A Model for Optimizing Data Cashing of Dual Mode Handheld Devices

Mohamed Grida[1], **Hasnaa Soliman**[2] **and Mohamed Hassan**[2]

[1]**Industrial Engineering Department, Faculty of Engineering, Zagazig University**
Zagazig, Egypt

[2]**Information System Departments, Faculty of Computers and Informatics, Zagazig University**
Zagazig, Egypt

Abstract

Handheld apparatus are widely used especially when equipped with an automatic data capturing solutions. They are usually connected to the main database server using either an economic fixed location mode or a more expensive mobile mode. In order to quantify the gap of the two modes, a metric called mobility coefficient is introduced. Then, a mathematical model is developed to optimize the handheld data caching. This model considers several factors affect the cache performance such as the data item size, its access frequency, its update rate, the cache size, and the caching cost. Finally, a series of simulation experiments are conducted in order to validate of the proposed model under various system configurations.

Keywords: *Handheld performance, Caching optimization, dual mode gadget.*

1. Introduction

With the rapid evolution in mobile computing and wireless technologies, more smart handheld devices (such as laptops, palmtops, and personal digital assistants (PDAs)) are supported and used all over the world, and they become a significant part of everyday life and work. These devices were defined as small, mobile, and battery-powered computing devices, which have no mass storage system and all persistent data have to be kept in a limited battery-buffer RAM [1] [2][3][4]. The main motivation for using these devices is their ability to reduce data entry errors while increasing the data entry rate [1] [4]. At the same time, current handheld devices support various ways for communication with the mass data storage devices such as serial cables or Universal Serial Bus (USB) for direct connection, short range wireless connections such as Bluetooth, IrDA and Wi-Fi (IEEE 802.11) and cellular networks for long distances through GPRS or 3G [3] [5]. Handheld devices can be connected with the mass data storage devices (such as the database server) through two main modes, which are economic base mode (for connectivity inside the organization) and costly mobile mode (for connectivity outside the organization).Within the base mode, handheld can be connected to the database server through serial cables, USB, Bluetooth, or Wi-Fi by which they can obtain bulky amount of data with low cost, high bandwidth and little power consumption. On the other hand, within the mobile mode, a handheld can be connected to the database server through cellular network via GPRS or 3G technologies, which are costly and provide limited bandwidth [6] [7][8][9][10][11][12] [13]. In order to overcome the bandwidth and the cost limitation of the mobile mode, data caching at the handheld local memory is considered as an effective solution to overcome these limitations and to reduce the network traffic [14] [6][15][8][7][16][17] [13]. Due to the limited storage capacity and processing power of the handheld devices, it will be impossible to hold all the accessed data items at the handheld local memory. Consequently, a model for optimizing the handheld data caching is needed by which a user can determine which database items can be cached at handheld local memory in the base mode. This model should aim to reduce the total cost and traffic between the handheld device and the database server.

In the next sections, an overview of related and previous researches on caching area will be presented. After that, a model for optimizing the handheld data caching will be developed based on a brief description of the handheld system architecture. Finally, the performance of the proposed scheme will be examined before concluding and highlighting future work possibilities.

2. Literature survey

Through the past few years, cache optimality has been addressed by number of researchers to decide which data items should be kept in or evicted from the cache. In this section, a brief overview of the caching replacement schemas will be introduced along with the factors affecting the cache performance. Most of previously described cache replacement policies were represented as a function of the different factors such as access frequency, update rate and so on.

In 2000, Xu et al. [18] proposed a cache replacement policy called Stretch Access Rate Inverse Update – frequency (SAIU), which was represented by a gain function as shown in Eq.(1).

$$gain(i) = \frac{A_{delay(i)} \cdot P_{A_i}}{S_i \cdot U_i} \qquad (1)$$

This policy considered only four factors affect cache replacement decision of a data item i, which were access probability (P_{A_i}), update frequency (U_i), retrieval delay ($A_{delay(i)}$), and data size (S_i). It aimed to remove the data item with the least gain function value from the cache [18]. In 2004, Xu et al. added cache consistency to the former policy due to its importance for some applications such as financial transactions. An updated SAIU replacement policy was introduced by including (cache validation delay caused by the underlying cache invalidation scheme) as cache consistency metric. The updated policy was called Minimum Stretch integrated with Access rates, Update frequencies, and cache validation Delay (Min-SAUD) [15]. They introduced a new gain function considering two new factors, which are the cache validation delay ($V_{delay(i)}$) and the ratio of update rate ($\overline{u}_i$) to access rate ($\overline{a}_i$) for a data item i (X_i) as shown in Eq. (2).

