Information Retrieval Quantification of a Mobile Agent-Based Adaptive Information Service System

I Ahmed, M. J Sadiq

Department of Electrical and Computer Engineering
International Islamic University, Malaysia, 53100

a.iftikhar@iiu.edu.my, jafars@iiu.edu.my

Abstract

Background: As we migrate from platform-centric computing to network-centering computing, the current state of wide area network technologies are increasingly becoming inefficient to cope with the high demand of service provision and utilization. Users now look forward to anytime anywhere services on the Internet. Various techniques are being researched to address the issue. One such technique involves fading the information contents of a service provider away to various nodes in the network and let the users retrieve the relevant information by sending mobile agents. The previous work reported lacks the simulation of a network environment.
Methods: A wide are network environment was created by programming a simulation engine capable of simulating Faded Information Field (FIF) Architecture and its attributes. A number of information fading algorithms were tested to characterize the function of information fading and the subsequent retrieval in the information service system.
Results: The results show that the information dissemination on various nodes in the FIF can be preset by using a particular algorithm depending on the nature of the information service being provided on the network resulting in reducing the network congestion
Conclusions: The FIF may enable a user to find information closest to his/her location. This is particularly advantageous in an e-commerce application, but equally applicable to other areas of information provision and utilization as well. It was concluded that information fading does not have significant impact on retrieval effectiveness, but can reduce network traffic.
Keyword: Mobile agents, Information retrieval, wide area networks, electronic commerce.

I. Introduction

Information in today's business environment has become the prime ingredient of a successful business enterprise. The rapid advances in information and communication technology (ICT) have opened new avenues to experiment with e-commerce and information service provision and utilization on wide area networks. The traditional paradigms of business, learning and other essential services are undergoing an evolution. The internet is a huge data repository that is expanding without a central authority. The number of worldwide internet users is predicted to exceed one billion by the end of 2005 [1]. Moreover 300 terabytes of information is published online every year [2]. As the performance demands of the internet and its usage increase exponentially, the emphasis is shifting from platform-centric computing to network-centric computing. As ICT advances, the dynamics of e-commerce tend to be more data intensive and complex. Companies have to comprehend the trends and demands of the users quickly in order to survive in the competitive environment of today. The expectations of users and customers have soared high and they seek a flawless service – anytime, anywhere. In addition, these systems are expected to provide high assurance coupled with fault tolerance and timeliness [3,4].

It is becoming increasingly inefficient to handle the complexities of data-intensive computing by the client-server model. New techniques are needed to keep pace with the changing dynamics of information services. Fading of information around a service provider is one such techniques reported recently [5]. In faded information field (FIF) architecture, autonomous mobile agents are employed by the information provider to disseminate the information contents on the network nodes. The information seeker sends mobile agents to seek the desired information. The FIF has been researched with respect to its various attributes [6-9]. We evaluated the information dissemination process in a FIF by proposing a number of algorithms. Unlike the queuing theory model used in the previous work reported [5], a wide area network environment was programmed and the proposed algorithms were simulated to quantify the information discovery through mobile agents. The subsequent results indicated that each algorithm could be more efficiently employed in a particular application compared to the other.

This paper is organized as follows. The critical performance requirements of an adaptive information system will be presented in section II followed by a survey on mobile agents in section III. The algorithmic evaluation of Push MAs and simulation results will be discussed in section IV. The information fading strategies will be discussed in section V followed by conclusion in section VI.

II. Critical Performance Requirements of an Adaptive Information System

Internet services are gradually creeping into our daily lives with the number of users constantly on the rise. Information systems now involve a constantly changing environment where stringent performance demands are placed on the network. Businesses are taking advantage of the internet facility to offer bargains and promotions and put various items on sale, resulting in users having much more flexibility and choice of bargains.

The current information services on wide area networks are required to handle very high data rates and therefore these are required to be dynamic, efficient and effective. The information environment and the search system form the two essential components of an information service domain. A conventional communication interface linked these two components in the past [10]. Appropriate integration of these components is therefore needed to establish a dynamic interface between the user and the service provider for optimum provision and utilization of information services.

