

**International Journal of
Computer Science and Security
(IJCSS)**

ISSN : 1985-1553



VOLUME 3, ISSUE 1

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

Editor in Chief Dr. Haralambos Mouratidis

International Journal of Computer Science and Security (IJCSS)

Book: 2009 Volume 3, Issue 1

Publishing Date: 30- 02 - 2009

Proceedings

ISSN (Online): 1985 -1553

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

IJCSS Journal is a part of CSC Publishers

<http://www.cscjournals.org>

©IJCSS Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers

Table of Contents

Volume 3, Issue 1, January/February 2009.

Pages

- 1 -15 Integration of Least Recently Used Algorithm and Neuro-Fuzzy System into Client-side Web Caching
Waleed Ali Ahmed, Siti Mariyam Shamsuddin
- 16 - 22 A Encryption Based Dynamic and Secure Routing Protocol for Mobile Ad Hoc Network
Rajender Nath , Pankaj Kumar Sehgal
- 23 - 33 MMI Diversity Based Text Summarization
Mohammed Salem Binwahlan, Naomie Salim , Ladda Suanmali
- 34 - 42 Asking Users: A Continuous Evaluation on Systems in a Controlled Environment
Suziah Sulaiman, Dayang Rohaya Awang Rambli, Wan Fatimah Wan Ahmad, Halabi Hasbullah, Foong Oi Mean, M Nordin Zakaria, Goh Kim Nee, Siti Rokhmah M Shukri
- 43 - 61 Exploring Knowledge for a Common Man through Mobile Services and Knowledge Discovery in Databases
Mayank Dave, S. B. Singh, Sanjeev Manchanda

62 - 75

Testing of Contextual Role-Based Access Control Model (C-RBAC)

Muhammad Nabeel Tahir

Integration of Least Recently Used Algorithm and Neuro-Fuzzy System into Client-side Web Caching

Waleed Ali

*Faculty of Computer Science and Information System
Universiti Teknologi Malaysia (UTM)
Skudai, 81310, Johor Bahru, Malaysia*

prowalid_2004@yahoo.com

Siti Mariyam Shamsuddin

*Faculty of Computer Science and Information System
Universiti Teknologi Malaysia (UTM)
Skudai, 81310, Johor Bahru, Malaysia*

mariyam@utm.my

ABSTRACT

Web caching is a well-known strategy for improving performance of Web-based system by keeping web objects that are likely to be used in the near future close to the client. Most of the current Web browsers still employ traditional caching policies that are not efficient in web caching. This research proposes a splitting client-side web cache to two caches, short-term cache and long-term cache. Initially, a web object is stored in short-term cache, and the web objects that are visited more than the pre-specified threshold value will be moved to long-term cache. Other objects are removed by Least Recently Used (LRU) algorithm as short-term cache is full. More significantly, when the long-term cache saturates, the neuro-fuzzy system is employed in classifying each object stored in long-term cache into either cacheable or uncacheable object. The old uncacheable objects are candidate for removing from the long-term cache. By implementing this mechanism, the cache pollution can be mitigated and the cache space can be utilized effectively. Experimental results have revealed that the proposed approach can improve the performance up to 14.8% and 17.9% in terms of hit ratio (HR) compared to LRU and Least Frequently Used (LFU). In terms of byte hit ratio (BHR), the performance is improved up to 2.57% and 26.25%, and for latency saving ratio (LSR), the performance is improved up to 8.3% and 18.9%, compared to LRU and LFU.

Keywords: Client-side web caching, Adaptive neuro-fuzzy inference system, Least Recently Used algorithm.

1. INTRODUCTION

One of the important means to improve the performance of Web service is to employ web caching mechanism. Web caching is a well-known strategy for improving performance of Web-based system. The web caching caches popular objects at location close to the clients, so it is considered one of the effective solutions to avoid Web service bottleneck, reduce traffic over the Internet and improve scalability of the Web system[1]. The web caching is implemented at client, proxy server and original server [2]. However, the client-side caching (browser caching) is

economical and effective way to improve the performance of the Word Wide Web due to the nature of browser cache that is closer to the user [3,4].

Three important issues have profound impact on caching management namely: cache algorithm (passive caching and active caching), cache replacement and cache consistency. However, the cache replacement is the core or heart of the web caching; hence, the design of efficient cache replacement algorithms is crucial for caching mechanisms achievement [5]. In general, cache replacement algorithms are also called web caching algorithms [6].

Since the apportioned space to the client-side cache is limited, the space must be utilized judiciously [3]. The term “cache pollution” means that a cache contains objects that are not frequently used in the near future. This causes a reduction of the effective cache size and affects negatively on performance of the Web caching. Even if we can locate large space for the cache, this will be not helpful since the searching for object in large cache needs long response time and extra processing overhead. Therefore, not all Web objects are equally important or preferable to store in cache. The setback in Web caching consists of what Web objects should be cached and what Web objects should be replaced to make the best use of available cache space, improve hit rates, reduce network traffic, and alleviate loads on the original server.

Most web browsers still concern traditional caching policies [3, 4] that are not efficient in web caching [6]. These policies suffer from cache pollution problem either cold cache pollution like the least recently used (LRU) policy or hot cache pollution like the least frequently used (LFU) and SIZE policies [7] because these policies consider just one factor and ignore other factors that influence the efficiency the web caching. Consequently, designing a better-suited caching policy that would improve the performance of the web cache is still an incessant research [6, 8].

Many web cache replacement policies have been proposed attempting to get good performance [2, 9, 10]. However, combination of the factors that can influence the replacement process to get wise replacement decision is not easy task because one factor in a particular situation or environment is more important than others in other environments [2, 9]. In recent years, some researchers have been developed intelligent approaches that are smart and adaptive to web caching environment [2]. These include adoption of back-propagation neural network, fuzzy systems, evolutionary algorithms, etc. in web caching, especially in web cache replacement.

The neuro-fuzzy system is a neural network that is functionally equivalent to a fuzzy inference model. A common approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS) that has more power than Artificial Neural Networks (ANNs) and fuzzy systems as ANFIS integrates the best features of fuzzy systems and ANNs and eliminates the disadvantages of them.

In this paper, the proposed approach grounds short-term cache that receives the web objects from the Internet directly, while long-term cache receives the web objects from the short-term cache as these web objects visited more than pre-specified threshold value. Moreover, neuro-fuzzy system is employed to predict web objects that can be re-accessed later. Hence, unwanted objects are removed efficiency to make space of the new web objects.

The remaining parts of this paper are organized as follows: literature review is presented in Section 2, related works of intelligent web caching techniques are discussed in Section 2.1. Section 2.2 presents client-side web caching, and Section 2.3 describes neuro-fuzzy system and ANFIS. A framework of Intelligent Client-side Web Caching Scheme is portrayed in Section 3, while Section 4 elucidates the experimental results. Finally, Section 5 concludes the paper and future work.

2. LITERATURE REVIEW

2.1 Related Works on Intelligent Web Caching

Although there are many studies in web caching, but research on Artificial Intelligence (AI) in web caching is still fresh. This section presents some existing web caching techniques based on ANN or fuzzy logic.

In [11], ANN has been used for making cache replacement decision. An object is selected for replacement based on the rating returned by ANN. This method ignored latency time in replacement decision. Moreover, the objects with the same class are removed without any precedence between these objects. An integrated solution of ANN as caching decision policy and LRU technique as replacement policy for script data object has been proposed in [12]. However, the most important factor in web caching, i.e., recency factor, was ignored in caching decision. Both prefetching policy and web cache replacement decision has been used in [13]. The most significant factors (recency and frequency) were ignored in web cache replacement decision. Moreover, applying ANN in all policies may cause extra overhead on server. ANN has also been used in [6] depending on syntactic features from HTML structure of the document and the HTTP responses of the server as inputs. However, this method ignored frequency factor in web cache replacement decision. On other hand, it hinged on some factors that do not affect on performance of the web caching.

Although the previous studies have shown that the ANNs can give good results with web caching, the ANNs have the following disadvantage: ANNs lack explanatory capabilities, the performance of ANNs relies on the optimal selection of the network topology and its parameters, ANNs learning process can be time consuming, and ANNs are also too dependent on the quality and amount of data available [14, 15, 16].

On other hand, [17] proposed a replacement algorithm based on fuzzy logic. This method ignored latency time in replacement decision. Moreover, the expert knowledge may not always available in web caching. This scheme is also not adaptive with web environment that changes rapidly.

This research shares consideration of frequency, recency, size and latency time in replacement decision with some previous replacement algorithms. Neuro-Fuzzy system especially ANFIS is implemented in replacement decision since ANFIS integrates the best features of fuzzy systems and ANNs. On the contrary, our scheme differs significantly in methodology used in caching the web objects, and we concentrate more on client-side caching as it is economical and effective way, primarily due its close proximity to the user [3.4].

2.2 Client-side Web Caching

Caches are found in browsers and in any of the web intermediate between the user agent and the original server. Typically, a cache is located in the browser and the proxy [18]. A browser cache (client-side cache) is located in client. If we examine the preferences dialog of any modern web browser (like Internet Explorer, Safari or Mozilla), we will probably notice a cache setting. Since most users visit the same web site often, it is beneficial for a browser to cache the most recent set of pages downloaded. In the presence of web browser cache, the users can interact not only with the web pages but also with the web browser itself via the use of the special buttons such as back, forward, refresh or via URL rewriting. On other hand, a proxy cache is located in proxy. It works on the same principle, but in a larger scale. The proxies serve hundreds or thousands of users in the same way.

As cache size is limited, a cache replacement policy is needed to handle the cache content. If the cache is full when an object needs to be stored, then the replacement policy will determine which object is to be evicted to allow space for the new object. The goal of the replacement policy is to make the best use of available cache space, to improve hit rates, and to reduce loads on the original server. The simplest and most common cache management approach is LRU algorithm,

which removes the least recently accessed objects until there is sufficient space for the new object. LRU is easy to implement and proficient for uniform size objects such as in the memory cache. However, since it does not consider the size or the download latency of objects, it does not perform well in web caching [6].

Most web browsers still concern traditional replacement policies [3, 4] that are not efficient in web caching [6]. In fact, there are few important factors of web objects that can influence the replacement policy [2, 9, 10]: recency, i.e., time of (since) the last reference to the object, frequency, i.e., number of the previous requests to the object, size, and access latency of the web object. These factors can be incorporated into the replacement decision. Most of the proposals in the literature use one or more of these factors. However, combination of these factors to get wise replacement decision for improving performance of web caching is not easy task because one factor in a particular situation or environment is more important than others in other environments [2, 9].

2.3 Neuro-Fuzzy System and ANFIS

The neuro-fuzzy systems combine the parallel computation and learning abilities of ANNs with the human-like knowledge representation and explanation abilities of fuzzy systems [19]. The neuro-fuzzy system is a neural network that is functionally equivalent to a fuzzy inference model.

A common approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS), which has shown superb performance at binary classification tasks, being a more profitable alternative in comparison with other modern classification methods [20]. In ANFIS, the membership function parameters are extracted from a data set that describes the system behavior. The ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion. Jang's ANFIS is normally represented by six-layer feed forward neural network [21].

It is not necessary to have any prior knowledge of rule consequent parameters since ANFIS learns these parameters and tunes membership functions accordingly. ANFIS uses a hybrid learning algorithm that combines the least-squares estimator and the gradient descent method. In ANFIS training algorithm, each epoch is composed of forward pass and backward pass. In forward pass, a training set of input patterns is presented to the ANFIS, neuron outputs are calculated on the layer-by-layer basis, and rule consequent parameters are identified. The rule consequent parameters are identified by the least-squares estimator. Subsequent to the establishment of the rule consequent parameters, we compute an actual network output vector and determine the error vector. In backward pass, the back-propagation algorithm is applied. The error signals are propagated back, and the antecedent parameters are updated according to the chain rule. More details are illustrated in [21].

3. FRAMEWORK OF INTELLIGENT WEB CLIENT-SIDE CACHING SCHEME

In this section, we present a framework of Intelligent Client-side Web Caching Scheme. As shown in FIGURE 1, the web cache is divided into short-term cache that receives the web objects from the Internet directly, and long-term cache that receives the web objects from the short-term cache.

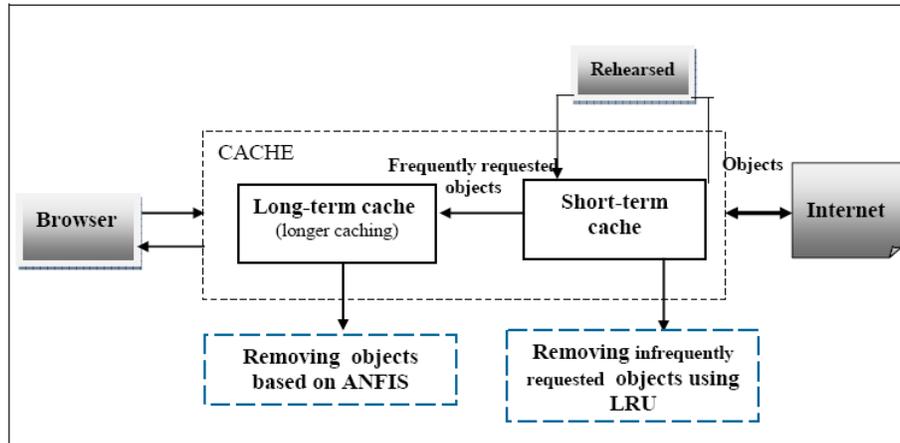


FIGURE 1: A Framework of Intelligent Client-side Web Caching Scheme.

When the user navigates specific web page, all web objects embedded in the page are stored in short-term cache primarily. The web objects that visited more than once will be relocated to long-term cache for longer caching but the other objects will be removed using LRU policy that removes the oldest object firstly. This will ensure that the preferred web objects are cached for longer time, while the bad objects are removed early to alleviate cache pollution and maximize the hit ratio. On the contrary, when the long-term cache saturates, the trained ANFIS is employed in replacement process by classifying each object stored in long-term cache to cacheable or uncacheable object. The old uncacheable objects are removed initially from the long-term cache to make space for the incoming objects (see algorithm in FIGURE 2). If all objects are classified as cacheable objects, then our approach will work like LRU policy.

In training phase of ANFIS, the desired output is assigned to one value and the object considered cacheable object if there is another request for the same object at a later point in specific time only. Otherwise, the desired output is assigned to zero and the object considered uncacheable object.

The main feature of the proposed system is to be able to store ideal objects and remove unwanted objects early, which may alleviate cache pollution. Thus, cache space is used properly. The second feature of the proposed system is to be able to classify objects to either cacheable or uncacheable objects. Hence, the uncacheable objects are removed wisely when web cache is full. The proposed system is also adaptive and adjusts itself to a new environment as it depends on adaptive learning of the neuro-fuzzy system. Lastly, the proposed system is very flexible and can be converted from a client cache to a proxy cache using minimum effort. The difference lies mainly in the data size at the server which is much bigger than the data size at the client.

```

If (size of new object > free space of long-term cache)
{
  // update objects' priorities using ANFIS

  For each object(i) in long-term cache
  { The common attributes of object(i) are forwarded as inputs of the trained
    ANFIS.

    If ( out of ANFIS > 0.5)

      priority of object(i)=1;          // cacheable object

    Else

      priority of object(i)=0;          // uncacheable object

  } // end for

  Do

  { // remove the oldest uncacheable object

    delete the oldest object that has priority=0 form long-term cache

  } while (size of new object > free space of long-term cache);

  If all objects have priority=1

  { Do {

    delete the oldest object form long-term cache

  } while(size of new object > free space of long-term cache);

  }

}

```

FIGURE 2: Intelligent long-term cache removal algorithm based on ANFIS.

4. Experimental Results

In our experiment, we use BU Web trace [22] provided by Cunha of Boston University. BU trace is composed of 9633 files, recording 1,143,839 web requests from different users during six months. Boston traces consist of 37 client machines divided into two sets: undergraduate students set (called 272 set) and graduate students set (called B19 set). The B19 set has 32 machines but the 272 set has 5 machines. In this experiment, twenty client machines are selected randomly from both 272 set and the B19 set for evaluating performance of the proposed method.

Initially, about one month data is used (December for clients from 272 set and January for clients from B19 set) as training dataset for ANFIS. The dataset is divided into training data (70%) and testing data (30%). From our observation; one month period is sufficient to get good training with small Mean Square Error (MSE) and high classification accuracy for both training and testing. The testing data is also used as validation for probing the generalization capability of the ANFIS at each epoch. The validation data set can be useful when over-fitting is occurred.

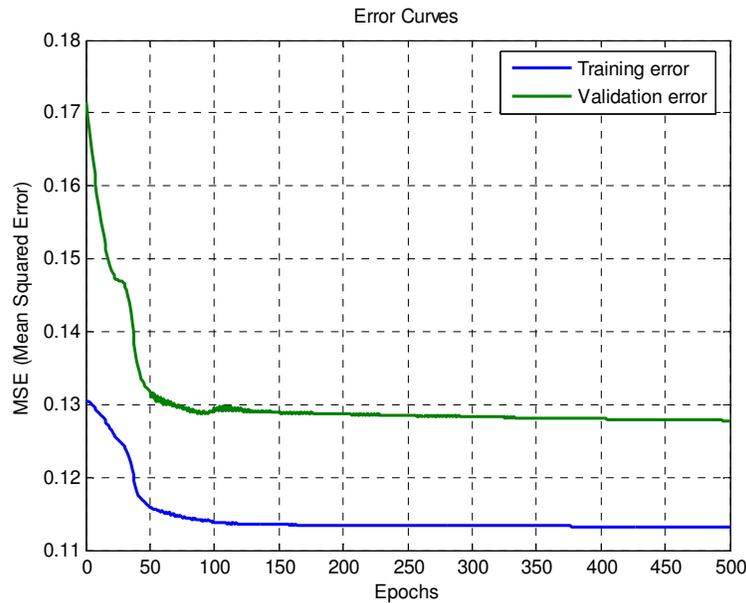


FIGURE 3: Error convergence of ANFIS.

FIGURE 3 shows error convergence of ANFIS for one of client machines, called beaker client machine, in training phase. As shown in FIGURE 3, the network converges very well and produces small MSE with two member functions. It has also good generalization ability as validation error decreases obviously. Thus, it is adequate to use two member functions for each input in training ANFIS. Table 1 summarizes the setting and parameters used for training ANFIS.

| ANFIS parameter | Value |
|-----------------------------------|------------------|
| Type of input Member function(MF) | Bell function |
| Number of MFs | 2 for each input |
| Type of output MF | linear |
| Training method | hybrid |
| Number of linear parameters | 80 |
| Number of nonlinear parameters | 24 |
| Total number of parameters | 104 |
| Number of fuzzy rules | 16 |
| Maximum epochs | 500 |
| Error tolerance | 0.005 |

TABLE 1: Parameters and their values used for training of ANFIS.

Initially, the membership function of each input is equally divided into two regions, small and large, with initial membership function parameters. Depending on used dataset, ANFIS is trained to adjust the membership function parameters using hybrid learning algorithm that combines the least-squares method and the back-propagation algorithm. FIGURE 4 shows the changes of the final (after training) membership functions with respect to the initial (before training) membership functions of the inputs for the beaker client machine.

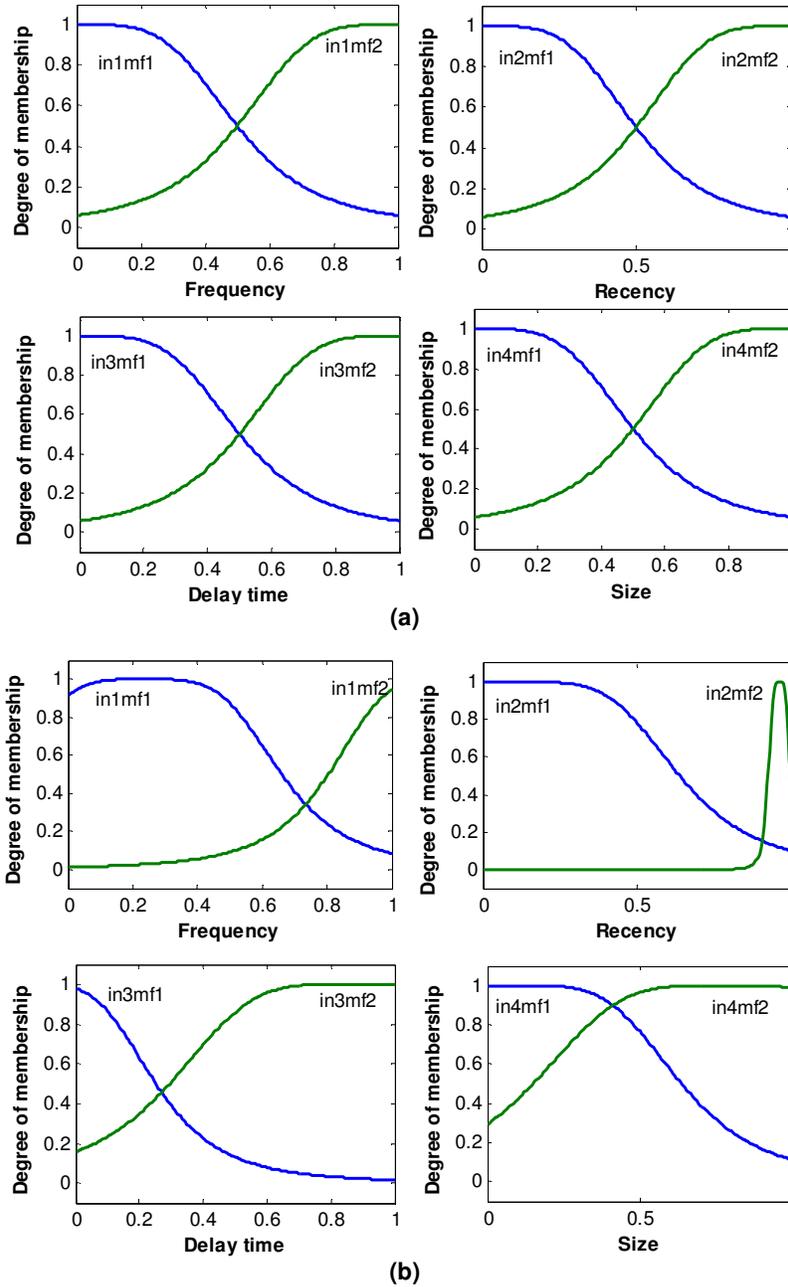


FIGURE 4: (a) Initial and (b) final generalized bell shaped membership functions of inputs of ANFIS.

Since ANFIS is employed in replacement process by classifying each object stored in long-term cache into cacheable or uncacheable object, the correct classification accuracy is the most important measure for evaluating training of ANFIS in this study. FIGURE 5 and FIGURE 6 show the comparison of the correct classification accuracy of ANFIS and ANN for 20 clients in both training and testing data. As can be seen in FIGURE 5 and FIGURE 6, both ANN and ANFIS produce good classification accuracy. However, ANFIS has higher correct classification for both training and testing data in most clients compared to ANN. The results in FIGURE 5 and FIGURE 6 also illustrate that ANFIS has good generalization ability since the correct classification ratios of the training data are similar to the correct classification ratios of the testing data for most clients.

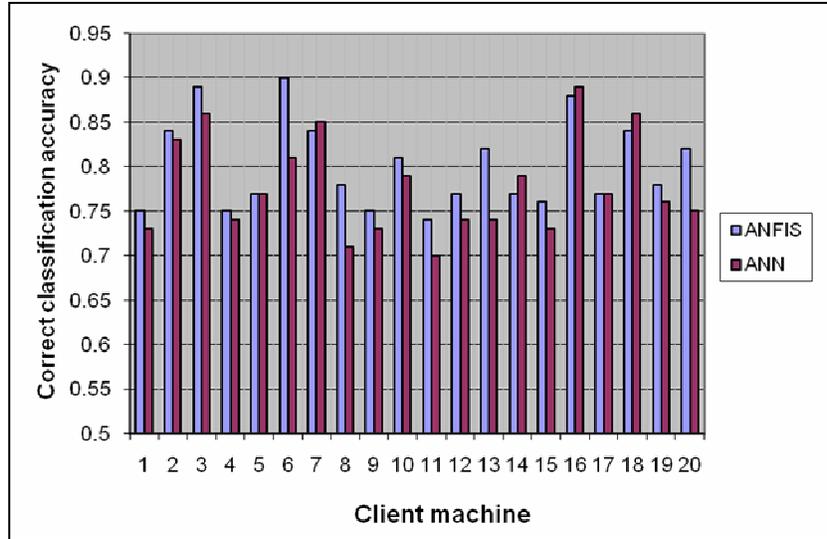


FIGURE 5: Comparison of the correct classification accuracy for training data.



FIGURE 6: Comparison of the correct classification accuracy for testing data.

