Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal of Computer Science & Technology

DISCOVERING THOUGHTS AND INVENTING FUTURE

Highlights

Efficient Smart Card Processing

Techniques against Various Attacks

Intentional Software Product Line

An Enhanced Cuckoo Search

Laboratory, Antarctica

Volume 12 | Issue 1 | Version1.0

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 12 Issue 1 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors: Open Association of Research Society

Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Open Association of Research Society, Marsh Road, Rainham, Essex, London RM13 8EU
United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Editorial Board Members (HON.)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Volume

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Architecture and Hardware Solutions Symbolic Information Processing

By Ibrahim Al-Turani, Dr. Eugene Titenko, Dr.Nabeel Zanoon
Applied University, Jordan

Abstract - The failure of some national projects AXES to expected results. According to experts one of the reasons is the lack of adequate theoretical apparatus for generating high-branching processes, unjustified detraction of opportunities enumerative models representative of AXES, which have their own laws paralleling computing, can not be reduced to algorithmic rules. This fact determines the need to develop innovative approaches to problem solving and organization of AXES interrelated levels of system design symbolic computation, from the linguistic level to the appropriate software and appropriate hardware level.

Keywords : *symbolic information, artificial intelligence, Flow control, Architecture.*

GJCST Classification: *B.1,H.3.2*

Strictly as per the compliance and regulations of:

Architecture and Hardware Solutions Symbolic Information Processing

Ibrahim Al-Turani[α], Dr. Eugene Titenko[Ω], Dr.Nabeel Zanoon[β]

Abstract - The failure of some national projects AXES to expected results. According to experts one of the reasons is the lack of adequate theoretical apparatus for generating high-branching processes, unjustified detraction of opportunities enumerative models representative of AXES, which have their own laws paralleling computing, can not be reduced to algorithmic rules. This fact determines the need to develop innovative approaches to problem solving and organization of AXES interrelated levels of system design symbolic computation, from the linguistic level to the appropriate software and appropriate hardware level.

Keywords : symbolic information, artificial intelligence, Flow control, Architecture.

I. Introduction

According to estimates of scientific authority in the modern computer technology , one of the leading trends of development of computer technology is the creation of homogeneous processing devices and information storage for the supercomputer and multiprocessor systems (MPS), focused on the implementation of parallel computing [1]. It is known that the main object processing symbolic information stands in various models of representation [2]. In this regard, the strategic relevance for the design supercomputers IBM and (MPS) are the questions of creating methods, architectures, and circuit design alternative for intelligent processing of information. The importance of models and techniques for intelligent processing of information associated with the emergence and rapid development of artificial intelligence (AI), in which we study a model of knowledge representation and processing. The main substantive aspects of AI problems and processes of AXES in them are the following. First, the basic format of knowledge representation and processing is currently a character format, essentially having a higher level of organization of parallel computations than the number format. Second, efficient processing of knowledge - is the realization of constructive processes of branching for multiple data with a variety of parameters, the dynamic variation of the structure and size of the data instances, by default require non-standard multi-processor architecture tour.

Originating in the 80th years of XX century as an independent branch of computer science, computer systems have traditionally included in the AXES interdepartmental, national and international scientific and technical programs and projects to create computer systems and new generation, that determines the strategic importance of research and processing of symbolic computation knowledge. Notable historical examples of such programs and projects are the West-European project ESPRIT, an American project ALVEY, Japanese project to create the fifth-generation machines (1982-1991) And the Japanese project to create a computer with a "fluid intelligence", the Russian project to create a supercomputer "Elbrus", etc. strategic importance to the problems of AXES gives a constant interest defense ministers of key countries, linking national security with the advanced development of parallel computing systems based on non-traditional (non-von Neumann) architecture.

Some national projects AXES did not lead to the expected results of my. According to experts one of the reasons is the lack of an adequate theoretical apparatus for generating high-branching processes, unjustified disparagement opportunities enumerative models representative of AXES, which have their own laws parallel computing, can not be reduced to algorithmic rules. This circumstance determines the need to develop fundamentally new approaches to solving problems of AXES and the organization of interrelated levels of engineering systems, symbolic computation, from the linguistic level to the appropriate software and hardware level.