It is thus imperative for an information system to respond to both user needs and service provider (SP) requirements to respond in time to these needs in a rapidly changing environment. The main desirable attributes of such a system therefore should be flexibility, reliability and quick reaction time. Only a system with these properties can successfully satisfy both users and SPs in a dynamic network environment

The traditional data retrieval methodologies focused on the optimization of digging in a huge data base to satisfy a search request. On the contrary, the FIF employs the concept of autonomous provision and utilization of information based on mobile agents. The amount of information to be faded away from SP is a function of network conditions like congestion, popularity of information contents and any other criteria programmed into mobile agents [5].

A. The FIF Architecture

The information in a FIF is distributed on various nodes in the system. The information pattern in the field is analogous to the electromagnetic radiation radiated from an antenna, high in intensity near

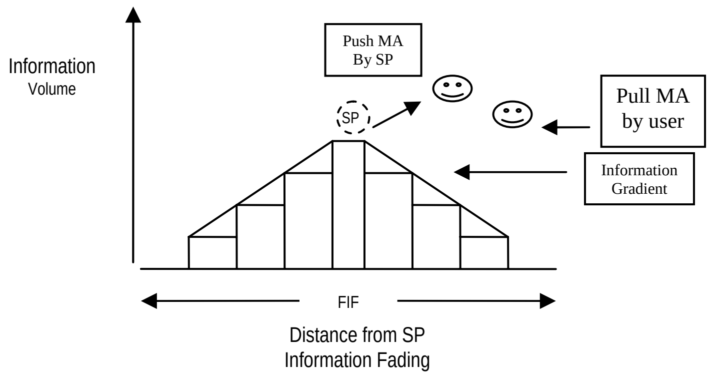

Fig 1. The general schematic of FIF architecture.

the antenna and low in intensity away from the antenna. The SP fades away the information contents to the nodes in the field with the richness in information contents inversely proportional to the distance of the node from the SP through push mobile agents (MAs). The user on the other hand sends out Pull MAs seeking the desired information [5,11]. The general schematic of FIF architecture is depicted in figure 1. The SP is located in the center of the FIF with network nodes (servers) around it making a field of information. The boundary of the FIF is defined by the outermost nodes. The interaction between the users and service providers is carried out by using mobile agents. The fading of information contents by the service provider is shown by the information gradient falling off in value from the SP.

B. FIF system components

The system essentially consists of logically connected nodes through which users and service providers correspond. Mobile agents are used by both parties to acquire and provide information respectively, under evolving/changing situations. The mobile agents (MA) generated by service providers are termed as push mobile agents (Push MAs). Push MAs carry out the function of autonomous coordination and negotiation with other nodes for information fading according to the network traffic status and the level of importance attached to the particular information contents. The level of importance of particular information content is based on its popularity, determined from a high hit rate of query by the information seeker on the network. Push-MAs are sent periodically to update the field. Thus, new information will not need to be indexed, but can be readily made available in this recursive pruning process in the information service system.

Once an information field has been created, users need a means to retrieve this information. Since there is no central index for the field, the users send their own agents to traverse the field and gather information. The pull mobile agents (Pull MAs) are generated by users and they autonomously navigate in search of the required information on the network nodes in a step by step fashion. Since a pull-MA is autonomous, it can navigate the network until it meets its user's requirements, even if the user goes offline. The pull-MA can simply deliver the message through other means, or when the user comes back online.

Once the required information is located, these agents report back to the information seeking source. The push and pull MAs have no direct correspondence with each other. A user will issue a pull-MA with his/her search criteria. The degree of complexity of pull-MAs may vary to cater for both advanced and less advanced users searching for the information. This owes to the inherent flexibility of mobile agents being software programs. Thus, pull-MAs can potentially search for exactly what the user wants and filter out the unwanted sites, resulting in an efficient search.

The third important subsystem of a FIF is the node itself. An information provider sends out push-MAs, which distribute the information provider's information to a group of computers in the network termed as "nodes". It is a platform for both storage of information and program execution. The users seek the desired information by sending out pull MAs that visit the network nodes. It monitors the local information-based system conditions and autonomously makes decisions for allocation requests by the SP. The information contents are handled at the node by a resident software program providing the necessary execution and navigation environment for mobile agents [12-13]. The nodes are responsible for negotiating with the push MA to decide what and how much information is to be stored on the node. The nodes that store a server's information constitute the faded information field of the server. These nodes help reduce congestion as many user requests will be satisfied at the nodes, reducing the number of requests made to the service provider and thus reducing the load on the service provider. Each subsystem is autonomous in terms of control to execute its operations and coordination with other nodes under evolving network conditions.

