A Predictive Framework for Web Access over Heterogeneous Wireless Networks

Stephen Drew and Ben Liang
Department of Electrical & Computer Engineering
University of Toronto
{drews,liang}@comm.utoronto.ca

***Abstract*— We present a framework that uses both prefetching and location prediction together to effectively lower the cost and total access time for mobile web users. A new prefetching algorithm is presented that dynamically predicts the likelihood of the mobile moving into a different network. We examine a heterogeneous system composed of two networks: a cheaper and faster WLAN network, and a slower, more expensive cellular data network. Using current industry network rates, our model demonstrates that mobility can significantly raise or lower the prefetching threshold. An accurate mobility prediction scheme can therefore achieve a lower total cost for end users.**

I. Introduction

There are many different types of wireless data networks employed worldwide. These networks often overlap the same coverage area, such that any mobile user could potentially choose from one of many available. Each network can have different properties, and having knowledge of these properties can offer the user a specific advantage. At one extreme there is the cellular data network, such as one built under the 3GPP or 3GPP2 standards that would offer a global-scale coverage and a data rate from 64 Kbps to 2 Mbps. At the other extreme there is the WLAN, which typically supports data rates from 10 Mbps to 100 Mbps, but is limited to a coverage range of a few hundred meters from the wireless access point. It is envisioned that these wireless access technologies will be strategically integrated in order to provide network services efficiently to end users[1].

In this work, we consider a Web user who is roaming in a heterogeneous network composed of both a cellular data network and a WLAN. For Web users, the system of choice would be WLAN. This network is more reliable, faster, and cheaper. However, the roaming user may exit the WLAN coverage area thus leaving her no choice but to continue on the cellular network. Our goal is to demonstrate how a mobile can use location awareness to minimize the more costly and slower cellular access. This can be done through location prediction and prefetching.

Prefetching is a technique used in association with caches to increase the hit rate of system data accesses. For prefetching to be a viable option, there must exist a time of dormant network use for the prediction and fetching to take place. Web page access is a perfect candidate, due to the reading time between user page requests. During this down time, an item deemed worthy of prefetching will be fetched and stored into the cache. The benefit of using prefetching is that it can reduce the total perceived time at the user by turning a long data-access time into a relatively short cache-access time. Previous analysis on Web prefetching include [3] and [4]

A popular approach to prefetching is the concept of a prefetching threshold to determine which items to prefetch. The threshold is essentially a probability level that represents the break-even cost point of the algorithm. If the probability of accessing a specific item is higher than the threshold, then using prefetching will be beneficial to the user. Items that have an access probability lower than the threshold should not be prefetched. In [4], Jiang and Kleinrock introduce a prefetching algorithm that takes into consideration user-access patterns, the impact of multiple-users on the load of the network, and the cost incurred by the user. They optimize their threshold by minimizing the total user cost. Our paper follows a similar approach to prefetching, but instead we consider the probability of the user roaming into a new network in our threshold. This gives us a network-aware prefetching algorithm which we present in the following sections.

II. System Model

We consider a mobile Web user in a two-network environment who wishes to access web documents. He is assumed to access his next Web page at a geometrically distributed time interval as modeled in [2]. The mobile user is free to physically roam anywhere within the system. There are two networks: multiple statistically identical WLAN networks, and a cellular wireless data network which covers the remaining area. Figure 1 illustrates our 2-dimensional model pattern which is repeated to infinity along both axes. The location and velocity of the mobile, as well as the boundaries of the networks are assumed to be known to the mobile (but not necessarily the user) through means of some location awareness scheme such as GPS. The radius of each WLAN network is fixed at size r. The distance between WLAN networks is also fixed at size R. Thus our topology can be summarized by the ratio of r/R. We denote the initial mobile location (x_0, y_0).

To build the mobility aware prefetching algorithm there are three steps : Derive the probability of being in each network for the next web access, compute the threshold level based upon the network probabilities and current network parameters, and determine the files, if any, that exceed the threshold level. The steps should be calculated at discrete time intervals, since the location and velocity of the user is assumed to be dynamic.