II. Methods of Managing the Processing of Symbolic Information

The tasks of AXES with the elements of intelligent computing are understood as a search problem and the parallel generation of new states from the available set of initial states and a set of mathematical rules that are permissive nature of the execution, i.e. based on enumerative systems. On one side, the term "permissive rules" enumerative system is specified in accordance with the agreement of Post as an alternative to firing rules. For this reason, many treatment options, and constructive processes enumerative branching system simulates the algorithmic

Author[α] : Teacher, Al- Balqa' Applied University, Jordan.
E-mail : traini110@yahoo.com, Mobile: +962-078815009; Fax: +962-962- 3- 2019628
Author[Ω] : Head of Department of Special Projects, South-West State University, Russia. E-mail : johntit@mail.ru
Author[β] : Teacher, Al- Balqa' Applied University, Jordan
Email : nabeel@bau.edu.jo, Mobile: +962-0779827915; Fax: +962-962- 3- 2019628

system with a serial product of branching processes in a linear space-time. Other hand, enumerative systems, the term "permissive rules" can be specified as equal to the firing rules [4, 5]. The consequence of this method of refinement is a parallel implementation of branching processes with a structural mechanism for the generation of new states permanently to the desired number of copies of copies of data.

Nevertheless, the implementation of parallel computations on an equal basis associated with the dynamic generation of specific objects that provide a quantitative assessment of the branching process along different trajectories computation. Static methods parallelism inherent in processing numerical information, based on the placement of a dynamically modifiable set of branching processes in homogeneous computing modules. For problems of AXES such methods are not applicable in due to the lack of reliable information about the structure of a graph problem AXES. Obscurity graph structure calculations also leads to the substitution of enumerative system on its equivalent algorithmic model and unproductive expenditures of time series-return mechanism of generation of new states in the search graph [6]. These differences between numeric and symbolic computations on a theoretical level, make it necessary to use different system architecture solutions, and micro-level implementation of the A subsystem of parallel symbolic computation.

A subsystem at the level of implementation of parallel computing are three control method of the computing process, consisting of many interacting flows: flow control commands (the traditional von Neumann method), flow control, flow control requirements (switching).

Flow control commands have limited opportunities to engage in branching processes, AXES, as additional time required for the dynamic placement costs between sub-cores of the system and loss of time for data synchronization. At the same time with increasing number of processor cores total load factor of the system is significantly reduced and the problems of real complexity is reduced to 5-10% [3]. These values do not allow us to consider how the flow control commands as a promising option for the computer systems AXES

For parallel computing is a potentially attractive model calculations, flow control [3]. According to this model, any computational process is directed graph of data flow. In this graph nodes (vertices) are computer operators, and the arcs of the graph moving special data structures – Tokens. Special structures (tokens) contain field offices, describing the formats and types of operands. The coincidence of the operands to the format and automatically determines the type of command being executed, and its readiness to perform. Refusal of addressing memory cells and the transition to management through conformity the various data fields token characterized the fundamental difference between computers of AXES from the machines with the von Neumann architecture for A subsystem level. The detection of all operands relating to the common vertex of the graph is executed by named tops is an indication of initialization and computation processing unit for the given tops. Such principle of the manage eliminates the problem of synchronization and racing flows, provides an asynchronous data flow promotion by pipelined to the ring of computers that control the flow of data. Composition operators computing, communications between nodes in a graph are defined in advance at the stage of writing the program, thereby setting the graph structure. Computers Architecture Data Flow Implement the direct execution of the graph. It is provides parallelism of computing processes, given program, and excludes the conflict situations in these. The main feature of the model calculations, flow control, command execution is not on the counter, and when ready input operands for the current nodes of the graph. Execution of this Rule leads to an asynchronous execution of multiple commands at the nodes of the graph, which resulted in its input from the arcs are absorbed, and the output arcs are generated by the results of computations in a node.

This way, parallel computing model on the data flow using a limiting parallelism peculiar task that meets the requirements of the tasks of the AXES in the generation of high-branching processes with varying duration of execution. Another feature of the competitive method and models of flow control is to use a of homogeneous set of devices on an equal basis, that provides the maximization of load devices in the asynchronous command execution on parallel graph computation, At the same time known the data flow machines (MIT SDA, MDFM, MIT TTDA (England), LAU System (France), NEC Image Pipelined Processor (Japan), and others) are still oriented to the processing of numerical information, which is characterized by an explicit task graph computations In contrast, the AXES the task in most cases are not finished making a graph that defines the limitations of direct accepts of the method of flow control.

Flow control requirements is a hybrid variant control that is based on the union of a sequence of commands in a single unit with a control in its flow control and flow control between blocks of commands. In essence, task is described poorly connected graph macro level (block commands) having a low rate of exchange flows between macro level.