It may be pointed out that a node may be shared among many information providers, who may all use the node as part of their own faded information field. In such a case, faded information fields may overlap, resulting in an aggregated global faded information field.

C. Communication Format in the FIF

As the user demand for information search is increasingly becoming dynamic, the various traits of a heterogeneous network environment; the state of the nodes, the status of the SP and the network traffic becomes highly unpredictable. A different form of communication is therefore for communication among the entities in such a changing network environment. FIF employs the Content Code Communication format [5] as illustrated in figure 3. The information about a university is structured according to the degree of importance. SP1 specifies the university name followed by CM1 indicating the location of university with respect to a particular country and city. CM2 depicts the programs offered by the university. The message format components thus lead to the breakdown of the principal source of information available on the SP into its uniquely defined characteristic codes CMs. Push MAs are sent out in the FIF by SPs specifying Content Codes (CMs) of information to the nodes using the message format as shown in Figure 2. The selection of information storage/allocation is autonomously carried out by the nodes based on CMs. Similarly the Pull MAs sent out by the users search for the required information based on CMs.

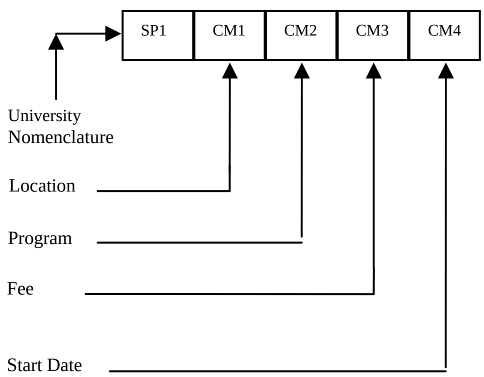

Fig 2. The message format in FIF architecture.

III. Why Mobile Agents

A distributed environment is most suitable for the employment of mobile agents. The details of such applications in a distributed environment can be found elsewhere [14-16]. The term mobile agent is often context-dependent and has two separate and distinct concepts: mobility and agency. The term agency implies having the same characteristics as that of an agent. These are self-contained and identifiable computer programs that can move within the network from node to node and act on behalf of a user or other entity. These can halt execution from a host without human interruption [17]. A mobile agent does not need to maintain communication with its source. Therefore, it is an autonomous entity. It may have some or all of the following properties [18].

i. *Autonomous* – The program exercises control over its actions.
ii. *Reactive* –It responds in a timely fashion to changes in environment.
iii. *Goal-oriented* – It does not simply react to the environment.
iv. *Communicative* – It communicates with other agents or with devices.
v. *Adaptive* – It changes its behavior based on past experience.
vi. *Mobile* – It is able to transfer itself from one host to another.
vii. *Flexible* – It is able to cope with unexpected situations.

The current distributed network environment is based on the traditional client-server paradigm. In the case of mobile agents employed in a network, the service provision and utilization can be distributed in nature and is dynamically configured according to the changing network performance metrics like congestion. Mobile agents are typically suited to applications requiring structuring and coordinating wide area network and distributed services involving a large number of remote real time interactions [19].

IV. Information System Simulation Parameters

The FIF structure was simulated using the following parameters:
*A total of 50 web servers acting as service providers (SPs) in the FIF.
* Each web server stores between 7 and 10 categories of information out of a total of 50 categories of information.
*Routers were generated at random using polar coordinates with maximum radius specified as 50 units. A total of 200 routers were created in the simulated information system.
* The inter-nodal distance is computed by using Djikstra's algorithm [20] which is a fair approximation in a WAN. Then, depending on the field determination algorithm, the maximum distance that a push-MA is required to hop among nodes was computed from the Djikstra costs.

