International Journal of Computer Science and Security (IJCSS)

ISSN : 1985-1553

VOLUME 4, ISSUE 1

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

Editor in Chief Dr. Haralambos Mouratidis

International Journal of Computer Science and Security (IJCSS)

Book: 2010 Volume 4, Issue 1
Publishing Date: 30-03-2010
Proceedings
ISSN (Online): 1985-1553

IJCSS Journal is a part of CSC Publishers
http://www.cscjournals.org

Published in Malaysia

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

CSC Publishers

Table of Contents

Volume 4, Issue 1, March 2010.

Pages

Cache Coherency in Distributed File System

Anagha Kulkarni anagha.kulkarni@cumminscollege.in
Department of Information Technology,
Cummins College of Engineering for Women,
Pune, 411052, India.

Radhika Bhagwat radhika.bhagwat@cumminscollege.in
Department of Information Technology,
Cummins College of Engineering for Women,
Pune, 411052, India.

Abstract

Principle of locality states that most memory references are made to a small number of memory locations. Not only that, memory locations near most recently referenced locations are more likely to be referenced than one further away. To take advantage of this, cache memory is inserted between memory and CPU [1]. Better utilization of cache is crucial for good performance of distributed file system; even in case of remote file accesses.

Not caching a file during writes prolongs the session, thereby increasing write-sharing time, leading to slow performance especially on WANs. This paper introduces a technique to reduce miss penalty during remote file writes and allows write sharing in LAN. It uses the principle of divide-and-rule and arranges the system into hierarchical domains and then gives ownerships to the writers.

Keywords: Cache Coherency, Distributed file system, Performance, WAN.

1. INTRODUCTION

Caches are used extensively in practice. It is a store of recently used data objects that is closer than the objects themselves. When a client needs an object it first checks its cache for the object. If not found or if it does not have a valid copy, it is fetched from the target computer and it is added or replaced in cache. This is done to increase the availability and performance of the service by reducing the latency in fetching the object [2].

Sharing data is fundamental to distributed systems. Therefore, distributed file systems form a very important part of distributed systems. They allow processes to access data stored at a server in secure way similar to data on local disk [3]. Due to this, existence of same data or collocation of a file on multiple client caches is normal in distributed systems.

This sharing of files comes at price. Shared files must remain consistent. When a file is being written and read simultaneously by different clients, there is a potential for different versions of same file in every clients' cache. This is when we say caches are not consistent. Changes made by a client have to be propagated to the appropriate file server, but it involves a finite amount of delay. In addition, if there are replicas of files, maintaining stronger degree of consistency is more

time consuming. Therefore, managing cache consistency in distributed file systems is very challenging.

Synchronization is the solution to ensure consistency of files.

The remainder of this paper discusses related work in section 2, proposed cache coherency model in section 3 and conclusion of the paper in section 4.

2. RELATED WORK

Sun Network File System (NFS) [4] does not support remote file locking. That means locks are maintained locally by the file server. When a client wants to obtain a write/read lock on a file, it has to request lock manager on the file server. This prevents writing of one file by different clients simultaneously. It uses UNIX semantics for maintaining file consistency. All reads and writes go to the file server, which are processed sequentially. So maintaining consistency is easier. Drawback of this mechanism is that it does not scale well.

Whenever a client needs a file, it sends a request to the Andrew File System (AFS) [5] server (file may be replicated). The file is sent to the client. The client then operates on the file locally. If the file is updated, the changes are sent back to the server when the file is closed. The server then sets the 'valid flag' to 'cancelled' of all the clients which have the file in their cache. This prompts the client to re-fetch the latest version of the file. This makes AFS scalable. But if two or more clients are writing into the same file simultaneously, the one who finishes early wins and 'valid flag' for that file of other clients is set to 'cancelled'.

Cache consistency is maintained in both the file systems, but altogether in a different way. NFS server is stateless and does not transfer the file to a client machine. AFS server is stateful and transfers the file to each and every client.

CODA [6] file system is a descendant of AFS that is more resilient to failures. It introduces a concept called Volume Storage Group (VSG) which contains set of replicas of a volume. Ideally, a client is connected to VSG and modifications are propagated to all replicas in VSG. In case one or more replicas are not up or if there is a network partition, then a client sends the modifications to Accessible Volume Storage Group (AVSG). If the client gets disconnected from all the servers (i.e. AVSG is a null set), it continues to use locally available file and when the connection is established, it propagates the changes to its AVSG, if any.

Mobile Agent-based Distributed File System (MADFS) [7] is conceptually a new Distributed File System suitable even in Wide Area Network (WAN). It improves the performance of distributed system by reducing the network transfer and the overhead of cache management.

MADFS divides the whole system into number of domains, each managed and controlled by domain server (DS). All the hosts within a domain are connected by high-speed Local Area Network (LAN). All the DSes are connected to each other by low-speed WAN. The DSes are managed and controlled by a main server (MS). Thus the whole system is hierarchically managed which reduces load on a single server. Every host, DS and MS has a mobile agent capable of accepting request from client process, moving the order to target server to execute and also responsible for name, cache, access management (in case of domain management agent).

Advantages of MADFS are: It works well in WAN by reducing traffic in WAN and reduces overhead in cache coherency management by using Hierarchical and Convergent Cache Coherency Mechanism (HCCM). The disadvantages are: It does not support write sharing.

Sprite File System [8] lets each host cache the file blocks for reading by multiple readers or by single writer at a time. It uses system memory for caching the blocks and not the disk. During

write sharing Sprite File system disables caching altogether whenever there is a writer with any other readers or writers.

Ownership-based Cache Consistency Model in a Distributed File System [9], based on Sprite File System, improves the performance of the file system during write sharing. Advantages are: It supports write sharing and works well only in LAN. The disadvantage is: It does not have good performance in WAN.

3. THE PROPOSED CACHE COHERENCY MODEL

The proposed distributed file system is based on applying the technique of ownership based file system to MADFS. It has been observed that when a host wants to read or write a file, almost 2/3rd of the times it has been accessed by another host within the same domain [10]. This paper presents a model for getting better performance during write sharing within a domain.

As in MADFS, the whole distributed system should be divided into domains; all hosts within a domain should be connected by high-speed link and controlled by DS. All DSes should be connected to each other by low-speed links and controlled by MS.

Every file in the proposed file system should have read and write locks as in the case of MADFS. All the servers using this technique should be stateful. MS should maintain a repository to record the details of file accesses. It should have following details:

File name, version no, DS id, type of lock (read/write), current owner (indicating DS id)

DS, too, has to maintain a repository similar to the one maintained by MS. It must have following details:

File name, version no, host id, type of lock (read/write), current owner (indicating host id)

When a host needs to read a file, it should send request to its DS. One of the two circumstances may arise.

4. DS may already have a valid read lock on that file - it forwards the cached file to the host.

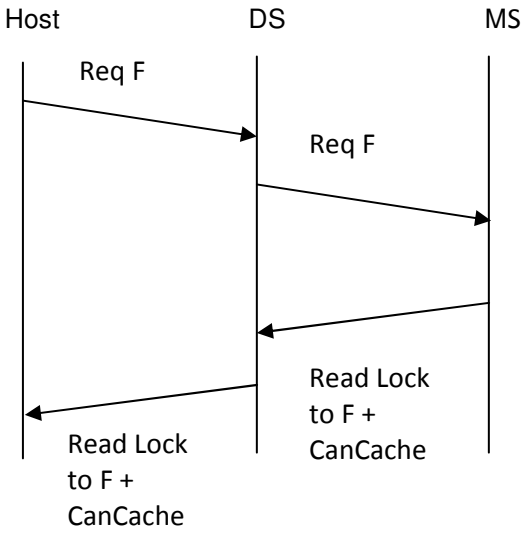


FIGURE 1: Exchange of messages between host, DS and MS for obtaining Read Lock

5. DS may not have a valid read lock on that file – As shown in figure 1, DS forwards the request to MS. MS makes entry into its repository and assigns a read lock to requesting DS. It also sends a 'CanCache' to DS. Upon acquiring the read lock on the requested file and 'CanCache', it grants read lock and sends 'CanCache' to requesting host. DS makes entry into its repository. Valid 'CanCache' allows the requesting entity to cache the file.

When a host needs to write into a file, it sends request to its DS. One of the two circumstances may arise.

1. DS may not have a valid write lock on that file – As shown in figure 2, DS forwards the request for write lock to MS. MS makes entry into its repository and assigns a write lock to requesting DS. It also sends a 'CanCache' as before and 'TempOwner' to DS. 'TempOwner' indicates that the requesting DS holds all ownership rights to the file temporarily (i.e. as long as it is writing the file). Upon acquiring the write lock on the requested file, 'CanCache' and 'TempOwner', it grants write lock and sends 'CanCache' and 'TempOwner' to requesting host. DS makes entry into its repository.

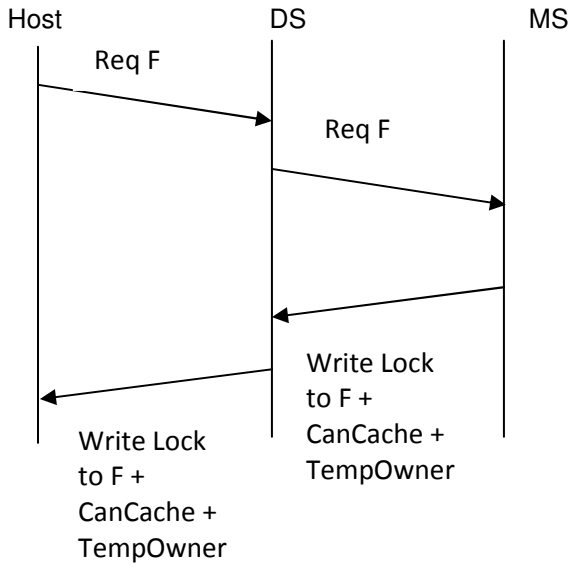


FIGURE 2: Exchange of messages between host, DS and MS for obtaining Write Lock

2. DS may have a valid write lock on that file – This means some other host in the same domain is writing the file. Hence DS should tell requesting host about the 'NewOwner'. If the requesting host receives 'NewOwner', it means it cannot cache the file into its system memory but has to access it remotely.

After writing is completed by the writer, it flushes all the dirty data to the server and releases lock on the file, informs DS which in turn informs MS. DS and MS should remove the respective records from the repository.

Write sharing is said to occur when there is a writer with other readers or writers. It occurs in two cases:

1. Writer and its DS have obtained a write lock and other readers or writers in the same domain want to open the file.

Assume h1 in DS1 has obtained a write lock on file F. h4 wants to obtain a read (or write) lock.

h4 will request DS1. DS1 knows h1 is owner of F, so it tells h4 about 'NewOwner' h1 and maintains h4's record into its repository. h4 now cannot cache F, but should read (or write) F from h1's cache ('CanCache' is not valid for h4), as shown in figure 3.

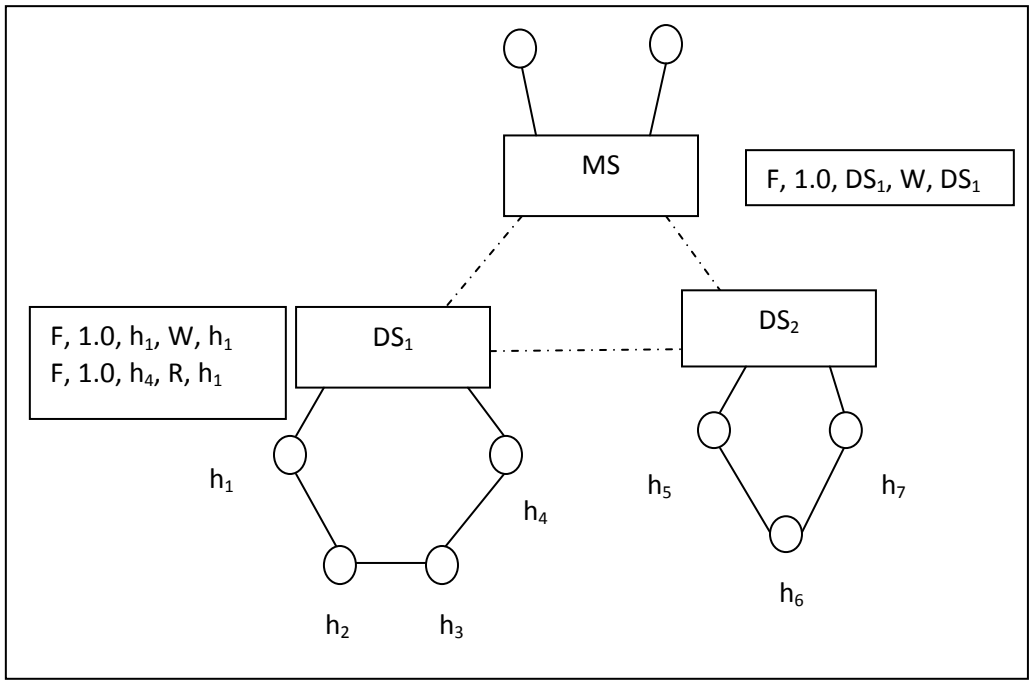


FIGURE 3: Repositories of MS and DS1

2. Writer and its DS have obtained a write lock and other readers or writers from other domain(s) want to open the file.

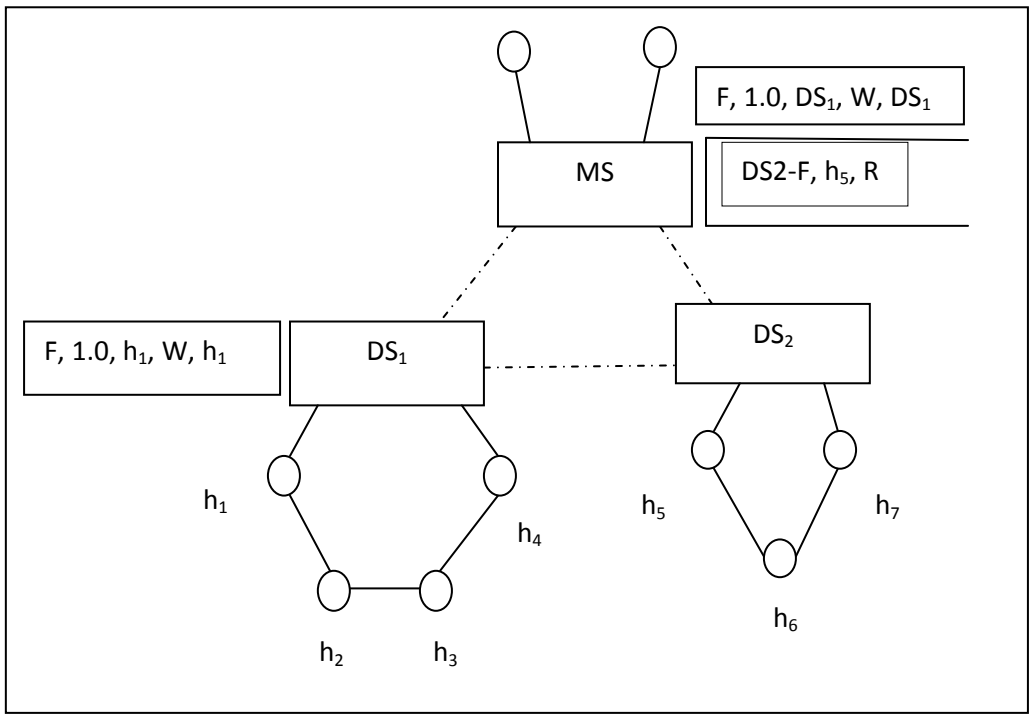


FIGURE 4: Repositories of MS and DS1 and PendingRequestQueue of MS

Assume h1 in DS1 has obtained a write lock on file F. h5 in DS2 wants to obtain a read (or write) lock on F.

It requests DS2. DS2 requests MS. Since the ownership of the file is transferred to DS1, DS2's request is kept pending in the 'PendingRequestQueue' maintained by MS, as shown in figure 4.

3. Readers across different domains and their respective DSes have obtained a read lock and a host wants to obtain a write lock.

Assume hosts h4 from DS1, h5 and h7 from DS2 have obtained read lock on F, as shown in figure 5. h1 in DS1 wants to obtain write lock on F.

h1 requests DS1 for write lock. DS1 requests MS. MS sets 'CanCache' of DS2 to invalid. DS2 sets 'CanCache' of h5 and h7 to invalid. It then grants write lock to DS1. MS's repository changes to:

F, 1.0, DS1, W, DS1

DS1 now grants write lock to h1 along with 'CanCache' and 'TempOwner'. Repository of DS1 changes to:

F, 1.0, h4, R, h1
F, 1.0, h1, W, h1

All the requests to MS from all other DSes are kept pending into 'PendingRequestQueue' as explained before.

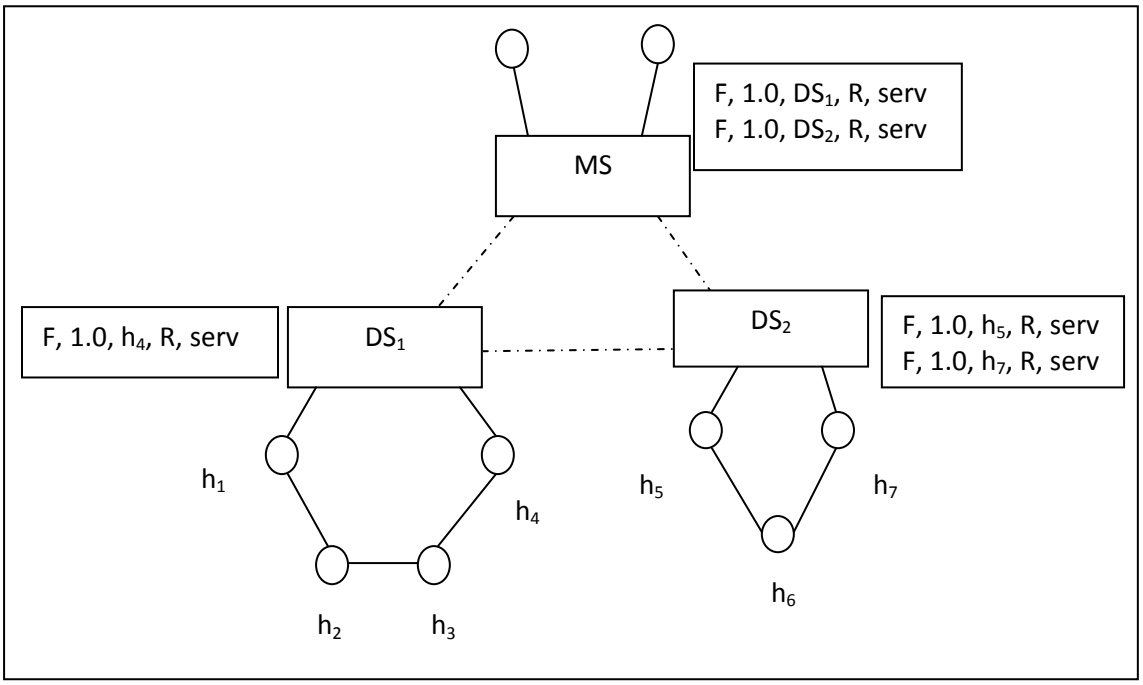


FIGURE 5: Repositories of MS, DS1 and DS2

In all the cases above, dirty data should be flushed to the server and lock on the file should be released by informing DS and MS. DS and MS should remove the respective records from the repository.

Before sending the dirty data to the server and releasing the write lock, no operation should be kept pending by the writer. If a request is received from another reader or writer during flushing of

dirty data to the server, the writer needs to send a 'Negative Acknowledge' to the requesting client. Client understands that it has to delay its operation till it receives 'CanCache' or 'TempOwner' from the server.

4. DISCUSSION

In MADFS, if a host has obtained write lock on a file, no other host (within or outside the domain) can obtain read or write lock on same file. But if there is no write sharing this file system gives good performance.

Ownership-based cache consistency model is suitable for LAN. It allows write sharing. But it does not handle file accesses in WAN.

The proposed file system allows write sharing within a domain. The performance of this file system will be the same for simple reads and writes as in case of MADFS because no change is proposed in the working from this point of view. In case of write sharing, however, the performance will be slightly poorer than MADFS as the file will now be shared for reading/writing by other clients in the domain while a client has held it for writing. During write sharing load of the writer who is 'TempOwner' increases slightly because it has to satisfy the requests of other readers or writers within the same domain. Reader/writer clients (not 'TempOwner') will not have the copy locally available in cache, but will be accessed remotely in LAN, thereby degrading the performance slightly. However, the performance will be the same as ownership based file.

When a client within a domain asks for a file, only for the first time DS has to contact MS and then fetch it from the file server. Any further requests for the same file within that domain are handled by DS. Thus communication overhead is reduced. Also, as the file is available in local cache memory, while being read by multiple clients across different domains simultaneously, the communication overhead is reduced. Thus the protocol works very efficiently [7].

During write-sharing, only one client ('TempOwner') holds the file and rest of the clients reading (or writing) the file, access it from 'TempOwner'. In this case, there is some amount of communication overhead but consistency is maintained.

5. CONCLUSION AND FUTURE WORK

This proposed technique, as discussed above, divides the network into domains. It allows write-sharing within a domain. It does not degrade the performance of MADFS in case of only writing or reading. In case of occasional write sharing, there is a slight degradation of performance.

If the write sharing should be allowed outside domain, then 'TempOwner' rights should not be given to the individual host. They should only be maintained at DS level. Also an attempt could be made to maintain distributed Main Server, so that there is no bottleneck with the Main Server.

6. ACKNOWLEDGEMENT

We would like to thank Dr. Sadashiv Bhide for his constant encouragement. His comments were very helpful.

7. REFERENCES

1. Hennessy and Patterson. *"Computer Architecture a Quantitative Approach*". 3rd Edition.

2. George Coulouris, Jean Dollimore and Tim Kindberg. *"Distributed Systems Concepts and Design".* Third Edition, Pearson Edition, 2006.

3. Andrew Tanenbaum and Maarten van Steen. *"Distributed Systems Principals and Paradigms".* Prentice-Hall of India Pvt Ltd. 2007.

4. Russel Sandberg, David Goldberg, Steve Kleiman, Dan Walsh and Bob Lyon. *"Design and Implementation of Sun Network Filesystem".* Summer Techn Conf USENIX, 1985.

5. John Howard. *"An overview of the Andrew file-system".* USENIX Winter Conference, 1988.

6. M. Satyanarayanan, J. Kistler, P. Kumar, M. Okasaki, E. Siegel and D. Steere. *"Coda: A Highly Available File System for a Distributed Workstation Environment*". IEEE Transactions on Computer, 1990.

7. Jun Lu, Bin Du, Yi Zhu and DaiWei Li. *"MADFS: The Mobile Agent-based Distributed Network File System".* 2009.

8. M. Nelson, B. Welch and J. K. Ousterhout. *"Caching in the Sprite Network File System.* ACM TOCS 1988.

9. ByungGi Hong and TaeMu Chang. *"Ownership based Cache Consistency Model in Distributed File System".* IEEE Tencon 1993.

10. M Blaze and R Alonso. *"Towards Massive Distributed Systems".* Proceedings of 3rd Workshop on Workstation Operating Systems, 92.

A Secured Smart Card Using a Pseudorandom Affine Transformation Based Cipher and a Secured LIRKES

Ehab Mahmoud Mohamed ehab@mobcom.is.kyushu-u.ac.jp
Faculty of Engineering/
Advanced Information Technology Dept/
Wireless Communication Section/Kyushu University
Motooka 744, Nishi-ku, Fukuoka-city 819-0395, Japan
Phone +81-92-802-3573, Fax +81-92-802-3572,

Yassin Mahmoud Yassin Hasan ymyhasan@aun.edu.eg
Faculty of Engineering /Electrical Dept/
Electronics and Communication Section
Assuit University
Assuit, Egypt.

Hiroshi Furukawa furuhiro@is.kyushu-u.ac.jp
Faculty of Engineering/
Advanced Information Technology Dept/
Wireless Communication Section/Kyushu University
Motooka 744, Nishi-ku, Fukuoka-city 819-0395, Japan
Phone +81-92-802-3573, Fax +81-92-802-3572,

Abstract

The RKES (Remotely Keyed Encryption Schemes) are greatly useful in solving the vital problem of how to do bulk encryption/ decryption for high-bandwidth applications (like multimedia and video encryption) in a way that takes advantage of both the superior power of the host and the superior security of the smart card. According to this issue, we propose a novel length increasing (LI) RKES, in which, the output ciphertext length is larger than input plaintext length. In this scheme, an extra ciphertext block is used as a self validation or signature of the whole ciphertext, so an adversary can't forge the scheme.

The proposed LIRKES needs a strong pseudorandom permutation (PRP) as its basic building block, so we introduce a new symmetric-key block cipher, with variable block and key lengths, referred to as PATFC (Pseudorandom Affine Transformation based Feistel Cipher), appropriate for software and hardware implementations. PATFC adopts the 3-round Luby-Rackoff construction (a compact form of the Feistel network structures) for fusing pseudorandom functions of the plaintext partitions to obtain a pseudorandom permutation.

PATFC mainly makes use of a novel keyed pseudorandom function (PRF) that is based on a pseudorandom affine transformation (constructed using a highly nonlinear pseudorandom sequence generator) followed by a data and key dependent encoding and a simple hashing scheme.

Extensive statistical tests of PATFC and its underlying round function consistently demonstrated their competitive diffusion, confusion and pseudorandomness characteristics. Furthermore, PATFC is provably secure and not vulnerable to known/chosen/adaptive plaintext/ ciphertexts attacks.

At the end of this paper, we show how we can apply PATFC as a strong PRP in the suggested LIRKES to be used for smart cards.

Keywords: pseudorandom function (PF), pseudorandom permutation (PRP), Luby-Rackoff ciphers, Feistel Network (FN), LIRKES.

1. INTRODUCTION

Smart cards provide an effective tool for portable safe hardware storage of secret keys critically needed in many recent multimedia applications such as real time access control, software license management, e-technology, e-commerce and e-services [1]. Smart cards are mainly reliable because of their distinctive features of tamper-resistant packaging, loose coupling to the host and low cost [2]. However, with their computationally limited resources, smart cards cannot process large data blocks as fast as the host may need.

The Remotely Keyed Encryption Protocol (RKEP), first introduced by Blaze, addressed how to do bulk encryption/decryption taking advantage of both the superior computational power, speed and resources of the (high bandwidth) host (trusted with plaintexts/ciphertexts) and the superior security of the slow (low bandwidth) smart-card (trusted with the key) [2]. Although of the interesting approach of Blaze, it suffers from some drawbacks. Its drawbacks basically result from the low security of the protocol. Lucks gave three attacks on the blaze's RKEP, namely a chosen plaintext attack, a two sided attack and a forgery attack (working on the decrypt only smart-card) [3]. In addition, Lucks specified three conditions, that Blaze's RKEP does not satisfy any of them, to make a secure RKE scheme (RKES). Moreover, Lucks suggested the RaMaRK "Random Mapping based RKES" which is based on the Luby-Rackoff construction. Although RaMaRK is based upon Lucks' criteria, a critical weakness was found in RaMaRK [4]. Consequently, Blaze, Feigenbaum and Naor suggested an efficient Length Preserving (LP) RKES named BFN-LPRKES, in which the length of the output ciphertext is equal to the length of the input plaintext. Although the BFN-LPRKES is the most efficient scheme from the security point of view, it is not efficient from card computations and keys storages point of views which are critical requirements for inexpensive smart cards. The authors suggested a new LPRKS based upon a general view of the Feistel Network (FN), in which they only used 2-round PRP instead of the 4-round used by Luby-Rackoff. Our proposed LPRKES is more secure than the previous literature, and more efficient from complexity, card computations, and keys storages point of views [5] [6].

In addition to the BFN-LPRKES, Blaze, Feigenbaum and Naor suggested the length Increasing (LI) RKES named BFN-LIRKES as an alternative to solve the RKES problem. Their proposal is based upon adding a signature of the whole ciphertext to the output ciphertext which cannot be computed by an adversary without running the encryption protocol. So, the length of the resulting ciphertext is larger than the length of the input plaintext that why it is called LIRKES [4]. Although Blaze, Feigenbaum and Naor are considered the pioneers in introducing the LIRKES schemes, their proposal contains some security and complexity drawbacks that get it a little bit efficient solution for smart cards security problem.
In this research, we propose a secure and computationally efficient LIRKES. The proposed scheme withstands dictionary attack which can be easily applied to the BFN-LIRKES. In addition, it is suitable for cheap smart cards with a limited computational power.

Because of the requirement for a strong PRP in the proposed LIRKES, we introduce PATFC: Pseudorandom Affine Transformation Based Feistel Cipher as variable block-size symmetric-key block cipher. Block cipher is a PRP that maps a block of bits called plaintext into another block called ciphertext using the key bits. Pseudorandomness implies being not distinguishable form truly random permutation (TRP). In a well designed block cipher, a plaintext bit change should change each bit of the output ciphertext with a probability of 0.5. Also, there should be no plaintext/ciphertext-to-ciphertext correlations. Thus, secure block ciphers should essentially exhibit high degree of pseudorandomness, diffusion, and confusion [7]. In addition, a block cipher

is most practically qualified as secure if it has survived after being extensively exposed to proficient cryptanalysis. The structure of a block cipher may be a substitution-permutation network (SPN) or Feistel network (FN). The Advanced Encryption Standard AES-Rijndael is currently the most famous SPN cipher [8]. Alternatively, the FN structure, which is a universal method for converting a round function into a permutation, is adopted in several ciphers such as the DES, DESX, DEAL, FEAL, GOST, Khufu and Khafre, LOKI, CAST, and Blowfish [7], [8]. Rather than the use of many rounds, such as 16 in the DES, Luby and Rackoff introduced a 3-round FN construction used in designing a provably secure PRP from pseudorandom functions (PRF) [9]. Further analysis and several block ciphers are designed based on the Luby-Rackoff construction [5], [10]–[13].
By adopting the Luby-Rackoff construction, we propose PATFC which is a novel variable block-size symmetric-key block cipher. PATFC mainly makes use of a novel keyed PRF that is based upon a PR affine transformation (PRAT), constructed using a highly nonlinear Pseudorandom Number Generator (PRNG), and followed by a data and key dependent encoding and simple hashing scheme.
Extensive confusion, diffusion and pseudorandomness tests based upon the NIST statistical tests on PATFC and its underlying PRAT-based PRF consistently demonstrated their effectiveness. Furthermore, PATFC is not practically vulnerable to known, chosen and adaptive plaintext/ciphertext as well as dictionary and brute force attacks. It is also suitable for both software and hardware implementations.

Although PATFC is introduced to be used in the proposed LIRKES, it can be used to strengthen wireless mesh networks clients security by applying it as a candidate with a good pseudorandom and security properties in the well known WPA2 protocol used in IEEE 802.11i standard [14], [15]. In addition, we can exploit the whole scheme (PATFC and the LIRKES) to build a smart card based wireless mesh network to enhance its authentication and security in general [16].

The rest of the paper is organized as follows. Section 2 describes the Luby-Rackoff construction in more details, section 3 introduces PATFC and its experimental work, section 4 gives the suggested LIRKES with its cryptanalysis, section 5 shows how we can apply PATFC in the LIRKES, and section 6 gives the conclusions and future work.

2. PRELIMINARIES

Let "$\oplus$"denote the bit-wise XOR operation and $f_1, f_3 : \{0,1\}^r \to \{0,1\}^l$ and $f_2 : \{0,1\}^l \to \{0,1\}^r$ be a keyed PRFs. Given a *k*-bit key $K \in \{0,1\}^k$, a plaintext message $P = (L,R) \in \{0,1\}^{l+r}$ is divided into an *l* -bit (left) block L and *r*-bit (right) block R. Let $C = (U,T) \in \{0,1\}^{l+r}$ be its corresponding ciphertext. In case of *l*=*r* (balanced structure), Luby and Rackoff described how to construct a secure (against known / chosen plaintext attacks) PRP $\psi(f_1, f_2, f_3)(L,R) = (U,T)$ over $\{0,1\}^{l+r}$, from *r*-bit PRF's using a 3-round balanced Feistel network, rather than the use of 16 rounds as in the DES algorithm [9], with U and T computed as follows Fig.1:

$$\begin{aligned} S &= L \oplus f_1(K_1, R), \\ T &= R \oplus f_2(K_2, S), \\ \text{and } U &= S \oplus f_3(K_3, T) \end{aligned} \qquad (1)$$

where $S, U \in \{0,1\}^l$ and $T \in \{0,1\}^r$. Likewise, $\psi(f_3, f_2, f_1)$ yields the inverse PRP.
Note that because the entropy of the required permutation is (*l*+*r*)-bit, at least two rounds of PRFs are needed. But, using two rounds only, the attacker can distinguish the outputs from truly random permutation, if he simply chooses two different inputs with the same R. Luby and Rackoff even suggested the use of 4 rounds to prevent adaptive hosen plaintext-ciphertext attacks. Also unbalanced Luby-Rackoff construction $l \neq r$ is presented [10].

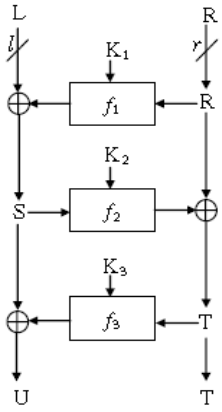


FIGURE 1: Luby-Rackoff cipher construction

3. The Proposed Cipher: PATFC

PATFC is a 3-round balanced (*l=r*) FN cipher, like Luby-Rackoff construction, based on a core PRAT based PRF. The following has motivated building PATFC using the proposed core PRF:

- In matrix-vector multiplication, it is evident that a change in an element of the input vector or major changes in the elements of the transformation matrix diffuse in all elements of the obtained output vector.
- Highly nonlinear PRNG's generate PR sequences that are very sensitive to the key. So, matrices successively constructed form such PR sequences dynamically pseudorandomly change their elements and significantly change with the key.

Thus, PR matrix-vector multiplication implies pseudorandomness, diffusion, and hence confusion [5] [17]. Pre-XORing the input vector with a PR vector (before matrix multiplication) yields the overall PRAT. Actually, the PRAT pseudorandomly substitutes the binary input vector with a vector of PR decimal numbers. Consequently, processing the obtained vector of PR decimal values to obtain the binary output, which incorporating proper additional nonlinearities to the PRF, complicates the cryptanalysis of the (underlying) PRF and the overall FN cipher constructed from it.

In implementing the PRAT, we use the RC4 PR bytes generator, which is a highly nonlinear generator with no public cryptanalytic results against it [7] [8]. The users of PATFC cipher could modify the algorithm to be suitable for any other PRNG such as the SEAL algorithm [7]. But we recommend RC4 because it makes use of a variable length key, and is PR (can likely be in 256×256^2 feasible states), fast, highly secure and invulnerable to linear/differential attacks [8].
Before presenting the PATFC internal construction, it is worth mentioning the balanced and homogeneous structure of PATFC. The FN is unbalanced or balanced if $r \neq l$ or $r = l$, respectively. Since the security of Luby-Rackoff ciphers depends on *min (l,r)*, the balanced structure provides more security [11][12]. Accordingly, to achieve optimal security, PATFC considers (without loss of generality) only the balanced structure, i.e., *l=r*. In addition, a FN is homogeneous if the same PRF is used in all rounds [18]. From the complexity (especially in hardware implementations) and computational burden points of view, it is recommended to reduce the number of different PRF's used in multi-round networks [11]. Therefore, to make the construction more efficient, PATFC uses the same PRF in the three rounds (but with different keys and consequently different PR sequences for the PRAT).

3.1 The Proposed PATFC Round Function *f*

The keyed PR round function f_{Ki} , $i \in \{1,2,3\}$, is the core of PATFC. f_{Ki} consists of PRAT followed by a dynamic (data and key dependent) *b*-bit encoding and simple hashing (Fig.2). Its inputs are an *r*-bit data, a 256-byte key K_i and a parameter representing the internally employed sub-block length $n \le min(r, N_{max})$, where $N_{max} \le r$ is a user specified number.

3.1.1 The PRAT construction

First, K_i is used to trigger an RC4 PRNG to successively generate *n* PR bits and n^2 PR bytes, i.e., $n+64n^2$ PR bits, to construct an $n \times 1$ -bit PR vector $\underline{\boldsymbol{v}}_{\boldsymbol{Ki}}$ and $n \times n$ byte PR matrix $\boldsymbol{G}_{\boldsymbol{Ki}}$.

The input *r*-bit is divided into sub-blocks each of size *n* bits. Then each *n*-bit sub-block of the input data is bitwise XORed with its associated RC4 generated PR bit-vector and then multiplied by its associated PR bytes matrix to obtain a vector of PR decimal values $\in \{0,1,...,255\boldsymbol{n}\}^n$. The actual maximum obtained value per sub-block dynamically changes depending on the patterns of the input data and the RC4 PR sequence.

To sum up, for the j^{th} *n*-bit input sub-block $\underline{\boldsymbol{x}}^{(j)}$, the PART can be represented as follows:

$$\underline{\boldsymbol{y}}^{(j)} = \boldsymbol{G}_{\boldsymbol{Ki}}^{(j)} (\underline{\boldsymbol{x}}^{(j)} \oplus \underline{\boldsymbol{v}}_{\boldsymbol{Ki}}^{(j)}) \tag{2}$$

Where $\underline{\boldsymbol{v}}_{\boldsymbol{Ki}}^{(j)}$ and $\boldsymbol{G}_{\boldsymbol{Ki}}^{(j)}$ are the j^{th} RC4 generated $n \times 1$ - bit PR vector and $n \times n$ byte PR matrix, respectively.

The bit wise XOR operation yields PR substitution whereas the multiplication by a PR matrix (which is actually equivalent to selective addition controlled by $\underline{\boldsymbol{x}}^{(j)} \oplus \underline{\boldsymbol{v}}_{\boldsymbol{Ki}}^{(j)}$) contributes to both diffusion and confusion.

Increasing the value of *n* results is achieving more randomness, diffusion, and confusion by the employed PRAT and the overall *f* (but with more computations).

3.1.2 Binary Encoding

We put each obtained decimal value from the PART into a binary format. The number of encoding bits used *b* dynamically changes from a sub-block to another, depending on its bit-pattern and the associated PR bit/byte patterns of the RC4 PR sequence, i.e., enforcing data and key dependency of the encoding process.

For the j^{th} sub-block, it is computed as

$$b^{(j)} = \max(1, \lceil \log_2 (y_{\max}^{(j)} + 1) \rceil) \tag{3}$$

Where $y_{\max}^{(j)}$ is the maximum decimal number in the j^{th} obtained vector. This step should yield a *br*-bit stream.

3.1.3 Simple Hash (Compression) Process

The *br*-bit stream R_1,obtained from the binary encoding step, is partitioned into *b r*-bit sections and compressed using *r b*-input XOR's working as a simple hash function to yield an *r*-bit output R' as follows:

$$\begin{aligned} R'(\xi)_{\xi=0\to r-1} &= \oplus_{j=0}^{b-1} R_1(\xi + jr) \\ &= R_1(\xi) \oplus R_1(\xi + r) \oplus .... \oplus R_1(\xi + (b-1)r) \end{aligned} \tag{4}$$

3.2 PATFC Key Scheduling Algorithm

The RC4 PRNG requires a 256-byte (2048-bit) key [7]. Hence, a total of 6144-bit key is required for the 3- round functions f_{K1}, f_{K2} and f_{K3}. So, PATFC can work in the 2048-bit key length mode in which the input key length is 2048-bit. Then, a simple key-scheduling algorithm, for example an RC4 based one, can be applied to generate the 3 round keys, K_1, K_2 and K_3, each of length 2048-bit. In other words, with K_1 only, the 3 PRF's may serially use a single RC4 PR sequence, while employing 3 distinct sections of the sequence. In addition, PATFC can work in the variable key length mode, in which the algorithm can accept any length key, and by the simple expansion

process suggested by the authors in [5], the key schedule algorithm can generate the 3-round keys used by PATFC.

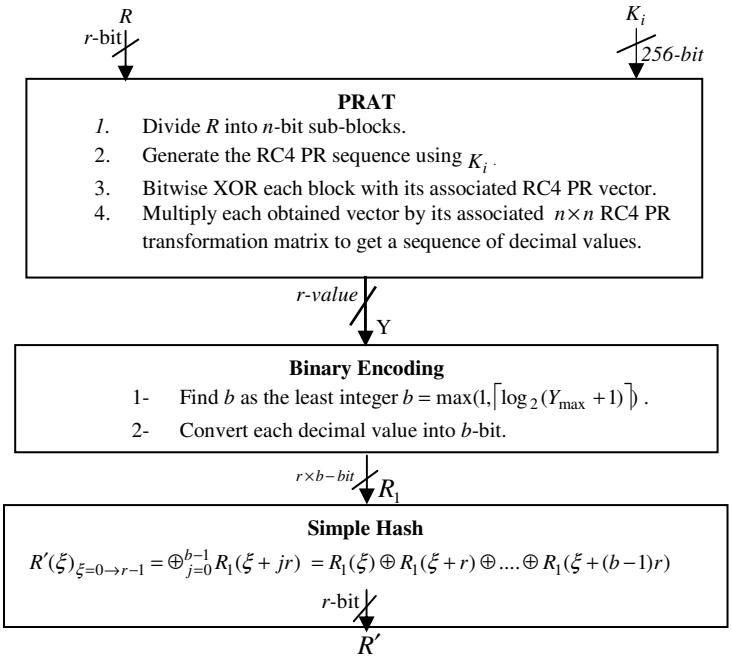


Figure 2: The proposed PATFC round function (*f*)

3.3 PATFC Cryptanalysis

In this section, we consider the performance of PATFC under several attacks types.

1- Possibility to attack the round function *f* :

In equation 2, $\underline{v}_{Ki}$ and G_{Ki} are successively generated using the PRNG, so they differ from a block to another. Choosing the PRNG to be highly nonlinear and secure leaves no helpful information to the attacker to predict its output sequence without knowing the seed value. i.e., the key used. Based on equation 2, the mean value of the elements of $\underline{y}$ is as follows:

$$E[y_i] = E\ [\boldsymbol{G}_K(\boldsymbol{i},:)\underline{\boldsymbol{v}}_K] = E\ [\sum_{j=1}^{n} G_K(i,j)v_K(j)] =$$

$$\sum_{j=1}^{n} E[G_K(i,j)v_K(j)] = \sum_{j=1}^{n} E[G_K(i,j)]E[v_K(j)] = \sum_{j=1}^{n} \frac{255}{2}\left(\frac{1}{2}\right) = \frac{255n}{4}, \qquad 1 \le i \le n\ . \tag{5}$$

Where *E* means the expected value.

It is appeared from equation 5 that the values of $\underline{y}$ almost widely spreads around $\frac{255n}{4}$. Therefore, it is a formidable task for the attacker to guess the values of $\underline{y}$ especially for large values of *n*, so every *r*-bit block output of the *f* function has almost the probability of $\frac{1}{2^r} + \varepsilon$, where ε depends upon the PRNG used.

2- Exhaustive key search attack (brut search attack):

In this attack, the attacker has many plaintext-ciphertext pairs encrypted under the same key and his job is to search all possible keys to find the key used in the encryption process. But, PATFC can accept a variable key ≤ 2048 -bit. So, practically, PATFC can effectively withstand the exhaustive key search attack.

3- Dictionary attack:
In this attack, the attacker makes a look up table (LUT) containing all possible plaintexts/ciphertexts pairs encrypted under all possible keys. Due to PATFC design as a variable block-size and key-length cipher, in case of dividing the plaintext into randomly sized blocks, the attacker neither knows the input plaintext length nor the key length. So he cannot practically make such a dictionary. Also if the whole plaintext is encrypted as a single block, the block size (and hence the codeword and number of entries of the LUT) is too large to practically try to construct the needed LUT.

4- Linear and Differential Cryptanalysis:
After the RC4 PRAT, the data/key-dependent encoding followed by the hashing scheme all together represents a highly PR nonlinear operation. So, even if some plaintext-ciphertext pairs (for the same key) are available to the attacker, the high PR nonlinearity of PATFC makes it invulnerable to linear and differential cryptanalysis. However, more analysis needs to be done to confirm our claim. On the other hand, since in linear and differential attacks [7], the attacker wants to know multiple distinct plaintexts-ciphertexts pairs for the same key, to know some of the key bits, we can encrypt the whole message at once using a different key each time or simply keep the employed PRNG in the PRAT step running and use its successive outputs for encoding the successive blocks.