Table 1: Ways to control sub system level

	Flow control commands	Data flow control	Flow control requirements
description	The usual execution of the operators at their place of in the control system	"Greedy" execution of all operators for which all operands are available	"Lazy" execution of the operators, without which there can be the results of the further calculations
advantages and	Full control Easy to realization of complex data structures and control structures	The high degree of parallelism high performance	Execution only the necessary operators The independence of computing
disadvantages	low efficiency complexity programming	Complexity of control data structures Costs of unnecessary storage resources on the operands complexity of control	Time costs for the transfer of markers The complexity of public access to the structures of local representation

III. Functional Nodes and Circuit Solutions for Ibm Axis

The known of hardware solutions can be classified as AXES in their relation to micro-level hardware solutions (circuit design implementation), A subsystem (structural and functional organization), and systemic levels (a common system architecture AXES).

At the micro level we are talking about the functional nodes, blocks, and device-properties of that support basic operations, elements of the programming language in their simplest form. On the one hand, these functional units have a rigid specialization and poorly suited for general-purpose microprocessors with software control. On the other hand, commonality of processes manipulation of symbols as with abstract images, belonging to the basic pattern of thinking, "condition-action" allow us to consider these functional units as the basis for computer AXES . Availability and use of abstract computing systems (machines) for manipulating the symbols will lead to the formation of self-class operating devices with non-traditional organization extend the instruction set of modern microprocessors. Digital Converters character-oriented branching symbolic computation and the generation of a set of symbolic structures (form image shape), ultimately, justify the existence of the justification of individual devices, high-performance machines and systems that process symbolic information and knowledge, as opposed to parallel processing of numeric data and numeric computer IBM.

As a promising technical solutions for computer systems AXES level devices and functional units [1, 3] the leading scientists in the field of view Tues hardware blocks (**Table 2**) for standard operations of AXES. Continue to comment on possible hardware units for the axis and branching of computational processes.

Table2 : AXES operations circuitry-level realization

AXES operations	hardware blocks
Calling functions recursion	hardware stacks register window
typing of data	Operating register-memory tagging of memory The apparatus of parallel testing tags
sorting	VLSI-graders
Pattern matching, identification	finite state machines associative memory device matrix Converters
Modification of the fragments of character structure, reconfiguration	associative memory device character tasovateli The positional shift memory
garbage Collection	Multifunctional VLSI-graders
handling multiple response	Distributed hierarchical arbiters
Binary substitution character data	Iterative schemes for processing unitary codes

Hardware stacks, and methods for quick access to the stacks are designed to speed up function calls. This is especially useful for functional paradigm of AXES and the family of programming languages LISP. Quick operations with the stack are also useful in the implementation of Prolog. When backtracking made numerous write operations on the stack and reading from the stack.

The current stage of development of the functional units of AXES involves the use of new abstract data structures (Table 1), such as deck. To control the pointer on the group of available vertices:

Another circuitry organization associated with the creation of hybrid structures to store and access items on a stack-based organization, complemented by:

1. The reconfiguration of the stack to the associative

structure with a parallel search and access.

2. Reconfiguration of the stack in a hierarchical structure of the shear to jump to the "deep" elements of the stack.

The following non-standard functional unit intended for computer IBM AXES is tagged memory. In conventional computer IBM Von Neumann type does not distinguish between data and program, which are stored as binary strings of fixed length. Semantics of the data determines only manipulated by the program rather than the actual contents of memory, which are stored as binary strings of fixed length. Semantics of the data determines only manipulated by the program rather than the actual contents of memory. In contrast, involves a self-sufficient representation tagged representation at all levels of memory. Currently, tagging- a powerful mechanism and hardware-software tool data typing, management calculations. The most significant use of tagged memory associated with the model calculations, flow control, and machines and data streams (first of all Manchester Data Flow Machine [3]).

The most common method of hardware realization of data tagging is to add a few bits of each word, determine its type. Check the type of data in the process can be supported by additional hardware, first of all associative memory, and perform the selection of priorities and amputation unpromising branches in the graph algorithms. A distinctive feature of tagging is the task type the command being executed. Special structures (tokens) contain field offices, describing the sizes and types of operands. The coincidence of the operands by size and type of data fields in memory command automatically determines the type of command being executed. Refusal of addressing memory cells and the transition to management through compliance data fields characterizes the fundamental difference between computers AXES from the machines with the Von Neumann architecture.

The third major operation AXES - hardware support for pattern matching. Analysis of empirical data shows, Up to 90% of the time of the enumerative production system (PS) in expert systems can be spent on the process of mapping, iterative nature of the bearing. In such a way hardware realization of this operation leads to more efficient generation of branching processes and the character generating a set of candidates solutions to common principles.