$$gain(i) = \frac{P_{A_i}}{S_i}\left(\frac{A_{delay(i)}}{1 + X_i} - V_{delay(i)}\right) \qquad (2)$$

In 2005, Yin et al. introduced a cache replacement policy based on the generalized cost function expressed in Eq. (3).

$$Cost(i) = P_{A_i}\left(A_{cost(i)} - V_{cost(i)} - P_{U_i}U_{cost(i)}\right) \qquad (3)$$

This policy brought in three additional factors affecting the decision of caching a data item or not, which were the cost of fetching a data item to the cache ($A_{cost(i)}$), the cost of validating the cache consistency ($V_{cost(i)}$) and the cost of getting the updated data from the server ($U_{cost(i)}$). It aimed to evict the data items with the least value of $[\mathrm{cost(i)}\ /s_i]$ from the cache [8].

In 2007, Chand et al. developed a cache replacement policy called Least Utility Value with Migration (LUV-Mi), which aimed to maximize the benefit of caching the data items by measuring a utility value for each data item kept in the cache as shown in Eq. (4). The data item with a high utility value will be kept in the cache. This policy introduced an additional factor affecting the decision of replacing a data item, which is time to life value (TTL_i) that indicates the freshness or validity of a data item. Moreover, it introduced the distance (δ_i) that calculates the number of hops between the requesting client and the responding client (data source) as a form of the data transmission cost [19].

$$utility_i = \frac{P_{A_i} .\delta_i .TTL_i}{S_i} \qquad (4)$$

Based on the generalized cost function proposed by Yin et al. (2005), Chand et al. (2009) proposed a novel cache replacement policy called Least Profitable Value (R-LPV), which based on a profit function expressed by Eq. (5).

$$PF_i = A_i \times \left(A_{delay(i)} - V_{delay(i)} - P_{U_i} \times U_{delay(i)}\right) \quad (5)$$

This replacement policy divided the delay factor into three sub factors, which are the data access delay ($A_{delay(i)}$), the cache validation delay ($V_{delay(i)}$) and data update delay ($U_{delay(i)}$) affecting the decision of caching the data items. Moreover, it used the access frequency(A_i) as a metric of the data access probability (P_{A_i}) representing the expected number of accesses to item (i) per a certain period of time. This aimed to maximize the total profit value for cached data items and it evicted an item (i) with minimum $[\mathrm{PF_i}/s_i]$ value from the cache [20].

In 2010, Pant et al. used only two of previously factors in order to remove the item with least important value from the cache and to introduced a cache replacement policy for wireless sensors networks called item replacement policy (IRP) represented by Eq. (6)[21].

$$Imp_i = P_{A_i} \times TTL_i \qquad (6)$$

In 2011, Dimokas et al. used some of previously described parameters and developed a cost based cache replacement policy for wireless sensor network. The data items with the greatest costs were removed from the cache [17]. This policy was represented as cost function as shown in Eq. (7).

$$\mathrm{cost(i)} = \frac{\mathrm{A_{delay(i)}} \times \mathrm{s_i}}{\mathrm{TTL_i} \times \mathrm{A_i}} \qquad (7)$$

Finally, all of the previously described cache replacement policies were summarized in Table 1.

Table 1: The most commonly factors affecting the cache performance along with the author who used each factor in a study

Decision Factors	***Authors***						
	Xu et al. (2000)	***Xu et al. (2004)***	***Yin et al. (2005)***	***Chand et al. (2007)***	***Chand et al. (2009)***	***Pant et al. (2010)***	***Dimokas et al. (2011)***
Item Size(s_i)	Yes	Yes	Yes	Yes	Yes		Yes
Access probability (P_{A_i})*or Access frequency* (A_i)	Used P_{A_i}	Used P_{A_i}	Used P_{A_i}	Used P_{A_i}	Used A_i	Used P_{A_i}	Used A_i
Update probability (P_{U_i}) *or Update frequency* (U_i)	Used U_i	Used P_{U_i}	Used P_{U_i}		Used P_{U_i}		
Distance (δ_i) *between the requesting client and the data source*				Yes			
Data retrieval delay	Yes	Yes			Yes		Yes
Data retrieval cost			Yes				
Time-To-Live value associated with each item(TTL_i)				Yes		Yes	Yes

3. The proposed model

The mobile computing system for the proposed model mainly consists of a central server/s, which stores a large database and a number of mobile clients (handheld devices), which host limited size mobile databases as shown in Fig. 1. These devices may not be always connected to the server due to their operational constraints. Therefore, these devices can connect to the database server periodically in order to obtain their required data using one of two modes the base economic mode and the costly mobile mode.

