Explication et réduction de l'impact des violations d'inégalités triangulaires dans Vivaldi

François Cantin* — **Bamba Gueye*** — **Mohamed Ali Kaafar*** — **Guy Leduc*** — **Laurent Mathy****

* *Université de Liège*
Institut Montefiore B 28, B-4000, Liège Belgique
{francois.cantin, cabgueye, ma.kaafar, guy.leduc}@ulg.ac.be

** *University of Lancaster*
Lancaster, UK LA1 4WA
laurent@comp.lancs.ac.uk

RÉSUMÉ. *Les systèmes de coordonnées sont des systèmes distribués ayant pour but, à partir de mesures de distance (par exemple RTT) entre certaines paires de nœuds, d'associer des coordonnées à chaque nœud dans un espace métrique. Toutefois, de tels systèmes ne fonctionnent pas correctement lorsque les distances mesurées ne respectent pas les inégalités triangulaires. Or, les violations de ces inégalités, appelées TIV, sont fréquentes dans l'Internet. Nous proposons une étude approfondie de l'impact des TIV sur le système de coordonnées Vivaldi. Nous quantifions et expliquons les erreurs de prédiction de distance et l'instabilité des coordonnées causées par les TIV selon leur fréquence et leur sévérité. Nous montrons aussi que la distance entre deux nœuds, mesurée par le RTT, est corrélée à la probabilité d'existence d'une TIV. Enfin, nous recommandons un système de coordonnées hiérarchique, et nous montrons, par des simulations sur une matrice de délais réelle, qu'une telle approche réduit l'impact des TIV.*

ABSTRACT. *Network coordinate systems embed delay measurements (e.g. RTT) between Internet nodes into some metric space. These systems often assume the triangle inequality holds for Internet delays. However, the reality is that the triangle inequality is violated by Internet delays. In a first step, we explore the ways in which TIVs impact the Vivaldi coordinate system using different metrics in order to quantify various levels of TIV's severity. In a second step, we study TIVs existing in the Internet. Our results show that path lengths do have an effect on the impact of these TIVs. In particular, we observed correlation between the (in)stability and high effective error of nodes' coordinates with respect to their involvement in TIVs situations. Finally, we propose a Two-Tier architecture that does mitigate the effect of TIVs on the distance predictions.*

MOTS-CLÉS : *Systèmes de coordonnées, Performance, Violation de l'inégalité triangulaire*

KEYWORDS: *Internet Coordinate Systems, Performance, Triangle Inequality Violations*

2e soumission à *CFIP'08*, le 08/02/2008.

1. Introduction

Nous avons assisté ces dernières années à une énorme croissance d'applications Internet [ROW 01, KUB 00, CHU 00, SKY], basées et/ou bénéficiant des réseaux de recouvrement (ou overlay). Ces applications tiennent compte de la topologie du réseau physique pour la construction du réseau de recouvrement. En particulier, la plupart de ces applications (si ce n'est toutes) et leur réseau de recouvrement associé, se basent sur la notion de proximité réseau, typiquement définie en termes de délais aller-retour (RTT), pour l'optimisation de la sélection de voisins. Cependant, les mesures de proximité peuvent s'avérer extrêmement coûteuses en termes de consommation en bande passante. En effet, l'existence simultanée de plusieurs réseaux de recouvrement, peut entraîner un surcoût de communications élevé, dû aux mesures de proximité individuelles menées par chaque nœud de ces réseaux.

De plus, traquer la proximité au sein d'un groupe dynamique, nécessite une fréquence de mesure très grande. Cela induit encore un surcoût de mesures plus important. Afin de pallier ce problème, des systèmes de positionnement "Internet", comme [NG 02, PIA 03, COS 04, SHA 03, DAB 04], ont été introduits. Dans ces systèmes, l'idée principale est que si chaque nœud peut être associé à des coordonnées virtuelles dans un espace approprié, la distance entre les nœuds est trivialement calculée sans pour autant avoir recours à des mesures directes. En d'autres termes, ces systèmes plongent les mesures des temps de latence (délais) entre nœuds dans un espace métrique et associent un vecteur de coordonnées (ou coordonnées) dans cet espace à chaque nœud, dans le but de permettre des prédictions de distances précises et peu onéreuses pour n'importe quelle paire de nœuds dans le réseau. L'avantage premier de ces systèmes est que si les distances réseaux (dans le sens de délais) sont plongées dans un espace de coordonnées où une position raisonnablement précise pour chaque nœud est établie, le surcoût de mesures produit par le positionnement, est amorti sur plusieurs prédictions de distances. Ceci réduit énormément le coût en termes de mesures de distances du système entier.