Hardware processing blocks unitary binary codes. Finally, a distinguishing feature of the processes and objectives AXES is to replace character-format data unweighted binary code, in particular in the problems of searching, comparing, and comparison, identification of character data and other operations of the higher forms of computing. Such codes were named as the unweighted unitary codes, they are characterized by the possibility of applying logic and arithmetic operations, and specific processing is not dependent on the size of the bit lengths of codes. Typical examples are the functional units of digital compressors, comparators, code converters. These sites, along with high-speed distributed by the arbitrators, are the basis of responses of multiple processing units.

Character shuffle for problems-axis is not sufficient to investigate the organization of branching in symbolic computation is the use of switching converters Data, oriented on structural change in the relations between the elements. Switching converters information regarding the functional units of numerical processing is not widely used, while symbolic calculations are based structural transformation of local or global. They are associated with dynamic changes in the computation of relations of subordination or repetition of elements of symbolic structures. Structural reforms could be considered as the composition of the reconfiguration of the data structures and inter-element rearrangements controlled in these structures.

Hardware support for this operation seems justified to use the matrix and Hypercube organizations operating parts to create two-and multi-operand switching shuffle - Switching Networks Kautsa, Stone, manipulators Fan, banyan network, shuffle register with cubes of memory [3].

IV. Conclusion

The current stage of development of computer systems, IBM AXES has a short but vivid history. It is characterized by the accumulation of quantitative theoretical and hardware and software organization of symbolic computing and knowledge processing. Intellectualization of calculations, i.e. transition from data processing to knowledge processing systems in the near future will lead to massive use of computer and telecommunications equipment to enhance the intellectual capabilities of man. The basis of a new class of computing, process-oriented analysis, understanding and synthesis of new knowledge will make their own circuit solutions, based on the basic elements of the future - the optics. Optical components and functional units will ask a variety of data due to reconfiguration and compression, as well as to parallel processing on non-specific operations, while the spatial reconfiguration of the elements of the matrix, and associative processing fragments of characters in a smart storage devices.

References References Referencias

1. VA, B.U.,"IBM Computers to symbolic information processing", V.U VA M.B Loughran. Lee guojia, *Proceedings of the Institute of Electrical and Electronics Engineers*. - 1989. - T.77, N 4. - S. 5-40.
2. Dovgal, V.M., "Methods for modification of formal systems of symbolic information processing", V.M. Dovgal, *Kursk STU 1996*. - 114.
3. Burtsev, V.S., "The parallelism of computational processes and the development of supercomputer

architecture": Sat. Articles / comp. V.P. Torchigin, Y.N. Nikolskaya, Y.V. Nikitin. - Moscow: *TORUS PRESS*, 2006. - 416.

4. Titenko, E.A., "Product system for the implementation of parallel symbolic computation", E.A. Titenko, V.M. Dovgal, *control systems and information technology,* 2006. - № 1 (23) - S. 185-187.
5. Titenko, E.A., "Structural-linguistic approach definitions of Production Systems for the job enumerative nondeterministic computational processes", E.A. Titenko, M. Shilenkov, *Info communication system,* 2009 - № 3. - P.77-80.
6. Dovgal, V.M., "Strategies for fast symbolic computations to enumerative systems of productions", V.M .Dovgal, V.S. Titov, E.A. Titenko *News institute of higher education*. Instruments. - 2008. -№ 2. - P.44-48.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Survey of Elliptic Curve Cryptography Implementation Approaches for Efficient Smart Card Processing

By Jayabhaskar Muthukuru, Prof. Bachala Sathyanarayana
Sri Krishnadevaraya University.Ananthapur,A.P. India

Abstract - Smart cards have been used for many different purposes over the last two decades, from simple prepaid credit counter cards used in parking meters, to high security identity cards intended for national ID programs. This has increased data privacy and security requirements. Data protection and authentication is now demanded for performing Electronic payment and allow secure multi-level access to private information. ECC uses smaller key sizes compared to traditionally used RSA based cryptosystems. Elliptic Curve Cryptography is especially suited to smart card based message authentication because of its smaller memory and computational power requirements than public key cryptosystems. It is observed that the performance of ECC based approach is significantly better than RSA and DSA/DH based approaches because of the low memory and computational requirements, smaller key size, low power and timing consumptions.