5- Adaptive chosen plaintext/ciphertext attack:
The 3-round Luby-Rackoff ciphers may not prevent the adaptive chosen plaintext/ciphertext (two-sided) attack, which is the strongest attack against any symmetric key block cipher (despite being of little practical availability where the attacker can reach both the encryption and decryption engines). So, as suggested by Luby and Rackoff [9], a 4-round PATFC successfully prevents such type of attack.

Input plaintext 64-bit (In Hex)	User-key 64-bit (In Hex)	Output ciphertext64-bit (In Hex)
{0000-0000-0000-0000}	{0000-0000-0000-0000}	{78 EC-00C1-8915-8318}
{0000-0000-0000-0000}	{0000-0000-0000-0001}	{D67B-52F4-0F3F-E73E}
{0000-0000-0000-0001}	{0000-0000-0000-0000}	{3161-B32C-88BE-98D6}
{0000-0000-0000-0000}	{FFFF-FFFF-FFFF-FFFF}	{B3B9-3458-9307-D1E7}
{FFFF-FFFF-FFFF-FFFF}	{FFFF-FFFF-FFFF-FFFF}	{F2BA-89F5-6A60-4383}
{0000-0000-0000-0001}	{FFFF-FFFF-FFFF-FFFF}	{AE8F-0874-354D-F6B6}
{FFFF-FFFF-FFFF-FFFE}	{FFFF-FFFF-FFFF-FFFF}	{277F-0BDE-66E5-7926}
{FFFF-FFFF-FFFF-FFFF}	{FFFF-FFFF-FFFF-FFFE}	{AE04-F8DB-37F2-A7E5}
{FFFF-FFFF-FFFF-FFFF}	{0000-0000-0000-0000}	{3570-B0DA-3126-B6A3}

TABLE 1: Examples of 64-bit test vectors (in Hex) for PATFC

3.4 PATFC Experimental Work

We fully software implemented PATFC as a variable block-size variable key-length cipher with a simple effective key scheduling scheme. Table.1 presents examples of plaintext-key-ciphertext PATFC test vectors, especially including low and high density and correlated plaintext and key patterns, assuming 64-bit plaintext/key that shows PATFC excellent diffusion and confusion properties.

As in all Luby-Rackoff ciphers, security and pseudorandomness of the cipher is based upon the PR of the employed keyed round PRF f_K. The diffusion and confusion properties as well as pseudorandomness of the proposed PRF and the overall PATFC have been verified using extensive statistical diffusion and confusion as well as NIST tests [19].

Diffusion Test: 100 64-bit (32-bit for testing the round function) PR plaintexts P_i, i=1,2,.. ..,100 and 100 64-bit key K_i, i=1,2,.. .., 100, are generated using the SEAL algorithm. For each P_i, 64 1-perturbed-bit plaintexts $\{P_{i,j}, j=1,2,.. ..,64\}$, with the *jth* bit inverted, are generated. Then, the histogram, mean value and variance of the 6400 hamming distances $d_{i,j}=\sum(E_{Ki}(P_i) \oplus E_{Ki}(P_{i,j}))$ are computed, where $E_{Ki}(P_i)$ means the encryption of plaintext P_i using the K_i key.

Confusion Test: For the $P_{i,j}$*'s* mentioned above, the histogram, mean value and variance of the 6400 plaintext-ciphertext correlation coefficients $\rho_{i,j}= corr(P_{i,j},E_{Ki}(P_{i,j}))$ are computed. Also, for the P_i*'s* and $P_{i,j}$*'s* the histogram, mean value and variance of the 6400 ciphertext-ciphertext (of correlated plaintexts) correlation coefficients $\rho_{ij} = corr(E_{Ki}(Pi),E_{Ki}(P_{i,j}))$ are computed.

The results of the confusion and diffusion tests (summarized in Table.2 and Fig.3, 4 and 5) illustrate the competitive performance of PATFC compared with the DES and IDEA ciphers [7] as the correlations are almost zero and the percentage of the changing bits due to 1-bit perturbations is almost 50%.

NIST Pseudorandomness tests: The NIST Test Suite is a statistical package composed of 16 tests, basically developed to test the randomness of PRNG sequences. To use the NIST tests for testing the pseudorandomness (and implicitly the diffusion and confusion) of a block cipher, 7 data types are generated, following the procedure suggested in [20]. Of each data type, 100 4096-bit binary sequences were analyzed. These data types include: Plaintext-Avalanche, Key-Avalanche, Plaintext-Ciphertext Correlation, Low-Density Plaintext, Low-Density Key, High-Density Plaintext and High-Density Key data types.

The following 13 tests, with 32 *p*-values, of the 16 NIST tests were applied, namely the frequency (monobit), frequency within a Block (using a 128-bit block length), runs, longest run-of-1's in a block (using a 128-bit block length), binary matrix rank (with a 3×3 size), discrete Fourier transform, overlapping template matching (using a template of 9 1's, with a block length of 512-bit), Maurer's "universal statistical" (with 4-bit per block with 60 blocks for the initialization sequence), linear complexity (with a 20-bit block length), serial (with a 3-bit block length), approximate entropy (with a 2-bit block length), cumulative sums (Cusums), and random excursions variant tests.

Cipher Algorithm	Diffusion block length=64	Confusion tests block length=64	
		plain /cipher texts Corr.	Ciphertexts Corr.
	mean/64, var/64	Mean, var	Mean, var
PATFC	0.49, 0.24	2.546e-4, 9.82e-4	8.93e-5, 9.65e-4
DES	0.50, 0.24	-1.05e-5, 9.46e-4	-2.93e-4, 9.67e-4
IDEA	0.50, 0.25	-4.43e-4, 9.65e-4	-6.17e-4, 9.78e-4

TABLE 2: Comparison between the PATFC, DES, and IDEA.

Significance level of 0.01 indicates that one would expect 1 sequence out of 100 sequences to be rejected. A *p*-value ≥ 0.01 means that the sequence can be considered as random with a confidence of 99%. For each *p*-value, either success or failure evaluation was made based on being either above or below the pre-specified significance level of α=0.01 [19]. For each 100

sequences, two quantities were determined: the proportion of binary sequences passing the statistical test and an extra uniformity *p*-value based on a chi *χ2* test (with 9 degree of freedom) applied to the *p*-values of the 100 sequences. A sample (of 100 sequences) was considered to be passed a statistical test if its proportion of success exceeded

$$(1-\alpha)-\sqrt[3]{\frac{\alpha(1-\alpha)}{\boldsymbol{m}}}=0.99-\sqrt[3]{\frac{0.99\times 001}{100}}\approx 0.94 \quad (6)$$

i.e., 94%, and the uniformity test *p*-value exceeds 0.0001 [19]. The obtained results of the 32 *p*-values of the NIST tests successfully verified the pseudorandomness, diffusion and confusion properties of the proposed PRF and the overall PATFC with more than 94% proportion of succeeded sequences. Figures 6-8 illustrate samples of the obtained results, specifically the proportion of succeeded sequences for the 32 NIST tests applied to PATFC with Plaintext-Avalanche, Key-Avalanche, and Plaintext-Ciphertext Correlation generated data types.

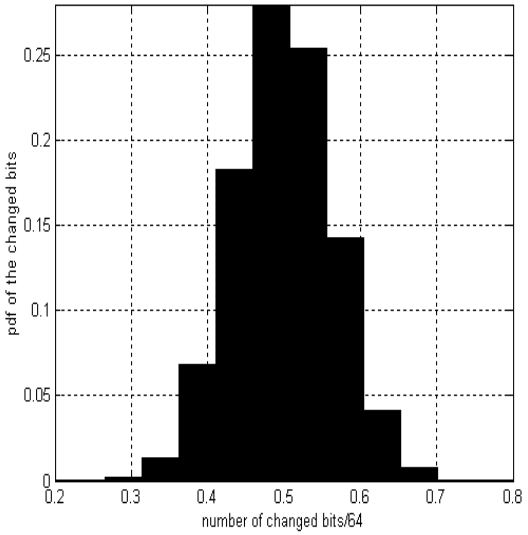


FIGURE 3: Diffusion test histogram: PATFC

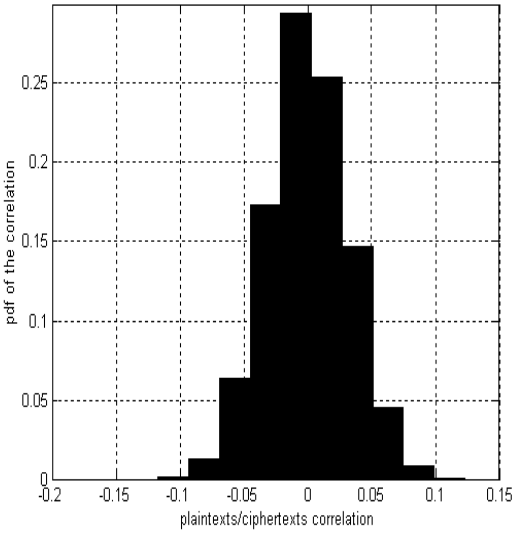


FIGURE 4: Confusion test: PATFC Plaintexts-Ciphertexts Correlations histogram

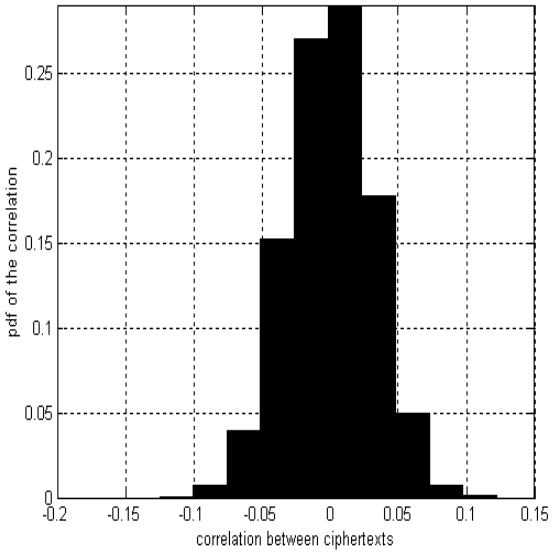


FIGURE 5: Confusion test: PATFC ciphertexts Correlations histogram

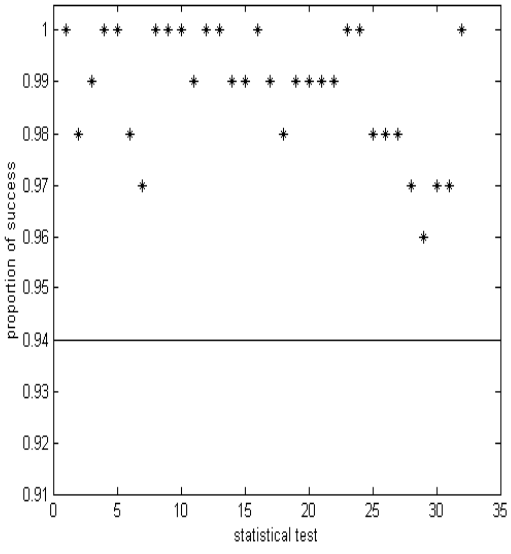


FIGURE 6: NIST tests using Plaintext-Avalanche data: Proportion of succeeded sequences for PATFC

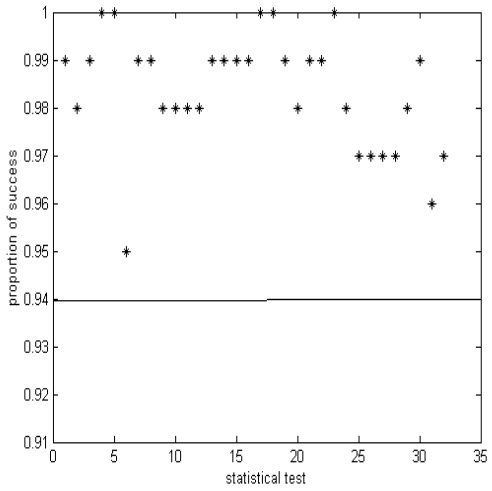


FIGURE 7: NIST tests using
Plaintext- Ciphertext correlation:
Proportion of succeeded sequences for PATFC

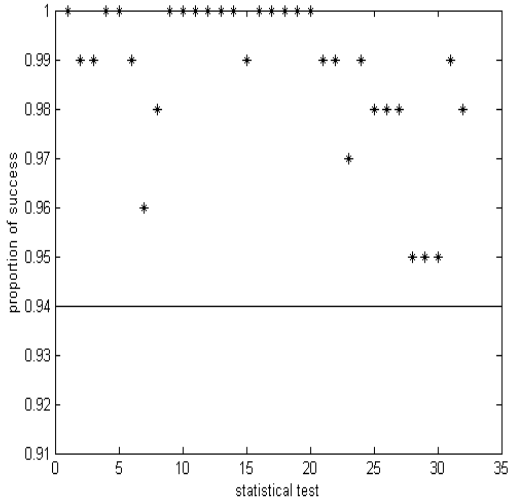


FIGURE 8: NIST tests using
key-Avalanche data:
Proportion of succeeded sequences for PATFC

4. A Novel LIRKES for Smart Cards

4.1 Overview of the BFN-LIRKES

As an attempt to solve the RKES problem, Blaze, Feigenbaum and Naor suggested a new trend different from the previous proposals [4]. Their trend is based upon the idea of self validation; Self validation means adding a signature ciphertext block to the original ciphertext, so, the resulting ciphertext length after adding the signature block is larger than the input plaintext length; as a result, their scheme is a length increasing (LI) RKES.

By using this idea, they suggested two schemes; one of them is insecure that any adversary can easily forge, and the other is a secure one that an adversary cannot forge. We will focus on the secure one.

The details of this scheme are as follows:

Secured BFN-LIRKES

Encryption protocol: input $P_1, P_2, ..., P_n$;output $t, C_0, C_1, C_2, ..., C_n$.

1. Generate session key S.
2. Host: $C_i = E_S^i(P_1, P_2, ..., P_n)$, $i \in \{1, ..., n\}$.
3. Host: $h = H(C_1, C_2, ..., C_n)$.
4. Host $\rightarrow$ Card: S, h.
5. Card: $C_0 \leftarrow E_{K_1}(S)$.
6. Card: $t \leftarrow F_{K_4}(F_{K_3}(C_0) \oplus F_{K_2}(h))$.
7. Card $\rightarrow$ Host: C_0, t .

Decryption protocol: input $t, C_0, C_1, C_2, ..., C_n$; output $P_1, P_2, ..., P_n$ or “invalid”.

1. Host: $h = H(C_1, C_2, ..., C_n)$.
2. Host $\rightarrow$ Card: C_0, h, t .
3. Card: if $t \neq F_{K_4}(F_{K_3}(C_0) \oplus F_{K_2}(h))$ Then $S \leftarrow$ “invalid”

Else $S \leftarrow D_{K_1}(C_0)$.

4. Card $\rightarrow$ Host: S.
5. Host: if S $\neq$ "invalid"
   Then $\{P_i = D_S^i(C_1, C_2, ..., C_n)$; output $(P_1, P_2, ..., P_n)\}$.
   Else output "invalid".

4.2 A Novel LIRKES

In this section, we will introduce a new LIRKES that overcomes the drawbacks of the BFN-LIRKES. Our scheme is also based upon the idea of self validation, but it is more secure and more efficient from card computations and key storages point of views than the BFN-LIRKES.

The proposed LIRKES

Encryption protocol: input $P_1, P_2, ..., P_n$; output $C_0, C_1, C_2, ..., C_n$.

1. Generate session key S by a best disposal.
2. Host: $C_i = E_S^i(P_1, P_2, ..., P_n)$, $i \in \{1, ..., n\}$.
3. Host: $h = H(C_1, C_2, ..., C_n)$.
4. Host $\rightarrow$ Card: S, h.
5. Card: $Z = H(K_1 \mid h)$. | Means concatenation
6. Card: $C_0 = E_Z(S)$.
7. Card $\rightarrow$ Host: C_0.

Decryption protocol: input $C_0, C_1, C_2, ..., C_n$; output $P_1, P_2, ..., P_n$.

1. Host: $h = H(C_1, C_2, ..., C_n)$.
2. Host $\rightarrow$ Card: C_0, h.
3. Card: $Z = H(K_1 \mid h)$.
4. Card: $S = D_Z(C_0)$.
5. Card $\rightarrow$ Host: S.
6. Host: $P_i = D_S^i(C_1, C_2, ..., C_n)$.

4.3 Security Analysis and Advantages of the proposed LIRKES

Theorem 1: The proposed LIRKES is forgery secure with probability $\frac{q(q-1)/2}{2^s} + \varepsilon$.

The proposed length increasing scheme is forgery secure in the sense that any probabilistic polynomial time adversary who access to the scheme during the HOST phase and makes *q* encryptions/decryptions with arbitrarily chosen inputs, can know no more than *q* valid plaintexts/ciphertexts pairs.

Proof: The collision resistance of *H* implies that $H(X_1) \neq H(X_2)$, for $X_1 \neq X_2$. So the chance for the adversary to find $(C_1, ..., C_n) \neq (C_1', ..., C_n')$ such that $H(C_1, ..., C_n) = H(C_1', ..., C_n')$ is negligibly small. Then the probability that $Z = Z'$, for $h \neq h'$ is also negligibly small. Also by assuming that the encryption function *E* is a strong invertible pseudorandom permutation. Then the probability that: $P_r(E_Z(S) = E_{Z'}(S')) = \frac{1}{2^s} + \varepsilon$, where $E(.) = \{0,1\}^z \times \{0,1\}^s \rightarrow \{0,1\}^s$, *z* and *s* are the lengths of Z and S respectively, and ε is a small number depends upon the pseudorandomness of *E*, and If *E* is truly random then ε=0. In addition, If the attacker makes *q* encryptions then there are *q(q-1)/2* different messages pairs then $P_r(E_Z(S) = E_{Z'}(S')) \leq \frac{q(q-1)/2}{2^s} + \varepsilon$.

Theorem 2: The proposed LIRKES is pseudorandom
Proof: From the above analysis, we can conclude that our proposed scheme is also pseudorandom.

Advantages of the proposed scheme over the BFN-LIRKES

1. The length of output ciphertext in the proposed scheme is shorter than the BFN-LIRKES. While the BFN-LIRKES uses two fields (t, C_0) to define the self validation (ciphetrext signature), we only use on field (C_0) to do that.
2. Our scheme is a self checking scheme; that is the checking step is inherited in the signature block C_0, i.e., if C_0 is incorrect the decryption protocol will output a random plaintext other than the correct one. consequently, there is no need for the checking steps used in the BFN-LIRKES which increases the complexity of the scheme.
3. We can attack BFN-LIRKES scheme by using a dictionary attack, i.e. if an adversary can access to the scheme many times, assuming the card uses the same K_1 every time, he can make a dictionary contains all values of S and its corresponding values of C_0. The size of this dictionary is 2^S, where *s* is the length of the session key S. So if the attacker successes in making such a dictionary, he can easily get the value of S for any C_0, so he can decrypt any message contains C_0. Therefore in BFN-LIRKES S must be very large. In contrast, in the proposed scheme the value of C_0 don't depend only on S but also upon *h*, for constant K_1, where *h* is the output length of the hash function, so the dictionary size will be $2^S \times 2^h$. As a result, the dictionary size is very large which gets such type of attacks computationally infeasible.
4. The proposed scheme is more efficient than the BFN-LIRKES from the card computation and key storage point of views. In BFN-LIRKES, the card uses four different keys but in our scheme we only use one key. In addition, the BFN-LIRKES requires from the card to evaluate four different functions, but in our scheme we require from the card to evaluate only two functions. In conclusion, the proposed scheme is suitable for cheap smart cards while BFN-LIRKES requires expensive ones.

5. The Application of PATFC in the Proposed LIRKES.

Figure 9 shows how we can apply PATFC as a strong PRP in the suggested LIRKES.

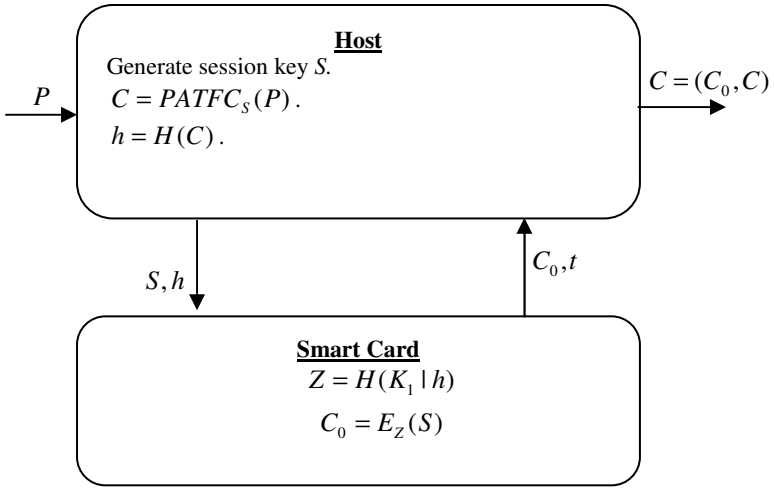


FIGURE 9: The proposed LIRKS using PATFC

6. CONSLUSION & FUTURE WORK

This paper deals with cryptographic smart cards protocols which are used to organize the bulk encryption process between the host and the smart card. In an attempt to solve this important issue, we introduce a new LIRKES that overcomes the drawbacks of the previous proposals. In addition we analyze this scheme from security and smart card efficiency point of views.
Because the suggested LIRKES is highly depending upon a strong PRP, we also present PATFC: Pseudorandom Affine Transformation based Feistel Cipher; a novel Luby-Rackoff construction-based variable block and key lengths symmetric-key block cipher. Its core function is a new pseudorandom function that consists of a pseudorandom affine transformation followed by binary encoding and a simple hashing scheme. Extensive simulations, diffusion, confusion, and NIST pseudorandomness tests proof that PATFC and its round function are good PRP and PR function respectively. However, PATFC needs a complexity analysis beside the security analysis, but we believe that PATFC is less complex.
Also, we show how PATFC can be applied as a PRP in the suggested LIRKES. For future development, we will try to apply our cipher and LIRKES in enhancing the security and authentication of the wireless mesh networks especially the wireless backhaul system.

7. REFERENCES

[1] S. Yuan and J. Liu, “Proceedings of the IEEE international conference on e-tech, e-commerce and e-services,” pp.91–110, 2004.

[2] M. Blaze, “High-bandwidth encryption with low-bandwidth smartcards,” Lecture Notes in Computer Science, vol.1039, pp.33–40, 1996.

[3] S. Lucks, “On the security of remotely keyed encryption,” Proceedings of the Fast Software Encryption Workshop, pp.219–229, Springer, 1997.

[4] M. Blaze, J. Feigenbaum, and M. Naor, “A formal treatment of remotely keyed encryption,” Lecture Notes in Computer Science, vol.1403, pp.251–265, 1998.

[5] E. M. Mohamed, Y. Hasan, H. Furukawa,” A Novel Luby-Rackoff Based Cipher in a New Feistel-Network Based LPRKES for Smart Cards”, International Journal of Computer Science and Security IJCSS, vol 3, pp 66- 81, 2009.

[6] Yasien M. Yasien, E. M. Mohamed “Two-Round Generalized FEISTEL Network Key-Linking Block Ciphers For Smart Card Applications”, Information Security Symposium (ISS), Al-Madinah Al-Munawwarah, Saudi Arabia, 2-4 May 2006.

[7] A. Menezes, P. Van Oorschot, and S. Vanstone, Handbook of Applied Cryptography, CRC press, 2001.

[8] A. Biryukov, “Block ciphers and stream ciphers: The state of the art,” Lecture Notes in Computer Science, Proc. COSIC Summer Course, 2003.

[9] M. Luby and C. Rackoff, “How to construct pseudorandom permutations from pseudorandom functions,” SIAM Journal on Computing, vol.17, no.2, pp.373–386, 1988.

[10] M. Naor, “On the Construction of Pseudorandom Permutations: Luby-Rackoff Revisited,” Journal of Cryptology, vol.12, no.1, pp.29–66, 1999.

[11] R. Anderson and E. Biham, “Two practical and provably secure block ciphers: BEAR and LION,” Lecture Notes in Computer Science, pp.113–120, 1996.

[12] P. Morin, “A critique of BEAR and LION,” Manuscript, citeseer. nj. nec. Com/124166. html.

[13] Y. Hasan, “YC: A Luby-Rackoff ciphers family driven by pseudorandom vector/matrix transformations,” Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on, pp.1–4, 2007.

[14] S. Frankel, B. Eydt, L. Owens, and K. Kent, “Guide to ieee 802.11 i: Establishing robust security networks,” Technical Report 800-97, National Institute of Standards and Technology Administration US Department of Commerce, Computer Security Division Information Technology Laboratory National Institute of Standards and Technology Gaithersburg, MD 20899-8930, 2006.

[15] F. Martignon, S. Paris, and A. Capone, “MobiSEC: a novel security architecture for wireless mesh networks,” Proceedings of the 4th ACM symposium on QoS and security for wireless and mobile networks, pp.35–42, ACM New York, NY, USA, 2008.
[16] M. Siddiqui and C. Hong, “Security issues in wireless mesh networks,” IEEE intl. conf. on multimedia and ubiquitous engineering, 2007.
[17] Y. Hasan, “From stream to provably secure block ciphers based on pseudorandom matrix transformations,” Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on, pp.260–265, 2008.
[18] U. Maurer, “A simplified and generalized treatment of Luby- Rackoff pseudorandom permutation generators”, Proceedings Advances in Cryptology- EUROCRYPT 92, LNCS, vol.658, pp.239-255, Springer-Verlag, 1992.
[19] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, and E. Barker, “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications,” , 2001.
[20] J. Soto and L. Bassham, “Randomness Testing of the Advanced Encryption Standard Finalist Candidates. National Institute of Standards and Technology (NIST),” Computer Security Division, 2000.

An Overview of Registration Based and Registration Free Methods for Cancelable Fingerprint Template

Radhika Bhagwat radhika.bhagwat@cumminscollege.in
Lecturer/IT Department
Pune University
Pune, India

Anagha Kulkarni anagha.kulkarni@cumminscollege.in
Lecturer/IT Department
Pune University
Pune, India

Abstract

Cancelable biometric techniques are becoming popular as they provide the advantages of privacy and security, not provided by biometric authentication system. It transforms a biometric signal or feature into a new signal or feature by some transformation. These are non invertible transforms to make sure that the original biometric template cannot be recovered from them. Most of the existing methods for generating cancelable fingerprint templates need an absolute registration of the image. Therefore they are not robust to intra user variations. But there also exists methods that do not require registration of the image. This paper provides a comparison between two such methods, one that needs registration and other that does not need registration.

Keywords: Cancelable biometrics, non invertible transformation, registration, registration free.

1. INTRODUCTION

The three fundamental techniques used in authentication systems are:

a. Something you know – refers to passwords and PINs.
b. Something you have – refers to tokens and cards.
c. Something you are – refers to biometrics.

The first two techniques used in traditional authentication systems are very famous but have certain disadvantages such as, passwords and PINs can be guessed or disclosed through accident or can be intentionally shared, like passwords, cards or tokens can be stolen and passwords need to be memorized. Moreover it cannot distinguish between an authentic user and a user that has gained access to password. To cater these problems, biometric authentication systems are used. Biometric technologies have automated the identification of people by one or more of their distinct physical or behavioral characteristics. Instead of depending on things that an individual may have or may know, it depends on the attributes of people. Biometric verification techniques try to match measurements from individuals like

fingerprint, hand, eye, face or voice, to measurements that were previously collected. Biometric authentication systems have advantages over traditional authentication schemes. The advantages are, biometric information cannot be obtained by direct covert observation, it is impossible to share and difficult to reproduce, it enhances user's convenience by alleviating the need to memorize long and random passwords and it protects against repudiation by the user. But even with all these advantages biometric techniques have security and privacy problems. Biometrics like voice, fingerprint, signature etc. can be easily recorded and misused without user's consent. PINs and passwords, if compromised, can be reset, but biometrics once compromised is compromised forever. If a biometric is compromised, then all the applications using that biometric are compromised. Cross matching of the stored information can be used to track individuals without their consent.

Cancelable biometric overcomes these disadvantages. Cancelable biometric is an intentional and systematic repeatable distortion of biometric features in order to protect user specific data. In this, the application does not store the original biometric but transforms it using a one way function and stores the transformed version. This method gives privacy and security as it is computationally very difficult to recover the original template from the transformed version. The transformation can be done either in signal domain or in feature domain. In signal domain, the raw biometric signal acquired from sensor is transformed (e.g. images of faces and fingerprint), while in feature domain, the processed biometric signal is transformed (e.g. minutiae of fingerprint).During the enrollment process, the fingerprint template is distorted by a one way transform, using a user specific key. Then, instead of storing the original fingerprint template, its distorted version is stored in the database. During verification, the query fingerprint template is distorted using the same function and then the distorted version is compared with the original, to give a similarity score.

Several approaches have been proposed regarding cancelable biometrics. This paper focuses on comparison between two methods used to generate cancelable fingerprint template. There are many approaches that construct cancelable fingerprint template and need absolute registration of the image before transformation [1], [7], [8], [9], while there also exist approaches where registration is not an absolute requirement and purely local measurements are sufficient for this purpose [3], [15]. Further part of the paper is organized as follows. The requirements for generating cancelable transform are explained, then the registration process which is the most important step in fingerprint matching is explained. Further part presents the registration based method and registration free method for generating cancelable fingerprint template followed by a comparison between the two methods and conclusion.

2. REQUIREMENTS FOR GENERATING CANCELABLE TRANSFORM

There are several challenges to overcome before successfully designing a cancelable transform that transforms the fingerprint template into a cancelable template. They are:

1. If two fingerprint templates x_1 and x_2 do not match, as they do not belong to the same individual, then, even after applying the transformation they should not match.
2. If two fingerprint templates match, as they belong to same person, then they should match even after applying the transformation.
3. Transformed version of the biometric should not match with the original biometric.
4. Two transformed versions of same template should not match.

3. REGISTRATION

One more very important requirement for generating cancelable fingerprint template is 'registration'. But this step is not always required. This depends on which method is used for generating the cancelable fingerprint template. It is required when the method used is registration based and not required when the method is registration free. In this paper, two methods, one registration based and other registration free are studied and are compared to review their characteristics.

Fingerprint registration explained in [6], [12] is a very critical step in fingerprint matching. Although a variety of registration alignment algorithms have been proposed [10], [11], accurate fingerprint registration

remains an unsolved problem. Fingerprint registration involves finding the translation and rotation parameters that align two fingerprints. In order to determine the degree of similarity between two fingerprints, it is first necessary to align the prints so that corresponding features may be matched. Aligning two images can be done in a number of ways like extracting the minutiae and then aligning, using orientation field for aligning, aligning based on generalized Hough transform [14], identifying distinctive local orientations and using them as landmarks for alignment, etc. Alignment has to be explored first, for matching the corresponding components of two templates or images. Traditional approach of fingerprint registration is based on aligning minutiae features. Given two fingerprint images all of the minutiae are extracted from each print and their location, orientation and type are recorded. Registration is based on aligning these two minutiae sets. For two sets of minutiae M1 and M2, ideal case of transformation is

$$f(M1) = M2 \tag{1}$$

However, ideal transformation does not exist since it is practically impossible for a user to place exactly the same part of his/her finger on a sensor and exert the same pressure on the sensor during two different fingerprint capture occasions. The error between the transformed version and the original fingerprint template $E(f(M1), M2)$ has to be minimized and for this optimal transformation has to be found out. Matching minutiae sets has following limitations:

1. Every time a fingerprint is obtained, a different area of the finger surface may be captured. Therefore alignment should be based only on the overlap area of the print and the corresponding minutiae subsets.
2. Missing and spurious minutiae are common when the fingerprint image quality is low. Therefore the alignment algorithm must allow some minutiae to be unmatched even in the area of overlap.

It is known that fingerprint deforms when pressed against a flat surface. This deformation changes the locations and orientations of the minutiae making it impossible to find a perfect alignment of the subsets. Therefore most registration algorithms attempt to find an alignment that minimizes these errors. But finding the optimal alignment is very difficult. Due to large number of possible translations, rotations and distortions, aligning fingerprint has a high computational overhead. One way to deal with these complexities is to use supplementary information from other fingerprint features to help the alignment process. Other features that can be used are local structural features, ridge shape, pixel intensities etc.

4. REGISTRATION BASED GENERATION OF CANCELABLE FINGERPRINT TEMPLATE

Ratha et al [1], [2] pioneered the concept of cancelable biometrics where they have proposed three transformation methods.

In the first method, i.e. the Cartesian coordinate transformation method, the image plane is divided into rectangles and then the rectangles are shuffled based on the user password such that any two rectangles can map to a single rectangle. Figure (1) shows that more than two cells can be mapped to the same cell.

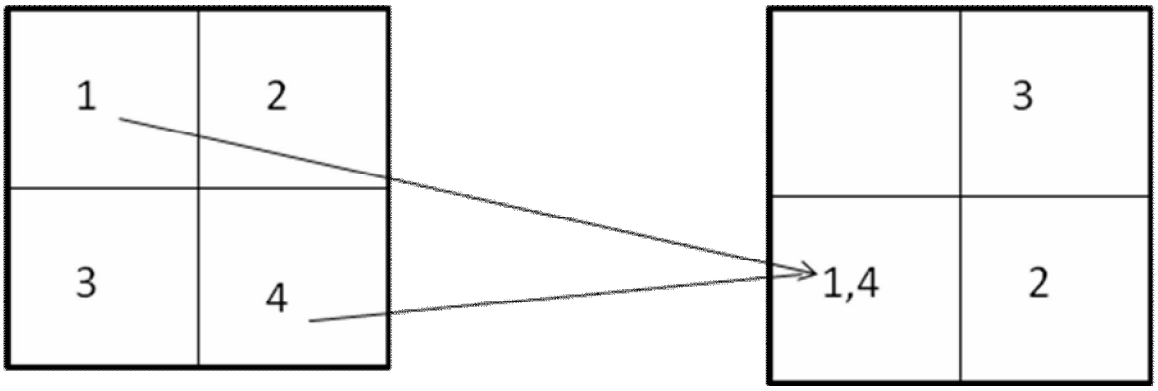


Figure 1: Cartesian Transformation

In the second, i.e., polar transform method, the same technique is applied but now the minutiae positions are measured in polar coordinates. The process of transformation consists of changing the sector position. But in polar coordinates the size of sectors can be different (sectors near the center are smaller than the ones far from the center). Restrictions are placed on the translation vector generated from the key so that the radial distance of the transformed sector is not very different from the original. Figure (2) explains the polar transformation.

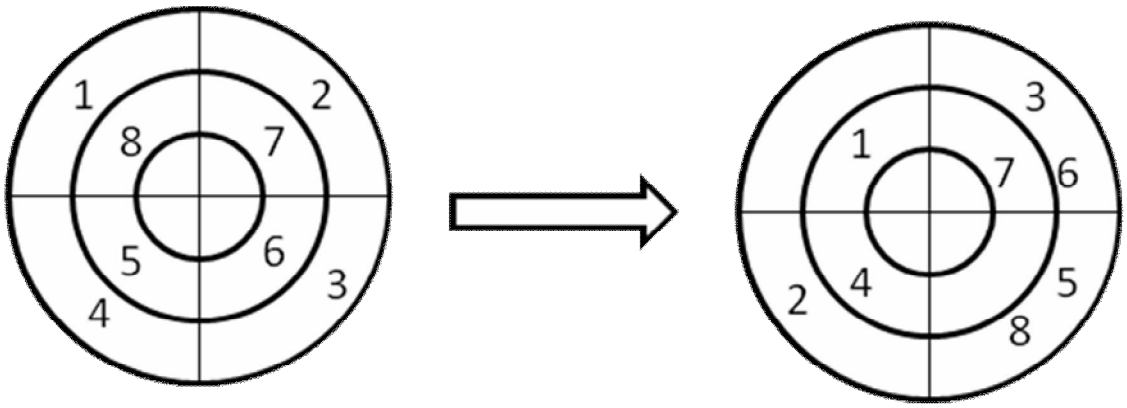


Figure 2: Polar Transformation

As there is 'many to one' mapping, it is impossible to tell which minutiae in the resulting block are from which original cell even if, both transformation and the transformed pattern are known. But the disadvantage with these two methods is that a small change in the minutia position in the original template can lead to a large change in the minutia position after transformation if the point crosses a sharp boundary. This can happen due to intra user variations i.e. variations occurring when the fingerprint of the same person taken at two different instances are different.

In the third method i.e surface folding, a smooth but non invertible functional transform is used to give high performance. Several constrains are put on the non invertible function. They are:

1. The transformation should be locally smooth but not globally smooth.
2. The transformation should be 'many to one' to make sure that it cannot be uniquely inverted to recover the original minutiae pattern.
3. Each minutiae position must be pushed outside the tolerance limit of the matcher after transformation.

In this method the minutiae positions are moved using two dimensional Gaussian functions. Each user is given a unique key which specifies the centers and the shapes of Gaussian kernels. These Gaussian

kernels are mixed to generate two functions $F(x, y)$ and $G(x, y)$. They are used to decide the direction and amount of shift for each minutia at (x, y). The direction of translation (phase) is represented as the gradient of the mixture and the extent of translation (magnitude) is represented as the scaled value of the mixture. The Gaussian mixture $F(z)$ is given as

$$\left| \vec{F}\left(z \right) \right| = \sum_{i=1}^{K} \frac{\pi_i}{\left| 2\pi\Lambda_i \right|} \exp\left\{ -\frac{1}{2}\left(z-\mu_i \right)^T \Lambda_i^{-1} \left(z-\mu_i \right) \right\} \tag{2}$$

$$\Phi_F\left(z \right) = \frac{1}{2} \arg\left\{ \nabla \vec{F} \right\} + \Phi_{rand} \tag{3}$$

Where $z = x + iy$ is the position vector K is a random key that defines the parameters of distribution such as the weights π_i, covariances Λ_i, the centers of kernels μ_i and the random phase offset Φ_{rand}. Another function $G\left(z \right)$ and its phase $\Phi_G\left(z \right)$ are defined in a similar way. Then a transformation $\left(x, y, \Theta \right) \rightarrow \left(X', Y', \Theta' \right)$ is given by

$$X' = x + K\left| \vec{G}\left(x, y \right) \right| + K\cos\left(\Phi_F\left(x, y \right) \right) \tag{4}$$

$$Y' = y + K\left| \vec{G}\left(x, y \right) \right| + K\sin\left(\Phi_F\left(x, y \right) \right) \tag{5}$$

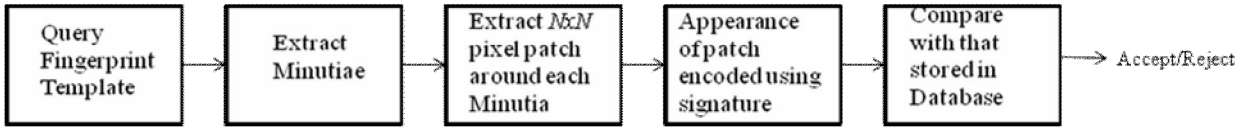
$$\Theta' = \mathrm{mod}\left(\Theta + \Phi_G\left(x, y \right) + \Phi_{rand}, 2\pi \right) \tag{6}$$

The Surface folding method is preferred over the other two methods due to their limitation in handling the intra user variation. The Surface folding method performs better than the Cartesian version and is comparable to the polar version.

4. REGISTRATION FREE GENERATION OF CANCELABLE FINGERPRINT TEMPLATE

Ratha et al [3] explained a registration free construction of cancelable fingerprint template. They have presented a new fingerprint representation based on localized, self aligned texture features. Most of the existing methods for generating cancelable fingerprint template need absolute registration process. But finding the optimal alignment is very difficult. Due to large number of possible translations, rotations and distortions, aligning fingerprint have high computational overhead. Although there are methods for getting accurate registration [10], [11], a small error in the process can lead to a faulty cancelable template leading to high 'false reject' during authentication. Also, absence of singular points can lead to failure. In this paper they have shown that absolute registration is not required and that purely local measurements are sufficient for this purpose. The process of enrollment and verification are shown in the figure (3).

(a)



(b)

Figure 3: (a) Enrollment Process (b) Verification Process

Enrollment

In the first stage, minutiae are extracted from the template. Then instead of storing the information regarding the minutiae, a $N \times N$ pixel patch around the minutia is extracted. The patch orientation is aligned with that of the minutia. This approach is based on the fact that each patch provides information about the unique identity of the individual. Common patches are non informative but patches with rare appearances have strong association with the identity of the person. The appearance (texture) of each patch is encoded using a compact signature. Each patch and its signature are stored in the database along with the identity of the person associated with the fingerprint.

Verification

During the verification process, minutiae are extracted from the query fingerprint. The $N \times N$ pixel patch around each minutia is encoded to generate a signature similar to the enrollment process. Then the set of signatures generated from the query fingerprint are compared with that stored in the database. The fact, that the distances are preserved under cancelable transformation, is used in this approach. Given two sets of minutiae signatures $\{x_1, x_2, ...\}$ and $\{y_1, y_2, ...\}$ and the distance between each match $D(x_i, y_j)$, the optimal minutiae correspondence is obtained by minimizing $\sum_i D(x_i, y_{T(i)})$, where $T(i)$ represents the index of the minutia in set $\{y_i\}$ that corresponds to x_i in the set $\{x_i\}$. Once the minutiae correspondence is established, the similarity measures across all matching minutiae signatures are aggregated to either accept or reject the query fingerprint.

Implementation Details

The implementation is done by representing the aligned patch compactly using a Gabor basis expansion. Similarity metric is derived from the normalized dot product distance metric d (). Some of the similarity measures described are: simple count, log weighting and inverse weighting. During verification, the reference set of signatures is compared with the query set of signatures. The evidences from each

matching pair are combined to generate the similarity measure for the fingerprint as a whole. The transform is made cancelable with the help of user specific projection matrix (B_k)

$$T(x,k) = B_k{}^T x \tag{7}$$

The distances will be preserved if $B_k B_k{}^T = I$. For this, the matrix B_k has to be orthogonal matrix, which can be synthesized from a random matrix by some orthogonal matrix decomposition method. The linear transformation $B_k{}^T x$ is invertible transformation. To make it non invertible, non-linearities are introduced in the transformation. A discretized projection is used as the patch signature, but this reduces the individuality of the transformed signature. Another technique, two factor key, where the transformation matrix B_k is split into two components can also be used to make the transform non invertible. This splitting can be achieved by SVD decomposition on a random matrix.

5. Discussion

In [3] the set of signatures generated from the query fingerprint are compared with that stored in the database. This comparison has two technical challenges: 1) How to measure similarity between signatures and 2) How to establish minutiae correspondence. As registration of image is done prior to transformation, the problem of minutiae correspondence does not occur in [1]. However, perfect registration itself is a big challenge.

In [1], all the three methods of transformation need absolute registration. Fingerprint registration as described earlier is a critical step in fingerprint matching. Accurate fingerprint registration is very difficult to achieve. Aligning two sets of minutiae needs a perfect transformation function. Achieving ideal transformation is almost impossible due to intra user variations. Although algorithms exist for accurate registration, any error in the process can lead to a 'false reject' during authentication. Absence of singular points can also lead to failure. Due to these limitations for getting accurate registration, in [3], [15] registration free method for generation of cancelable fingerprint templates is described. The method for generating cancelable template is free of any registration process as it is based on the information of neighboring local regions around minutiae.

In [1], in surface folding technique, although the process of aligning has high computational overhead, numbers of calculations during actual transformation are less compared to the calculations required in the patch based technique [3]. In patch based technique, two sets of minutiae 'signatures' being available, the distance measure from each match has to be calculated to find the optimal minutiae correspondence. The folding technique is a more compact representation making it suitable for memory limited applications.

In [1], in surface folding method, the transformation used is non invertible. But in [3] the patch based method, the proposed transformation is invertible. To make it non- invertible, non- linearities are added to the transformation.

In [1], the surface folding method is preferred over the other two. It performs noticeably better than Cartesian version and is comparable to the polar version. In [3], the localized patch based representation does not require registration and also provides a viable verification scheme. The patch based method is developed further to make the representation cancelable and it is also shown that it is resilient to adversarial attacks.

6. CONCLUSION

Two techniques for generating cancelable fingerprint templates are compared. The patch based technique is registration free while the surface folding technique needs absolute registration, so for fingerprints without singular points, it will fail. The surface folding technique has a non invertible transform while the patch based technique has to be made non invertible as the transform used is invertible. The surface folding technique is a compact way of representation and is suitable for memory limited applications.
Cancelable biometric provides a solution to address the privacy and security concerns about biometric authentication as it is computationally very difficult to recover the original template from the transformed version.