Cependant, les politiques de routages et l'existence de congestion sur certain chemins, peuvent entraîner des *violations du principe de l'inégalité triangulaire* ("Triangle Inequality Violation" - TIV) sur les délais dans l'Internet [ZHE 05] . Ces violations seraient la cause de distorsions et d'erreurs de prédiction pour les systèmes de coordonnées. Prenons l'exemple de trois nœuds A, B et C tels que $d(A,B)$ est de $1\ ms$, $d(B,C)$ est de $2\ ms$ et $d(A,C)$ est de $5\ ms$ où $d(X,Y)$ dénote le délai existant entre le nœud X et le nœud Y. Dans ce cas, le principe de l'inégalité triangulaire est violé car $d(A,C) > d(A,B) + d(B,C)$. Ainsi, l'immersion (ou "*embedding*") exacte des délais dans un espace métrique, où l'inégalité triangulaire doit être respectée, est forcément impossible. Face à des situations de TIV, les nœuds d'un système de coordonnées auront tendance à alterner entre des sous-estimations et des sur-estimations de la distance réelle, sans jamais parvenir à se positionner dans l'espace métrique d'une manière parfaite.

Exploiter les systèmes de coordonnées pour diverses opérations de prédiction de distances au niveau applicatif nécessite en revanche, que les coordonnées de ces systèmes soient aussi précises que stables.

Afin d'améliorer la précision des systèmes de coordonnées, certains travaux ont envisagé l'exclusion des nœuds qui forment des TIV entre eux [LED 07, Y.B 03]. Toutefois, sacrifier ne fusse qu'une infime partie des nœuds ne nous semble pas justifiable car les TIV sont une propriété naturelle inhérente à l'Internet. Les auteurs de [WAN 07] proposent d'éliminer parmi l'ensemble des voisins d'un nœud A tous les nœuds B_i telle que la distance $d(A, B_i)$ prédite à partir des coordonnées est fortement sous-estimée. Ainsi, la majorité des voisins de A sont choisis parmi les nœuds qui lui sont très proches. Ceci entraîne une sélection non hybride (*i.e.* proches et éloignés) des voisins, alors qu'une telle propriété est une condition nécessaire pour le bon fonctionnement des systèmes de coordonnées.

Dans cet article, nous étudions d'abord l'impact des TIV sur le système de coordonnées Vivaldi. Cette étude montre que les nœuds les plus impliqués dans les TIV sont en général deux fois moins précis comparés aux autres nœuds, quant à leur positionnement dans le système de coordonnées. Leurs coordonnées sont également moins stables, avec de fortes amplitudes d'oscillations. Dans une deuxième étape, nous analysons la distribution des TIV dans l'Internet et par la suite, caractérisons leur sévérité en considérant différentes métriques. Nous montrons que les chemins les plus longs sont ceux qui sont les plus impliqués dans des TIV sévères. Ainsi, nous proposons une structure hiérarchique, opposée à la structure plate de Vivaldi, se basant sur un regroupement de nœuds proches. Les nœuds appartenant à un même groupe calculent des coordonnées locales pour prédire les distances entre eux. En revanche, au niveau supérieur de la hiérarchie, nous considérons un Vivaldi "plat" afin de prédire les distances entre les nœuds de différents groupes. Cette approche hiérarchique limite l'impact des TIV sur les estimations de distances, et permet une meilleure prédiction pour les petits chemins.

2. Vivaldi et les violations de l'inégalité triangulaire

2.1. *Description de Vivaldi*

Vivaldi [DAB 04] est un système de coordonnées complètement décentralisé. Il est basé sur une simulation de ressorts, où la position du nœud correspond à celle de l'extrémité d'un ressort qui minimiserait l'énergie potentielle des ressorts, et donc minimiserait l'erreur de positionnement. Un nœud se joignant au système, calcule ses coordonnées en collectant des informations de positions et de mesures de délai à partir de quelques autres nœuds. Spécifiquement, Vivaldi place un ressort entre chaque paire de nœuds (i, j) dans le système, avec une longueur "au repos" correspondant à la mesure RTT entre ces deux nœuds. La longueur réelle du ressort est considérée comme étant l'estimation de distance entre les deux positions des nœuds. L'énergie potentielle d'un tel ressort est proportionnelle au carré du déplacement par rapport à sa longueur