Keywords : *symbolic Elliptic Curve Cryptography, finite fields, smart cards, Biometrics.*

GJCST Classification: *C.3*

Strictly as per the compliance and regulations of:

A Survey of Elliptic Curve Cryptography Implementation Approaches for Efficient Smart Card Processing

Jayabhaskar Muthukuru[α], Prof. Bachala Sathyanarayana[Ω]

Abstract - Smart cards have been used for many different purposes over the last two decades, from simple prepaid credit counter cards used in parking meters, to high security identity cards intended for national ID programs. This has increased data privacy and security requirements. Data protection and authentication is now demanded for performing Electronic payment and allow secure multi-level access to private information. ECC uses smaller key sizes compared to traditionally used RSA based cryptosystems. Elliptic Curve Cryptography is especially suited to smart card based message authentication because of its smaller memory and computational power requirements than public key cryptosystems. It is observed that the performance of ECC based approach is significantly better than RSA and DSA/DH based approaches because of the low memory and computational requirements, smaller key size, low power and timing consumptions.

Keywords : Elliptic Curve Cryptography, finite fields, smart cards, Biometrics.

I. Introduction

Smart card is a credit-card sized plastic card with an embedded computer chip. Smart cards play an increasingly important role in everyday life. We encounter them as credit cards, loyalty cards, electronic purses, health cards, and as secure tokens for authentication or digital signatures. Their small size and the compatibility of their form make them ideal carriers of personal information such as secret keys, passwords, customization profiles, and medical emergency information. Electronic Payment is one of the most widely used applications of the smart card and is the most familiar among the average user. There are several different types of smart cards in this category, all of which deal with currency or a fiscal value. Smart cards can provide multi-factor authentication by using PIN/Biometrics combination with the card.

Multi-factor authentication approach is recommended in which security requirements are intended for highly secure installation and mandate a robust solution. Multi-factor authentication ensures verification and validation of a user identity using multiple authentication mechanisms. It often combines two or more authentication methods—for example, a three-factor authentication is based on password (Something you know), smart card (Something you have), and fingerprints (Something you are). For example, in addition to what the user knows (such as a PIN), the card can provide authentication using the card owner's digital certificate with the card owner's public key. The digital certificate associates the card owner's identity to the person's public key. The smart card also contains the card owner's private key, which can be used for digitally signing e-mail or documents. With the support of biometric technologies, the smart card can also be used to store biometric templates of the card owner, which can be used to verify the card owner by acquiring a biometric sample (such as a fingerprint) and matching it to the reference template stored on the card or off the card using a biometric authentication server. Using biometric templates can be considered for security-sensitive applications where PINs can be stolen [1]. Unlike standard public-key methods that operate over integer fields, the elliptic curve cryptosystems operate over points on an elliptic curve. Cryptographic algorithms based on discrete logarithm problem can be efficiently implemented using elliptic curves [21]. ECC is emerging as an attractive public-key cryptosystem for smart cards because compared to traditional cryptosystems like RSA/DH, it offers equivalent security with smaller key sizes, faster computation, lower power consumption, as well as memory and bandwidth savings [2].

II. Smart Card & Architecture

Smart cards come in two varieties: memory and microprocessor. Memory cards simply store data and can be viewed as a small floppy disk with optional security. A microprocessor card, on the other hand, can add, delete and manipulate information in its memory on the card. Similar to a miniature computer, a microprocessor card has an input/output port operating system and hard disk with built-in security features.

a) Contact Vs. Contactless

Smart cards have two different types of interfaces: contact and contactless. Contact smart

Author α : PhD Scholar, Department of Computer Science & Technology, Sri Krishnadevaraya University.
E-mail : jayabhaskarm@gmail.com
Author Ω : Professor, Department of Computer Science & Technology Sri Krishnadevaraya University. Ananthapur, A.P. India.
E-mail : bachalasatya@yahoo.com

cards are inserted into a smart card reader, making physical contact with the reader. However, contactless smart cards have an antenna embedded inside the card that enables communication with the reader without physical contact. A combi card combines the two features with a very high level of security.

b) Basic Smart Card Chip Architecture

The basic smart card architecture is shown on Figure 1. It is a complete set of a microcontroller. It is a small embedded computer with low processing power (8-bit CPU, 5 MHz clock) and small memory (4 Kb RAM, 16 Kb EEPROM, 64 Kb ROM).It is secure and inexpensive [20].