To evaluate the performance of the proposed method, trace-driven simulator is developed (in java) which models the behavior of Intelligent Client-side Web Caching Scheme. The twenty clients' logs and traces for next two months of the training month are used as data in trace-driven simulator. BU Traces do not contain any information on determining when the objects are unchanged. To simulate correctly an HTTP/1.1 cache, size is a candidate for the purpose of consistency. Thus, the object that has the same URL but different size is considered as the updated version of such object [23].

Hit ratio (HR), Byte hit ratio (BHR) and Latency saving ratio (LSR) are the most widely used metrics in evaluating the performance of web caching [6,9]. HR is defined as the percentage of requests that can be satisfied by the cache. BHR is the number of bytes satisfied from the cache as a fraction of the total bytes requested by user. LSR is defined as the ratio of the sum of download time of objects satisfied by the cache over the sum of all downloading time.

In the proposed method, an obvious question would be the size of each cache. Many experiments were done to show the best size of each cache to ensure better performance. The simulation results of hit ratio of five clients with various sizes of short-term cache are illustrated in FIGURE 7. The short-term cache with size 40% and 50% of total cache size performed the best performance. Here, we assumed that the size of the short-term cache is 50% of the total cache size.

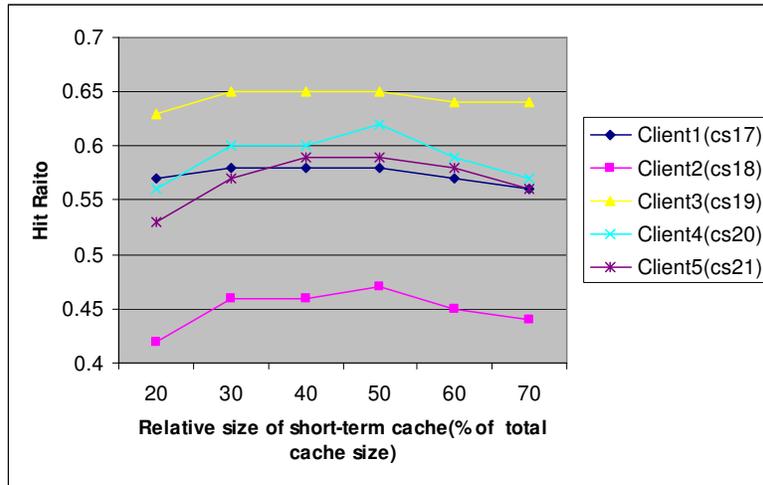


FIGURE 7: Hit Ratio for different short-term cache sizes.

For each client, the maximum HR, BHR, and LSR are calculated for a cache of infinite size. Then, the measures are calculated for a cache of size 0.1%, 0.5%, 1%, 1.5% and 2% of the infinite cache size, i.e., the total size of all distinct objects requested, to determine the impact of cache size on the performance measures accordingly. We observe that the values are stable and close to maximum values after 2% of the infinite cache size in all policies. Hence, this point is claimed as the last point in cache size.

The performance of the proposed approach is compared to LRU and LFU policies that are the most common policies and form the basis of other web cache replacement algorithms [11]. FIGURE 8, FIGURE 9 and FIGURE 10 show the comparison of the average values of HR, BHR and LSR for twenty clients for the different policies with varying relative cache size. The HR, BHR and LSR of the proposed method include HR, BHR and LSR in both short-term cache and the long-term cache.

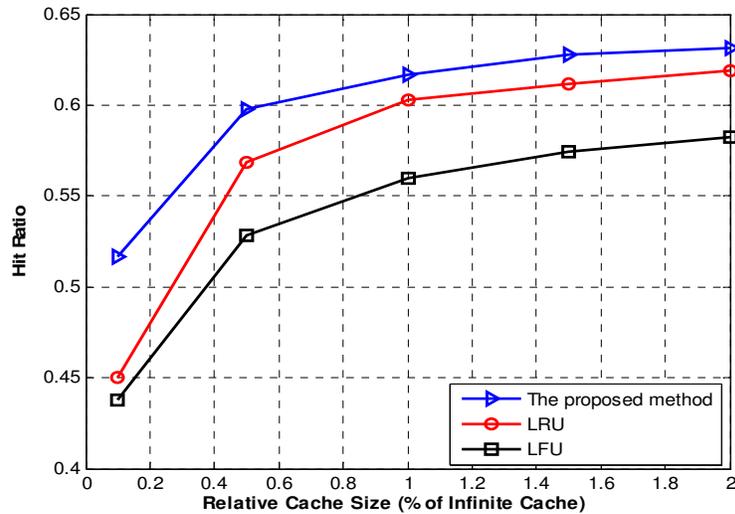


FIGURE 8: Impact of cache size on hit ratio.

As can be seen in FIGURE 8, when the relative cache size increases, the average HR boosts as well for all algorithms. However, the percentage of increasing is reduced when the relative cache size increases. When the relative cache size is small, the replacement of objects is required frequently. Hence, the impact of the performance of replacement policy appears clearly. The results in FIGURE 8 clearly indicate that the proposed method achieves the best HR among all algorithms across traces of clients and cache sizes. This is mainly due to the capability of the proposed method in storing the ideal objects and removing the bad objects that predicted by ANFIS.

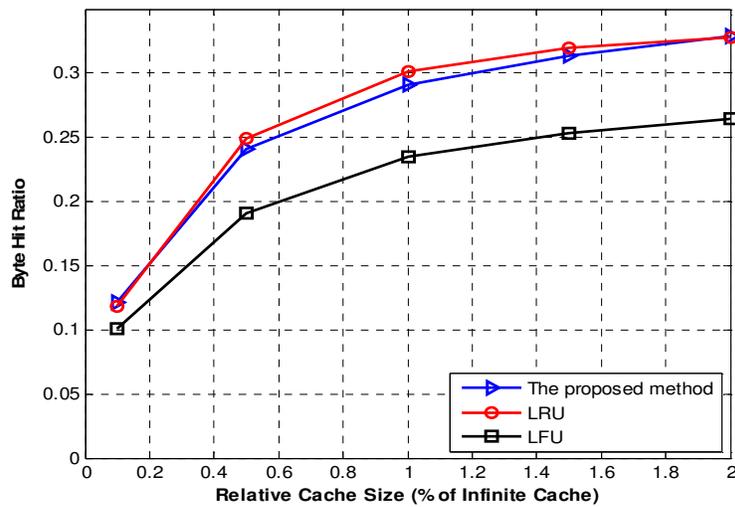


FIGURE 9: Impact of cache size on byte hit ratio.

Although the proposed approach has superior performance in terms of HR compared to all other policies, it is not surprising that BHR of the proposed method is similar or slightly worse than LRU (FIGURE 9). This is because of training of ANFIS with desired output depends on the future request only and not in terms of a size-related cost. These results conform to findings obtained by [11] since we use the same concept as [11] to acquire the desired output in the training phase.

The results also indicate that our approach concerns with ideal small objects to be stored in the cache that usually produces higher HR.

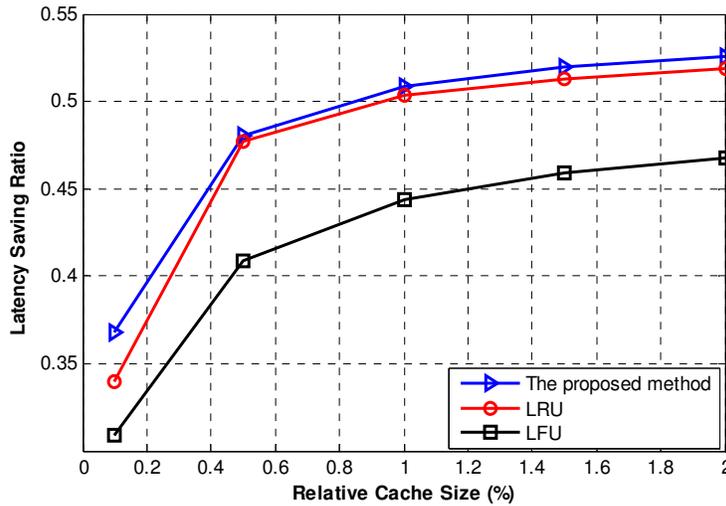


FIGURE 10: Impact of cache size on latency saving ratio.

FIGURE 10 illustrates the average LSR of the proposed method and the common caching schemes as the function of cache sizes. FIGURE 10 depicts that LSR increases rapidly for all policies. However, the proposed method outperforms others policies. LSR of the proposed approach is much better than LFU policy in all conditions. Moreover, LSR of the proposed approach is significantly better than LSR of LRU when the cache sizes are small (less or equal than 0.5% of the total size of all distinct objects). In these situations, many replacements occur and good replacement algorithm is important. On other hand, LSR of the proposed is slightly higher than LSR of LRU when the cache size is larger. Although the desired output through the training phase concern just on future request regardless delay-related cost, the proposed method outperforms the other policies in terms of LSR. This is a result of the close relationship between HR and LSR. Many studies have shown that increasing of HR improves LSR [24, 25].

In all conditions, LFU policy was the worst in all measures because of the cache pollution in objects with the large reference accounts, which are never replaced even if they are not re-accessed again. From FIGURE 8, FIGURE 9 and FIGURE 10, percents of improvements of the performance in terms of HR, BHR and LSR achieved by our approach over the common policies can be concluded and summarized as shown in TABLE 2.

| Relative Cache Size (%) | Percent of Improvements (%) | | | | | |
|-------------------------|-----------------------------|------|------|------|-------|-------|
| | LRU | | | LFU | | |
| | HR | BHR | LSR | HR | BHR | LSR |
| 0.1 | 14.8 | 2.57 | 8.3 | 17.9 | 20.30 | 18.9 |
| 0.5 | 5.2 | - | 0.81 | 13.3 | 26.25 | 17.48 |
| 1 | 2.32 | - | 1.04 | 10.2 | 24.04 | 14.53 |
| 1.5 | 2.58 | - | 1.41 | 9.32 | 24.11 | 13.29 |
| 2 | 1.96 | 0.38 | 1.35 | 8.33 | 20.30 | 12.5 |

TABLE 2: Performance improvements achieved by the proposed approach over common policies.

The results in TABLE 2 indicate that the proposed approach can improve the performance in terms of HR up to 14.8% and 17.9%, in terms of BHR up to 2.57% and 26.25%, and in terms of

LSR up to 8.3% and 18.9% compared to LRU and LFU respectively. TABLE 2 also shows that no improvements of BHR achieved by the proposed approach over LRU policy with relative cache sizes: 0.5%, 1%, and 1.5%.

5. CONCLUSION & FUTURE WORK

Web caching is one of the effective solutions to avoid Web service bottleneck, reduce traffic over the Internet and improve scalability of the Web system. This study proposes intelligent scheme based on neuro-fuzzy system by splitting cache to two caches, short-term cache and long-term cache, on a client computer for storing the ideal web objects and removing the unwanted objects in the cache for more effective usage. The objects stored in short-term cache are removed by LRU policy. On other hand, ANFIS is employed to determine which web objects at long-term cache should be removed. The experimental results show that our approach has better performance compared to the most common policies and has improved the performance of client-side caching substantially.

One of the limitations of the proposed Intelligent Client-side Web Caching Scheme is complexity of its implementation compared to LRU that is very simple. In addition, the training process requires extra computational overhead although it happens infrequently. In the real implementation, the training process should be not happened during browser session. Hence, the user does not fell bad about this training. In recent years, new solutions have been proposed to utilize cache cooperation on client computers to improve client-side caching efficiency. If a user request misses in its local browser cache, the browser will attempt to find it in another client's browser cache in the same network before sending the request to proxy or the original web server. Therefore, our approach can be more efficient and scalable if it supports mutual sharing of the ideal web object stored in long-term cache.

Acknowledgments. This work is supported by Ministry of Science, Technology and Innovation (MOSTI) under eScience Research Grant Scheme (VOT 79311). Authors would like to thank Research Management Centre (RMC), Universiti Teknologi Malaysia, for the research activities and *Soft Computing Research Group (SCRG)* for the support and incisive comments in making this study a success.

6. REFERENCES

1. L. D. Wessels. *"Web Caching"*. USA: O'Reilly. 2001.
2. H.T. Chen. *"Pre-fetching and Re-fetching in Web Caching systems: Algorithms and Simulation"*. Master Thesis, TRENT UNIVESITY, Peterborough, Ontario, Canada, 2008.
3. V. S. Mookerjee, and Y. Tan. *"Analysis of a least recently used cache management policy for Web browsers"*. Operations Research, Linthicum, Mar/Apr 2002, Vol. 50, Iss. 2, p. 345-357.
4. Y. Tan, Y. Ji, and V.S Mookerjee. *"Analyzing Document-Duplication Effects on Policies for Browser and Proxy Caching"*. INFORMS Journal on Computing. 18(4), 506-522. 2006.
5. T. Chen. *"Obtaining the optimal cache document replacement policy for the caching system of an EC website"*. European Journal of Operational Research, Amsterdam, Sep 1, 2007, Vol. 181, Iss. 2; p. 828.
6. T.Koskela, J.Heikkonen, and K.Kaski. *"Web cache optimization with nonlinear model using object feature"*. Computer Networks journal, elsevier , 20 December 2003, Volume 43, Number 6.

7. R .Ayani, Y.M. Teo, and Y. S. Ng. "Cache pollution in Web proxy servers". International Parallel and Distributed Processing Symposium (IPDPS'03). ipdps, p. 248a,2003.
8. C. Kumar, and J.B Norris. "A new approach for a proxy-level Web caching mechanism. *Decision Support Systems*". 46(1), 52-60. Elsevier Science Publishers. 2008.
9. A.K.Y. Wong. "Web Cache Replacement Policies: A Pragmatic Approach". IEEE Network magazine, 2006 , vol. 20, no. 1, pp. 28–34.
10. S.Podlipnig, and L.Böszörmenyi. "A survey of Web cache replacement strategies". ACM Computing Surveys, 2003, vol. 35, no. 4, pp. 374-39.
11. J.Cobb, and H.EIAarag. "Web proxy cache replacement scheme based on back-propagation neural network". Journal of System and Software (2007),doi:10.1016/j.jss.2007.10.024.
12. Farhan. "Intelligent Web Caching Architecture". Master thesis, Faculty of Computer Science and Information System, UTM University, Johor, Malaysia, 2007.
13. U.Acharjee. "Personalized and Artificial Intelligence Web Caching and Prefetching". Master thesis, Canada: University of Ottawa, 2006.
14. X.-X . Li, H .Huang, and C.-H .Liu." *The Application of an ANFIS and BP Neural Network Method in Vehicle Shift Decision*". 12th IFToMM World Congress, Besançon (France), June18-21, 2007.M.C.
15. S.Purushothaman, and P.Thrimurthy. "Implementation of Back-Propagation Algorithm For Renal Data mining". International Journal of Computer Science and Security. 2(2), 35-47.2008.
16. P. Raviram, and R.S.D. Wahidabanu. "Implementation of artificial neural network in concurrency control of computer integrated manufacturing (CIM) database". International Journal of Computer Science and Security. 2(5), 23-35.2008.
17. Calzarossa, and G.Valli. "A Fuzzy Algorithm for Web Caching". SIMULATION SERIES journal, 35(4), 630-636, 2003.
18. B. Krishnamurthy, and J. Rexford. "Web protocols and practice: HTTP/1.1, networking protocols, caching and traffic measurement". Addison-Wesley, 2001.
19. Masrah Azrifah Azmi Murad, and Trevor Martin. "Similarity-Based Estimation for Document Summarization using Fuzzy Sets". International Journal of Computer Science and Security. 1(4), 1-12. 2007.
20. J. E. Muñoz-Expósito,S. García-Galán,N. Ruiz-Reyes, and P. Vera-Candeas. "Adaptive network-based fuzzy inference system vs. other classification algorithms for warped LPC-based speech/music discrimination". Engineering Applications of Artificial Intelligence,Volume 20 , Issue 6 (September 2007), Pages 783-793,Pergamon Press, Inc. Tarrytown, NY, USA, 2007.
21. Jang. "ANFIS: Adaptive-network-based fuzzy inference system". IEEE Trans Syst Man Cybern 23 (1993) (3), pp. 665.
22. BU Web Trace,<http://ita.ee.lbl.gov/html/contrib/BU-Web-Client.html>.
23. W. Tian, B. Choi, and V.V. Phoha . "An Adaptive Web Cache Access Predictor Using Neural Network". Published by :Springer- Verlag London, UK. 2002.

24. Y. Zhu, and Y. Hu. "*Exploiting client caches to build large Web caches*". The Journal of Supercomputing. 39(2), 149-175. Springer Netherlands. .2007.
25. L. SHI, L. WEI, H.Q. YE, and Y.SHI. "*MEASUREMENTS OF WEB CACHING AND APPLICATIONS*". Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, 13-16 August 2006. Dalian,1587-1591.

A Encryption Based Dynamic and Secure Routing Protocol for Mobile Ad Hoc Network

Pankaj Kumar Sehgal

*Lecturer, MM Institute of Computer
Technology and Business Management,
MM University, Mullana (Ambala), Haryana, India*

pankajkumar.sehgal@gmail.com

Rajender Nath

*Reader, Department of Computer Science and
Applications, Kurukshetra University,
Kurukshetra, Haryana, India*

rnath_2k3@rediffmail.com

ABSTRACT

Significant progress has been made for making mobile ad hoc networks secure and dynamic. The unique characteristics like infrastructure-free and absence of any centralized authority make these networks more vulnerable to security attacks. Due to the ever-increasing security threats, there is a need to develop algorithms and protocols for a secured ad hoc network infrastructure. This paper presents a secure routing protocol, called EDSR (Encrypted Dynamic Source Routing). EDSR prevents attackers or malicious nodes from tampering with communication process and also prevents a large number of types of Denial-of-Service attacks. In addition, EDSR is efficient, using only efficient symmetric cryptographic primitives. We have developed a new program in c++ for simulation setup.

Keywords: mobile network, ad hoc network, attacks, security threats

1. INTRODUCTION

While a number of routing protocols [9-17] have been proposed in the Internet Engineering Task Force's MANET working group in the last years, none of the address security issues satisfactorily. There are two main sources of threats to routing protocols. The first is from nodes that are not part of the network, and the second is from compromised nodes that are part of the network. While an attacker can inject incorrect routing information, reply old information, or cause excessive load to prevent proper routing protocol functioning in both cases, the latter case is more severe since detection of compromised nodes is more difficult. A solution suggested by [6] involves relying on discovery of multiple routes by routing protocols to get around the problem of failed routes. Another approach is the use of diversity coding [19] for taking advantage of multiple paths by transmitting sufficient redundant information for error detection and correction. While these approaches are potentially useful if the routing protocol is compromised, other work exists for augmenting the actual security of the routing protocol in ad hoc networks. The approach proposed in [25] complements Dynamic Source Routing (DSR, [17]), a popular ad hoc routing protocol, with a "watchdog" (Malicious behavior detector), and a "pathrater" (rating of network paths), for enabling nodes to avoid malicious nodes, the approach has the unfortunate side effect of rewarding them by reducing their traffic load and forwarding their messages. The approach in

[24] applies the concept of local “neighborhood watch” to identify malicious nodes, and propagate this information such that malicious nodes are penalized by all other nodes.

Efficient and reliable key management mechanisms are arguably the most important requirement for enforcing confidentiality, integrity, authentication, authorization and non-repudiation of messages in ad hoc networks. Confidentiality ensures that information is not disclosed to unauthorized entities. Integrity guarantees that a message between ad hoc nodes to ascertain the identity of the peer communicating node. Authorization establishes the ability of a node to consume exactly the resources allocated to it. Non-repudiation ensures that an originator of a message cannot deny having sent it. Traditional techniques of key management in wire-line networks use a public key infrastructure and assume the existence of a trusted and stable entity such as a certification authority (CA) [1-3] that provides the key management service. Public keys are kept confidential to individual nodes. The CA's public key is known to every node, while it signs certificates binding public keys to nodes. Such a centralized CA- based approach is not applicable to ad hoc networks since the CS is a single point of failure, which introduces the problem of maintaining synchronization across the multiple CAs, while alleviating the single-point-of-failure problem only slightly. Many key management schemes proposed for ad hoc networks use threshold cryptography [4-5] for distributing trust amongst nodes. In [6], n servers function as CAs, with tolerate at most t compromised servers. The public key of the service is known to all nodes, while the private key of the service is known to all nodes, while the private key is divided into n shares, one for each server. Each server also knows the public keys of all nodes. Share refreshing is used to achieve proactive security, and scalable adapt to network changes.

This proposal also uses DSR for illustration, describes DSR vulnerabilities stemming from malicious nodes, and techniques for detection of malicious nodes by neighboring nodes.

2. ENCRYPTED-DSR

EDSR has two phases: route discovery and route utilization phases. We give an overview of each phase below. A snapshot of the simulator for the same is shown in FIGURE 1.

2.1 Route Discovery

In the route discovery phase, a source node S broadcasts a route request indicating that it needs to find a path from S to a destination node D . When the route request packets arrive at the destination node D , D selects three valid paths, copy each path to a route reply packet, signs the packets and unicasts them to S using the respective reverse paths. S proceeds with the utilization phase when it receives the route reply packets.

2.2 Route Utilization

The source node S selects one of the routing paths it acquired during the routing discovery phase. The destination node D is required to send a RREP (Route Reply) packet to S . Then S sends data traffic to D . S assumes that there are selfish or malicious nodes on the path and proceeds as follows: S constructs and sends a forerunner packet to inform the nodes on the path that they should expect a specified amount of data from the source of the packet within a given time. When the forerunner packet reaches the destination, it sends an acknowledgment to S . If S do not receives an acknowledgment. If there are malicious agents in the path and they choose to drop the data packet or acknowledgment from D , such eventuality is dealt with cryptography so that malicious node can not get the right information.

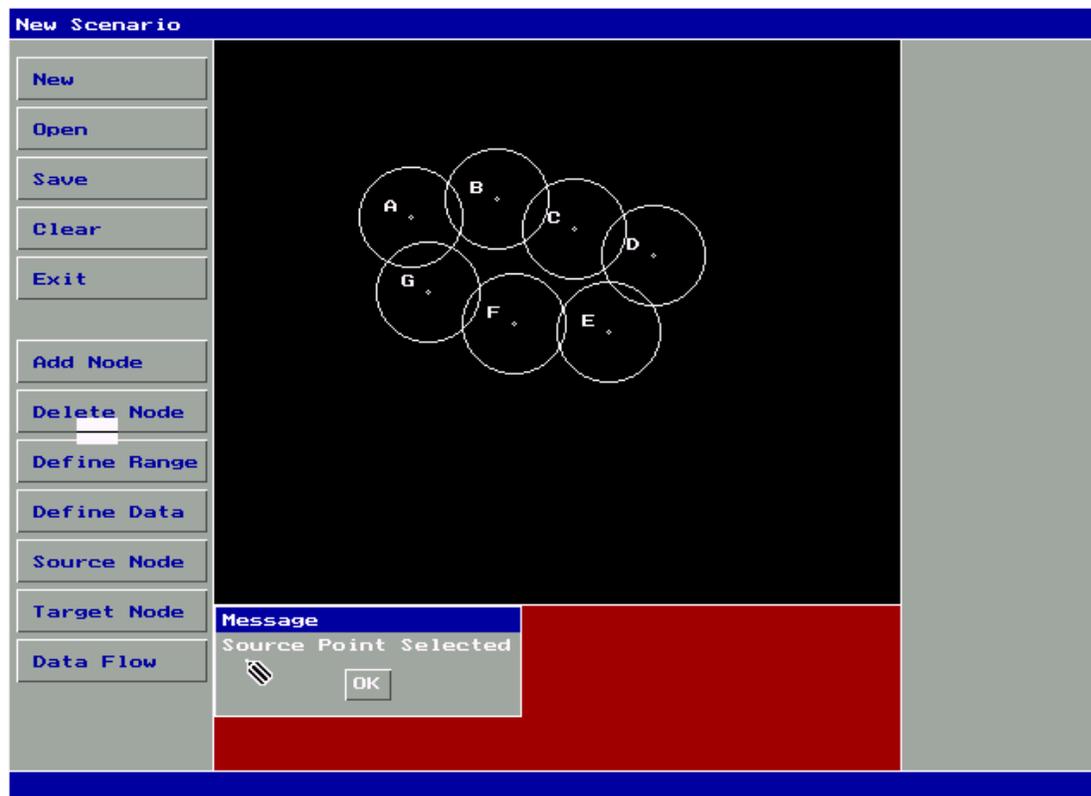


FIGURE 1: An ad hoc network environment for E-DSR

2.3 Network Assumptions

E-DSR utilizes the following assumptions regarding the targeted MANETs:

- Each node has a unique identifier (IP address, MAC address).
- Each node has a valid certificate and the private keys of the CAs which issued the certificates of the other network peers.
- The wireless communication links between the nodes are symmetric; that is, if node N_i is in the transmission range of node N_j , then N_j is also in the transmission range of N_i . This is typically the case with most 802.11 [23] compliant network interfaces.
- The link-layer of the MANET nodes provides transmission error detection service. This is a common feature of most 802.11 wireless interfaces.
- Any given intermediate node on a path from a source to a destination may be malicious and therefore cannot be fully trusted. The source node only trusts a destination node, and visa versa, a destination node only trusts a source node.