7. Acknowledgement

The authors would like to thank Dr. Sadashiv Bhide of Cummins College of Engineering for his valuable comments and suggestions.

8. REFERENCES

1. Nalini K. Ratha, Sharat Chikkerur, Jonathan Connell, and Ruud Bolle, “Generating cancelable fingerprint templates”, IEEE Trans. on PAMI, April 2008.
2. N.K. Ratha, J.H.Connell, and R.Bolle, “Enhancing Security and Privacy in Biometrics Based Authentication System,” IBM Systems J., vol. 40, no. 3, pp. 614-634, 2001.
3. Chikkerur, S.; Ratha, N.K.; Connell, J.H.; Bolle, R.M,” Generating Registration – free Cancelable Fingerprint Templates “, 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, 2008. BTAS 2008.Sept. 29 2008-Oct. 1 2008 pp :1 - 6
4. D. Maio, Maltoni, A.K.Jain and S. Prabhakar. “Handbook of Fingerprint Recognition”. Spinger Verlag 2003.
5. N.Ratha, J.Connell, R. Bolle, and S. Chikkerur,” Cancelable Biometrics: A Case Study in Fingerprints,”Proc. Int’l Conf. Pattern Recognition, 2006.
6. N.K.Ratha, K.Karu, S.Chen, and A.K Jain,” A Real Time Matching System for Large Fingerprint Database,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.18, no. 8, pp 799-813, Aug. 1996.
7. Andrew Teoh Beng Jin, David Ngo Chek Ling and Alwyn Goh, “Biohashing: Two factor authentication featuring fingerprint data and tokenized random number”. Pattern Recognition, vol. 37 no. 11, pp 2245-2255, 2004.
8. C. Soutar, D. Roberge; A. Stoianov; R. Gilroy and B.V.Kumar,” Biometric Encryption using Image Processing”, In SPIE, volume 3314, pages 174-188, 1998.
9. U. Uludag and A.K.Jain, “A fuzzy fingerprint vault. In Workshop: Biometrics: Challenges arising from theory to practice”, pages 13-16, 2004.
10. S. Chikkerur, S. Pankanti, N. K. Ratha, and V. Govindaraju. “Singular point detection in fingerprint images using linear phase portraits “. In AVBPA, 2005.
11. K. Nilsson and J. Bigun, “Localization of corresponding points in fingerprints by complex filtering”. Pattern Recognition Letters, 24, 2003.
12. Moon, Y.S. Yeung, H.W. Chan, K.C. Chan, S.O.,”Template synthesis and image mosaicking for fingerprint registration: An experimental study”, IEEE International Conference on ICASSP '04 Vol 5, pp 409-12.
13. Neil Yager and Adnan Amin,” Coarse Fingerprint Registration Using Orientation Fields”, EURASIP Journal on Applied Signal Processing, Volume 2005 (2005), Issue 13, Pages 2043-2053.
14. D.H. Ballard and C.M. Brown, “Computer Vision”. Englewood Cliffs, NJ: Prentice-Hall, 1982.
15. C. Lee, J. Y Choi, K. A Toh, S. Lee,” Alignment Free Cancelable Fingerprint Templates Based on Local Minutiae Information”, IEEE Trans. On System, Man, and Cybernetics – Part B: Cybernetics, vol. 37, no. 4, Aug 2007.

Verifying ODP Computational Behavioral Specification by using B-Method

Jalal Laassiri laassiri.jalal@gmail.com
Faculty of Science/Department of Mathematic And Informatics/ Laboratory of Mathematic and Informatics and Applications Mohamed V University -Agdal Rabat/ BP 1014/ Morocco

Saïd El Hajji elhajji@fsr.ac.ma
Faculty of Science/Department of Mathematic and Informatics/ Laboratory of Mathematic and Informatics and Applications Mohamed V University -Agdal Rabat/ BP 1014/ Morocco

Mohamed Bouhdadi bouhdadi@fsr.ac.ma
Faculty of Science/Department of Mathematic and Informatics/ Laboratory of Mathematic Pand Informatics and Applications Mohamed V University -Agdal Rabat/ BP 1014/ Morocco

Abstract

Reference Model for Open Distributed Processing (RM-ODP) defines a framework for the development of Open Distributed Processing (ODP) systems in terms of five viewpoints. Each viewpoint language defines concepts and rules for specifying ODP systems from the corresponding viewpoint. However the ODP viewpoint languages are abstract and do not show how these should be represented and specified. We treat in this paper the need of formal notation and specification for behavior al concepts in the Computational language. Using the Unified Modeling Language (UML)/OCL (Object Constraints Language) we define a formal semantics for a fragment of ODP behavior concepts defined in the RM-ODP foundations part and in the Computational language. We mainly focus on time, action, behavior constraints (sequentiality, non determinism and concurrency constraints), and policies (permission, obligation, prohibition). We also give a mapping of the considered concepts to Event-B. This will permit the verification of such specifications. Finally we explore the benefits provided by the new extension mechanisms of B-Method for verifying the ODP computational specifications.

Keywords: RM-ODP, Computational Language, computational specifications, Behavior Semantics, UML/OCL, B-Method.

1. INTRODUCTION

The Reference Model for Open Distributed Processing (RM-ODP) [1]-[4] provides a framework within which support of distribution, networking and portability can be integrated. It consists of

four parts. The foundations part [2] contains the definition of the concepts and analytical framework for normalized description of arbitrary distributed processing systems. These concepts are grouped in several categories which include structural and behavioral concepts. The architecture part [3] contains the specifications of the required characteristics that qualify distributed processing as open. It defines a framework comprising five viewpoints, five viewpoint languages, ODP functions and ODP transparencies. The five viewpoints are Computational, information, computational, engineering and technology.
Each viewpoint language defines concepts and rules for specifying ODP systems from the corresponding viewpoint. However, RM-ODP is a meta-norm [5] in the sense that it defines a standard for the definition of other ODP standards. The ODP standards include Modeling languages, specification languages and verification.
In this paper we treat the need of formal notation of ODP viewpoint languages. The languages Z [6], SDL, LOTOS, and Esterel are used in RM-ODP architectural semantics part [4] for the specification of ODP concepts. However, no formal method is likely to be suitable for specifying every aspect of an ODP system.
Elsewhere, there had been an amount of research for applying the Unified Modeling Languages UML as a notation for the definition of syntax of UML itself [7]-[9]. This is defined in terms of three views: the abstract syntax, well-formedness rules, and modeling elements semantics. The abstract syntax is expressed using a subset of UML static Modeling notations. The well-formedness rules are expressed in Object Constrains Language OCL [10]. A part of UML meta-model has a precise semantics [11],[12] defined using denotational meta-Modeling semantics approach. A denotational approach [13] is realized by a definition of the form of an instance of every language element and a set of rules which determine which instances are and are not denoted by a particular language element.
Furthermore, for testing ODP systems [2-3], the current testing techniques [14, 15] are not widely accepted and especially for the Computational viewpoint specifications. A new approach for testing, namely agile programming [16, 17] or test first approach [18] is being increasingly adopted. The principle is the integration of the system model and the testing model using UML meta-Modeling approach [19-20]. This approach is based on the executable UML [21]. In this context OCL can be used to specify the invariants [12] and the properties to be tested [17].
In this context we used the meta-Modeling syntax and semantics approaches in the context of ODP systems. We used the meta-Modeling approach to define syntax of a sub-language for the ODP QoS-aware Computational viewpoint specifications [5]. We also defined a UML/OCL meta-model semantics for structural concepts in ODP computational language [22]. In this paper we use the same approach for behavior al concepts in the foundations part and in the Computational language. We also show how the ODP considered concepts could be specified in the Event-B method.
The paper is organized as follows. In Section 2, we define a meta-model semantics of core behavior concepts (time, action, behavior, role, process). Section 3 defines a meta-model semantics for behavior concepts of RM-ODP foundations part namely, time, and behavior al constraints. We focus on sequentiality, non determinism and concurrency constraints. In Section 4 we introduce the behavior concepts defined in the Computational language. We give precise definitions for behavior al policies. In section 5 overview the correspondence of the main concepts with the B-Method method constructs. A conclusion and perspectives end the paper.

2. Meta-Modeling Core Behavior Concepts in RM-ODP Foundations Part

We consider the minimum set of modeling concepts necessary for behavior specification. There are a number of approaches for specifying the behavior of distributed systems and considering different aspects of behavior. We represent a concurrent system as a triple consisting of a set of states, a set of action and a set of behavior. Each behavior is modeled as a finite or infinite sequence of interchangeable states and actions [23]. To describe this sequence there are mainly two approaches [24].
1. “Modeling systems by describing their set of actions and their behaviors”.
2. “Modeling systems by describing their state spaces and their possible sequences of state changes”.

These views are dual in the sense that an action can be understood to define state changes, and state occurring in state sequences can be understood as abstract representations of actions [24]. We consider both of these approaches as abstraction of the more general approach based on RMODP. We provide the formal definition of this approach that expresses the duality of the two mentioned approaches.

We mainly use concepts taken from the clause8 “Basic Modeling concepts” of the RM-ODP part 2. These concepts are: behavior, action, time, constraints and state (see figure 1). the latter are essentially the first-order propositions about model elements. We define concepts (type, instance, pre-condition, post-condition) from the clause 9 “Specification concepts”. Specification concepts are the higher-order propositions applied to the first-order propositions about the model elements. Although basic Modeling concepts and generic specification concepts are defined by RM-ODP as two independent conceptual categories [25].

The behavior definition uses two RM-ODP modeling concepts: action and constraints (RM-ODP, part 2, clause 8.6):

Behavior (of an object): “A collection of actions with a set of constraints on when they may occur”.

Action: “something which happens”.

RM-ODP does not give the precise definition of behavioral constraints. These are part of the system behavior and are associated with actions. This can be formally defined as follows:

Context c: constraint inv: c.constrained_act -> size > 0

Context m: model behavior inv: m.behavior->includesAll(m.Actions->union(m.constraints))

For any element b from Behavior. ”if b is an Action and has at least one constraint , this constraint is a Behavior element.” Similarly when b is a Constraint and has at least one action, this action is a Behavior element.

Context b: behavior inv :m.behavior->forall(b |(m.actions->includes(m.b) and b.constraints->notempty) or(m.constraints->includes(m.b) and b.actions->notempty)

To formalize the definition, we have to consider two other modeling concepts: time and state. We can see how these concepts are related with the concept of action by looking at their definitions. Time is introduced in the following way (RM-ODP, part 2, clause 8.10):

Location in time: “An interval of arbitrary size in time at which action can occur.”

instant_begin: each action has one time point when it starts.

instant_end: each action has one time point when it finishes [26].

State (of an object) (RM-ODP, part 2, clause 8.7): At a given instant in time, the condition of an object that determines the set of all sequences of actions in which the object can take part. Hence, the concept of state is dual with the concept of action and these modeling concepts cannot be considered separately: This definition shows that state depends on time and is defined for an object for which it is specified.

Context t: time inv: b.actions->exists (t1,t2| t1 =action.instant_beging ->notempty and t2 =action.instant_end ->notempty and t1<> t2).

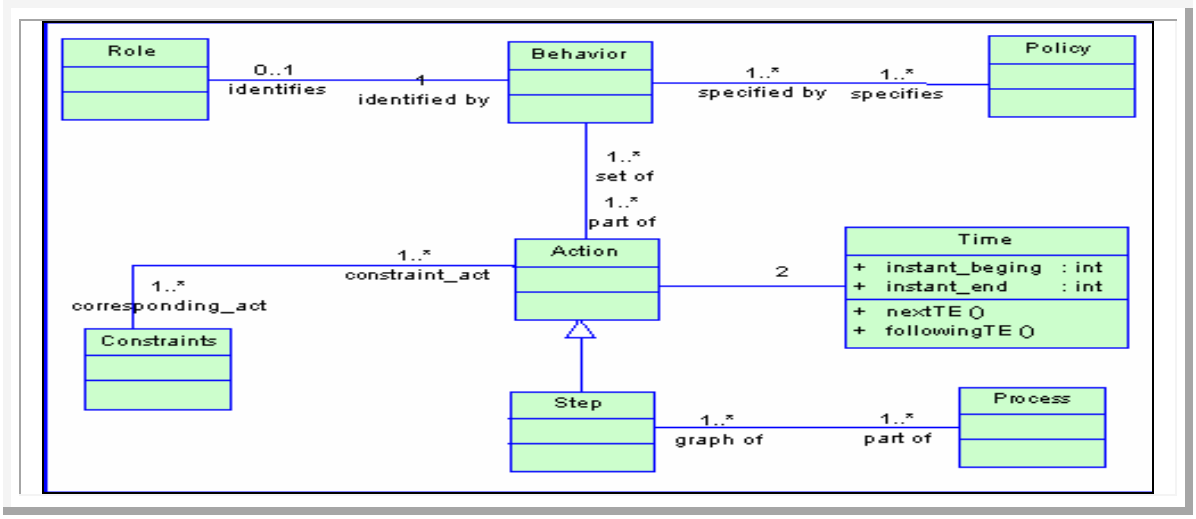


FIGURE 1: Core Behavior Concepts

3. Meta-Modeling Time and Behavioral Constraints

"Behavioral constraints may include sequentiality, non-determinism, concurrency, real time" (RM-ODP, part 2, clause 8.6). In this work we consider constraints of sequentiality, non-determinism and concurrency. The concept of constraints of sequentiality is related with the concept of time.

3.1 Time

Time has two following important roles in system design [26]:

•It serves for the purpose of synchronization of actions inside and between processes, the synchronization of a system with system users, the synchronization of user requirements with an actual performance of a system.

•It defines sequences of events (action sequences)

To fulfil the first goal, we have to be able to measure time intervals. However, a precise clock that can be used for time measurement does not exist in practice but only in theory [27]. So the measurement of the time is always approximate. In this case we should not choose the most precise clocks, but ones that explain the investigated phenomena in the best way. Simultaneity of two events or their sequentiality, equality of two durations should be defined in the way that the formulation of the physical laws is the easiest" [27]. For example, for the actions synchronization, internal computer clocks can be used and, for the synchronization of user requirements, common clocks can be used that measure time in seconds, minutes and hours.

We consider the second role of time. According to [27] we can build some special kind of clock that can be used for specifying sequences of actions. RM-ODP confirms this idea by saying that "a location in space or time is defined relative to some suitable coordinate system" (RM_ODP, part 2, clause 8.10). The time coordinate system defines a clock used for system Modeling. We define a time coordinate system as a set of time events. Each event can be used to specify the beginning or end of an action. A time coordinate system must have the following fundamental properties [26]:

•Time is always increasing. This means that time cannot have cycles.

•Time is always relative. Any time moment is defined in relation to other time moments (next, previous or not related). This corresponds to the partial order defined for the set of time events.

We use the UML (fig1) and OCL to define time: Time is defined as a set of time events.

nextTE: defines the closest following time events for any time events [26].

We use the followingTE relation to define the set of the following time events or transitive closure for the time event t over the nextTE relation:

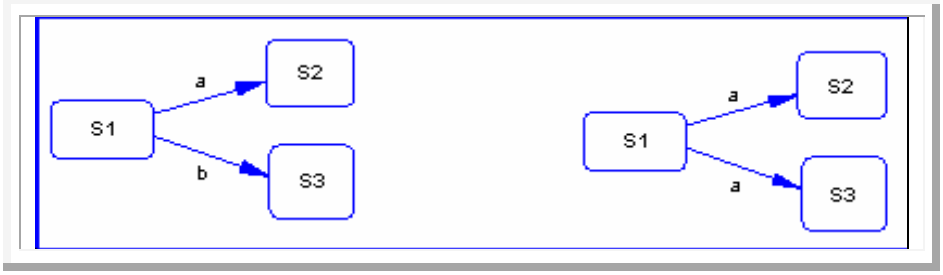
followingTE: defines all possible following time events Using followingTE we can define the following invariant that defines the transitive closure and guarantees that time event sequences do not have loops :

Context t: time Inv: Time->forAll(t:Time | (t.nextTE->isempty implies t.follwingTE->isempty)

and (t.nextTE->notempty and t.follwingTE->isempty implies t.follwingTE =t.nextTE) and (t.nextTE->notempty and t.follwingTE->notempty implies t.follwingTE-> includes(t.nextTE.follwingTE->union(t.nextTE)) and t.follwingTE->exludes(t)).
This definition of time is used in the next section to define sequential constraints.

3.2 Behavioral constraints

We define the behavior like a finite state automaton (FSA). For example, figure 2 shows a specification that has constraints of sequentiality and non determinism. The system is specified using constraints of non-determinism since state S1 has a non-deterministic choice between two actions a and b.
Based on RM-ODP, the definition of behavior must link a set of actions with the corresponding constraints. In the following we give definition of constraints of sequentiality, of concurrency and of non-determinism.



(a) (b)

FIGURE 2: a - Sequential deterministic constraints;
b - Sequential non deterministic constraints.

3.2.1 Constraints of sequentiality

Each constraint of sequentiality should have the following properties [26]:
•It is defined between two or more actions.
•Sequentiality has to guarantee that one action is finished before the next one starts. Since RM-ODP uses the notion of time intervals it means that we have to guarantee that one time interval follows the other one:
Context sc: constraintseq inv:
Behavior.actions-> forAll(a1,a2 | a1<> a2 and a1.constraints->includes(sc) and a2.constraints->includes(sc)and((a1.instant_end.followingTE->includes(a2.instant_begin)
or(a2.instant_end.followingTE->includes(a1.instant_begin))
For all SeqConstraints sc, there are two different actions a1, a2, sc is defined between a1 and a2 and a1 is before a2 or a2 is before a1.

3.2.2 Constraints of concurrency

Figure 3 shows a system specification that has constraints of concurrency since state a1 has a simultaneous choice of two actions a2 and a3.

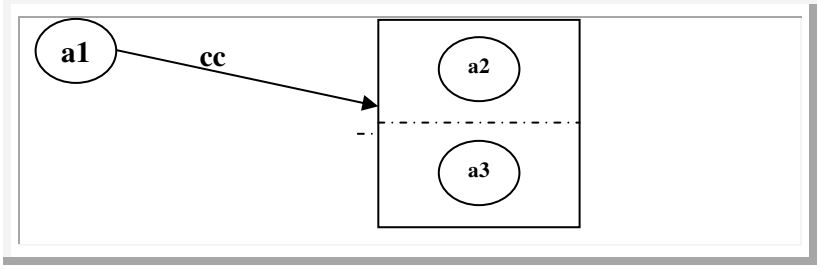


FIGURE 3: RM-ODP diagram: Example constraints of concurrency

For all concuConstraints cc there is a action a1, there are two different internal actions a2, a3, cc is defined between a1 and a2 and a3, a1 is before a2 and a1 is before a3
Context cc: constraintconc inv:

Behavior.actions-> forAll(a1 :Action ,a2 ,a3 : internalaction | (a1 <> a2) and (a2 <> a3) and (a3 <> a1) and a1.constraints->includes(cc) and a2.constraints->includes(cc) and a3.constraints->includes(cc) and a1.instant_end.followingTE-> includes(a2.instant_begin) and a1.instant_end.followingTE-> includes(a3.instant_begin))

3.2.3 Constraints of non-determinism

In order to define constraints of non-determinism we consider the following definition given in [24]: "A system is called non-deterministic if it is likely to have shown number of different behavior, where the choice of the behavior cannot be influenced by its environment". This means that constraints of non-determinism should be defined between a minimum of three actions. The first action should precede the two following actions and these actions should be internal (see figure 4).

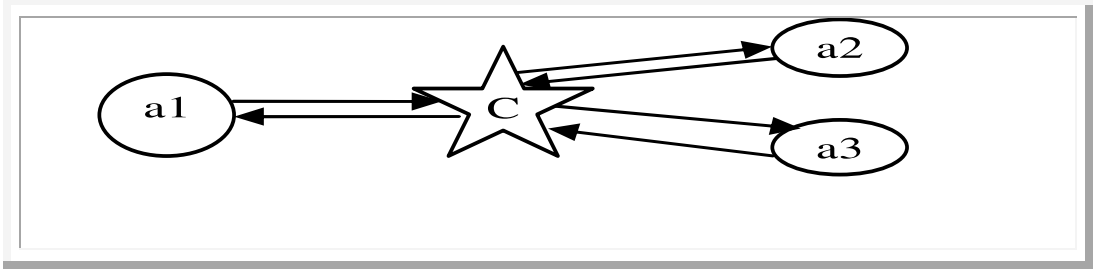


FIGURE 4: Example Constraints example of non-determinism

We define this constraint as follows:
Context ndc: NonDetermConstraints inv:Behavior.actions-> forAll(a1 :Action ,a2 ,a3 : internalaction | (a1 <> a2) and
(a2 <> a3) and (a3 <> a1) and a1.constraints->includes(ndc) and
a2.constraints->includes(ndc) and a3.constraints->includes(ndc) and a1.instant_end.followingTE-> includes(a2.instant_begin) or a1.instant_end.followingTE-> includes(a3.instant_begin)) .
We note that, since the choice of the behavior should not be influenced by environment, actions a2 and a3 have to be internal actions (not interactions). Otherwise the choice between actions would be the choice of environment [26].

4. Modeling Behavior constraints Specifications in Event-B

In this last section, we treat the question of verifying ODP specifications. For this we begin by defining how to use the formal method B-Method to specify the RM-ODP concepts. Event-B is a simplification as well as an extension of de B formalism [31] which has been used in number of large industrial projects. The objective of this formal method is use the refinement calculus to define and prove in the step by step fashion so that the system in question will be correct by construction. This will be very adequate in our context since each specification is a refinement of another. This will be done by using the propositional language, the predicate language, the set-theoretic language, and arithmetic language ,such they presents some mathematical justifications to proof obligation rules used in this approach.
In the previous section we specified the behavior constraints (Sequentiality, non-determinism, concurrent), here we presents how we can develop these concepts by using the Event-B and the tools of the open source RodinPlatform.
This section introduces a Event-B concepts which supports Modeling with a set of semantic constructs that correspond to those in behavior concepts, defined in enterprise language (see table 1).

Behavior Concepts	***Event-B Construct***
Behavior	***Machine***

State	***State static (constant with axioms) or State dynamic(variable with invariants)***
Action	***Event with guards(necessary conditions for event to occur)***
Constraint	***Invariants + guards***

Table 1: T Sample table

We develop the initial model of the sequential constraint by both essentials construct of Event-B: machine and context.

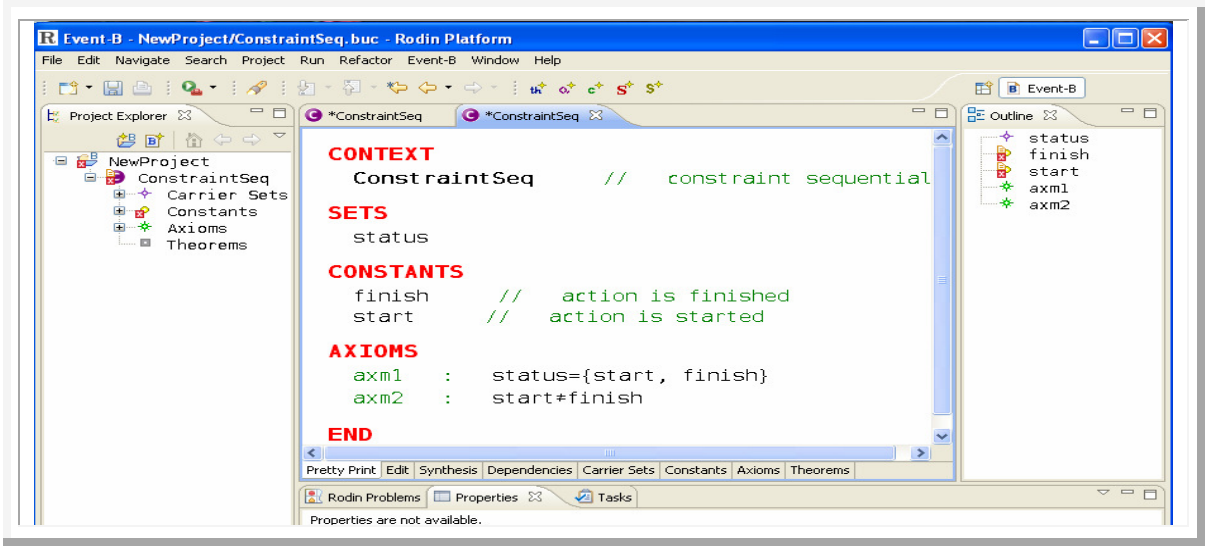


FIGURE 5: A context of sequential constraint

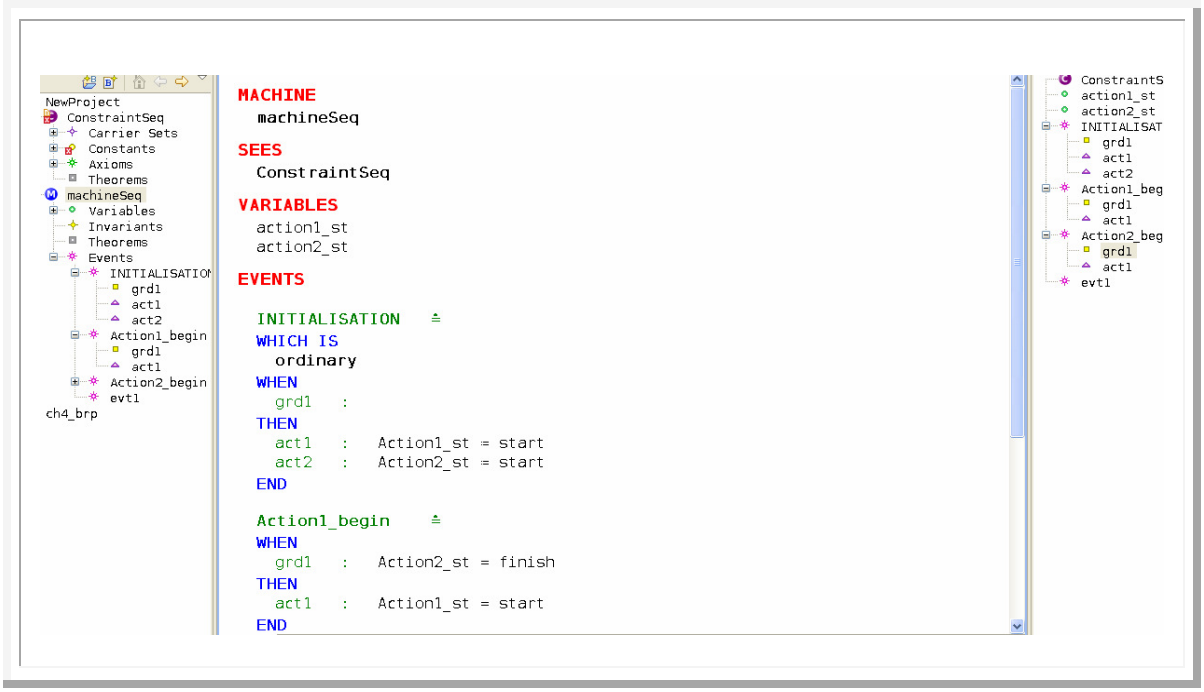


FIGURE 6: A machine of sequential constraint

5. CONSLUSION & FUTURE WORK

We address in this paper the need of formal ODP viewpoint languages. Using the meta-modeling semantics, we define a UML/OCL based semantics for a fragment of behavior concepts defined in the foundations part (time, sequentiality, non determinism and concurrency) and in the Computational viewpoint language (behavioral policies). These concepts are suitable for describing and constraining the behavior of open distributed processing Computational specifications.

The initial model of sequential constraint is developed by using Event-B, Each model will be analyzed and proved to be correct. The next step is the refinement of this model. We are applying the same approach for other ODP Computational behavior concepts (real time).

6. REFERENCES

1. ISO/IEC, ''Basic Reference Model of Open Distributed Processing-Part1: Overview and Guide to Use, ''ISO/IEC CD 10746-1, 1994
2. ISO/IEC, ''RM-ODP-Part2: Descriptive Model, '' ISO/IEC DIS 10746-2, 1994.
3. ISO/IEC, ''RM-ODP-Part3: Prescriptive Model, '' ISO/IEC DIS 10746-3, 1994.
4. ISO/IEC, ''RM-ODP-Part4: Architectural Semantics, '' ISO/IEC DIS 10746-4, July 1994.
5. M. Bouhdadi et al., ''A UML-Based Meta-language for the QoS-aware Enterprise Specification of Open Distributed Systems'' IFIP Series, Vol 85, Springer, 255-264 (2002).
6. Abhishek Dixit and al. "Applying UML and Z to Extended Basic Interoperability Data Model", International Journal of computer science and security (IJCSS), June 2007.
7. B. Rumpe, ''A Note on Semantics with an Emphasis on UML, '' Second ECOOP Workshop on Precise Behavioral Semantics, LNCS 1543, Springer, 167-188 (1998).
8. A. Evans et al., ''Making UML precise, '' Object Oriented Programming, Systems languages and Applications, (OOPSLA'98), Vancouver, Canada, ACM Press (1998)
9. A. Evans et al. The UML as a Formal Modeling Notation, '' UML, LNCS 1618, Springer, 349-274 (1999)
10. J. Warmer and A. Kleppe, the Object Constraint Language: Precise Modeling with UML, Addison Wesley, (1998).

11. S. Kent, and al. ‘’A meta-model semantics for structural constraints in UML,, In H. Kilov, B. Rumpe, and I. Simmonds, editors, Behavioral specifications for businesses and systems, Kluwer , (1999). chapter 9
12. E. Evans and al., Meta-Modeling Semantics of UML, In H. Kilov, B. Rumpe, and I. Simmonds, eds, Behavioral specifications for businesses and systems, Kluwer , (1999). ch. 4.
13. D.A. Schmidt, ‘’Denotational semantics: A Methodology for Language Development, ‘’ Allyn and Bacon, Massachusetts, (1986)
14. G. Myers, ‘’The art of Software Testing, ‘’, John Wiley &Sons, (1979)
15. Binder, R. ‘’ Testing Object Oriented Systems. Models. Patterns, and Tools, ‘’ Addison-Wesley, (1999)
16. A. Cockburn, ‘’Agile Software Development. ‘’Addison-Wesley, (2002).
17. B. Rumpe, ‘’ Agile Modeling with UML, ‘’ LNCS vol. 2941, Springer, 297-309 (2004).
18. Beck K. Column on Test-First Approach. IEEE Software, Vol. 18, No. 5, 87-89 (2001)
19. L. Briand, ‘’A UML-based Approach to System testing, ‘’ LNCS Vol. 2185. Springer, 194-208 (2001).
20. B. Rumpe, ‘’ Model-Based Testing of Object-Oriented Systems; ‘’ LNCS Vol.. 2852, Springer; 380-402 (2003).
21. B. Rumpe, Executable Modeling UML. A Vision or a Nightmare?, In: Issues and Trends of Information technology management in Contemporary Associations, Seattle, Idea Group, London, 697-701 (2002).
22. M. Bouhdadi, Y. Balouki, E. Chabbar. ‘’ Meta-Modeling Syntax and Semantics of Structural Concepts for Open Networked Enterprises”, ICCSA 2007, Kuala Lumpor, 26-29 August, LNCS 4707, Springer, 45-54 (2007)
23. Lamport, L. and N.A. Lynch, Distributed Computing: Models and Methods, in Handbook of Theoretical Computer Science, Volume B: Formal Models and Semantics. 1990, Elsevier and MIT Press.
24. Broy, M., “Formal treatment of concurrency and time,’’ in Software Engineer's Reference Book,J. McDermid, Editor, Oxford: Butterworth-Heinemann pp 23, (1991).
25. Wegmann, A. and al. ‘’ Conceptual Modeling of Complex Systems Using RMODP Based Ontology‘’ . in 5th IEEE International Enterprise Distributed Object Computing Conference - EDOC (2001). September 4-7 USA. IEEE Computer Society pp. 200-211
26. P. Balabko, A. Wegmann, “From RM-ODP to the formal behavior representation” Proceedings of Tenth OOPSLA Workshop on Behavioral Semantics ¨Back to Basics¨, Tampa, Florida, USA , pp. 11-23 (2001).
27. Henri Poincaré, The value of science, Moscow «Science», 1983
28. Harel, D. and E. Gery, “Executable object modeling with statecharts“, IEEE Computer.30(7) pp. 31-42 (1997)
29. Jean-Raymond Abrial: A System Development Process with Event-B and the Rodin Platform. ICFEM (2007) 1-3.
30. A.R.M Nordin and al. Managing Software Change Request Process: Temporal Data Approach,. International Journal of Computer Science and Security, (IJCSS) Volume (3):January 01, 2009.

A Novel Image Steganography Method With Adaptive Number of Least Significant Bits Modification Based on Private Stego-Keys

Yogendra Kumar Jain ykjain_p@yahoo.co.in
Head of the Department
Computer Science & Engineering Department
Samrat Ashok Technological Institute
Vidisha (M.P) 464001 India

R. R. Ahirwal ram2004_ahirwal2004@rediffmail.com
Computer Science & Engineering Department
Samrat Ashok Technological Institute,
Vidisha (M.P) 464001 India

Abstract

To enhance the embedding capacity of image steganography and provide an imperceptible stego-image for human vision, a novel adaptive number of least significant bits substitution method with private stego-key based on gray-level ranges are proposed in this paper. The new technique embeds binary bit stream in 24-bits color image (Blue channel) or in 8-bits gray-scale image. The method also verifies that whether the attacker has tried to modify the secret hidden (or stego-image also) information in the stego-image. The technique embeds the hidden information in the spatial domain of the cover image and uses simple (Ex-OR operation based) digital signature using 140-bit key to verify the integrity from the stego-image. Besides, the embedded confidential information can be extracted from stego-images without the assistance of original images. The proposed method can embed 4.20 bits in each pixel of gray-scale image and 4.15 bits in each pixel of color image. The presented method gives better results than the existing methods.

Key-words: Steganography, stego-key, data hiding, digital image, PSNR (Peak-Signal-to-Noise-Ratio).

1. INTRODUCTION

The emergent possibilities of modern communication need the exceptional way of security, especially on computer network. The network security is becoming more important as the number of data being exchanged on the internet increases. Therefore, the confidentiality and data integrity are essential to protect against unauthorized access. This has resulted in an explosive growth of the field of information hiding. Moreover, the information hiding technique could be used extensively on applications of military, commercials, anti-criminal, and so on [1]. To protect secret message from being stolen during transmission, there are two ways to solve this problem in general. One way is encryption, which refers to the process of encoding secret information in such a way that only the right person with a right key can decode and recover the original information successfully. Another way is steganography, steganography literally means covered writing. Its goal is to hide the fact that communication is taking place. In the field of steganography some terminology has been developed. The term cover is used to describe the original, innocent message, data, audio, still video, and so on. If the cover media is a digital image hidden with secret data, this image is called stego-image. Steganography hides the secret message with the host data set and its presence is imperceptible [2]. PCs facilitated sending and exchanging photographs, greeting cards, birthday cards, etc. in a manner that thousand of these are exchanged on the internet on the daily basis. It is not only economical, but users can choose cards from a vast Varity of them freely available and takes no time taken to send to them.

Additionally audio and video files are also exchanged freely. This exchange of cards and files has further given strength to steganography.

Watermarking, another way of data hiding aims at different purposes from steganography. Copyright protection and authentication is on primary target of image watermarking and it is required that embedded information can be prevented, resisted, or altered up to some degrees of distortion while the watermarked image is attacked or damaged. Because of this requirement, robustness becomes the main benchmark emphasized by the image watermarking techniques. Unlike watermarking, capacity, security and invisibility are the benchmarks needed for data hiding techniques of steganography.

2. TYPES AND MEDIA

Steganography may be classified as pure, symmetric, and asymmetric. While pure steganography does not need any exchange of information, symmetric and asymmetric need to exchange of keys prior to sending the messages [3]. Symmetric steganography is employed in our proposed method in which stego-key is exchanged. Steganography is highly dependent on type of medium being used to hide the information. Medium being commonly used include, text, images, audio-files, and network protocols used in network communication [4].

Image steganography is generally more preferred media because of its harmlessness and attraction. Image steganography may classify according to working domain: (a) Spatial domain and, (b) Frequency domain. Spatial domain steganography work on the pixel value directly and modify the pixel gray-value [5]. In Frequency domain based methods [6], images are first transformed into the frequency domain and then message are embedded in the transform coefficients.

A digital image is an array of numbers that represent light intensities of various points [7]. The light intensities or pixels are combines to form the images raster data. The images can be grayscale (8-bits) or color (24-bits). Although larger size image file facilitate larger amount of data to be hidden but transferring require more bandwidths and therefore increases the cost. Two types of file compression generally used to overcome above said problem are lossy compression and lossless compression. JPEG (Joint photographic group) is an example of lossy compression. Its advantage is that it saves more space but in doing so loses its originality. On the other hand GIF, PNG and BMP are examples of lossless compression which is in general recommended media types. Since both of these retain their originality [8]. Our algorithm is simple and flexible using LSBs technique. We have selected the formats that commonly use lossless compression that is BMP, PNG, TIF and GIF. We can make use of any of these formats or convert BMP into any of the above said format.

3. REVIEW OF RELATED WORK

The usage of a stego-key is important, because the security of a protection system should not be based on the secrecy of the algorithms itself, instead of the choice of a secret key [9] as shown in Fig. 1. The steganographer's job is to make the secretly hidden information difficult to detect given the complete knowledge of the algorithm used to embed the information except the secret embedding key. This so called kerckhoff's principle is the golden rule of cryptography and is often accepted for steganography as well [10]. Some steganographic methods [11] [12] uses a stego-key to embed message for achieving rudimentary security. Mehboob et. al. proposed technique uses predictive position agreed between two parties as stego-key [3]. Same position used only once to enhance security. But drawback of the algorithm is small amount of data to be embedded.

The most common and simplest steganographic method [13] [14] is the least significant bit insertion method. It embeds message in the least significant bit. For increasing the embedding capacity two or more bits in each pixel can be used to embed message. At the same time not only the risk of making the embedded statistically detectable increase but also the image fidelity degrades. So how to decide the number of bits of each pixel used to embed message becomes an important issue of image steganography.

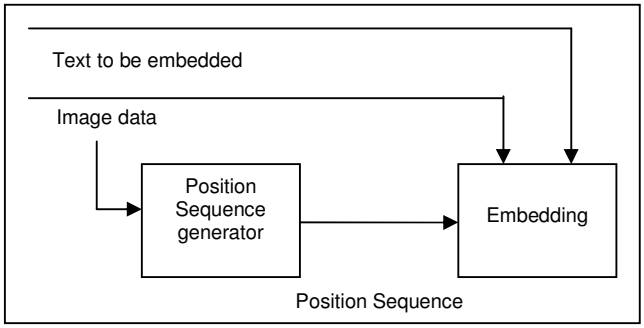


Fig. 1: Generalized Stego-key system

Cheeldod et al. [15] proposed an adaptive steganography that select the specific region of interest (ROI) in the cover image. Where safely embeds data. The choice of these regions based on human skin tone color detection. Adaptive steganography are not an easy target for attacks especially when the hidden message is small [16]. The tri-way pixel value differencing method proposed by ko-chin-chang can successfully provide embedding capacity and outstanding imperceptibility for the stego-images.

Suresh Babu et al. [17] Proposed steganographic model authentication of secret information in image steganography, that can be used to verify the integrity of the secret message from the stego-image. In this method payload is transformed from spatial domain to discrete wavelet transform. The DWT coefficients are then permuted with the verification code and then embedded in the special domain of the cover image. The verification code is generated using to special coefficient in the DWT domain. Thus the method can verify each row has been modified or forget by attacker.

Moon and Kavitkar [18] proposed a fixed 4LSB method to embedding an acceptable amount of data; 4LSB embedding data can easily be implemented and do not visually degrade the image to the point of being noticeable. But drawback of the scheme is that the encoded message can be easily recovered and even altered by 3rd party. So techniques must be developed to solve above said problems. Lie et al. [19] proposed an adaptive method based on using variable amount of bits substitution instead of fixed length for adjusting the hiding capacity.

Adnan Gutub, et al. proposed a steganography technique for RGB color images [20]. They proposed an image-based steganography technique called triple-A algorithm. The algorithm adds more randomization by using two different seeds generated from a user-chosen key in order to select the component (s) used to hide the secret bits as well as the number of the bits used inside the RGB image component. This randomization adds more security especially if an active encryption technique is used such as AES. While Enayatifar et al. proposed a method, in which two chaotic signals for specifying the location of the different parts of the message in the picture [21]. An 80-bit key was used to reach the preliminary measures of the two chaotic signals and this caused a kind of scattering format for the data embedding place in the image, as they are randomly selected. But one can easily find the place and order of the data embedding by knowing the chaotic function and the key values (two 5 bit keys). It is noticeable that a minor change in the key values (primary values of the keys) will bring about a drastic change in the produced values of the chaotic functions.

4. PROPOSED METHODOLOGY

The proposed scheme works on the spatial domain of the cover image and employed an adaptive number of least significant bits substitution in pixels. Variable K-bits insertion into least significant part of the pixel gray value is dependent on the private stego-key K_1. Private stego-key consists of five gray-level ranges that are selected randomly in the range 0-255. The selected key shows the five ranges of gray levels and each range substitute different fixed number of bits into least significant part of the 8-bit gray value of the pixels (in gray image and in color image blue channel). After making a decision of bits insertion into different ranges, Pixel p(x, y) gray value “g” that fall within the range Ai-Bi is changed by embedding k-message bits of secret information into

new gray value “g’ ”. This new gray value “g’ ”of the pixel may go beyond the range Ai-Bi that makes problem to extract the correct information at the receiver. Specific gray value adjustment method is used that make the new gray value “g’ ” fall within the range Ai-Bi. Confidentiality is provided by the private stego-key k_1 and to provide integrity of the embedded secret information, 140-bit another key K_2 is used. Digital signature of the secret information with the key K_2 were obtained and appended with the information. The whole message plus signature is embedded into the cover image that provides some bit overheads but used to verify the integrity. At the receiver key K_1 is used to extract the message and key K_2 is used to verify the integrity of the message.

4.1 Private stego-key generation

Private stego-key K_1 play an important role in proposed scheme to provide security and deciding the adaptive K bits insertion into selected pixel. For a gray scale image (or RGB color image blue channel) 8-bit used to represent intensity of pixel, so there are only 256 different gray values any pixel may hold. Different pixels in image may hold different gray values. We may divide the pixels of images into different groups based on gray ranges. Based on this assumption let five ranges of gray levels are $< A_1\text{-}B_1, A_2\text{-}B_2, A_3\text{-}B_3, A_4\text{-}B_4, A_5\text{-}B_5 >$ each range starting and ending value are in 8-bits, total 80-bits are used to make a key K_1. If the difference of each range is denoted by Di=Bi-Ai (for i=1, 2, 3, 4, 5; Ai denote starting value and Bi denote ending value of the range), it should not be less than 32 gray values and any range should not be overlap with other ranges. For Example selected key K_1: 2-36, 38-73, 74-102, 105-170, and 178-245. Difference $D_2=B_2\text{-}A_2$ will be $D_2 = 73\text{-}38=35 \geq 32$, and any range is not overlapping. Hence key is usable.

4.2 Method to decide Bits insertion in each range

Let the five gray ranges decided by the stego-key are $<A_1\text{-}B_1, A_2\text{-}B_2, A_3\text{-}B_3, A_4\text{-}B_4, A_5\text{-}B_5 >$ and number of pixel count from cover image in each range are $< N_1, N_2, N_3, N_4, N_5 >$. Range with maximum pixel count will hold maximum bits insertion let five bits, second maximum count will hold four bits insertion and so on. In this way we decide the fixed number of bits insertion into each range and adaptive number of bits insertion into different ranges based on pixel count of cover image in different ranges. In similar way we decide the bits extraction from each range. For Example assume key K_1 is 2-36, 38-73, 74-102, 105-170, 178-245 and let pixel count in each range from any image are 300,100,34,4000,700. Then range first insert three message bits in the pixel that comes within the range, range second insert two message bits in the pixel ,range third insert one bit in the pixel ,range four insert five bits in the pixel and range five insert four message bits in the pixel that comes in this range. In this manner we decide the bits insertion into each range.

4.3 LSB substitution

Least significant substitution is an attractive and simple method to embed secret information into the cover media and available several versions of it. We employ in propose scheme adaptive LSB substitution method in which adaptive K-bits of secret message are substituted into least significant part of pixel value. Fig.2 shows entire method for K-bits insertion.