V. Information Fading Strategies – Algorithms and Simulations

Since the information is faded in the FIF away from the SP and it is relatively easier for the push MAs to reach the nearby nodes, these are more frequently updated than the further nodes and thus contain the up-to-date information contents. Information fading therefore has the effect of increasing the reliability of information on the network. The degree of information fading depends on the distance of a node from the SP. We propose three algorithms that we later tested and evaluated by programming a wide area network simulation engine. These algorithms simulated the behavior of the SP in a wide area network information retrieval application by fading the information contents thru push MAs. In all the three cases the nodes randomly decide the nature of information contents to store in their memory after negotiation with push-MAs.

- **Straight Line Fading**– The information is faded proportionally with the distance away from the SP, where a node at zero distance would have 100% of the information and a node at the FIF edge would have 0% of the information. However, since the furthest node of the FIF is still a member of the FIF, it should still retain a portion of the information. The following function was used for the purpose:

P = (L – 100) / W * D + 100

Where P = percentage of information of the push-MA that is stored, L = minimum percentage of information that a node in the FIF is required to store, D = distance of the current node from the source and W = width of the FIF. The resultant simulated distribution is shown in Figure 3.

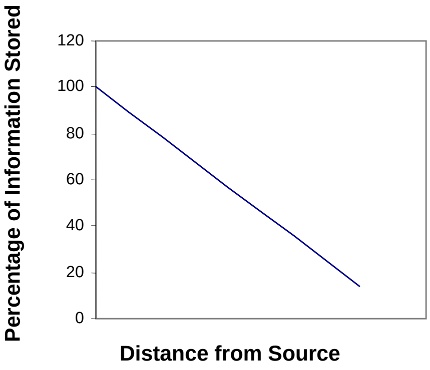

Figure 3. Straight line fading.

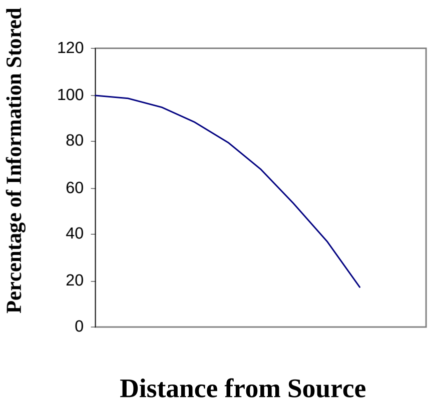

Figure 4. The quadratic fading.

- **Quadratic –** In order to increase the volume of information available in the FIF, the fading may be carried out with a quadratic function. In this strategy, each node in the FIF stores more information than the straight line fading as a result of negotiation with push MA. Thus the likelihood of finding the information by the user sending out a pull MA is increased. The following function was used to realize the quadratic fading;
  $P = (1 - (D / (M * W))^2) * 100$ where M = multiplier to increase field size so that the perimeter node stores at least some of the push-MA’s information. The fading function is shown graphically in figure 4.
- **No Fading –** The extreme case to ensure the highest probability of information discovery would involve all the information held by the SP to be stored in every node in the FIF, implying that none of the information content is faded on the nodes. The nodes will therefore need to allocate more space for the storage of information than the straight line and quadratic fading strategies.

The performance of the fading strategies was observed by simulations. There are two important aspects of information discovery in this analysis; the percentage of requests satisfied by the system and the percentage of information discovered. The simulation results are depicted in Figures 5 and 6 Percentage respectively.