2.4 Threat Model

In this work, we do not assume the existence of security association between any pair of nodes. Some previous works, for example [7, 20] rely on the assumption that protocols such as the well known Diffie-Hellman key exchange protocol [18] can be used to establish secret shared keys on communicating peers. However, in an adversarial environment, malicious entities can easily disrupt these protocols - and prevent nodes from establishing shared keys with other nodes-by simply dropping the key exchange protocol messages, rather than forwarding them. Our threat model does not place any particular limitations on adversarial entities. Adversarial entities can intercept, modify or fabricate packets; create routing loops; selectively drop packets; artificially delay packets; or attempt denial of service attacks by injecting packets in the network with the goal of consuming network resources. Malicious entities can also collude with other malicious entities in attempts to hide their adversarial behaviors. The goal of our protocol is to detect selfish

or adversarial activities and mitigates against them. Examples for malicious nodes shown in FIGURE 2 and FIGURE 3.

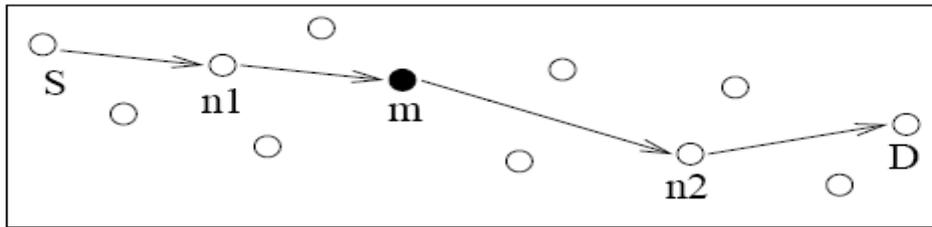


FIGURE 2: One malicious node on a routing path

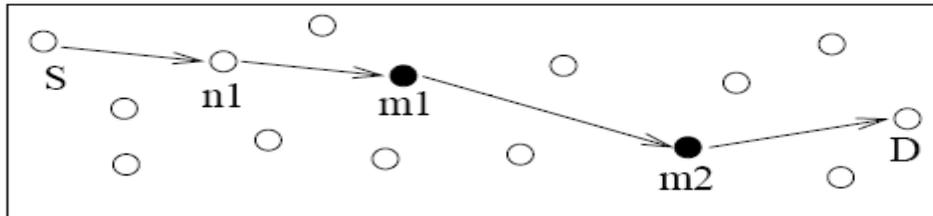


FIGURE 3: Adjacent colluding nodes on a routing path

2.5 Simulation Evaluation

We implemented EDSR in a network simulator using C language shown in Fig.5.1. For the cryptographic components, we utilized Cryptlib crypto toolkit [26] to generate 56-bit DES cryptographic keys for the signing and verification operations. In the simulation implementation, malicious nodes do not comply with the protocol. The simulation parameters are shown in TABLE 1.

| Parameter | Value |
|---------------------------|-----------------|
| Space | 670m × 670m |
| Number of nodes | 26 |
| Mobility model | Random waypoint |
| Speed | 20 m/s |
| Max number of connections | 12 |
| Packet size | 512 bytes |
| Packet generation rate | 4 packets/s |
| Simulation time | 120 s |

TABLE 1: Simulation Parameter Values

2.7 Performance Metrics

We used the following metrics to evaluate the performance of our scheme.

2.7.1 Packet Delivery Ratio

This is the fraction of data packets generated by CBR (Constant Bit Rate) sources that are delivered to the destinations. This evaluates the ability of EDSR to deliver data packets to their destinations in the presence of varying number of malicious agents which selectively drop packets they are required to forward.

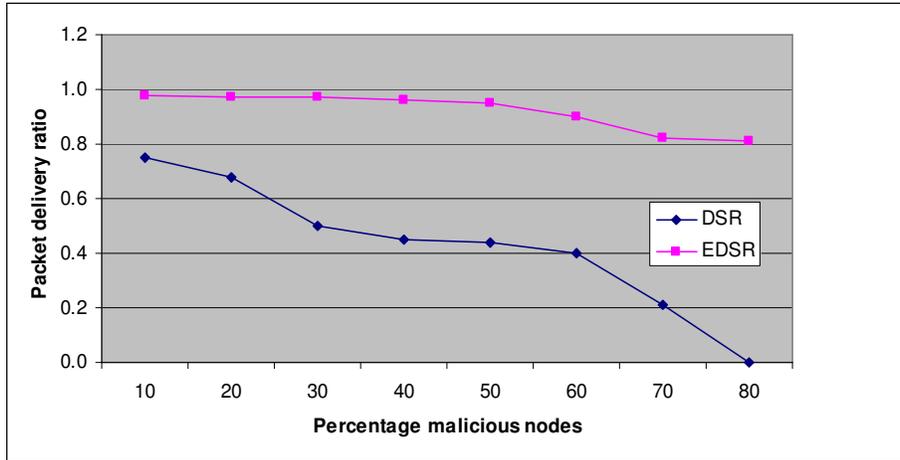


FIGURE 4: Data packet delivery ratio

2.7.2 Number of Data Packets Delivered

This metric gives additional insight regarding the effectiveness of the scheme in delivering packets to their destination in the presence of varying number of adversarial entities.

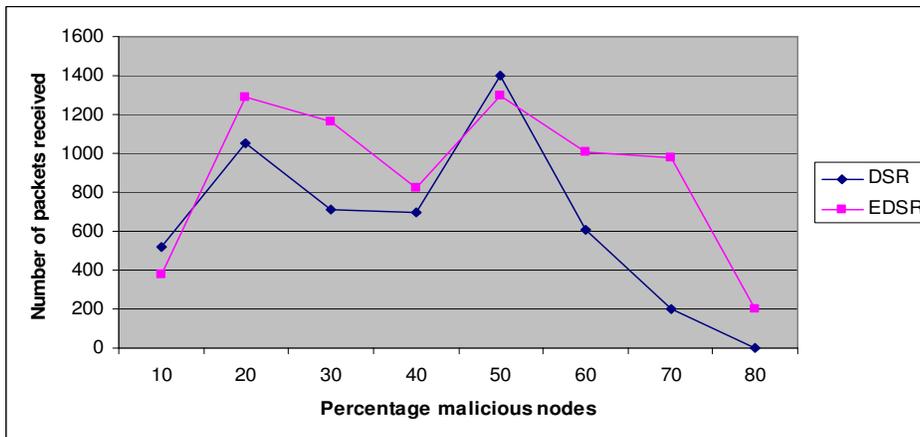


FIGURE 5: Number of packets received over the length of the simulation

2.7.3 Average End-to-end Latency of the Data Packets

This is the ratio of the total time it takes all packets to reach their respective destinations and the total number of packets received. This measures the average delays of all packets which were successfully transmitted.

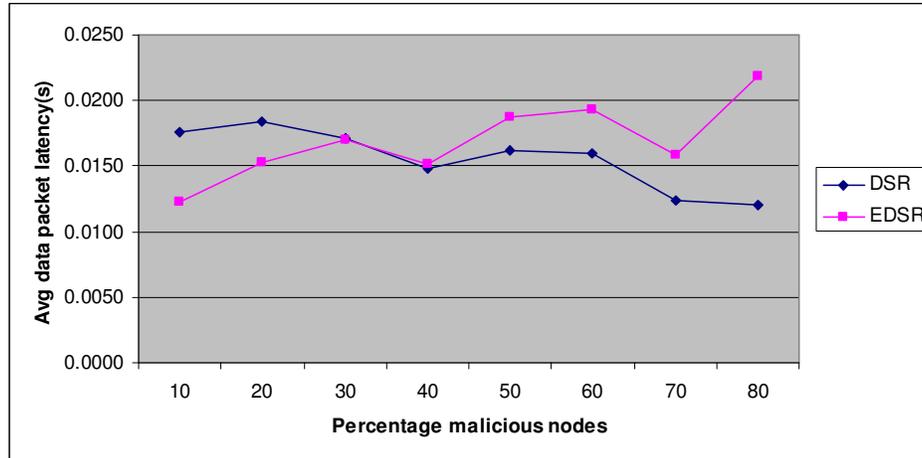


FIGURE 6: Average data packets latency

The results of the simulation for EDSR is compared to that of DSR [17], which currently is perhaps the most widely used MANET source routing protocol.

3. CONSLUSION & FUTURE WORK

Routing protocol security, Node configuration, Key management and Intrusion detection mechanisms are four areas that are candidates for standardization activity in the ad hoc networking environment. While significant research work exists in these areas, little or no attempt has been made to standardize mechanisms that would enable multi-vendor nodes to inter-operate on a large scale and permit commercial deployment of ad hoc networks. The standardization requirements for each of the identified areas will have to be determined. Based on the identified requirements, candidate proposals will need to be evaluated. Care has to be taken to avoid the trap that the MANET working group is currently in, which is of having of large number competing mechanisms. A protocol has been presented by us to standardized key management area. We have presented a simulation study which shows that E-DSR works better than DSR when malicious node increases.

In the future, complex simulations could be carried out which will included other routing protocols as well as other cryptography tools.

4. REFERENCES

1. M. Gasser et al., "The Digital Distributed Systems Security Architecture", Proc. 12th Natl. Comp. Security Conf., NIST, 1989.
2. J. Tardo and K. Algappan, "SPK: Global Authentication Using Public Key Ceriticates", Proc. IEEE Symp. Security and Privacy, CA, 1991.
3. C Kaufman, "DASS: Distributed Authentication Security Service", RFC 1507, 1993.
4. Y. Desmedt and Y. Frankel, "Threshold Cryptosystems", Advances in Cryptography- Crypto' 89, G. Brassard, Ed. Springer- Verlag, 1990.
5. Y. Desmedt "Threshold Cryptography", Euro. Trans. Telecom., 5(4), 1994.
6. L. Zhou and Z. Haas, "Securing Ad Hoc Networks", IEEE Networks, 13(6), 1999.
7. Y. -C. Hu, D. B. Johnson and A. Perrig, "SEAD: Secure Efficient Distance Vector Routing for Mobile Wireless Ad Hoc Networks", Fourth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'02), Jun. 2002
8. C. S. R. Murthy and B. S. Manoj, "Ad Hoc Wireless Networks: A Architectures and Protocols", Prentice Hall PTR, 2004.

9. S. Das et Al. "Ad Hoc On-Demand Distance Vector (AODV) Routing", draft-ietf-manet-aodv-17, February, 2003.
10. J. Macker et Al., "Simplified Multicast Forwarding for MANET", draft-ietf-manet-smf-07, February 25, 2008.
11. I. Chakeres et. Al., "Dynamic MANET On-demand (DYMO) Routing", draft-ietf-manet-dymo-14, June 25, 2008.
12. I. Chakeres et. Al., "IANA Allocations for MANET Protocols", draft-ietf-manet-iana-07, November 18, 2007.
13. T. Clausen et. Al., "The Optimized Link State Routing Protocol version 2", draft-ietf-manet-olsrv2-06, June 6, 2008.
14. T. Clausen et. Al., "Generalized MANET Packet/Message Format", draft-ietf-manet-packetbb-13, June 24, 2008.
15. T. Clausen et Al., "Representing multi-value time in MANETs", draft-ietf-manet-timetlv-04, November 16, 2007.
16. T. Clausen et Al., "MANET Neighborhood Discovery Protocol (NHDP)", draft-ietf-manet-nhdp-06, March 10, 2008.
17. D. Johnson and D. Maltz., "Dynamic source routing in ad-hoc wireless networks routing protocols", In Mobile Computing, pages 153-181. Kluwer Academic Publishers, 1996.
18. C. R. Davis., "IPSec: Securing VPNs", Osborne/McGraw-Hill, New York, 2001.
19. E. Ayannoglu et al., " Diversity Coding for Transparent Self-Healing and Fault-Tolerant Communication Networks", IEEE Trans. Comm. 41(11), 1993.
20. P. Papadimitratos and Z. J. Haas, "Secure Routing for Mobile Ad hoc Networks", In Proc. of the SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), Jan. 2002.
21. Y. Zhang and W. Lee, "Intrusion Detection in Wireless Ad Hoc Networks", Proc. Mobicom, 2000.
22. R. Droms, " Dynamic Host Configuration Protocol", IETF RFC 2131, 1997.
23. IEEE-SA Standards Board. IEEE Std 802.11b-1999, 1999.
24. S. Buchegger and J. LeBoudec, " Nodes Bearing Grudges: Towards Routing Security, Fairness, and Robustness in Mobile Ad Hoc Network," Proc. 10th Euromicro PDP , Gran Canaria, 2002.
25. S. Marti et al., "Mitigating Routing Behavior in Mobile Ad Hoc Networks", Proc. Mobicom, 2000.
26. P. Gutmann. Cryptlib encryption tool kit. <http://www.cs.auckland.ac.nz/~pgut001/cryptlib>.
27. Rashid Hafeez Khokhar, Md Asri Ngadi, Satria Mandala, "A Review of Current Routing Attacks in Mobile Ad Hoc Networks", in IJCSS: International Journal of Computer Science and Security, "Volume 2, Issue 3, pages 18-29, May/June 2008.
28. R. Asokan , A.M. Natarajan, C. Venkatesh, "Ant Based Dynamic Source Routing Protocol to Support Multiple Quality of Service (QoS) Metrics in Mobile Ad Hoc Networks", in IJCSS: International Journal of Computer Science and Security, "Volume 2, Issue 3, pages 48-56, May/June 2008.
29. N. Bhalaji, A. Shanmugam, Druhin mukherjee, Nabamalika banerjee, "Direct trust estimated on demand protocol for secured routing in mobile Adhoc networks", in IJCSS: International Journal of Computer Science and Security, "Volume 2, Issue 5, pages 6-12, September/ October 2008.

MMI Diversity Based Text Summarization

Mohammed Salem Binwahlan

*Faculty of Computer Science and Information Systems
University Teknologi Malaysia
Skudai, Johor, 81310, Malaysia*

moham2007med@yahoo.com

Naomie Salim

*Faculty of Computer Science and Information Systems
University Teknologi Malaysia
Skudai, Johor, 81310, Malaysia*

naomie@utm.my

Ladda Suanmali

*Faculty of Science and Technology,
Suan Dusit Rajabhat University
295 Rajasrima Rd, Dusit, Bangkok, 10300, Thailand*

ladda_sua@dusit.ac.th

Abstract

The search for interesting information in a huge data collection is a tough job frustrating the seekers for that information. The automatic text summarization has come to facilitate such searching process. The selection of distinct ideas “diversity” from the original document can produce an appropriate summary. Incorporating of multiple means can help to find the diversity in the text. In this paper, we propose approach for text summarization, in which three evidences are employed (clustering, binary tree and diversity based method) to help in finding the document distinct ideas. The emphasis of our approach is on controlling the redundancy in the summarized text. The role of clustering is very important, where some clustering algorithms perform better than others. Therefore we conducted an experiment for comparing two clustering algorithms (K-means and complete linkage clustering algorithms) based on the performance of our method, the results shown that k-means performs better than complete linkage. In general, the experimental results shown that our method performs well for text summarization comparing with the benchmark methods used in this study.

Keywords: Binary tree, Diversity, MMR, Summarization, Similarity threshold.

1. INTRODUCTION

The searching for interesting information in a huge data collection is a tough job frustrating the seekers for that information. The automatic text summarization has come to facilitate such searching process. It works on producing a short form for original document in form of summary. The summary performs a function of informing the user about the relevant documents to his or her need. The summary can reduce the reading time and providing quick guide to the interesting information.

In automatic text summarization, the selection process of the distinct ideas included in the document is called diversity. The diversity is very important evidence serving to control the redundancy in the summarized text and produce more appropriate summary. Many approaches have been proposed for text summarization based on the diversity. The pioneer work for diversity based text summarization is MMR (maximal marginal relevance), it was introduced by Carbonell and Goldstein [2], MMR maximizes marginal relevance in retrieval and summarization. The sentence with high maximal relevance means it is highly relevant to the query and less similar to the already selected sentences. Our modified version of MMR maximizes the marginal importance and minimizes the relevance. This approach treats sentence with high maximal importance as one that has high importance in the document and less relevance to already selected sentences.

MMR has been modified by many researchers [4; 10; 12; 13; 16; 21; 23]. Our modification for MMR formula is similar to Mori et al.'s modification [16] and Liu et al.'s modification [13] where the importance of the sentence and the sentence relevance are added to the MMR formulation. Ribeiro and Matos [19] proved that the summary generated by MMR method is closed to the human summary, motivating us to choose MMR and modify it by including some documents features. The proposed approach employs two evidences (clustering algorithm and a binary tree) to exploit the diversity among the document sentences. Neto et al. [17] presented a procedure for creating approximate structure for document sentences in the form of a binary tree, in our study, we build a binary tree for each cluster of document sentences, where the document sentences are clustered using a clustering algorithm into a number of clusters equal to the summary length. An objective of using the binary tree for diversity analysis is to optimize and minimize the text representation; this is achieved by selecting the most representative sentence of each sentences cluster. The redundant sentences are prevented from getting the chance to be candidate sentences for inclusion in the summary, serving as penalty for the most similar sentences. Our idea is similar to Zhu et al.'s idea [25] in terms of improving the diversity where they used absorbing Markov chain walks.

The rest of this paper is described as follows: section 2 presents the features used in this study, section 3 discusses the importance and relevance of the sentence, section 4 discusses the sentences clustering, section 5 introduces the document-sentence tree building process using k-means clustering algorithm, section 6 gives full description of the proposed method, section 7 discusses the experimental design, section 8 presents the experimental results, section 9 shows a comparison between two clustering algorithms based on the proposed method performance. Section 10 concludes our work and draws the future study plan.

2. SENTENCE FEATURES

The proposed method makes use of eight different surface level features; these features are identified after the preprocessing of the original document is done, like stemming using porter's stemmer¹ and removing stop words. The features are as follows.

a. Word sentence score (WSS): it is calculated using the summation of terms weights (TF-ISF, calculated using eq. 1, [18]) of those terms synthesizing the sentence and occur in at least in a number of sentences equal to half of the summary length (LS) divided by highest term weights (TF-ISF) summation of a sentence in the document (HTFS) as shown in eq. 2, the idea of making the calculation of word sentence score under the condition of occurrence of its term in specific number of sentences is supported by two factors: excluding the unimportant terms and applying the mutual reinforcement principle [24]. MAN'A-LO'PEZ *et al.* [15] calculated the sentence score as proportion of the square of the query-word number of a cluster and the total number of words in that cluster.

¹ <http://www.tartarus.org/martin/PorterStemmer/>

Term frequency-inverse sentence frequency (TF-ISF) [18], term frequency is very important feature; its first use dates back to fifties [14] and still used.

$$W_{ij} = tf_{ij} \times isf = tf(t_{ij}, s_i) \left[1 - \frac{\log(sf(t_{ij}) + 1)}{\log(n + 1)} \right] \quad (1)$$

Where W_{ij} is the term weight (TF-ISF) of the term t_{ij} in the sentence s_i .

$$WSS(S_i) = 0.1 + \frac{\sum_{t_j \in S_i} W_{ij}}{HTFS} \quad | \text{no. of sentences containing } t_j \geq \frac{1}{2} LS \quad (2)$$

Where 0.1 is minimum score the sentence gets in the case its terms are not important.

b. Key word feature: the top 10 words whose high TF-ISF (eq. 1) score are chosen as key words [8; 9]. Based on this feature, any sentence in the document is scored by the number of key words it contains, where the sentence receives 0.1 score for each key word.

c. Nfriends feature: the nfriends feature measures the relevance degree between each pair of sentences by the number of sentences both are similar to. The friends of any sentence are selected based on the similarity degree and similarity threshold [3].

$$Nfriends(s_i, s_j) = \frac{|s_i(friends) \cap s_j(friends)|}{|s_i(friends) \cup s_j(friends)|} \quad | i \neq j \quad (3)$$

d. Ngrams feature: this feature determines the relevance degree between each pair of sentences based on the number of n-grams they share. The skipped bigrams [11] used for this feature.

$$Ngrams(s_i, s_j) = \frac{|s_i(ngrams) \cap s_j(ngrams)|}{|s_i(ngrams) \cup s_j(ngrams)|} \quad | i \neq j \quad (4)$$

e. The similarity to first sentence (sim_fsd): This feature is to score the sentence based on its similarity to the first sentence in the document, where in news article, the first sentence in the article is very important sentence [5]. The similarity is calculated using eq. 11.

f. Sentence centrality (SC): the sentence has broad coverage of the sentence set (document) will get high score. Sentence centrality widely used as a feature [3; 22]. We calculate the sentence centrality based on three factors: the similarity, the shared friends and the shared ngrams between the sentence in hand and all other the document sentences, normalized by n-1, n is the number of sentences in the document.

$$SC(S_i) = \frac{\sum_{j=1}^{n-1} sim(S_i, S_j) + \sum_{j=1}^{n-1} nfriends(S_i, S_j) + \sum_{j=1}^{n-1} ngrams(S_i, S_j)}{n-1} \quad | i \neq j \text{ and } sim(S_i, S_j) \geq \theta \quad (5)$$

Where S_j is a document sentence except S_i , n is the number of sentences in the document. θ is the similarity threshold which is determined empirically, in an experiment was run to determine the best similarity threshold value, we have found that the similarity threshold can take two values, 0.03 and 0.16.

The following features are for those sentences containing ngrams [20] (consecutive terms) of title where n=1 in the case of the title contains only one term, n=2 otherwise:

g. Title-help sentence (THS): the sentence containing n-gram terms of title.

$$THS(s_i) = \frac{s_i(ngrams) \cap T(ngrams)}{|s_i(ngrams) \cup T(ngrams)|} \quad (6)$$

h. Title-help sentence relevance sentence (THSRS): the sentence containing ngram terms of any title-help sentence.

$$THSRS(s_j) = \frac{s_j(ngrams) \cap THS(s_i(ngrams))}{|s_j(ngrams) \cup THS(s_i(ngrams))|} \quad (7)$$

The sentence score based on THS and THSRS is calculated as average of those two features:

$$SS_NG = \frac{THS(s_i) + THSRS(s_i)}{2} \quad (8)$$

3. THE SENTENCE IMPORTANCE (IMPR) AND SENTENCE RELEVANCE (REL)

The sentence importance is the main score in our study; it is calculated as linear combination of the document features. Liu *et al.* [13] computed the sentence importance also as linear combination of some different features.

$$IMPR(S_i) = avg(WSS(S_i) + SC(S_i) + SS_NG(S_i) + sim_fsd(S_i) + kwrds(S_i)) \quad (9)$$

Where *WSS*: word sentence score, *SC*: sentence centrality, *SS_NG*: average of THS and THSRS features, *Sim_fsd*: the similarity of the sentence s_i with the first document sentence and *kwrds*(S_i) is the key word feature.

The sentence relevance between two sentences is calculated in [13] based on degree of the semantic relevance between their concepts, but in our study the sentence relevance between two sentences is calculated based on the shared friends, the shared ngrams and the similarity between those two sentences:

$$Rel(s_i, s_j) = avg(nfriends(s_i, s_j) + ngrams(s_i, s_j) + sim(s_i, s_j)) \quad (10)$$

4. SENTENCES CLUSTERING

The clustering process plays an important role in our method; it is used for grouping the similar sentences each in one group. The clustering is employed as an evidence for finding the diversity among the sentences. The selection of clustering algorithm is more sensitive needing to experiment with more than one clustering algorithm. There are two famous categories of the clustering methods: partitional clustering and hierarchical clustering. The difference between those two categories is that hierarchical clustering tries to build a tree-like nested structure partition of clustered data while partitional clustering requires receiving the number of clusters

then separating the data into isolated groups [7]. Example of the hierarchical clustering methods is agglomerative clustering methods like Single linkage, complete linkage, and group average linkage. We have tested our method using k-means clustering algorithm and complete linkage clustering algorithm.

5. DOCUMENT-SENTENCE TREE BUILDING (DST) USING K-MEANS CLUSTERING ALGORITHM

The first stage for building the document-sentence tree is to cluster the document sentences into a number of clusters. The clustering is done using k-means clustering algorithm. The clusters number is determined automatically by the summary length (number of sentences in the final summary). The initial centroids are selected as the following:

- Pick up one sentence which has higher number of similar sentences (sentence friends).
- Form a group for the picked up sentence and its friends, the maximum number of sentences in that group is equal to the total number of document sentences divided by the number of clusters.
- From the created group of sentences, the highest important sentence is selected as initial centroid.
- Remove the appearance of each sentence in the created group from the main group of document sentences.
- Repeat the same procedure until the number of initial centroids selected is equal to the number of clusters.

