Matthias C. M. Troffaes and Gert de Cooman. Reliable interval estimates for download times. In *Fourth FTW PhD Symposium*. Faculty of Engineering, Ghent University, Dec 2003.

Reliable Interval Estimates for Download Times

Matthias C. M. Troffaes, Gert de Cooman

Abstract—**We propose an algorithm for calculating interval estimates for download times, based on an imprecise version of the linear regression algorithm and a discounted imprecise Dirichlet model.**

Keywords— **imprecise probabilities, statistics, estimation, imprecise Dirichlet model, download time, prefetching**

I. Introduction

WHY are reliable download time estimates important? Obviously, an aesthetic reason cannot be ignored: any web browser should be able to provide the user with a reliable waiting time estimate: "users are often more irritated by variances in their waiting times than by a high mean value of waiting time" [1]. Of a more practical importance, download time estimates are also used in page prefetching and cache removal algorithms [1]. Prefetching reduces page load delays by preloading web pages before a user even selects them. Many prefetch and cache removal algorithms rely on download time estimates for deciding what pages to add to or to remove from the cache. Having a more reliable estimate should therefore also improve overall web browsing performance.

Point estimators for download times abound. However, under particular circumstances, these estimators produce a strongly varying estimate. There are several causes for variations in estimated download times. Sudden changes in network activity, often experienced during peaking hours, have a huge impact on the response time of the server. It is typical for these phenomena that changes are hard to predict, especially on the client side. If a server does not provide a stable transfer speed to all of its clients, then unreliable, and sometimes even oscillating download time estimates are a well-known result.

Surprisingly, no serious effort seems to have been done in developing algorithms for reliable download time estimation. The popular web browser Lynx [2], for instance, uses a straightforward estimate of the connection speed based only on the ongoing download (see Figure 1). This estimate is clearly based on a steady-state assumption—constant transfer speed—which is hardly ever satisfied in practice. In order to provide the end-user with a reliable estimate, we should also take into account possible variations of the future transfer speed.

```
if (LYTransferRate == rateEtaBYTES
  || LYTransferRate == rateEtaKB) {
  if (now - last_active >= 5)
    HTSprintf (&line,
      gettext(" (stalled for %ld sec)"),
      (long)(now - last_active));
  if (total > 0 && transfer_rate)
    HTSprintf (&line,
      gettext(", ETA %ld sec"),
      (long)((total - bytes)/transfer_rate));
}
```

Fig. 1

An extract from the download time estimation algorithm used by Lynx Version 2.8.4 (`HTAlert.c`, lines 266–272)

II. A Statistical Approach

It is often argued that the main aim of statistics is to model variations in the outcome of a repeatable experiment (typically, drawing a sample from a population). No experiment really is repeatable of course. More precisely speaking, statistical methods apply when we are unsure about the exact cause of the variation, when these causes are hard to identify, or when they are hard to measure. In the frequentist approach to statistics, we model our knowledge about such repeatable experiments through a so-called probability density that summarises the frequency of each possible outcome.

Sending a packet over the network is such a repeatable experiment. Naively, we could do the following. In the frequentist spirit, measure frequencies of observed transfer speeds. The measured frequencies will hopefully converge after a sufficient amount of measurements. We thus obtain the probability density of the transfer speed. From this density, the expected value of the download time, or a download time confidence interval, can be easily derived. For example, $E[T] = E[s/V]$, where T is the time needed to download s bytes, and V is the transfer speed (T and V are random variables; the number of bytes s is assumed to be known). Actually, this naive approach may work very well in small networking environments such as local area networks, or homogeneous networks in which the number of users is not too high.

In the more general case, at least two serious objections arise. First of all, the method does not take into account the particular server or network neighbourhood to which the transfer applies. But this is a very important, and easily identified cause of transfer speed variation: different servers are likely to deliver at different transfer speeds. In order to reduce the variance of the transfer speed probability density, and hence, to improve the reliability of the estimator, it is worth considering distinct transfer speed probability densities for each network server, or at least,

Matthias C. M. Troffaes is with the Department of Electrical Energy, Systems and Automation, Ghent University (UGent), Gent, Belgium. E-mail: Matthias.Troffaes@UGent.be .

Gert de Cooman is with the Department of Electrical Energy, Systems and Automation, Ghent University (UGent), Gent, Belgium. E-mail: Gert.deCooman@UGent.be .

for a number of server categories.