au repos. La somme de ces erreurs à travers tous les ressorts est la fonction d'erreur que Vivaldi tente de minimiser. Une procédure identique tourne sur tous les nœuds Vivaldi. Celle-ci est basée sur des échantillons récoltés par les nœuds qui fournissent les informations pour leur positionnement. Un échantillon, utilisé par un nœud i est ainsi constitué de la mesure vers un nœud j, de la coordonnée du même nœud j, ainsi que de l'erreur locale reportée par j. L'algorithme traite les nœuds à haute erreur, en assignant des poids à chaque échantillon collecté. L'erreur relative de cet échantillon, e_e, est alors calculée comme suit :

$$e_e = \frac{|\parallel x_j - x_i \parallel - RTT(i,j)_{mesure}|}{RTT(i,j)_{mesure}}$$

où x_i et x_j représentent respectivement les coordonnées des nœuds i et j. Le nœud i met à jour son erreur locale comme suit : $e_i = e_e \times w_e + e_i \times (1 - w_e)$ où $w_e = e_i/(e_i + e_j)$ est le poids de l'échantillon calculé à partir de son erreur locale e_i et de l'erreur locale e_j du nœud j. Le nœud utilise aussi ce poids pour calculer la valeur $\delta_e = C_c \times w_e$, avec $0 < C_c < 1$, qui définit l'amplitude du déplacement du nœud dans la direction spécifiée par cet échantillon. A l'aide de δ_e, le nœud i met à jour ses coordonnées de la manière suivante :

$$x_i = x_i + \delta_e \cdot (RTT(i,j)_{mesure} - \parallel x_i - x_j \parallel) \cdot u(x_i - x_j)$$

où $u(x_i - x_j)$ est un vecteur unitaire indiquant la direction dans laquelle le noeud i doit se déplacer.

2.2. *Critères de sévérité des TIV*

Des études antérieures tels que [SAV 99, ZHE 05, LEE 06] ont caractérisé les TIV dans l'Internet en considérant le ratio et le pourcentage de triangles victimes de TIV. Soit un triangle ABC tel que, par convention, AB représente le côté le plus long de ce triangle. Si $d(A,B) > d(A,C) + d(C,B)$, alors ABC forme une TIV car le principe de l'inégalité triangulaire est violé. Nous proposons deux critères de sévérité des TIV. Le premier critère est la *sévérité relative* et est définie comme suit :

$$G_r = \frac{d(A,B) - (d(A,C) + d(C,B))}{d(A,B)} \quad (1)$$

La valeur de G_r varie entre 0 (sévérité minimale) et 1 (sévérité maximale) et permet de caractériser le gain relatif obtenu en termes de distance en empruntant le chemin indirect. Toutefois, lorsque la distance $d(A,B)$ est petite, la sévérité relative peut être grande sans pour autant que la TIV soit réellement sévère. Nous définissons alors un second critère dénomé *sévérité absolue*, comme suit :

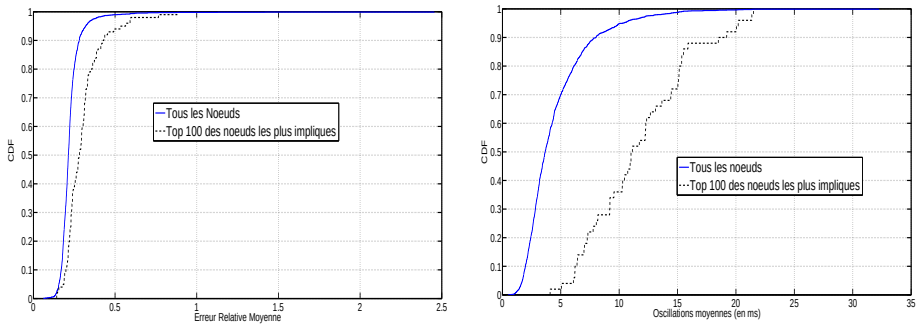
$$G_{ar} = \frac{d(A,B) - (d(A,C) + d(C,B))}{\mathcal{D}} \quad (2)$$

G_{ar} représente le gain absolu en termes de distance normalisée par le diamètre de l'espace dénoté $\mathcal{D}$. G_{ar} varie également entre 0 (sévérité minimale) et 1 (sévérité maximale). Par la suite, nous fixons différents seuils de sévérité th_r et th_{ar} et considérons toutes les TIV telles que $G_r \geq th_r$, ou $G_{ar} \geq th_{ar}$, ou bien celles supérieures aux deux seuils à la fois.