To calculate the sentence similarity between two sentences s_i and s_j , we use *TF-ISF* and *cosine* similarity measure as in eq. 11 [3]:

$$sim(s_i, s_j) = \frac{\sum_{w_i \in s_i, s_j} tf(w_i, s_i)tf(w_i, s_j) \left[\frac{1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}}{1} \right]^2}{\sqrt{\sum_{w_i \in s_i} \left(tf(w_i, s_i) \left[\frac{1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}}{1} \right] \right)^2} \times \sqrt{\sum_{w_i \in s_j} \left(tf(w_i, s_j) \left[\frac{1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}}{1} \right] \right)^2}} \quad (11)$$

Where tf is term frequency of term w_i in the sentence s_i or s_j , sf is number of sentences containing the term w_i in the document, n is number of sentences in the document.

Each sentences cluster is represented as one binary tree or more. The first sentence which is presented in the binary tree is that sentence with higher number of friends (higher number of similar sentences), then the sentences which are most similar to already presented sentence are selected and presented in the same binary tree. The sentences in the binary tree are ordered based on their scores. The score of the sentence in the binary tree building process is calculated based on the importance of the sentence and the number of its friends using eq. 12. The goal of incorporating the importance of sentence and number of its friends together to calculate its score is to balance between the importance and the centrality (a number of high important friends).

$$Score_{BT}(s_i) = impr(s_i) + (1 - (1 - impr(s_i)) \times friendsNo(s_i)) \quad (12)$$

Where $Score_{BT}(s_i)$ is the score of the s_i sentence in the binary tree building process, $impr(s_i)$ is importance of the sentence s_i and $friendsNo(s_i)$ is the number of sentence friends.

Each level in the binary tree contains 2^{ln} of the higher score sentences, where ln is the level number, $ln=0, 1, 2, \dots, n$, the top level contains one sentence which is a sentence with highest

score. In case, there are sentences remaining in the same cluster, a new binary tree is built for them by the same procedure.

6. METHODOLOGY

The proposed method for summary generation depends on the extraction of the highest important sentences from the original text, we introduce a modified version of MMR, and we called it MMI (maximal marginal importance). MMR approach depends on the relevance of the document to the query, and it is for query based summary. In our modification we have tried to release this restriction by replacing the query relevance with sentence importance for presenting the MMI as generic summarization approach.

Most features used in this method are accumulated together to show the importance of the sentence, the reason for including the importance of the sentence in the method is to emphasize on the high information richness in the sentence as well as high information novelty. We use the tree for grouping the most similar sentences together in easy way, and we assume that the tree structure can take part in finding the diversity.

MMI is used to select one sentence from the binary tree of each sentence cluster to be included in the final summary. In the binary tree, a level penalty is imposed on each level of sentences which is 0.01 times the level number. The purpose of the level penalty is to reduce the noisy sentences score. The sentences which are in the lower levels are considered as noisy sentences because they are carrying low scores. Therefore the level penalty in the low levels is higher while it is low in the high levels. We assume that this kind of scoring will allow to the sentence with high importance and high centrality to get the chance to be a summary sentence, this idea is supported by the idea of PageRank used in Google [1] where the citation (link) graph of the web page or backlinks to that page is used to determine the rank of that page. The summary sentence is selected from the binary tree by traversing all levels and applying MMI on each level sentences.

$$MMI(S_i) = Arg \max_{S_i \in CS \setminus SS} \left[(Score_{BT}(S_i) - \beta(S_i)) - \max_{S_j \in SS} (Rel(S_i, S_j)) \right] \quad (13)$$

Where $Rel(S_i, S_j)$ is the relevance between the two competitive sentences, S_i is the unselected sentence in the current binary tree, S_j is the already selected sentence, SS is the list of already selected sentences, CS is the competitive sentences of the current binary tree and β is the penalty level.

In MMR, the parameter λ is very important, it controls the similarity between already selected sentences and unselected sentences, and where setting it to incorrect value may cause creation of low quality summary. Our method pays more attention for the redundancy removing by applying MMI in the binary tree structure. The binary tree is used for grouping the most similar sentences in one cluster, so we didn't use the parameter λ because we just select one sentence from each binary tree and leave the other sentences.

Our method is intended to be used for single document summarization as well as multi-documents summarization, where it has the ability to get rid of the problem of some information stored in single document or multi-documents which inevitably overlap with each other, and can extract globally important information. In addition to that advantage of the proposed method, it maximizes the coverage of each sentence by taking into account the sentence relatedness to all other document sentences. The best sentence based on our method policy is the sentence that has higher importance in the document, higher relatedness to most document sentences and less similar to the sentences already selected as candidates for inclusion in the summary.

7. EXPERIMENTAL DESIGN

The Document Understanding Conference (DUC) [6] data collection became as standard data set for testing any summarization method; it is used by most researchers in text summarization. We have used DUC 2002 data to evaluate our method for creating a generic 100-word summary, the task 1 in DUC 2001 and 2002, for that task, the training set comprised 30 sets of approximately 10 documents each, together with their 100-word human written summaries. The test set comprised 30 unseen documents. The document sets D061j, D064j, D065j, D068j, D073b, D075b and D077b were used in our experiment.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) toolkit [11] is used for evaluating the proposed method, where ROUGE compares a system generated summary against a human generated summary to measure the quality. ROUGE is the main metric in the DUC text summarization evaluations. It has different variants, in our experiment, we use ROUGE-N (N=1 and 2) and ROUGE-L, the reason for selecting these measures is what reported by same study [11] that those measures work well for single document summarization.

The ROUGE evaluation measure (version 1.5.5²) generates three scores for each summary: recall, precision and F-measure (weighted harmonic mean, eq. 14), in the literature, we found that the recall is the most important measure to be used for comparison purpose, so we will concentrate more on the recall in this evaluation.

$$F = \frac{1}{\left(\alpha \times \left(\frac{1}{P} \right) + (1 - \alpha) \times \left(\frac{1}{R} \right) \right)} \quad (14)$$

Where P and R are precision and recall, respectively. Alpha is the parameter to balance between precision and recall; we set this parameter to 0.5.

8. EXPERIMENTAL RESULTS

The similarity threshold plays very important role in our study where the most score of any sentence depends on its relation with other document sentences. Therefore we must pay more attention to this factor by determining its appropriate value through a separate experiment, which was run for this purpose. The data set used in this experiment is document set D01a (one document set in DUC 2001 document sets), the document set D01a containing eleven documents, each document accompanied with its model or human generated summary. We have experimented with 21 different similarity threshold values ranging from 0.01 to 0.2, 0.3 by stepping 0.01. We found that the best average recall score can be gotten using the similarity threshold value 0.16. However, this value doesn't do well with each document separately. Thus, we have examined each similarity threshold value with each document and found that the similarity threshold value that can perform well with all documents is 0.03. Therefore, we decided to run our summarization experiment using the similarity threshold 0.03.

We have run our summarization experiment using DUC 2002 document sets D061j, D064j, D065j, D068j, D073b, D075b and D077b where each document set contains two model or human generated summaries for each document. We gave the names H1 and H2 for those two model summaries. The human summary H2 is used as benchmark to measure the quality of our method summary, while the human summary H1 is used as reference summary. Beside the human with human benchmark (H2 against H1), we also use another benchmark which is MS word summarizer

² <http://haydn.isi.edu/ROUGE/latest.html>

The proposed method and the two benchmarks are used to create a summary for each document in the document set used in this study. Each system created good summary compared with the reference (human) summary. The results using the ROUGE variants (ROUGE-1, ROUGE-2 and ROUGE-L) demonstrate that our method performs better than MS word summarizer and closer to the human with human benchmark. For some document sets (D061j, D073b and D075b), our method could perform better than the human with human benchmark. Although the recall score is the main score used for comparing the text summarization methods when the summary length is limited³, we found that our method performs well for all average ROUGE variants scores. The overall analysis for the results is concluded and shown in figures 1, 2 and 3 for the three rouge evaluation measures. MMI average recall at the 95%-confidence interval is shown in Table-1.

| Metric | 95%-Confidence interval |
|---------|-------------------------|
| ROUGE-1 | 0.43017 - 0.49658 |
| ROUGE-2 | 0.18583 - 0.26001 |
| ROUGE-L | 0.39615 - 0.46276 |

Table 1: MMI average recall at the 95%-confidence interval.

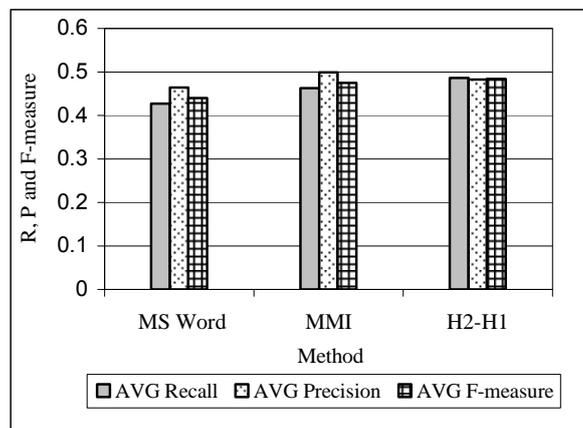


FIGURE 1: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-1.

³ <http://haydn.isi.edu/ROUGE/latest.html>

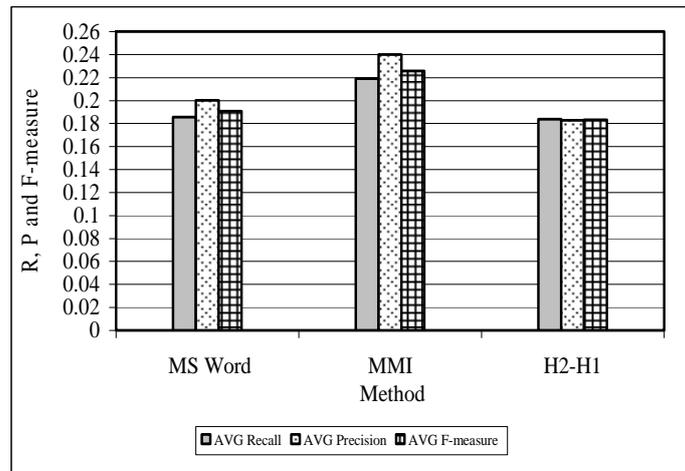


FIGURE 2: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-2.

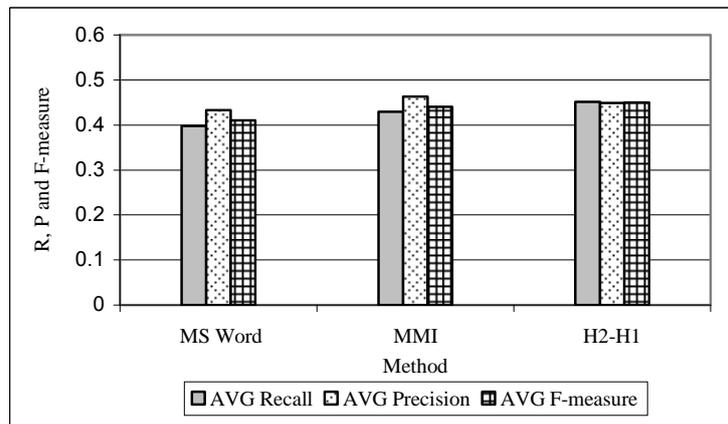


FIGURE 3: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-L.

For ROUGE-2 average recall score, our method performance is better than the two benchmarks by: 0.03313 and 0.03519 for MS word summarizer and human with human benchmark (H2-H1) respectively, this proves that the summary created by our method is not just scatter terms extracted from the original document but it has meaningful. For ROUGE-1 and ROUGE-L average recall scores, our method performance is better than MS word summarizer and closer to human with human benchmark.

9. COMPARISON BETWEEN K-MEANS AND C-LINKAGE CLUSTERING ALGORITHMS BASED ON MMI PERFORMANCE

The previous experiment was run using k-means as clustering algorithm for clustering the sentences, we also run the same experiment using c-linkage (complete linkage) clustering algorithm instead of k-means to find out the clustering method which performs well with our method. The results show that c-linkage clustering algorithm performs less than k-means clustering algorithm. Table 3 shows the comparison between those clustering algorithms.

| ROUGE | Method | R | P | F-measure |
|-------|---------------|---------|---------|-----------|
| 1 | MMI-K-means | 0.46293 | 0.49915 | 0.47521 |
| | MMI-C-linkage | 0.44803 | 0.48961 | 0.46177 |
| 2 | MMI-K-means | 0.21885 | 0.23984 | 0.22557 |
| | MMI-C-linkage | 0.20816 | 0.23349 | 0.21627 |
| L | MMI-K-means | 0.42914 | 0.46316 | 0.44056 |
| | MMI-C-linkage | 0.4132 | 0.45203 | 0.42594 |

Table 2: comparison between k-means and c-linkage clustering algorithms.

10. CONCLUSION AND FUTURE WORK

In this paper we have presented an effective diversity based method for single document summarization. Two ways were used for finding the diversity: the first one is as preliminary way where the document sentences are clustered based on the similarity - similarity threshold is 0.03 determined empirically - and all resulting clusters are presented as a tree containing a binary tree for each group of similar sentences. The second way is to apply the proposed method on each branch in the tree to select one sentence as summary sentence. The clustering algorithm and binary tree were used as helping factor with the proposed method for finding the most distinct ideas in the text. Two clustering algorithms (K-mean and C-linkage) were compared to find out which of them performs well with the proposed diversity method. We found that K-means algorithm has better performance than C-linkage algorithm. The results of our method supports that employing of multiple factors can help to find the diversity in the text because the isolation of all similar sentences in one group can solve a part of the redundancy problem among the document sentences and the other part of that problem is solved by the diversity based method which tries to select the most diverse sentence from each group of sentences. The advantages of our introduced method are: it does not use external resource except the original document given to be summarized and deep natural language processing is not required. Our method has shown good performance when comparing with the benchmark methods used in this study. For future work, we plan to incorporate artificial intelligence technique with the proposed method and extend the proposed method for multi document summarization.

References

1. S. Brin, and L. Page. "The anatomy of a large-scale hypertextual Web search engine". Computer Networks and ISDN System. 30(1-7): 107-117. 1998.
2. J. Carbonell, and J. Goldstein. "The use of MMR, diversity-based reranking for reordering documents and producing summaries". SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 24-28 August. Melbourne, Australia, 335-336. 1998
3. G. Erkan, and D. R. Radev. "LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization". Journal of Artificial Intelligence Research (JAIR), 22, 457-479. AI Access Foundation. 2004.
4. K Filippova, M. Mieskes, V. Nastase, S. P. Ponzetto and M. Strube. "Cascaded Filtering for Topic-Driven Multi-Document Summarization". Proceedings of the Document Understanding Conference. 26-27 April. Rochester, N.Y., 30-35. 2007.
5. M. K. Ganapathiraju. "Relevance of Cluster size in MMR based Summarizer: A Report 11-742: Self-paced lab in Information Retrieval". November 26, 2002.
6. "The Document Understanding Conference (DUC)". <http://duc.nist.gov>.

7. A. Jain, M. N. Murty and P. J. Flynn. "Data Clustering: A Review". ACM Computing Surveys. 31 (3), 264-323, 1999.
8. C. Jaruskulchai and C. Kruengkrai. "Generic Text Summarization Using Local and Global Properties". Proceedings of the IEEE/WIC international Conference on Web Intelligence. 13-17 October. Halifax, Canada: IEEE Computer Society, 201-206, 2003.
9. A. Kiani –B and M. R. Akbarzadeh –T. "Automatic Text Summarization Using: Hybrid Fuzzy GA-GP". IEEE International Conference on Fuzzy Systems. 16-21 July. Vancouver, BC, Canada, 977 -983, 2006.
10. W. Kraaij, M. Spitters and M. v. d. Heijden. "Combining a mixture language model and naive bayes for multi-document summarization". Proceedings of Document Understanding Conference. 13-14 September. New Orleans, LA, 109-116, 2001.
11. C. Y. Lin. "Rouge: A package for automatic evaluation of summaries". . Proceedings of the Workshop on Text Summarization Branches Out, 42nd Annual Meeting of the Association for Computational Linguistics. 25–26 July. Barcelona, Spain, 74-81, 2004b.
12. Z. Lin, T. Chua, M. Kan, W. Lee, Q. L. Sun and S. Ye. "NUS at DUC 2007: Using Evolutionary Models of Text". Proceedings of Document Understanding Conference. 26-27 April. Rochester, NY, USA, 2007.
13. D. Liu, Y. Wang, C. Liu and Z. Wang. "Multiple Documents Summarization Based on Genetic Algorithm". In Wang L. et al. (Eds.) Fuzzy Systems and Knowledge Discovery. (355–364). Berlin Heidelberg: Springer-Verlag, 2006.
14. H. P. Luhn. "The Automatic Creation of Literature Abstracts". IBM Journal of Research and Development. 2(92), 159-165, 1958.
15. M. J. MAN`A-LO`PEZ, M. D. BUENAGA, and J. M. GO` MEZ-HIDALGO. "Multi-document Summarization: An Added Value to Clustering in Interactive Retrieval". ACM Transactions on Information Systems. 22(2), 215–241, 2004.
16. T. Mori, M. Nozawa and Y. Asada. "Multi-Answer-Focused Multi-document Summarization Using a Question-Answering Engine". ACM Transactions on Asian Language Information Processing. 4 (3), 305–320 , 2005.
17. J. L. Neto, A. A. Freitas and C. A. A. Kaestner. "Automatic Text Summarization using a Machine Learning Approach". In Bittencourt, G. and Ramalho, G. (Eds.). Proceedings of the 16th Brazilian Symposium on Artificial intelligence: Advances in Artificial intelligence. (pp. 386-396). London: Springer-Verlag ,2002.
18. J. L. Neto, A. D. Santos, C. A. A. Kaestner and A. A Freitas. "Document Clustering and Text Summarization". Proc. of the 4th International Conference on Practical Applications of Knowledge Discovery and Data Mining. April. London, 41-55, 2000.
19. R. Ribeiro and D. M. d. Matos. "Extractive Summarization of Broadcast News: Comparing Strategies for European Portuguese". In V. M. sek, and P. Mautner, (Eds.). Text, Speech and Dialogue. (pp. 115–122). Berlin Heidelberg: Springer-Verlag, 2007.
20. E. Villatoro-Tello, L. Villaseñor-Pineda and M. Montes-y-Gómez. "Using Word Sequences for Text Summarization". In Sojka, P., Kopeček, I., Pala, K. (eds.). Text, Speech and Dialogue. vol. 4188: 293–300. Berlin Heidelberg: Springer-Verlag, 2006.
21. S. Ye, L. Qiu, T. Chua and M. Kan. "NUS at DUC 2005: Understanding documents via concept links". Proceedings of Document Understanding Conference. 9-10 October. Vancouver, Canada, 2005.
22. D. M. Zajic. "Multiple Alternative Sentence Compressions As A Tool For Automatic Summarization Tasks". PhD theses. University of Maryland, 2007.
23. D. M. Zajic, B. J. Dorr, R. Schwartz, and J. Lin. "Sentence Compression as a Component of a Multi-Document Summarization System". Proceedings of the 2006 Document Understanding Workshop. 8-9 June. New York, 2006.
24. H. Zha. "Generic summarization and key phrase extraction using mutual reinforcement principle and sentence clustering". In proceedings of 25th ACM SIGIR. 11-15 August. Tampere, Finland, 113-120, 2002.
25. X. Zhu, A. B. Goldberg, J. V. Gael and D. Andrzejewski. "Improving diversity in ranking using absorbing random walks". HLT/NAACL. 22-27 April. Rochester, NY, 2007.

Asking Users: A Continuous Usability Evaluation on a System Used in the Main Control Room of an Oil Refinery Plant

Suziah Sulaiman
Dayang Rohaya Awang Rambli
Wan Fatimah Wan Ahmad
Halabi Hasbullah
Foong Oi Mean
M Nordin Zakaria
Goh Kim Nee
Siti Rokhmah M Shukri

*Computer and Information Sciences Department
Universiti Teknologi PETRONAS
31750 Tronoh, Perak, Malaysia*

suziah@petronas.com.my
roharam@petronas.com.my
fatimhd@petronas.com.my
halabi@petronas.com.my
foongoimean@petronas.com.my
nordinzakaria@petronas.com.my
gohkimnee@petronas.com.my
sitirohkmah@petronas.com.my

ABSTRACT

This paper presents a case study that observes usability issues of a system currently used in the main control room of an oil refinery plant. Poor usability may lead to poor decision makings on a system, which in turn put thousands of lives at risk, and contributes to production loss, environmental impact and millions dollar revenue loss. Thus, a continuous usability evaluation on an existing system is necessary to ensure meeting users' expectation when they interact with the system. Seeking users' subjective opinions on the usability of a system could capture rich information and complement the respective quantitative data on how well the system is in supporting an intended activity, as well as to be used for system improvement. The objective of this survey work is to identify if there are any usability design issues in the systems used in the main control room at the plant. A set of survey questions was distributed to the control operators of the plant in which 31 operators responded. In general, results from the quantitative data suggest that respondents were pleased with the existing system. In specific, it was found that the experienced operators are more concerned with the technical functionality of the system, while the lesser experienced are towards the system interface. The respondents' subjective feedback provides evidences that strengthen the findings. These two concerns however, formed part of the overall usability requirements. Therefore, to continuously improve the usability of the systems, we strongly suggest that the system be embedded with these usability aspects into its design requirements.

Keywords: usability, evaluation, continuous improvement, decision making.

1. INTRODUCTION

Issues pertaining to user interface design are not new. It started as early as the 1960s and has evolved ever since. However, designing a user interface especially for systems in a control room is still a challenging task. Having an appropriate user interface design that includes the display and control design, console layout, communications, and most importantly, the usability of the system to be addressed by and made to help users is important in control room systems [1]. A huge amount of information needs to be presented on the screen in order for the users to monitor the system. Therefore, designers need to be careful as not to impose a cognitive workload to the users when interacting with the system. A continuous assessment on the users' performances may help in determining if such an issue exists [2]. In this case, understanding the users' subjective experience interacting with the existing system in order to capture qualitative information is necessary [3] in order to decide if any improvements are needed; hence, ensuring the usability of the system.

One important preparation before evaluating an existing system is addressing the question of what to evaluate from the system. The phrase usability and functionality as two sides of the same coin could possibly provide an answer to this issue. The usability of the system and, the requirements analysis on the functionality are two aspects in the system development lifecycle that need to be emphasized [4,5]. The evaluation should focus on a thorough approach that provides a balance between the meaning of the visualization elements that conform to the mental model of an operation, and what lies beneath these visual representations i.e. functionality from a technical engineering perspectives.

This paper attempts to examine operators' opinions when interacting with an interface design of systems used in a control room of an oil refinery. The intention is to provide a case study that emphasises on the importance of a continuous assessment. The paper includes a section on related work and follows with a survey study conducted to elicit users' opinions on the user interface design. The survey study uses both quantitative [6] and qualitative data [7,8] in evaluating an existing system [9,10]. The evaluation involves assessing the technical functionality and usability of the systems. A claim based on the study findings is suggested and discussed at the end of the paper.

2. RELATED WORK

Studies that involve evaluation of user interface design in various control room environments such as in the steel manufacturing industry [11], transportation [12], power plant [13], and refineries [12,13,14,15] have frequently been reported. Even though there are many challenges involved in the evaluation process, there is still a pattern in terms of the study focuses that could be found. Two main focuses from these studies are: those pertaining to designing for an interactive control room user interface, and applying various types of interface design into industrial applications.

Designing the user interface for control rooms is the most common topic found. In most cases, the new design is an enhancement based on existing systems after seeking for users' feedback [2]. The methods and procedures for developing the graphical user interfaces of a process control room in a steel manufacturing company have been described in Han et al.'s [11]. A phase-based approach was used in the study after modifying the existing software development procedures to emphasize the differences between the desktop tasks and control room tasks. With GUI-based human computer interface method, the users were able to monitor and control the manufacturing processes. A more explicit explanation that details out the approach when designing the user interface could be found in Begg et al.'s [13]. A combination of techniques i.e. questionnaires, interviews, knowledge elicitation techniques, familiarization with standard operating procedures, and human factors checklist was used in order to obtain the user requirements for the control system. Similar to Begg et al.'s [13] approach, Chen et al. [16] included an initial study that consists of several techniques to gather information for the user requirements. Chen et al.'s work

is more extensive in which it involves a development and an evaluation of a user interface suitable for an Incident Management System. Although Guerlain et al. [12] mainly described how several design principles could be applied to represent data into hierarchical data layers, interviews and observations on the operators using the controllers were still conducted. A common feature found in all studies reported here is the users' subjective experience sought in order to improve the user interface design of the system.

Applying various types of interface design into industrial applications [14] is another area in user interface design for control room environments. The work reported involves applying an ecological interface design into the system, aiming towards providing information about higher-level process functions. However, the work did not involve eliciting user subjective experience as it was not within the scope of study.