According to the previously described survey, there are seven factors affect the decision of caching data or request them from the server when required through the mobile networks. These factors can be categorized into two main categories. First, factors related to system, which are the data retrieval delay and data retrieval cost in addition to the distance between the mobile client and the server. Second, factors related to data, which are the data item size, its access frequency, its update rate and its time to live. The proposed model addresses the first category through introducing a metric called mobility coefficient in order to quantify the system connectivity effectiveness. Then, it utilizes the second category in order to optimize the handheld data caching.

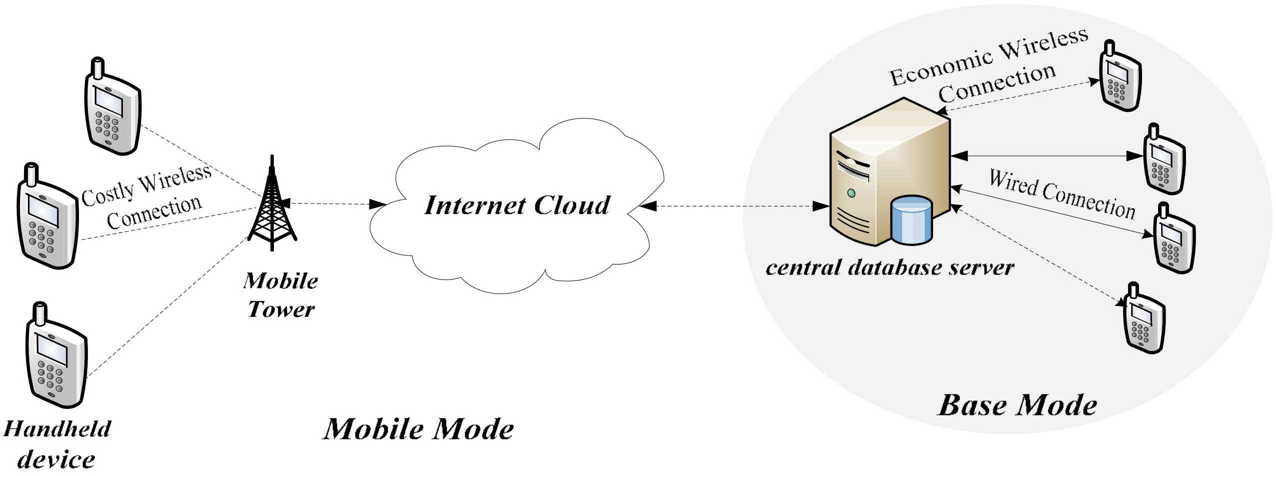

Fig.1Theproposed system architecture

3.1 Mobility coefficient

According to system architecture, a mobile client with a handheld device can connect to the central database server either through the base mode or through mobile mode. The operational conditions of the mobile computing environment may have some obstacles that may not exist in non-mobile environments. In order to quantify such obstacles, a metric called Mobility Coefficient will be introduced. Based on this measure, an organization can decide the return on its investment in caching resources. Thus, mobility coefficient can be represented by an equation annotated by M_C as shown in Eq. (8).

$$\mathrm{M_C} = \frac{\text{Attributes of mobile mode}}{\text{Attributes of base mode}} \tag{8}$$

According to previously described survey, there are three important factors affecting the connectivity between the handheld device and the database server either in base mode or mobile mode, which are data retrieval delay, data retrieval cost and the distance between requesting node and the server. As this distance was introduced as a form of data transmission cost, it will not be used in the proposed model. Moreover, [6] [8] [9] [10][11] [7] [16][12] illustrated that there are two reasons for the data retrieval delay, which are the bandwidth (transmission speed) and the availability of the service. Consequently, it can be concluded that there are three factors affecting the connectivity between the handheld device and the server, which are the data transmission cost, the transmission bandwidth and the service availability. Therefore, the mobility coefficient can be represented as a function of the three previously mentioned factors as shown in Eq. (9).

$$M_C = \left(\frac{BW_b}{BW_m}\right) \times \left(\frac{COST_m}{COST_b}\right) \times \left(\frac{AV_b}{AV_m}\right) \tag{9}$$

Where:

- BW_m: Bandwidth of mobile mode.
- BW_b: Bandwidth of base mode.
- $COST_m$: Transmission cost for the mobile mode.
- $COST_b$: Transmission cost for the base mode.
- AV_m : Availability for mobile mode.
- AV_b: Availability for base mode.