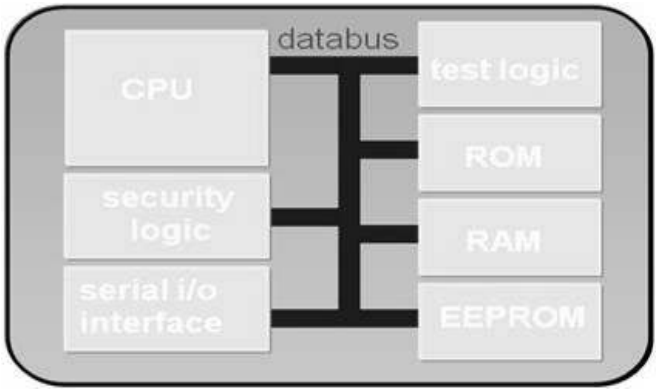


Fig.1 : Basic Smart card Chip Architecture

Smart card components are:

CPU (Central Processing Unit): The heart of the chip, all computational work like implementing cryptographic algorithms and data exchange goes via this function.

Test Logic: A verification function only used during the production process to test all internal circuits for manufacturing faults.

Security Logic: A continuous function that checks environmental conditions that could jeopardise the security of the smart card.

I/O Interface: A communication function that takes care of receiving external commands and sending back responses using a serial communication protocol.

ROM: The permanent memory of the chip. It can contain parts of the operating system and self test procedures.

RAM: The CPU's scratch pad memory. This is used for storing temporary or intermediate data like session keys, internal variables and stack data.

EEPROM: Non-volatile updateable memory. It is used for storing application data like keys, PINs, balances, phone numbers, Biometric template and sometimes application or even operating system code.

Data Bus: The transfer channel within the chip. All information exchanged between the various functions passes through this channel.

III. Biometric Authentication

Biometric technique is an automated methodology for the recognition of a person based on behavioral or physiological characteristics. These characteristics include features such as hand geometry, handwriting, face, fingerprints, vein, voice, retina, and iris. Biometric technologies are now the key to an extensive array of highly secured identification and personal verification solutions. Biometric system is a pattern recognition technology that makes personal identification of an individual by determining the authenticity of a specific physiological or behavioral characteristics possessed by the user [3].

a) Biometric Based Implementation on Smart Card

The use of biometrics within the card itself will mean that biometric features (fingerprint, retina, voice etc) can reliably identify a person. The use of some of these features has already been implemented in many applications. Table 1 below gives the required bytes for various biometric types. Additional information about biometric technology and standards can be found from the following organizations: The Biometric Consortium (www.biometrics.org), International Biometric Industry Association (www.ibia.rg), or BioAPI Consortium (www.iapi com) [4].

Table 1: No. of Bytes required for various Biometric systems

Biometric System	No. of Bytes Required
Finger scan	300-1200
Finger geometry	14
Hand geometry	9
Iris recognition	512
Voice verification	1500
Face recognition	500-1000
Signature verification	500-1000
Retina recognition	96

b) Classification of Biometric Approaches

Main Biometric based smart card implementation approaches are "match-off-card" and "match-on-card".

Match-off-card: For this type of implementation, the enrolled template is initially loaded onto the smart card and then transferred from the smart card via either contact or contactless interface when requested by the external biometric system. The external equipment then compares a new live template of the biometric with the

one retrieved from the smart card. This implementation clearly has some security risks associated with transmitting the enrolled template off of the smart card for every biometric comparison. Appropriate security measures should be implemented to ensure the confidentiality and integrity of the released template.

Match-on-card: This implementation technique initially stores the enrolment template in the smart card's secure memory. When a biometric match is requested, the external equipment submits a new live template to the smart card. The smart card then performs the matching operation within its secure processor and securely communicates the result to the external equipment.

Biometric match-on-card approach can provide more private and secure identity verification system compare to match-off-card approach [5].

V. Elliptic Curve Arithmetic

Elliptic curves are not like an ellipse or curve in shape. They look similar to doughnuts. Geometrically speaking they somehow resemble the shape of torus, which is the product of two circles when projected in three-dimensional coordinates. ECC makes use of elliptic curves in which the variables and coefficients are restricted to elements of a finite field. There are two families of elliptic curves defined for use in cryptography: prime curves defined over odd prime field F_P and binary curves defined over Galois field *GF* (2^m).

a) Geometrical Definition of Point Addition and point Doubling using chord-and-tangent rule

For any two points $P(x_1, y_1) \neq Q(x_2, y_2)$ on an elliptic curve, EC group law point addition can be defined geometrically (Figure 2) as: "If we draw a line through P and Q, this line will intersect the elliptic curve at a third point (-R). The reflection of this point about x-axis, $R(x_3, y_3)$ is the addition of P and Q".