Despite many enhancements on the user interface designs done based upon evaluation of the existing systems, those reported work [11,12,13,16] lacks attention on summative evaluation [17]. Such a pattern could result in lesser emphasis given on the evaluation of the whole system development life cycle; hence, not fully meeting the goal of a summative evaluation that is judging the extent to which the system met its stated goals and objectives and the extent to which its accomplishments are the result of the activities provided. In order to ensure that the usability of the system is in place, a continuous evaluation even after the deployment of a system to the company is needed. Rather than checking only on certain features of the system, such an evaluation should involve assessing the functionality as well as the usability of the interface design. Thus, summative evaluation should be performed even after beta testing and perhaps beyond the product released stage.

Recently, the way in which usability evaluation is performed by the HCI communities has been heavily criticized because at times the choice of evaluation methods is not appropriate to the situations being studied. Such a choice could be too rigid that hinders software designers from being creative in expressing their ideas; hence the designs [18]. Greenberg and Buxton in [18] suggest for a balance between objectivity and subjectivity in an evaluation. By being objective, one is seeking for assurance on the usability of the system through quantitative empirical evaluations while being subjective focuses on qualitative data that is based on users' expressed opinions. A similar argument of objectivity over subjectivity has also been raised in other design disciplines as noted in Snodgrass and Coyne [19]. The issue raised signals for a need to incorporate both quantitative and qualitative approaches during the summative evaluation.

3. SURVEY STUDY

The pattern from reported work presented in the earlier section indicates that a new system is usually developed based on the limitations found in the system currently being used. These limitations could be identified when evaluations that include some forms of a qualitative approach are used. Based on this pattern, a survey was conducted at an oil refinery in Malaysia.

The objective of the survey was to identify if there are any usability issues in the systems used in the main control room at the plant. The idea is to ensure usability and user experience goals [13] are maintained throughout the system life cycle. By asking users through a survey, an example on how both quantitative and qualitative data could complement one another could be demonstrated; hence, assisting in achieving the objective.

The target group for the survey was the panel operators at the plant. 31 operators responded to the survey. The survey questions could be divided into three parts, in which a mixture of both quantitative and qualitative questions was used to benefit from the study. Part 1 covers demographic questions regarding the panel operator's background working in the plant. Part 2 involves seeking quantitative data from the users. It investigates the usefulness of the system(s) to the panel operators (users). The questions in this part could be divided into two groups i.e.

those pertaining to the technical functionalities and, that on the usability of the system. Finally, Part 3 involves those that elicit their subjective experience when interacting with the user interface of the system(s) used in the control room.

4. RESULTS AND ANALYSIS

For easy referencing and clarity purposes, the study findings presented in this paper will follow the sequence of parts as described in Section 3. In this case, the results and analysis of findings in Part 1 that covers the demographic questions will be described in Section 4.1. Similarly, quantitative data collected from Part 2 will be presented in Section 4.2. Likewise, the qualitative data obtained in Part 3 will be discussed in Section 4.3. These findings are primarily used as a basis to justify and complement those found in Part 2.

4.1 Findings on the respondents' background

All 31 respondents were male. The average age was between 31 to 40 years old. Most of them have been working at the company for more than 10 years (Figure 1) but majority has only 1-3 years experience as panel operator (Figure 2). From this finding, 2 groups of panel operators are formed: more experienced and less experienced to represent the former and latter groups, respectively. These groupings will be analysed and referred to frequently in this paper.

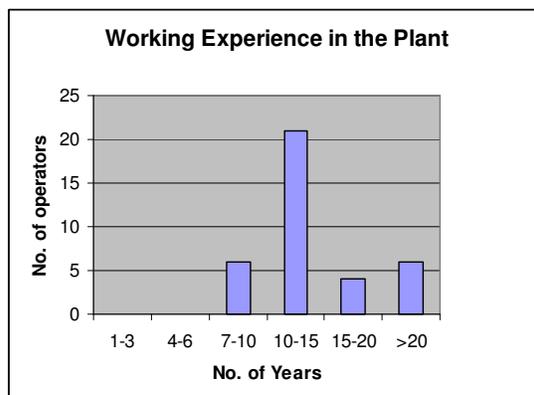


FIGURE 1: No. of years working in a plant

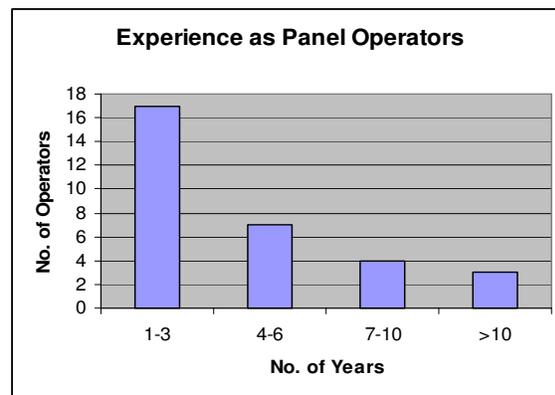


FIGURE 2: Experience as a panel operator

There are two main systems currently used to control the processes in the plant: Plant Information System (PI) and Distributed Control System (DCS). Based on the survey findings, about 90% of the respondents interact frequently with DCS while the rest with PI system.

4.2 Checking the usefulness of the existing systems – quantitative data

The usefulness of the system, mainly the DCS was initially determined based on the quantitative data collected. Each usability feature was rated by the respondents using a 5 ranking scale i.e. '1' indicates 'never', '2' is for 'almost never', '3' means 'sometimes', '4' is 'often', and '5' indicates 'always', accordingly. The study findings reveal that none of the respondents rated 'never' and very few rated 'almost never' in the survey, indicating that overall the respondents are happy with the existing system. One possible reason could be due to their familiarity interacting with the applications. This may be the only systems that they have been exposed to when working as a panel operator. Quite a number of responses were received for the 'Sometimes' category but these are not very interesting to be analysed further as they may signal a neutral view from the respondents. In this case, only those rated for 'Often' and 'Always' categories are being analysed in this study as they imply definite responses from the respondents. The summary of findings is presented in Table 1.

| Element | Usability Features | Frequency of rating according to years of experience | | | | | | | | | |
|--------------------------------|---|--|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | | Never | | Almost Never | | Sometimes | | Often | | Always | |
| | | Less experienced | More experienced | Less experienced | More experienced | Less experienced | More experienced | Less experienced | More experienced | Less experienced | More experienced |
| <i>Technical Functionality</i> | The system provides info I was hoping for | 0 | 0 | 2 | 1 | 1 | 3 | 8 | 6 | 6 | 4 |
| | The system provides sufficient content for me | 0 | 0 | 1 | 0 | | 2 | 13 | 8 | 3 | 4 |
| | The content of the system is presented in a useful format | 0 | 0 | 1 | 0 | 1 | 3 | 9 | 6 | 6 | 3 |
| | | Average | | | | | | 10 | 6.7 | 6.3 | 3.7 |
| <i>Usability of the System</i> | The content presentation is clear | 0 | 0 | 1 | 1 | 3 | 1 | 10 | 9 | 3 | 3 |
| | The system is user friendly | 0 | 0 | 1 | 0 | 1 | 1 | 13 | 7 | 2 | 6 |
| | The system is easy to use | 0 | 0 | 1 | 0 | 1 | 2 | 10 | 6 | 4 | 6 |
| | It is easy to navigate around the system | 0 | 0 | 1 | 0 | 3 | 3 | 9 | 7 | 4 | 3 |
| | I can easily determine where I am in the system at any given time | 0 | 0 | 2 | 2 | 3 | 2 | 9 | 7 | 3 | 3 |
| | The system uses consistent terms | 0 | 0 | 2 | 1 | 2 | 1 | 8 | 8 | 5 | 4 |
| | The system uses consistent graphics | 0 | 0 | 2 | 1 | 1 | 2 | 9 | 8 | 5 | 3 |
| | | Average | | | | | | 10 | 7.4 | 3.7 | 4 |

TABLE 1: Quantitative Data Analysed

Table 1 shows the frequency of each usability feature being rated based on the respective ranking i.e. from 'never' until 'always'. For each rank, a further grouping is provided to cater for respondents who have experience working as a panel operator for less than 3 years and, those that have more.

As our main concerns are on those who rated for the 'Often' and 'Always' categories, so only the highlighted columns in Table 1 will be discussed. The average values for category 'Often' are higher than that for 'Always' in both experienced and less experienced panel operators. This indicates that although the operators are satisfied with the current system, there are still some features that require improvement.

Comparing the average values in the 'Technical Functionality' element for experienced operators with less experience (i.e. 10 and 6.7 for 'Often' and '6.3 and '3.7' for 'Always'), we could conclude that with regards to the technical functionality of the system, those who have more experience tend to feel that the technical content is not adequate as compared to those have less experience. This is when the average values for experienced operators are lesser than the less experienced in both categories. However, this is not the case in the 'Usability of the System' group whereby the more experienced operators rank slightly higher (average = 4) than the less experienced (average = 3.7) in the 'Always' category. This could signal that the more experienced operators felt that the usability of the system is more important as compared to the lesser experienced operators. One possible reason for this pattern could be due to the familiarity aspect

whereby the more experienced operators are more familiar with the system and looking for more usefulness features of the system as compared to those with lesser experienced.

When comparing the 'Often' and 'Always' columns to examine the 'Technical Functionality' and 'Usability of the System' elements, the findings from the 'Always' columns indicate that the experienced group is more concerned with the technical functionality of the system while the less experienced is on the usability of the interface. This is derived from the average value that shows a slightly lesser value for more experienced as compared to the less experienced in the 'Technical Functionality' section and vice-versa for the 'Usability of the System'.

4.3 User Subjective Experience on the Usability of the Existing Systems

The qualitative data collected from the survey is used to strengthen and provide justifications for the findings presented in Section 4.2. The subjective data were compiled, analysed, and grouped according to the common related themes. These groupings are presented as categories shown in Table 2.

| <i>Categories</i> | <i>Details</i> |
|--------------------|---|
| Ease of use | The current systems have met usability principles that include: <ul style="list-style-type: none"> • user friendly (Respondent Nos. 5, 6, 7, 12, 15, 20, 21, 24, 25, 26) • easy to learn (Respondent No. 12) • easy to control (Respondent Nos. 6, 7, 8, 13, 15, 16, 17, 18, 20, 22) • and effective (Respondent No. 14) |
| Interaction Styles | The touch screen provides a convenient way of interacting with the systems (Respondent Nos. 15, 17, 24). |
| User Requirements | The systems are functioning well and following the specifications. <i>“Parameters of the system is (are) mostly accurate”</i> (Respondent No. 14). Thus, resulted in <i>“achieve (achieving) greater process stability, improved product specifications and yield and to increase efficiency.”</i> (Respondent No. 10) |
| Information Format | The information has been well presented in the systems. Most of the process parameters are displayed at DCS (Respondent No. 11).The users have positive feeling towards the visual information displayed on the screen. They are pleased with the trending system (Respondent Nos. 7, 21, 24), graphics (Respondent Nos. 3, 27), colour that differentiates gas, hydrocarbon, water etc. (Respondent Nos. 8, 13, 14). |

TABLE 2: Subjective responses on the positive aspects

Table 2 shows four main categories that were identified from the subjective responses regarding the existing systems. The 'Ease of Use' category that refers to the usability principles and the 'Interaction Styles', mainly about interacting with the system, should be able to support the positive feedback given for the user interface design element presented in Section 4.2. On the other hand, both the 'User Requirements' and 'Information Format' categories could be a reason why the operators are happy with the content provided in the system.

4.4 Improvement for the current system

The subjective user feedback from the survey could also be used to explain the reasons for the slight differences between the more experienced operators and those with less experienced. From the previous section, it has been indicated that overall, the more experienced operators have some reservations towards the content of the system. Based on the data compiled as mentioned in Section 4.3, the subjective findings pertaining to issues raised by the respondents were analysed and categorised. These categories and their details are presented in Table 3.

| <i>Categories</i> | <i>Details</i> |
|----------------------------|--|
| Information representation | <p>The current systems may not be able to provide adequate information to the users; hence, they require input from other resources.</p> <p><i>“Receive less feedback or information compared to my senior panelman.”</i> - (Respondent No. 18) <i>“I receive less feedback and information compared to my senior panelman.”</i>- (Respondent No. 5) <i>“Not 100% but can access at other source.”</i>- (Respondent No. 7)</p> <p>Several respondents expressed their opinions to improve the information representation. Respondent No. 15 said: <i>“Need extra coaching about the system”</i> while Respondent No. 11 commented: <i>“During upset condition, we need SP, PU and OP trending. Add (link) PI system to DCS”</i>.</p> |
| Trending | <p>Trend information is used to assist panel operators to monitor and control the processes in the plant. Feedbacks received pertaining to this issue are <i>“no trending of output in History Module of Honeywell DCS”</i> (Respondent #11) and the <i>“need to learn how to do the graphic”</i> (Respondent No. 15).</p> |

TABLE 3: Subjective responses on the content

From Table 3, both categories presented are related to the content of the system. With the exception of Respondent No. 5, the rest of the respondents who commented on the content have less than 3 years experience being a panel operator. This could imply that overall, majority of the respondents in the survey feel that additional features should be made available in the existing system in order to increase their work performances.

Similarly, several issues were also raised by the panel operators when interacting with the system’s user interface design. The qualitative data that reveals this information is presented in Table 4.

| <i>Categories</i> | <i>Details</i> |
|-------------------|--|
| System Display | <p>The black background color of the current system interfaces is causing discomfort to the users (Respondent Nos. 20, 21, 24, 25). Some of the discomfort reported includes glaring and eye fatigue. Such background color drowns down the information on the screen such as the font and color of the texts. Improvement to adjust the contrast setting (Respondent No. 24) and to <i>“change the screen background color”</i> (Respondent No. 13) are proposed.</p> |
| System Design | <p>With respect to the system design, the most frequent comments were on managing alarms (Respondent Nos. 2, 6, 10, 17, 22, 28). Whenever there are too many alarm signals, the alarm management system will malfunction (Respondent Nos. 2, 17). Users expressed their frustration when this happens (Respondent No. 10). One main concern arises when interacting with PI and DCS systems is <i>“so many alarms and how to manage alarm”</i> (Respondent No. 28). This is strengthened by another claim that says <i>“cannot handle repeatability alarm”</i> (Respondent No. 6).</p> |

TABLE 4: Subjective responses on the user interface design

Table 4 consists of ‘System Display’ and ‘System Design’ categories that are identified from comments made by a balanced mixture of panel operators from the more experienced and less

experienced groups. The comments made on the user interface design in the system are mainly pertaining to the screens and interface elements (e.g. colour and font of the text). This corresponds to Ketchel and Naser [1] findings to emphasize on the importance of choosing the right color and font size for information presentation. Managing the alarm system is also an issue raised in this category. The frequency of the alarms frustrates the operators especially when the warnings are of minor importance. This issue needs addressing in order to reduce operators' cognitive workload in an oil refinery plant.

Besides the feedback received pertaining to the content, and user interface design, another important issue raised by the panel operators is mainly on the working environment. The current situation could affect the performance of the workers. Respondent No. 27 commented on the "contrast; lack (of) light" in the control room. Improvement on the system may be able to reduce the workers' negative moods. "As panel man, you (are) always in (a) positive mood; you must put yourself in (a) happy/cool/strategic when you face a problem" (Respondent No. 14). He added that so far the system is good but an improved version would be better still as panel operators could get bored if they have to interact with the same thing each time.

5. CONCLUSIONS & FUTURE WORK

The main aim of this paper is to emphasise that a continuous assessment on existing systems is necessary to maintain the system usability and at the same time examine if any improvements are required. This has been demonstrated in a case study that uses a survey to elicit comments from a group of panel operators from the main control room of an oil refinery plant. In doing so, the capability of both quantitative and qualitative data have been utilised. The combination of these two approaches benefits evaluation activities as findings from each complement the other.

The study result has suggested that although in general the panel operators are happy with the existing system in terms of its technical functionality and user interface design, there are still enhancement to be made on the system. While the more experienced panel operators are more concern about the technical functionality issues of the system, the less experienced tend to focus on the system interface. It could be argued that should both concerns are addressed the overall user requirements could be better met. This is supported from the fact that usability and functionality are two elements of equal importance in a system.

A future work could be suggested based on the issues raised from the study findings. Users' feedback indicates that "automation" of knowledge based on previous experience is necessary to assist them in their work. This could be made available by having an expert system and should be made accessible to all panel operators. Such a system may be necessary especially when most respondents in the survey have less than 3 years experience working as a panel operator. In developing the expert system, collective opinions from both experienced and lesser experienced operators is required in order to obtain a more complete set of design requirements.

Acknowledgement

The authors would like to thank all the operators who took part in the survey study.

6. REFERENCES

1. J. Ketchel and J. Naser. "A Human factors view of new technology in nuclear power plant control rooms". In Proceedings of the 2002 IEEE 7th Conference, 2002
2. J.S. Czuba and D.J. Lawrence. "Application of an electrical load shedding system to a large refinery". IEEE, pp. 209-214, 1995

Suziah Sulaiman, Dayang Rohaya Awang Rambli, Wan Fatimah Wan Ahmad, Halabi Hasbullah, Foong Oi Mean, M Nordin Zakaria, Goh Kim Nee & Siti Rokhmah M Shukri

3. S. Faisal, A. Blandford and P. Cairns. *"Internalization, Qualitative Methods, and Evaluation"*. In BELIV '08, 2008
4. J. Preece, Y. Rogers and H. Sharp, *"Interaction Design: Beyond Human Computer Interaction"*, John Wiley & Sons Inc. (2002)
5. S. Sulaiman. *"Usability and the Software Production Life Cycle"*, In Electronic Proceedings of ACM CHI, Vancouver, Canada, 1996.
6. C.M. Burns, Jr. Skraaning , G. Jamieson, G. A. Lau, N. Kwok, J. R. Welch and G. Andresen, *"Evaluation of Ecological Interface Design for Nuclear Process Control: Situation Awareness Effects"*, International Journal of Human Factors, vol. 50, pp. 663-679, 2008
7. Vicente, K. J., Mumaw, R. J., & Roth, E. M., "Operator monitoring in a complex dynamic work environment: A qualitative cognitive model based on field observations", Journal of Theoretical Issues in Ergonomics Science, vol. 5, pp. 359-384, 2004.
8. P. Savioja and L. Norros . "Systems Usability -- Promoting Core-Task Oriented Work Practices" ,In Law E. et.al. (eds.) *Maturing Usability*. London: Springer(2008).
9. L. Norros and M. Nuutinen. *"Performance-based usability evaluation of a safety information and alarm system"*, International Journal of Human-Computer Studies 63(3): 328-361, 2005.
10. P. Savioja, L. Norros and L. Salo. "Evaluation of Systems Usability", Proceedings of ECCE2008, Madeira, Portugal, September 16-19, 2008.
11. S.H. Han, H. Yang and D.G. Im. *"Designing a human-computer interface for a process control room: A case study of a steel manufacturing company"*. International Journal of Industrial Ergonomics, 37(5):383-393, 2007
12. S. Guerlain, G. Jamieson and P. Bullemer. *"Visualizing Model-Based Predictive Controllers"*. Proceedings of the IEA 2000/HFES 2000 Congress, 2000
13. I.M. Begg, D.J. Darvill and J. Brace. "Intelligent Graphic Interface Design: Designing an Interactive Control Room User Interface", IEEE, pp. 3128 – 3132, 1995
14. G.A. Jamieson. *"Empirical Evaluation of an Industrial Application of Ecological Interface Design"*, In Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA:Human Factors and ergonomics Society, October 2002
15. G.A. Jamieson, "Ecological interface design for petrochemical process control: An empirical assessment," IEEE Transactions on Systems, Man And Cybernetics, vol. 37, pp. 906-920, 2007
16. F. Chen, E.H.C. Choi, N. Ruiz, Y. Shi and R. Taib, R. "User Interface Design and Evaluation For Control Room, *Proceedings of OZCHI 2005*, Canberra, Australia, November 23-25, 2005, pp. 1-4.
17. J. Preece, Y. Rogers and H. Sharp *"Interaction Design: Beyond Human-Computer Interaction"*, John Wiley & Sons Inc.(2002)
18. S. Greenberg and B. Buxton. *"Usability Evaluation Considered Harmful?"* In Proceedings of ACM CHI. Florence, Italy, 2008
19. A. Snodgrass and R. Coyne. "Interpretation in Architecture: design as a way of thinking", London:Routledge (2006)

Exploring Knowledge for a Common Man Through Mobile Services and Knowledge Discovery in Databases

Sanjeev Manchanda*

*School of Mathematics and Computer Applications,
Thapar University, Patiala-147004 (INDIA)*

**Corresponding author*

smanchanda@thapar.edu

Mayank Dave

*Department of Computer Engg.,
National Instt. Of Technology, Kurukshetra, India*

mdave67@yahoo.com

S. B. Singh

*Department of Mathematics,
Punjabi University, Patiala, India*

sbsingh69@yahoo.com

ABSTRACT

Every common man needs some guidance/advice from his/her friends/relatives/known one, whenever he/she wants to buy something or want to visit somewhere. In this paper authors propose a system to fulfill the knowledge requirements of a common man through mobile services using data mining and knowledge discovery in databases. This system will enable us to furnish such information to a common man without any cost or at a low cost and with least effort. A comparison of proposed system is provided with reference to other available systems.

Keywords: Data mining, Knowledge Discovery in Databases, Global Positioning System, Geographical Information System, Mobile Services.

1. INTRODUCTION

Science and technology in its roots is meant for serving human community in masses. But latest trends are away from their original thought. If we look towards Indian scenario, we have started working for big organizations, which are meant for making big profits rather working for the uplifting the society. This paper is an effort to help a common human being through automating knowledge discovery to contribute him/her in its day-to-day requirements. Information Technology and communication has facilitated a common person to connect one to whole world. It is one of the biggest achievements of science and technology. In continuation of same technology and incorporation of data mining and knowledge discovery in databases can help anyone to gain some knowledge through the information stored in databases scattered worldwide. This paper will furnish a base to fulfill the information/knowledge requirements of a common human being in its day-to-day proceedings. Present paper presents architecture of system to develop a sophisticated system to furnish customized information needs of users. Present paper also compares proposed system with other major competitive systems. This paper is organized as following section introduces the foundational concepts behind the development of proposed system. Third section contains literature review, Fourth section describes the problem, Fifth section briefly describes possible solution, Sixth section presents system's architecture including algorithm of the system, Seventh section discusses issues involved in developing and commercializing proposed system, Eighth section includes the description of technology used for development of system, Ninth section compares proposed system with other systems as well as other results, Tenth section discusses current state of the system, Eleventh section concludes the paper as well as discusses future aspects and last but not the least Twelfth section contains references used in present paper.

1 2. KDD, GPS, GIS and Spatial Databases

2

3 2.1 Data mining and knowledge discovery in databases

Knowledge discovery in databases (KDD) is a drift to search new patterns in existing databases through the use of data mining techniques. The techniques used earlier to search data in response to queries having fixed domains have changed to search unknown patterns or vaguely defined queries.

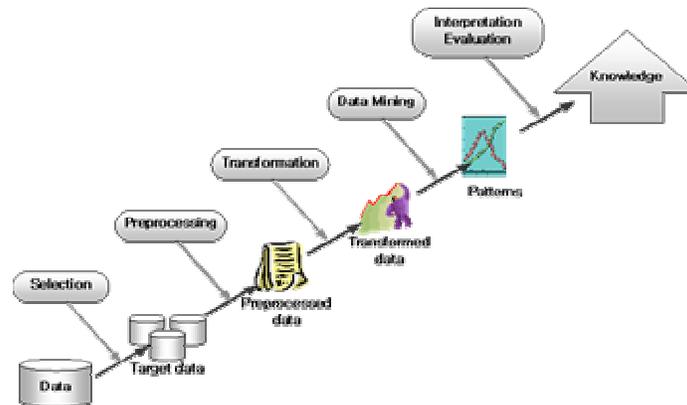


FIGURE 1: KDD Process

2.2 Geographic Information System

Geographic Information System (GIS) is a computer based information system used to digitally represent and analyse the geographic features present on the Earth' surface and the events

(non-spatial attributes linked to the geography under study) that taking place on it. The meaning to represent digitally is to convert analog (smooth line) into a digital form.

2.3 Global Position System

The Global Positioning System (GPS) is a burgeoning technology, which provides unequalled accuracy and flexibility of positioning for navigation, surveying and Geographical Information System data capture. The GPS NAVSTAR (Navigation Satellite timing and Ranging Global Positioning System) is a satellite-based navigation, timing and positioning system. The GPS provides continuous three-dimensional positioning 24 hrs a day throughout the world. The technology seems to be beneficiary to the GPS user community in terms of obtaining accurate data up to about 100 meters for navigation, meter-level for mapping, and down to millimeter level for geodetic positioning. The GPS technology has tremendous amount of applications in GIS data collection, surveying, and mapping.