According to Eq. (9), there is one positive relationship between the connectivity effectiveness and the ratio of transmission cost for mobile mode to the transmission cost for the base mode$\left(\frac{COST_m}{COST_b}\right)$. While, there are two negative relationships between the connectivity effectiveness and the ratio of the bandwidth of two modes$(\frac{BW_m}{BW_b})$, and the ratio of availability of two modes$(\frac{AV_m}{AV_b})$.

Based on the value of M_C, the organization can decide to size of the handheld cache. For systems with high mobility coefficient ($M_C \gg 1$), the organization should invest the hardware of handheld devices in order to increase the amount of data being cached and reduce the connection to the server using mobile mode. While for systems with low mobility coefficient ($M_C \cong 1$), smaller cache size should be considered because requesting data from the server through mobile mode instead of caching them will not affordable.

3.2 The proposed Caching model

In this section, a simple client side data-caching model will be introduced. Fig.3 shows the data flow diagram (DFD) of model. In the base mode, the handheld connect to the server to acquire the data to be cached. In mobile mode, it connects to the server either to request non-cached data or to update the cached data.

The previously described survey illustrated that there are four important factors related to data items and can affect the decision of caching a data item, which are the data item size (S_i), its access frequency (A_i), its update rate (U_i), andits time-to-live (TTL_i). However, the value of $\mathrm{TTL_i}$ can be expressed in term of update rate for each data item ($\mathrm{U_i}$). There is a negative relationship between the gain and the data item size and its update rate resulting in reducing the opportunity of caching data [18]. However, there is a positive relationship between the gain and the access frequency resulting in increasing the opportunity of caching. Moreover, increasing the data item size will result in increasing the traffic between the client and the server. Consequently, the proposed model aims to reduce the network data traffic denoted by "$\mathrm{M_{traffic}}$", which can be calculated as following:

$$\begin{aligned} M_{traffic} &= [(1 - X_i) * S_i * A_i] + [X_i * S_i * U_i] \\ &= S_i * A_i - X_i * S_i * A_i + X_i * S_i * U_i \\ &= S_i * A_i + X_i(S_i * U_i - S_i * A_i) \\ &= S_i * A_i + \big(S_i * X_i(U_i - A_i)\big) \end{aligned} \tag{10}$$

- $\mathrm{X_i}$: A binary caching decision variable.

If a data item (i) is cached in base mode ($X_i = 1$), it will need to be updated periodically by the server. If it isn't cached ($X_i = 0$), it will be requested from the server when required through mobile networks.