2.3. *Impact des TIV dans Vivaldi*

Pour nos simulations, nous avons utilisé le simulateur *p2psim* [P2P] qui fournit une implémentation de Vivaldi. Chaque nœud Vivaldi choisit 32 voisins et la moitié d'entre eux sont choisis parmi les nœuds les plus proches. La constante C_c (voir section 2.1) est fixée à 0,25 (valeur recommandée dans [DAB 04]). Le système est considéré comme "ayant convergé" ou "s'étant stabilisé" si toutes les erreurs relatives des nœuds convergent vers des valeurs qui varient au maximum de 0,02 pour 10 pas de simulation consécutifs. Toutes les simulations convergent en moins de 1800 pas (ce qui correspond à peu près à 8 heures, un pas représentant environ 17 secondes).

Pour nos simulations, nous avons considéré un espace euclidien à deux dimensions. Nous utilisons l'ensemble de données "King" [GUM 02] qui contient les RTT mesurés entre 1740 serveurs *DNS* [P2P]. Pour étudier l'impact des TIV sur la précision de la prédiction de Vivaldi, nous définissons le degré d'implication d'un nœud dans une TIV. Pour chaque nœud C, ce degré d'implication est le nombre de paires (A_i, B_j), $i \neq j$, existantes où A_iB_jC est un triangle violant l'inégalité triangulaire. Sur la base de ce degré d'implication, nous comparons la précision de la prédiction de distance des 100 nœuds les plus impliqués dans des TIV à celle de l'ensemble des nœuds du système.



(a) CDF de l'erreur relative moyenne (b) CDF des oscillations

Figure 1. *Impact des TIV sur la précision de la prédiction de Vivaldi.*

La figure 1(a) montre la CDF (*Cumulative Distribution Function*) de l'erreur relative moyenne des 100 nœuds les plus impliqués dans des TIV comparée à celle de l'ensemble des nœuds du système (S). L'*erreur relative moyenne* (ERM) pour un

nœud est calculée comme étant la moyenne des erreurs relatives de prédiction entre ce nœud et tous les autres nœuds du système lors des derniers pas de simulation.

$$ERM(i) = \frac{\sum_{j \neq i \in S} \frac{|(RTT(i,j) - ||x_i - x_j||)|}{RTT(i,j)}}{|S|}$$

La figure 1(a) montre que la prédiction de distance est moins précise pour les nœuds les plus impliqués dans des TIV. En effet, en considérant l'ensemble des nœuds, plus de 90 % d'entre eux ont une ERM inférieure à 0,3, alors que pour les 100 nœuds les plus impliqués dans des TIV le pourcentage de nœuds atteignant cette valeur est seulement de 50 %.

Nous avons également observé l'oscillation des coordonnées des nœuds dans le système Vivaldi. L'oscillation d'un nœud est définie comme étant la distance entre les coordonnées du nœud lors de deux pas successifs. Nous représentons l'oscillation moyenne d'un nœud comme étant la moyenne de ses oscillations lorsque l'on considère les 500 derniers pas de la simulation. La figure 1(b) montre la CDF des oscillations moyennes des 100 nœuds les plus impliqués dans des TIV comparée à celle de l'ensemble des nœuds du système. Nous remarquons une fois de plus que l'influence des TIV sur la prédiction de distance est assez importante puisque les nœuds les plus impliqués dans des TIV ont une oscillation moyenne beaucoup plus importante que celle des autres nœuds. 70 % des nœuds ont une oscillation inférieure à $5\ ms$ alors qu'uniquement 3 nœuds parmi les 100 les plus impliqués dans des TIV voient leur coordonnées osciller en moyenne de moins de $5\ ms$.

Nous venons de montrer que les TIV ont à la fois un impact considérable sur la précision du système de coordonnées Vivaldi, mais aussi sur la stabilité de ses coordonnées. Dans la section suivante, nous étudions la distribution des TIV existant dans l'Internet et caractérisons leur sévérité en utilisant les deux critères de sévérités présentés dans la section 2.2.

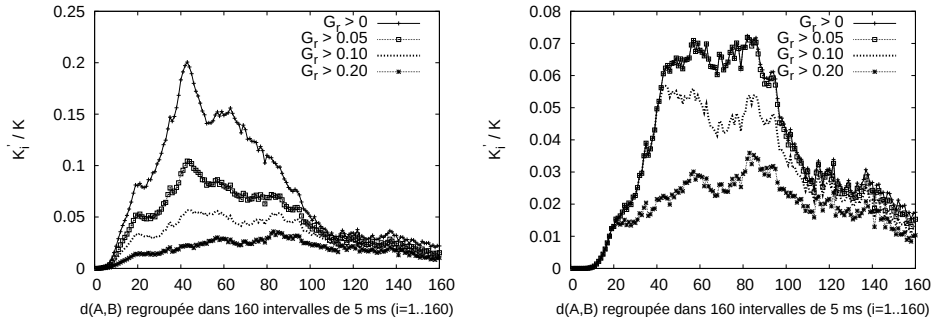
3. Analyse des violations de l'inégalité triangulaire dans l'Internet