2.4 Spatial Database

A spatial database is defined as a collection of inter-related geospatial data, that can handle and maintain a large amount of data which is shareable between different GIS applications. Required functions of a spatial database are consistency with little or no redundancy, maintenance of data quality including updating, self descriptive with metadata, high performance by database, management system with database language and - security including access control

3. Literature Review

Recently, the data mining and spatial database system has been subjects of many articles in business and software magazines. However, two decade before, few people had heard of the terms data mining and spatial/geographical information system. Both of these terms are the evolution of fields with a long history, the terms themselves were only introduced relatively recently, in late 1980s. Data mining roots are traced back along three family lines viz. statistics, artificial intelligence and machine learning, whereas roots of Geographical Information System lies from the times when human being started traversing to explore new places. Data mining, in many ways, is fundamentally the adaptation of machine learning techniques to business applications. It is best described as the union of historical and recent developments in statistics, artificial intelligence and machine learning. These techniques are then used together to study data and find hidden trends or patterns within. Data mining and Geographical Information systems are finding increasing acceptance in science and business areas that need to analyze large amounts of data to discover trends, which they could not otherwise find. Combination of geography and data mining has generated a great need to explore new dimensions as Spatial Database Systems.

Formal foundation of data mining lies from a report on (International Joint Conference on Artificial Intelligence) IJCAI-89 Workshop (Piatetsky 1989), in which he emphasized on the need of data mining due to growth in amount of large databases. This report confirmed the recognition for the concept of Knowledge Discovery in Databases. Data mining term got its recognition from this report and became familiar in scientific community soon. Initial work was being done on classification and logic programming in 1991. Han et al., 1991 emphasized on concept based classification in Relational database. He devised a method for the classification of data in relational databases by concept-based generalization and proposed an algorithm concept-based data classification algorithm called DBCLASS, whereas Bocca, 1991 discussed about design and engineering of the enabling technologies for building Knowledge Based Management Systems (KBMS). He showed the problems with the required solutions of existing technologies commercially available i.e., relational database systems and logic programming. Aggarwal et al., 1992 discussed about the problems with classifying data populations and samples into m group intervals. They proposed a tree based interval classifier, which generates a classification function for each group that can be used to efficiently retrieve all instances of the specified group from the population database.

The term "spatial database system" has become popular during the last few years, to some extent through the *Symposium on Large Spatial Databases*, which has been held biannually

since 1989 (Buchmann et al. 1990, Giinther et al. 1991 and Abel et al. 1993). This term is associated with a view of a database as containing sets of objects in space rather than images or pictures of a space. Indeed, the requirements and techniques for dealing with objects in space that have identity and well-defined extents, locations, and relationships are rather different from those for dealing with raster images. It has therefore been suggested that two classes of systems, *spatial database systems* and *image database systems*, be clearly distinguished (Buchmann et al. 1989, Frank 1991). Image database systems may include analysis techniques to extract objects in space from images, and offer some spatial database functionality, but they are also prepared to store, manipulate, and retrieve raster images as discrete entities.

Han et al., 1994 tried to explore whether clustering methods had a role to play in spatial data mining. They developed a clustering algorithm called CLARANS, which is based on randomized search and two data mining algorithms Spatial Dominant Approach (SD) and Non-Spatial Dominant Approach (NSD) and showed the effectiveness of algorithms. Mannila et al., 1994 revised the solution to the problem raised by Agarwal Rakesh et al. 1993 and proposed an algorithm to improve the solution. Han et al., 1994 studied the construction of Multi Layered Databases (MLDBs) using generalization and knowledge discovery techniques and the application of MLDBs to cooperative/intelligent query answering in database systems. They proposed an MLDB model and examined in their study and showed the usefulness of MLDB in cooperative query answering, database browsing and query optimization. Holsheimer et al., 1994 surveyed the data mining researches of that time and presented the main underlying ideas behind data mining such as inductive learning, search strategies and knowledge representations used in data mine systems, also described important problems and suggested their solutions. Kawano et al., 1994 integrated knowledge sampling and active database techniques to discover interesting behaviors of dynamic environments and react intelligently to the environment changes. Their studies showed firstly that data sampling was necessary in the collection of information for regularity analysis and anomaly detection, Secondly knowledge discovery was important for generalizing low-level data to high-level information and detecting interesting patterns, Thirdly active database technology was essential for the real time reaction to the changes in real time environment and lastly the integration of three technologies forms a powerful tool for control and management of large dynamic environments in many applications.

Data Classification technique contributed by Han et al. 1995, Kohavi 1996, Micheline et al. 1997, Li et al. 2001 and others, Han et al., 1995, 1998 stated that their research covered a large wide spectrum of knowledge discovery, which included the study of knowledge discovery in relational, object oriented, deductive, spatial, active databases and global information systems and development of various kinds of knowledge discovery methods including attribute-oriented induction, progressive deepening for mining multiple level rules, meta-guided knowledge mining *etc.* and also studied algorithms for the techniques of data mining. Later he investigated the issues on generalization-based data mining in object-oriented databases in three aspects generalization of complex objects, class based generalization and extraction of different kinds of rules. Their studies showed that set of sophisticated generalization operators could be constructed for generalization of complex objects, dimension based class generalization mechanism could be developed for class based generalization and sophisticated rules formation method could be developed for extraction of different kinds of rules.

The development of specialized software for spatial data analysis has seen rapid growth since the lack of such tools was lamented in the late 1980s by Haining, 1989 and cited as a major impediment to the adoption and use of spatial statistics by GIS researchers. Initially, attention tended to focus on conceptual issues, such as how to integrate spatial statistical methods and a GIS environment (loosely vs. tightly coupled, embedded vs. modular, etc.), and which techniques would be most fruitfully included in such a framework. Familiar reviews of these issues are represented in, among others, Anselin et al., 2000, Goodchild et al. 1992, Fischer et al., (1993, 1996, 1997), Fotheringham et al., (1993, 1994). In geographical analysis by Monmonier, (1989) made operational Spider/Regard toolboxes of Haslett, Unwin and associates (Haslett et al. 1990, Unwin 1994). Several modern toolkits for exploratory spatial data analysis (ESDA), also incorporate dynamic linking and to a lesser extent brushing. Some of these rely on interaction with a GIS for the map component, such as the linked frameworks combining XGobi or XploRe with ArcView (Cook et al. 1996, Symanzik et al. 2000), the SAGE toolbox, which uses ArcInfo

(Wise et al., 2001), and the DynESDA extension for ArcView (Anselin, 2000), GeoDa's immediate predecessor. Linking in these implementations is constrained by the architecture of the GIS, which limits the linking process to a single map (in GeoDa, there is no limit on the number of linked maps). In this respect, GeoDa is similar to other freestanding modern implementations of ESDA, such as the cartographic data visualizer, or cdv (Dykes, 1997), GeoVISTA Studio (Takatsuka et al., 2002) and STARS (Rey et al., 2004). These all include functionality for dynamic linking, and to a lesser extent, brushing. They are built in open source programming environments, such as Tcl/Tk (cdv), Java (GeoVISTA Studio) or Python (STARS) and thus easily extensible and customizable. In contrast, GeoDa is (still) a closed box, but of these packages it provides the most extensive and flexible form of dynamic linking and brushing for both graphs and maps.

Common spatial autocorrelation statistics, such as Moran's I and even the Local Moran are increasingly part of spatial analysis software, ranging from CrimeStat (Levine, 2004), to the spdep and DCluster packages available on the open source Comprehensive R Archive Network (CRAN),³ as well as commercial packages, such as the spatial statistics toolbox of the forthcoming release of ArcGIS 9.0 (ESRI, 2004). Continuous space in spatial data was discussed and presented join less approach for mining spatial patterns (Yoo et al. 2006).

One major aspect of any such systems is user satisfaction. User satisfaction depends on many aspects like usability, Accuracy of product, information quality etc. Usability is one of the most important factors in the phases of designing up to selling a product. (Nielsen, 1993). But the efficiency of a product is influenced by the acceptance of the user. Usability is one basic step to acceptance and finally to efficiency of a product. A new approach is the "User Centred Design", UCD. The prototyping is described with the ISO standard 13407: "Human centred design process for interactive systems". The main mantras used here are "Know your user!" and "You aren't the user!" Both slogans describe the importance of the user.(Fröhlich et al., 2002) Concluding from the own experience as a user to other user groups is precarious and should be avoided. It is only possible to understand the user groups and the context of usage by careful analysis. (Hynek, 2002). The User Centred Design focuses on the user and their requirements from the beginning of achieving a product. "Usability goals should drive design. They can streamline the design process and shorten the design cycle." (Mayhew, 2002) Factors like reliability, compatibility, cost, and so on affect the user directly. Usability factors influence the decision of the user indirectly and can lead to subconscious decisions that are hardly traceable. The Information Quality (IQ) is the connector between Data Quality and the user. General definitions for IQ are "fitness for use" (Tayl et al., 1998), "meets information consumers needs" (Redman, 1996) or "user satisfaction" (Delone et al., 1992). This implies data that are relevant to their intended use and are of sufficient detail and quantity, with a high degree of accuracy and completeness, consistent with other sources, and presented in appropriate ways. Many criteria depend on each other and in this case not all criteria will be used. The information quality is a proposal to describe the relation between application, data, and the users. (Wang et al., 1999).

After a lot of development in this area, phenomenal success has been registered by the entry of world's best IT organizations like Google, Oracle, Yahoo, Microsoft etc. A lot of online services have been made available by these organizations like Google Maps, Google Earth, Google Mobile Maps, Yahoo Maps, Windows Live Local (MSN Virtual Earth) and mapquest.com (2007) etc., present study will compare its features with many of these services and will elaborate the comparative advantage of proposed system.

4. Problem Definition

Every man/woman in this world need some guidance or advice from its friends/relatives/known people about some purchase/visit/traveling, related expenditure, prices and path to be followed to reach the destination, best outlets/places etc. One tries its best to explore such information from its means. Reliability is always a matter of consideration in this regard; still one manages either to visit one's house or picks up telephone/mobile phone to enquire the required information. Now question arises, whether these services can't be automated? Answer is why not and up to certain extent such services have been automated, which may furnish information regarding targeted products/services/destinations e.g. on-line help-lines, Internet etc. On-line help is available to

facilitate anyone information regarding intended products/services. Search engines may help anyone to retrieve information from hypertext pages scattered all around the world. One can collect information in traces from such sources and can join them to get some knowledge from them. Again reliability is on stake. e. g. health help lines in any city are provided to furnish information regarding health services available in the city. One can enquire regarding these services available in the city, but many things aspects which one will not be in a position to clarify *i.e.* nearest service center, quality of service, comparable pricing, way to reach the destination, contact numbers *etc.* So problem can be stated that every person in this world seeks some guidance or advice for day-to-day purchases/servicing/traveling and some service, which can furnish such service on demand, can be an interesting solution.

5. Proposed Solution

An automated solution to such day-to-day problem can be formed as follows:

Consider a scenario, when a common man picks up its mobile device (cell phone or laptop *etc.*) and types a short message and sends it to a knowledge service center number and in return gets a bundle of information within moments through responding messages.

Now we'll see how this can be made possible through the merger of mobile services and data mining. In previous sections we have explained the concept of data mining and knowledge discovery in databases. Now we'll present a system to find the solution of above discussed problem. In last we'll look for into issues involved in implementation of this solution. In last the conclusion of the paper will be presented.

4

5 6. System Architecture

We present a system that can help in finding solution to above discussed problem.

Higher level view of this system is just like any other Client/Server architecture connected through Internet or a Mobile Network.

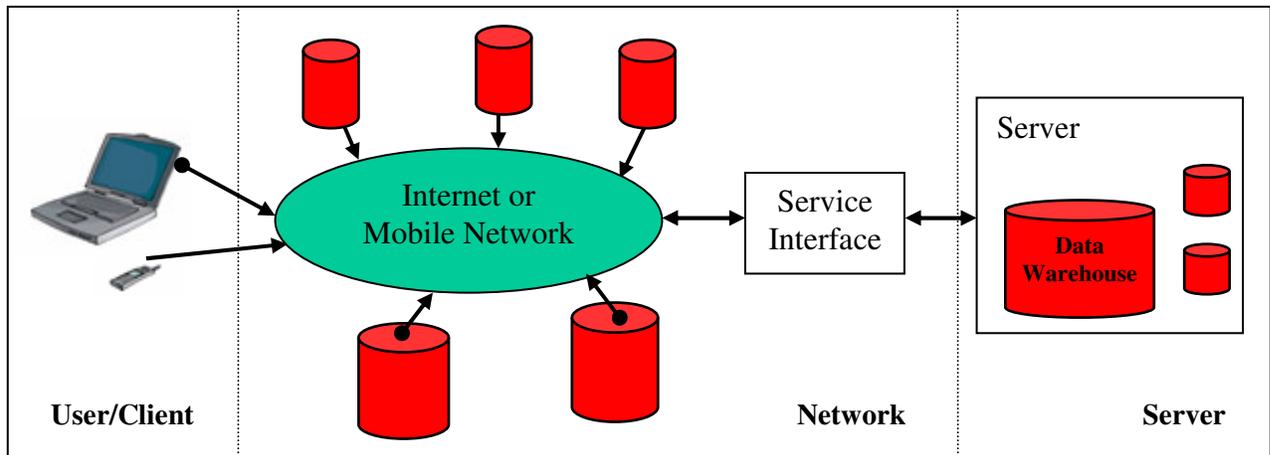


FIGURE 2: Client Server Architecture of the system

Client sends a message to Server through Internet or its mobile network. Network transfers this message further to server through a service interface. Service interface is connected to server through three interfaces. Depending upon the content of the message and forwards it any of the three interfaces of the server. If content has been received from a mobile phone or laptop/computer and its real time position/location is known, then message is forwarded to Client/Target Location Resolution Module, else if a mobile has sent a message without any information about its location then message is forwarded to Mobile Location Tracking Module, so that mobile's real time location can be identified else if a laptop/computer has sent a message without any information about its location then message is forwarded to Geographic IP Locator

Module , so that mobile’s real time location can be identified. If message is forwarded to either Mobile Location Tracking Module or Geographic IP Locator Module, then after finding sender’s current location message is forwarded to Client/Target Location Resolution Module to find the client’s current and targeted location from spatial data warehouse. Incoming message may be in the form of an SMS (Short Message Service) from mobile or in the form of a query if obtained from a computer/laptop, so there may be need to convert the message into a query, where Query Conversion and Optimization module will help the system to fill the gap between actual message and internal representation of query. After converting message into suitable query Client/Target Location Resolution Module takes help from other modules to resolve the query. Processing of this query will use following algorithm.

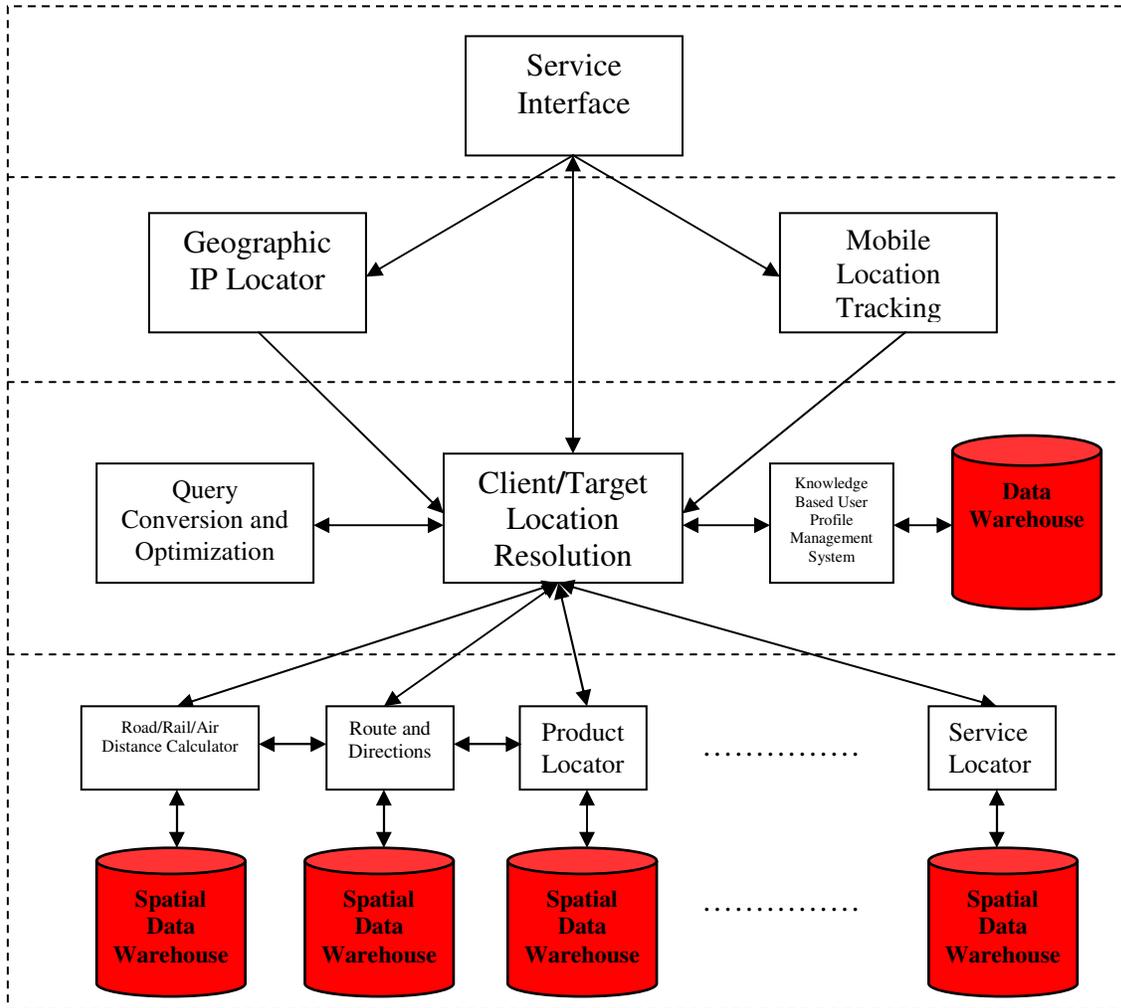


FIGURE 3: Server’s Architecture

We shall discuss this algorithm in three aspects i. e. input, processing and output

6.1 Input

Input for this algorithm will be a message typed as a query in somewhat following formats or some similar representation that may need to be converted into SQL query.

product television, mcl india.punjab.patiala.tu, size ‘300 ltrs’; (1)

or

service travel, mtl india.punjab.patiala.bus_terminal; (2)

or

service food, quality best, mcl india.punjab.patiala.bus_terminal; (3)

[mcl/mtl stands for my (user's) current/target location]

This format includes target product or service, user's current location as well as extra information, which may be furnished on user's discretion, this extra information may involve many parameters, which may require to be standardized or may be defined as per algorithms implementation. Processing of Input from user will decide the response, so right kind of query increases more suitable response. After typing such message user sends it message to a knowledge service center, which initiates a process in response. Target of the process will be to return at least one success and at most a threshold number of successful responses.

6.2 Processing

Different modules will be initiated in response to the input. We discuss major aspects of resolving the problems related to search required information in databases.

6.2.1 User's Target Location Resolution

With the help of Mobile Location Tracking Module or Geographic IP Locator Module, we can identify client's current location. Many a times client's target location will be included in message itself, if we see at query number (2), then we find that client's targeted location is specified and finding such a location in spatial database is quite easy if this location is already included in database. Problem becomes complicated if client's target location is to be determined by the system itself e. g. in queries (1) and (3), for which a search process is initiated to find nearest neighbor that fulfills the demanded product path. Following modules within Client/Target Location Resolution Module will be initiated.

Module 1 First of all responding system searches for the user's current location and targets the place to search the solution. Whole spatial database is divided into four levels i. e. place level (Lowest Level i.e. to find target location at adjacent places), city level, state level and country level (Highest Level i. e. to find target location at adjacent countries). The user will define the level at which solution is required, if user enquire at some particular place level or city level or state level or national level, the search will also be implemented at same level, e.g. current search is implemented at place level within a city, search space will be adjacency of different places, if enquiry is at city level, search space will be at city level, algorithm will search in adjacent cities and so on up to national level.

Location will be targeted hierarchically as if a Domain Name System works as follows:

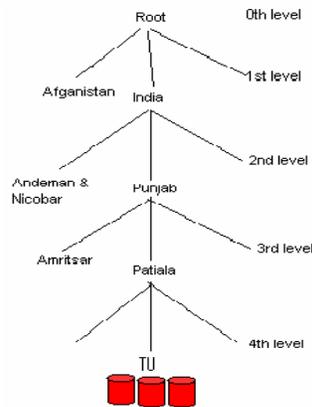


FIGURE 4: User's Current Location Identification.

User's requested search domain will search the databases at fourth level and will try to find the solution within the place first of all. If search finds related records from the databases of same place e.g. TU related records have information regarding the query, then system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the in user's desired format, otherwise search will involve next module.

Module 2 In this module a system search for the enquiry related records of the adjacent places and continues the search until all the adjacent places aren't searched for the solution of the query. Then system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and this module will continue search until all the adjacent places are not searched. If system is unable to retrieve the required knowledge, then it will involve the next module otherwise output will be furnished to the user in desired format.

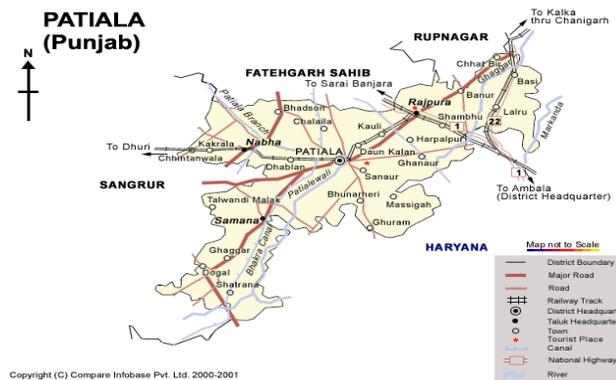


FIGURE 5: Search within User's Current City i.e adjacent places.

Module 3 As this module is being involved, when the search has failed within the city, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 3rd level and starts finding solution of enquiry at state level by retrieving the records of adjacent cities and searches for the related records in the databases of these cities. This search continues until the success is not achieved by finding the records of related queries in adjacent cities and continues its expansion until databases of all the cities of the state aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will involve the next module.



FIGURE 6: Search for databases in adjacent cities within user's current state.

Module 4 As this module is being involved, when the search has failed within the state, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 2nd level and starts finding solution of enquiry at national level by retrieving the records of adjacent states and searches for the related records in the databases of these states. This search continues until the success is not achieved by finding the records of related queries in adjacent states and continues its expansion until databases of all the states of the country aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will involve the next module.



Figure 7: Search for databases at National Level.

Module 5 As this module is being involved, when the search has failed within the country, now search expands its search domain one step higher in search hierarchy i.e. it expands its search to 1st level and starts finding solution of enquiry at international level by retrieving the records of adjacent countries and searches for the related records in the databases of these countries. This search continues until the success is not achieved by finding the records of related queries in adjacent countries and continues its expansion until databases of all the countries of the world aren't exhausted. If the related records are found, the system will retrieve the records and then it will process the information as per the intelligence of the system and requirements of the query and output will be furnished to the user in desired format, otherwise system will report that the related information isn't available worldwide.



FIGURE 8: Search for databases at International Level.

Above discussed modules are to be implemented recursively, if we view the last three modules, one will observe that all three are doing the same thing then what is the necessity of having three different modules, answer lies in the sophistication involved in different levels, so these modules are required to be maintained separately or recursively in such a way that these can handle the complexity involved.

After fetching the records related to queries, task is to convert this information into useful knowledge for the user. Till now information collected is regarding product/service and the path to be traversed to reach the destinations. Now system will add its intelligence through its learned knowledge about the usefulness of the information, it will depend upon the kind of user of service is to be served. In this regard categories will be registered users and occasional users. System will serve both kind of users, but the implementation will differ as registered users will be furnished information on the basis of its profile and habit through continuously learning from the behaviour and choices of the user, whereas occasional users will be served that information, which the system will optimize to be important for the user. In this way it will optimize the information to be furnished for the user. Then system will calculate the distances total costs etc. to furnish the user important information.