Secondly, the method relies on the fact that measured frequencies have converged. It is well-known that one needs a huge amount of independent measurements before, for instance, the strong law of large numbers applies accurately. In such a case, a reliable estimate is only obtained after a long time, and only if the measurements are actually independent. The applicability of this algorithm is questionable. For example, based on what density should a reliable estimate rely when connecting to a new server?

III. Imprecise Estimation

Imprecise probability theory [3] is a recent extension of the classical probability theory that is able to produce estimates even when only a scarce amount of information is available. Instead of a point estimate, imprecise probability theory naturally leads to an interval estimate that grows narrower as more information becomes available. The theory has already been applied successfully in learning (see [4]), reliability theory (see for instance [5]) and estimation (see for instance [6]).

Imprecise probability theory can be considered as a theory that works with convex sets of probability distributions. The imprecise Dirichlet model provides a convenient method for updating such a set as new observations become available. We propose the following strategy for estimating download times.

A. Learning transfer speeds

Consider the packet transfer speed from a particular server to the client as a multinomial process (the possible transfer speeds are discretized). Each time the client receives a packet, we update the client's imprecise model for the transfer speed through an imprecise Dirichlet model—which may discount observations proportional to their age, in order to make the model more robust against changes in the network.

The imprecise Dirichlet model keeps track of the overall behaviour of the client-server transfer speed. We now wish to take the most recent measurements into account in order to obtain a more reliable and more precise estimate of the download time of a single file.

B. Estimating time to arrival

To estimate the time to arrival (ETA, remaining download time) the imprecise model is invoked in an imprecise linear regression algorithm. Let $V_1, \dots, V_n$ be the transfer speeds of n subsequently received recent packets. These transfer speeds are assumed to be described by the same imprecise model as the overall transfer speed V. We then identify a subset A^* of

$$A = \left\{(\alpha_1, \dots, \alpha_n) \colon \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i = 1\right\}$$

for which $[V - \sum_{i=1}^{n} \alpha_i V_i]^2$ is minimised with respect to a *partial* preference ordering that is determined through the imprecise model. The estimate $\sum_{i=1}^{n} \alpha_i V_i$ has the interpretation that a fraction α_i of the total download will be transferred with transfer speed V_i. Contrary to most classical regression methods, the observations $V_1, \dots, V_n$ are not assumed to be uncorrelated—simply no assumption about correlation or independence is made. Assume transfer speeds $v_1, \dots, v_n$ have been observed in the download, a reliable interval estimate for the waiting time for s bytes to be transferred is the convex hull of

$$\left\{s/ \textstyle\sum_{i=1}^{n} \alpha_i v_i \colon (\alpha_1, \dots, \alpha_n) \in A^*\right\}.$$

IV. Conclusions

The at first sight simple problem of estimating download times in a reliable way proves to be a challenging problem both on a theoretical as well as on a practical level. We have argued in what sense classical statistical methods fail to give an overall reliable estimate. We have therefore proposed a more general model that is able to cope with scarce amounts of information. This occurs for example in estimating transfer speeds when connecting to new servers.

Our algorithm consists of two components: a learning component in which data about the transfer speed is accumulated, and an estimation component in which an interval estimate for the download time is given. Our method is constructed in such a way that learning and estimation can be done in a reliable way, and at the same time. This is in strong contrast with classical statistical methods that can only guarantee a reliable result, after a (sufficiently long) learning stage, and under sometimes questionable assumptions such as independence. The main idea underlying our approach is that the imprecise Dirichlet model keeps track of the overall behaviour of the client-server connection, and the linear regression takes local variations into account.

The algorithm might be extended in order to estimate installation times, times to failure, etc.

Acknowledgements

The authors would like to thank the FWO for funding the research reported in this paper. The first author also would like to thank Rene Boel and Erik Quaeghebeur for stimulating discussions.

References

[1] Roland P. Wooster and Marc Abrams, “Proxy caching that estimates page load delays,” *Computer Networks and ISDN Systems*, vol. 29, pp. 977–986, 1997.

[2] “Lynx Version 2.8.4,” URL, July 2001, http://lynx.browser.org/.

[3] Peter Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.

[4] Peter Walley, “Inferences from multinomial data: Learning about a bag of marbles,” *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 3–34, 1996.

[5] Lev V. Utkin and Sergey V. Gurov, “Imprecise reliability for some new lifetime distribution classes,” *Journal of Statistical Planning and Inference*, vol. 105, no. 1, pp. 215–232, June 2002.

[6] Marcus Hutter, “Robust estimators under the imprecise dirichlet model,” in *ISIPTA ’03 – Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon, Eds. July 2003, pp. 274–289, Carleton Scientific.