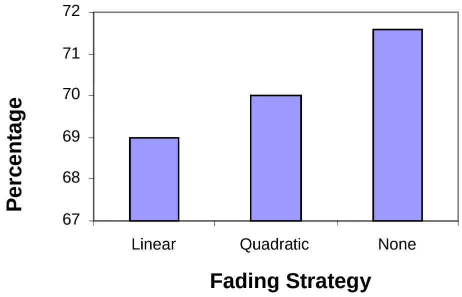

Figure 5.Percentage of requests satisfied

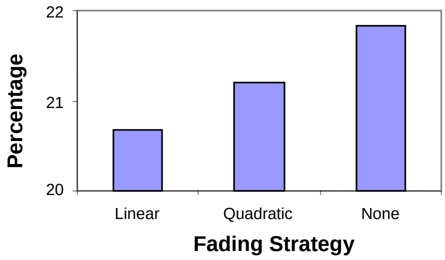

Figure 6. Percentage of information discovered

These results are the averages of simulations with different network parameters in the FIF. As expected, the "no fading" strategy gives the best results, however, it is not significantly better than the other two algorithms. It is less processor intensive since there is no calculation required to store the information on a node. The reason; as the SP continues to send push-MAs to various nodes in the FIF, eventually more and more information is stored in the nodes until the other strategies approximate a situation with no fading.
It was observed that only 21% of the information requested by the user is discovered, although a larger percent of requests are satisfied. Increasing the number of pull-MA hops can alleviated the problem as it would increase the probability of finding the required information. In the case of random strategy employed in another simulation run, the pull-MA chooses its next node at random from all available unvisited nodes in the FIF. The results are depicted in figure 7 and figure 8.

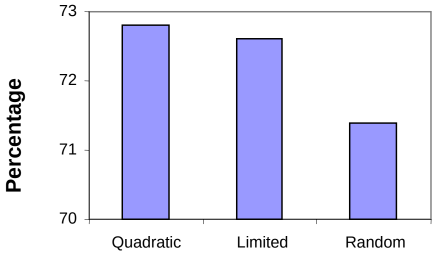

Figure 7. Percentage of requests satisfied

This algorithm is inspired by the discovery that random selection of pages in a CPU's cache is as good as other algorithms in improving the cache's performance. This is the least computationally intensive strategy. The results show that using the quadratic function to choose hops is the best, but the difference in performance between the worst algorithm – random – and the best is only approximately 1% which is not significant. The random strategy is the least computationally intensive since there are no complex computations involved.

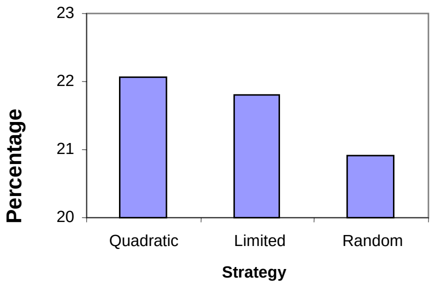

Figure 8. Percentage of Information Discovered.

VI. Conclusion

Information fading by the service provider in a faded information field has been characterized in this paper in addition to a brief review of the technique. The FIF may enable a user to find information closest to his/her location. This is particularly advantageous in an e-commerce application where different vendors in a geographical area fairly distribute their product information on a wide area FIF. However, the concept is equally applicable to other areas of information provision and utilization as well.

The fading of information was tested by the proposal of three fading algorithms. These were subsequently simulated to quantify their effectiveness in the information retrieval function. It is concluded that the three algorithms did not differ much in discovering the requisite information. For instance the difference in the output from the quadratic and random strategies turned out to be around 1%. The random strategy is the least computationally intensive since there are no complex computations involved. Its use in the information search application can therefore, reduce the processing overhead required in the other proposed algorithms. It is also worth noting that increasing the number of pull MA hops on various nodes in the system also helps in increasing the probability of finding the requisite information. Also, not all information providing businesses possess information-rich contents like the basic knowledge of an entity whether available on the network. In that case the ‘no fading strategy’ would be the optimum to use. Thus information fading does not have significant impact on retrieval effectiveness, but can reduce network traffic.

This research was carried out in the information provision and utilization domain of a FIF. However there are a number of issues which could be investigated in future research. The management of various FIFs overlapping with each other’s fields for example could be a starting point. In addition the inter-cooperation between various service providers in terms of sharing their FIFs can also be researched. Further research may involve the characterization of a fully integrated business information system with simulation of the system performance with respect to data throughput handling and efficiency.

References

[1] Wallys W. Conhaim,” The Global Net,” Information today, 2004, Vol.21, Issue 9, pp.35-39.

[2] P. Lyman and H. R. Varian, “How much Information” Journal of Electronics Publishing, December 2000, Vol. 6, Issue 2.

[3] D. B Ingham and S. K Shrivatava,” Constructing Dependable Web Services,” IEEE Computing Magazine, pp. 25-33, Jan-Feb, 2000.

