

TP1 : analyse de Traces

Le but du TP est d'effectuer l'analyse statistique d'une trace réelle de trafic HTTP. Les traces fournies proviennent de UC Berkeley et les originaux sont disponibles sur <http://ita.ee.lbl.gov/html/contrib/UCB.home-IP-HTTP.html>. Pour le TP, une version allégée et instrumentée vous est fournie sur la page du cours http://mescal.imag.fr/membres/nicolas.gast/enseignements/polytech-eval_perf/tds/td1-trace.tgz.

Téléchargez et décompressez l'archive contenant les traces et les outils.

Cette archive contient les outils permettant de manipuler les traces binaires (fichiers .tr). Les programmes d'origine fournis avec les traces contiennent notamment :

- un parseur : logtrace.c
- un exemple d'utilisation : showtrace.c

Vous trouverez une description des traces et de leur utilisation dans le fichier README.html.

Dans ce TP on vous demande d'analyser les tailles des messages "HTTP response" (header + data) ainsi que les temps de réponse. A cet effet, vous pouvez utiliser l'utilitaire objectsizes.c qui affiche pour chaque transfert la taille et le temps de réponse.

Exercice 1. Pour commencer

- a. Pourquoi le fichier de trace UCB-home-IP-848278026-848292426.tr n'est-il pas lisible ?
- b. Quelles informations peut-on trouver dans cette trace ?
- c. Quelles sont les limitations de l'échantillon ?

Exercice 2. Observation de l'échantillon

- a. À l'aide de l'utilitaire showtrace, générer un fichier contenant les tailles de réponse et les temps de réponse associés avec l'utilitaire objectsizes.
- b. Visualiser les temps de réponses à l'aide de gnuplot ou R. Il y a-t-il des valeurs aberrantes ?
- c. Observer l'échantillon. Observe-t-on une dépendance temporelle ?
- d. Faire la même opération avec la taille des paquets.

Exercice 3. Analyse de la distribution

On commencera par étudier la distribution de la taille des paquets puis on passera à la distribution des temps de réponse.

- a. Tracer la distribution de la fonction.
- b. A quelle distribution typique cette distribution ressemble-t-elle ?
- c. Choisir un ou plusieurs indicateurs représentatifs de votre échantillon (moyenne, médiane, mode, variance...) et les calculer.
- d. Calculer les intervalles de confiance relatifs à la moyenne et à la variance.

Exercice 4. Corrélation entre taille et temps de réponse

- a. Tracer la taille des réponses en fonction de la taille de paquets et commenter le résultat.
- b. Étudier la corrélation entre les tailles et les temps de réponse.