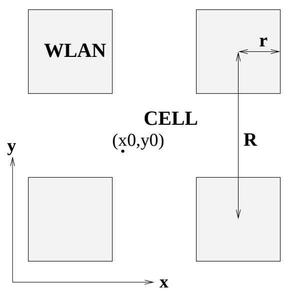

Fig. 1. A 2-Dimensional model of the network topology. The distance between networks is R, and the width of WLAN networks is $2r$. This pattern is repeated in all directions.

The candidate files for prefetching are possible documents that may be accessed based upon the currently viewed page. The file access probabilities can be determined either from examining user history or from the server or a shared proxy cache. The determining of file access probabilities lies outside the scope of this paper. An example method is examined in [4].

A mobility model must also be chosen. We have selected the discrete-time Gauss-Markov model, as found in [5] , which is a probabilistic estimation of future velocities over discrete-time intervals. The n-step velocity is determined by

$$v_n = e^{-\beta\Delta t}v_{n-1}+(1-e^{-\beta\Delta t})\mu+\sigma\sqrt{1-e^{-2\beta\Delta t}}w_{n-1} \ , \quad (1)$$

where μ is the mean velocity, σ is the root-variance, Δt is the discrete time interval, and w_{n-1} is a Normal random variable. The Gauss-Markov model uses a parameter, β, which adjusts according to the degree the model varies from past velocities. A near-zero value of β therefore tends towards a constant velocity pattern, whereas a large value of β tends towards a random walk. We will use different values of β throughout our analysis.

To construct a model of displacement s_k in k time intervals, the velocity model is recursively expanded along an axis. The distribution is Gaussian with mean

$$E[\boldsymbol{s_k}] = \frac{1-e^{-k\beta\Delta t}}{1-e^{-\beta\Delta t}}v_0 + \mu(k - \frac{1-e^{-k\beta\Delta t}}{1-e^{-\beta\Delta t}}) \quad (2)$$

and variance

$$var[\boldsymbol{s_k}] = \sigma^2(1-e^{-2\beta\Delta t})var[\boldsymbol{w}] \quad (3)$$

where

$$var[\boldsymbol{w}] = E\Big\{\big(\sum_{i=1}^{k-1}\sum_{j=0}^{i-1} e^{-\beta\Delta t+i-j-1}\boldsymbol{w_j}\big)^2\Big\} \ . \quad (4)$$

The computation of $var[\boldsymbol{w}]$ follows a similar approach as in [5].

A distribution for k, which is the number of time intervals between web page accesses, can be chosen from literature such as [2]. The probability distribution of the location along any axis for the next web page access is then simply

$$P\{\boldsymbol{x} \leq x\} = \sum_{k=1}^{\infty} P[k]P\{\boldsymbol{s_k} \leq x \mid k\} \ . \quad (5)$$

Taking the initial location x_0, the probabilities along each axis, we can integrate along the bounds of a network to uniquely determine the probability that the next web access lies within a specific network.

To compute our prefetching threshold to minimize user cost, we take an approach building on that used by Jiang and Kleinrock[4]. Setting the HTTP startup time to zero, we obtain for our two network system the following threshold equation

$$H = \frac{\alpha_{B_{curr}}}{\alpha_T(\frac{p_1}{b_1}+\frac{1-p_1}{b_2})+\alpha_{B_1}p_1+(1-p_1)\alpha_{B_2}} \ , \quad (6)$$

where α_{B_1} is the cost per byte of network 1, α_{B_2} is the cost per byte of network 2, α_T is the cost per unit time for the user, b_1 is the bandwidth capacity in bytes for the document using network 1, b_2 is the bandwidth capacity in bytes for the document using network 2, p_1 is the probability the next access will be in network 1 as computed in the previous section, and $\alpha_{B_{curr}}$ is the cost per byte of the current network.