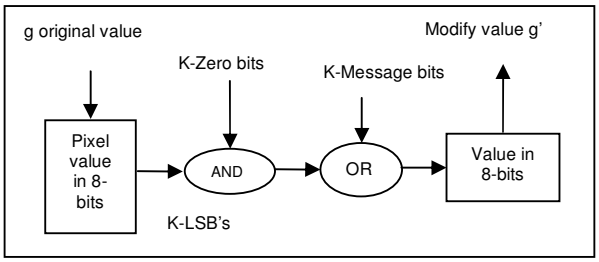


Fig 2: Method for K-bits insertion

To decide arbitrary k-bits insertion into pixel, first we find the range of pixel value and then find the number of bits insertion decided by method given in section IV (b) and insert K-message bits into least significant part of pixel using LSB. After embedding the message bits the changed gray value g’ of pixel may go beyond the range. To make value within the range, reason is that receiver side required to count pixels to extract message, pixel value adjusting method is applied to make changed value within range.

4.4 Pixel value adjusting method

After embedding the K-message bits into the pixel gray value g new gray vale g' may go outside the range. For example let our range based on key is 0-32. Let the gray value g of the pixel is 00100000 in binary forms (32 in Decimal), decided K-bits insertion is 3-bits are 111. The pixel new gray value g' will be 00100111 in binary forms after inserting three bits (39 in Decimal). Modified value is outside the range. To make within the range 0-32, K+1 bits of g' is changed from 0 to 1 or via- versa. And checked again to fall within range if not K+2 bit is changed and so on until gray value fall within range. For example, 00100111- 00101111- 00111111- 00011111. Figure 3 shows the whole process.

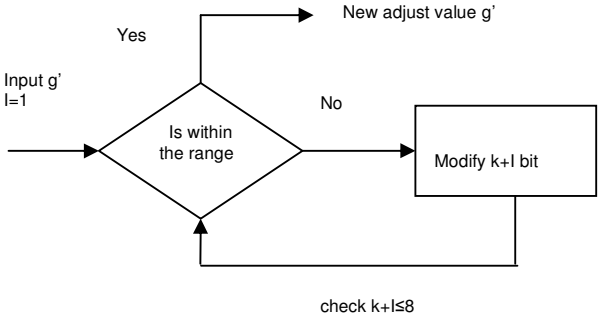


Fig. 3: Pixel value adjusting method

4.5 Digital signature

To verify the integrity of the stego-image and secret information, a simple Ex-OR method to find signature of secret message with random stego-key of 140 bits is used and appended with the message, some overheads occurs but integrity of the message is checked at the receiver. Block Diagram of whole process is given in Fig. 4 (a) and 4 (b). Algorithm for coding and decoding the secret information is given below.

Algorithms: Coding

Input: Cover-image, secret message, keys K_1, K_2.
Output: Stego-image.
Step1: Read key K_1 based on gray-Level ranges.
Step2: Read cover image (8-bit gray Image or 8-bit color image blue Channel)
Step3: Decide No. of bits insertion into each range describe in section IV (b).
Step4: Read the secret message and Convert it into bit stream form.
Step5: Read the key K_2.
Step6: Find the signature using K_2 and append with the message bits.
Step7: For each Pixel
7.1: Find gray value g.
7.2: Decide the K-bits insertion based on gray ranges.
7.3: Find K-message bits and insert using method given in section IV(c).
7.4: Decide and adjust new gray Value g' using method described in sec. IV (d)
7.5: Go to step 7.
Step 8: end

Algorithm: Decoding

Input: Stego-image, keys K_1, K_2;
Output: Secret information;
Step1: Read key K_1 based on gray-level ranges.
Step2: Read the stego image.
Step3: Decide No. of bits extraction into each range. Describe in section IV (b).
Step4: For each pixel, extract the K-bits and save into file.
Step5: Read the key K_2 and find the signature of bit stream
Step6: Match the signature.
Step7: End

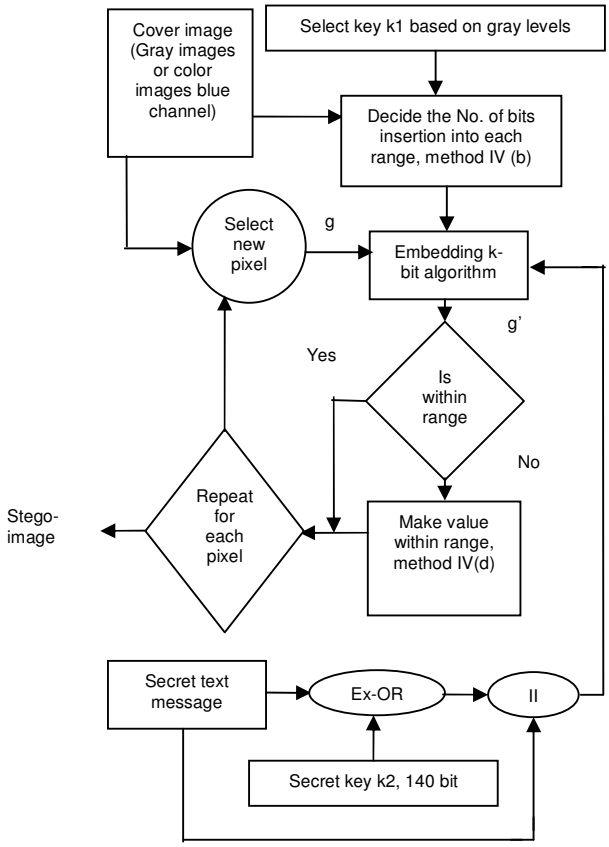


Fig. 4 (a): Message Embedding with signature

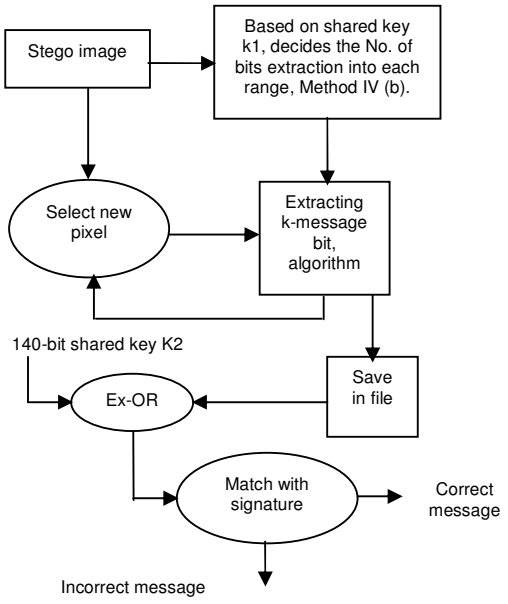


Fig. 4 (b): Message extraction and Integrity check

5. RESULTS AND DISCUSSIONS

To demonstrate the accomplished performance of our proposed approach in capacity and imperceptibility for hiding secret data in the cover-image, we have conducted different experiments using different images to compare the proposed approach with fixed 4 LSB method [18] and the method given in [19]. According to invisibility benchmark PSNR 30dB is acceptable. Results are considered for each image (gray image and color image) size 150x150 with 100% capacity using different stego-keys (five ranges in each key).

The well known Peak-Signal-to-Noise Ratio (PSNR) is used as performance measurement criteria, which is classified under the difference image distortion metrics, is applied on the Stego and the Original images. It is defined as [22]:

$$PSNR = 10 \log_{10} \left(\frac{C_{max}^{2}}{MSE} \right) \qquad \text{---------(1)}$$

Where, C_{max} holds the maximum value in the original images and MSE denotes Mean Square Error and given as:

$$MSE = \frac{1}{MN} \sum_{m=1} \sum_{n=1} (S_{xy} - C_{xy})^2 \quad \text{(For Grayscale Images)} \qquad \text{-------(2)}$$

Where, x and y are the image coordinates, M and N are the dimensions of the image, S_{xy} is the generated Stego image and C_{xy} is the cover image

$$MSE = \left[\frac{MSE(R) + MSE(G) + MSE(B)}{3} \right] \quad \text{(For Color RGB Images)} \qquad \text{---------(3)}$$

As a performance measurement for embedding capacity, the average number of bits embedded into each pixel is calculated as:

$$Capacity = \left(\frac{\text{Total Number of bits embedded into image}}{\text{Total Number of Pixels in image}} \right) \text{ (bits/pixel)} \qquad \text{------(4)}$$

The embedding capacity and PSNR results of proposed method for the different grayscale and color images are shown in Table-1, Table-2. Table-1 shows the results when the message is embedded into gray scale images and Table-2 shows the result when the message is embedded into the blue channel of the RGB color images using different key.

Different keys (Using five ranges)	**Grayscale Images (8-bit)**							
	Cameraman		**Shadow**		**Baboon**		**Pout**	
	CAP	**PSNR**	**CAP**	**PSNR**	**CAP**	**PSNR**	**CAP**	**PSNR**
0-33, 34-70, 71-105, 106-170, 171-255	4.11	34.4979	4.1373	32.7129	4.16	37.8410	4.5031	32.8608
2-35, 37-73, 74-105, 106-170, 171-255	4.1178	33.9581	4.1260	32.5237	4.1469	37.9279	4.5030	31.5815
2-35, 37-73, 74-115, 116-170, 171-250	3.9836	34.1044	3.9698	33.4759	4.15	37.6350	4.5850	31.4671
0-45, 47-85, 86-143, 144-190, 191-255	4.1211	32.2653	4.0009	34.0415	4.1248	38.6471	4.6652	31.4089
0-45, 47-85, 86-143, 144-188, 189-255	4.1077	32.6375	4.0076	34.0831	4.1305	38.5906	4.6650	31.4585
Average Values	4.0881	33.4926	4.0483	33.3674	4.1424	38.1283	4.5842	31.7553

Table 1: Results in terms of Embedding Capacity and Image Quality (In PSNR) using different keys for different grayscale images [(CAP- Embedding Capacity in bits/pixel), (PSNR-Peak-Signal-to-Noise-Ratio)].

Different keys	**Color Images (24 bit)**

(Using five ranges)	Lena		Onion		Football		Baboon	
	CAP	PSNR	CAP	PSNR	CAP	PSNR	CAP	PSNR
0-33, 34-70, 71-105, 106-170, 171-255	4.1568	36.9145	4.2208	37.0962	4.0685	45.9343	4.0685	45.9343
2-35, 37-73, 74-105, 106-170, 171-255	4.1284	36.8187	3.8610	37.4604	4.0361	45.9434	4.0361	45.9434
2-35, 37-73, 74-115, 116-170, 171-250	4.1968	36.8365	3.8574	37.4886	4.0157	45.9260	4.0157	45.9260
0-45, 47-85, 86-143, 144-190, 191-255	4.3933	37.4985	4.3999	37.2020	4.1693	44.7775	4.1693	44.7775
0-45, 47-85, 86-143, 144-188, 189-255	4.3902	37.5066	4.3992	37.1964	4.1695	44.5496	4.1695	44.5496
Average Values	4.3131	37.1149	4.1476	37.2887	4.0918	45.4261	4.0918	45.4261

Table 2: Results in terms of Embedding Capacity and Image Quality (In PSNR) using different key for different color images [(CAP-Embedding Capacity in bits/pixel), (PSNR-Peak-Signal-to-Noise-Ratio)].

Table-3 shows the comparison of results in terms of Embedding Capacity (in bits/pixel) and Image Quality (PSNR in dB) of Proposed Method with 4LSB method and Adaptive Method. The 4LSB Method [18] can embeds upto 4 bits/pixel for gray-scale and color images, while Adaptive Method [19] can embeds upto 4.025 bits/pixel for gray-scale and color images. On the average case, our proposed method can embed 4.20 bits in each pixel of gray-scale image and 4.15 bits in each pixel in blue channel of color image. Hence, the embedding capacity is better than the existing 4LSB Method and Adaptive Method. Also, the image quality attained in proposed method is better than the existing methods.

Images	Embedding methods					
	4LSB Method		Adaptive Method		Proposed Method	
	CAP	PSNR	Average CAP	PSNR	Average CAP	PSNR
Gray-Scale images	4	31.71	4.025	32.57	4.20	34.18
Color images using Blue Channel	4	--	4.025	--	4.15	40.99

Table 3: The Comparative Results in terms of Embedding Capacity and Image Quality (In PSNR) of Proposed Method with 4LSB method [18] and Adaptive Method [19] [(BC-Blue channel of color image), (CAP-Capacity in bits/pixel)].

The Comparison of existing methods with Proposed Method in terms of in term of embedding capacity, image quality for grayscale and color images are shown in figure 5. Therefore, it is clearly seen from the experimental results that the performance of proposed method is better than the existing methods. In addition to that, the advantage of proposed method is that employment of stego key in embedding process provides better security.

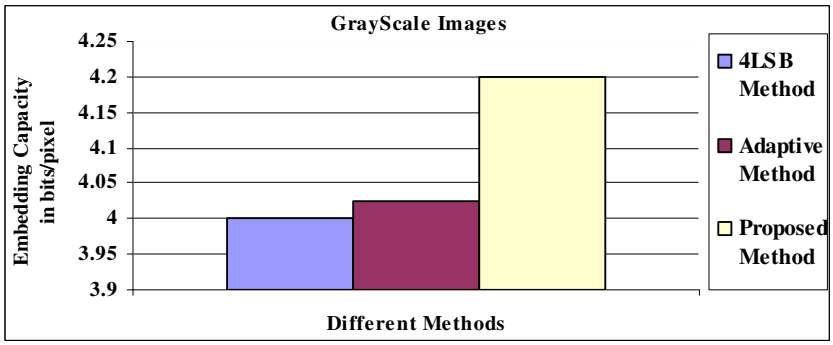


Fig. 5 (a): Comparison of different methods with Proposed Method in term of embedding capacity

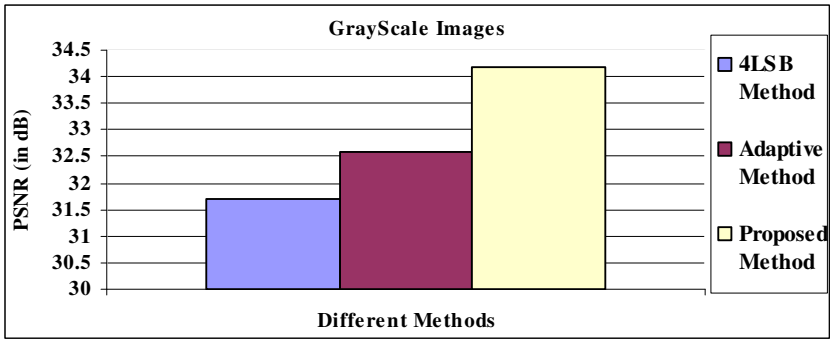


Fig. 5(b): Comparison of different methods with Proposed Method in term of image quality

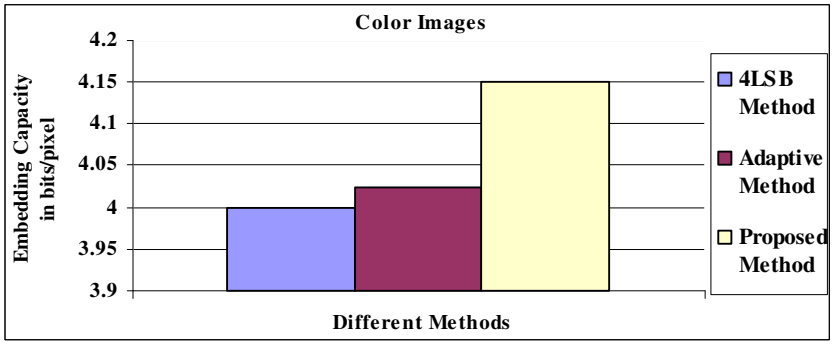


Fig. 5 (c): Comparison of different methods with Proposed Method in term of embedding capacity

6. CONCLUSION

We have introduced a novel image steganographic model with high-capacity embedding/extracting module that is based on the Variable-Size LSB substitution. In the embedding part based on stego-key selected from the gray value range 0-255. We used the pixel value adjusting method to minimize the embedding error and adaptive 1-5 bits to embed in the pixel to maximize average capacity per pixel. Using the proposed method, we embedded at least four message bits in each pixel while maintaining the imperceptibility. For the security requirement we have presented two different ways to deal with the issue. The major benefit of supporting these two ways is that the sender can use different stego-keys in different sessions to Increase difficultly of stegano analysis on these stego images. Using only the stego-keys, which is used to count the number of pixel in each range and second 140-bit key to verify the integrity of the message, the receiver can extract the embedded messages exactly. Experimental results verify that the proposed model is effective and efficient.

7. REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, M. G. Kuhn, *"Information Hiding - A Survey"*, Proceeding of the IEEE, vol. 87, issue 7, pp. 1062-1078, July 1999.

[2] S. Dumitrescu, W. X. Wu and N. Memon, *"On steganalysis of random LSB embedding in continuous-tone images"*, Proceeding of International conference on image Processing, Rochester, NY, pp. 641-644, 2002.

[3] B. Mehboob and R. A. Faruqui, *"A steganography Implementation"*, IEEE – International symposium on biometrics & security technologies, ISBAST'08, Islamabad, April 2008.

[4] K. Ahsan and D. Kundur, *"Practical data hiding in TCP/IP"*, Proceeding of the workshop on multimedia security at ACM multimedia, 2002.

[5] A. Westfeld, *"F5- A steganographic algorithm: High capacity Despite Better Steganalysis"*, Proceeding of 4th Int. Information Hiding Workshop, Springer-Verlag, vol. 2137, 2001.

[6] I. Cox, J. Kilian, T. Leighton and T. Shamoon, *"Secure spread spectrum watermarking for multimedia"*, IEEE Transaction on Image processing, vol. 6, issue 12, pp. 1673-1687, 1997.

[7] Neil F. Johnson, and Sushil Jajodia, *"Exploring Steganography: Seeing the Unseen"*, IEEE computer society press vol. 31, issue 2, pp 26-34, Feb. 1998.

[8] D. E. Denning, E. Dorothy, *"Information Warfare and Security",* Boston, MA: ACM Press, pp. 310-313, 1999.

[9] Jian Zhao, E. Koch, *"Embedding Robust Lables into Images for Copyright Protection"*, Proceeding of the international Conference on Intellectual property Right for specialized information, Knowledge and New Technologies, Vienna, August 1995.

[10] Jiri Fridrich. "*A New Steganographic Method for Palette-Based Images",* Center for Intelligent Systems, SUNY Binghamton, Binghamton, NY 13902-6000, U.S Government, a grant number F30602-98-C-0009.

[11] F. A. P. Petitcolas, R. J. Anderson, *"On the Limit of Steganography"*, IEEE J. Sel. Areas Communication, vol. 16, issue 4, pp. 474-481, 1998.

[12] M. Kutter, E. Jordan and E. Bossin, *"Digital signature of Color images using amplitude modulation"*, Journal of Electronics imaging, vol. 7, issue 2, pp. 326-332, 1998.

[13] E.T. Lin, E.J. Delp, *"A review of data hiding in images "*, Proceedings of the conference on image processing image quality image capture systems, PICS'99, pp. 274-278, April 1999.

[14] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, *"Techniques for data Hiding"*, IBM, syst. J., vol. 35, issue 3&4, pp. 313-336, 1996.

[15] A. Cheddad, J. Condell, K. Curran and P. McKevitt, *"Enhancing Steganography in digital images"*, IEEE - 2008 Canadian conference on computer and Robot vision, pp. 326-332, 2008

[16] Ko-Chin Chang, Chien-Ping Chang, Ping S. Huang, and Te-ming Tu, *"A novel image steganographic method using Tri-way pixel value Differencing"*, Journal of multimedia, vol. 3, issue 2, June 2008.

[17] K. S. Babu, K. B. Raja, K. Kiran Kumar, T. H. Manjula Devi, K. R. Venugopal, L. M. Pataki, *"Authentication of secret information in image steganography"*, IEEE Region 10 Conference, TENCON-2008, pp. 1-6, Nov. 2008.

[18] S. K. Moon and R.S. Kawitkar, *"Data Security using Data Hiding"*, IEEE International conference on computational intelligence and multimedia applications, vol. 4, pp. 247-251, Dec. 2007.

[19] W. N. Lie and L. C. Chang, *"Data Hiding in images with adaptive numbers of least significant bits based on human visual system"*, IEEE international conference on image processing, vol. 1, pp. 286-290, 1999.

[20] Adnan Gutub, Ayed Al-Qahtani, Abdulaziz Tabakh, *"Triple - A: Secure RGB Image Steganography Based on Randomization"*, IEEE/ACM international conference on computer systems and applications, pp. 400 - 403, 2009.

[21] R. Enayatifar, S. Faridnia, H. Sadeghi, *"Using the Chaotic Map in Image Steganography"*, IEEE international conference on signal processing systems, pp. 754 - 757, 2009.

[22] J. Zollner, H. Federrath, H. Klimant, et al., *"Modeling the Security of Steganographic Systems"*, in 2nd Workshop on Information Hiding, Portland, pp. 345-355, April 1998.

A Multi-Operator Based Simulated Annealing Approach For Robot Navigation in Uncertain Environments

Hui Miao Hui.Miao@microchip.com
Microchip Australia Design Centre
Microchip Technology Inc.
Brisbane, 4108, Australia

Abstract

Optimization methods such as simulated annealing (SA) and genetic algorithm (GA) are used for solving optimization problems. However, the computational processing time is crucial for the real-time applications such as mobile robots. A multi-operator based SA approach incorporating with additional four mathematical operators that can find the optimal path for robots in dynamic environments is proposed in this paper. It requires less computation times while giving better trade-offs among simplicity, far-field accuracy, and computational cost. The contributions of the work include the implementing of the simulated annealing algorithm for robot path planning in dynamic environments, and the enhanced new path planner for improving the efficiency of the path planning algorithm. The simulation results are compared with the previous published classic SA approach and the GA approach. The multi-operator based SA (MSA) approach is demonstrated through case studies not only to be effective in obtaining the optimal solution but also to be more efficient in both off-line and on-line processing for robot dynamic path planning.

Keywords: Optimization, MSA, SA, GA, Dynamic Environments

1. INTRODUCTION

Mobile robots have been widely used in many industrial areas such as aerospace systems, nuclear applications, and mining equipment. How to find an absolute safe path in a dangerous environment for a mobile robot is one of the most important aspects in robot navigation. The main goal of the robot path planning is to search a safe path for a mobile robot, to make the robot move from the start point to the destination point without collision with obstacles. Also, the path is often required to be optimal in order to reduce processing times, communication delay, and energy consumption. According to [18], existing methods for robot path planning could be classified in different ways. Depending on the environment where the robot is located in, they can be categorised into the following two types:

1) Path planning in a static environment with static obstacles in the map; and
2) Path planning in a dynamic environment with both static and dynamic obstacles in the map.

Each of these two types could be further divided into two sub-groups depending on how much the robot knows about the entire information of the surrounding environment:

- Path planning in a clearly known environment, in which the robot already knows the location of the obstacles before it starts to move. Because the environment is fully known, the path for the robot could be the globally optimised result.
- Path planning in a partially known or uncertain environment, in which the robot probes the environment using sensors to acquire the information about the location, shape, and size of the obstacles, and then uses the information for local path planning.

This paper proposes a multi-operator simulated annealing (MSA) approach incorporating with multiple mathematical operators for robot path planning in dynamic environments. In this work, the MSA approach will be shown that the approach gives much improved performance than existing approaches for dynamic path planning results that have been presented in ICARCV'2008 conference [18].

2. RELATED WORKS AND MOTIVATIONS

2.1 Related Works on Dynamic Path Planning

Given the entire information of the environment in which a robot is, the globally optimal or near-optimal path could be found by using optimisation algorithms, e.g., the GA [1], [3], [20] and Fuzzy logic [2]. The A* algorithm [4] is developed to help a robot to find the optimal path in grid decomposed static maps. The A* algorithm uses the heuristic based Dijkstra algorithm to obtain the optimal result for the robot. The drawback of this method is the use of a uniformed grids representation, which demands allocation of large amounts of memory, even for those regions that may never be traversed or may not contain any obstacles, implying that the efficiency of the method may be low.

Evolved from the improved A* shortest-path algorithm, an improved algorithm is developed by Hu and Gu [5] to solve the problem of optimum route planning in vehicle navigation systems. It is based on the standard GA and the lambda-interchange local search method. However, in many practical applications such as those described in [6] and [7], it is difficult for a robot to get the full information of the surrounding environment at any one time because the status and the movement of the obstacles in the environment change all the time in the map. A one time global path planning made by the robot may become an infeasible solution due to the changes in the environment.

For dynamic environments with moving obstacles, limited work has been reported on optimal path planning for mobile robots. Chakravorty and Junkins [8] have introduced a methodology for intelligent path planning in uncertain environments with vision-like sensors. Recently, Wang, Sillitoe and Mulvaney [9] have presented a GA planner to determine optimal or near-optimal path solutions for mobile robots in dynamic environments. The GA based approach is shown to be a promising tool for the path planning problem of mobile robots in dynamic environments with moving obstacles.

Zhang, L¨ u, and Song [10] have developed an artificial potential field algorithm for dynamic path planning for soccer robots in dynamic environments where both the target and obstacles are moving. The D* method [11], the dynamic A*, is a typical method for path planning in dynamic and unknown environments. It plans optimal traverses in real-time by incrementally repairing the paths to the robot's state when new environment information becomes known to the robot, making it possible to reduce the computational cost significantly. When the robot gathers new information about the environment, it re-plans new paths based on the new information.

To further enhance the performance of the D* algorithm, improvements have been made to the D* algorithm. Representative works include the framed-quadatree D* method [12], the field D* mehtod [13], and others such as [14], [15]. The framed-quadatree D* method uses the quadatree structure to represent the dynamic environments. In order to minimise the search space, different

dimensions of grids are used in the quadatree structure and border cells are added for connecting the grids. The field D* method [13] employs an interpolation-based planning and re-planning algorithm to generate smooth paths through non-uniform cost grids. It uses linear interpolation during the planning to calculate accurate path cost estimates for arbitrary positions within each grid cell and to produce paths with a continuous range of headings. It can produce a smooth optimal path for a robot to overcome the sub-optimal problem appearing in other non-uniformed-girds methods. Willms and Yang [16] proposed a dynamic system for real-time robot path planning. Recently, they further developed a grid-based algorithm for real-time robot path planning via a distance- propagating dynamic system [17].

2.2 **Motivations of This Work**

SA based method [18] is published previously proving that SA method can offer better performance on both path length and processing time than the GA method [9]. The performance of the previous proposed SA [18] still deteriorates significantly as the problem size increases. We believe that the performance of the SA approach in [18] can be further improved because of the operator that generates new solutions is relatively simple in [18]. Only two operators have been used, which switching or deleting some bits of the result to generate a new solution. This means that the possibility of jumping out of the local minimum is small. Therefore, more mathematical operators (switching, deleting, mutating and repairing) are implemented with the existing SA approach to improve the efficiency of the SA approach.

In this work, an SA approach incorporating multiple mathematical operators is developed for robot path planning in dynamic environments. It will be shown that the approach gives much improved performance than existing approaches for dynamic path planning. Some preliminary results of this work have been presented in ICARCV'2008 conference [18].

3. Multi-Operator Simulated Annealing Approach

3.1 Dynamic Environments

As in previous work [18]. The obstacles in the environments are represented by bounded polygons. Thus, the movement and trajectory of a dynamic obstacle are constituted by a series of polygons with their positions being updated along with the time. The vertices of the obstacles form the search space of our path planning algorithm. Following assumptions are made in this work: The movement and trajectory of moving obstacles in the environment are unknown to the robot; The motion parameters, such as speed and direction, of the dynamic obstacles can be sensed by the robot if the obstacle is in the range of the robot sensors; The robot could change its moving direction at any time when necessary; As in [9], all obstacles in the map are enlarged by a fixed value so that the robot could approach the obstacles without collision; and the dimension of the robot is neglected, and consequently the robot is regarded as a single point.

FIGURE 1 shows a dynamic environment, where the black polygons represent the static obstacles and the hollow polygons are moving obstacles. All the obstacles are enlarged by some values through creating additional margins. The vertices of the enlarged polygons are the search space for robot path planning.

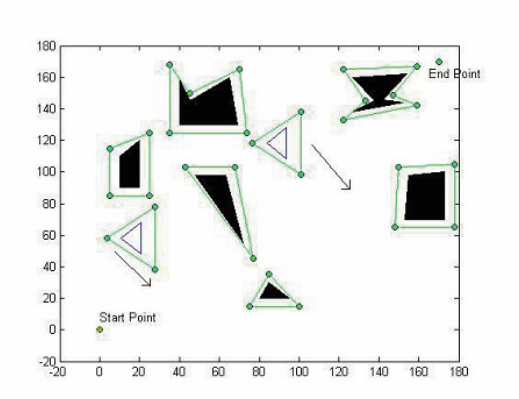


FIGURE 1: A Dynamic Environment with Static and Moving Obstacles

3.2 Architecture of the Multi-Operator SA Approach

The SA algorithm begins with the off-line path computation, in which the locations of the vertices of the static obstacles are fully known to the robot. Once the off-line computation is complete, the robot can start to travel through the stationary obstacles. When a moving obstacle enters the detection range of the robot sensors, the obstacle together with its moving speed and direction will be detected. Then, the robot calculates the possibility of clashing of the robot with the moving obstacle. The calculation result determines what the robot does next. If the moving obstacle will not hit the robot, the robot will keep travelling using the current path. Otherwise, if the robot will likely collide with the obstacle with the current movement, the SA algorithm will be triggered to find an alternative path from the current location of the robot to the destination.

As shown in FIGURE 2, a mathematical model is established for calculating the possibility of clashing of the with a moving obstacle. It is seen from FIGURE. 2 that the first crossing point of the current robot path and the predicted trajectory of the moving obstacle can be calculated. Then, the time t required for the robot to travel from its current location to the first cross point can be derived; and the location and consequently the corresponding exclusion area of the moving obstacle can also be estimated after the obstacle moves forward for the same amount of time t. If the robot path segment from the first crossing point to either of the two directions of the path crosses the edges of the moving obstacle odd times, then a collision will likely occur between the robot and the moving obstacle; otherwise, a collision will unlikely happen.

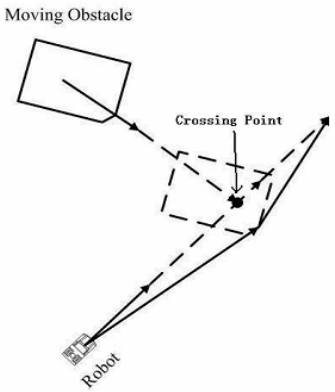


FIGURE 2: Calculation of the collision possibility (dotted lines: the original trajectories of the robot and dynamic obstacle; solid line: alternative robot path to avoid collision with the obstacle).

3.3 MSA Architecture

As in [18], a feasible path solution is expressed by a series of vertices linking the start point through to the end point. Each vertex of the obstacles has its series number; and thus a path is

represented by a sequence of vertex numbers. Therefore, a feasible path solution X is described as:

$$X = \{V_{start}, V_{start+1}, V_{start+2}, \cdots, V_{end-1}, V_{end}\} \quad (1)$$

Where Vi means the i^{th} vertex.

Traditionally, the length of the path, E_f, is used as a criterion to quantify the quality of the path solution derived from a path planning algorithm: the shorter the path, the better the solution. The evaluation function E_f is given by:

$$E_f = \sum_{i=start}^{i=end-1} D(V_i, V_{i+1}) \quad (2)$$

Where $D(V_i, V_{i+1})$ is the direct distance from V_i to V_{i+1}.

FIGURE 3 shows the top-level algorithm structure and pseudo-code of the multi-operator SA for dynamic path planning.

MSA Algorithm:

$T = T_{initial}$;
while $(T > T_{terminate})$
    Randomly generate a feasible solution X_s;
    Evaluate X_s, $E_f = f(X_s)$;
    $count = 1$;
    while $(count < Threshold)$
        Generate a new solution X_n base on X_s;
        Evaluate X_n; $E_n = f(X_n)$;
        If $f(X_n) < f(X_s)$
            $X_s = X_n$;
        Elseif $rand(1) < \exp\left(\frac{f(X_s)-f(X_n)}{T}\right)$
            $X_s = X_n$;
            $count = count + 1$;
        End-if
    End-while
    $T = cool\_rate * T$;
    Update X_s at each reduction of temperature T
End-while
X_s is the optimal path solution.
Return

FIGURE 3: Pseudo-code of the MSA Algorithm

3.4 Random Multi-Operator Path Planner

The MSA procedure is demonstrated in FIGURE 3. Different from the simple path planners that were previously used in [18], more complicated random path planners are developed in this work. Deleting, switching, mutation and repairing operators are used in the planner; and the planner randomly chooses one operator to generate a new path Xn from the initial path Xs. FIGURE 4 shows that how the operators randomly generate a new path Xn from the initial path Xs.

Similar to the procedure for initial path selection, the feasibility of each path segment generated by the operators is tested. This is to ensure that the path segment does not intersect with any edges in the map. When the length of the path is chosen as the evaluation criterion, randomly deleting vertices could help improve the performance of the path solution. Therefore, the possibility of choosing the deleting operator is set to be higher than other operator. As will be

seen later in case studies, the possibility of choosing the deleting operator is set to be 0.70 in this work.

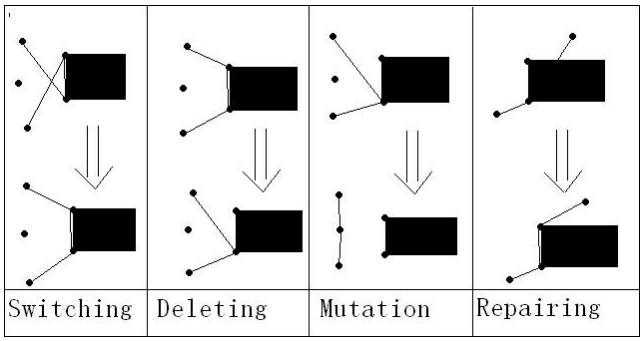


FIGURE 4: Multi-Operator Planner

After a new path solution is generated by using the operators, it is evaluated using the evaluation criterion, i.e., the length of the path. It is accepted if it is better than the previous one. It may also be accepted in a certain probability defined by the current temperature even if it is not better than the previous one.

3.5 Online Path Planning

While a robot uses the route generated by off-line planning to travel through static obstacles, the on-line path planner is triggered automatically to calculate an alternative path when a dynamic obstacle is detected. As no particular brand or configuration of the sensors is specified, the sensing range of the robot sensors is set to be a fixed value. If the distance between the robot and every vertex of the moving obstacle is shorter than the fixed value, the moving obstacle enters the sensing range of the robot and thus can be detected by the robot sensors.

It has assumed that robot sensors can monitor the shape and trajectory of a moving obstacle as well as the moving speed and direction. After acquiring the moving information of the moving obstacle, the robot could infer the possibility of collision with the moving obstacle. When it is inferred that the robot will collide with the moving obstacle, the SA based dynamic path planning algorithm will be triggered for finding an alternative path for the robot to travel from its current location to the destination. The search space, the current status of the robot, and location and moving information of the moving obstacle will be updated to enable the dynamic path planning.

4. Experiment Results

4.1 Simulation Environments and Parameters

Our case studies are carried in Matlab [19] under Windows XP on a computer with 2.8GHz Pentium Core 2 Duo CPU and 2GB memory. Four dynamic environments are designed to test the performance of the dynamic path planning approaches. They contain both static and dynamic obstacles, as shown in Table 1. For each of these four environments, the number of the vertices of the static obstacles is also tabulated in TABLE 1; it is used for offline path planning before the robot starts to travel.

Environment	Static Obstacles	Dynamic Obstacles	No. of Static Vertices
1	3	2	10
2	6	2	25
3	9	4	53
4	14	6	82

TABLE 1: Four Environments

The dynamic obstacles in all the environments have random shapes. The first two environments simulate simple scenarios where the dynamic obstacles appear simultaneously and simply move forward in the same direction. With more static and dynamic obstacles, the last two environments are more complicated scenarios where the dynamic objects each appears at a random time and moves either forward or backward. Each environment was tested in fifty times. The termination conditions of the approaches are tabulated in TABLE 2.

Initial Temperature	Termination Temperature	Cooling Rate	Deleting Operator Rate	Other Operator Rate
9999	5555	0.97	70%	30%

TABLE 2: Control Parameters for MSA

4.2 Simulation Results for Case One

Environment One contains two dynamic obstacles which will appear simultaneously as well as three static obstacles with ten vertices. It is depicted in FIGURE 5, where the black and fully filled blocks represent static obstacles; the hollow triangles show the trajectories of the dynamic obstacles. The sequence of the points in the figure is the travel trajectory of the robot from the start to the end points. The arrows indicate the moving directions of the dynamic obstacles.

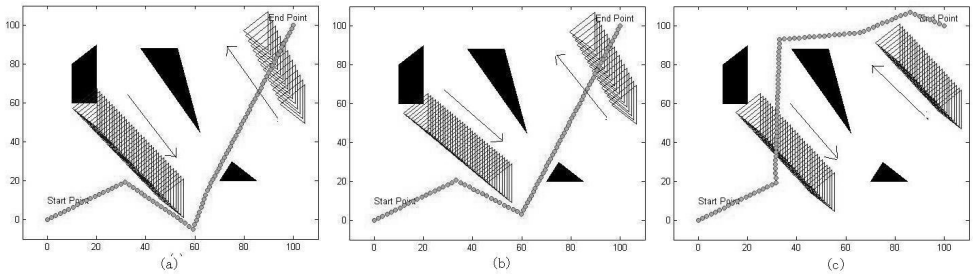


FIGURE 5: Environment One (left to right: MSA, SA, GA)

All four approaches can re-plan the path successfully when moving obstacles are detected and thus no collision has occurred in FIGURE 5. TABLE 3 lists the simulation results of off-line processing time, on-line processing time, and path length. The results show that all approaches have similar path length. For offline and online processing times, MSA approach have the minimum processing time, the normal SA approach outperforms the GA method.

4.3 Simulation Results for Case Two

Compared with Environment One, Environment Two also contains two dynamic obstacles which will appear simultaneously. The total number of the vertices of the static obstacles is 25, compared to 10 in Environment One. FIGURE 6 shows Environment Two and its simulation results for the MSA, normal SA, GA approaches.

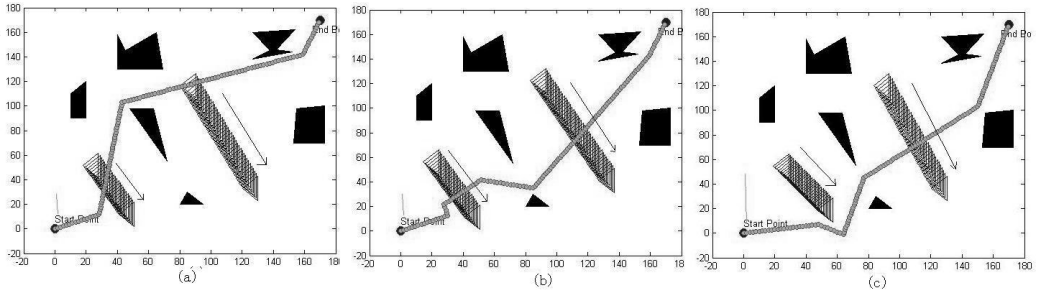


FIGURE 6: Environment Two (left to right: MSA, SA, GA)

TABLE 3 gives some quantitative performance results of the four approaches in off-line processing time, on-line processing times, and path length for Environment Two. It is seen from

this table that the performance of the path length can be considered to be comparable for all the approaches. However, for both off-line and on-line processing times, MSA has the best performance, the normal SA is superior to the GA method.

4.4 Simulation Results for Case Three

Environment Three is more complicated than Environment Two. Additional four static and two dynamic obstacles are present in the environment. There are nine static obstacles and four dynamic obstacles altogether; and the number of the vertices of the static obstacles reaches 53. Unlike what we have simulated in the last two environments, the dynamic obstacles in Environment Three do not appear simultaneously. Furthermore, the trajectory of one dynamic obstacle is not a strait line, i.e., the dynamic obstacle changes its direction during movement.

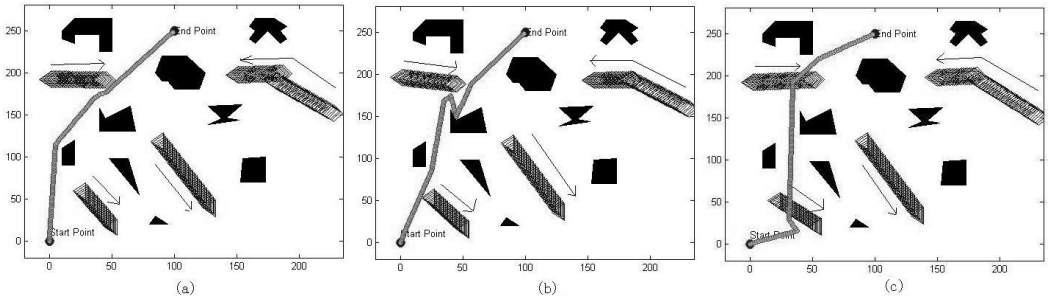


FIGURE 7: Environment Three (left to right: MSA, SA, GA)

FIGURE 7 shows Environment Three and its simulation results for the MSA, normal SA, GA approaches. No collision has occurred in all four approaches, implying that all approaches can re-plan the path successfully when moving obstacles are detected. TABLE 3 gives some quantitative simulation results for Environment Three. It is seen from this table that in all performance criteria (off-line processing time, on-line processing time, and path length), the normal SA is significantly better than the GA method; and the MSA performs much better than the normal SA.

4.5 Simulation Results for Case Four

Environment Four is the most complicated scenario in our simulation studies. There are fourteen static obstacles and six dynamic obstacles altogether in the environment. The number of the vertices of the static obstacles is 82. The dynamic obstacles appear randomly at different times and move in different directions. Also, the dynamic obstacles can change their moving directions during movement.

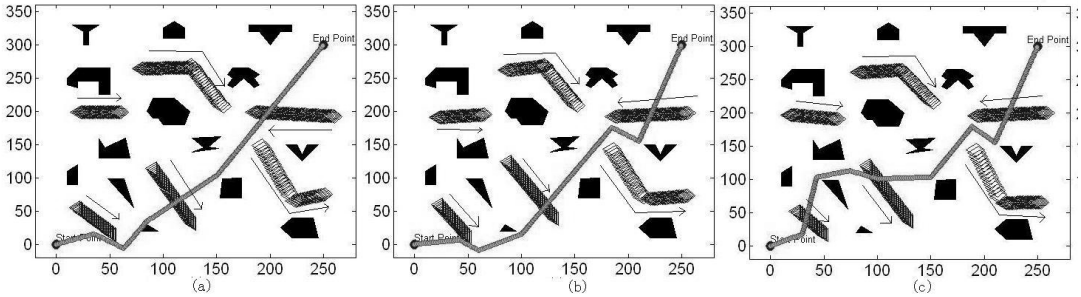


FIGURE 8: Environment Four (left to right: MSA, SA, GA)

FIGURE 8 shows Environment Four and its simulation results for the MSA, normal SA, GA approaches. Again, the robot does not collide with any obstacles in all the approaches, implying that all approaches work well in re-planning of the robot path. TABLE 3 summarizes some quantitative results for Environment Four. It is seen from these results that among the three

approaches, the MSA performs the best and the GA method gives the worst performance on the processing time, and the MSA approach gives the optimal path length result.

Environment/Performance	MSA	Normal SA	GA
Environment One			
Offline Processing Time	0.1421	0.1497	0.9147
Online Processing Time	0.1445	0.1514	1.1247
Path Length	145.78	145.78	145.78
Environment Two			
Offline Processing Time	0.2281	0.3817	1.7912
Online Processing Time	0.2411	0.4015	2.0713
Path Length	246.248	256.76	261.46
Environment Three			
Offline Processing Time	0.3418	0.9012	3.3452
Online Processing Time	0.3519	1.1075	3.9716
Path Length	280.85	290.36	305.43
Environment Four			
Offline Processing Time	0.3918	1.4098	4.1214
Online Processing Time	0.4125	1.6987	4.3123
Path Length	410.24	443.67	460.67

TABLE 3: Summary of Performance Results

5. Experiment Results Evaluation

5.1 Path Length Evaluation

FIGURE 9 graphically compares the path length performance of the three approaches for all four environments. Taken from the quantitative simulation results shown in TABLE 3, the values of the path length in the figure are median values obtained in offline path planning. It is seen from FIGURE 9 that the path length performance of all three approaches deteriorates when the environment becomes more complicated; while the MSA approach performs the best in all cases.