6.2.2 Calculation of Distance between Client's Current and Targeted location

After identifying Client's Current and Targeted locations it is obvious to calculate the distances between these locations. Usually any Global Positioning System is used to calculate Air distances of locations, it is due to the reason that Air distances can easily be calculated by finding the distance between pixels of these locations and multiplying it by scale used for differentiating two neighboring pixels, but calculating road and rail distances is very complex task. For a system to be practical for human beings, Air distances are not sufficient, Rail and Road distances need to be calculated to ease the users of the system. We have divided calculating distances into three parts i. e. Air, Rail and Road. Air distance calculating modules calculates Air as well as Water distances, whereas Rail and Road distance calculating module calculate their respective distances individually or as a combination of both. Calculating individual distances for Rail and Road distances is obvious, whereas combination of both is required when user is presently at a place away from railway station and it has to travel certain path through road. So client's total distance from its current location to targeted location will be a combination road and rail travel. Euclidean spaces are the most important metric for calculating distances between locations. We can define Euclidean space for arbitrary dimension n , but usually $n=2, 3$ is common in spatial information system as world can be defined in two or three dimensions. Thus, n -dimensional Euclidean space is defined as (R^n, Δ) , where

$$R^n = \{(x_1, x_2, \dots, x_n) : x_i \in R, \forall 1 \leq i \leq n\} \text{ and } \Delta(X, Y) = \left\{ \sum_{i=1}^n (x_i - y_i)^2 \right\}^{\frac{1}{2}}, \forall X, Y \in R^n.$$

Most of the systems use Euclidean distance for calculating distances between two coordinates that is their Air distance. Distance between two pixels can be calculated quite easily, but situation becomes more complex when distances are not straight forward e. g. for Rail and Road.

Database can be maintained to map and to calculate the distances on Roads and Rail tracks or are made available by the authorities of respective countries, so that distances between two points on a road can be calculated. But problem appears with client's location or target location where no road or rail track is available. For example in Figure 9 a person wants to move from location 'A' to location 'B'. Air distances in both situation is lesser than the required path to be travelled by that person. In situation (a) person has to travel a distance that is a sum of distance from point 'A' to nearest road and through that road up to point 'B', whereas in situation (b) person has to travel distance from 'A' to rail track, through rail track up to a point near to 'B' on rail track and from that point through road up to point 'B'. Distance calculating module of proposed system calculates all these distances through Euclidean distance formula for different sub intervals of total that is from point 'A' to nearest road, then through road up to 'B' etc.

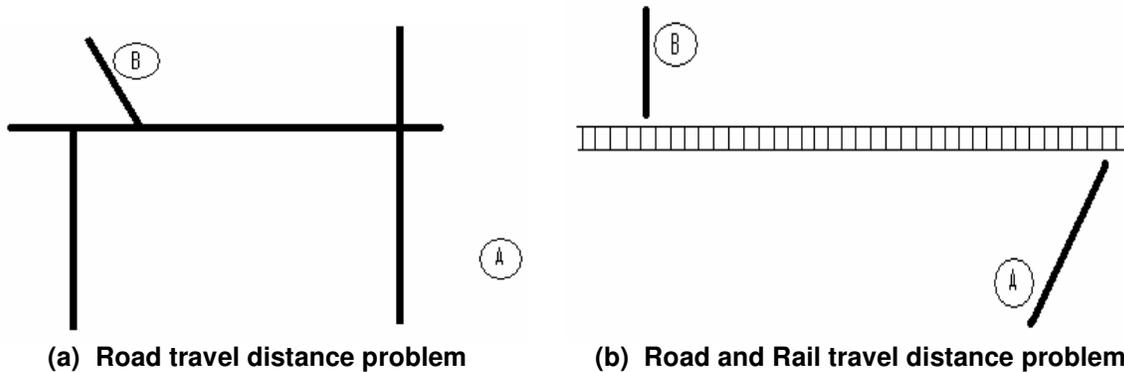


FIGURE 9: Two situations where a person wants to move from A to B.

6.2.3 Databases required for acquiring information

After finding Client's Current/Target locations and calculating distances between them, question arises about the kind of products and services that can be provided to a user and responses provided to users about their queries need to be customized according to their preferences. For example if a customer is looking for a hotel to have meal, then responses to such queries must be based upon knowledge extracted from the past experiences of system with users, distance cannot be the only criteria for finding a solution and system must be equipped with some knowledge based implementation that can handle such aspects. We have equipped our system with a knowledge based implementation for registered customers. Registered customers provide their profile and behavior of registered people is monitored by the system, so that suitable responses can be provided facilitated to them. System includes different spatial databases for different kind of products and services as well as databases containing registered users' profiles.

If we view whole process broadly, then two types of major subsystems are combined to furnish the important information to the user; first one is enhanced Global Positioning System, which will help in targeting users' locations as well as location targeted by users and calculating distances between these locations, whereas second one is Knowledge Discovery System which will help in managing profiles of users, retrieving and facilitating information to users by adding its intelligence to process the information.

6.3 Output of the Queries

Output of the queries discussed in section 6.1 will be somewhat in following format, it contains information in map as well as textual format and this format has potential to be changed as per the enhancements in state-of-the-art technologies. Directions are provided visually as well as in textual format.

| | |
|---|---|
|  | Product Television Available at 22 No Railway crossing Route TU → 22 No Railway Crossing Price Wide Range 10K-25K Distance 1.5 Kms Contact No. 91-175-2390000,0175-2390001 Links www.abc.com |
|  | Travel 'Bus Terminal' Current Location 'Fountain Square' Route Fountain Square → SBP Square → Railway Station → Bus Terminal Distance 3.25 Kms Enquiry Number 91-175-2390000 |
|  | Service Food Quality Best Target Location XYZ Restaurant, Fountain Square Route Bus Terminal → SBP Square → Fountain Square Distance 3 Kms Contact No. 91-175-2390002,0175-2390004 Links www.xyz.com |

FIGURE 10: Output of queries on mobile devices.

In this way information can be furnished to the user on its mobile device with proper directions and information.

7. Issues in Implementing Knowledge Services to user through Mobile Devices

There are certain issues involved in commercializing proposed system.

- 7.1 Most important issue is regarding availability of such databases, their access, their connectivity, continuous updating and their schema. There is requirement to make such databases available on-line in some standard way, so that information can be exchanged and produced on demand. Governments maintain such databases; telephone exchanges, city help-lines, yellow pages services *etc.* can be of great use. But standardization will be a big trouble in this regard, which can be facilitated by implementing XML standards.
- 7.2 Whole world is required to be divided into identifiable places as places form cities, cities form states, states form countries and countries form the world. If required one more level can be incorporated by dividing cities into sub-cities and sub-cities into places. It will be very cumbersome task to implement such a thing and name as well as overlapping areas' conflict is also involved.
- 7.3 Another issue is regarding initiation of search, when system starts searching in other city, state or country, from which place, city or state respectively it should start searching. One possible solution may be to start the search from centralized place e.g. bus stand or railway stations for cities and capitals for states and countries *etc.*

- 7.4 Development of efficient algorithms is also an issue of concern.
- 7.5 Another issue is regarding knowledge center regulating authorities, commercial authorities can't be trusted to furnish information impartially.
- 7.6 Security issue and prompt delivery of service is pre-requisite in any on-line service.
- 7.7 Another issue is regarding Input/Output format standardization. Lack of standardization will not make this concept popular and individually implemented protocols will create much bigger problems.
- 7.8 Another issue is regarding privacy of the users; it may be the case information made available to such a service can be misused easily, so privacy of user is also an issue of concern.

6 8. Implementation

Above discussed algorithm is implemented in prototype of the system using J2EE 5.0 and Java 2 Platform Micro Edition (J2ME) is implemented and a database (using different packages like Oracle 10g, SQL Server etc.) is made available for implementing the above discussed system. The results have been obtained from the queries being invited by the campus students and the results are tested in highly simulated environment by observing different user selecting the outcomes. Maximum number of responses to be provided by system to user for each query was fixed at five. On the basis of such queries different results were being collected and it was being reviewed pilot testing and different responders are asked to value different results. On the basis of different queries and testing the response on the basis of some objective, descriptive criteria and satisfaction from responses of different queries have been given the percentage value to response and following experiments have been conducted within the campus.

7

8 9. Comparative and Experimental Results

Around fifty queries have been fed into the system to test the learning of systems on repetition of similar queries. System responds to the queries is on the basis of least distance, but responses are analyzed on the basis of user's satisfaction from responses as different users may prefer different responses and their review on the output is being given weightage and their choices are fed into system as per their behavior on responses, which is being noted by the system automatically. This prototype of the system is repeated for the processing of queries and results are observed, which are described here. First we present a comparative study of proposed system and other systems available yet, then we analyze the learning aspect of system.

9.1 Comparison of Proposed System with other Systems

If only mobile devices based such systems are considered then Google Mobile Maps is available and considering internet based such systems we can find systems that there are many systems like Google Maps, Google Earth, Yahoo Maps, Windows Live Local (MSN Virtual Earth) and Mapquest.com etc. are widely being made available. First biggest difference between proposed system and these systems is that no other system provides a client oriented service, only general information is provided, whereas proposed system is designed to furnish customized services to clients. Second difference is with respect to calculation of distances, as only Air distances are calculated by other systems, whereas proposed system will calculate actual distance. Google has started calculating real distances, but combination of different transportation modes, their distances, charges etc. are defined by proposed system, which is unavailable in any system present worldwide. Third foremost difference is the kind of services to be provided to users, Google earth can be considered as capable enough to be compared, as it is providing information related to many products and services, but problem with these services is that only limited number of options are made available, whereas proposed system is designed to involve exhaustive list of products and services as well as large number of options is planned to be included so that every human being can be benefitted through this. Fourth major difference is facilitating directions to users to reach up to destination, MSN Virtual Earth is providing this

feature at a small level, but proposed system is designed to furnish directions extensively so that user may be guided appropriately. Proposed system will go beyond these capabilities and will provide a lot more information about target product/service like telephone/mobile number, website link, travelling/communication modes available etc. Complexity of proposed system is also very high as is for most of other systems. Above comparison is included in Table 1. This table is prepared through general observations and expert opinion about different systems. Table 1 provides an overview of comparative advantage of proposed system over other systems. Analysis of table 1 indicates that Google is close competitor of proposed system, still proposed is comparatively much ahead of others.

| System \ Dimensions of Comparison | Google Maps | Google Earth | Google Mobile Maps | Map Quest | Windows Live Local (MSN Virtual Earth) | Yahoo Maps | Proposed System |
|--|-------------|--------------|--------------------|-----------|--|------------|-----------------|
| Real Distances | Moderate | High | Moderate | Low | Moderate | Low | High |
| Product Availability | Moderate | High | Low | Low | Low | Moderate | High |
| Customized Service | Moderate | Moderate | Moderate | Low | Low | Moderate | High |
| Combination of Multiple Transportation Modes | Moderate | Moderate | Low | Low | Low | Low | High |
| Directions | Low | Moderate | Moderate | Low | Moderate | Moderate | High |
| Targeted Users | Web | Web | Mobile | Web | Web | Web | Web/Mobile |
| Complexity of System | High | High | High | Moderate | Moderate | Moderate | High |

Table 1: Comparison of Proposed system with other major competitive systems

With above comparison there are certain unique features with proposed systems as well as other system. For example Google has started providing real time traffic information, whereas in proposed system can find real time location of web user, which is unavailable to any other system yet.

9 9.2 Performance

System prototype is tested for upper threshold of responses to be three and five responses and the results are analyzed in the light of many parameters involving user's preferences and others. These parameters on the basis of different weights assigned to them are evaluated and following graphs are being prepared.

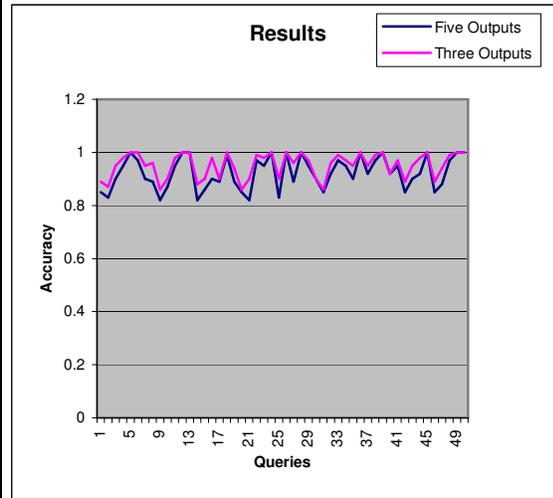


FIGURE 11: Accuracy of 50 Queries with Accuracy

FIGURE 12: Repetition of processing and Results

On analyzing the results of the queries and their accuracy in figures 11 and 12, we found that accuracy of results varies from 70% to 100% for five results for a single query and it varies from 82% to 100% for three outputs. As experiments were being conducted on a sample size of 50 queries one after the other and the number of responses were being kept at five and three. Decreasing the number of responses improved the results significantly. Also the learning process helped the system to improve the results significantly, when the same queries were being implemented on repetition system responded with an accuracy varying 82% to 100% for five outputs and an accuracy of 86% to 100% being observed. More experiments are undergoing to test the system with large number of queries and more repetition and more and more accuracy is expected to achieve in the range of 95% to 100%. With these results we are hopeful for the system to respond accurately in beta testing of prototype of system.

10

11 10. Work in Progress

One of the most important challenges before implementing this system is the challenges of identifying databases scattered worldwide for the search and their integration as different databases may differ in their configurations. So this prototype is undergoing the process of data integration by connecting different database standards available in the market. There is need to achieve an accuracy of 98% to 100%, so there is a need for strengthening the learning process. More learning processes are under development. Here decision tree is being implemented for searching the outputs. Other techniques (or hybrids) are also under consideration. For implementation of such a system public awareness is highly required, so that real implementation can catch up and the utility of system can be understood well by users. There is a need for marketing the system in this regard. This system will be implemented on several multiprocessing machines and server to server transactions will be experimented on them.

11. Future Expectations and Conclusion

We are very hopeful for the accomplishment of this system at Authors' affiliated place and future researches as well as standardization for its implementation and its acceptance worldwide. Such system may contribute for the up-liftment of society and also to bring science and technology to masses at large, So that a common person can be benefited through it.

12 12.Bibliography

1. Abel, D.J. and Ooi, B.C.(1993). An Introduction to Spatial Database Systems. Proceedings of the Third International Symposium on Large Spatial Databases, Singapore.
2. Aggarwal Rakesh, Ghosh Sakti, ImielinskiTomasz, Iyer Bala and Swami Arun (1992). An interval classifier for database mining applications. VLDB.
3. Agrawal Rakesh, Imielinski Tomasz and Swami Arun N. (1993). Mining Association Rules between Sets of Items in Large Databases. SIGMOD Conference: 207-216.
4. Anselin, L. (2000). Computing environments for spatial data analysis. Journal of Geographical Systems, 2(3):201–220.
5. Buchmann A. and Giinther O. (1990). Research issues in spatial databases. ACMSIGMOD Record, 19:61-68.
6. Buchmann A., Gfinther O., Smith T.R., and Wang Y.E. (1989), An Introduction to Spatial Database Systems. Proceedings of the First International Symposium on Large Spatial Databases, Santa Barbara,.
7. Cook, D., Majure, J., Symanzik, J., and Cressie, N. (1996). Dynamic graphics in a GIS: A platform for analyzing and exploring multivariate spatial data. Computational Statistics, 11:467–480.
8. Delone, W. H. and E. R. McLean (1992). "Information systems success: the quest for the dependent variable." Information Systems Research 3(1): 60-95.
9. Dykes, J. A. (1997). Exploring spatial data representation with dynamic graphics. Computers and Geosciences, 23:345–370.
10. ESRI (2004). An Overview of the Spatial Statistics Toolbox. ArcGIS 9.0 Online Help System (ArcGIS 9.0 Desktop, Release 9.0, June 2004). Environmental Systems Research Institute, Redlands, CA.
11. Fischer, M. and Nijkamp, P. (1993). Geographic Information Systems, Spatial Modelling and Policy Evaluation. Springer-Verlag, Berlin.
12. Fischer, M. M. and Getis, A. (1997). Recent Development in Spatial Analysis. Springer-Verlag, Berlin.
13. Fischer, M. M., Scholten, H. J., and Unwin, D. (1996). Spatial Analytical Perspectives on GIS. Taylor and Francis, London.
14. Fotheringham, A. S. and Rogerson, P. (1993). GIS and spatial analytical problems. International Journal of Geographical Information Systems, 7:3–19.
15. Fotheringham, A. S. and Rogerson, P. (1994). Spatial Analysis and GIS. Taylor and Francis, London.
16. Frank A. (1991). Properties of geographic data: Requirements for spatial access methods. Proceedings of the Second International Symposium on Large Spatial Databases, Zurich.
17. Fröhlich, M. and J. Mühlig (2002). Usability in der Konzeption. Usability Nutzerfreundliches Web-Design. Berlin.
18. Giinther, O. and Schek, H.-J. (1991). An Introduction to Spatial Database Systems. Proceedings of the Second International Symposium on Large Spatial Databases, Zurich.

19. Goodchild, M. F., Haining, R. P., Wise, S., and others (1992). Integrating GIS and spatial analysis — problems and possibilities. *International Journal of Geographical Information Systems*, 6:407–423.
20. Google Earth - Home(2007): <http://earth.google.com>
21. Google Maps(2007). <http://maps.google.com>
22. Google Mobile Maps(2007). <http://www.google.com/gmm/index.html>
23. Haining, R. (1989). Geography and spatial statistics: Current positions, future developments. In Macmillan, B., editor, *Remodelling Geography*, pages 191–203. Basil Blackwell, Oxford.
24. Han J., Fu Y., and Ng R. (1994). Cooperative Query Answering Using Multiple Layered Databases, Proc. 2nd Int'l Conf. on Cooperative Information Systems (CoopIS'94), Toronto, Canada, May, pp. 47-58.
25. Han J., Fu Y., Koperski K., Melli G., Wang W., Zaïane O. R. (1995). Knowledge Mining in Databases: An Integration of Machine Learning Methodologies with Database Technologies, *Canadian Artificial Intelligence*, October.
26. Han J., Nishio S., Kawano H., and Wang W. (1998). Generalization-Based Data Mining in Object-Oriented Databases Using an Object-Cube Model. *Data and Knowledge Engineering* , 25(1-2):55-97.
27. Han Jiawei, Cai Yandong and Cercone Nick (1991). Concept-Based Data Classification in Relational Databases. Workshop Notes of 1991 AAAI Workshop on Knowledge Discovery in Databases (KDD'91), Anaheim, CA, July, pp. 77-94.
28. Haslett, J., Wills, G., and Unwin, A. (1990). SPIDER — an interactive statistical tool for the analysis of spatially distributed data. *International Journal of Geographic Information Systems*, 4:285–296.
29. Holsheimer Marcel and Siebes Arno (1994). Data Mining: The Search for Knowledge in Databases. Technical Report CS-R9406, CWI, Amsterdam.
30. Hynek, T. (2002). User experience research- treibende Kraft der Designstrategie. Usability Nutzerfreundliches Webdesign. Berlin, Springer Verlag: 43-59.
31. Jorge B. Bocca (1991). Logic Programming Environments for Large Knowledge Bases: A Practical Perspective . *VLDB*: 563.
32. Kawano H., Nishio S., Han J., and Hasegawa T. (1994). How Does Knowledge Discovery Cooperate with Active Database Techniques in Controlling Dynamic Environment?, in Proc. 5th Int'l Conf. on Database and Expert Systems Applications (DEXA'94), Athens, Greece, September, pp. 370-379.
33. Kohavi Ron and Sommerfield Dan (1996). Data Mining using MLC++ A Machine Learning Library in C++. *Tools With AI*, pages 234–245.
34. Levine, N. (2004). The CrimeStat program: Characteristics, use and audience. *Geographical Analysis*. Forthcoming.
35. Li W., Han J., and Pei J. (2001). CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules, Proc. 2001 Int. Conf. on Data Mining (ICDM'01), San Jose, CA, Nov..
36. Mannila Heikki, Toivonen Hannu and Verkamo A. (1994). Inkeri: Efficient Algorithms for Discovering Association Rules. *KDD Workshop*: 181-192.
37. Mapquest.com (2007): <http://city.mapquest.com>

38. Mayhew, D. J. (2002). *The Usability Engineering Lifecycle - A practitioner's handbook for user interface design*. San Francisco, California, Morgan Kaufmann Publishers Inc.
39. Micheline Kamber, Winstone Lara, Gon Wang and Han Jiawei (1997). *Generalization and Decision Tree Induction: Efficient Classification in Data Mining*. RIDE.
40. Monmonier, M. (1989). Geographic brushing: Enhancing exploratory analysis of the scatterplot matrix. *Geographical Analysis*, 21:81–4.
41. MSN Virtual Earth(2007). <http://maps.live.com>
42. Ng R. and Han J. (1994). Efficient and Effective Clustering Method for Spatial Data Mining, Proc. of 1994 Int'l Conf. on Very Large Data Bases (VLDB'94), Santiago, Chile, September, pp. 144-155.
43. Nielsen Jakob, N. N. G. (1993). *Usability Engineering*. Cambridge, Academic press limited.
44. Piatetsky-Shapiro Gregory(1989). *Knowledge Discovery in Real Databases*. IJCAI-89 Workshop.
45. Redman, T. (1996). *Data Quality for the Information Age*. Boston, London, Artech House.
46. Rey, S. J. and Janikas, M. V. (2004). STARS: Space-time analysis of regional systems. *Geographical Analysis*. forthcoming.
47. Symanzik, J., Cook, D., Lewin-Koh, N., Majure, J. J., and Megretskaja, I. (2000). Linking ArcView and XGobi: Insight behind the front end. *Journal of Computational and Graphical Statistics*, 9(3):470–490.
48. Takatsuka, M. and Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers and Geosciences*, 28:1131–1141.
49. Tayi, G. K. and Ballou D. p. (1998). Examining data quality. *Communications of the ACM*. 41 (2): 54-57.
50. Unwin, A. (1994). REGARDing geographic data. *Computational Statistics*, pages 345–354. Physica Verlag, Heidelberg.
51. Wang, R. Y., Strong D. M., (1999). An information quality assessment methodology. *Proceedings of the International Conference on Information Quality*, Cambridge MA.
52. Wise, S., Haining, R., and Ma, J. (2001). Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. *International Journal of Geographic Information Science*, 15(3):239–254.
53. Yahoo Maps, Driving Directions, and Traffic (2007). <http://maps.yahoo.com>
54. Yoo J. S. and Shekhar S. (2006). A join less approach for mining spatial collocation patterns. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 10, October.

Testing of Contextual Role-Based Access Control Model (C-RBAC)

Muhammad Nabeel Tahir
Multimedia University, Melaka
75450, Malaysia

m_nabeeltahir@hotmail.com

Abstract

In order to evaluate the feasibility of the proposed C-RBAC model [1], the work in this paper presents the prototype implementation of C-RBAC model. We use eXtensible Access Control Markup Language (XACML) as a data repository and to represent the extended RBAC entities including purpose and spatial model.

Key words: C-RBAC Testing, XACML and C-RBAC, Policy Specification Languages

1 INTRODUCTION

1.1 EXtensible Access Control Markup Language (XACML)

The OASIS eXtensible Access Control Markup Language (XACML) is a powerful and flexible language for expressing access control policies used to describe both, policy and access control decision request / response [2]. XACML is a declarative access control policy language implemented in XML and a processing model, describing how to interpret the policies. It is a replacement for IBM's XML access control language (XACL) which is no longer in development. XACML is a language primarily aimed at expressing privacy policies in a form such that computer systems can enforce them. The XACML has been widely deployed and there are several implementations of XACML in various programming languages available [3]. The XACML is designed to support both centralized and decentralized policy management.

1.2 Comparison Between EPAL, XACML and P3P

Anderson [3] suggested that a standard structured language for supporting expression and enforcement of privacy rules must meet the following requirements:

Rq1. The language must support constraints on who is allowed to perform which action on which resource;

Rq2. The language must support constraints on the purposes for which data is collected or to be used;

Rq3. The language must be able to express directly-enforceable policies;

Rq4. The language must be platform-independent; and

Rq5. The language used for privacy policies must be the same as or integrated with the language used for access control policies.

Keeping in mind the above requirements, the comparison of P3P, EPAL, and XACML are summarized in Table 1 in which “√” means the language can satisfy the requirement, “×” means the language cannot satisfy the requirement and “?” means it is an unknown feature for the corresponding requirement and may depend on the language extension and implementation.