Soit K le nombre total de triangles ABC obtenus en considérant l'ensemble des nœuds de la matrice de délai King. Nous obtenons K = 854 773 676 parmi lesquels 105 329 511 triangles sont victimes de TIV (12 % du nombre total de triangles). Les RTT mesurés entre les nœuds sont répartis dans 160 intervalles de $5\ ms$ chacun (les RTT variant de $5\ ms$ à $800\ ms$ dans la matrice de délai). Nous supposons que le triangle ABC est dans l'intervalle i ($0 \leq i < 160$), si d(A, B) (son côté le plus long) appartient à cet intervalle. Nous dénoterons par K_i le nombre de triangles appartenant à l'intervalle i et par K'_i le nombre de triangles qui forment des TIV à l'intérieur de l'intervalle i.

Soit P_i la probabilité qu'un triangle ABC choisi au hasard soit victime d'une TIV et appartienne à l'intervalle i. Cette probabilté est calculée comme suit :

$$P_i = \frac{K'_i}{K_i} \times \frac{K_i}{K} = \frac{K'_i}{K} \quad (3)$$

La figure 2 montre la distribution des P_i en considérant les deux critères de sévérité définis dans la section 2.2.



(a) th_{ar} = 0 (b) th_{ar} = 0,025

Figure 2. *Distribution des TIV pour différents seuils de sévérité.*

Sur la figure 2(a), nous ne considérons que la sévérité relative ($th_{ar} = 0$) alors que sur la figure 2(b), nous ne tenons pas compte des TIV dont la sévérité absolue est inférieure à th_{ar} = 0,025. La figure 2 montre qu'il est peu probable d'avoir une TIV sévère pour les triangles tels que $d(A, B) < 150\ ms$ ($i \leq 30$). Quel que soit le degré de sévérité des TIV, les triangles tels que $d(A, B) > 200\ ms$ ($i \geq 40$) ont plus de chance d'être des TIV sévères. Au vu de ces observations nous proposons une approche hiérarchique pour limiter l'impact des TIV. Ainsi, en définissant des groupes dont le diamètre est inférieur à $150\ ms$, il est peu probable que les TIV à l'intérieur des groupes soient sévères. Les coordonnées calculées à l'intérieur d'un groupe seraient dès lors plus précises de même que la prédiction des chemins à l'intérieur d'un groupe.

4. Vivaldi hiérarchique

Cette section est constituée de trois parties. Dans un premier temps, nous donnons un aperçu de l'approche hiérarchique de Vivaldi. Ensuite, nous décrivons l'algorithme de regroupement et les groupes utilisés lors de notre étude. Enfin, nous comparons les résultats de notre approche hiérarchique à ceux obtenus avec le Vivaldi "plat".

4.1. *Aperçu*

Nous proposons un Vivaldi hiérarchique dans lequel les nœuds sont regroupés en ensemble de nœuds proches. Chaque nœud choisit 32 voisins dont 16 sont à l'inté-

rieur de son cluster (voisins proches) et 16 autres en dehors. Les nœuds se trouvant dans chaque groupe calculent deux vecteurs de coordonnées : des coordonnées locales et des coordonnées globales. Les coordonnées locales sont mises à jour en réalisant des mesures avec les 16 voisins membres du même groupe. Elles sont utilisées uniquement pour la prédiction des chemins intra-groupe. Pour la mise à jour des coordonnées globales, chaque nœud considère ses 32 voisins. Celles-là sont utilisées pour la prédiction des chemins inter-groupe. Nous décrivons par la suite comment les clusters considérés dans notre étude ont été construits.

4.2. *Le processus de regroupement*

Dans un premier temps nous avons basé la construction des groupes sur les coordonnées fournies par un Vivaldi "plat" dans un espace euclidien à deux dimensions. Ces coordonnées $2D$, obtenues en appliquant Vivaldi sur la matrice de délais, permettent d'observer 5 principaux nuages de points (groupes) distincts. Pour cette étude, nous avons sélectionné les trois groupes les plus peuplés en fixant un diamètre maximum. Celui ci variant de $60\ ms$ à $140\ ms$ pour les 3 groupes. Les nœuds qui ne font pas partie d'un des 3 groupes participent uniquement au niveau supérieur du Vivaldi hiérarchique et ne calculeront donc que des coordonnées globales.