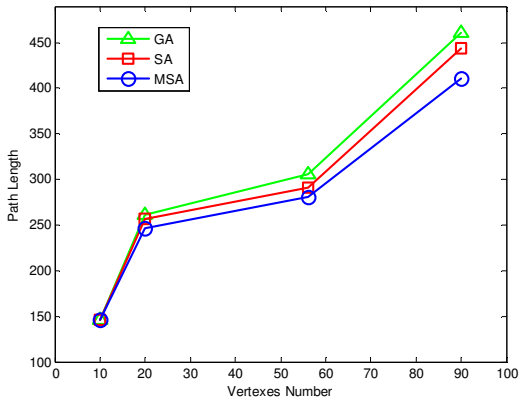


FIGURE 9: Path Length Evaluation

In the simplest environment, i.e., Environment One, all three approaches give the same path length. However, for Environments Two through to Four, the MSA approach improves the path length performance over the normal SA and GA approaches. For example, for Environment Four, the path length of the MSA is 10.9% shorter than that of the GA approach.

5.2 Offline Processing Time Evaluation

FIGURE 10 graphically demonstrates the offline processing time performance of the three approaches for all four environments. The offline planning is conducted based on the static obstacles in the environments. It is seen from FIGURE 10 that the MSA approach is significantly superior to the other two approaches. This is further verified by the quantitative comparison results in TABLE 3. For example, for Environment Four, the MSA improves the offline processing time performance by 72.5% and 90.9% over the normal SA and GA approaches, respectively.

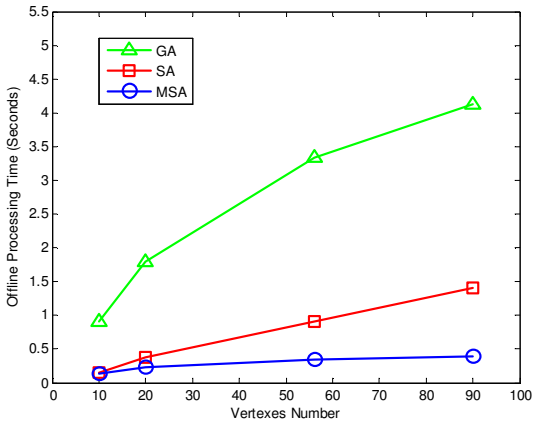


FIGURE 10: Offline Processing Time Evaluation

5.3 Online Processing Time Evaluation

Once a dynamic obstacle is detected, the on-line planner will calculate if a collision is likely to happen. If no collision is likely to occur, the robot keeps traveling along its current path; otherwise, the online planner will re-plan an alternative path for the robot. FIGURE 11 compares the processing time of the online path planning of the three approaches in all four environments. It clearly shows the superiority of the MSA approach to the other two approaches. As an example, in Environment Four with total about 90 vertices, the MSA approach consumes 81.5% less time to re-plan the path than the normal SA approach, and 92.1% less time than the GA approach.

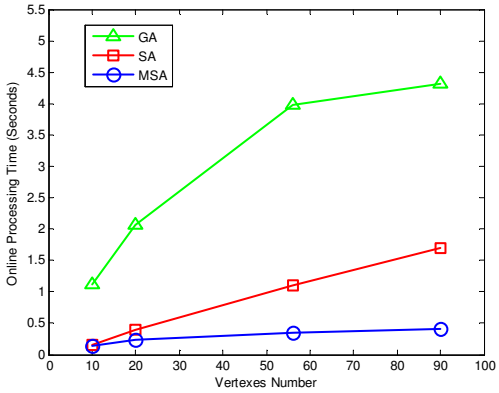


FIGURE 11: Online Processing Time Evaluation

6. CONSLUSION

A Multi-Operator SA (MSA) approach has been proposed for robot path planning in dynamic environments with both static and dynamic obstacles. The contributions of the work include the implementing of the simulated annealing algorithm for robot path planning in dynamic environments, and the enhanced new path planner for improving the efficiency of the path-

planning algorithm. Comprehensive case studies and statistical analysis have been carried out to demonstrate the proposed approach in four dynamic environments with different complexities. The MSA has been shown to be capable of giving an optimal or near-optimal path solution in various dynamic environments, and to consume much less processing time than the standard SA with two operators. Unlike the popular A* or D* based approaches, it uses the vertices of the obstacles as the search space. Compare to the previous published SA method; the proposed MSA approach introduces two more additional mathematical operators to ensure the quality of the path solutions in the evolutionary computation. With comparisons with the normal SA and GA, the MSA approaches has been shown through case studies for four dynamic environments to be effective in getting quality path solution and computationally efficient in deriving the path solution. As a result of the significant improvement in the computational efficiency, real-time and on-line applications of the developed approach in dynamic path planning become possible.

7. REFERENCES

[1] P. S. Y. Wang, J. Mulvaney, "*Genetic-based mobile robot path planning using vertex heuristics*," in Proceedings of the Conference on Cybernetics and Intelligent Systems, vol. 1, Bangkok, Thailand, June 7–9, 2006, pp. 1 – 6.

[2] Ahmed Mustafa, Aisha-Hassan A, *"Adaptive Emotional Personality Model based on Fuzzy Logic Interpretation of Five Factor Theory,"* International Journal of Computer Science and Security, vol. 3, no. 3, pp. 210–215, Sept. 2009.

[3] Dzulkifli Mohamad, *"Multi Local Feature Selection Using Genetic Algorithm For Face Identification,"* International Journal of Computer Science and Security, vol. 1, no. 2, pp. 1–10, Sept. 2007.

[4] J. N. Russell, "*Artificial Intelligence: A Modern Approach*." Berkeley, CA, USA: Prentice Hall, 2003.

[5] L. Hu and Z. Q. Gu, "*Research and realization of optimum route planning in vehicle navigation systems based on a hybrid genetic algorithm,*" Proceedings of the Institution of Mechanical Engineers Part D – Journal of Automobile Engineering, vol. 222, no. D5, pp. 757–763, May 2008.

[6] J. Ayers, "*Underwater walking,*" Arthropod Structure and Development, vol. 33, no. 3, pp. 347–360, July 2004.

[7] B. Williams and I. Mahon, "*Design of an unmanned underwater vehicle for reef surveying*," in Proceedings of the IFAC 3rd Symposium on Mechatronic Systems. Manly NSW, Australia: IEEE, Sept. 15, 2004.

[8] S. Chakravorty and J. L. Junkins, "*Motion planning in uncertain environments with vision-like sensors,*" Automatica, vol. 43, no. 12, pp. 2104–2111, Dec. 2007.

[9] Y. Wang, P. W. Sillitoe, and J. Mulvaney, "*Mobile robot path planning in dynamic environments,*" in Proceedings of the International Conference on Robotics and Automation, vol. 1. Roma: IEEE, Apr. 10–14, 2007, pp. 71–76.

[10] P.-Y. Zhang, T.-S. L¨ u, and L.-B. Song, "*Soccer robot path planning based on the artificial potential field approach with simulated annealing,*" Robotica, vol. 22, no. 5, pp. 563–566, Aug. 2004.

[11] A. Stentz, "*Optimal and efficient path planning for partially-known environments,*" in Proceedings of the IEEE International Conference on Robotics and Automation, vol. 4, San Diego, CA, USA, May 8–13, 1994, pp. 3310–3317.

[12] A. Yahia, A. Stentz, S. Singh, and B. Brummit, "*Framed-quadatree path planning for mobile robots operating in sparse environments,*" in Proceedings of the IEEE Conference on Robotics and Automation, vol. 1. Leuven, Belgium: IEEE, May 16–20, 1998, pp. 650–655.

[13] D. Ferguson and A. Stentz, "*Field D*: An interpolation-based path planner and replanner,*" in Proceedings of International Symposium on Robotics Research, San Francisco, CA, USA, Oct. 12, 2005, pp. 239–253.

[14] A. Stentz, "*The focussed D* algorithm for real-time replanning,*" In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, pp. 1652–1659, Aug. 20–25, 1995.

[15] A. Yahja, S. Singh, and A. Stentz, "*An efficient online path planner for outdoor mobile robots operating in vast environments,*" Robotics and Autonomous Systems, vol. 32, pp. 129–143, 2000.

[16] A. R. Willms and S. X. Yang, "*An efficient dynamic system for real-time robot path planning,*" IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 36, no. 4, pp. 755–766, 2006.

[17] A. R. Willms and S. X. Yang, "*Real-time robot path planning via a distance-propagating dynamic system with obstacle clearance,"* IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 38, no. 3, pp. 884–893, June 2008.

[18] H. Miao and Y.-C. Tian, "*Robot path planning in dynamic environments using a simulated annealing based approach,*" in Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision - ICARCV'2008, Hanoi, Vietnam, Dec. 17–20, 2008, pp. 1253–1258.

[19] Mathworks, "Matlab," http://www.mathworks.com, retrived on 18 Feb 2009.

[20] Sufal Das, Banani Saha, *"Data Quality Mining using Genetic Algorithm"*, International Journal of Computer Science and Security, vol. 3, no. 2, pp. 105-112. May 2009.

Estimation of Ready Queue Processing Time Under SL Scheduling Scheme in Multiprocessors Environment

D. Shukla diwakarshukla@rediffmail.com
Deptt. of Mathematics and Statistics
Dr. H.S.Gour Central University
Sagar (M.P.),470003, INDIA

Anjali Jain anjalidcsa@rediffmail.com
Deptt. of Computer Science and Applications
Dr. H.S.Gour Central University
Sagar (M.P.), 470003, INDIA

Amita Chowdhary amita411@gmail.com
Deptt. of Physics and Electronics
Dr. H.S.Gour Central University
Sagar (M.P.), 470003, INDIA

Abstract

CPU Scheduling is an open area of research where computer scientists used to design efficient scheduling algorithms for CPU processes in order to get output in the efficient manner. There are many CPU scheduling schemes available in literature. Lottery scheduling is one of them which adopts random choice of processes by the processors. This paper presents a new CPU scheduling scheme in the form of SL Scheduling which is found useful and effective. By virtue of this, an attempt has been made to estimate the total processing time of all the processes present in ready queue waiting for their processing. A numerical study is incorporated in the content to support the mathematical findings related to the estimation of processing time.

Keywords: CPU, Ready Queue, Scheduling, SL Scheduling (SLS), Lottery Scheduling.

1. INTRODUCTION

The scheduling is a methodology of queue of processes to minimize delay and to optimize performance of the system in the multiple processor environment where queues of processes exist with servers. A scheduler is part of an operating system module whose primary objective is to optimize system performance according to the criteria set by the system designers. It refers to a set of policies and mechanism, built into the operating system, which governs the order in which work to be done by computer system [see Silberschatz and Galvin [13], Stalling [9] and Tanenbaum and Woodhull [15]]. There are many CPU scheduling schemes available like FIFO, Round Robin, LIFO, DRRA etc. The lottery scheduling is one more, based on a probabilistic scheduling algorithm for in which processes are assigned some numbers in the form of lottery tickets, and the scheduler draws a random ticket to select the process. The distribution of tickets need not be uniform; granting a process more tickets to provide a relatively higher chance of selection. This technique can be used to approximate other scheduling algorithms, such as

shortest- job – next and fair- share scheduling etc.. In other words, lottery scheduling is highly responsive because it solves the problem of starvation also, giving each process at least one lottery ticket which guarantees that it has non- zero probability of being selected at each scheduling operation. Suppose that there are many processors and each fetches a process at a time from the ready queue under lottery scheduling scheme. Then this may be treated as a random sample from the long ready queue of processes. There are techniques available in the literature sampling theory by which one can improve upon the quality of sample. This paper presents a new scheduling scheme as SL scheduling (modified form of lottery scheduling) and the approach has been adopted to estimate total processing time likely to consume if entire ready queue becomes empty.

2. A REVIEW

Lottery Scheduling by Waldsparger et al. [3] has recently introduced proportional share scheduler that enables flexible control over the relative rates at which CPU- bound work loads consume processor time. David et al. [5] extended lottery scheduling, a proportional share resource management algorithm, to provide the performance assurances present in traditional non-real time process schedulers. They used dynamic tickets adjustments to incorporate into a lottery scheduler the specialization present in the Free BSD scheduler to improve interactive response time and reduce kernel lock contention, which enables flexible control over relative process execution rates with a ticket abstraction and provides load insulation among group of processes using concurrencies. Shukla and Jain [7, 8] examined the multilevel queue scheduling scheme and examined the deadlock property using stochastic process. Shukla and Jain [9] presented deficit round robin alternated (DRRA) scheduling algorithm under Markov chain model and examined variety of scheduling scheme and their relative mutual comparisons by simulation study. Raz et al. [6] described *n* jobs to service, *p* class of priority, and m servers for the queue which holds tasks to execute and introduce some simulation results for the formula for dynamic priority calculation for CMPQ. The goal is to assure that even in worst case situations starvation does not occur. Cochran [4] contains an introduction to the methods of sampling theory with applications over multiple data. One more contribution is due to Tanenbaum and Woodhull [15].

3. MOTIVATION

Deriving an idea from all these contributions, this paper is an attempt to estimate possible time duration in case when a bank server or power supply is suddenly shut down to avoid disaster for few minutes. If some processes are running on different machines then it is not wise to stop them all of a sudden. In such a case one may desire know after what time they all will be finished from ready queue, then after estimating time duration we will be able to stop processing. Therefore, it is an open problem for researcher to estimate the total time of all processes in the ready queue likely to be consumed before closing the systems. Efficient sampling methodologies could be useful at this level to develop computational technique.

4. SL SCHEDULING SCHEME

SL Scheduling (SLS) scheme employs a technique in which the complete and up-to-date list of the processes is available in the Ready Queue of the system. It selects only the first process in random manner and the rest being automatically selected according to some predetermined pattern. The random number *'i'* is random start whose value is determined by CPU logic unit. The CPU then estimates duration of possible processing time of all *N* processes at the end of a session. The SL scheduling is laid down as under:

a) Assume *N* processes in the ready queue and the number *N* is such that *N=nk* holds for any positive number *n* and *k*. The system has *k* processors in multiprocessor environment. Every process in ready is assigned a token of serial number 1 to *N* while arrival.

b) The CPU restricts a session in which all *N* ready queue processes are available for execution.

c) Scheduling chooses randomly a serial number *i (1 ≤ i ≤ n)*.This process is assigned to the first processor Q_1.

d) The other processors $Q_2 \ldots . Q_k$ are assigned processes having serial number [*i+n, i+2n, i+3n……i + (n-1) k*].

e) At the end of the first job processing session CPU computes mean time of all *k* jobs processed in a session.

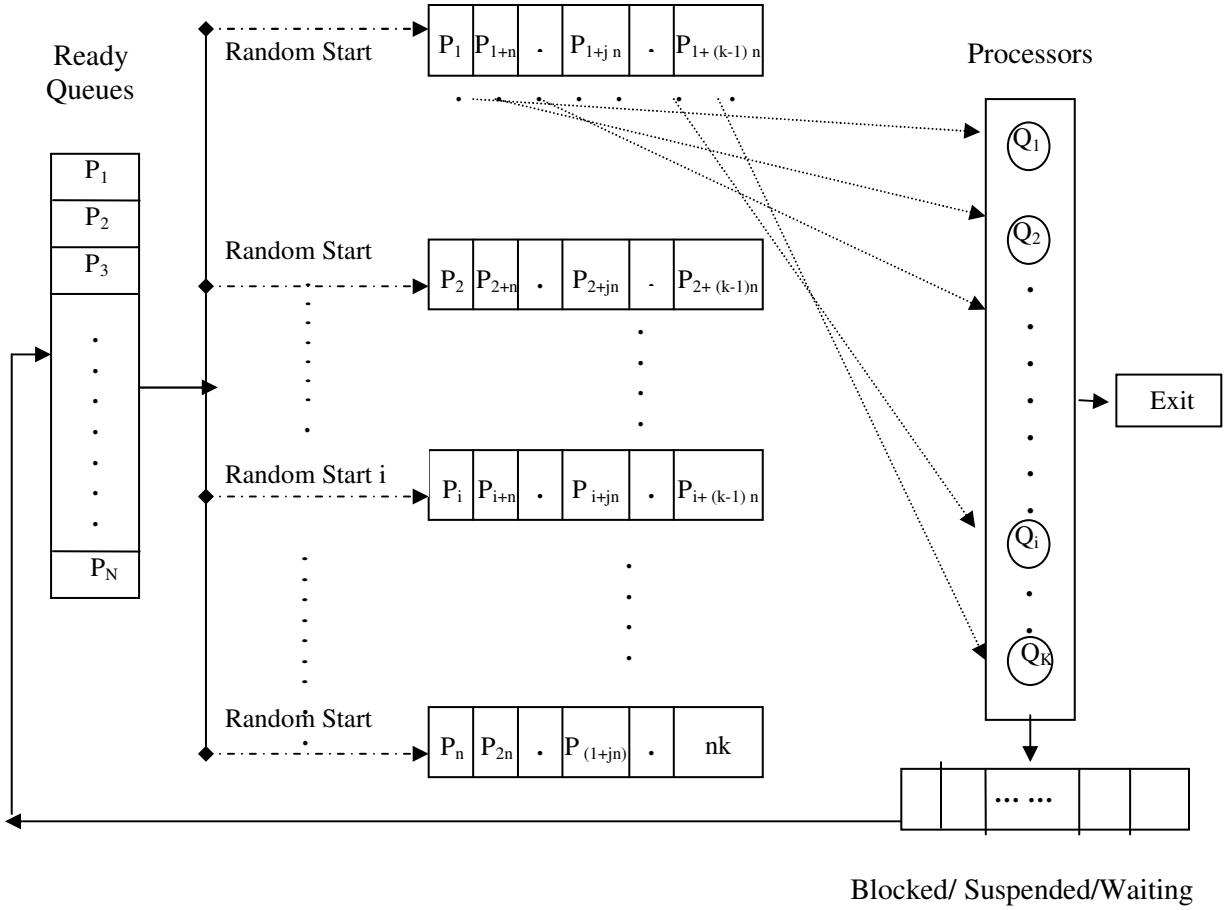


FIGURE 1: Processing of Ready Queue under Systematic Lottery Scheduling Scheme

5. ESTIMATION OF READY QUEUE PROCESS TIME IN A SESSION

Let t_{ij} denote the time of processing consumed for j^{th} process of the i^{th} sample, (*i=1,2,……n; j=1,2,…….k*).

$\bar{t}_{i.}$ = Mean of the i^{th} systematic sample

$$= 1/k \sum_{j=1}^{k} t_{ij} \quad \text{.... (5.1)}$$

$\bar{t}..$ = Overall process mean time of *N* processes in ready queue

$$= 1/nk \sum_{i=1}^{k} \sum_{j=1}^{n} t_{ij} = 1/n \sum_{i=1}^{n} \bar{t}_{i.} \quad \text{.... (5.2)}$$

S^2 = Mean square of processing time for all N processes in ready queue

$$= 1/(N-1) \sum_{i=1}^{k} \sum_{j=1}^{n} \left(t_{ij} - \bar{t}_{..}\right)^2 = 1/(nk-1) \sum_{i=1}^{k} \sum_{j=1}^{n} \left(t_{ij} - \bar{t}_{..}\right)^2 \quad \text{.... (5.3)}$$

TABLE 1: The *k* possible systematic samples together with their means

Random Start	Sample Composition *(Units in the sample)*	Probability	Mean
1	1 1+n.....1+jn.....1+ (k-1) n	1/n	$\bar{t}_{1.}$
2	2 2+n.....2+jn.....2+ (k-1) n	1/n	$\bar{t}_{2.}$
.			
i	i i+n..... i+jn.....i+ (k-1) n	1/n	$\bar{t}_{i.}$
.			
n	K 2n..... (1+j) n..... nk	1/n	$\bar{t}_{k.}$

Thus *k* rows of the table 1 gives the *k*-systematic random samples. The probability of selecting i^{th} group of processes as the systematic sample is *1/n.* The $\bar{t}_{i.}$ is sample mean time consumed by K processors each to process one job in a session. The expected value of sample mean is

$$E\left(\bar{t}_{i.}\right) = 1/k \sum_{i=1}^{k} \bar{t}. = \bar{t}_{..} \quad \text{....(5.4)}$$

So if *N= nk*, the process sample mean provides an unbiased estimate of the entire processes ready queue mean. Let $\bar{t}_{sys}$ is mean time of one systematic sample of size *k* units. Then $\bar{t}_{sys}$ is estimator of ready queue mean time and

$$\bar{t}_{sys} = \bar{t}_{i.}$$

5.1 Variance of the Estimated Mean

$$\text{Var}\left(\bar{t}_{sys}\right) = 1/n \sum_{i=1}^{n} \left(\bar{t}_{i.} - \bar{t}_{..}\right)^2 \quad \text{.... (5.5)}$$

$$\text{Var}\left(\bar{t}_{sys}\right) = ((N-1)/N)S^2 - ((n-1)k/N)S^2_{sys} \quad \text{.... (5.6)}$$

where $$S^2_{sys} = 1/k(n-1)\sum_{i=1}^{n}\sum_{j=1}^{k}\left(t_{ij} - \bar{t}_{i.}\right)^2 \quad \text{.... (5.6a)}$$

Which is the mean square among process time k units which lie within the same systematic samples.

6. NUMERICAL ILLUSTRATION

TABLE 2: Data Set

Processes	P_1	P_2	P_3	P_4	P_5
CPU Time	30	20	112	40	59

Processes	P_6	P_7	P_8	P_9	P_{10}
CPU Time	60	33	43	101	69

Processes	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}
CPU Time	138	43	109	26	74

Processes	P_{16}	P_{17}	P_{18}	P_{19}	P_{20}
CPU Time	89	123	67	58	84

Processes	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}
CPU Time	143	29	147	94	131

Processes	P_{26}	P_{27}	P_{28}	P_{29}	P_{30}
CPU Time	79	46	59	72	22

Considered *30* processes in the ready queue and their CPU time as shown in table 2 with *n=5*, *k=6* and *N=nk* holds.

6.1. Under Systematic Lottery Scheduling (SLS) Scheme

We have taken random samples of *6* processes from given *30* processes as shown in table 2 and find their sample mean time as shown in table 3.

TABLE 3: Computation of Sample Mean Time for SLS

Sample number for random start n=5	**Sampled Process (*k*=6) Sampled Processing Time**	**Sample Mean Time**
i=1	P_1 =30, P_6 = 60, P_{11} = 138, P_{16} = 89, P_{21} = 143, p_{26}=79	89.83

i=2	P_2 = 20, P_7 = 33, P_{12} = 43, P_{17} = 123, P_{22} = 29, P_{27} =46	49
i=3	P_3 = 112, P_8 = 43, P_{13} = 109, P_{18} = 67, P_{23} = 147, P_{28} = 59	89.5
i=4	P_4 = 40, P_9 = 101, P_{14} = 26, P_{19} = 58, P_{24} = 94, P_{29} = 72	65.16
i=5	P_5 = 59, P_{10} = 69, P_{15} = 74, P_{20} = 84, P_{25} = 131, P_{30} = 22	73.16

TABLE 4: Computational Values for Total Processes

Total Numbers of Processes *N*	30
Mean Time $\bar{t}..$	73.33
Square of Mean Time	5377.28
Total Sum of Squares $\sum_{i=1}^{n}\sum_{j=1}^{k} t_{ij}^2$	203712
Mean Square S^2	1461.8390
Variance of SL Scheduling $Var(\bar{t}_{sys})$	238.48

Confidence Interval: The 99% confidence interval is $\left[\bar{t}_{sys} - 1.96\sqrt{V(\bar{t}_{sys})}, \bar{t}_{sys} + 1.96\sqrt{V(\bar{t}_{sys})}\right]$

TABLE 5: Computation of Confidence Intervals

Random Sample	Sampled Processing Time	Total Time	Sampled Mean	Confidence Interval of Time for per process	Confidence Interval for Total Time for complete Ready Queue
1.	30,60,138,89,143,79	539	89.83	(59.57,120.09)	(1787.1,3627)
2.	20,33,43,123,29,46	294	49	(18.74,79.26)	(562.2,2377.8)

3.	112,43,109,67,147,59	537	89.5	(59.24,119.76)	(1777.2,3592)
4.	40,101,26,58,95,72	391	65.16	(34.9,95.42)	(1047,2862.6)
5.	59,69,74,84,131,22	439	73.16	(42.9,103.42)	(1287,3102.6)

7. CONCLUDING REMARKS

It is observed that SL scheduling is a more scientific way of representing algorithm than usual lottery scheduling. The unique feature it has, to provide procedure of estimating ready queue processing time. Since sample representation is better by this procedure, so the queue time estimation is also sharper. In table 5, most of confidence intervals contain true value within the 99% confidence limits. It seems SL scheduling helps to estimate ready queue time processing length in advance. These estimates are useful when suddenly the system needs to shut down due to unavoidable reasons.

8. REFERENCES

1. Ankur Agarwal "System-Level Modeling of a Network-on-Chip", International Journal of Computer Science and Security (IJCSS), 3(3):154-174, 2009.
2. Agarwal, Rinki and Kaur Lakhwinder "On flexibility analysis of fault tolerant multistage interconnection networks, International Journal of Computer Science and Security (IJCSS),2(4), 01 – 08,2008.
3. Carl A. Waldspurger William E. Weihl "Lottery Scheduling a flexible proportional-share resource management", Proceedings of the 1st USENIX Symposium on Operating Systems Design and Implementation (OSDI): 1-11, 1994.
4. Cochran, W.G "Sampling Technique", Wiley Eastern Publication, New Delhi 2005.
5. David Petrou, Garth A. Gibson, John W. Milford "Implementing Lottery Scheduling: Matching the specializations in Traditional Schedulers", Proceedings of the USENIX Annual Technical Conference USA: 66-80, 1999.
6. Raz, D., B. Itzahak, H. Levy "Classes, Priorities and Fairness in Queuing Systems". Research report, Rutgers University, 2004.
7. Shukla, D. and Jain, Saurabh. "*A Markov chain model for multilevel queue scheduler in operating system*, Proceedings of International Conference on Mathematics and Computer Science, ICMCS-07, pp. 522-526, 2007.
8. Shukla, D. and Jain, Saurabh. "*Deadlock state study in security based multilevel queue scheduling scheme in operating system"*, Proceedings of National Conference on Network Security and Management, NCNSM-07, pp. 166-175, 2007.
9. Shukla, D. and Jain, Saurabh "A Markov chain model for Deficit Round Robin Alternated (DRRA) scheduling algorithm", Proceedings of the International Conference on Mathematics and Computer Science, ICMCS-08: 52-61, 2008.
10. Shukla D., Tiwari Virendra, Thakur Sanjay, And Deshmukh, A. K. "Share Loss Analysis of Internet Traffic Distribution in Computer Networks" International Journal of Computer Science And Security(IJCSS), 3(5): 414- 427,2009.
11. Shukla D., Tiwari Virendra, Thakur Sanjay, And Tiwari Mohan "A comparison of methods for internet traffic in computer network", International Journal of Advance Networking and Applications, 1(3): 164- 169, 2009.
12. Shukla D., Ojha, Shweta, And Jain, Sourabh, "Analysis of multilevel queue with the effect of data model approach", Proceedings of the National Conference on Research and Development Trends in ICT (NCRTICT – 10), 245- 251, 2010.

13. Silberschatz, A. and Galvin, P. "*Operating System Concepts"*, Ed.5, John Wiley and Sons (Asia), Inc. (1999)
14. Stalling, W. "*Operating System"*, Ed.5, Pearson Education, Singapore, Indian Edition, New Delhi. (2004)
15. Tanenbaum, A. and Woodhull "*Operating system"*, Ed. 8, Prentice Hall of India, New Delhi,(2000).

Detecting and Localizing Wireless Network Attacks Techniques

Iyad Aldasouqi iyad@rss.gov.jo
Princess Sumaya University for Technology
The King Hussein School for Information Technology

Walid Salameh walid@psut.edu.jo
Princess Sumaya University for Technology
The King Hussein School for Information Technology

Abstract:

In order to increase employee productivity within a feasible budget, we have to track new technologies, investigate and choose the best plan and implementation of these technologies.

WLAN is vulnerable to malicious attacks due to their shared medium in unlicensed frequency spectrum, thus requiring security features for a variety of applications.

This paper will discuss some techniques and approaches which can help to detect, localize and identify wireless network attacks, which present a unique set of challenges to IT and security professionals. All efforts were focusing on the ability to identity based attacks in which a malicious device uses forged MAC addresses to masquerade as a specific client or to create multiple illegitimate identities. Also, to be sure that the network is able to robustly identify each transmitter independently of packet contents, allowing detection of a large class of identity-based attacks with high probability.

The attacker can listen to all wireless traffic, compromise encryption and Use attenuators, amplifiers, directional antennas, software radios, but he cannot be at the location of user or at the location of access points. However, we have to choose the best design, implementation, and evaluation techniques in order to secure our network from attackers, where our choice will depend on a technical implementation to mitigate the risk on the enterprise network infrastructure.

Keywords : *Security, Sensors, Access points, wireless, Authentication*

1.Introduction:

Wireless Local Area Network (WLAN) which became increasingly viable for many reasons, the same wireless technology that can erase the physical limitations of wired communications to increase user flexibility, boost employee productivity, and lower cost of wireless network ownership.

Security becomes a key factor and boosts employee demand for access to their enterprise's wireless network beyond the area of their office workstation. In addition, wireless access to a network can represent the entry point for various types of attacks, which can crash an entire network, render services unavailable, and potentially subject the enterprise to legal

liabilities, so we can understand that there are many factors affected on the quality and strength of the security, such as the signal propagation characteristics, limited bandwidth, weak processing capability, and various other reasons.

Wireless Network

Wireless frequencies are designed to be used by anyone with a wireless receiver – anyone can connect to a wireless network in the same way that they can tune into a radio station.

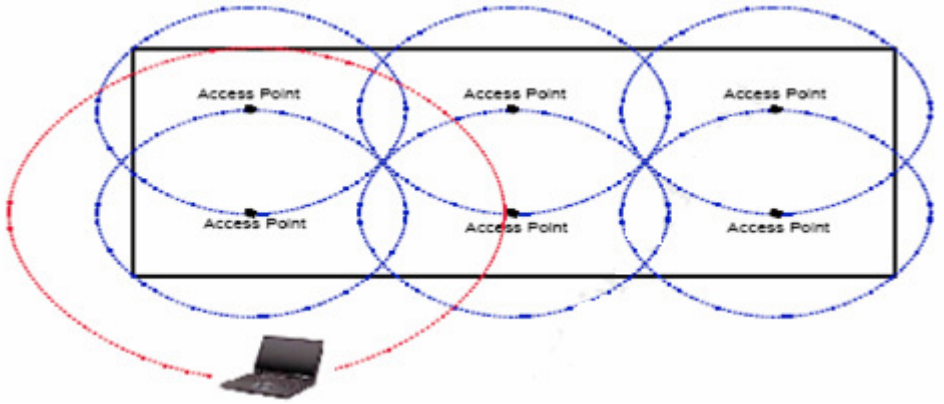


Figure 1: WLAN Coverage can often overrun a building's boundaries.

A wireless local area network (WLAN) is a flexible data communications system that can use either infrared or radio frequency technology to transmit and receive information over the air. In 1997, 802.11 was implemented as the first WLAN standard. It is based on radio technology operating in the 2.4 GHz frequency and has a maximum throughput of 1 to 2 Mbps. The currently most spread and deployed standard, IEEE 802.11b, was introduced late 1999. It still operates in the same frequency range, but with a maximum speed of 11 Mbps.

WLAN has been widely used in many sectors ranging from corporate, education, finance, healthcare, retail, manufacturing, and warehousing. According to a study by the Gartner Group, approximately 50 percent of company laptops around the world will be equipped for WLAN by 2006 [3]. It has increasingly becoming an important technology to satisfy the needs for installation flexibility, mobility, reduced cost-of-ownership, and scalability.

1.1 Intrusion Detection

"For an enterprise to protect itself from abuse of its information, it must monitor the events occurring in its computer system or network and analyze them for signs of intrusion. To do this, the enterprise must install an Intrusion Detection System (IDS)." *Ant Allen, research director at Gartner.*

IDS watch the wired and wireless network from the inside and report or alarm depending on how they evaluate the network traffic they see. They continually monitor for access points to the network and are able, in some cases, to do comparisons of the security controls defined on the access point with pre-defined company security standards and either reset or closedown any non conforming AP's they find. The distinction between placing IDS sensors on both wired and wireless networks is an important one as large corporate networks can be worldwide.
IDS systems can also identify and alert to the presence of unauthorized MAC addresses on the networks. This can be an invaluable aid in tracking down hackers.[1]
However, IDS is a vital component in auditing a network installation.

MAC Address spoofing

MAC addresses can be easily changed through device drivers, effective attacks can be implemented with some equipment available on the market. IEEE 802.11 facing many security threats, which represented by a class of attacks which can be known as masquerading attacks.[3] With such tools, the attacker modifies either the MAC or the IP address of the victim in order to adopt another identity in the network. By this technique the intruder will be able to operate as a trustworthy node and can advertise incorrect routing information to other participants of the

network. Another example is creation of loops in the routing computation which result in unreachable nodes.

To prevent and secure the network from spoofing, the specialist divided the techniques into three categories:

1. Sequence number analysis: by modifying the MAC address header, so each device will have a serial number(SN)
2. Transceiver fingerprinting: where each radio transceiver has its unique shape and pattern.
3. Signal strength analysis: which depends on the strength of the coming signals from the clients.

Physical Layer
Physical layer is hard to frog and not easy as the MAC address; because the information in this layer is inherent to radio characteristics and the physical environment, in addition it is used to differentiate devices. Hall uses the frequency-domain patterns of the transient portion of radiofrequency (RF) signals, as a fingerprint, to uniquely identify a transceiver [5].

This paper is divided into three sections. Starting by describing available methods to eliminate attacks; secondly, comparing between available techniques from different perspectives; and thirdly, are my suggestions which are depending on the first two sections in order to bet better results. The rest of the paper organized as follows: survey 802.11 spoofing-based attacks and related detection methods in Section 2. Then describe the key observation regarding section 2 techniques and compare between them in section 3. The suggested technique, which is a hybrid technique from previous two techniques and finally the conclusion, will be in Section 5.

2.Spoofing Attack and related work

It is very important to distinguish between two terms localization and spoof detection, actually they are different types of problems. Localization is based on the assumption that all measurements gathered received signal strength (RSS) are from a single station and, based on this assumption, the localization algorithm matches a point in the measurement space with a point in the physical space. But Spoofing detection distinguish if all matched measurements are from a single station, and tries to determine whether they are definitely from the same station.

2.1. Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength

This method is using “air monitors” (AMs), which is a device available on the market used to passively sniff wireless traffic, without cooperation with other devices (Access Points (APs), computers).
An AM is an embedded device and may not capture all frames sent by transmitters in its range, due to limited resources. Their own AM sniffing software, basset, passively captures wireless frames and forwards the key frame features to a centralized merger, which removes duplicates and synchronizes timestamps to construct a more complete and coherent frame sequence that is stored for further analysis [6].

They developed a RSS profiling algorithm based on the Expectation-Maximization (EM), in which they referenced to Gaussian Mixture Model (GMM) [7]. Once the RSS is ready to receive from any transmitter in normal conditions, it will distinguish any difference in the RSS signals and it will consider it as a potential spoofing attack. After a set of signals received they did some of hypothesis, algorithms and calculations (Ratio Test) as a detection tool each AM in order to increases detection accuracy.

In addition they developed two global detection algorithms which are focusing on:
1. Combine local statistics from multiple AMs.
2. Works on the frame sequence output by the merger.

This method has a role in improving networking intrusion detection via some contributions:

1. Discovered that antenna diversity is the major cause of multimodal RSS patterns.
2. Presented a new GMM profiling algorithm.

2.2. Detecting Identity Based Attacks in Wireless Networks Using Signal-prints

Faria and Cheriton propose to detect spoofing attacks using a signal-print, which is the vector of median RSS for a MAC address measured at multiple AMs [8]. They believed in that a transmitting device can be robustly identified by its signal print, a stream of signal strength values reported by access points acting as sensors. In addition they proved that, different from MAC addresses or other packet contents, attackers do not have as much control regarding the signal prints they produce. Signal-print can be represented by a signal strength characterization of a packet transmission. Each signal-print is represented as a vector of signal strength measurements, with one entry for each access point acting as sensor.

They restricted themselves to 802.11 networks, but as they said the ideas presented can be equally applied to other wireless LAN technologies. Regarding the network architecture they suggested to use the network as in figure2, which composed of multiple access points (APs) distributed across the environment that feed traffic information to a centralized server, which we call a wireless appliance (WA). In addition they focused on the access points deployed as sensors: by observing the traffic on a channel specified by the WA and collect information such as the received signal strength level for each packet successfully received. This information is then forwarded to the WA, which is able to create a signal-print for each packet of interest. [8]

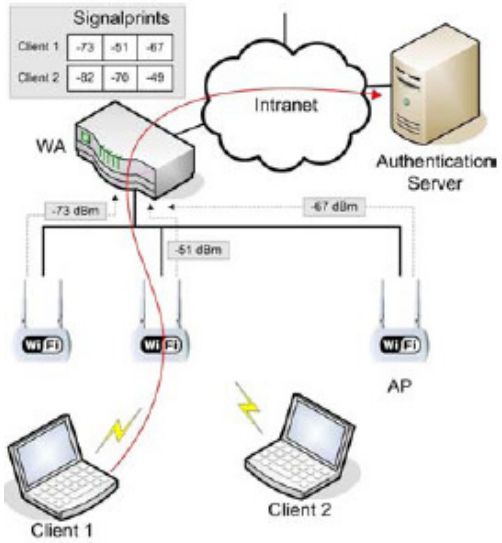


Figure 2: Signal-print creation

Signal-print Properties:

-Signal-prints are hard to spoof. Signal attenuation is a function of the distance between clients and access points, with a strong dependence on environmental factors such as construction materials and obstacles such as furniture items [9, 10].

-Signal-prints are strongly correlated with the physical location of clients, with similar signal-prints found mostly in close proximity.

-Packet bursts transmitted by a stationary device generate similar signal-prints with high probability.

-Signal-prints allow a centrally controlled WLAN to reliably single out clients. Instead of identifying them based on MAC addresses or other data they provide, signal-prints allow the system to recognize them based on what they look like in terms of signal strength levels.

MATCHING SIGNALPRINTS:

In order to distinguish between different based attacks signals matching rules are specified.

These rules can be categorized into:

- Differential Values: which represent the absolute values (In dBm) of the difference between the value at a given position and the maximum value found in that signal-print.

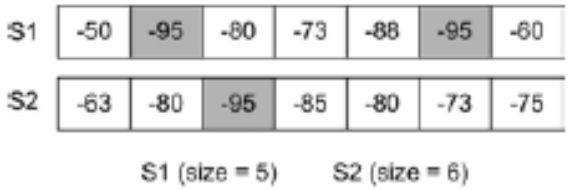
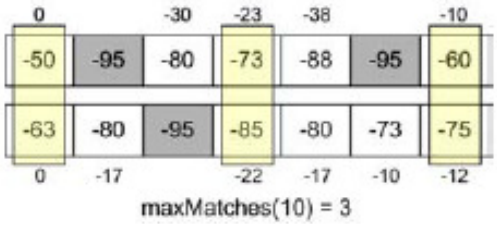


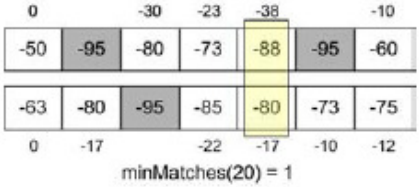
Figure 3: Shows two signal-prints and their corresponding sizes.

- Max-Matches: Matches are found by comparing values at the same position in two different signal-prints.



Figures 4: Demonstrate how max-matches are computed.

- Min-Matches: Analogous to a max-match, which is found whenever values differ by at least a certain value in dB.



Figures 5: Demonstrate how min-matches are computed.

Matching Rules: a pair of signal-prints matches if they satisfy a specified matching rule, a Boolean expression involving numbers of max-matches and min-matches, and possibly signal-print properties such as size. Example: The matching rule max-Matches(S1; S2; 5) ≥ 4 requires two signal-prints to have RSSI values within 5 dB of each other in at least 4 positions.

Finally, attack detection has three properties which are important for the analysis of this method: R denotes the rate in packets per second (pps) required for a given DoS attack to be effective. S denotes the speed of the device. A denotes the number of antennas under the control of the attacker.

2.3. Wireless Client Puzzles in IEEE 802.11 Networks: Security by Wireless

It is a protection method which assists an AP to preserve its resources by discarding fake requests, while allowing legitimate clients to successfully join the network. Rather than conditioning a puzzle's solution on computational resources of highly heterogeneous clients, the puzzles utilize peculiarities of a wireless environment such as broadcast communication and signal propagation which provide more invariant properties. [13]

The puzzle is a question about which other stations are in the client's signal proximity as in figure.6, and can thus be labeled as neighbors. The received signal strength of neighbors is strong, contrary to non-neighbors which are received weakly in relation to a certain signal value. In other words it is security by wireless application, since it is exploit the chaotic and erratic character of radio communications, describing the radio of the neighborhood, do the mutual verification via the broadcasting as in figure 7.and depending on the new location of the client (N) there will be different solutions as in figure 8.

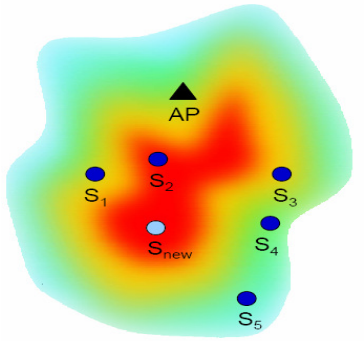


Figure 6: Signal Proximity

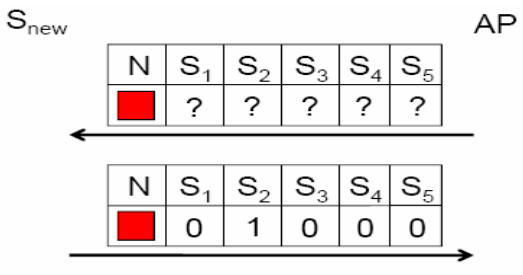


Figure 7: Mutual verification

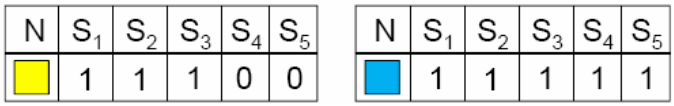


Figure 8: Solutions for different N's

Asymmetries and noise in the wireless channel can cause wrong solutions for honest requests; which caused by small deviations as in figure 9.

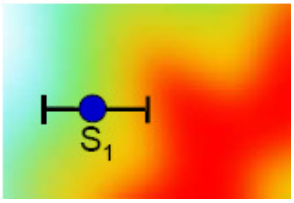


Figure 9: Small deviation gives wrong solutions

Therefore, the attackers can't exploit these tolerance intervals, which make it hard for them to attack the network.

The puzzle experiment started with the AP broadcasting the NST (The Neighborhood Signal Threshold) within a beacon frame. The NST was randomly chosen by the AP from values between -55 dBm and -95 dBm in steps of 5 dBm and changed every 7 seconds. The joining station monitors the channel and computes the sample median (choose a sample size of 20 received frames) that, after receiving a beacon frame and identifying the NST, is used to create a region by selecting those stations as neighbors whose signal strength is greater or equal to the current NST. The region is then sent along with the authentication requests to the AP. If no warnings arrive (the timer was set to 1 second) and no such region has already been used by another associated station, the AP responds with an authentication successful frame and proceeds with the association procedure. On the other hand, if a warning arrives the joining station is declined and it must wait for a different NST to re-attempt the authentication procedure. [13]

An AP has a decision role in selecting a subset of its associated stations to participate in wireless client puzzles in order to avoid increasing the number of false positives in larger networks, which will eliminate the number of warnings and false positives resulting from unsymmetrical channels. So if these subsets changed randomly, it will be too difficult for an attacker to guess which stations are currently monitoring the channel.

2.4. Advancing Wireless Link Signatures for Location Distinction (AWLS)

The authors of this technique want to show that: Detecting whether a transmitter is changing its location or not. In other words, unlike localization or location estimation, location distinction does not attempt to determine where a transmitter is. Therefore it is useful in many applications; especially it can enforce physical security by identifying illegal transmitter.

In this technique they use sophisticated physical-layer measurements in wireless networking systems for location distinction. First they compared two existing location distinction methods

1. Channel gains of multi-tonal probes: where the channel frequency response is sensitive to each multipath. An impulse in the time domain is a constant in the frequency domain, and thus a change to a single path may change the entire multiple tone link signature.
2. Channel Impulse Response (CIR): it uses a time domain signature, which support it with more robust against channel small changes.