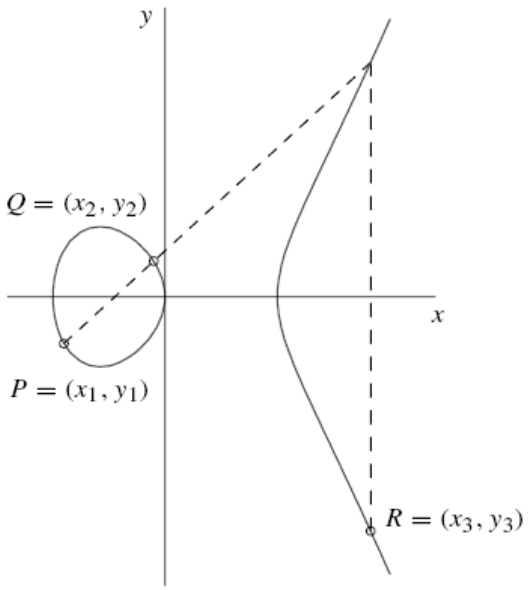


Fig.2 : Addition: R=P+Q

For P=Q, point doubling, geometrically (Figure 3) if we draw a tangent line at point P, this line intersects elliptic curve at a point (-R). Then, R is the reflection of this point about x-axis.

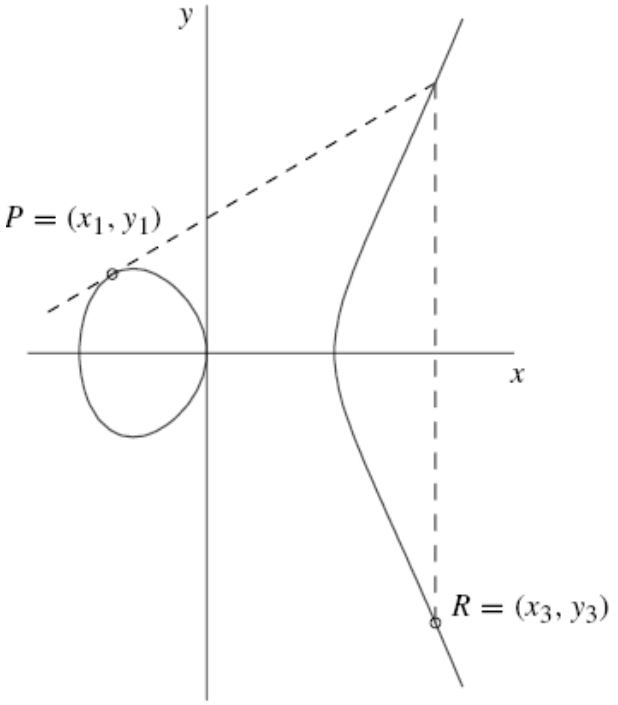


Fig.3 : Doubling: R=P+P

b) Point Multiplication

The dominant operation in ECC cryptographic schemes is point multiplication. This is the operation which is the key to the use of elliptic curves for asymmetric cryptography---the critical operation which is itself fairly simple, but whose inverse (the elliptic curve discrete logarithm) is very difficult. ECC arranges itself so that when you wish to performance operation the cryptosystem should make easy encrypting a message with the public key, decrypting it with the private key the operation you are performing is point multiplication. Scalar multiplication of a point P by a scalar k as being performed by repeated point addition and point doubling for example 7P=(2((2P)+P)+P.

c) Elliptic Curve Over F_P and F_2^m

Definition of elliptic curve over F_P as follows [6].

Let p be a prime in F_P and a, b$\in F_P$ such that $4a^3 + 27b^2 \neq 0$ mod p in F_P, then an elliptic curve $E(F_P)$ is defined as

$$E(F_P) := \{ p(x, y), x, y \in F_P \}$$

Such that $y2 = x^3 + ax + b$ mod p together with a point O, called the point at infinity. Below is the definition of addition of points P and Q on the elliptic curve $E(F_P)$. Let $P(x_1, y_1)$ and $Q(x_2, y_2)$ then

$$R = P+Q = \begin{cases} O & \text{If } x_1 = x_2 \text{ and } y_2 = -y_1 \\ Q = Q+P & \text{If } P = O \\ (x_3, y_3) & \text{otherwise} \end{cases}$$

Where

$$x_3 = \begin{cases} \lambda^2 - x_1 - x_2 & \text{If } P \neq \pm Q \text{ (Point Addition)} \\ \lambda^2 - 2x_1 & \text{If } P = Q \text{ (Point Doubling)} \end{cases}$$

$y_3 = \lambda(x_1 - x_3) - y_1$, and

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \text{If } P \neq \pm Q \text{ (Point Addition)} \\ \dfrac{3x_1^2 + a}{2y_1} & \text{If } P = Q \text{ (Point Doubling)} \end{cases}$$

The point $p(x, -y)$ is said to be the negation of $p(x, y)$.