Toutefois, il est important de noter que les groupes tels que définis à ce niveau sont basés sur des estimations de RTT faites grâce aux coordonnées fournies par un Vivaldi "plat". Ainsi, il peut arriver que certains nœuds soient mal placés, et par conséquent engendrent un diamètre beaucoup plus grand que prévu pour les groupes. Dans une seconde étape, nous vérifions alors dans chaque groupe si les nœuds respectent les contraintes de délai fixées en tenant compte des mesures de RTT réelles. Pour chaque groupe, nous appliquons de manière récursive l'algorithme ci-dessous jusqu'à ce qu'aucune paire de nœuds appartenant au même groupe ne viole le diamètre maximum fixé :
1. Pour chaque nœud, nous vérifions si le RTT avec les autres nœuds du groupe ne dépasse pas le diamètre fixé ; chaque dépassement entraîne l'incrémentation d'un compteur pour ce nœud.
2. Nous retirons ensuite du groupe le nœud qui compte le plus de violations[1].

Le tableau 1 montre les caratéristiques des groupes obtenus.

Dans ce qui suit, nous comparons les performances obtenues en utilisant le Vivaldi hiérarchique et le Vivaldi "plat" et ce vis-à-vis de la précision de positionnement des nœuds et de l'amplitude moyenne des oscillations de leurs coordonnées.

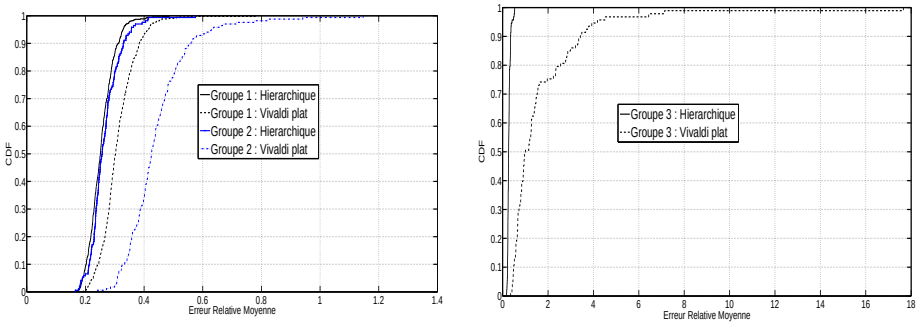
1. Ce nœud a très probablement une grande erreur relative occasionnant une erreur de positionnement importante.

	Nombre de nœuds	Diamètre
Groupe 1	565	140 ms
Groupe 2	169	100 ms
Groupe 3	93	60 ms

Tableau 1. *Caractéristiques des groupes*

4.3. *Evaluation des performances du Vivaldi hiérarchique*

Afin de caractériser la précision de positionnement des nœuds, nous utilisons l'erreur relative moyenne des nœuds (ERM - voir section 2.3) comme indicateur de performance. Les figures 3(a) et 3(b) montrent les CDF des erreurs relatives moyennes pour les différents groupes. Pour un groupe donné, seuls les nœuds de ce groupe sont considérés et les ERM sont calculées uniquement sur la base des erreurs de prédiction des chemins intra-groupe. Pour chaque groupe, nous avons une courbe représentant le Vivaldi hiérarchique, où les coordonnées locales des nœuds sont utilisées pour estimer les erreurs de prédiction, et une représentant le Vivaldi plat en considérant le même ensemble de nœuds.



(a) groupes 1 et 2 (b) groupe 3

Figure 3. *Comparaison des erreurs relatives moyennes.*

Nous remarquons clairement que les erreurs relatives moyennes calculées à l'aide des coordonnées locales sont plus faibles que celles calculées à l'aide des coordonnées du Vivaldi "plat". Par exemple, dans le groupe 2, plus de 90 % des nœuds ont une ERM inférieure à 0,3 en considérant les coordonnées locales alors qu'en considérant les coordonnées du Vivaldi plat, un peu moins de la moitié des nœuds du groupe ont une ERM supérieure à 0,5. C'est dans le groupe 3 que nous observons la plus grande différence entre le Vivaldi hiérarchique et le Vivaldi plat. En effet, sur la figure 3(b), la quasi totalité des nœuds du groupe ont une ERM inférieure à 0,5 en considérant les coordonnées locales. Par contre, 70 % de ces nœuds ont une très mauvaise erreur relative moyenne lorsque les coordonnées du Vivaldi "plat" sont utilisées.