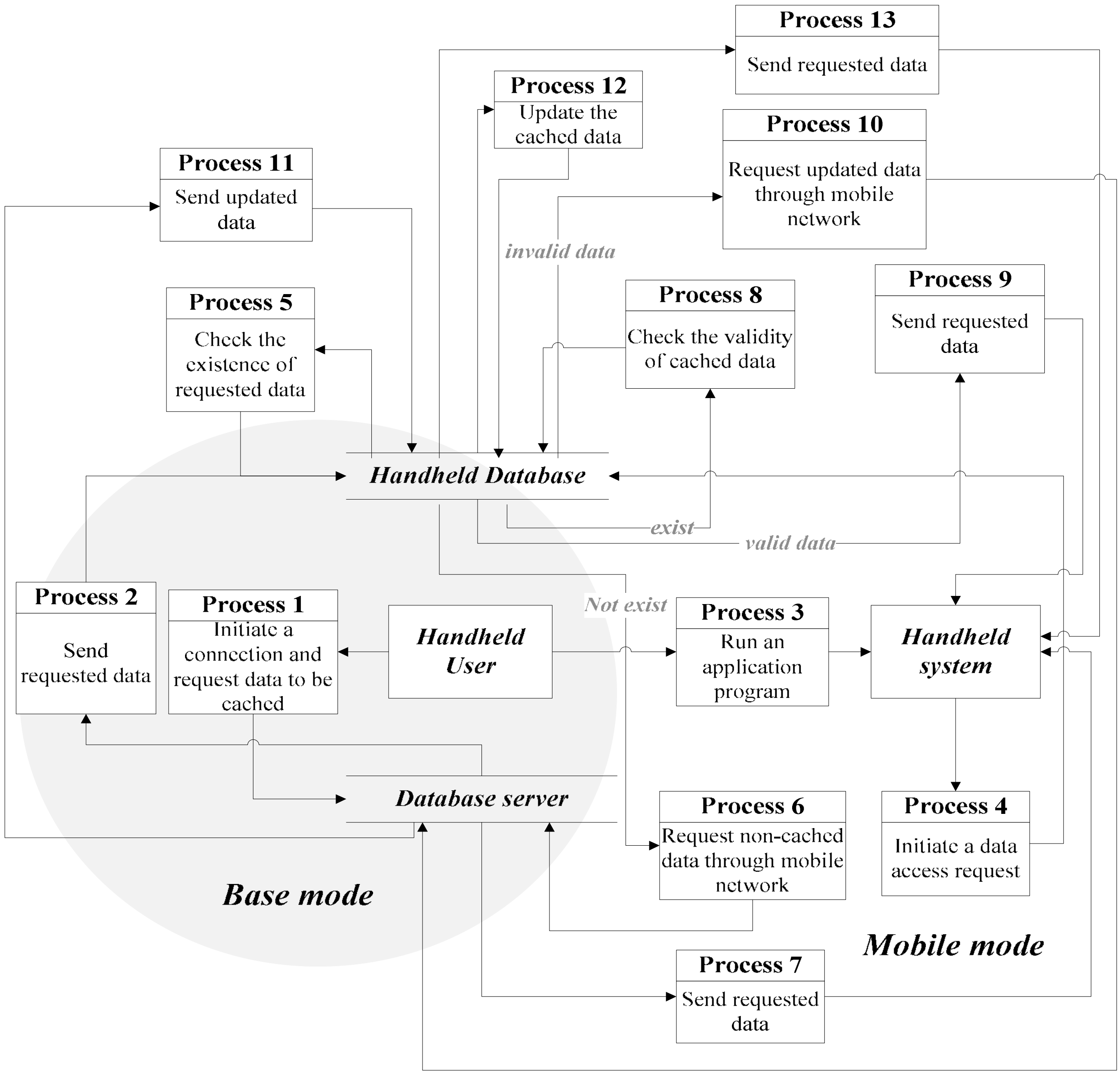

Fig. 2 The DFD Diagram for the system model

As the Eq. (9) for the mobility coefficient and Eq. (10) for computing mobile traffic affect the decision of caching a data item at the handheld local memory or not, both equations should be integrated into one formula in order to introduce the final representation for the proposed model. This formula is denoted by weighted mobile traffic as follows:

Weighted Mobile Traffic = Mobility Coefficient × Mobile Traffic

$$= M_C * \left[\sum_{i=1}^{N}\left(S_i * A_i + \left(S_i * X_i(U_i - A_i)\right)\right)\right] \tag{11}$$

The literature introduced an additional factor affecting the decision of caching data locally at the handheld or not which is the size of the handheld cache [15] [8] [19][10] [11] [20] [16] [17] [22] [13]. Therefore, Eq. (11) representing the proposed model should be updated to include the new factors as follows:

$Total\ operating\ cost$

$$= [Tran_{cost} * \text{Weighted Mobile Traffic}] + [Cach_{cost} * Cach_{size}]$$

$$= Tran_{cost} * M_C * \left[\sum_{i=1}^{N} \left(S_i * A_i + \left(S_i * X_i(U_i - A_i)\right)\right)\right] + [Cach_{cost} * (\sum_{i=1}^{V} X_i * S_i)] \qquad (12)$$

In order to minimize the total operating cost, Eq. (12) is differentiated with respect to X_i as following:

$$\frac{d(Total\ operating\ cost)}{d(X_i)} = 0$$

$$(A_i - U_i) = \frac{Cach_{cost}}{Tran_{cost} * M_C} \qquad (13)$$

The right side term of the above equation tend to be very small due to the great value of M_C. Therefore, it is clear that the value of $(A_i - U_i)$ is critical to decide whether to cache a data item into the handheld or not. Items with update rate higher than its access frequency should not be cached and vice versa.

4. Case Study

In order to test the proposed model validity, a set of Matlab functions was developed to simulate the whole system along with the proposed cache optimization model. A set of data representing a two mode handled operation of a food distribution company in Cairo, Egypt was obtained in order to test the system, as shown in Table 2. The company maintains its main Enterprise Resource Planning (ERP) database within its headquarter, where all the sales and distribution representatives connects their handheld units through a Wi-Fi-network. After that, the team used to connect to the ERP server in order to update the products, roots, orders, and customers info through 3G mobile network.