Table 1: Comparison of P3P, EPAL, and XACML (Anderson, 2005).

| | P3P | EPAL | XACML |
|------------------------------------|-----|------|-------|
| Rq1: Constraints on subject | × | √ | √ |
| Rq2: Constraints on the purposes | √ | √ | √ |
| Rq3: Directly-enforceable policies | × | √ | √ |
| Rq4: Platform-independent | √ | ? | √ |
| Rq5: Access control | × | × | √ |

Although P3P is a W3C recommended privacy policy language that supports purpose requirements and is platform-independent, P3P does not support directly-enforceable policies. P3P policies are not sufficiently fine-grained and expressive to handle the description of privacy policies at the implementation level. P3P mainly focuses on how and for what purpose information is being collected rather than on how and who can access the collected information. Thus, P3P is not a general-purpose access control language for providing technical mechanisms to check a given access request against the stated privacy policy especially in ubiquitous computing environment. EPAL supports directly-enforceable policies but it is a proprietary IBM specification without a standard status. According to a technical report comparing EPAL and XACML by Anderson [3], EPAL does not contain any privacy-specific features that are not readily supported in XACML. EPAL does not allow policies to be nested as each policy is separate with no language-defined mechanism for combining results from multiple policies that may apply to a given request whereas XACML allows policies to be nested. A policy in XACML, including all its sub-policies, is evaluated only if the policy's Target is satisfied. For example, policy “A” may contain two sub-policies “B1” and “B2”. These sub-policies could either be physically included in policy “A” or one or both could be included by a reference to its policy-id, a unique identifier associated with each XACML policy. Thus making XACML more powerful in terms of policy integration and evaluation. EPAL [4] functionality to support hierarchically organized resources is extremely limited whereas XACML core syntax directly supports hierarchical resources [data-categories] that are XML documents. In an EPAL rule, obligations are stated by referencing an obligation that has been defined in the (vocabulary) element associated with the policy; in XACML, obligations are completely defined in the policy containing the rule itself. EPAL lacks significant features that are included in XACML and that are important in many enterprise privacy policy situations. In general, XACML is a functional superset of EPAL as XACML supports all the EPAL decision request functionality. XACML provide a more natural way of defining role hierarchies, permissions, permission-role assignment and it support the idea of complex permissions that are used in the systems implementing role-based access control models for distributed and ubiquitous environments. As a widely accepted standard, it is believed that

XACML is suitable for expressing privacy specific policies in a privacy-sensitive domain as healthcare.

2 CORE C-RBAC IMPLEMENTATION USING XACML

The implementation of core RBAC entities (USERS, ROLES, OBJECTS, OPS, PRMS) in XACML are presented in table 2.

Table 2: Core RBAC Entities in XACML.

| Core RBAC Entities | XACML Implementation |
|--------------------|-------------------------|
| USERS | <Subjects> |
| ROLES | <Subject Attributes> |
| OBJECTS | <Resources> |
| OPS | <Actions> |
| PRMS | <Policyset> <Policy> |

The current XACML specification does not include the work for extended RBAC model but it has the core RBAC profile to implement the standard RBAC model. Therefore, XACML is further investigated and extended to support the proposed privacy access control model C-RBAC and privacy policies. Table 3 shows the proposed XACML extension for the privacy access control model.

Table 3: Extended Entities of C-RBAC Model.

| C-RBAC ENTITIES | XACML/XML IMPLEMENTATION |
|-----------------------------|--------------------------|
| PHYSICAL LOCATION | <PLOC> |
| LOGICAL LOCATION | <LLOC> |
| LOCATION HIERARCHY SCHEMA | <LHS> |
| LOCATION HIERARCHY INSTANCE | <LHI> |
| SPATIAL DOMAIN OVER LHS | <SSDOM> |
| SPATIAL DOMAIN OVER LHI | <ISDOM> |
| PURPOSE | <PURPOSE> |
| SPATIAL PURPOSE | <SP> |
| SPATIAL PURPOSE ROLES | <SPR> |

2.1 Experimental Evaluation

We created different healthcare scenarios to analyze behavior of the proposed C-RBAC entities. By simulating different healthcare scenarios, we calculated response time including the access time (with and without authorization) and response time to derive spatial granularity, spatial purpose and spatial purpose role enabling and activation, have showed that the time required to collect contextual attributes, to generate a request and to authorize an access request have been in milliseconds and seconds that are considered to be tolerable in real time situations.

The use of XML as a tool for authorization raises questions as to expressiveness versus efficiency, particularly in a large enterprise. Ideally, authorization should account for a negligible amount of time per access but it is necessary that all access conditions be expressed and context be checked completely. In this implementation, all authorization policies are loaded into memory, independent of request comparison. Therefore, the time to read policies is not included into access time. Instead, authorization time consists of formal request generation, request parsing, contextual attribute gathering, request-policy comparison and context evaluation, response building, and response parsing. The experiments have been performed on a 2.66 GHz Intel machine with 1 GB of memory. The operating system on the machine is Microsoft Windows XP Professional Edition, and the implementation languages used is Microsoft C-Sharp (C#).

For the experimental evaluation, different healthcare scenarios that are mentioned throughout the thesis (the one presented in chapter 5 and section 7.3) have been executed to analyze the performance and expected output of C-RBAC model (Tahir, 2009a). According to those healthcare scenarios, contextual values including purpose setup, location modeling that include locations, location hierarchy schemas and instances, spatial purposes, spatial purpose roles and privacy policies have been defined in the system with their selectivity to 100 percent i.e. all policies, operations, purposes, locations and spatial purpose roles have been set to allow access for every access request. After creating the necessary objects and relations the response has been analyzed in order to verify that whether the proposed model correctly grant or deny access according to the privacy rules or not. Moreover, the response time has been also calculated at different levels to measure the computational cost for monitoring and evaluating the dynamic contextual values like purpose, location and time.

Figure 1 shows purpose inference algorithm based on the contextual values of the user. It includes time, location, motion direction, distance and user motion direction with measurement unit as meter, centimeters etc.

```
PurposeInference (s, pos1, pos2) {  
  // s ∈ SESSIONS, pos1 and pos2 are user's current position and the position  
  to which user is heading to;  
  
  //Step 1: Getting the subject roles through the active session  
  SPR spr = SessionSPR(s);  
  
  //Step 2: Getting the current time  
  Time t = DateTime.Now;  
  
  //Step 3: Getting ploc in which user is located  
  PLOC ploc1 = Ploc(pos1);  
  PLOC ploc2 = Ploc(pos2);  
  
  //Step 4: Getting motion direction  
  DIRECTION dir = PlocDir(ploc1, ploc2);  
  
  //Step 5: Getting distance measurement unit  
  DUnitPloc(ploc2) → dunit  
  
  //Step 6: Getting distance between the two physical locations  
  Distance dval = DisPloc(ploc1, ploc2)  
  
  //Step 7: Retrieving the corresponding spatial purposes from the spatial  
  purpose global file (refer to figure 7.10)  
  
  Purpose p = Get_Purpose(spr, t, dir, pos1, dval, DUnit)  
  
  Return p;  
}
```

Figure 1: Purpose inference algorithm

Figure 2 shows the response time of purpose inference algorithm. As shown, the response time increases as the number of purpose inference requests increase. This is because of the constant movement of the user over the space defined within the system. For a single request, the system takes approximately 38 milliseconds to compute the purpose from the collected contextual attributes that are necessary input to the purpose inference algorithm.

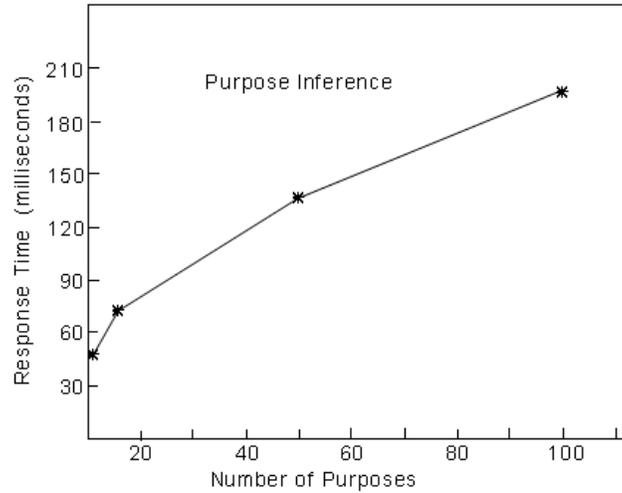


Figure 2: Purpose Inference Response Time.

Figure 3 shows the response time in general for purpose collection based on the user's current contextual attributes. Figure 4 shows the response time for purpose collection at location hierarchy schema and instance level. As shown, the response time increases as the number of logical or physical locations defined in schema or instances increases. It also shows that the response time at schema level is less than that of instance. This is because for each instance, the system collects the spatial purposes defined not only at an instance level but also from its corresponding schema from which it is instantiated (lhi is instance of lhs). Thus, the response time increases as the location granularity becomes finer.

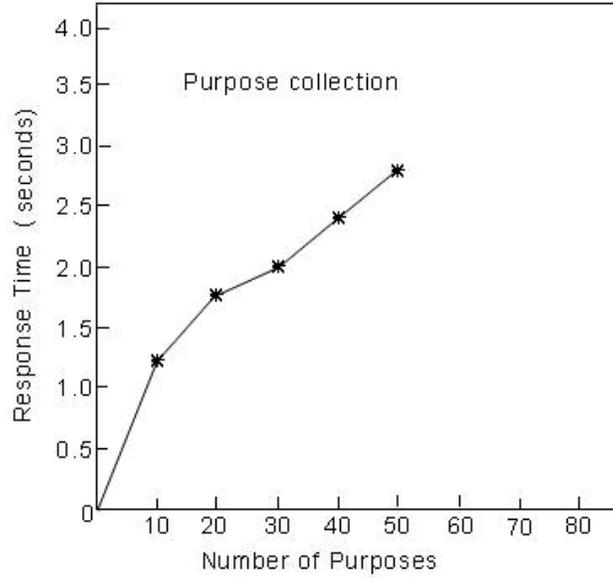


Figure 3: Purpose Collection Response Time in General.

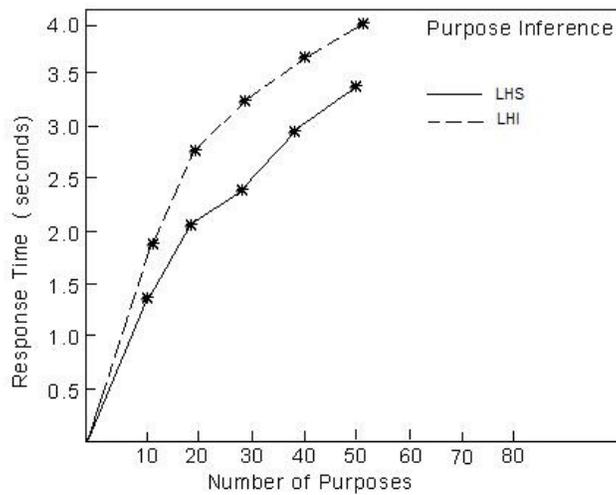


Figure 4: Purpose Collection Response Time at LHS and LHI Level.

Figure 5 shows spatial granularity mapping from LHS to logical locations lloc defined within the schema. It also shows the mapping response time to generate a set of physical locations ploc that are derived from lloc defined within the given LHS. Figure 6 shows the response time to derive physical locations from a given LHI.

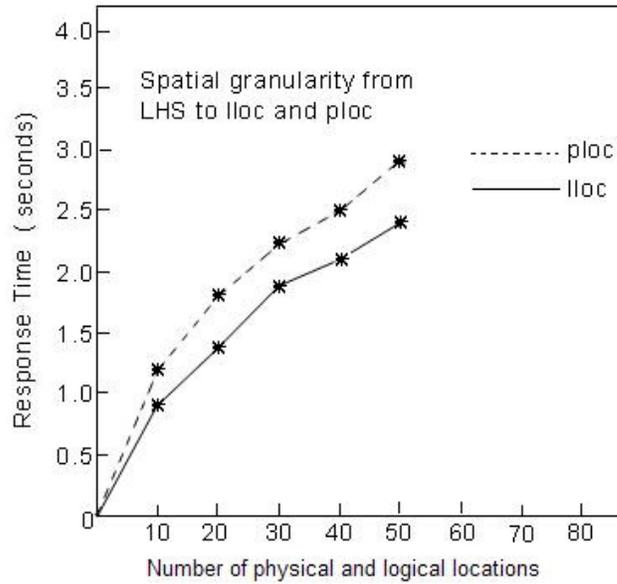


Figure 5: Response Time to Derive Physical and Logical Locations from a Given LHS.

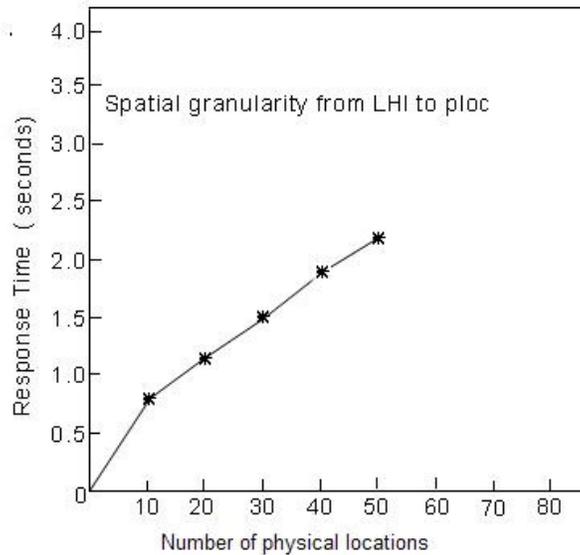


Figure 6: Response Time to Derive Physical Locations from a Given LHI.

Figure 7 shows the response time to activate a spatial purpose through C-RBAC constraints defined within the system. It has been observed that the activation of spatial purposes depends on the spatial granularity. For example the spatial purposes defined at location hierarchy schema level took more time to activate as compared to spatial purpose at physical location level. This is because at physical level, the system directly activate the spatial purpose for the given purpose and physical location whereas in case of location hierarchy schema, the system had to

derive all logical locations and then to its corresponding physical locations first and then activate those corresponding physical locations with the given purpose.

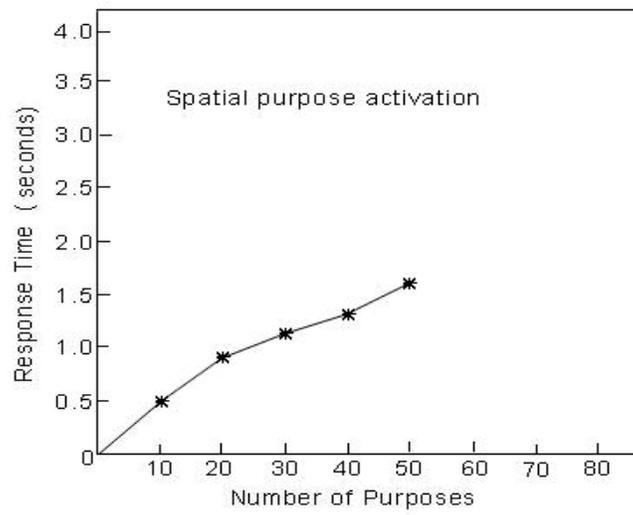


Figure 7: Response Time to Activate Spatial Purposes.

Figure 8 shows the response time to enable spatial purpose roles defined with different spatial granularities and purposes. The results have been analyzed by enabling a single spatial purpose role spr (without spatial purpose role hierarchy) and multiple spr in the presence of hierarchy. It is noticed that the enabling of roles defined without hierarchical relationships is less than to those defined with hierarchical relationships. This is because in case of hierarchical relationships, constraints are applied and evaluated based on the contextual values of the user before the system enable/disable spatial purpose roles defined within the C-RBAC implementation.

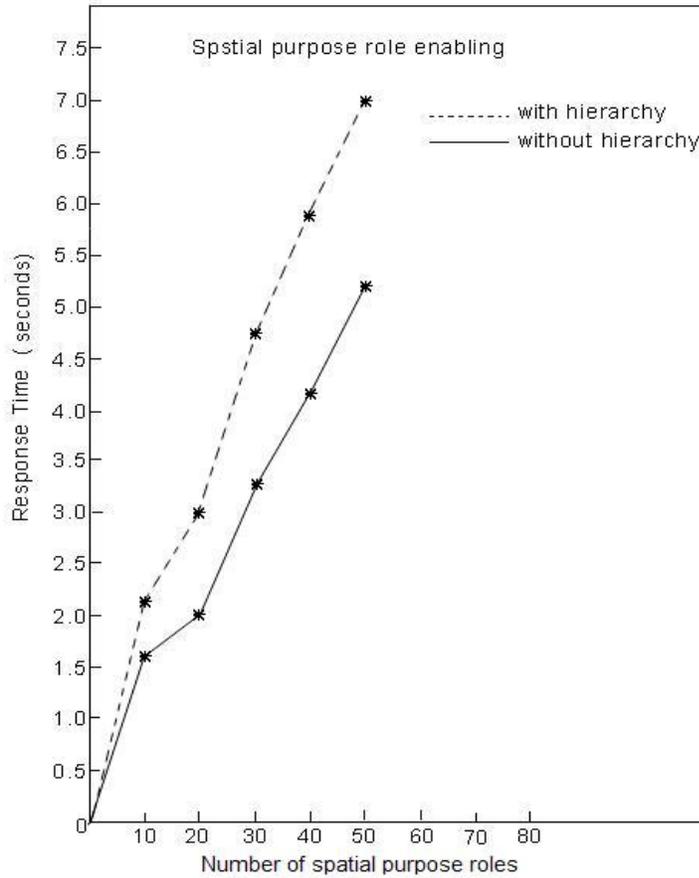


Figure 8: Response Time for Spatial Purpose Roles Enabling (with and without Hierarchical Relationships).

Figure 9 and 10 shows the response time for spatial purpose role activation and mapping of user session onto enabled and active spatial purpose roles respectively.

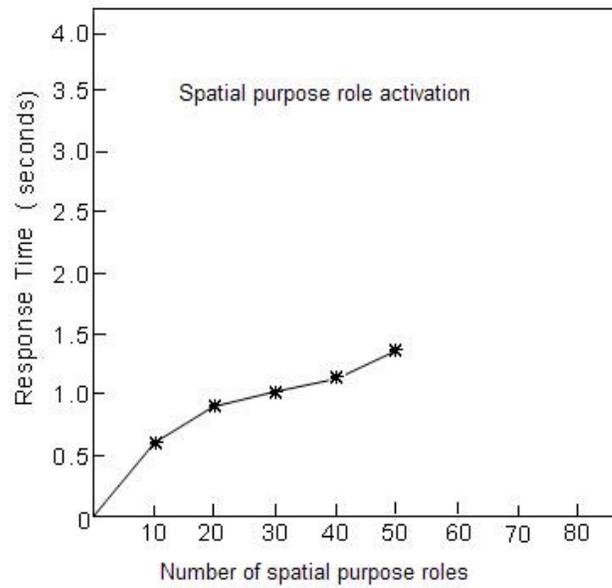


Figure 9: Response Time for Spatial Purpose Roles Activation.

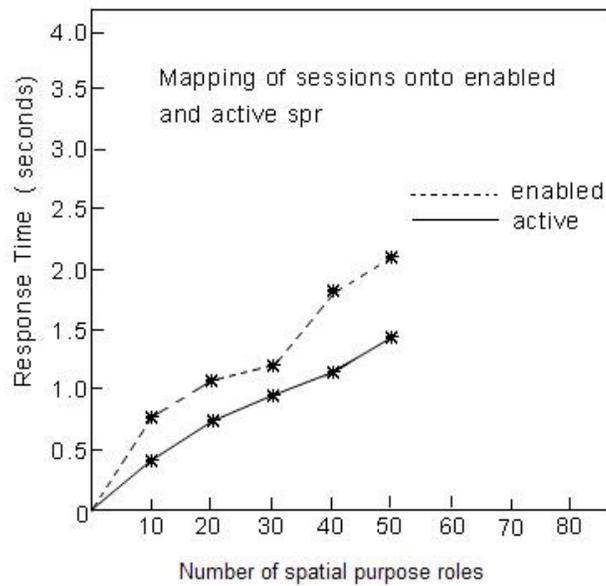


Figure 10: Response Time for Mapping a User Session Onto Enabled and Active Spatial Purpose Roles.

```

while(true){
    //Step 1: Gets the access requests from the subject
    If request(SUBJECTS s, OPS op, OBJECTS o, PURPOSES {p1,p2 ..., pn },
    RECIPIENTS {rp1,rp2 ..., rpn}) {

        //Step 2: Processes the request
        //Step 2.1: Checks the object ownership
        OWNERS owr = object_owner(o)

        //Step 2.2: Checks the subject role
        ROLES r = subject_roles(s)

        //Step 2.3: Retrieves the corresponding privacy rules
        PRIVACY-RULES rule = GetPrivacyRules(r, op, o, {p1,p2 ..., pn},{rp1,rp2 ...,
        rpn})

        //Step 3: Makes a decision by
        DECISIONS d = deny or allow;

        //Step 3.1: Checks permission from the core C-RBAC model
        PRMS prms = assigned_permission(sprloc_type, p)

        //Step 3.2: Checks legitimate purposes
        If(p' ∧ rule.p = {p1,p2 ..., pn}){
            //Step 3.3: Checks legitimate recipients
            If(rule.rp = {rp1,rp2 ..., rpn}){

                //Step 3.4: Checks the location granularity
                If (loc_type ∧ rule.loc_type = {lloc, ploc, lhs, lhi, sdomlhs, sdomlhi}) {

                    //Step 3.5 Checks ssod and dsod constraints
                    If (rloc_type, p) {
                        Apply_SSoDConstraints(rloc_type, p);
                        Apply_DSoDConstraints(rloc_type, p);

                        //Step 3.6 Final decision
                        d = rule.decisions
                        OBLIGATIONS {obl1, obl2 ..., obln} = rule.obligations
                        RETENTIONS rt = rule.retentions
                    } } }

                //Step 4: Returns a response and an acknowledgement
                If(d = allow){
                    //Step 4.1: Returns: allow, Obligations, Retention policy
                    Response(d, {obl1, obl2 ..., obln},rt)
                } Else {
                    //Step 4.1: Returns deny, null, null
                }
            }
        }
    }
}

```

Figure 11: Access Control Decision Algorithm for the Proposed Privacy Based C-RBAC

It is observed that the response time to enable spatial purpose roles is more than that of activation and mapping time. This is because of object/classes based implementation in C# of the proposed C-RBAC model. During the execution of different healthcare scenarios, it is observed that at the time of login, the system has evaluated the contextual values of the user and enabled all the spatial purpose roles assigned by the administrator. From implementation point of view, role enabling means that the system loads all the assigned roles into the memory based on the contextual values and SSoD constraints. Then for each change in the user's context, the system decides whether to activate or deactivate the spatial purpose role by based on the DSoD constraints and new contextual values. Figure 11 shows the access control algorithm to evaluate the user's request and to grant/deny access based on the contextual values, enabled and activated roles.

For authorization, request generation time is approximately 2 seconds. The request parsing time is 1.28 seconds. The average time for the PDP to gather attributes and authorize a formal request is 3.5 seconds. All local transfer times are less than 1 seconds. Therefore, the total time to authorize an access is 6.78 seconds.

The average total time to determine which regular spatial purpose roles a user has assigned is 776 ms. Role assignment is trivially parallelizable because each role can be checked independently, so taking a distributed approach or using multi-threads could reduce this number to a fraction of this original value. If the time is reduced to a tenth of the original, it would take 77 ms to determine a user's roles.

Without authorization, the average time to perform an access is 703 ms. When authorization is added into this system, the total time for an authorized access is 7483 milliseconds ($6.78 * 1000 + 703 = 7483$ milliseconds = 7.5 seconds approximately). The 6.78 seconds access authorization time is 89% of the total system time. This additional time is easily tolerated in a system where tens of milliseconds are not critical. Role assignment can be determined per session or per access. The 77 milliseconds this process took is invisible during the login process. Per access, this 77 milliseconds added to the 6780 milliseconds (7.78 seconds) for authorization would account for 88% of the 7483 milliseconds (7.5 seconds) total access time. This result is still tolerable. Based on the results generated by measuring the response time for spatial granularity derivation, spatial purpose and spatial purpose role enabling and activation, request generation and evaluation and response time, it is concluded that the extensions introduced by C-RBAC are reliable and due to very less overheads, the model can be effectively used for dynamic context-aware access control applications.

3. CONCLUSION

In this paper, we simulated the different healthcare scenarios to analyze the behavior and to calculate the response time of the proposed C-RBAC model. Our findings include the access time (with and without authorization) and response time to derive spatial granularity, spatial purpose and spatial purpose role enabling and activation, have showed that the time required to collect contextual attributes, to generate a request and to authorize an access request have been in milliseconds and seconds that are considered to be tolerable in real time situations. The model implementation and its results also showed that the extensions introduced by C-RBAC have been reliable and due to very less overheads, the model can be effectively used for dynamic context-aware access control applications.

4. REFERENCES

- [1] Tahir, M. N. (2007). Contextual Role-Based Access Control. *Ubiquitous Computing and Communication Journal*, 2(3), 42-50.

- [2] OASIS (2003). A brief introduction to XACML. Retrieved November 14, 2008, from http://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.htm.
- [3] Anderson, A. (2005). A comparison of two privacy policy languages: EPAL and XACML. Sun Microsystems Laboratory Technical Report #TR-2005-147, November 2005. Retrieved November 14, 2008, from <http://research.sun.com/techrep/2005/abstract-147.html>.
- [4] IBM (2003). Enterprise privacy authorization language (EPAL). IBM Research Report June 2003. Retrieved November 14, 2008, from <http://www.zurich.ibm.com/security/enterprise-privacy/epal>.

COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA