Efficient WISH in an Indoor Environment

M. Angelaccio, B. Buttarazzi, R. Giuliano, and G. Guidoni

***Abstract*—With the diffusion of wireless connections to Internet, the number of complex operations carried out from mobile users is increasing. To cope with bandwidth limitations and with disconnections, data caching is the most used technique. However for complex operation like dynamic searching a better solution is to take advantage of the multichannel property offered by CDMA protocol. In this case, cached documents can be allocated on distinguished channels in a dynamic way to obtain a better utilization of the radio communication links. We study a particular caching strategy suitable to be integrated with a radio-channel policy. We consider a semantic caching for intranet queries (or intranet searching) that takes advantage of data semantics by caching query answers instead of pages in order to exploit similarities between different queries. In fact, in a WLAN scenario, Internet activity is frequently composed by intranet searching operations characterized by local queries that aim to explore documents stored in a neighbor of the home site. We study benefits from a channel allocation strategy applied to intranet searching with semantic caching. Simulation experiments are carried out by considering an indoor scenario model where mobile clients perform keyword-based queries answered by local Web servers running application we refer to as WISH (Wireless Intranet SearcHing), an intranet searching tool based on semantic caching. The results show a 12% improvement in radio channel usage for 20% of users that share cached documents.**

I. Introduction

WITH the convergence of wireless communications and the Internet, the number of wireless connections to Websites is growing, thus demanding a broadband wireless access for Web queries [7]. Page caching at client level is an effective way of reducing network traffic and Web latency [1], [5]. However, even if traditional proxy caching efficiently also supports mobile clients that browse through Web documents, a different approach is needed when the Web is viewed as a database. In fact, in this case, mobile clients tend to submit queries instead to specify page URL directly and with the growing diffusion of wireless Internet, is becoming important to keep into account this different Web view. In the mobile database literature [6], [9], semantic caching has been proposed as an alternative to traditional caching. The idea of semantic caching is to maintain in mobile client cache both semantics descriptions and results of previous queries. This enforces a query processing that aims to determine what data are locally available and what data are needed from the server. Recently, semantic caching has been proposed for Web queries like Web databases wrappers [8] and meta-search engine queries [10]. Instead, different types of Web queries often have been submitted by wireless clients. In fact, as shown in [4], mobile clients usually ask for local queries that are restricted to a site or to its neighbor instead to perform global search engines queries. We call this type of searching intranet searching and we study its performance in a multichannel (CDMA) wireless scenario. As a searching platform we use WISH (Wireless Internet SearcHing System) [3] as an example of intranet searching toll based on semantic caching. It is a wireless access protocol proxy for local Web queries managed by a Java search engine. We study how the wireless internet searching efficiency is improved by using WISH as a wireless internet searching for WLAN-based intranets.

Manuscript received February 4, 2003. The review of this letter was arranged by Guest Editor Roberto Sorrentino.

M. Angelaccio and B. Buttarazzi are with the Dipartimento di Informatica, Sistemi e Produzione, Università degli Studi di Roma “Tor Vergata,” 00133 Rome, Italy (angelaccio@disp.uniroma2.it).

R. Giuliano and G. Guidoni are with the Dipartimento di Ingegneria Elettronica, Università degli Studi di Roma “Tor Vergata,” 00133 Rome, Italy.

Digital Object Identifier 10.1109/LMWC.2003.815811

II. Intranet Searching

We consider a WLAN-based intranet where mobile clients performs Internet searching restricted to the intranet (or its neighbor sites). For instance, let us consider as intranet a university campus area with a local Website for each faculty. Users ask for documents through clients connected to a local WLAN network. We assume a typical distributed searching scheme composed by a local *site searching* tool for each faculty server. However, it may happen that a client wants to ask for extra-faculty documents by submitting a query in a different way. He perhaps might use a global search engine or a local search tool that points out of the faculty subintranet. If this type of *extra-site searching* is not well supported by a caching scheme, it may hold that excessive extra-WLAN connections can reduce the query-performance.