Then, they combined the benefits of these two methods to develop a new link measurement that called the complex temporal signature. They used a 2.4 GHz link measurement data set, to evaluate the three location distinction methods. They found that the complex temporal signature method performs significantly better compared to the existing methods. They also perform new measurements to understand and model the temporal behavior of link signatures over time. They integrated their model in location distinction mechanism and significantly reduced the probability of false alarms due to temporal variations of link signatures. [14]

The link signatures in the multiple tones probing method and in the temporal link signature method both make measurements of the multipath channel and use them to quantitatively identify a link.

In addition, AWLS improved the multiple tone probing method by developing a new link signature using the strengths of the two existing methods. The proposed improvements includes:

1. A new metric related to the first method, that improve its robustness to changing received powers.
2. Come with a new method which combines the strength of the two methods, and show that a simple metric is robust to uninformative, random phase shifts, which will give us an accurate measured distance between two link signatures.

2.5. PARADIS: Physical 802.11 Device Identification with Radiometric Signatures

This technique used passive radio-frequency analysis to identify the location. They measure artifacts of individual wireless frames in the modulation domain, identify a suite of differentiating features, and apply efficient 802.11-specific machine-learning based classification techniques to achieve significantly higher degrees of accuracy than prior best known schemes. [17]

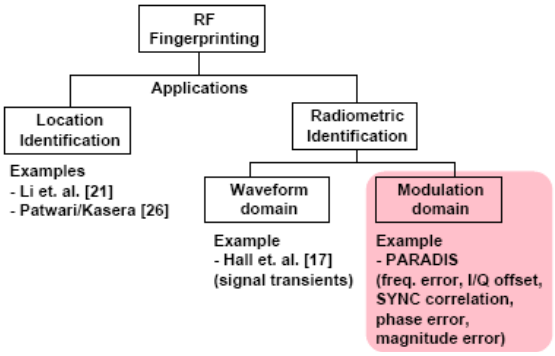


Figure 10: Radiometric identification and PARADIS.

This system built to distinguish between 802.11 nics and to achieve significantly improved identification accuracy when compared to schemes operating over transient signal characteristics. Furthermore Paradis uses distinct features from the modulation domain, frequency error, magnitude error, phase error, I/Q offset, and sync correlation of the corresponding wireless frame.

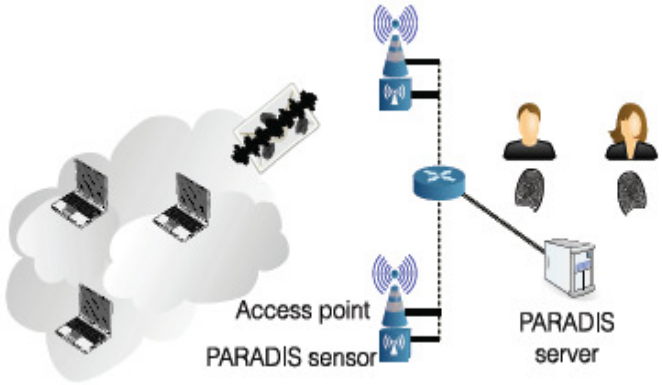


Figure 11: PARADIS schematic

Every radio transceiver can be presented by a unique physical signal, which guided them to build a library of patterns. To distinguish between these pattern they used wavelet and fuzzy neural networks as in figure 12. Therefore, to implement this technique the requirement cost will be high for both measurement and analysis devices, and thus limits the use of this technique.

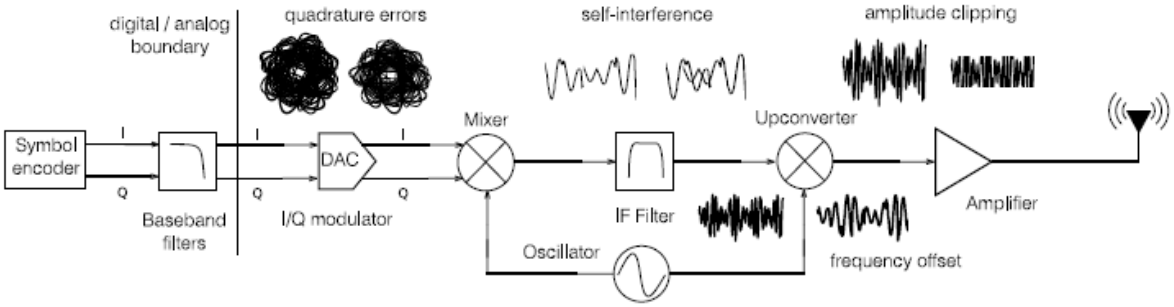


Figure 12: Common transmitter impairments and their sources

3. Analysis and Comparison

3.1 Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength versus other techniques using RSS

Detecting 802.11 techniques believed that these RSS-based detection methods are not effective due to recent advances in wireless hardware. And they proved that via conducting a series of large scale experimental studies of RSS measurements.

There are wireless networks classes that provide automatic reconfiguration of APs, adjusting power levels and channel assignments to optimize coverage while minimizing contention between neighbors. Most such systems reconfigure infrequently. By comparing the detecting 802.11 method with other methods using network management software we can see that this method can re-compute an AP's profile whenever it is reconfigured.

3.2 Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength versus Detecting Identity Based Attacks in Wireless Networks Using Signal-prints

Refers to detecting 802.11, Signal-print demonstrated above 95% detection accuracy in their test bed. False positive rate is not reported. They did observe some missing RSS measurements for AMs, and for signal-print-matching they propose to ignore any AMs with missing RSS values. However, they did not use statistical methods. On the other hand the Detecting 802.11 they did like signal-print's work; they also build a normal profile for a transmitter, and detect spoofing attacks by matching to the profiles. In addition their detector works even if the genuine station is quiet or absent, or there are multiple attackers. Unlike signal-print, their algorithm uses per-frame RSS measurements and multiple AMs. They re-implemented signal-print's algorithms, to the best of understanding.

3.3 Detecting Identity Based Attacks in Wireless Networks Using Signal-prints versus client puzzles

Puzzles technique enforced any incoming request to send back computational puzzles, which require CPU- or memory intensive operations. In addition puzzles demand that both clients and servers be modified, increasing deployment overhead when compared to a signal-print based mechanism, implemented solely at the WA. On another side Signal-print gives a similar efficiency with less cost and equipment. The puzzle weak point is if an attacker finds a physical position, or is able to find a signal strength for transmitting a region such that k stations tolerate it, it can generate as many as 2k different regions that will not result in warnings. But in Signal-print the weak point is when two clients are very close to each other WA can't distinguish them from each other.

Furthermore, puzzle technique has an alternative approach which is to use dedicated devices (ex. Sensors (similar to Signal-print technique)) installed by a network operator to implement wireless client puzzles instead of associated stations. These sensors play the same role and cover more regions with a small number of stations.

3.4 Advancing Wireless Link Signatures for Location Distinction (AWLS) versus other techniques

Importantly, the multiple tone link signatures are a complex measurement, while the temporal link signature is a real-valued measurement. The inclusion of phase information in the multiple tone signature effectively increases the richness of the measurement space. The temporal link signature, with only magnitude information, does not retain some identifiable information about a link captured by the channel phase response, and thus we would expect it to lose some ability to uniquely identify links. [14] Also, unlike localization or location estimation, the objective of location distinction is only to distinguish one link signature from another, and not to map the signature to a particular physical coordinate as in other techniques.

Furthermore, both AWLS and Signalprint are using RSS-based signalprints to prevent impersonation in wireless local area networks, which is readily available in commodity wireless cards. But it fails to capture the rich multipath characteristics of wireless channels. After that Patwari et al [15] solved that problem by proposing the use of temporal channel impulse response, which captures the multipath characteristics of wireless channels, as a link signature for location distinction, and Li et al [16] proposed the use of complex channel gains by multi-tonal probes, that also captures multipath effects, for securing wireless systems.

4. Improvements and solution

In this section 2 propose improvements to the signal print method described in Section 3.
First, as seen in section one and two, most of Identification techniques are referred to Farias[8] as a reference and tried to compare their results with his result. Therefore I also put my suggestion depending on Farias[8]results. Second, I present how to modify this technique by using two approaches:

First Approach:

Starting from the week point (limitation):

"Due to the use of RSSI levels to characterize wireless clients, one inherent limitation of our mechanism is that it may be unable to distinguish two devices located physically close to each other. Masquerading attempts can be detected if there is a noticeable difference in RSSI with respect to at least one access point. This happens even for some locations in close range, possibly due to obstacles that affect one location more than the other. In some situations - such as multiple clients in a conference room - the system may not have compelling evidence that packets are coming from different devices, making masquerading attacks possible." [8]

As known it is very hard for an attacker in any location to get close enough to the victim in offices or working area and do masquerading attack. But it will be easier for him to do that in meeting rooms or at cafeterias.

In order to prevent this from happing, I suggest doing the following:

1. Depending on the size of (cafeteria or meeting room), I suggest to have at least two AP's (it is an additional cost, but compared with it is benefit it is acceptable), which can help in showing the variances of signal print between closed clients.
2. Some times it is not applicable to have more than two AP in a small location (meeting room or cafeteria), in this case since there is a server (Authentication server), we can get benefit from the response time to calculate the distance from the access point. This addition can be added as a logical statement in the authentication application (program), first by determine which AP gives the highest signal-print, then calculate the distance when received many request from the same location. The distance will help us in determining if the signal is coming from the same client or not, so if there is a difference it means from different clients, in this case matching rules can be applied to a new list which consists of the signal print and the distance. In the existing technique if many requests received from one signal-print, this client will be considered as an attacker, which most of the time is correct, but if there is clients who are very close to each other they will reduce the same signal-print but they are not attackers. From this suggestion, it will be easy to distinguish between closed clients and attackers. (example. channel impulse response)

Second Approach:

In this approach I suggest another solution to overcome Farias[8] limitation by using some ideas from another technique (puzzle technique [13]).

The puzzle technique is using the signal print as an alternative approach, so what I\m suggesting is to use puzzle as an alternative technique for the Signal-print.

In puzzle technique, since it depends on functionality of neighborhood monitoring, so it is centralized /decentralized where each station plus the server can distinguish its neighbors, this will affect not only on the server, but also on the station (authentication on station bases and server), this will secure the network but will exhaust the resources.

In Signal-print the entire load is only on the server and nothing on the stations; so this is the main reason of the limitation.

My suggested solution is get benefit from the authentication technique in the puzzle and uses it in the signal-print. The authentication in the puzzle is not only the username and password, but also part of the packet will represent the puzzle (The frames consist of an IEEE 802.11 frame header and an additional custom puzzle header that contains all required information (Defined within a frame's custom Information Elements)[13]), in addition to get benefit from puzzle zones (without neighbors authentication) to confuse the attacker, so he can't guess to which AP the client related.

By using first approach which can overcome Signal-print limitation with little affect on the server, but by using second approach which can increase the security level by additional affect on the server.

5. Test bid

As an implementation of previous works and explaining the idea in more details, I did the following survey in order to choose the type and direction of the antenna; therefore the zone of the access point can be specified.

This will be a prototype and can be applied to a complex network, so the hacker can't know to which access point the victim is connected. In addition, reference to my second approach (using puzzle authentication technique in signal print), each zone can have its covered area and its range of IP's.
After the survey done, the boundaries of the access point zone can be specified, not only that, but it added strength to the algorithm used to determine the locations of access points (Gaussian Mixture), so this will guide us to choose the optimal number and the most appropriable location of the access points.

To implement this survey I used the area shown in figure.12, I call it Outdoor Test Facility (OTF), which is used for testing and evaluation of wireless video system and ground sensors, the tower is used to hang the camera and wireless system on, where both of them can be powered by electricity of battery charged by solar panel (which is enough for three windy days). To implement this I used the following tools:

- CISCO Aironet 350
- 13.5db antenna
- Laptop with Network Stumbler software
- External either net card
- GPS

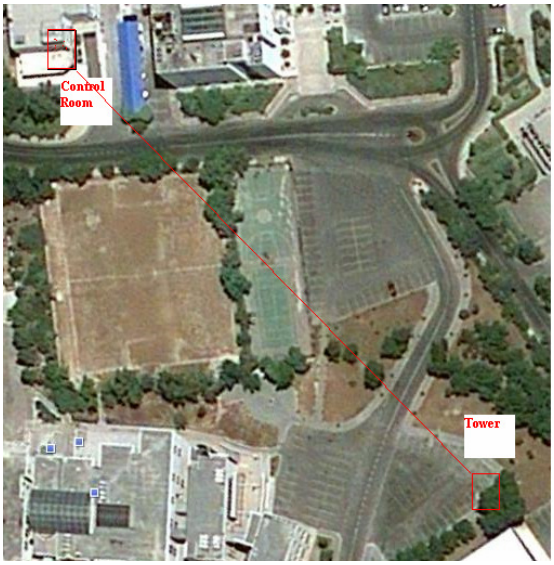


Figure 12: Outdoor Test Facility (OTF)

The distance between source (tower) and destination (control room) is 200m.

As shown in table.1 there are five columns, distance and bearing are readings from GPS, and the rest are from the software. Different types of graphs can be generated which can describe the relation between different readings. As an example, also generate a graph that represents the relation between signal and noise as in figure 13; the signal strength decreases as the noise increases.

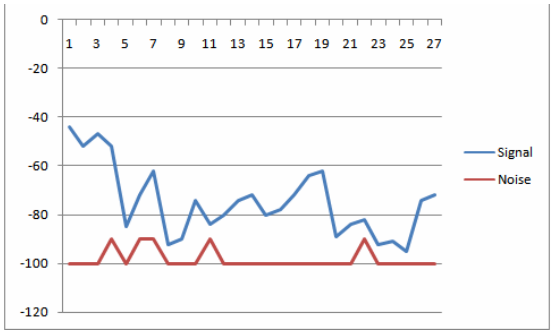


Figure 13: Signal / noise relationship

Another relation can be built between the signal and the data rate. The relation between them is a direct correlation as shown in figure 14

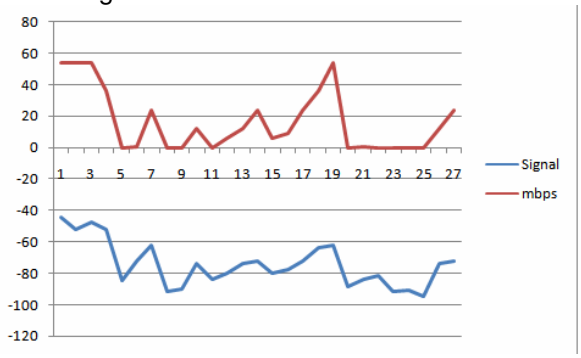


Figure 14: Signal / Data rate relationship

A third relationship can be built between the signal and distance; the signal strength becomes weaker as go further from the source as in figure 15.

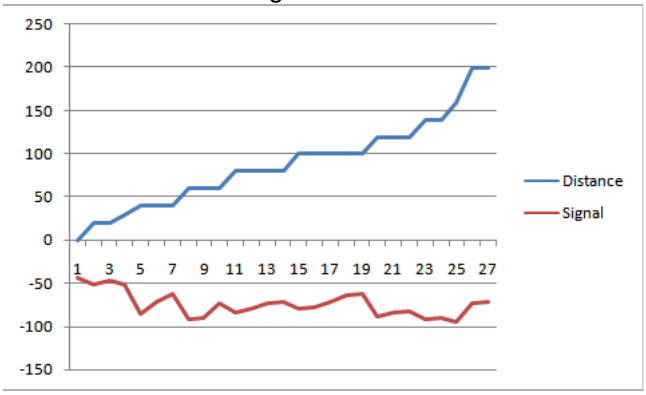


Figure 15: Signal / distance relationship

In order to choose the appropriate type of antenna, not all readings are taken into consideration. However readings are needed to verify our assumption. Only distance and signal will be used to draw the output, fist order the readings depending on distance then on signal. After that draw a radar chart of the signal readings, finally the output will be as shown in figure.16

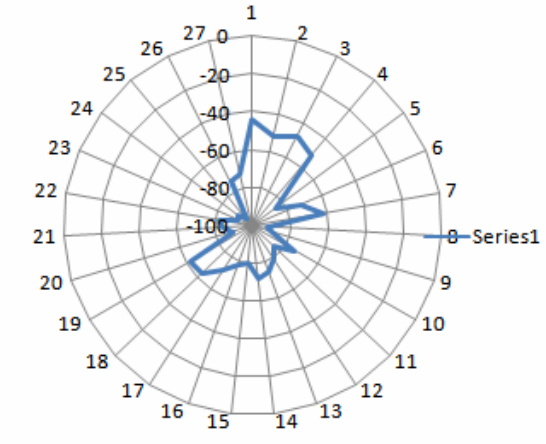


Figure 16: Signal out put

Then by comparing figure 16 with figure 17 recognize that used antenna is a directional antenna.

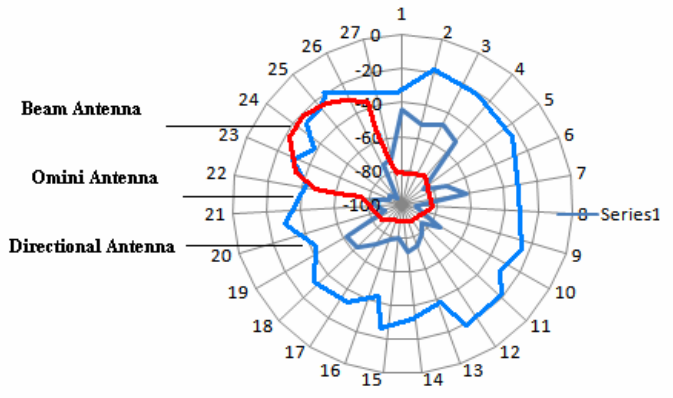


Figure 17: Antennas types

Distance	Bearing	Signal	Noise	mbps
0	240	-44	-100	54
20	270	-52	-100	54
20	250	-47	-100	54
30	290	-52	-90	36
40	80	-85	-100	0
40	340	-72	-90	1
40	20	-62	-90	24
60	100	-92	-100	0
60	120	-90	-100	0
60	140	-74	-100	12
80	160	-84	-90	0
80	180	-80	-100	6
80	200	-74	-100	12
80	220	-72	-100	24
100	330	-80	-100	6
100	300	-78	-100	9
100	280	-72	-100	24
100	240	-64	-100	36
100	260	-62	-100	54
120	20	-89	-100	0
120	0	-84	-100	1
120	340	-82	-90	0
140	60	-92	-100	0
140	40	-91	-100	0
160	60	-95	-100	0
200	230	-74	-100	12
200	240	-72	-100	24

Table1: Survey readings

As a result of this survey, the knowing of the boundaries of the access point can help us to monitor and secure our network. In addition, it will help us in our planning and future expansion, since this survey achieved our goal/assumption, it can be applied to a complex network and which can be considered as an additive security layer.

6. CONCLUSION:

MAC addresses of wireless frames can be easily forged, imposing a serious security challenge. After many experiments and researches published regarding this problem, the Received Signal Strength (RSS) which is related to the Physical-layer is most appropriate tool and it is hard to forge, in another words it can be used to detect such spoofing.

In this paper compared many existing location distinction methods. I also suggest some improvement for signal-print [8] method by using different approaches, response time approach and used the strengths of the two existing methods [8] and [13]to develop a new approach.

Nevertheless, there are still various interesting issues left open for further investigation; because until now and after all of these researches the number of false positives warning in large networks. If it is possible to control this issue the attacker will be confused and can't predict which stations are currently monitoring the channel.

Signal-prints give a good indication for the relation between mobile devices in wireless network and their physical location. The challenges are for both, the network administrator and for the attacker. For the attacker, it is how to masquerade the victim, and for the network administrator is how to protect the network without extra load and overhead on the network infrastructure.

Finally, Security methods and techniques are like antibiotics', it kills the germs. Meanwhile it has side effects on the body. In other words security slows down the network speed, but without it, we can't run our networks.

7. References:

1. **FOR CONFERENCES:** Wireless Intrusion Detection Systems, SANS, Ken Hutchison, 2004

2. **FOR JOURNALS:**Mobile and Wireless Network Security and Privacy, Edited by S. Kami Makki, Peter Reiher, Kia Makki, Niki Pissinou, Shamila Makki. 2007 Springer

3. **FOR CONFERENCES:** Swisscom.com. "Swisscom Mobile to launch Public Wireless LAN on 2, December 2002." 2 Jan. 2003. URL: http://www.swisscom.com/mr/content/media/20020924_EN.html (9 Dec. 2002).

4. **FOR CONFERENCES:** LAN MAN Standards Committee of the IEEE Computer Society. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Speci_cations - Amendment 6: Medium Access Control (MAC) Security Enhancements. Technical Report 2004 Edition, IEEE Std 802.11i, July 2004.

5. **FOR CONFERENCES:** J. Hall, M. Bareau, and E. Kranakis, "Using transceiverprints for anomaly based intrusion detection," in Proceedings of 3rd IASTED, CIIT, Nov. 2004, pp. 22–24.

6. **FOR CONFERENCES:** Y. Sheng, G. Chen, K. Tan, U. Deshpande, B. Vance, C. McDonald, H. Yin, T. Henderson, D. Kotz, A. Campbell, and J. Wright, "Securing 802.11 wireless networks through fine-grained measurements," Submitted to IEEE Wireless Communications Magazine.

7. **FOR JOURNALS:** R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," SIAM Review, vol. 26, no. 2, pp. 195–239, 1984.

8. **FOR JOURNALS:** D. B. Faria and D. R. Cheriton, "Detecting identity-based attacks in wireless networks using singalprints," in Proceedings of WiSe'06: ACM Workshop on Wireless Security, Sept. 2006, pp. 43–52.

9. **FOR JOURNALS:** H. Hashemi. The Indoor Radio Propagation Channel. Proceedings of IEE, 81(7):943-968, July 1993.

10. **FOR BOOKS:** T. S. Rappaport. Wireless Communications – Principles and Practice. Prentice Hall PTR, 2nd edition, Jan. 2002.

11. **FOR JOURNALS:** K. J. Ellis and N. Serinken. Characteristics of Radio Transmitter Fingerprints. Radio Science, 36:585-598, 2001.

12. **FOR JOURNALS:** M. Gruteser and D. Grunwald. Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis. Mobile Networks and Applications, 10(3):315-325, June 2005.

13. **FOR JOURNALS:** Wireless Client Puzzles in IEEE 802.11 Networks: Security by Wireless, Ivan Martinovic, Frank A. Zdarsky, Matthias Wilhelm, Christian Wegmann, and Jens B. Schmitt

14. **FOR JOURNALS:** Advancing Wireless Link Signatures for Location Distinction, by Junxing Zhangy Mohammad H. Firoozz Neal Patwariz Sneha K. Kaseray

15. **FOR CONFERENCES:** N. Patwari and S. K. Kasera. Robust location distinction using temporal link signatures. In ACM Intl. Conf. on Mobile Computing Networking (Mobicom'07), Sept. 2007.

16. **FOR CONFERENCES:** Z. Li, W. Xu, R. Miller, and W. Trappe. Securing wireless systems via lower layer enforcements. In Proc. 5th ACM Workshop on Wireless Security (WiSe'06), pages 33-42, Sept. 2006.

17. **FOR CONFERENCES:** PARADIS: Physical 802.11 Device Identification with Radiometric Signatures by Vladimir Brik, Suman Banerjee, Marco Gruteser, Sangho Oh

An ID-based Blind Signature Scheme from Bilinear Pairings

B.Umaprasada Rao buprasad@yahoo.co.in
Research scholar
Dept. of Engineering Mathematics
A.U. College of Engineering
Andhra University
Visakhapatnam. A.P, INDIA

K.A.Ajmath
Research scholar kaajmath@yahoo.com
Dept.of Mathematics
Sri Venkateswara University
Tirupati. A.P, INDIA

Dr.P.Vasudeva Reddy vasucrypto@yahoo.com
Associate Professor
Dept. of Engineering Mathematics
A.U. College of Engineering
Andhra University
Visakhapatnam, A.P, INDIA

T.Gowri gowri3478@yahoo.com
Associate Professor
Dept. of Electronics and Communication Engineering
Audisankara College of Engineering & technology
Gudur, Nellore Dist. A.P. INDIA.

Abstract

Blind signatures, introduced by Chaum, allow a user to obtain a signature on a message without revealing any thing about the message to the signer. Blind signatures play an important role in plenty of applications such as e-voting, e-cash system where anonymity is of great concern. Identity based(ID-based) public key cryptography can be a good alternative for certificate based public key setting, especially when efficient key management and moderate security are required. In this paper, we propose an ID-based blind signature scheme from bilinear pairings. The proposed scheme is based on the Hess ID- based digital signature scheme. Also we analyze security and efficiency of the proposed scheme.

Keywords: Public key cryptography, Blind signature scheme, Hess ID based digital signature scheme, Bilinear pairing, CDH problem.

1. INTRODUCTION

Digital signature is a cryptographic tool to authenticate electronic communications. Digital signature scheme allows a user with a public key and a corresponding private key to sign a document in such a way that anyone can verify the signature on the document (using her/his public key), but no one can forge the signature on any other document. This self-authentication is required for some applications of digital signatures such as certification by some authority.

Blind signature is a variant of digital signature scheme. Blind signatures play a central role in digital cash schemes. A user can obtain from a bank a digital coin using a blind signature protocol. The coin is essentially a token properly signed by the bank. The blind signature protocols enable a user to obtain a signature from a signer so that the signer does not learn any information about the message it signed and so that the user can not obtain more than one valid signature after one interaction with the signer. The concept of blind signatures provides anonymity of users in applications such as electronic voting, electronic payment systems etc.

The concept of a blind signature scheme was introduced by Chaum[1], since then many blind signature schemes have been presented in the literature[2,3,4,5].Blind signature scheme allows a user to acquire a signature from the signer without revealing message content for personal privacy. The basic idea is as follows. The user chooses some random factors and embeds them into the message to be signed, while the signer cannot recover the message. Using the blind signature scheme, the user gets the blinded signature and removes the random factors. Then the user outputs a valid signature. This property is very important for implementing e-voting, e-commerce, and e-payment systems, etc.

In public key cryptosystem, each user has two keys, a private key and a public key. The binding between the public key(PK) and the identity(ID) of a user is obtained via a digital certificate. However, in certificate-based system before using the public-key of a user, the participant must first verify the certificate of the user. As a consequence, this system requires a large amount of computing time and storage when the number of users increases rapidly.

In 1984, Shamir [6] introduced the concept of ID-based cryptography to simplify key management procedures in public key infrastructures. Following Joux's [7] discovery on how to utilize bilinear pairings in public key cryptosystems, Boneh and Franklin [8] proposed the first practical ID-based encryption scheme in Crypto 2001. Since then, many ID-based encryption and signature schemes have been proposed that use bilinear pairings. ID-based cryptography helps us to simplify the key management process in traditional public key infrastructures. In ID-based cryptography any public information such as e-mail address, name, etc., can be used as a public key. Since public keys are derived from publicly known information, their authenticity is established inherently and there is no need for certificates in ID-based cryptography. The private key for a given public key is generated by a trusted authority and is sent to the user over a secure channel.

In this paper, a blind signature scheme in the identity-based setting is presented. The scheme is based on the Hess ID-based signature scheme. The proposed signature scheme is validated and its security is proven under the assumption that the hardness of the Computational Diffie-Hellman problem.

The rest of the paper is organized as follows. Section 2 briefly explains the bilinear pairings and some computational problems, on which of our scheme is based .The syntax and security model of ID-based Blind signature Scheme is given in Section3. We then present our ID-based Blind Signature Scheme from bilinear pairings in Section 4. The correctness and security analysis of the proposed scheme is given in Section 5. Section 6 concludes this paper

2. PRELIMANARIES

In this section, we will briefly review the basic concepts on bilinear pairings and some related mathematical problems, and then we present ID-based public key setting from pairings.

2.1 Bilinear Pairings

Let G_1 be a additive cyclic group generated by P whose order is a prime q and G_2 be a multiplicative cyclic group of the same order q. A bilinear pairing is a map $e: G_1 \times G_1 \to G_2$ with the following properties:

1. Bilinear: $e(aP, bQ) = e(P, Q)^{ab}$, for all $P, Q \in G_1$ and all $a, b \in Z_q^*$.
2. Non –degenerate: There exists $P, Q \in G_1$ such that $e(P, Q) \neq 1$.
3. Computable: There is an efficient algorithm to compute $e(P, Q)$, for all $P, Q \in G_1$.

2.2 Computational problems

Now, we give some computational problems, which will form the basis of security for our scheme.

-Discrete Logarithm Problem (DLP): Given two group elements P and Q, find an integer n such that $Q = nP$ whenever such an integer exists.

-Decisional Diffie-Hellman Problem (DDHP): For $a, b, c \in_R Z_q^*$, given P, aP, bP, cP decide whether $c \equiv ab \bmod q$.

-Computational Diffie-Hellman Problem (CDHP): For $a, b, c \in_R Z_q^*$, given P, aP, bP, Compute abP.

We assume through this paper that CDHP and DLP are intractable. When the DDHP is easy but the CDHP is hard on the group G, we call G a Gap Diffe-Hellman (GDH) group. Such groups can be found on super singular elliptic curves or hyper elliptic curves over finite field and the bilinear pairings can be derived from the Weil or Tate pairing.

2.3 ID- based public key setting using pairings

In ID-based public key cryptosystems (IDPKC), everyone's public key is predetermined by information that uniquely identifies them, such as name, social security number, email address, etc., rather than an arbitrary string. This concept was first proposed by Shamir [6]. Since then, many researchers devote their effort on ID-based cryptographic schemes.

ID-based public key setting involves a Key Generation Centre (KGC) and users. The basic operations consists of Setup and Private Key extraction (simply Extract). When we use bilinear pairings to construct IDPKC, **Setup** and **Extract** can be implemented as follows:

Let P be a generator of G_1. Remember that G_1 is an additive group of prime order q and the bilinear pairing is given by $e: G_1 \times G_1 \to G_2$. Define two cryptographic hash functions $H_1: \{0,1\}^* \to G_1^*$, $h: \{0,1\}^* \times G_2 \to Z_q^*$.

-Setup: KGC chooses a random number $s \in Z_q^*$ and sets $P_{pub} = sP$. The center publishes system parameters $params < G_1, G_2, e, P, P_{pub}, H_1, h >$ and keeps $< s >$ as the master-key, which is known only by itself.

-Extract: A user submits his/her identity information ID to KGC. KGC computes the user's public key as $Q_{ID} = H_1(ID)$, and returns $d_{ID} = sQ_{ID}$ to the user as his/her public key.

2.4 Review of Hess-ID- based signature scheme

To prepare for the proposed scheme, we first give a review of the Hess ID-based signature scheme [9].

-Setup: The Private Key Generator (PKG) chooses $s \in_R Z_q^*$ as his master secret key and computes the global public key $P_{pub} = sP$. The PKG also selects a map-to-point hash function $H_1 : \{0,1\}^* \to G_1^*$ and another cryptographic hash function $h : \{0,1\}^* \times G_2 \to Z_q^*$. PKG publishes system parameters $params < G_1, G_2, e, P, P_{pub}, H_1, h >$ and the master key $< s >$ is kept secret.

-Extract: Given the public identity information on ID, compute the secret key for the identity ID as $d_{ID} = sQ_{ID}$. The component $Q_{ID} = H_1(ID)$ plays the role of the corresponding public-key.

-Signature: To sign a message $M \in \{0,1\}^*$, using the secret key d_{ID}, the signer chooses an arbitrary $P_1 \in G_1^*$ and picks a random integer $k \in z_q^*$.

Then signer computes

$$R = e(P_1, P)^k,$$
$$V = h(M, R),$$
$$U = Vd_{ID} + kP_1.$$

The signature on message M is $\sigma = (U,V) \in G_1 \times Z_q^*$.

-Verification: To verify the signature $\sigma = (U,V)$ of an identity ID on a message M, the verifier computes $R = e(U, P)e(Q_{ID}, -P_{pub})^V$. He accepts the signature if and only if $V = h(M, R)$.

3. SYNTAX AND STRUCTURE OF BLIND SIGNATURE SCHEME

The formal definition of a blind signature is presented below.

Blind Signatures: A blind signature scheme consists of three algorithms and two parties (the user and the signer). The details are as follows.

-System Key Generation: This is a probabilistic polynomial time algorithm (PPT algorithm). It takes a security parameter *k* as its input and outputs a pair of public key and private key *{y, x}* for the blind signature scheme, where *x* is preserved secretly by the signer.

-Generation of Blind Signatures: This is an interactive and probabilistic polynomial time algorithm protocol, which is operated by the user and the signer. The user first blinds the message *m* and obtains a new version m' of *m*, and then sends it to the signer. The latter utilizes his/her private key to sign on m' and obtains S', and then sends it to the user. The user unblinds it to obtain S which is a blind signature on *m*.

-Verification of Blind Signatures: This is a deterministic polynomial time algorithm. Given a message *m* and its alleged blind signature S, anyone who knows the public key *can* verify the validity of S. If it is valid, then the algorithm outputs '1'; otherwise outputs '0'.

The blindness property of a signature scheme may be formally defined as follows: A blind signature scheme possesses the blindness property, if the signer's view (m', S') and the message-signature pair (m, S) are statistically independent.

A secure blind signature scheme must satisfy the following three requirements:

-Correctness: If the user and the signer both comply with the algorithm of blind signature generation, then the blind signature S will be always accepted.

-Unforgeability of Valid Blind Signatures: It is with respect to the user especially, i.e. the user is not able to forge blind signatures which are accepted by the algorithm of Verification of Blind Signatures.

-Blindness: While correctly operating one instance of the blind signature scheme, let the output be (*m, S*) (i.e. message-signature pair), and the view of the protocol V. At a later time, the signer is not able to link V to (*m, S*).

4. PROPOSED ID-BASED BLIND SIGNATURE SCHEME:

In this section, we present an ID-based blind signature scheme from the bilinear pairings.

Setup: The PKG chooses $s \in Z_q^*$ as his master key and computes the global public key P_{pub} as sP. The PKG also selects a map-to-point hash function $H_1 : \{0,1\}^* \to G_1^*$ and another cryptographic hash function $h : \{0,1\}^* \times G_2 \to Z_q^*$. PKG publishes system parameters $params < G_1, G_2, e, P, P_{pub}, H_1, h >$ and keeps the master key $< s >$ as secret.

Extract: Given signer's public identity $ID \in \{0,1\}^*$, compute the public key $Q_{ID} = H_1(ID)$ and the corresponding private key $d_{ID} = sQ_{ID}$.

Initialization: The signer randomly chooses $k \in Z_q^*$, compute $R = e(P, P)^k$ and sends R to the user as a commitment.

Blinding: The user randomly chooses $a, b \in Z_q^*$ as blinding factors, compute $R' = e(bQ_{IDs} + aP, \mathrm{P_{pub}}).R$, $V = h(m, \mathrm{R'}) + \mathrm{b}$ and sends V to the signer.

Signing: The signer computes $S = Vd_{IDs} + kP$, and send S to the user

Unblinding: The user compute $S' = S + aP_{pub}$, $V' = V - b$ and outputs (m, S', V'), then (S', V') is the blind signature of the message m.

Verification: Accept the signature if and only if $V' = h(m, e(S', P).e(Q_{IDs}, P_{pub})^{-V'})$.

5. Analysis of the proposed scheme

5.1 Correctness

The following equations give the correctness of the proposed scheme.

$$
\begin{aligned}
& h(m, e(S', P).e(Q_{IDs}, P_{pub})^{-V'}) \\
& = h(m, e(S + aP_{pub}, P).e(Q_{IDs}, P_{pub})^{-V'}) \\
& = h\left(m, e(S, P).e(aP_{pub}, P).e(Q_{IDs}, P_{pub})^{-V'}\right) \\
& = h\left(m, e(Vd_{IDs} + kP, P).e(aP_{pub}, P).e(Q_{IDs}, P_{pub})^{-V'}\right) \\
& = h\left(m, e(Vd_{IDs}, P).e(kP, P).e(aP_{pub}, P).e(Q_{IDs}, P_{pub})^{-V+b}\right)
\end{aligned}
$$

$$= h\left(m, e(d_{IDs}, P)^{V}.e(P,P)^{k}.e\left(aP, P_{pub}\right).e\left(Q_{IDs}, P_{pub}\right)^{-V}.e\left(Q_{IDs}, P_{pub}\right)^{b}\right)$$

$$= h\left(m, e\left(Q_{IDs}, P_{pub}\right)^{V}.e\left(P,P\right)^{k}.e\left(aP, P_{pub}\right).e\left(Q_{IDs}, P_{pub}\right)^{-V}.e\left(Q_{IDs}, P_{pub}\right)^{b}\right)$$

$$= h\left(m, e\left(P,P\right)^{k}.e\left(aP + bQ_{IDs}, P_{pub}\right)\right)$$

$$= h\left(m, R'\right)$$

$$= V - b$$

$$= V'.$$

5.2 Security

In the following, we will show that our ID-based Blind signature scheme satisfies all the requirements stated in Section 3.

Blindness property: To prove the blindness we show that given a valid signature $\left(m, S', V'\right)$ and any view $\left(R, V, S\right)$, there always exists a unique pair of blinding factors $a, b \in Z_q^*$. Since the blinding factors $a, b \in Z_q^*$ are chosen randomly, the blindness of the signature scheme naturally satisfies. We can find more formal definition about the blindness [10, 11, 12, 13].

Given a valid signature $\left(m, S', V'\right)$ and any view $\left(R, V, S\right)$, then the following equations must hold for $a, b \in Z_q^*$:

$$R' = e\left(bQ_{IDs} + aP, P_{pub}\right).R \qquad (1)$$

$$\mathrm{V}=\mathrm{h}\left(\mathrm{m},\mathrm{R}'\right)+b \qquad (2)$$

$$\mathrm{S}' = S + aP_{pub} \qquad (3)$$

$$\mathrm{b}=\mathrm{V}-\mathrm{h}\left(\mathrm{m},\mathrm{R}'\right) \text{ and } \mathrm{aP}_{\mathrm{pub}}=\mathrm{S}'-\mathrm{S}$$

It is obvious that $a, b \in Z_q^*$ is existed uniquely from (2) and (3). Next we show that such $a, b \in Z_q^*$ satisfy the first equation too. Obviously, due to the non-degenerate of the bilinear pairings we have $R' = e\left(bQ_{IDs} + aP, P_{pub}\right).R \Leftrightarrow e(R', P_{pub}) = e\left(e\left(bQ_{IDs} + aP, P_{pub}\right), P_{pub}\right)$. So we only need to show that such a and b satisfy $e(R', P_{pub}) = e\left(e\left(bQ_{IDs} + aP, P_{pub}\right), P_{pub}\right)$.

We have

$$e\left(e\left(bQ_{IDs} + aP, P_{pub}\right).R, P_{pub}\right) =$$

$$=\mathrm{e}\left(\mathrm{e}\left(\mathrm{bd}_{\mathrm{IDs}} + aP_{pub}, P\right).R, P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{e}\left(\left(\mathrm{V}-\mathrm{h}\left(\mathrm{m},\mathrm{R}'\right)\right)d_{IDs}, P\right).e\left(aP_{pub}, P\right).R, P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{e}\left(\left(\mathrm{V}-\mathrm{h}\left(\mathrm{m},\mathrm{R}'\right)\right)d_{IDs}, P\right)e\left(S' - S, P\right).R, P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{e}\left(\mathrm{Vd}_{\mathrm{IDs}}, P\right)e\left(-h\left(m, R'\right)d_{IDs}, P_{pub}\right)e\left(S', P\right)e\left(S, P\right)^{-1}.R, P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{e}\left(\mathrm{S}-\mathrm{kP},\mathrm{P}\right)e\left(-h\left(m, R'\right)d_{IDs}, P\right)e\left(S', P\right)e\left(S, P\right)^{-1}.R, P_{pub}\right)$$

$$=\mathrm{e}\left(e\left(-h\left(m, R'\right)d_{IDs}, P\right)R'e\left(Q_{IDs}, P_{pub}\right)^{V_i}, P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{e}\left(-\mathrm{h}\left(\mathrm{m},\mathrm{R}'\right)Q_{IDs}, P_{pub}\right)R'e\left(h\left(m, R'\right)Q_{IDs}, P_{pub}\right), P_{pub}\right)$$

$$=\mathrm{e}\left(\mathrm{R}', P_{pub}\right)$$

Thus the blinding factors always exist which lead to the same relation defined in the signature issuing protocol.

Unforgeability: Assume that ***A*** is the adversary (he/she can be a user or any third party) holding the system parameters $params < G_1, G_2, e, P, P_{pub}, H_1, h >$ and the identity public key Q_{IDs} of the signer IDs . ***A*** tries to forge a valid message-signature of the signer.

First, we assume that ***A*** performs the ID attack, i.e. ***A*** queries **Extract** qE (qE>0) times with (PARAMS, $ID_i \neq ID$) for i=1….qE. **Extract** returns to ***A*** the qE corresponding secret key d_{IDs_i} . We assume that qE is limited by a polynomial in k. If ***A*** can get a $\left(ID_{s_i}^{'}, d_{IDs_i}^{'} \right)$ such that $H_1\left(ID_{s_i}^{'} \right) = H_1\left(IDs \right) = Q_{IDs}$, then he/she can forge a valid blind signature of the signer ID. But since H_1 is random oracle, **Extract** generates random numbers with uniform distributions. This means that ***A*** learns nothing from query results.

Next we assume that ***A*** had interacted with the signer ID, and let $\left(R, V, S \right)$ be the view in the blind signature issuing phase. Since $S = Vd_{IDs} + kP$ and ***A*** knows S,V, from S to get d_{IDs} , ***A*** must know k, but k is chosen randomly by signer. ***A*** Knows $R = e\left(P, P \right)^k$, but from R to get k, this is CDHP in G_1 . We assume that CDHP in G_1 in intractable, so ***A*** cannot get the private information of the signer at the blind signature issuing phase.

On the other hand, the signature and the verifying equation are same as Hess ID- based signature scheme. For any message m, if ***A*** can construct $S^{'} and\ V^{'}$ such that $V^{'} = h\left(m, e\left(S^{'}, P \right) e\left(Q_{IDs}, P_{pub} \right)^{-V^{'}} \right)$, then ***A*** can forge a valid signature of Hess ID-based signature scheme on the message m. Due to Hess proof on their ID-based signature scheme (i.e., Hess ID-based signature scheme is proven to be secure against existential forgery on adaptive chosen message and ID attacks, under the hardness assumption of CDHP and the random oracle model), we claim that this attack is impossible.

Efficiency: We compare our blind signature scheme with the Zhang- Kim ID-based blind signature scheme [11] from computation overhead and summarize the result in Table1. We denote p_a the pairing operation, p_m the point scalar multiplication on G_1, ***Ad*** the point addition on G_1, ***Mu*** the multiplication in Z_q^*, and ***Mu*** G_2 the multiplication in G_2, ***Me*** exponentiation in G_2.

Schemes	Blind signature issuing	Verification
Proposed scheme	User : 1***Pa***+3***Pm***+1***Mu***+3***Ad*** Signer: 1***Pa***+1***Me***+2***Mu***+1***Ad***	2***Pa***+1***Me***
The scheme [11]	User : 1***Pa***+3***Pm***+3***Ad*** Signer: 3***Pm***+1***Ad***	2***Pa***+1***Pm***+1***Mu*** G_2

Table 1. Comparision of our scheme with Zhang-Kim scheme

The efficiency of the system of paramount importance when the number of verifications is considerably large (e.g., when a bank issues a large number of electronic coins and the customer wishes to verify the correctness of the coins). Assuming that $\left(S_1^{'}, V_1^{'} \right), \left(S_2^{'} V_2^{'} \right), - - -, \left(S_n^{'}, V_n^{'} \right)$ are ID-based blind signatures on messages $m_1, m_2, - - -, m_n$ respectively, which are issued by the signer with identity ID. The verification of each signature is as follows:

$V_i^{'} = h\left(m_i, e\left(S_i^{'}, P \right) e\left(Q_{IDs}, -P_{pub} \right)^{V_i^{'}} \right)$, *for* i=1,2,...,n .

To verify these signatures individually, our scheme requires only (n+1) pairing operations, where as the Zhang-Kim scheme requires 2n pairing operations. So, with the proposed scheme we can save (n-1) pairing operations. In particular, here, we consider only computations of pairing operation (Pa), we need not consider the remaining operations as they are cheaper than the computation of pairings. We note that the computation of pairing is the most time consuming. Although there has been many papers discuss the complexity of pairings and how to speedup the pairing computation [14, 15], the computation of pairing is still time consuming.