Table 2: Default system parameter settings

General parameters notations	***Description***	***Default value***
N	Number of data items in the database	28 data items
BW_m	The bandwidth for mobile mode	14.4 Kbps
BW_b	The bandwidth for base mode	115 Kbps
$COST_m or Tran_{cost}$	The cost for transmitting 1GB of Data in mobile mode	10 dollars
$COST_b$	The cost for transmitting 1GB of Data in base mode	1 dollar
AV_m	Service availability for the mobile mode	30 %
AV_b	Service availability for the base mode	70%
Max_{cache}	Maximum cache size for handheld	32kB
S_i	The data item size	Ranging from .78 to 10000 KB
U_i	The update rate for each data item per day	Ranging from 1 to 1000 times daily
A_i	The access frequency for each data item per day	Ranging from 1 to 300 times daily
$Cach_{cost}$	The cost of acquiring and running 1GB of handheld cache	1 dollar

4.1 Using default system parameters

In this section, the performance of the proposed model is investigated by using default system parameters shown in Table 2 and the simulation results are displayed in Fig. 3. According to Fig. 3, it is obvious that the proposed model doesn't cache all data items but only a small subset of the data items whose value of cache coefficient $(U_i - A_i)$ is less than zero. Moreover, some of the negative cache coefficient data items were not cached due to the limited size of the cache.

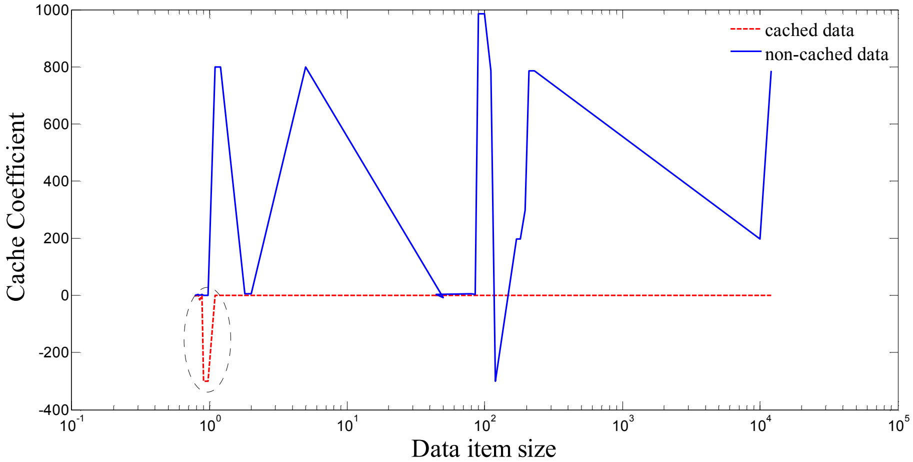

Fig. 3 The Proposed model performance using default data items' parameter settings.

4.2 The impact of increasing the cache size

According to the proposed model, not only the data item's access frequency and its update rate but also the cache size can affect the cache performance. In order to investigate the impact of the cache size on the proposed model performance, in this section, the proposed model is examined using a different cache size. The simulation results are shown in Fig. 4. By comparing the simulation results in Fig. 3 and Fig. 4, it can be concluded that increasing the cache size to 32GB instead of 32 KB results in caching all of more frequently accessed data items whose Cache coefficient value less than zero. Therefore, it can be concluded that the proposed model balances between caching more frequently accessed items and the cache size. When the cache size is limited, only the more frequently accessed small sized data items are cached. On the other hand, when a plenty of cache is available, all of the frequently accessed data items rare cached regardless their size as shown in Fig. 4.