For our model of WLAN and cellular networks, we use approximate industry values for b_1, b_2, α_{B_1}, and α_{B_2}. For α_T, the value will vary among users based on how valuable they perceive time wasted on web access to be. A geometric distribution was used for modeling the time intervals between web page accesses, using an expected value of k=12 as found in [2].

III. Numerical Analysis and Discussions

To get a feel for how the users initial position changes the prefetching threshold, we consider a few different mobility scenarios and examine the behaviour at various points in our network. First, we use a scenario where the average velocity is north-east with a mid range value of β as plotted in Figure 2. As expected, the threshold values in WLAN are much lower than those in cellular. The highest threshold values exist in the cellular region along the south and west edges whereas the lowest threshold values exist in the north and east edges of the WLAN. This is because at these points the mobile has the greatest chance of leaving the current network.

To estimate the average cost savings, we assume that the user has 10 different links to choose from, and we generate 20 different document access probability functions ranging uniformly from equal chance to one document predetermined. Again we take the average over each position in the network. We compare the total cost in our scheme to the cost for using no prefetching. We also compute the gain for a prefetching scheme without mobility awareness [4]. Our mobile prefetching scheme reports gains of 15 to 25 percent for low values of α_T, the user assessed monetary cost per unit time. Figure 3 illustrates the cost gain for varying levels of α_T and mobility parameter β. At high levels of α_T, the two prefetching schemes converge. However, these very high values of α_T are somewhat unrealistic assumptions.

Next, we vary the topology radius r as defined in Figure 1 to show how the prefetching system behaves under different amounts of area coverage. We use radii from 0.05, which has a WLAN area of 1%, up to a radius of 0.5, which has a WLAN area of 100%. Figure 4 plots the cost gain for each of these

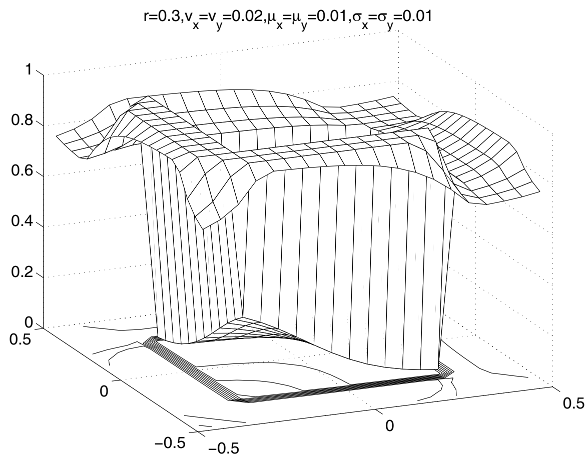

Fig. 2. Prefetch threshold by initial position (x_0, y_0) in the network. Values range from 0-0.3 in WLAN, and 0.7-1.0 in the cellular network. $\beta = 1$,$(\alpha_{BWLAN}, \alpha_{BCELL}) = (0.001, 0.05)\$/KB$, $\alpha_T = 300\$/hr$.

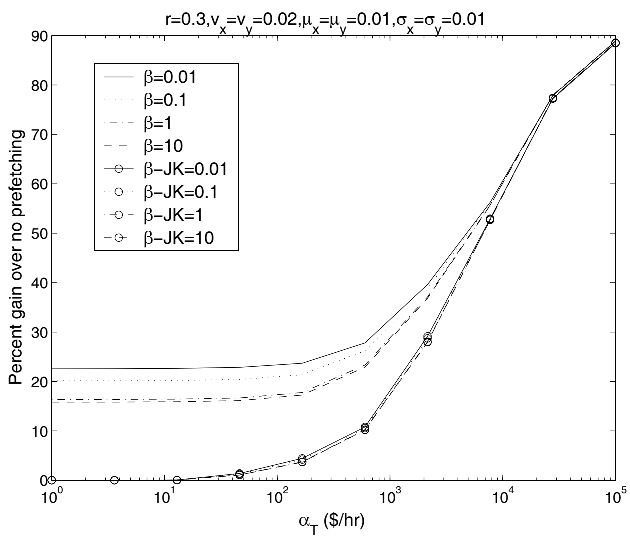

Fig. 3. Average cost gain from using mobility-aware prefetching algorithm versus mobility unaware prefetching algorithm from [4]. Different mobility patterns β are displayed. The two algorithms converge for high values of α_T.