6. CONCLUSIONS

In this paper, we proposed an ID-based blind signature scheme from bilinear pairings. The proposed scheme is based on Hess ID-based signature scheme with the assumption CDH Problem is hard. We have discussed the correctness and security analysis of the proposed scheme. The proposed scheme is efficient when the number of blind signature verifications is considerably large.

REFERENCES

[1] D. Chaum, “Blind signatures for untraceable payments”, In Proc. CRYPTO 82, pp.199-203, NY, Plenum, 1983.

[2] T.Okamoto, “Provable, Secure and Practical Identification Schemes and Corresponding signature schemes”, In Advances in Cryptology-CRYPTO 1984, Springer-Verlag, LNCS 740, pp.31-53,1992.

[3] D.Pointcheval and J.Stern, “Provably Secure Blind signature Schemes”, In Advances in Cryptology – ASIACRYPT 1992, Springer-Verlag, LNCS 1163, pp.252-26,1992.

[4] D.Pointcheval and J.Stern, “New Blind Signature Signatures Equilent to Factorization”, In proceedings of the 4th ACM Conference on Computer and Communications Security, pp.92-99, Zurich, Switzerland, 1997.

[5] C.P. Schnorr, “Efficient Identification and Signatures for Smart cards”, In G.Brassard(ed), In proceedings of CRYPTO 1989, Springer-Verlag,LNCS 435,pp.239-252,1990.

[6] A,Shamir, “Identity-based cryptosystems and signature schemes”, In Proc. of CRYPTO’84, LNCS 196, pp. 47-53 Springer-verlag, 1984.

[7] A.Joux, “A one round protocol for tripartite diffie-Hellman” In proc.of ANTS-IV, LNCS 1838, pp.385-394, Springer-Verlag, 2000.

[8] D.Boneh and M.Franklin, “Identity-based encryption from the Weil pairing”, In Proc.of CRYPTO’01, LNCS 2139, pp.213-229, Springer-verlag, 2001.

[9] F. Hess, “Efficient identity based signature schemes based on pairings”, SAC 2002, LNCS2595, pp.310-324, Springer-Verlag, 2002.

[10] A. Juels, M. Luby and R. Ostrovsky, “Security of blind digital signatures”, Advances in Cryptology-Crypto 97, LNCS 1294, pp.150-164, Springer-Verlag, 1997.

[11] F. Zhang and K. Kim, “ID-based blind signature and ring signature from pairings”, Proc. of Asiacrpt 2002, LNCS 2501, pp. 533-547, Springer-Verlag, 2002.

[12] F. Zhang and K. Kim, “*Efficient ID-based blind signature and proxy signature from bilinear pairings”*, ACISP 03, LNCS 2727, pp. 312-323, Springer-Verlag, 2003.

[13] D. Pointcheval and J. Stern, “Security arguments for digital signatures and blind Signatures”, Journal of Cryptology, Vol.13, No.3, pp.361-396, 2000.

[14] P.S.L.M.Barreto, H.Y.Kim, B.Lynn and M.Scott, “Efficient algorithms for pairing- based cryptosystems”, Advances in Cryptology-Crypto 2002, LNCS 2442, pp.354-368, Springer-Verlag, 2002.

[15] S.D.Galbraith, K. Harrison, and D.Soldera, “Implementing the Tate pairing”, ANTS 2002, LNCS 2369, pp.324-337, Springer-Verlag, 2002.

Recognition of Non-Compound Handwritten Devnagari Characters using a Combination of MLP and Minimum Edit Distance

Sandhya Arora sandhyabhagat@yahoo.com
Assistant Professor, Department of CSE & IT
Meghnad Saha Institute of Technology
Kolkata-107, India

Debotosh Bhattacharjee debotoshb@hotmail.com
Department of Computer Science and Engg.
Jadavpur University
Kolkata-107, India

Mita Nasipuri mitanasipuri@yahoo.com
Department of Computer Science and Engg.
Jadavpur University
Kolkata-107, India

D. K. Basu dipakkbasu@gmail.com
Department of Computer Science and Engg.
Jadavpur University
Kolkata-107, India

M. Kundu mkundu@cse.jduv.ac.in
Department of Computer Science and Engg.
Jadavpur University
Kolkata-107, India

Abstract

This paper deals with a new method for recognition of offline Handwritten non-compound Devnagari Characters in two stages. It uses two well known and established pattern recognition techniques: one using neural networks and the other one using minimum edit distance. Each of these techniques is applied on different sets of characters for recognition. In the first stage, two sets of features are computed and two classifiers are applied to get higher recognition accuracy. Two MLP's are used separately to recognize the characters. For one of the MLP's the characters are represented with their shadow features and for the other chain code histogram feature is used. The decision of both MLP's is combined using weighted majority scheme. Top three results produced by combined MLP's in the first stage are used to calculate the relative difference values. In the second stage, based on these relative differences character set is divided into two. First set consists of the characters with distinct shapes and second set consists of confused characters, which appear very similar in shapes. Characters of distinct shapes of first set are classified using MLP. Confused

characters in second set are classified using minimum edit distance method. Method of minimum edit distance makes use of corner detected in a character image using modified Harris corner detection technique. Experiment on this method is carried out on a database of 7154 samples. The overall recognition is found to be 90.74%.

Keywords :- Harris corner detector, Classification, Multilayer Perceptron, feature extraction, Minimum Edit Distance method.

1. INTRODUCTION

Optical Character Recognition (OCR) is the most crucial part of Electronic Document Analysis Systems. The solution lies in the intersection of the fields of pattern Recognition, image and natural language processing. Although there has been a tremendous research effort, the state of the art in the OCR has only reached the point of partial use in recent years. Nowadays, clearly printed texts in documents with simple layouts can be recognized reliably by off-the-shelf OCR software. There is only limited success in handwriting recognition, particularly for isolated and neatly hand-printed characters and words for limited vocabulary. However, in spite of the intensive effort of more than thirty years, the recognition of free style handwriting continues to remain in the research arena.

An OCR has variety of commercial and practical applications in processing bank cheques, government records, credit card imprints and postal code reading, reading commercial forms, manuscripts and their archival etc. Such a system facilitates a key board less user computer interaction also the text which is either printed or handwritten can be directly transferred to the machine. An elaborate list of OCR applications has been presented by Govindan[1].

Historically, Devnagari is the script used by Sanskrit, Hindi, Marathi and Nepali. Hindi is the world's third most commonly used language after Chinese and English. Thus research on Devnagari script is gaining importance because of their large market potential. With the explosion of information technology there has been a dramatic increase of research in this field since the beginning of 1980.

OCR work on printed Devnagari Script started in early 1970's. Sinha and Mahabala[2] presented a syntactic pattern analysis system with an embedded picture language for the recognition of handwritten and machine printed Devnagari characters. Veena described Devnagari OCR in her doctoral Thesis [3]. Performance of 93% accuracy at character level is reported after post processing. Pal and Chaudhuri [4] reported a complete OCR system for printed Devnagari with 96% accuracy. Hanmandlu and Murthy [5,6] proposed a Fuzzy model based recognition of handwritten Devnagari numerals and characters and they obtained 92.67% accuracy for Handwritten Devnagari numerals and 90.65% accuracy for Handwritten Devnagari characters. Sinha et al [2,7] have reported various aspects of Devnagari script recognition. Bajaj et al [8] employed three different kinds of features namely, density features, moment features and descriptive component features for classification of Devnagari Numerals. They proposed multi-classifier connectionist architecture for increasing the recognition reliability and they obtained 89.6% accuracy for handwritten Devnagari numerals. Kumar and Singh [9] proposed a Zernike moment feature based approach for Devnagari handwritten character recognition. They used an artificial neural network for classification. Sethi et. al. [10,11] has described Devnagari numeral recognition based on structural approach. The primitive used are horizontal line segment, vertical line segment, right slant and left slant. A decision tree is employed to perform analysis based on presence/absence of these primitives and their interconnection. A similar strategy was applied to constrained hand printed Devnagari character. Bansal et. al. [12], have used translation and scaling invariant moments and structural description of a character and reported accuracy of 93%

at character level of printed Devnagari characters. Bhattacharya et al [13] proposed a Multi-layer perceptron (MLP) neural network based classification approach for the recognition of Devnagari handwritten numerals and obtained 91.28% results. They considered a multi- resolution features based on wavelet transform in their proposed system. N. Sharma and U. Pal [14,15,16] proposed a directional chain code features based quadratic classifier and obtained 80.36% accuracy for handwritten Devnagari characters and 98.86% accuracy for handwritten Devnagari numerals. Few more work[17,18,19] is going on for handwritten devnagari characters. In our previous work [20] we proposed a MLP based classifier designed with three different features namely: Intersection, Shadow, Chain code histogram. The recognition accuracy 69.37% achieved by considering top 1 choices results on 4900 samples. In this paper, we purpose a system based on MLP and minimum edit distance for the recognition of offline Handwritten Devnagari character recognition.

While a large amount of literature is available for recognition of English script, relatively less work has been reported for the recognition of Indian languages. Main reason for this slow development could be attributed to the complexity in the shapes of Indian scripts, and also the large set of different patterns that exists in these languages, as opposed to English.

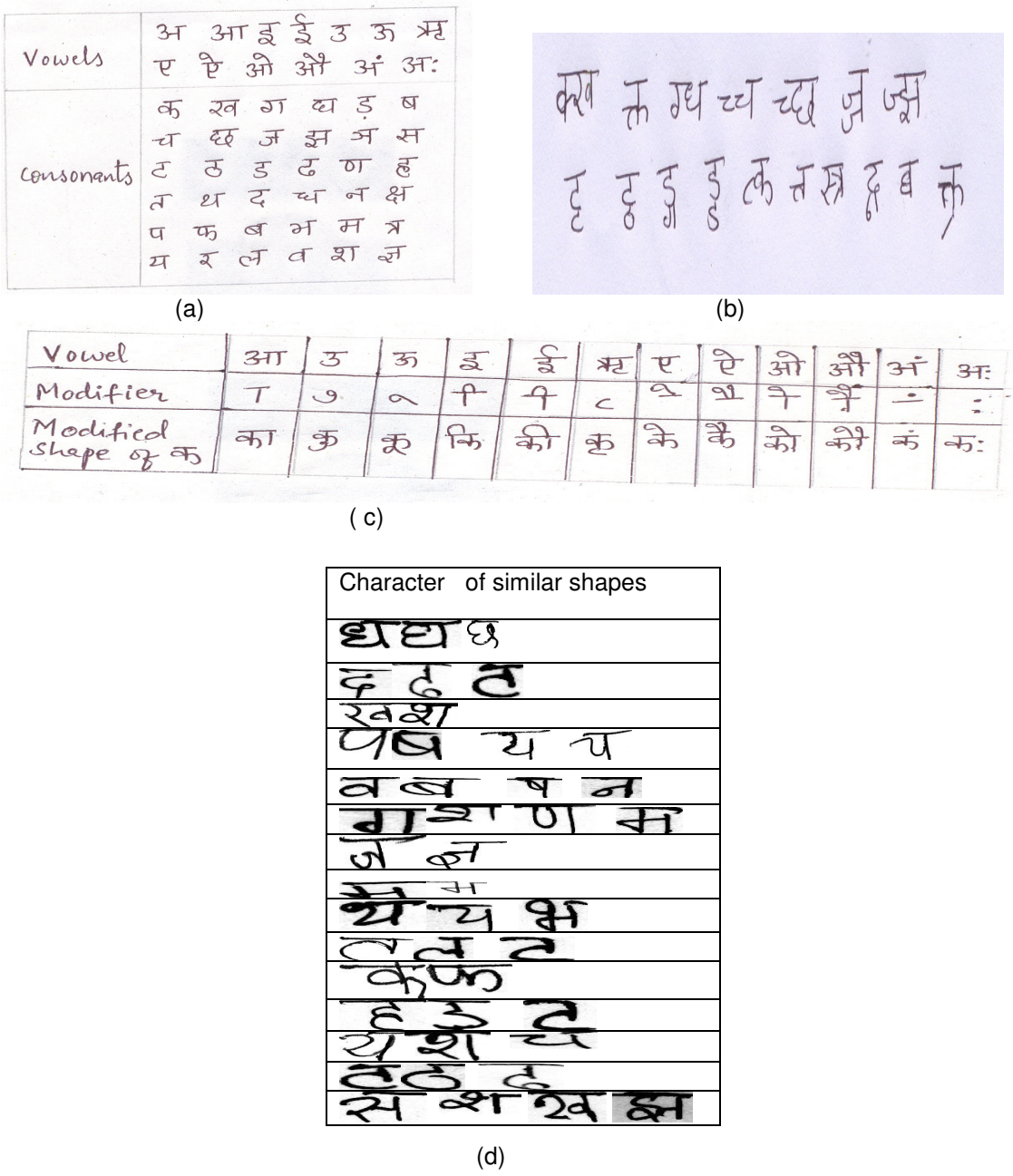
Most of the work reported above [2,4,12] were on printed Devnagari characters. For handwritten Devnagari character recognition system [3,5,6,9] , accuracy reported is not high and dataset used are not large. We worked on 7154 samples. As no standard database is available for handwritten Devnagari characters, we created some samples of our own and some we collected from ISI, Kolkata.

Rest of the paper is organized as follows. In section 2, peculiarities of Devnagari Script are discussed. Overall approach used, is discussed in section 3. Feature extraction techniques are reported in section 4. Section 5, deals with the classifiers used for the recognition purpose. The experimental results are discussed in section 6.

2. PECULIARITIES OF DEVNAGARI SCRIPT

Devnagari script is different from Roman script in several ways. This script has two-dimensional compositions of symbols: core characters in the middle strip, optional modifiers above and/or below core characters. Two characters may be in shadow of each other. While line segments (strokes) are the predominant features for English, most of the characters in Devnagari script is formed by curves, holes, and also strokes. In Devnagari language scripts, the concept of upper-case, the lower-case characters is absent. However the alphabet itself contains more number of symbols than that of English.

Devnagari script has around 13 vowels and 36 consonants resulting in a total of 49 basic characters as shown in Figure 1a. Vowels occur either in isolation or in combination with consonants. Apart from vowels and consonants characters called basic characters, there are compound characters in Devnagari script alphabet system, which are formed by combining two or more basic characters (shown in Figure 1b). The shape of compound character is usually more complex than the constituent basic characters. Coupled to this in Devnagari script there is a practice of having more than twelve forms each for 36 consonants , giving rise to modified shapes which, depending on whether the vowel modifier is placed to the left, right, top or bottom of the consonants as shown in Figure 1c. They are called modified characters. The net result is that there are several thousand different shapes or patterns in the script, some of them are almost similar in shapes. Even with the basic character same problem about their shapes exists. Some basic characters have distinct shapes (Figure 1a) and can be identified with certainty. Some groups of basic characters have almost similar shapes (Figure 1d) causing confusion and need special attention in recognizing them. The most of the confusing pair of Devnagari characters are from the Figure specified in 1d.



(a) (b) (c) (d)

FIGURE 1. Samples of handwritten devnagari a)Vowels and Consonants b) some compound characters c) Modifiers with their corresponding vowel and a sample character image of "ka" modified with modifier d) Confusing characters

3. OVERALL APPROACH

Scheme of our proposed method is shown in Figure 2. We perform scaling of character bitmap and after that we extract two different features. First, 24 shadow features are extracted from eight octants of the scaled binarized character image. Second, 200 chain code histogram features are obtained by first detecting the contour points of original scaled binarized character image, and dividing the contour image into 25 segments. For each segment chain code histogram features are obtained. Here, the character recognition is done in two stages. In the first stage, two MLP's are designed using these two different feature sets. Outputs of individual MLP classifiers [20] are combined using weighted majority scheme and the character classes corresponding to top three

values are considered. A relative difference measure is computed from these top three values. If this measure is greater than some threshold value, we infer that the top choice determines the class of the sample character with certainty. On the other hand, if the relative difference measure is less than or equal to the threshold value, we infer that the sample character belongs to a group of confusing character identified by top three choices. In the second stage , the true class of the sample character belonging to a confusing group are identified by applying minimum edit distance method, on detected corners of the sample character using a modified form of Harris corner detector.

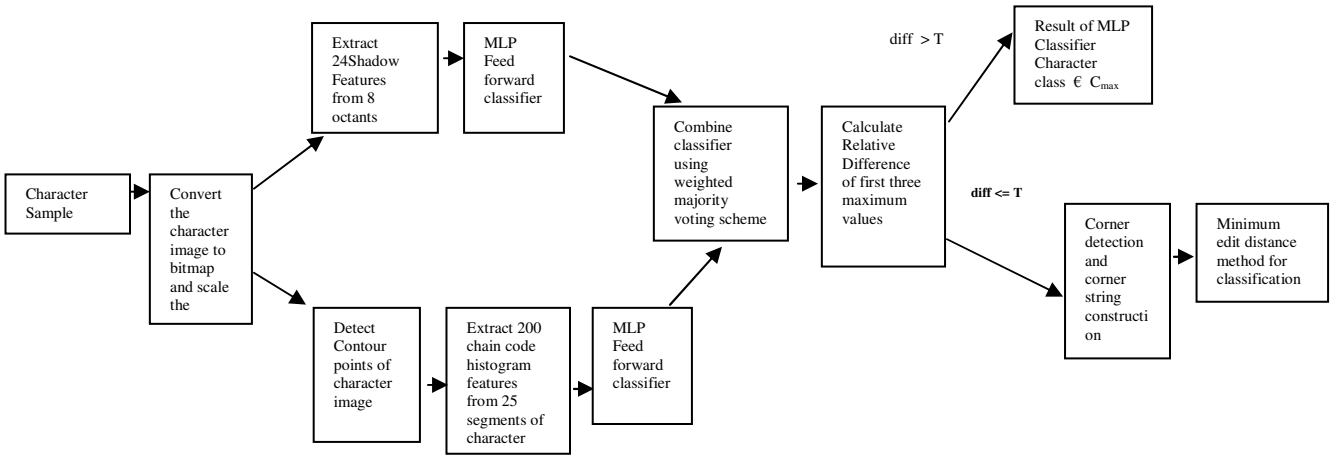


FIGURE 2. Overall scheme of proposed Technique

4. FEATURE EXTRACTION

In the following section we give a brief description of the two feature sets used in our proposed system. Shadow features are extracted from scaled binarized character image. Chain code histogram features are extracted by chain coding the contour points of the scaled character binarized image.

4.1 Shadow Features of a character image

Shadow is basically the length of the projection on the sides as shown in Figure 3. For computing shadow features [21], the rectangular boundary enclosing the character image is divided into eight octants. For each octant shadows or projections of character segment on three sides of the octant dividing triangles are computed so, a total of 24 shadow features are obtained. Each of these features is divided by the length of the corresponding side of the triangle to get a normalized value.

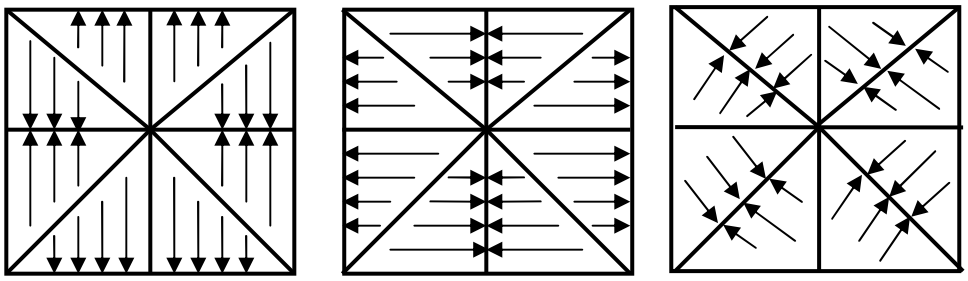


FIGURE 3. Shadow features

4.2 Chain Code Histogram of Character Contour

Chain code provides the direction of the next pixel in the image. Given a scaled binary image, we first find the contour points of the character image. We consider a 3 × 3 window surrounded by the object points of the image. If any of the 4-connected neighbor points is a background point then the object point (P), as shown in Figure 4 is considered as contour point.

	X	
X	P	X
	X	

FIGURE 4. Contour point detection

The contour following procedure is used to trace the contour and a contour representation called "chain coding" as proposed by Freeman [22], shown in figure 5a. Each pixel of the contour is assigned a different code that indicates the direction of the next pixel that belongs to the contour in some given direction. Chain code provides the points in relative position to one another, independent of the coordinate system. In this methodology of using a chain coding of connecting neighboring contour pixels, the points and the outline coding are captured. Contour following procedure may proceed in clockwise or in counter clockwise direction. Here, we have chosen to proceed in a clockwise direction.

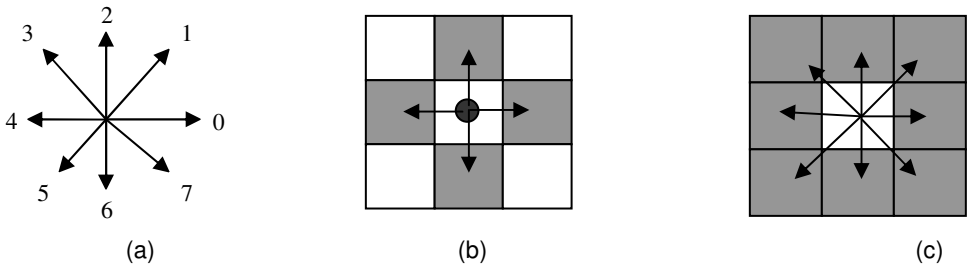


FIGURE 5. Chain Coding: (a) direction of connectivity, (b) 4-connectivity, (c) 8-connectivity. Generate the chain code by detecting the direction of the next-in-line pixel

The chain code for the character contour will yield a smooth, unbroken curve as it grows along the perimeter of the character and completely encompasses the character. When there is multiple connectivity in the character, then there can be multiple chain codes to represent the contour of the character. We chose to move with minimum chain code number first.

We divide the contour image in 5 × 5 blocks. In each of these blocks, the frequency of the direction code is computed and a histogram of chain code is prepared for each block. Thus for 5 × 5 blocks we get 5 × 5 × 8 = 200 features for recognition.

5. CHARACTER RECOGNITION

Devnagari character recognition is done in two stages using Multilayer Perceptron (MLP) and method of minimum edit distance, each applied at different stage. We divided the character set in two set using relative difference value discussed in section 6.1. First set consists of characters with certainty and second set consists of confused characters. Characters of first set are classified using MLP discussed in section 5.1 and characters of second set are classified using minimum edit distance method applied on corner detected character image, is discussed in section 5.2. We rejected some samples in MLP classifier and the corner detection method with minimum edit distance is applied to these confused rejected samples to increase the accuracy.

5.1 MLP Classifier

We designed different MLP with 3 layers including one hidden layer for two different feature sets consisting of 24 shadow features and 200 chain code histogram features. The experimental results obtained while using these features for recognition of handwritten Devnagari characters is

presented in section 6. At this stage all characters are non-compound, single characters so no segmentation is required.

Each MLP is trained with Backpropagation learning algorithm with momentum [13]. It minimizes the sum of squared errors for the training samples by conducting a gradient descent search in the weight space. As activation function we used sigmoid function. Learning rate and momentum term are set to 0.8 and 0.7 respectively. As activation function we used the sigmoid function. Numbers of neurons in input layer of MLPs are 24 or 200, for shadow features and chain code histogram features respectively. Number of neurons in Hidden layer is not fixed, we experimented on the values between 20-70 to get optimal result and finally it was set to 30 and 70 for shadow features and chain code histogram features respectively. The output layer contained one node for each class., so the number of neurons in output layer is 49.

5.1.1 Combining Multiple Classifiers

The ultimate goal of designing pattern recognition system is to achieve the best possible classification performance. This objective traditionally led to the development of different classification scheme for any pattern recognition problem to be solved. Classifiers producing crisp, single class labels (SCL) provide the least amount of useful information for the combination process. However, they are still well performing classifiers and the sets of patterns misclassified by the different classifiers does not necessarily overlap. This suggested that different classifier designs potentially offered complementary information about the pattern to be classified which could be harnessed to improve the performance of the selected classifier. So instead of relying on a single decision making scheme we can combine classifiers.

Voting strategies can be applied to a multiple classifier system assuming that each classifier gives a single class label as an output. There are a number of approaches to combination of such uncertain information units in order to obtain the best final decision [21]. We applied the voting definition. For convenience let the output of the classifiers form the decision vector d defined as $d = [d_1, d_2, ..., d_m]^T$ where $d_i \in \{c_1, c_2, ..., c_m, r\}$, c_i denotes the label of the i-th class and r the rejection of assigning the input sample to any class. Let binary characteristic function be defined as follows:

$$B_j(c_i) = \begin{cases} 1 & \text{if } d_j = c_i \\ 0 & \text{if } d_j \neq c_i \end{cases}$$

Then the general voting routine can be defined as:

$$E(d) = \begin{cases} c_i & \text{if } \underset{t \in \{1,...,m\}}{\forall} \sum_{j=1}^{n} B_j(c_t) \leq \sum_{j=1}^{n} B_j(c_t) \geq \alpha.m + k(d) \\ r & \text{otherwise} \end{cases}$$

Where α is a parameter and k(d) is a function that provides additional voting constraints. The most conservative voting rule is given if k(d) =0 and α =1 ,meaning that the class is chosen when all classifiers produce the same output. This rule can be liberalized by lowering the parameter α. Function k(d) is usually interpreted as a level of objection to the most often selected class and refers mainly to the score of the second ranked class. This option allows adjusting the level of collision that is still acceptable for giving correct decision.

$$\alpha = \frac{p_k}{\sum_{k=1}^{2} p_k}$$

In our present work, we have used two MLP classifiers with recognition performances

p_1=73.33% (Success rate of MLP1 using chain code histogram feature)
p_2= 68.10% (Success rate of MLP2 using shadow feature)

5.2 Confused characters classification
For classifying characters of similar shapes we took different approach. We detected corners in character image using modified form of Harris corner detector [23], discussed in section 5.2.1 and 5.2.2. After detecting corners we divided the character image in 25 segments and in each segment we counted the number of corners. On this corner detected string we applied minimum edit distance discussed in section 5.2.3

5.2.1 Corner Detection Algorithm
Corner Detector can be considered interest point corner detector as they assign a measure of cornerness to all pixels in an image. The brute force method of comparing every pixel in the two images is computationally prohibitive. Intuitively one can relate two images by matching only locations in the image that are in some way interesting. Such points are referred to as interest points and are located using an interest point detector. Finding a relationship between images is then performed using only these points. This drastically reduces the required computation time. Corner points are interesting as they are formed from two or more edges and edges usually define the boundary between two different objects or parts of the same object.

Many Corner Detectors[23] are available but we chose Harris/Plessey corner detector with some modification. This corner detector is computationally demanding, but directly addresses many of the limitations of the other corner detectors.

Algorithm for detecting corners using Harris Corner detector in all confused characters is as follows:-

1. For each pixel (x, y) in the image calculate the autocorrelation matrix M:-
   $$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$
   Where $A = \left(\frac{\partial I}{\partial x}\right)^2 \otimes w$
   $B = \left(\frac{\partial I}{\partial y}\right)^2 \otimes w$
   $C = \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right)^2 \otimes w$
   $\otimes$ is the convolution operator , W is the Gaussian window of size 5
   $\left(\frac{\partial I}{\partial x}\right), \left(\frac{\partial I}{\partial y}\right), \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right)$ are horizontal, vertical and diagonal intensity variations respectively.

2. Construct the cornerness map by calculating the cornerness measure C(x, y) for each pixel (x, y):
   $C(x,y) = det(M)-k(trace(M))^2$
   $det(M) = \lambda_1 \lambda_2 = AB-C^2$
   $trace(M) = \lambda_1 + \lambda_2 = A + B$
   k = constant
   λ_1 and λ_2 eigenvalues of M

3. Threshold the interest map by setting all C(x, y) below a threshold T to zero.
4. Perform non-maximal suppression to find local maxima.

The basic idea behind detecting corners of an image in this algorithm is to estimate the measurement of local autocorrelation so intensity variation is measured in different directions for that purpose. Intensity variation calculation for Harris operator is approximated using the gradient of the image so the intensity variation in horizontal, vertical and diagonal direction can be written as a function of the gradient of the image. Here A, B, C in step 1 is defined as:-

A= Weighted horizontal intensity variation
B= Weighted vertical intensity variation
C= Weighted diagonal intensity variation

Here the intensity variation is convolved with the Gaussian window. Gaussian window is a circular window that puts more weight on measurement made closer to the centre of the window. This is desirable so that the Euclidean distance from the centre pixel to the edge is same in all directions. This improves the estimate of the local intensity variation.

Weighted horizontal intensity variation A is calculated by convolving below specified window in figure 6 with Gaussian window of size 5. Vertical Intensity Variation B is calculated by convolving below specified window in figure 7 with Gaussian window of size 5. Diagonal Intensity Variation in upward direction C is calculated by convolving below specified window in figure 8 with Gaussian window of size 5.

A1	A2 B1	A3 B2	B3
A4	A5 B4	A6 B5	B6
A7	A8 B7	A9 B8	B9

FIGURE 6. Horizontal intensity variation window(A)

B1	B2	B3
A1 B4	A2 B5	A3 B6
A4 B7	A5 B8	A6 B9
A7	A8	A9

FIGURE 7. Vertical intensity variation window(B)

	B1	B2	B3
A1	A2 B4	A3 B5	B6
A4	A5 B7	A6 B8	B9
A7	A8	A9	

FIGURE 8. Diagonal upward intensity variation window(C)

W1 .004	W2 .015	W3 .026	W4 .015	W5 .004
W6 .015	W7 .059	W8 .095	W9 .059	W10 .015
W11 .026	W12 .095	W13 .15	W14 .095	W15 .026
W16 .015	W17 .059	W18 .095	W19 .059	W20 .015
W21 .004	W22 .015	W23 .026	W24 .015	W25 .004

FIGURE 9. Gaussian Window of size 5

5.2.2 Modification to Harris Corner Detection Algorithm

We modified this algorithm because it just considers weighted intensity variations in three directions i.e. in horizontal, vertical and diagonal upward direction. We took one more measure of C i.e. weighted diagonal intensity variation in downward direction D, because previous A, B, C

measures alone does not detect corners properly for Devnagari characters. Now the Diagonal Intensity variation in downward direction D is calculated by convolving below specified window in figure 10 with Gaussian window of size 5.

	A1	A2	A3
B1	B2 A4	B3 A5	A6
B4	B5 A7	B6 A8	A9
B7	B8	B9	

FIGURE 10 Diagonal downward intensity variation window (D)

So the autocorrelation matrix M in step 1 is modified as

$$M = \begin{bmatrix} A & C+D \\ C+D & B \end{bmatrix}$$

Steps 2, 3 and 4 are performed as specified above. All non-zero points remaining in the corner ness map are corners. We used k=0.04 in our algorithms.

5.2.3 Method of Minimum edit Distance

After detecting the corners in the character image we segmented it into 25, and in each segment we counted number of corner points. So for each character we got a corner string of 25 which is utilized for calculating the distance among the characters. Minimum edit distance method [17] is used for this purpose. Distance is a measure of similarity between two strings which is referred as source string and target string. The distance is the number of deletions, insertions or substitutions required to transform s into t. The basic idea behind calculating the distance D(i,j) between two corner string s1[1…i] and s2[1….j] is as follows:-

D(i,0)=i

$$D(i, j) = \text{minimum}\begin{pmatrix} D(i-1, j)+1, \\ D(i, j-1)+1, \\ D(i-1, j-1)+t(i, j) \end{pmatrix}$$

$$t(i, j) = \begin{cases} 0 & \text{if } s1(i) = s2(j) \\ 1 & \text{if } s1(i) != s2(j) \end{cases}$$

6. RESULTS

The experiment evaluation of the above technique was carried out on 7154 isolated samples of Devnagari basic characters (vowels as well as consonants) out of which 4900 samples were collected from ISI, Kolkata [24] and rest 2254 samples were collected from different people in our organization. A total of 65% characters are used for the training and rest is used for testing purpose. We have used 3-fold cross validation schema for recognition result computation. Data

set is divided into 3 subsets and testing is done on each subset using the rest two subsets for learning. The recognition rates for all the test subsets are averaged to calculate recognition accuracy.

6.1 Recognition Results

The recognition accuracy obtained from our above discussed classifiers separately is shown in Table 1. The overall global recognition accuracy of our system using combined MLP is 76.67% when zero percent confusion was considered. 93.27% accuracy was obtained when we considered top 5 choices of the recognition result and with zero percent confusion.

Classifier	Accuracy
Combined MLP (top 1 choices)	76.67%
Combined MLP (top 5 choices)	93.27%
Minimum Edit distance classifier	85%

TABLE 1. Individual accuracy of MLP and minimum edit distance method

6.2 Relative Difference versus Confused Characters

Confused character samples are separated out based on the relative difference value. Relative difference measure used is computed as

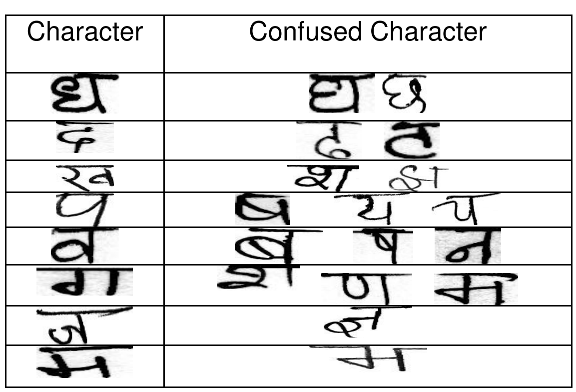
Diff = (2* max2-max1-max3) / (2*max2)

Where max1, max2, max3 are top3 values of combined MLP classifier. If this relative difference is greater than some threshold value, we infer that the top choice determines the class of the sample character with certainty. On the other hand, if the relative difference measure is less than or equal to the threshold value, we infer that the sample character belongs to a group of confusing character identified by top three choices.

Considering the lowest relative difference value 57.3% characters were separated in characters with certainty set which have distinct shapes and rest 42.7% characters were in confused character set which are of very similar shapes. For classifying 57.3% characters MLP classifier was used which gave 97.27% accuracy and for 42.7% characters minimum edit distance classifier was used which gave 82% accuracy. The combined accuracy we achieved is 90.74%.

6.3 Confused Character Pair

From experiment we noticed that mainly the error occurred because of the similar shaped characters. Since the shape of the handwritten characters in this pair is very similar, most of the pair is misclassified. Some pair of Devnagari Characters which forms the main confusion pairs of characters is listed in table 2. Maximum error occurred between फ and ढ, फ and ट, ध and घ, ध and छ, म and म, य and य, य and श्र, क and फ, ट and ठ, ट and ढ, त and ल, त and ट, प and ब, प and य, प and प pairs.



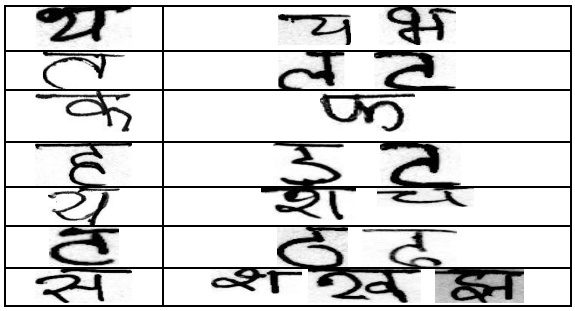


TABLE 2 . Confused Characters pair

6.4 Comparison of results

We compared our current results with those existing pieces of work. Details comparative results are given in Table 3. In our previous work [20] we obtained 69.37% accuracy as top 1 choice result. Using this proposed technique we obtained 90.74% accuracy as top 1 choice result

S. no.	Method purposed by	Data Size	Accuracy Obtained
1.	Kumar and Singh [2]	200	80%
2.	N. Sharma, U. Pal, F. Kimura, and S. Pal [5]	11270	80.36%
3.	M. Hanmandlu, O.V. R. Murthy, V.K. Madasu[21]	4750	90.65%
4.	Proposed method	7154	90.74%

TABLE 3. Comparison of Results

7. CONCLUSION

The MLP and method of Minimum Edit Distance is tested on offline Handwritten Devnagari characters. MLP classifier is used on character set with certainty and minimum edit distance classifier is used on corner detected character image to classify the confused characters of similar shapes. Modified form of Harris corner detection algorithm is used for detecting corner in character image. Results are quite promising. In future we plan to experiment on other feature extraction methods and other classifiers to get higher recognition accuracy from our system.

8. REFERENCES

[1] V. K. Govindan and A. P. Shivprasad, "*Character Recognition a Review* ", Pattern Recognition , Vol. 23, no 7 pp 671-683, 1990

[2] Sinha R.K., Mahabala 1979 "*Machine Recognition of Devnagari Script*", IEEE Trans. System Man Cyber Pgs 435-441

[3] Veena Bansal 1999 " *Integrating Knowledge Source in Devnagari text Recognition* "Ph. D. Thesis, IIT Kanpur

[4] Pal U., B.B. Chaudhuri 1997 "*Printed Devnagari Script OCR System*", Vivek, vol 10, Pgs. 12-24

[5] M. Hanmandlu and O.V. Ramana Murthy, "*Fuzzy Model Based Recognition of Handwritten Hindi Numerals",* InProc. International Conference on Cognition and Recognition, 2005, pp. 490-496.

[6] M. Hanmandlu, O.V. Ramana Murthy, Vamsi Krishna Madasu, "*Fuzzy Model based recognition of Handwritten Hindi characters*", IEEE Computer society, Digital Image Computing Techniques and Applications , 2007

[7] R.M..K. Sinha, "*A syntactic pattern analysis system and its application to Devnagari script recognition*", Ph.D. Thesis , Electrical Engineering Department, Indian Institute of Technology, India, 1973.
[8] Reena Bajaj, Lipika Dey, and S. Chaudhury, "*Devnagari numeral recognition by combining decision of multiple connectionist classifiers*", Sadhana, Vol.27, part. 1, pp.-59-72, 2002
[9] S. Kumar and C. Singh, "*A Study of Zernike Moments and its use in Devnagari Handwritten Character Recognition*", *In Proc. International Conference on Cognition and Recognition*, 2005, pp. 514-520.
[10] I.K. Sethi and B. Chatterjee, "*Machine Recognition of constrained Hand printed Devnagari*", Pattern Recognition, Vol. 9, pp. 69-75, 1977.
[11] K. Sethi and B. Chattarjee, "*Machine recognition of Handprinted Devnagri Numerals*". Journal of Instituions of Electronics & Telecommunication Engineers, India Vol 22, pp 532-535, 1976.
[12] Bansal V, Sinha R. M. K., "*Integrating Knowledge Resources in Devnagri. Text recognition system*", IEEE Transaction on System, Man & Cybernatics Part A: Systems & Humans. V30 n 4 july 2000.p 500-505
[13] U. Bhattacharya, B. B. Chaudhuri, R. Ghosh and M. Ghosh, "*On Recognition of Handwritten Devnagari Numerals"*, In Proc. of the Workshop on Learning Algorithms for Pattern Recognition (in conjunction with the 18th Australian Joint Conference on Artificial Intelligence), Sydney, pp.1-7, 2005.
[14] U. Pal, N. Sharma, T. Wakabayashi and F. Kimura, "*Off-Line Handwritten Character Recognition of Devnagari Script"*, In Proc. 9th International Conference on Document Analysis and Recognition, 2007, pp. 496-500.
[15] U. Pal, T. Wakabayashi, N. Sharma and F. Kimura, "*Handwritten Numeral Recognition of Six Popular Indian Scripts"*, In Proc. 9th International Conference on Document Analysis and Recognition, 2007, pp. 749-753.
[16] N. Sharma, U. Pal, F. Kimura and S. Pal, "*Recognition of Offline Handwritten Devnagari Characters using Quadratic Classifier"*, In Proc. Indian Conference on Computer Vision Graphics and Image Processing, 2006, pp. 805-816
[17] Malik L., Deshpande P.S., "*Handwritten devnagari character recognition using connected segments and minimum edit distance"*, TENCON 2007 - 2007 IEEE Region 10 Conference Volume , Issue , Oct. 30 2007-Nov. 2 2007 Page(s):1 – 4
[18] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, "*A Novel Approach for Handwritten Devnagari Character Recognition"*, International Conference on Signal and Image Processing (ICSIP), Hubli, Karnataka, India, 2006
[19] P. S. Despande, Latesh Malik, "*Fine classification & recognition Of handwritten Devnagri Characters with regular expressions & minimum edit distance*", Journal computing Vol. 3 No. 5 Mar 2008 .
[20] S. Arora, D. Bhattacharjee, M. Nasipuri, D.K. Basu, M. Kundu, "*Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition*", IEEE Region 10 Colloquium and the Third ICIIS, IIT Kharagpur, India Dec 2008
[21] S. Basu, N.Das, R. Sarkar, M. Kundu, M. Nasipuri, D.K. Basu, "*Handwritten Bangla alphabet recognition using MLP based classifier*", NCCPB, Bangladesh, 2005
[22] Freeman, H., *On the Encoding of Arbitrary Geometric Configurations,* IRE Trans. on Electr. Comp. or TC(10), No. 2, June, 1961, pp. 260-268.
[23] http://www.cim.mcgill.ca/~dparks/cornerDetector/harris.html
[24] http://www.isical.ac.in/~ujjwal/download/appform.pdf
[25] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, "*A Two Stage Classification Approach for Handwritten Devanagari Characters"*, International Conference on Computational Intelligence and Multimedia Application(ICCIMA07), Sivkasi, Tamil Nadu, India 2007
[26] J. Hertz, A. Krogh, R.G. Palmer, "*An Introduction to neural Computation"*, Addison-Wesley (1991)
[27] E.R. Davies and A.P. Plummer, "*Thinning Algorithms: A critique and new Methodology"* ,Pattern Recognition 14, [1981]: 53-63
[28]D. Ruta, B. Gabrys "*An overview of classifier Fusion Methods"*, Computing and Information Systems,7(2000),p.1-10

[29] T. Nawaz, S.A. H.S. Naqvi, H. Rahman "*Optical Character Recognition System for Urdu using Pattern Matching Technique*", International Journal of Image Processing(IJIP), vol 3, issue 3, pp 92-104

[30] A.R. Khan, Z. Mohammed, "A Simple Segmentation Approach for Unconstrained cursive Handwritten words in Conjunction with the Neural Network", International Journal of Image Processing(IJIP), vol 2, issue 2, pp 29-35

Expert Search Engine - A New Approach for Web Environment

Laxmi Ahuja laxmiahuja@aiit.amity.edu

Amity Institute of Information Technology
Amity University, Uttar Pradesh
Sec 125 Noida (UP)

Dr Ela Kumar ela_kumar@rediffmail.com

School of Information Technology
Gautam Budha University
Greater Noida

Abstract

This paper develops an expert web search engine for Web Environment and uses Ajax based technology for this. Applications (Search Engine) of this expert Search system will be to give the user a choice of best Search results as per their need/requirement. From organizational point of view this knowledge can be used for devising various enhancements for search engine optimization. It applies a knowledge engineering based technique for the development of this expert system. To understand the basic functioning of search engine, various Web Forums and Blogs have been studied. Present work develops Intelligent Agent and Interaction Agent based knowledge base of Search Engine. The results produced depend upon what type of program it is using and details are produced according to it. This knowledge based Search Engine model thus developed may be useful in knowledge management and knowledge reuse. At user level it can be used for suggesting best Search results to the user and at organizational level it can be used for drawing various conclusions for managing quality database for better application use.

Terminology: Knowledge, Knowledge Engineering, Knowledge Management, Search Engine

1 INTRODUCTION

Search Engine deals with offering search results over WWW. In this paper we have developed knowledge based expert system of Web Search Engine, which simulates functionality of a Search Engine in a Web Environment. To get the information about actual working of concerned system

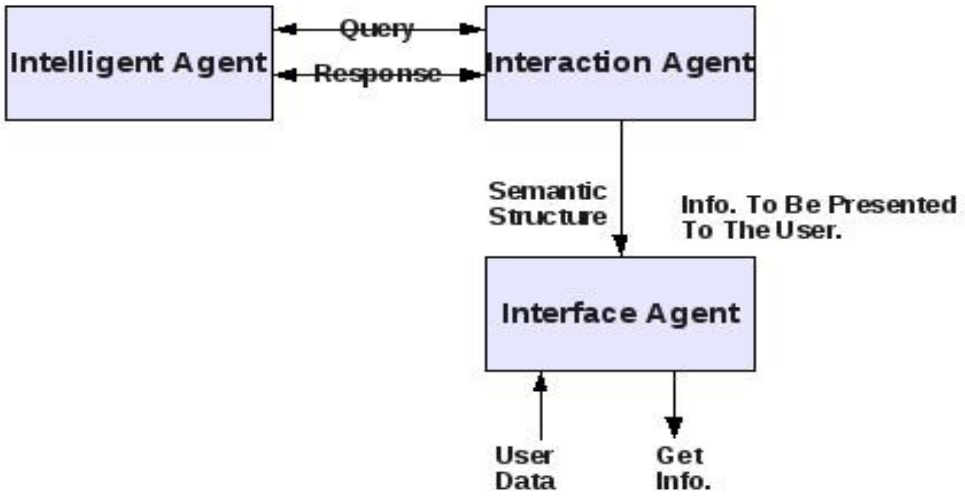


Fig: The Agent Based Diagram of Expert Search Engine

we have checked other search engines and different forums. The model has three components intelligent agent, Interaction agent and Interface agent interacting with each other for the purpose of drawing the decision for a search results. The schematic diagram of simulated model is shown in fig.