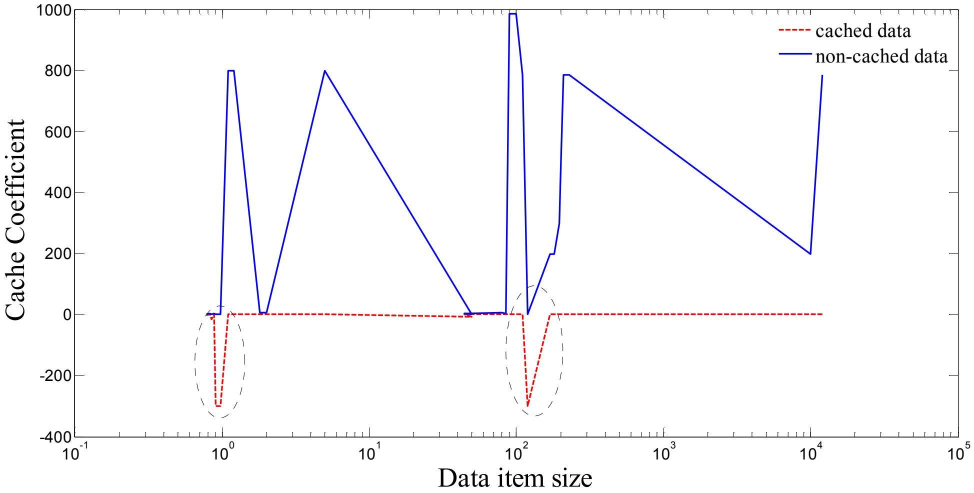

Fig.4 The impact of the cache size on the proposed model performance

4.3 Impact of the data transmission cost

According to Eq. (12) and Eq. (13), it can be concluded that not only the cache size, the data access frequency and update rate affect the caching decision, but also other factors such as data transmission cost can have an impact on the caching decisions. Therefore, the proposed model is examined with a theoretically very low transmission cost in order to force the right term side of Eq. (13) to have a positive value away from zero. The simulation results are displayed in Fig. 5. The proposed model cached only the data items with negative cashing coefficient that has an absolute value greater than the caching-transmission weighted value shown in the right side of Eq. (13).

5. Conclusions

A model for optimizing the handheld data caching in a dual mode environment is proposed. That model objective is to reduce traffic between the handheld used and the server. This model is based on a literature survey that determined the factors affecting the caching decisions, which are the data item size, its access frequency, its update rate and the cache size and the transmission cost.

In order to evaluate the performance of the proposed model, a set of MATLAB functions is created to simulate the system and proposed optimization model. These functions were used along with a data set obtained from an Egyptian food distribution company in order to test the validity the system outcomes. The tests showed that the system was able to determine the appropriate caching decisions under different caching sizes and transmission costs. There is a need to test the model with data obtained from different domains and system configurations in order to exploring its applicability with these configurations.

.

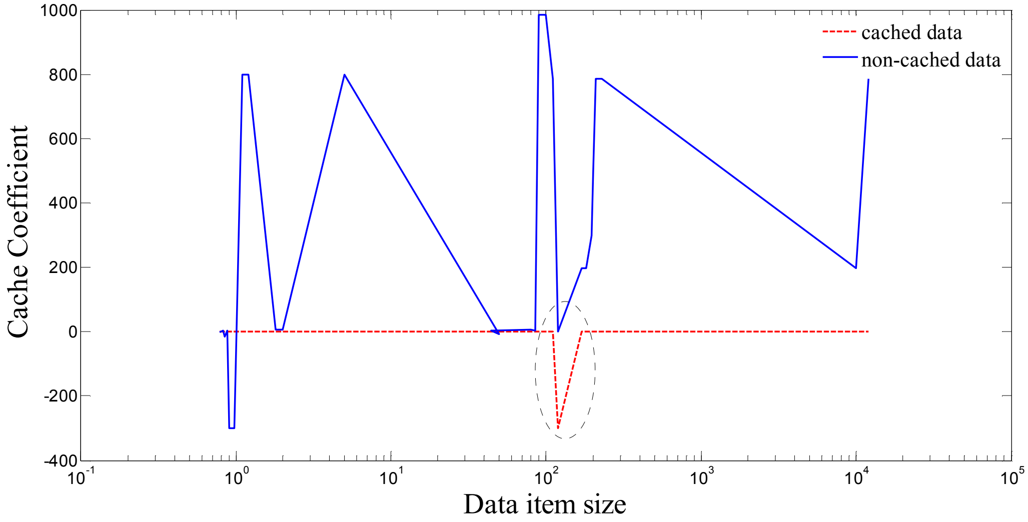

Fig.6 The impact of the data transmission cost on the proposed model performance under large cache size (32GB).

References

[1] J. A. Blackman, “The Usefulness of Handheld Computers in a Surgical Group Practice”, M.S. thesis, School of Medicine, Oregon Health Sciences University, May 1999.

[2] F. A. Adatia, and P. L. Bedard,“Palm reading”: 1. Handheld hardware and operating systems, Canadian Medical Association Journal, Vol. 167, Issue 7, 2002 October 1, PP. 775–780.

[3] W. C.Hu, J. h.Yeh, H. J. Chu, and C. w. Lee, “Internet-Enabled Mobile Handheld Devices for Mobile Commerce”, Contemporary Management Research, Vol. 01, No. 01, September 2005, PP.13-34.

[4] Y. C. Lu, Y. Xiao, A. Sears, and J.A. Jacko, “A review and a framework of handheld computer adoption in healthcare”, International Journal of Medical Informatics, Vol.74, Issue 5, June 2005, pp. 409-422.