radii against that of the mobility unaware prefetching scheme. The largest cost gains come from a radius around 0.45, which corresponds to a WLAN area of 81%. For a low value of α_T, the difference in performance gain for our scheme is much larger, although the highest performance gains occur at larger values of α_T. At the two extreme ends of the radius choice, the two prefetching schemes behave identically. This is because at these values, the probability of entering a different network is so remote (or does not exist at all) that the system behaves as a one-network model. Therefore, in these cases mobility will not offer any opportunity for system improvement.

IV. Conclusions

Introducing mobility awareness and prediction to the mobile Web environment gives users a means to more fully optimize their time and operation costs. We have presented a novel

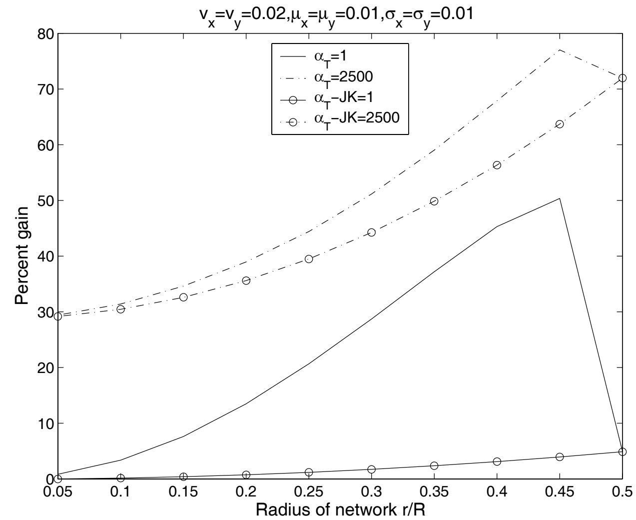

Fig. 4. Average cost gain across different network topology radii. The ratio r/R is adjusted from 0.05 (1% of area is WLAN) to 0.5 (100% of area is WLAN). Our mobility-aware prefetching algorithm is compared to the gain from a mobility unaware algorithm from [4]. $\beta = 0.1$.

threshold algorithm for Web page prefetching in heterogeneous wireless networks, one that dynamically computes the threshold value based on the user's current location and future predicted location. Our numerical results show how the prefetching decision can be altered by user mobility and the heterogeneous network configuration. The proposed prefetching algorithm provides significant performance gains over that of no prefetching, using current industry parameters. It also results in higher performance gain than mobility-unaware, non-predictive prefetching schemes such as that used in [4].

The model of square WLANs surrounded by a cellular area is only an example topology used for this paper. A similar approach can be applied to any topology and can be extended to three dimensions. Additional network models of varying QoS may also be added to the analysis for cases where more than two networks are available. Furthermore, the proposed mobility-aware framework of predictive prefetching not only can be used for Web applications, but also can be modified to suit any other type of application that obeys similar access patterns.

References

[1] R. Berezdivin, R. Breinig, and R. Topp. Next-generation Wireless Communications Concepts and Technologies. *IEEE Communications Magazine*, March 2002.

[2] E. Casilari, A. Reyes-Lecuona, F.J. Gonzalez, A. Daz-Estrella, and F. Sandoval. Characterisation of Web Traffic. *GLOBECOM '01*, 3:1862–1866, Nov 2001.

[3] M. Crovella and P. Barford. The network effects of prefetching. *INFOCOM '98*, pages 1232–1239, 1998.

[4] Z. Jiang and L. Kleinrock. An Adaptive Network Prefetching Scheme. *IEEE JSAC*, 16:358–368, 1998.

[5] Ben Liang and Zygmunt J. Haas. Predictive Distance-Based Mobility Management for Multidimensional PCS Networks. *IEEE/ACM Transactions on Networking*, 11(5), Oct 2003.