A. Semantic Query Caching

The main feature of WISH is the semantic caching mechanism which allows reuse of answers to previous queries, thus reducing the delivery time of answers and the traffic on the net. Semantic caching is based on the representation of cached data as semantic regions and the processing of queries by construction of probe queries for retrieving cached data and remainder queries for fetching data from remote servers. Semantic caching is considered a support for intranet query caching. The query scope could go beyond a local Web server as shown in the section above.

The detailed description of semantic caching for WISH is given in [3]. Fig. 1 shows two cases of *partially answered* queries that arise when the answers returned by WISH go beyond the spidering boundaries that have cached in a previously query. The cache is organized in semantic regions called query maps [Fig. 1(a)] and the way each query is split (query trimming) [Fig. 1(b)].

When a query is posed at a server with a semantic cache, the query is split into two pieces: 1) probe query, which retrieves the portion of the answer available in the cache and 2) remainder query, which retrieves the missing data from the Internet. When a new query comes in, 1) (cache-hit) the result of the new query could be contained in the cache; 2) (query trimming) it could intersect with the cache, or; 3) (cache-miss) it could be disjoint

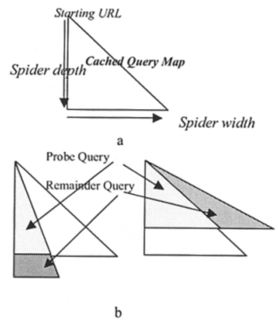

Fig. 1. Partially answered queries.

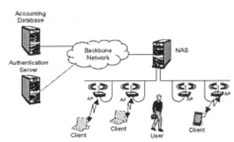

Fig. 2. Wireless architecture.

from the cache. Query trimming happens when similar queries are submitted. In this case, semantic caching can be used.

Using the terminology from semantic caching, the query evaluated over the cache is the ***probe query***, whereas the difference query sent to the Internet is the ***remainder query***. We want to take advantage of this type of caching by introducing a *channel allocation policy* based on this semantic caching scheme supported by WISH.

III. System Model

To study the benefits obtained by using WISH, we use a suitable scenario. We consider a WLAN hosting a Web-based application composed of Web servers and clients querying for documents by using WISH as example of semantic caching intranet searching system. We report a complete description of the system. In particular, the *indoor environment* where WISH is considered and the *pathloss* and *traffic model* used in this scenario are described. At the communication level, the system is composed of Master (Ms) and Slave (Sv). The Ms is the access point (AC) to the internal network or intranet (and to the external network) and the Sv are the intranet users, connected to the Ms in a wireless mode

In the wireless LAN (Fig. 2), we have five components: 1) AP or BTS *(base transceiver station)* or Ms; 2) Client or Sv; 3) NAS (network access server); 4) accounting database; 5) authentication server (RADIUS). The main components of the NAS server are firewall server, Web with support to SSL Client, RADIUS services of system, and script server DHCP.

A. Indoor Environment Description

Our scenario is an *indoor environment* [13]. The analyzed area is large $5000\ m^2 (50 \times 100\ m^2)$. It represent an office environment with some square rooms of side 10 m and corridors large $100 \times 5\ m^2$. In the system, three floors are considered with a height of 3 m. Statistics should only be collected in the middle floor to reduce the boundary effects. The number of Ms (the black bullet in Fig. 3) are less than the number of the Sv.

The Sv are deployed in the area uniformly. The mobility of the Sv is not considered. That is, a generic Sv starts a session

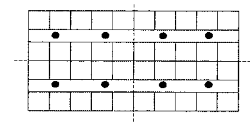

Fig. 3. Indoor office environment.

in a room (or corridor) connecting to the closest Ms and after a period of time he finishes at the same position. The antenna of both devices (Ms, Sv) is omnidirectional. For spectrum efficiency evaluation, quality statistics should only be collected in the middle floor.