[5] A. A. Economides and N. Nikolaou, “Evaluation of Handheld Devices for Mobile Learning”, International Journal of Engineering Education, 2008.

[6] S. K. Madria, M. Mohania, S. S. Bhowmick, and B. Bhargava, “Mobile Data and Transaction Management”, Information Sciences Journal, Vol.141, Apr 2002, PP. 279-309.

[7] R. B. Patel, and Anu, "A Mobile Transaction System for Open Networks," BVICAM's International Journal of Information Technology, Vol.1, Issue 1, January-July 2008, PP. 67-78.

[8] L. Yin, G. Cao and Y. Cai, “A Generalized Target-Driven Cache Replacement Policy for Mobile Environments”, Journal of Parallel and Distributed Computing, Vol. 65, No. 5, 2005, pp. 583-594.

[9] E. Pitoura, and P. K. Chrysanthis, “Caching and Replication in Mobile Data Management”, IEEE Data Eng, 2007, PP. 13-20.

[10] S. Yi, S. Jung, and J. Suh, “Better mobile client's cache reusability and data access time in a wireless broadcast environment”, Data & Knowledge Engineering, Vol. 63, Issue 2, November 2007, PP. 293–314.

[11] H. K. Chavan, and S. D. Patil, “A Survey of Mobile Database Cache Replacement Policies”.

[12] S. Fageeri1, R. Ahmad, and A. A. Mu'azu, “ Mobile Data availability Similarity between Cooperative Caching and Replication”, IOSR Journal of Computer Engineering, Vol. 1,Issue 4, May-June 2012, PP. 09-13.

[13] M. Akon, M.T. Islam, X. Shen, and A. Singh, “A Bandwidth and Effective Hit Optimal Cache Scheme for Wireless Data Access Networks with Client Injected Updates”, Computer Networks, Vol.56, Issue 7, 2012, PP. 2080–2095.

[14] I. CHUNG, L. GRUENWALD, C.HWANG, and S.Y. JUNG, “Adaptive Transaction Management Protocols for Mobile Client Caching DBMSs”, Journal of Information Science and Engineering, Vol. 18, No.3, May 2002, PP.445-457.

[15] J. Xu, Q. Hu, W. Lee, and D. Lee, “Performance Evaluation of an Optimal Cache Replacement Policy for Wireless data Dissemination”, IEEE Transactions on Knowledge and Data Engineering, Vol.16, No. 1, Jan. 2004, pp. 125-139.

[16] Y. Du, S. K. S. Gupta, G. Varsamopoulos, “Improving On-Demand Data Access Efficiency in MANETs with Cooperative Caching”, Elsevier Ad Hoc Journal, No.7, May 2009, PP.579–598.

[17] N. Dimokas, D. Katsaros, L. Tassiulas, and Y. Manolopoulos, “High performance, low complexity cooperative caching for wireless sensor networks”, Wireless Networks, Vol.17, Issue 3, April 2011, PP.717-737.

[18] J. Xu, Q. L. Hu, D.L. Lee, W. C. Lee, “SAIU: An Efficient Cache Replacement Policy for Wireless on-Demand Broadcasts,” Proc. Ninth ACM Int’l Conf. Information and Knowledge Management, Nov. 2000, PP. 46-53.

[19] N. Chand, R.C. Joshi and M. Misra, “Cooperative Caching Strategy in Mobile Ad Hoc Networks Based on Clusters”, International Journal of Wireless Personal Communications special issue on Cooperation in Wireless Networks, Vol.43, Oct 2007, PP.41–63.

[20] N. Chand, R.C. Joshi, and M. Misra, “"Energy Efficient Cache Replacement in Mobile Computing Environment”, Ubiquitous Computing and Communication Journal, 2009, pp. 1-12

[21] S. Pant, N. Chauhan, and P. Kumar, “Effective Cache based Policies in Wireless Sensor Networks: A Survey”, International Journal of Computer Applications, Vol.11, No.10, December 2010.

[22] L. Jia, “A Mobile Ad-hoc Network Data Cache Invalidation Method”, Procedia Engineering, Vol.23, 2011, PP.150–154.

Mohamed Osama Khalil Grida is an assistant professor in Industrial Engineering Department in Faculty of Engineering in Zagazig University.

Hasnaa Raafat Hamed Soliman is a teaching assistant in Information System Department in Faculty of Computers and Informatics in Zagazig University.

Mohamed Monier Hassan is a professor in Information System Department in faculty of Computers and Informatics in Zagazig University.