Ces résultats montrent que la prédiction de distance faite à l'intérieur des groupes est nettement meilleure en considérant le Vivaldi hiérarchique. Nous pouvons en déduire que plus le diamètre du groupe est petit, plus la différence de précision entre le Vivaldi hiérarchique et le Vivaldi plat est importante. La longueur des chemins est corrélée à la sévérité des TIV.

La meilleure précision de positionnement obtenue à l'intérieur des groupes est due au fait que pour calculer leurs coordonnées locales, les nœuds choisissent tous leurs voisins à l'intérieur de leur groupe. Ceci a pour effet de limiter au diamètre du groupe le délai maximal mesurable entre un nœud et son voisin. Ainsi, la valeur $G_a = d(A,B) - (d(A,C) + d(C,B)$ est limitée au diamètre du groupe. Si on considère le Vivaldi plat, cette valeur varie jusqu'à $800\ ms$. Des TIV ayant une valeur de G_a assez grande risquent d'introduire des erreurs d'estimation absolues importantes. Ainsi, lors de la mise à jour de leurs coordonnées, des nœuds risquent d'être confrontés à de grandes valeurs pour $RTT(i,j)_{mesure} - \parallel x_i - x_j \parallel$ (*voir* section 2.1). De tels nœuds oscillent sur de plus grandes amplitudes qui engendrent une imprécision de l'estimation de distance. Par contre, si les TIV ont toutes des petits G_a, alors les nœuds oscillent sur de petits intervalles permettant ainsi une meilleure prédiction des délais.

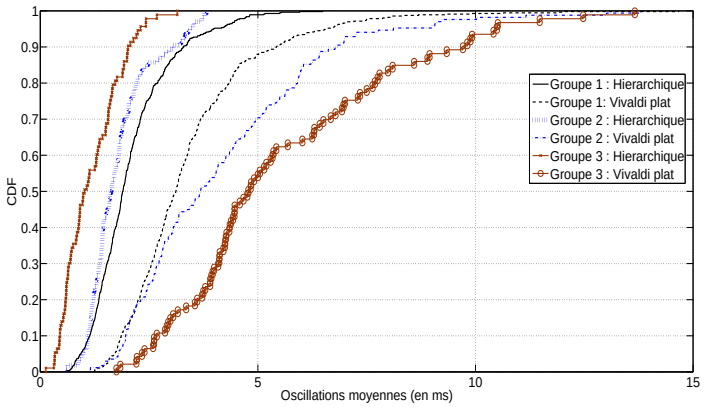


Figure 4. *Comparaison des oscillations des coordonnées pour le Vivaldi "plat" et le hiérarchique.*

Nous avons également observé les oscillations moyennes des coordonnées durant les 500 derniers pas de simulation. Les 3 CDF les plus à gauche de la figure 4 représentent les oscillations des coordonnées locales calculées dans nos 3 groupes. Ces résultats montrent clairement que les coordonnées locales oscillent avec une amplitude plus faible comparées aux coordonnées du Vivaldi plat. En effet, en considérant les coordonnées locales, plus de 80 % des nœuds ont une oscillation moyenne inférieure à $3\ ms$ alors que seuls 40 % des nœuds sont concernés si on considère les coordonnées du Vivaldi plat.

5. Conclusion

Nous avons proposé une approche hiérarchique qui permet d'atténuer l'impact des violations des inégalités triangulaires dans l'Internet. Jusqu'ici, les études sur les systèmes de coordonnées étaient généralement concentrées sur la recherche d'espaces ayant des propriétés géométriques particulières permettant un meilleur positionnement des nœuds. Notre proposition est plutôt basée sur un regroupement des nœuds proches de manière à atténuer l'impact des TIV. Les nœuds calculent des coordonnées locales plus précises pour estimer les chemins intra-groupes, et ils conservent des coordonnées globales pour estimer des chemins inter-groupes.

Même si cette étude s'est focalisée sur le système Vivaldi, l'architecture proposée peut être utilisée avec n'importe quel système de coordonnées. En effet, il suffit de faire tourner une instance du système de coordonnées dans chaque groupe et une instance au niveau global sans rien changer au protocole utilisé.

Lors de l'analyse de l'impact des TIV sur les estimations de distances, nous avons considéré qu'un nœud est impliqué dans une TIV s'il fait partie d'un triangle qui est une TIV (en tant que nœud A, B ou C). Toutefois, dans certaines situations, il se peut qu'un nœud ne soit pas (ou moins) affecté par les TIV existant avec les nœuds qu'il n'a pas choisi comme voisins. D'autres définitions de la notion de degré d'implication dans des TIV pourraient donc être envisagées de manière à analyser plus en détail l'impact des TIV sur la prédiction de distance. Par exemple, nous pourrions considérer qu'un nœud est impliqué dans une TIV uniquement s'il fait partie du TIV *et* s'il a choisi les deux autres membres du TIV comme voisins. Si nous arrivions à mettre en évidence un lien entre la relation de voisinage, les TIV et la précision de la prédiction, nous pourrions utiliser ces résultats pour choisir les voisins de manière à atténuer autant que possible l'impact des TIV au niveau supérieur de notre architecture.