B. Path Loss Model for Indoor Office Environment

Due to air separation distance between Ms and Sv, the signal undergoes losses in its propagation path. Our scenario is a typical floor (i.e., offices) with hollow pot tiles, reinforced concrete, and wall thickness less than 30 cm. The walls are mediated taking into account the numbers of holes (e.g., windows). Under the simplifying assumptions of the office environment, the indoor path loss model, based on the COST 231 model, has the form (in dB) [12]

$$L_{dB} = 37 + 10 \cdot \gamma \cdot Log_{10}(R) + 18.3 \cdot n^{\left(\frac{n+2}{n+1}0.46\right)} + \zeta \quad (1)$$

where n is number of the floor met by the signal in the path, R is transmitter–receiver separation in meters, and γ is propagation coefficient set to 3. In (1), a Gaussian random variable, ζ, has zero mean and standard deviation, σ, equal to 12 to take into account the slow variation of the shadowing effect.

C. Traffic Model

In the intranet, most of offered traffic concerns on nonreal time services, so we do not consider work speech traffic. The typical intranet request is a browsing *session*. In general, a packet service session contains one or several *packet calls* depending on the application. The user initiates a packet call when requesting an information entity. After the document arrives at the terminal, the user consumes a certain amount of time to study the information (*reading time*).

User request to start a session has an exponential distribution with mean λ. Number of packet call requests per session, N_{pc}, is a geometrically distributed random variable with a mean $\mu_{N_{pc}}$. Reading time between two consecutive packet call requests in a session, D_{pc}, is a exponentially distributed random variable with mean $\mu_{D_{pc}}$. Data volume for a single packet call is a lognormal random variable with mean and standard deviation μ_V and σ_V, respectively [11], [13].

IV. Channel Allocation Strategy

Simulation is carried out by modeling a *channel allocation scheme* based on the semantic caching scheme of WISH.

In a typical intranet, type of access to the internal server is correlated both in time and requested information. WISH allows us to reduce the duplication of the information to be sent. The probability, $P(k)$ that k users require the same part of information (equal to p_n) if n users are simultaneously connected to the Ms at a certain time is

$$P(k) = \binom{n}{k} \cdot P_n{}^k \cdot (1 - p_n)^{n-k}. \quad (2)$$

p_n is not fixed. The probability that a user requires the same information (or part of it) of another user increases as the amount of connected users increases at the same Ms. For this probability, $p_n = (a + bn + cn^2)p_0$ where the coefficients are taken assuming that $p(1) = 0, p(n_{\text{MAX}}/2) = 30\%$ and $p(n_{\text{MAX}}) = 1$ and where n_{MAX} is maximum number of active connections that a single Ms can manage and p_0 is the probability that a user requires the same information (or part of it) of another user if the n_{MAX} users are connected at the same Ms.

The rest of the $n-k$ users could require the same information (or part of it) different from the previous k users. The recursive algorithm is described in the following way.

CDMA-WISH Channel allocation algorithm

For each simulation step and for each Ms in the considered scenario, do:

1st step:

At Master, n Slaves are connected. The Master, thanks to WISH application, is able to know the k_1 Slaves which have common information between them (cache-hit or query trimming).

2nd step:

If k_1 is >1, at Master, $n - k_1$ Slaves are connected without considering WISH caching scheme. Then, applying WISH to them, the Master is able to know the k_2 Slaves which have common information between them and continues with ***2nd step*** changing k_1 with k_2 and $n - k_1$ with $n - k_1 - k_2$, and so on

If k_1 is equal to 1, the algorithm exits from the inner loop and the Master transmits using just the useful codes, without duplicating the information.

In order to take into account the duplicate information, found by WISH, the Ms is able to transmit the part of common information exploiting the same channel toward all interested Sv. This provides a reduction in system interference, better quality, and larger number of available simultaneously active Ms-Sv connections.

V. Simulation Results

A complete system simulator has been implemented to evaluate the performance of the CDMA-WISH allocation strategy. The scenario is that described in Section III-A. The considered service is an HTTP session with the parameters reported in the following: exponential distribution for the reading time with $\mu_{D_{pc}} = 30$ ms, lognormal distribution for data volume with $\mu_V = 7.47$ bit, and $\sigma_V^2 = 3.76$ bit. We suppose that the available connection between Ms and Sv is of 144 kb/s with a $(C/I)_{req} = -16.18$ dB.

A. Simulation Outputs

System quality has been evaluated by plotting the outage probability as function of the number of users per cell at the connection speed of 144 kb/s and reading time of 30 ms. Fig. 4 shows the plots obtained for different values of p_0 that measures the percentage of users submitting similar or equivalent queries that are answered by using WISH cache regions. When $p_0 = 0$, it holds that the semantic caching of WISH is not applied.

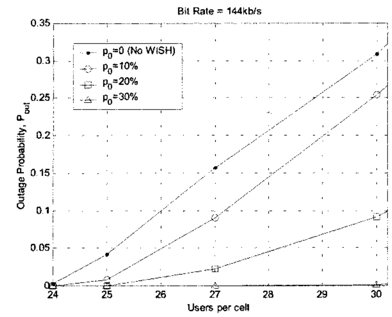

Fig. 4. Users per cell versus outage probability for different value of p_0 in HTTP service.

If we assume that the value of outage probability P under which system quality would be acceptable is equal to 5%, then we have that in absence of WISH $(p_0 = 0)$ at most 25 users per cell must be considered. On the contrary, when WISH is adopted with values of $p_0 = 10\%$ and $p_0 = 20\%$, respectively, the maximum number of users for cell increases up to 26 and 28, respectively. Therefore, we obtain a gain of 4% and 12%, respectively, for the two cases. In the third case of $p_0 = 30\%$ the obtained gain is beyond 60%.

VI. Conclusions

We have presented the experimental performance evaluation of a CDMA channel allocation strategy used in combination of semantic caching scheme for a wireless intranet searching application called WISH. The analysis is carried out by considering a common scenario of wireless connected clients submitting Web queries in an indoor environment. The results show that for more than 25 users per cell, the improvement in the quality of the system is obtained even for a small percentage (10%) of users that find documents in the cache. This can be used as motivating reason to include this type of caching as support for intranet searching in a wireless indoor office scenario like campus or similar environments that are logically associated to intranets.

References

[1] J. Wang, "A survey of Web caching schemes for the Internet," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 5, pp. 36–46, Oct. 1999.

[2] M. Angelaccio and B. Buttarazzi, "Local searching the Internet," *IEEE Internet Computing*, vol. 6, no. 1, pp. 25–33, Jan.–Feb. 2002.

[3] ——, "Semantic Caching for Wireless Internet Querying—Tech. Rep.,".

[4] O. Buyukkokten and H. Garcia-Molina, "A focused Web searching with PDAs," in *Proc. 9th WWW Conf.*, 2000.

[5] P.Pei Cao and S.Sandy Irani, "Cost-aware www proxy caching algorithms," in *Proc. 1997 USENIX Symp. Internet Technology and Systems*, 1997, pp. 193–206.

[6] S. Dar, M. J. Franklin, B. Jonsson, D. Srivastava, and M. Tan, Semantic Data Caching And Replacement VLDB, Sept. 1996.

[7] D. J. Goodman, "The wireless internet: promises and challenges," *IEEE Computer*, vol. 33, no. 7, July 2000.

[8] D. Lee and W. Chu, "Semantic caching via query matching for Web sources," in *Proc. 8th ACM Conf. CIKM 99*.

[9] Q. Ren and M. Dunham, Semantic Caching in Mobile Computing, 2000.

[10] C. Roncancio, B. Chidlovski, and M. Schneider, "Semantic cache mechanism for heterogeneous Web querying," in *Proc. WWW8*, K. Jones, J. Doe, and J. Smith, Eds., 1998.

[11] 3GPP, TR 101 112 V3.2.0 (UMTS 30.03 version 3.2.0) [Online]. Available: http://www.3gpp.org

[12] "COST 231 Final Rep.," Commission of the European Communities.

[13] A. Klemm, C. Lindemann, and M. Lohmann, "Traffic modeling and characterization for UMTS networks," in *Proc. GLOBECOM '01*, vol. 3, 2001.