In order to develop the Search Engine the offered results found are categorized into different levels. Various offered results are represented in form of a tree called as "offer tree". Part of offer tree is depicted as follows. However the complete offer tree consisting all results detail is stored in intelligent agent. The intelligent agent will consult this tree while suggesting appropriate results to Interaction agent.

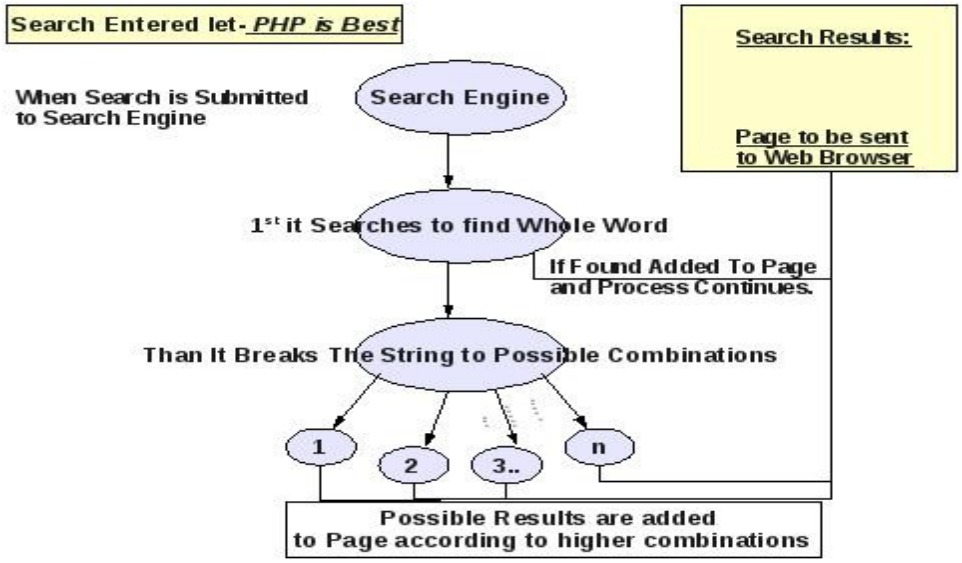


Fig. Part of Offer Tree

The Search Engine functions in two phases namely analysis phase (server-side) and customer interaction phase (client-side). Function performed in these phases is summarized below. Knowledge engineering can be used to develop system (Search Engine) with high level

processing layer that provides reasoning strategy to be considered by Search Engine for search results.

2 ANALYSIS PHASE

The functions performed in this phase are:
Find/identify appropriate Combinations of search results according to user requirements and constraints, considering factors such as client preferences, Meta keywords for a page, relationship between Search term and preference, set of constraints of a string.
Customer interaction phase – The activities performed in this phase are: analysis and management of query based conversation, client profile and the explanation demanded by the server/client. The agent based model of the prototype defined for the system allow the concurrency in distribution of data, information , knowledge, tasks/methods among interface agent, interaction agent and intelligent agent. The agent based model schematic is shown in following diagram:

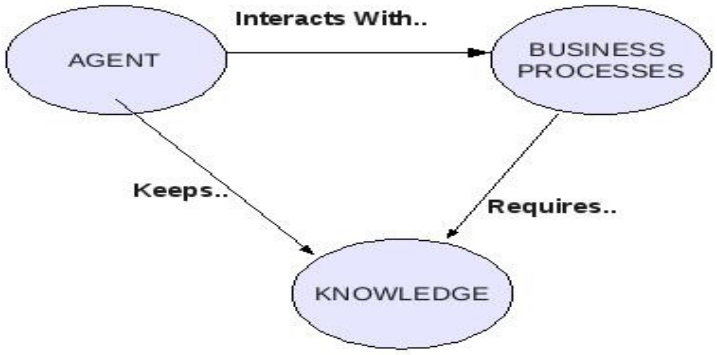


Fig. Agent Interaction in System

It's been analyzed that at present without following systematic approach of acquiring / gathering / manipulating of knowledge, Company is using less than 50 % of available knowledge in the company. This indicates the underutilization of resources and production capacity. In terms of Search Engine the production capacity is defined as "how many new development has been done in it" in the Web Application. The Web Standards are controlled by a Consortium called W3C. One has to validate his source code there. Here Keywords has a vital role in the system. These Keywords causes how right Your Search is? All Meta-keywords are stored in server's database to be matched with user entry.

Knowledge Engineering technology can be of great help in dealing with the complexity of real problems of Searching one's required Web page. It can automate the process of guiding the user to manually check each and every website related and which is best suited according to his requirements and should be used. The efforts of web searches can greatly be reduced if an expert system (Search Engine) can be developed exhibiting the intelligent behavior i.e it will include developing a system with high level processing layer. This leads to the need of improvement at interaction level in the user system relationship. This indicates that the system should be able to recognizing what are the key information required "of the user" (Search String) and then invoke the user adapted information generation process to provide the required information according to their conversation/communication.

This paper tries to give you simulated system which gives user better results to be adapted as per needs consisting of good interface for “Interaction agent". The objective is to design and implement an interface and intelligent computer system based on the analysis of the problem for a Web Application Developing company supporting natural language interface and graphics with the user(through Browser). It will perform a query based interaction supported by different keywords stored in server's database represented in form of “offer tree” provided by Intelligent Agent(as shown before). Intelligent agent is responsible for interaction and will collect user’s query and transform them into semantic structures. The intelligent agent is responsible of the generation of information required by user as well as information required to perform an intelligent decision to improve the best possibility of the interaction.

The Interaction Agent is responsible for adequate management of interaction between Interface agent and intelligent agent as well as with the user. This agent will ensure the consistency in conversation and pass the query of user to Intelligent Agent and Interface agent. In this paper we concentrate on the aspect of design of intelligent agent from the perspective of user.

3 DEVELOPMENT OF INTELLIGENT AGENT

It uses a hierarchical tree structure to develop knowledge model which can represent /characterize many classes of problems to be solved or tasks to be performed. The design of knowledge engineering model is based on the actual organization oriented principle. Knowledge model consist of three knowledge area and each knowledge sub areas defines a domain of expertise that explains a specific problem solving behavior, encapsulating both task and domain knowledge.

Every knowledge area is described by specific task domain knowledge and its functionality will be depicted by a set of tasks. Each task is also one descriptive entity that includes problem solving process that incorporates reasoning strategy involved in the problem set to achieve the objective defined by the task.

In our model three knowledge sub areas are :

- System specialist
- Search specialist
- Database specialist

Corresponding to three knowledge area the knowledge will be stored in a model called as domain model because the related knowledge will be specific to that particular domain. Accordingly there will be three domain models, corresponding to each knowledge area. Each area specialist will be expert in their domain.
System Specialist: They handle high level knowledge about the task that is to be performed in Web-application e.g. w.r.t Web Search Engine there is one regulatory body looking after the following tasks:

- Launch of new Features.
- Revision of old Features
- Speed of Search.
- Failure Handling etc.

Search Specialist: they deal with the knowledge required or related to the sector and what kind of Results is needed to accomplish/perform the task (search). Database specialist: They manage/manipulate knowledge about the features/benefits of the meta-keywords and it uses that knowledge to match the requirement/needs of user with the meta keywords for a website in the server's database.

Knowledge engineering approach is used for designing function model that allow the reuse of

knowledge base component. After coding the knowledge in form of three components described in the model, this model can be used for drawing decision to suggest best search results according to user requirement. The process of decision making is described in following diagram.

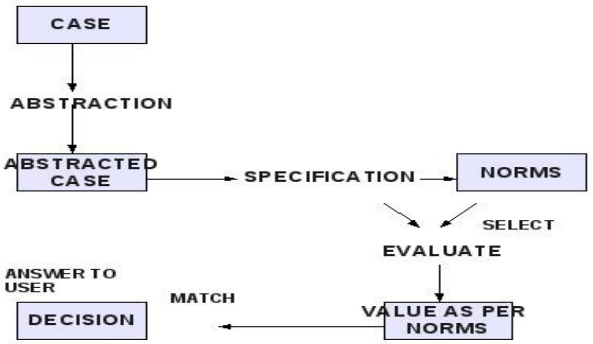


Fig. Dynamics of Decision taking

It is necessary to define a shared vocabulary to have unified schema that will interact with different knowledge are. Definition of generic vocabulary that should be instantiated in respect of a particular concepts, attributes and values of domain problem must be given. We have defined this in terms of generic vocabulary of keywords. In domain model of Search Engine the knowledge is incorporated using many different knowledge representation techniques and distributed between knowledge bases. Each knowledge will have an inference method. In this we can use rule based inference method for case based reasoning strategy. For example a complete model of one Search can be defined with the help of concept, attributes and values.

The rules will be evaluated based on values held by attributes and with the certainty of rule it will be matched. Certainty of rule will be evaluated with the average values. Whenever an Interaction agent sends a request to Intelligent Agent, it will start inference process and gives back a offer tree with the combinations of search results that meets the user requirements. Interaction agent passes information regarding all suitable results to user. User with the help of Interaction agent will prune the tree according to his psychology and will adapt the best result.

4 CONCLUSION

This paper automates the selection process of a Search Engine by developing an Intelligent System for Web Environment. It uses Ajax based technology for it. The objective which system fulfill is to satisfy the user as per their requirements by having query conversation in natural language(through strings). The same study can be used as an asset for the organization. It will also allow the reuse of knowledge. Besides this the knowledge which is maintained and manipulated inside the organization may leads to take better decision about finding a search result, revising some methods, features or withdrawing some features from the application that helps the organization to raise their users.

5 REFERENCES

[1] Elizabeth Liddy, How a Search Engine Works, a white paper available at http://www.infotoday.com/searcher/may01/liddy.htm.
[2] Sergey Brin and Lawrence Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine, available at http://infolab.stanford.edu/~backrub/google.html
[3] Search Engine, available at http://www.learnwebskills.com/search/engines.html
[4] Wrox Beginning PHP5, Dave W. Mercer, Allan Kent, Steven D. Nowicki, David Mercer, Dan Squier, W. Choi. ISBN: 81-265-0539-7

[5] C. Mitchell, G.Y. Tian, D. Gledhill, and D. Taylor, Web-based Interactive 3D Visualisation for Business and Building Management, Proceeding of Internet and Multimedia Systems and Applications - 2004, pp: 427-436.
[6] Google Search Engine http://google.stanford.edu/
[7] Harvest http://harvest.transarc.com/
[8] Mauldin, Michael L. Lycos Design Choices in an Internet Search Service, IEEE Expert Interview http://www.computer.org/ pubs/expert/1997/trends/x1008/mauldin.htm
[9] The Effect of Cellular Phone Use Upon Driver Attention http://www.webfirst.com/aaa/text/ cell/cell0toc.htmwww.bing.com
[10] www.wrox.com
[11] www.lycos.com
[12] www.php.net , PHP Manual
[13] www.mysql.com
[14] www.google.com
[15] www.debian.com
[16] www.phpclasses.com
[17] www.google.com/googleapi

On the use of continued fractions for electronic cash

Amadou Moctar Kane amadou-moctar.kane.1@ulaval.ca
Département de Mathématiques et de Statistiques
Université Laval
Québec G1V 0A6 Canada.

Abstract

This paper presents an electronic cash scheme using the theory of continued fractions. Continued fractions have already some utilities in cryptography such as in the cryptanalysis of RSA ***[17]*** or in the design of some stream ciphers ***[9]***. In order to achieve our prepaid e-cash scheme, we will use the continued fraction expansion of some irrationals numbers, although the same scheme can be obtain with a block cipher algorithm like AES or with some pseudo-random generators. Our e-cash scheme has two aims: the first one is to create a payment system independent of current constraints such as the revocation of anonymity (in the double spending case) or the obligation for those who want to use the e-cash, to have a bank account.
The second aim is to propose here a solution which prevents the copy of our e-coins and allows if necessary the reimbursement of the user with e-cash.

Keywords: continued fractions, cryptography, electronic commerce, electronic cash, prepaid card.

1. INTRODUCTION

The electronic cash aim is to permit an efficient trade in a total anonymity over networks. In the e-cash system, the user withdraws electronic coins from a bank, and pays a merchant using these coins. During the transaction, the merchant can verify the authenticity of the electronic coins (in some protocols the merchant does not need to interact with the bank before accepting a coin from the user); collects multiple coins spent by users and deposits them later at the bank. In our case the scheme is a bit different because the bank makes a prepaid card and the user buys this card in a shop.

The e-commerce is widely developed today, unfortunately many transactions are done with credit cards. And as we know it, making a transaction with a credit card can be dangerous when the web merchant is not well protected or when the web merchant is cheating. Recently, information from more than 130 million credit and debit cards was stolen in some big company in North America ***[2]***. Because of the increasing number of frauds, it is likely that holders of credit and debit cards will suffer the consequences of these frauds in the future.
We can list in the following some reasons which delay the emergence of the e-cash system in the world.

- People do not know how it is dangerous to use a credit card on-line.
- E-cash are often produced for holders of bank accounts.
- The existence of the e-cash is only known by passionate people.
- E-cash are accepted by a small number of traders.
- The e-cash cost is quite expensive.

- E-cash are produced by very few banks.
- There exists some fears regarding the traceability of e-cash.
- …

To correct these mistakes, we believe that the e-cash must be done in such a way as to be easy to get and to use. Unfortunately up to now, all the solutions given in the e-cash system are difficult to implement for a large public. Since the introduction of the e-cash by Chaum ***[3]*** in 1982, a lot of others protocols have been proposed. Among these protocols Okamoto and Ohta ***[6]*** have proposed the six desirables properties for an electronic currency system:

I. Hardware independence:
   The cash can be sent securely through computer networks.
II. Security:
   The ability to copy (reuse) and forge the cash must be prevented.
III. Anonymity:
   The identity of the user should be protected, but for some transactions weaker forms of anonymity should be use ***[1]***.
IV. Off-line:
   The transaction can be done off-line, meaning no communication with the bank is needed during the transaction.
V. Transferability:
   The e-cash can be transferred to others users.
VI. Dividability:
   A piece of cash can be divided into smaller amounts.

We can add to these six properties two new ones. The first one is: to be efficient, the transaction must be quick in some phases. The second one is the reimbursement of the user, which must be available if the client is not satisfied by goods or services.

To ease the use and the acquisition of the e-cash, we propose to increase the use of prepaid cards. In this paper the prepaid card is a piece of paper contending some hidden codes. This card can create some security issues because of the total anonymity which it gives. Hence, it will depend on each issuer to limit the value of the prepaid card or to forbid the use of the e-cash in the sale of some dangerous products. In some countries, some studies on prepaid cards have already been done ***[15]***, and some prepaid cards already exist.

In the e-cash area there exists now some efficient schemes such as the compact e-cash ***[4]*** which permits the withdraw of 2^lcoins in a short time. And more recently some news schemes can use the compact e-cash without the trusted third party ***[5]***.
Up to now, a lot of solutions such as the hash function, the random oracle model, the cut-and-choose technique, and the blind signature have been used in e-cash schemes. Our approach in the e-cash system is different, because of these following points.

I. We want to introduce the use of continued fractions in e-cash protocols.
II. In order to avoid some burdens imposed by the banks, we want to reduce their influence on the e-cash system.
III. We want to prevent the copy of the e-cash.
IV. As noticed previously, we believe that the reimbursement is important to solve in all e-cash schemes, because sending back the e-coin is not a solution.
V. Like the traditional cash, we want to create an e-cash which will be difficult to trace.
VI. We aim to present a very simple e-cash scheme.
VII. …

For the security aspect, we aim to cover these three issues, unforgeability, stating that valid coins can only be issued by the bank or an authorized entity; anonymity, ensuring that a user stays anonymous even if the complete system conspires against him; and exculpability, a malicious bank should not be able to conspire with malicious merchants to frame an honest user for double-spending.

Here we present an algorithm based on the difficulty of retrieving an irrational number from the sole knowledge of a part of its continued fraction expansion. As proved in ***[9]***, the continued fraction expansion can produce a pseudo-random sequence, hence our e-cash scheme is built around a pseudo-random sequence.

We recall that the use of pseudo-random function is already effective in some e-cash scheme like ***[4]***.

Continued Fractions: An expression of the form

$$\alpha = a_0 + \cfrac{b_0}{a_1 + \cfrac{b_1}{a_2 + \cfrac{b_2}{\ddots}}}$$

is called a generalized continued fraction. Typically, the numbers $a_1, , b_1, \ldots$ may be real or complex, and the expansion may be finite or infinite.

We will avoid the use of the continued fraction expansions involving $b_i = 1$ for most i's. However, in order to simplify our explanation we will use in some cases the classical continued fraction expansion, namely $b_i = 1$ for any i:

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$$

$$\alpha := a_0, a_1, a_2 \ldots$$

In this paper we denote by Γ the combined sets of algebraic irrationals of degree greater than 2 and transcendental numbers. Our algorithm, will use the irrational numbers which are in Γ, but we will avoid the use of transcendental numbers having a predictable continued fraction expansion (some examples of irrationals numbers given a predictable continued fraction expansion are presented in ***[7]***,***[10]***).

To calculate the classical continued fraction expansion of a number α, write down the integer part of α. Subtract this integer part from α. If the difference is equal to 0, stop; otherwise find the reciprocal of the difference and repeat. The procedure will halt if and only if α is rational.

We can enumerate some continued fractions properties:

I. The continued fraction expansion of a number is finite if and only if the number is rational.
II. The continued fraction expansion of an irrational number is unique.
III. Any positive quadratic irrational number α has a continued fraction which is periodic from some point onward, namely a sequence of integers repeat. (Lagrange Theorem)
IV. The knowledge of the continued fraction expansions of α and β cannot determine simply those of $\alpha + \beta$, or $\alpha\beta$.

Continued fractions were widely studied by C. Olds ***[13]*** and O. Perron ***[14]***, but cryptographic views are not explored by number theory specialists except in some fields like RSA cryptanalysis.

This paper is organized as follows. In section 2 we will propose and demonstrate some results concerning continued fractions; in section 3, we will introduce our e-cash scheme. Section 4 prove the security given by our scheme, and before the conclusion, we will study the efficiency of our design.

2. RESULTS

In order to show that the merchant or the user will not be able to make e-coins, we will prove the result 1 and 3. The result 2 will help us to exhibit an example of irrational number which we can use in our e-cash scheme.

Definition 1. An electronic cash scheme is secure if it has the unforgeability, the anonymity, the exculpability, and if the e-cash can not be copied.

Notation
Let $\alpha \epsilon \Gamma$ such that $a_1, \dots, a_m, \dots, a_{m+n}, \dots$ is the continued fraction expansion of α; m and n are two integers such that $m > 20, n \geq 1$. We denote by δ the vector made with the n partial quotients following the m first partials quotients in the continued fraction expansion.

Remark
The choice of $m > 20$ does not have a real influence on the security of this algorithm.

Result 1. *It is not possible to find α out of the knowledge of δ.*

Proof. Let $\alpha \epsilon \Gamma$. We suppose that we know a given part $a_{m+1}, \dots, a_{m+n}$ of α's continued fraction expansion. Can we find α with the knowledge of these n partial quotients?
The answer is negative, because there exists an infinite number of irrationals with these same partial quotients.
For instance we can exhibit infinitely many irrational numbers α_ρ which are different from α and which have the property that $a_{m+1}, .., a_{m+n}$ appears as a sequence of n consecutive partial quotients. As a matter of fact, when θ is an irrational number, it suffices to consider any sequence of m integers $(r_1, r_2, \dots, r_m)$ and to define α_ρ to be

$$\alpha_\rho = r_1 + \cfrac{1}{\ddots_{r_m} + \cfrac{1}{a_{m+1} + \cfrac{1}{\ddots_{a_{m+n-1}} + \cfrac{1}{a_{m+n} + \frac{1}{\theta}}}}}$$

Result 2. *For an integer r such that $r \geq 3$ and a real algebraic number A ($A \neq 0, 1$), the number $\sqrt[r]{log\ (A)}$ is transcendental.*

Proof. Assume that A is a real algebraic number such that $A \neq 0, 1$, then $\log(A)$ is transcendental number by Corollary 3.6 of ***[8]***.
If we suppose that $X = \sqrt[r]{\log(A)}$ is an algebraic number, then X^r is a algebraic number, which is absurd because $X^r = log(A)$ and $log(A)$ is transcendental.

Remark
The irrational number $\sqrt[r]{\log(A)}$ used in this paper is not a standard which we impose. It is an example which we choose in order to illustrate our scheme.

Result 3. *Let $\beta \epsilon \Gamma$ and let $a = [a_m, \dots, a_{m+n}]$ be a part of the continued fraction expansion of β. The knowledge of a does not allow to know any other partial quotient of continued fraction expansion.*

Proof. From the proof of *Result 1*, we can deduce the proof of *Result 3.*

For instance we can exhibit infinitely many irrational numbers γ such that the partial quotients of the continued fraction expansion of γ ($[u_0, u_1, u_2, \dots]$) verify $u_t \neq a_t$ for any $t \geq m+n+1$, $u_i \neq a_i$ for any $i < m$, and $u_j = a_j$ for any $m \leq j \leq m+n$.

3. DESIGN

Definitions
In this design, we have four entities the bank (***{B}***), the user (***U***), the merchant (***M***) and the trusted third party (***{TTP}***). We denote the concatenation by ||, and the hash function by h.

Remark
In order to reduce the influence of banks, the banking entity could be replaced in this scheme by a person who have a public key, a credit card and who is permitted to sell e-cash prepaid cards.

3.1 Design

We suppose that all communications between the protagonists are secure, we assume that the bank public key, the merchant public key, and the ***{TTP}*** public key are available for all entities. The integer e is fixed arbitrarily for all entities ($e = 7$). The solution proposed in this paper can be divided in five parts.

I. The bank preparation

a. ***{B}*** chooses randomly a number $z \in N^* \backslash \{1\}$ and computes the irrational number $\alpha = \sqrt[e]{\log(z)}$.
b. ***{B}*** computes the first partial quotients of α. For example if the price of the prepaid card is ten dollars, then the bank will compute the 1020 first partial quotients of the irrationals numbers α. In the following, ***{B}*** will ignore the 20 first partial quotients.
c. ***{B}*** chooses x_1 and q which are the prepaid card number.
d. The bank writes q and x_1 in the prepaid card.
e. The bank hides q.
f. ***{B}*** sends to the ***{TTP}*** z, q and x_1.
g. The bank sells the prepaid card to a shop.

II. The user preparation
We suppose that the user has already downloaded the software managing the prepaid card.
a. Before buying a prepaid card, the user and the seller ask to the ***{TTP}*** if the card number x_1 is already allocated?
b. If the ***{TTP}*** response is negative, the seller asks to the ***{TTP}*** to activate the prepaid card.
c. The ***{TTP}*** activates the card, chooses randomly a number $q_1 \in N^* \backslash \{1\}$ and sends q_1 to the user.
d. The vender sells the card to the user.
e. The user scratches the card and recovers q.
f. When the user connects the software, he chooses randomly a number $x \in N^* \backslash \{1\}$ and sends x, x_1, q_1 and q to the ***{TTP}***.

III. The ***{TTP}*** preparation
a. The ***{TTP}*** verifies if x_1, q and q_1 match.
b. If the verification is conclusive, the ***{TTP}*** computes y= $\sqrt[e]{\log(x)}$ and $\alpha = \sqrt[e]{\log(z)}$.
c. The ***{TTP}*** computes the first partial quotients of y and α. For example if the price of the prepaid card is ten dollars, then the ***{TTP}*** will compute

the 1020 first partial quotients of the irrationals numbers y and α. In the following, the ***{TTP}*** will ignore the 20 first partial quotients.

d. For all $i's$ (21 to 1020), let c_i be the concatenation of i-th partial quotient of y which is y_i and the i-th partial quotient of α which is α_i, hence $c_i = y_i \,||\alpha_i$.
e. The ***{TTP}*** computes the hash of the 100 first c_i's namely h_1, before computing it for the 100 c_i's following until the end of the c_i's.
f. The ***{TTP}*** sends the ten hash $\{h_1, h_2, \dots, h_{10}\}$ to the bank.
g. From the ten hash, the bank produces ten signatures $\{s_1, s_2, \dots , s_{10}\}$ with its private key.
h. The bank sends the signatures to the ***{TTP}***.
i. The ***{TTP}*** sends the signatures $\{s_1, s_2, \dots , s_{10}\}$ to the user.

IV. The Spending.
a. The user visits the merchant website and chooses for example a product which price is 3 dollars and 11 cents.
b. For the payment, the web merchant transfers the user to the ***{TTP}*** website.
c. The ***{TTP}*** asks to the user x_1, the 311 first useful partial quotients and the three first signatures mainly s_1, s_2, s_3.
d. The user ignores the 20 first partial quotients, sends the 311 following, x_1 and the signatures s_1, s_2, s_3.
e. The ***{TTP}*** rebuilds the c_i's after the calculation of the 420 first partial quotients of α.
f. The ***{TTP}*** verifies the validity of signatures and partial quotients. He generates with its private key s which is the signature of the 11 last partial quotients if all signatures and partial quotients are valid.
g. The ***{TTP}*** sends to the merchant the signatures s, s_1, s_2, s_3, the 311 partial quotients and x_1.
h. The merchant verifies the validity of the signatures received with public keys.
i. The merchant sends the product to the user if the partial quotients are valid.

V. The deposit
a. The merchant sends to the bank x_1, the signatures s, s_1, s_2, s_3 and the 311 partial quotients.
b. The bank verifies the validity of the partial quotients received.
c. If all the partial quotients are valid, the bank sends 3 dollars and 11 cents to the merchant.
d. When the bank is replaced by a person who sells prepaid cards, the person will pay the merchant with his credit card.

3.2 Remark

I. The order of the partial quotients is important to respect.
II. We suppose that the hash function used in this scheme is a collision resistant hash function.
III. In order to be effective, the bank signs by hundred partial quotients. But if the bank computing possibilities are high, the bank can sign by ten partial quotients (or less).
IV. In our scheme we assume that the smallest coin value is one cent, but depending on the value of the prepaid card the partial quotient value can for example correspond to 10 cents, 1 dollar…
V. For example in the case where a partial quotient value is one dollar, the ***{TTP}*** can divide the partial quotient. If we assume that c_1 is partial quotient to divide,

the ***{TTP}*** computes the partial quotients of $c = \sqrt[e]{\log(10^{40}\log(c_1)}$. The ***{TTP}*** ignores the 20 first partial quotients and takes the 100 partial quotients following, each value of these new partial quotients is one cent.

VI. Due to the rounding errors, the use of continued fractions must obey some rules. For example the ***{TTP}***, the user, and the bank must agree on their multiple precision library, on the rounding error, on the software used and on the architecture.

3.3 Reimbursement

A lot of electronic cash schemes omit or neglect this functionality which is a serious issue for the development of the e-cash system. We suppose that the user buys some goods with e-cash coins and he is unsatisfied when he receives the product. As the law allows it, he asks for a reimbursement to the web merchant. The anonymity of the user must be protected and we do not want the merchant to see the e-cash that he has to send back to the user (reuse).
The user can recover his e-cash in these following steps.

I. The user chooses randomly a transaction number N_1 and sends to the merchant a file $fic\ (x_1, N_1)$ contending x_1 (his prepaid card number), and N_1.
II. The merchant signs the file $fic(x_1, N_1)$ with his private key and sends to the user the signed file denoted by $fic_M\ (x_1, N_1)$.
III. The user verifies the validity of the merchant signature with the merchant public key. If the signature is valid, the user sends back to the merchant the goods with a system of acknowledgment of receipt (the user will receive a proof showing that he has sent the goods and the merchant has received it).
IV. The merchant sends the file $fic_M\ (x_1, N_1)$ to the bank as soon as he receives the goods.
V. The bank signs the file $fic_M\ (x_1, N_1)$ with its private key and sends to the merchant the signed file denoted by $fic_{MB}\ (x_1, N_1)$.
VI. The merchant verifies the validity of the bank signature with the bank public key. If the signature is valid, the merchant sends to the bank an signed order to debit the amount on his account.
VII. The bank sends to the ***{TTP}*** x_1, N_1, and the number of partial quotients needs for the reimbursement (for example 311).
VIII. The ***{TTP}*** computes the partial quotients from 1021 to 1332 of y and α which are $y_{1021}, \dots, y_{1332}$ and $\alpha_{1021}, \dots, \alpha_{1332}$. For all i's from 1021 to 1332, let c_i be the concatenation of i-th partial quotient of y which is y_i and the i-th partial quotient of α which is α_i, hence $c_i = y_i \,||\alpha_i$.

From $i = 1021$, the ***{TTP}*** computes the hash of the hundred first c_i's before computing it for the hundred c_i's following until the last bloc of hundred c_i's. The ***{TTP}*** computes the hash of 11 c_i's remaining.

The ***{TTP}*** writes in the same file, the four hash produced $(h_{11}, h_{12}, h_{13}, h_{14})$, x_1 and N_1. The ***{TTP}*** signs the file and sends to the bank the signed file denoted by f_{TTP}.

IX. ***{B}*** signs the four hash $(h_{11}, h_{12}, h_{13}, h_{14})$ with its private key, sends the four signatures produced $\{s_{11}, s_{12}, s_{13}, s_{14}\}$ and the file f_{TTP} to the web merchant.
X. The web merchant sends $\{s_{11}, s_{12}, s_{13}, s_{14}\}$ and the file f_{TTP} to the user.
XI. The user verifies the bank signature, verifies the ***{TTP}*** signature and recovers the four signatures required for his future transaction.

4. SECURITY ANALYSIS

Result 4. *According to definition 1 the e-cash scheme proposed in this paper is secure:*

- *The e-cash is unforgeable.*

- *The e-cash user is anonymous.*
- *The e-cash scheme has an exculpability property.*
- *The user is unable to copy the e-cash.*

Proof. In the following, we will prove that our scheme respect these four properties. We suppose that we have four potential attackers in this scheme. This attack can be tried by the user, by the merchant, by the bank or by someone else who is completely outside the scheme.

4.1 Unforgeability

The unforgeability is effective for someone who is outside the e-cash scheme under the assumption that the bank signature is unforgeable and the communication between the protagonists is secure.

If we assume that the user is not able to forge the bank signature, the unforgeability is effective for the user because of the following reasons.

- The user does not know the bank partial quotients which are α_i, so he will not be able to send corrects c_i's to a web merchant without the help of the ***{TTP}***.
- The user can not forge e-coins in addition to the good e-cash that he already has, because he needs the bank signature for these new partial quotients.

Due to the following points the merchant can not forge a valid e-cash.

- The merchant is not able to find the partial quotients corresponding to the signatures sent by the bank (in the case of a reimbursement).
- Assuming that the merchant knows the $c_i = y_i||\alpha_i$, can he guess the y_i and the α_i? The answer is negative because the number of digits of partial quotients is not fixed. In some cases a probability attack can find some y_i and α_i.
- Assuming that the merchant knows a lot of y_i (or α_i) can he find y (or α)? The answer is negative (proved in result 1).
- Assuming that the merchant knows y_i, can he find y_{i+1}? The answer is negative (proved in result 3).
- Even if we suppose that the merchant can find $y_{i+1}, y_{i+2}, \ldots, y_{i+100}$, he will not be able to forge the bank signature.

If we assume that the merchant and the user are accomplice.

- The user sends to the ***{TTP}*** his partial quotients and the bank signatures.
- The ***{TTP}*** computes α_i's, concatenates partial quotients, verifies bank signatures and sends signatures and c_i's to the merchant.
- The merchant sends to the user the c_i's.
- Since the user knows c_i he can find the α_i used but he can not find α (proved in result 1) and he can not find more α_i's (proved in result 3).
- The user and the merchant can not swindle the bank.

If we assume that the bank tries to swindle the user.

- The bank receives from the merchant the c_i's.
- Since the bank knows c_i he can find the y_i used but he can not find y(proved in result 1) and he can not find more y_i's (proved in result 3).
- The bank can not swindle the user.

4.2 Anonymity

The prepaid card is bought somewhere in the world, then the anonymity of the user is total. Even in the case of a reimbursement his anonymity is protected.

4.3 exculpability

The double spending is not feasible for the user, then the exculpabilty is effective in this scheme.

4.4 The e-cash can not be copied

The e-cash can not be copied by the user because of the ***{TTP}*** participation. If we suppose that the user sends these 400 partial quotients of y which are $y_{21}, \dots, y_{420}$ to the ***{TTP}***. The ***{TTP}*** computes the partial quotients corresponding which are $\alpha_{21}, \dots, \alpha_{420}$. When the user tries to send again the same 400 partial quotients, the ***{TTP}*** will assume that the user is sending the rest of his continued fraction expansion, then the ***{TTP}*** computes the 400 partial quotients following which are $\alpha_{421}, \dots, \alpha_{820}$. Hence, the signatures will not match with the partial quotients and this payment will be rejected.

If we suppose that the merchant copies the e-cash which he has received from the ***{TTP}***, and deposits it twice to the bank. The bank can easily find the cheater and refuses to send the concerned money twice. If the merchant copies the e-cash and gives the copy to another merchant, the ***{TTP}*** can show who had received in the first time the partial quotients which he sent.

4.5 Reimbursement security

Due to the total anonymity given in this scheme, the reimbursement must obey to some proprieties.

I. While protecting his anonymity, we must be sure that the user will receive his money back if he returns the undesirable product to the merchant.
II. We must be sure that the reimbursement can not be returned to someone else (someone who has stolen the product for example).
III. We must be sure that the seller has received its products and the condition of the product is good.
IV. We must be sure that the e-cash refunded is secure in the sense of *result 4*.

The first property is satisfied in our scheme, because to ensure that the user will receive his money back if he sends the goods to the merchant, the user has the file signed by the merchant $fic_M\ (x_1, N_1)$ and the acknowledgment of receipt.

To ensure that the bank will provide e-coins to the merchant, the merchant has the statement of his account showing that the money has been debited.

If after receiving the e-coins, the merchant denies and argues that he has not received the e-cash, the bank will send back the e-coins through the ***{TTP}***.

If after receiving the e-coins, the user denies and argue that he has not received the e-cash, the merchant will send back the e-coins through the ***{TTP}***.

If the bank tries to scam the user by sending an old refund where the user had chose (x_1, N), the merchant will exhibit $fic_{MB}\ (x_1, N_1)$ in order to prove to the bank that the user had sent (x_1, N_1).

If the merchant attempts to scam the user by sending to the user an old refund where the user had chose (x_1, N), the user will exhibit $fic_M\ (x_1, N_1)$ in order to prove that he had sent (x_1, N_1).

The second condition is solved by the fact that the user and the ***{TTP}*** are the only ones who know y, so if someone else return the merchandise to the merchant, he will not be able to use the signatures sent by the bank without knowing the partial quotients of the irrational number y.

The third condition is solved, since the acknowledgment of receipt allows the merchant to refuse the goods if it comes in a bad condition.

Finally, the fourth issue is resolved because the refunded e-cash is a part of the e-cash investigated in *result 4*.

Remarks
We recommend the use of the generalized continued fraction instead of the classical continued fraction, because the classical continued fraction produces a several partial quotients with only one digit ***[12]***. Even if the classical continued fraction is used, guessing 100 partial quotients in the right order will be very difficult. Another solution is to concatenate the partial quotients before the use.

5. EFFICIENCY ANALYSIS

We largely draw our inspiration from the evaluation method used in ***[11]*** in order to perform our critical discussion. We prove in this section that our scheme satisfy the specific characteristics from usability points of view, we compare our scheme with other e-cash models and with the traditional cash.

5.1 Usability

P2P Transferability: Our e-coin is transferable between users. If we suppose that user1 transfers 50cts to user2, then user1 has to send the 50 partial quotients concerned and the copy of the bank signature to user2. Hence user2 will not be able to spend more than 50cts because he can not guess the rest of the partial quotients (proved in result 3) and this transfer does not need the intervention of any authority (***{TTP}***, bank).

Interoperability: The use of continued fractions needs some precautions, so it can be difficult to inter-operate with e-cash using other formats.

Applicability and Cost: This scheme is very affordable because protagonists needs are: a simple calculator in order to compute the partial quotients; a connection between them (for example internet); and one software in order to manage the signatures. We recall that they need to agree on some rules in order to have the same partial quotients.

Ease of use: stages of preparation, expense and deposit are relatively short. The only point which could be cumbersome is the ***{TTP}***, however we believe that the ***{TTP}*** could be a server managed by the government. The ***{TTP}*** should not require an human intervention except during its audit or maintenance.

Efficiency: The time need for computing the partial quotients is low ***[9]***, and the time need for computing (verifying) the signature and the hash is low. We can add that the storage need is not important because it consists on storing approximately 1000 numbers for each user. The ***{TTP}*** capacities must be important because he plays a central role in this scheme.

Scalability: This system is scalable even for the ***{TTP}***.

Off-line usage: The system works in one-line mode because we need to prevent the copy of the partial quotients and to protect the anonymity of the user.

Mobility: The important things are bank signatures and integers (z, x), and they have the property of mobility.

5.2 Comparison

If we compare our solution to other existing e-cash schemes, we find that we meet the usual criteria such as security, mobility, portability, dividability, scalability, and P2P transferability. We

exceed by far these criteria with the introduction of the reimbursement that we have not seen yet in an e-cash scheme, and we prevent the copy of e-cash where most of the existing schemes break the anonymity of the user in a case of a double spending ***[3]***. The element that might play against our scheme would be the off-line use because we have reintroduced the ***{TTP}*** which most of the new schemes remove. The first reason which motivates the reintroduction is: the absence of the ***{TTP}*** in e-cash models is often offset by the introduction of a smart card ***[11]*** or an RFID tag. We have avoided the use of these cards because we have designed this e-cash scheme for internet transactions and using these cards on the Internet is often difficult, expensive and as dangerous as the traditional credit cards. The second reason is: we reintroduced the ***{TTP}*** entity because we take our inspiration from traditional cash where the bank distributes just the ticket which it has received from the central bank. We consider here the ***{TTP}*** as a sort of a central bank.

In Table 1 we compare our e-cash with the traditional cash

Characteristics	Our e-cash	Traditional cash
Verifiable origin and unforgeability	- Knowledge of the partial quotients. - Order of the partial quotients. - Authenticity of digitally signed.	Authenticity by means of material secure characteristics
Anonymity	- E-coins are anonymous. - Prepaid card is anonymous. - User anonymity can not be cancel.	Regular coins are anonymous. Payee is normally identifiable.
Untraceability	Serial numbers of prepaid cards can be recorded	Serial numbers of notes though, can be recorded
divisibility	Partial quotients are divisible.	Change can be done.
Mobility	Partial quotients and signatures have the property of mobility.	The paper is easy to transport.
Scalability	Everything can be extended in this scheme.	Only physical constrains of production, storage and transfer.
P2P Transferability	- Exchange of partial quotients and signatures. - No intermediate entities. - The merchant is able to exchange his partial quotients and signatures.	Physical objects exchange.
Off-line	One-line	No reference to authority while circulating
Life-cycle	- Coin expiration to reduce the ***{TTP}*** storage.	Material physical deterioration and destruction
Openness	-Open protocols (Continued Fraction Algorithm & Public Key	No infrastructure needed on the

	cryptography) - Common hardware (Calculator, Internet)	user's side
Reimbursement	The merchant sends back the e-coins to user The anonymity of the user is protected even in this case.	Physical objects can be given back.
Copy	The e-coin can be copied by the user but the presence of the ***{TTP}*** prevents the use of the copied e-coin. The e-coin can be copied by the merchant but bank will not send the concerned amount twice.	The cash can be copied by the user but the merchant and the bank are able to detect fake cash.

TABLE 1: Comparison between the designed e-cash (for Internet user) and the traditional cash.

6. CONCLUSION

We introduced in this paper an e-cash scheme using a prepaid card system, in summary the scheme presented here looks like a counter.
Assuming that it is the e-cash which has to adapt to users and not the opposite, we have designed a low cost system (scratch card) accessible to those who had not a bank account (prepaid card). At the same time, we have avoided to the Internet users some congestions such as readers of smart cards, and we have incorporated in our system most of the features existing in other e-cash models. On the other hand we have improved the patterns of existing e-cash, by proposing the use of the reimbursement, which is closer to reality since when a user buys something, he may have to return the thing purchased at the store if it does not suit him.
Finally, we introduced the use of continued fractions in order to create an alternative to mechanisms already used in the e-cash schemes (for example hash function and random oracle model).

Due the computer limits, the use of irrational numbers can be theoretical, but as proved in ***[9]***, we can use an approximation of irrational numbers.

It could be interesting in future research to find a prepaid card off-line without the ***{TTP}*** and where the communication between protagonists is not secure.

7. REFERENCES

1. E. Brickell, P. Gemmell, D. Kravitz, “Trustee-based Tracing Extensions to Anonymous Cash and the Making of Anonymous Change” In Proc. 6th Annual ACM-SIAM Symposium on Discrete Algorithms, 1995.
2. TJX suspect indicted in Heartland “Hannaford breaches” http://www.theregister.co.uk/2009/08/17/heartland_payment_suspect.
3. D. Chaum, “Blind Signatures for Untraceable Payments.” In Proceedings of CRYPTO 82, 1983, Plenum, New York.
4. J. Camenisch, S. Hohenberger, and A. Lysyanskaya, “Compact e-cash” In EUROCRYPT, 2005, pages 302-321.
5. Man Ho Au, Willy Susilo and Yi Mu, “Practical Anonymous Divisible E-Cash From Bounded Accumulators.” In Proceedings of Financial Cryptography and Data Security 2008 (FC 2008).
6. T. Okamoto, K. Ohta, “Universal Electronic Cash.” In Proceedings of Crypto 91, 1992.
7. Beeler M., Gosper R.W., and Schroeppel, R. Hakmen, “MIT Artificial intelligence memo 239”, Feb. 29, 1972.
8. E.B. Burger and R. Tubbs, “Making transcendence transparent: An intuitive approach to classical transcendental number theory”, Springer-Verlag, 2004.
9. Amadou Moctar Kane, “On the use of Continued Fractions for Stream Ciphers” In Proceedings of Security and Management 2009, Las Vegas, USA.
10. Donald E. Knuth, “The art of computer programming Volume 2: Seminumerical algorithms (3rd Edition)”, Addison-Wesley, 1997.
11. Dimitrios Lekkas and Diomidis Spinellis. “Implementing regular cash with blind fixed-value electronic coins”. Computer Standards & Interfaces, 29(3), March 2007, 277–288.
12. P. Levy, “Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction Continue”, Bull. Soc. Math. 57 (1929) 178-194.
13. C. D. Olds, “Continued Fractions”, Random House, 1963.
14. Oskar Perron, “Die Lehre Von Den Kettenbrüchen”, 3rd ed. (1954).
15. Report to the Council of The European Monetary Institute on PREPAID CARDS by the Working Group on EU Payment Systems, May 1994.
16. Sattar J Aboud "Secure E-payment Protocol". International Journal of Security,Volume 3, Issue 5:85-92, 2009.
17. Bruce Schneier, “Applied cryptography (2nd ed.): protocols, algorithms, and source code in C”, John Wiley & Sons, Inc., (1995).
18. G. Skinner. "Multi-Dimensional Privacy Protection for Digital Collaborations". International Journal of Security,Volume 1, Issue 1:22-31, 2007.
19. Michael J. Wiener, “Cryptanalysis of short RSA secret exponents”, IEEE Transactions on Information Theory, 36, 553-558, 1990.
20. A.Chandrasekar, V. Vasudevan, V.R. Rajasekar "Improved Authentication and Key Agreement Protocol Using Elliptic Curve Cryptography" International Journal of Computer Science and Security,Volume 3, Issue 4:272-333, 2009.

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