La technique de regroupement des nœuds proposée dans cet article est statique et nécessite la connaissance de la matrice de délais. Nous envisageons l'utilisation d'un processus de regroupement distribué et auto-organisé comme proposé dans [MIN 06].

Remerciements

François Cantin et Bamba Gueye sont financés respectivement par le Fonds pour la Recherche dans l'Industrie et l'Agriculture (*F.R.I.A*), et par le projet européen *ANA*.

6. Bibliographie

[CHU 00] H. CHU Y., RAO S. G., ZHANG H., « A case for end system multicast », *Proc. the ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, Santa Clara, jun 2000.

[COS 04] COSTA M., CASTRO M., ROWSTRON R., KEY P., « PIC : practical Internet coordinates for distance estimation », *Proc. International Conference on Distributed Computing*

Systems, 2004, p. 178–187.

[DAB 04] DABEK F., COX R., KAASHOEK F., MORRIS R., « Vivaldi : A decentralized network coordinate system », *Proc. ACM SIGCOMM*, Portland, OR, USA, août 2004.

[GUM 02] GUMMADI K. P., SAROIU S., GRIBBLE S. D., « King : Estimating latency between arbitrary Internet end hosts », *ACM Internet Measurement Workshop 2002*, Marseille, France, novembre 2002.

[KUB 00] KUBIATOWICZ J., ET AL., « OceanStore : An Architecture for Global-Scale Persistent Storage », *Proc. the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Cambridge, nov 2000.

[LED 07] LEDLIE J., GARDNER P., SELTZER M. I., « Network Coordinates in the Wild », *Proc NSDI*, Cambridge, apr 2007.

[LEE 06] LEE S., ZHANG Z., SAHU S., SAHA D., « On suitability of Euclidean embedding of internet hosts », *SIGMETRICS*, vol. 34, n° 1, 2006, p. 157–168, ACM Press.

[MIN 06] MIN S., HOLLIDAY J., CHO D., « Optimal Super-peer Selection for Large-scale P2P System », *Proc. the International Conference on Hybrid Information Technology*, Washington, DC, USA, nov 2006, p. 588–593.

[NG 02] NG T. S. E., ZHANG H., « Predicting Internet Network Distance with Coordinates-Based Approaches », *Proc. IEEE INFOCOM*, New York, NY, USA, juin 2002.

[P2P] « A simulator for peer-to-peer protocols », `http://www.pdos.lcs.mit.edu/p2psim/index.html`.

[PIA 03] PIAS M., CROWCROFT J., WILBUR S., HARRIS T., BHATTI S., « Lighthouses for Scalable Distributed Location », *Proc. the Second International Workshop on Peer-to-Peer Systems*, Berkeley, CA, USA, février 2003.

[ROW 01] ROWSTRON A., DRUSCHE P., « Pastry : Scalable, distributed object location and routing for large-scale peer-to-peer systems », *Proc. IFIP/ACM International Conference on Distributed Systems Platforms*, Heidelberg, Germany, Nov 2001.

[SAV 99] SAVAGE S., COLLINS A., HOFFMAN E., SNELL J., ANDERSON T., « The End-to-end Effects of Internet Path Selection », *Proc. of ACM SIGCOMM'99*, Cambridge, MA, USA, septembre 1999.

[SHA 03] SHAVITT Y., TANKEL T., « Big-Bang Simulation for Embedding Network Distances in Euclidean Space », *Proc. IEEE INFOCOM*, San Francisco, CA, USA, mars 2003.

[SKY] SKYPE, `www.skype.com/`.

[WAN 07] WANG G., ZHANG B., NG T. S. E., « Towards network triangle inequality violation aware distributed systems », *Proc. the 7th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2007, p. 175–188.

[Y.B 03] Y. BARTAL N. LINIAL M. M., NAOR A., « On metric ramsey-type phenomena », *Proc. the Annual ACM Symposium on Theory of Computing (STOC)*, San Diego, CA, june 2003.

[ZHE 05] ZHENG H., LUA E. K., PIAS M., GRIFFIN T., « Internet Routing Policies and Round-Trip-Times », *Proc. PAM Workshop*, Boston, MA, USA, avril 2005.