

M1 SIRIS	Evaluation de performances
<p style="text-align: center;">TP</p> <p style="text-align: center;">Analyse de données de simulation</p>	

Objectifs : analyse de données de simulation a posteriori ; outils statistiques ; utilisation de GNUplot

Généralités

Lorsque l'on exécute un scénario de simulation, on peut collecter un ensemble de résultats immédiats (i.e. des résultats directement interprétables à la fin de la simulation). On peut également générer une trace de la simulation, c'est-à-dire une liste d'événements survenus pendant l'exécution, qui sera traitée *a posteriori*.

Souvent, les traces de simulation sont des fichiers plats, dont la taille peut être très grande : leur analyse doit donc être effectuée en utilisant des algorithmes de faible complexité, et minimisant les accès au disque.

Un cas concret

On a lancé une simulation discrète décrivant un ensemble de paquets circulant sur un réseau IP, à partir de laquelle on a généré une trace décrivant l'ensemble des changements d'état du système simulé. En d'autres termes, chaque ligne de la trace correspond à un événement traité dans l'échéancier discret. Ces événements correspondent aux changements d'états de chacun des paquets IP simulés.

Topologie du réseau

La topologie du réseau est connue :

- le réseau comporte 26 nœuds et 50 liens bidirectionnels (l'existence du lien (i,j) implique celle du lien (j,i)) et symétriques (les débits sur les liens (i,j) et (j,i) sont égaux),
- le délai de propagation est nul sur chacun des liens : le délai de transmission d'un paquet sur un lien dépend donc uniquement de sa taille et du débit du lien,
- le temps de traitement dans les nœuds (lecture des données d'en-tête, lecture de la table de routage, mise à jour TTL et checksum..) est nul,
- les nœuds disposent d'une file d'attente pour chacune de leurs interfaces (liens) de sortie. Les files sont FIFO, de taille finie. Si l'on veut placer un paquet dans une file d'attente et que celle-ci est pleine, le paquet est simplement détruit.

La matrice d'adjacence est fournie : res26.txt (matrice des bandes passantes en Mbits/s) - voir moodle.

Trafic

Le trafic est composé d'un ensemble de flux UDP. Un flux représente un ensemble de paquets devant être transmis d'une source vers une destination. Chacun des flux démarre et se termine à des instants donnés. A un instant donné, il peut y avoir plusieurs flux actifs entre un même couple (source,destination). Par simplicité, tous les paquets d'un même flux sont de même taille ; en revanche, les tailles des paquets issus de différents flux sont différentes.

La matrice de trafic et les caractéristiques des flux ne sont pas fournis : ils pourront être déduits de la trace.

Trace de simulation

La trace de simulation est un fichier texte qui énumère l'ensemble des événements traités durant la simulation. Chaque ligne est de la forme :

t	code	pid	fid	tos	<bif>	s	d	pos
---	------	-----	-----	-----	-------	---	---	-----

- **t** indique l'heure où est survenu l'événement. La simulation débute à t=0, et se termine lorsque tous les paquets émis sont arrivés à destination (ou ont été détruits dans des files pleines), c'est-à-dire quand l'échéancier est vide. Dans l'analyse de la trace, on ne se préoccupe pas des périodes de warmup et de cooldown.
- **code** précise le type de l'événement traité
 - o 0 : départ de la source
 - o 1 : arrivée dans un nœud intermédiaire
 - o 2 : départ d'une file d'attente
 - o 3 : arrivée à destination
 - o 4 : destruction d'un paquet (placement dans une file pleine)
- **pid** : identifiant (unique) du paquet
- **fid** : identifiant du flux auquel appartient le paquet

- **tos** : type de service (vous n'utiliserez pas cette information)
- **bif** : opérateur de bifurcation (vous n'utiliserez pas cette information) ; cette information n'apparaît pas dans les événements dont le code est égal à 4
- **s** : nœud source du paquet
- **d** : nœud destination du paquet
- **pos**
 - o pour les codes 0, 1, 3, 4 : nœud dans lequel se trouve le paquet à l'instant t
 - o pour le codes 2 : prochain saut = nœud qui se trouve à l'extrémité du lien pour lequel le paquet a été placé en file d'attente.

Travail à effectuer

Le fichier de trace à analyser (136Mo, 3.502.252 événements) est disponible sur moodle (trace2650.txt.zip).

L'objectif est d'analyser ce fichier pour en déduire les résultats listés dans les sections suivantes. Les outils d'analyse (logiciels, langages...) sont libres, mais devront être précisés et justifiés. En revanche, les figures seront obligatoirement réalisées avec GNUplot.

Lorsque vous calculerez des valeurs moyennes, vous appliquerez impérativement les consignes vues en cours (calcul de l'écart-type ou de la variance, intervalle de confiance, etc.)

Données globales

Elles concernent l'ensemble de la simulation.

- nombre de paquets émis/reçus/routés
- nombre de flux
- nombre de paquets perdus, taux de perte
- localisation des pertes : proportion des paquets perdus dans chaque nœud
- délai moyen de bout en bout, pour chaque couple (source,destination)
- temps d'attente moyen dans chaque file d'attente
- taille des files d'attente

Echantillonnage

L'échantillonnage consiste à tracer une courbe représentant les valeurs prises par une variable au cours du temps. Vous pouvez générer ces courbes en créant des fichiers de données pour GNUplot.

Vous représenterez graphiquement la variation au cours du temps :

- du nombre de paquets en attente dans l'ensemble des files d'attente (ce qui permet de repérer les congestions)
- du nombre de paquets en transit dans le réseau
- du délai de bout en bout pour l'ensemble des paquets d'un flux ou plusieurs flux
- du nombre de flux actifs (utile pour repérer les périodes *warmup* et *cooldown*)
- du nombre de paquets perdus

Vous tracerez également des courbes représentant la distribution d'un ensemble de valeurs (densité de probabilité d'une ou plusieurs variables) :

- utilisation de l'ensemble des liens du réseau (temps ou proportion du temps durant lequel ils transmettent des paquets) : cela permet d'avoir une représentation graphique de l'utilisation des ressources,
- la charge des routeurs

Déductions

En observant les séries de données, vous pouvez déduire certains paramètres de la simulation qui ne vous ont pas été fournis :

- donnez la matrice de trafic utilisée pour cette simulation (vous préciserez bien l'unité que vous employez)
- comment (c'est-à-dire, en suivant quelle processus ou quelle loi) ont été générés les paquets de chaque flux ?

Résultat final

Avant la fin de la période précisée pendant la séance, vous déposerez sur moodle un rapport au format PDF comportant :

- la liste des outils utilisés ; vous prendrez soin de justifier le choix de ces outils (performances, passage à l'échelle, etc.)
- les résultats que vous avez obtenus, et les commentaires/interprétations que vous en déduisez

Evaluation

L'évaluation de ce TP reposera sur les critères suivants :

- justesse de vos résultats
- bon emploi des outils statistiques
- pertinence des outils employés
- bien-fondé de vos interprétations
- clarté de la présentation des résultats (par exemple des figures)