[4] I. L Yen et al,” Toward Integrated Methods for High Assurance Systems,” IEEE Computer Magazine, pp. 32-34, April 1998

[5] H. F. Ahmad, H. Arfaoui and K. Mori, “Autonomous Information Fading by Mobile Agents for improving User’s Access Time and Fault Tolerance,” Proceedings, 5th International Symposium on Autonomous Decentralized Systems,2001, pp. 279-283.

[6] I .Ahmed, J Sadiq," Characterization of Information Distribution Fairness in Mobile Agent-Based Adaptive Information System" presented in the Sixth International Conference on Web Integration and Services, September 27-29, 2004, Jakarta, Indonesia.

[7] I Ahmed, J. Sadiq," Nodal Congestion Analysis of an Autonomous Mobile Agent-Based Distributed Information Service System for Efficient Utilization of Information Contents," International Journal of Information Technology, Vol.10, No.1, 2004.

[8] I .Ahmed, J Sadiq," A Novel Autonomous Architecture for the Integration of Information and control in an Automated Manufacturing System," Presented in the International Conference on Intelligent Agents, Web Technologies and Internet Commerce, 12-14 July 2004, Gold Coast Australia.

[9] I Ahmed," Balancing the user demand and information provision in a Web-based linguistic learning system employing Faded Information Field Architecture," presented in the ASIA CALL International Conference on Information Technology and Linguistic Learning, Bangkok, Thailand, 3-5 December 2003.

[10] A. Bestavres,” WWW Traffic Reduction and Load Balancing through Server-Based Cashing,” IEEE Concurrency Magazine, pp. 56-67, Jan-March, 1997.

[11] H. Arafoui and K. Mori," Autonomous Navigation Architecture for Load Balancing User demands in Distributed Information Systems," IEICE Transactions on Communications, E84-B(10), pp. 1085-1093, 2001.

[12] M. Dalmeijer et al,” A Reliable Mobile Agent Architecture,” Proceedings OORTDCS, pp. 64-72, IEEE, 1998.

[13] F. A Silva et al,” An Approach for providing mobile Agent Fault tolerance,” Springer, 1998

[14] A. Karmouch and V. A. Pham, "Mobile Software Agents: An Overview," IEEE Communications Magazine, Vol. 36, No.7, July 1998.

[15] M. K. Perdikeas et al, "Mobile Agents Standards and Available Platforms," Comp. Net., Vol. 31, No. 19, pp.1999-2016, August 1999.

[16] D. B. Lange and M. Oshima, "Seven Good Reasons for Mobile Agents," Communications of the ACM, Vol. 42, No. 3, March 1999, pp. 88-89, March 1999

[17] J. Cao, G. H. Chan, W. Jia, and T. Dillon, "Check-pointing and Rollback of wide-area Distributed Applications using Mobile Agents,” Proceedings, International Parallel and Distributed Processing Symposium. San Francisco: IEEE Computer Press. April 2001.

[18] Covaci, S. 1999. Autonomous Agent Technology. Proceedings of The Fourth International Symposium on Autonomous Decentralized Systems. pp. 255-257.

[19] Applications using Mobile Agents," Proceedings International Parallel and Distributed Processing Symposium. San Francisco: IEEE Computer Society Press, April 2001.

[20] Streenstrup, M. 1995. Routing in Communications Networks. New Jersey: Prentice Hall.

Dr Iftikhar Ahmed graduated in Avionics Engineering from NED University of Engineering & Technology Pakistan in 1981. He then served as an engineer in Government of Pakistan till 1987 when he was awarded a National Talent Scholarship by The Ministry of Science & Technology Pakistan for higher studies abroad. He was awarded a MSc in Optical Electronics and a PhD in Electrical Engineering by Strathclyde University UK in 1989 and 1993 respectively. He served in various R&D Establishments in Pakistan and worked as an Associate Professor in Computer Science Department Bahria University Pakistan for two years before moving to Malaysia in 2000 where he is currently serving as an Associate Professor in Computer Engineering Department International Islamic University. His research interests include autonomous mobile agents in information service systems, fault tolerant computing and autonomous avionics systems.

Muhammad Jafar Sadeq graduated with a Bachelor's in Computer and Information Engineering from International Islamic University Malaysia in 2000. He is currently undergoing a Master's of Science Program in Computer and Information Engineering in International Islamic University Malaysia. His research interests include distributing computing and adaptive information systems.