entre anomalies légitimes ou non. Les courbes de la figure 8 donnent une indication sur les seuils à utiliser : vers 0,5 pour les divergence de Kullback, vers 20 pour D_α. Cependant les seuils obtenus ainsi varient entre les réalisations et il reste à établir une méthode automatique de sélection des seuils. Le tableau 2 résume, pour toutes les attaques réalisées, les probabilités de détection par rapport aux probabilités de fausses alarmes fixées à 10% et 20%. Ces tableaux sont obtenus en lisant sur les courbes COR

Méthode	#A	#B	#C	#G	#I	#II	#III	#IV	#V
DQM D_α	21	81	52	93	51	48	48	33	18
DK $K_{2^4}^{(1D)}$	50	78	91	93	25	35	74	56	87
DK $K_{2^7}^{(1D)}$	37	9	91	93	35	35	70	56	34
DK $K_{2^4,2^7}^{(2D)}$	53	78	91	93	25	35	74	45	90
DQM D_α	50	87	58	96	64	54	58	50	40
DK $K_{2^4}^{(1D)}$	78	78	91	93	64	58	93	67	96
DK $K_{2^7}^{(1D)}$	59	33	91	93	67	61	83	69	93
DK $K_{2^4,2^7}^{(2D)}$	81	81	91	93	51	61	93	66	96

Tableau 2. Performances des détections. *Pour chaque distance, pour chaque attaque, probabilités de détection obtenue pour une probabilité de fausse alarme fixée à* 10% *(en haut) ou* 20% *(en bas) pour les 4 distances pertinentes.*

les valeurs de probabilités de détection obtenues pour la probabilité de fausse alarme fixée a priori. Ces chiffres montrent dans tous les cas des niveaux de détection très satisfaisants. C'est en particulier vrai pour les attaques d'intensités faibles (comme les attaques A, B, I, II), donc ayant peu d'impacts sur le profil en volume du trafic. En effet, les IDS traditionnels ne détectent en général pas les attaques dans ces conditions. La méthode proposée offre ainsi une possibilité de détection des DdSD, même de faible intensité, comme c'est par exemple le cas sur un nœud encore loin de la cible finale. Les scores de détection obtenus sont donc encourageants.

5. Conclusions et perspectives

Dans ce travail, nous avons proposé des procédures de détection d'attaques de DdSD. Celles-ci sont intrinsèquement multirésolution (elles reposent sur les analyses conjointes du trafic agrégé à plusieurs niveaux) et consistent à seuiller des distances calculées entre une fenêtre d'observation courante et une référence. Nous avons également réalisé une campagne de DdS dont nous avons pu faire varier les intensités et caractéristiques de manière contrôlée. A partir du trafic collecté avant, pendant et après ces attaques, nous avons pu déterminer des performances statistiques de ces procédures de détection ; celles-ci sont encourageantes.

Ce travail sera poursuivi, d'une part, par la réalisation de nouvelles campagnes d'attaques mettant en jeu d'autres intensités, protocoles, caractéristiques ou mécanismes et d'autre part, par l'usage d'autres distances, exploitation renforcée de l'aspect multirésolution. Un aspect important à développer est la détermination automatique des seuils, de manière à le à fixer sans supervision constante par un opérateur pour une probabilité de fausse alarme donnée : l'usage de technique dite de *bootstrap* est à l'étude pour ce faire. Enfin, on explorera la possibilité de descendre la durée des fenêtre de détection en deçà de la minute, celle-ci faisant l'objet d'une dégradation du compromis détection correcte/fausse alarme.

6. Bibliographie

[BAR 02] BARFORD P., KLINE J., PLONKA D., RON A., « A signal analysis of network traffic anomalies », *ACM/SIGCOMM Internet Measurement Workshop*, Marseille, France, novembre 2002.

[BAS 89] BASSEVILLE M., « Distance measures for signal processing and pattern recognition », *Signal Processing*, vol. 18, 1989, p. 349–369.

[BRU 00] BRUTLAG J., « Aberrant behavior detection in time series for network monitoring », *USENIX System Administration Conference*, New Orleans, décembre 2000.

[EVA 00] EVANS M., HASTINGS N., PEACOCK B., *Statistical Distributions*, Wiley (Interscience Division), juin 2000.

[HOC 93] HOCHBERG J., JACKSON K., STALLINGS C., MCCLARY J., DUBOIS D., FORD J., « NADIR : an automated system for detecting network intrusion and misuse », *Journal of Computer Security*, vol. 12, n° 3, 1993, p. 235–248.

[HUS 03] HUSSAIN A., HEIDEMANN J., PAPADOPOULOS C., « A framework for classifying denial of service attacks », *SIGCOMM*, Karlsruhe, Germany, 2003.

[IPE] IPERF - THE TCP/UDP BANDWIDTH MEASUREMENT TOOL, http ://dast.nlanr.net/Projects/Iperf/.

[JAV 91] JAVITS, VALDES, « The SRI IDES Statistical Anomaly Detector », *ESORICS*, , 1991.

[JUN 02] JUNG J., KRISHNAMURTHY B., RABINOVICH M., « Flash Crowds and Denial of Service Attacks : Characterization and Implications for CDNs and Web Sites », *International WWW Conference*, Honolulu, HI, mai 2002.

[LAK 04] LAKHINA A., CROVELLA M., DIOT C., « Diagnosing Network-Wide Traffic Anomalies », *SIGCOMM*, août 2004.

[MOO 01] MOORE D., VOELKER G., SAVAGE S., « Inferring Internet Denial-of-Service activity », *Usenix Security Symposium*, 2001.

[PAR 96] PARK K., KIM G., CROVELLA M., « On the relationship between file sizes, transport protocols, and self-similar network traffic », *International Conference on Network Protocols*, Washington, DC, USA, 1996, IEEE Computer Society, page 171.

[PAX 99] PAXSON V., « Bro : a system for detecting network intruders in real-time », *Computer Networks Journal*, vol. 31, n° 23–24, 1999, p. 2435–2463.

[SCH 05] SCHERRER A., ABRY P., « Marginales non gaussiennes et longue mémoire : analyse et synthèse de trafic Internet », *Colloque GRETSI-2005*, Louvain-la-Neuve, Belgique, septembre 2005.

[SCH 06] SCHERRER A., LARRIEU N., BORGNAT P., OWEZARSKI P., ABRY P., « Non Gaussian and Long Memory Statistical Modeling of Internet Traffic », *4th Workshop IPS-MoMe*, Salzburg (Austria), février 2006, Salzburg Research, p. 176–185.

[VAC 89] VACCARO H., LIEPINS G., « Detection of Anomalous Computer Session Activity », *IEEE Symposium on Security and Privacy*, Oakland, California, mai 1989, p. 280–289.

[YE 00] YE N., « A Markov chain model of temporal behavior for anomaly detection », *Workshop on Information Assurance and Security*, West Point, NY, juin 2000.