The elliptic curves over F_2^m is defined as follows.

Denote the (non-super singular) elliptic curve over F_2^m by $E(F_2^m)$. If $a, b \in F_2^m$ such that $b \neq 0$ then

$$E(F_2^m) = \{p(x, y), x, y \in F_2^m\}$$

such that $y^2 + xy = x^3 + ax^2 + b \in F_P^m$ together with a point O, called the point at infinity.

The addition of points on $E(F_2^m)$ is given as follows: Let $P(x_1, y_1)$ and $Q(x_2, y_2)$ be points on the elliptic curve $E(F_2^m)$, then

$$R = P+Q = \begin{cases} O & \text{If } x_1 = x_2 \text{ and } y_2 = -y_1 \\ Q = Q+P & \text{If } P = O \\ (x_3, y_3) & \text{otherwise} \end{cases}$$

Where

$$x_3 = \begin{cases} \lambda^2+\lambda + x_2 + x_1 + a & \text{If } P \neq \pm Q \text{ (Point Addition)} \\ \lambda^2 + \lambda + a & \text{If } P = Q \text{ (Point Doubling)} \end{cases}$$

$$y_3 = \lambda(x_1 + x_3) + x_3 + y_1$$

and

$$\lambda = \begin{cases} \dfrac{y_2 + y_1}{x_2 + x_1} & \text{If } P \neq \pm Q \text{ (Point Addition)} \\ x_1 + \dfrac{x_1}{y_1} & \text{If } P = Q \text{ (Point Doubling)} \end{cases}$$

VI. Elliptic Curve Cryptography for Message Authentication

The use of Elliptic Curve Cryptography was initially suggested by Neal Koblitz [7] and Victor S. Miller [8]. Elliptic curve cryptosystems over finite field have some advantages like the key size can be much smaller compared to other cryptosystems like RSA, Diffie-Hellman since only exponential-time attack is known so far if the curve is carefully chosen [7] [6] and Elliptic Curve Cryptography relies on the difficulty of solving the Elliptic Curve Discrete Logarithm Problem ECDLP, which states that, "Given an elliptic curve E defined over a finite field F_P , a point $P \in E(F_P)$ of order n, and a point $Q \in E(F_P)$, find the integer $k \in [0, n-1]$ such that $Q = kP$. The integer k is called the discrete logarithm of Q to the base P, denoted $k = \log_P Q$".

a) Elliptic Curve Encryption/Decryption

Consider a message 'Pm' sent from A to B. 'A' chooses a random positive integer 'k', a private key 'n_A' and generates the public key $P_A = n_A \times G$ and produces the cipher text 'Cm' consisting of pair of points $Cm = \{kG, Pm + kP_B\}$ where G is the base point selected on the Elliptic Curve, $P_B = n_B \times G$ is the public key of B with private key 'n_B'.

To decrypt the cipher text, B multiplies the 1st point in the pair by B's secret & subtracts the result from the 2nd point $Pm + kP_B - n_B(kG) = Pm + k(n_B G) - n_B(kG) = Pm$.

VII. Various Ecc Implementation Approaches on Smart Card

In [14] Ahmad Khaled M. AL-Kayali demonstrated the advantages and disadvantages of using Prime/binary fields to implement ECC on smart cards. Prime fields are best for software applications where as Binary fields are suitable for Hardware applications [22]. To access remote information systems Password authenticated key agreement scheme [15] is very useful in limited computation and communication resource (smart card) environments. A two-phase authentication mechanism proposed [16] by Abhilasha, Anna Squicciarini, Elisa Bertino. In that first phase consists of a two-factor biometric authentication and second phase combines several authentication factors in conjunction with biometric to provide a strong authentication. A key advantage of this approach is that any unanticipated combination of factors can be used. Disadvantage of using existing remote user authentication schemes [17] [18] is if the smart card is lost and password is revealed then any one can impersonate to sever as authorized user. To overcome this K K Goyal and M S Chahar proposed a new scheme [19] using Biometrics. Table 2 presents ECC based implementations on Smart Card applications.

Table 2 : ECC implementation details on Smart Card

S.No	Implementation Approach	Implemented Field	Aim/Impact	Implemented smart card model
1	Smart cards do not require coprocessor to execute arithmetic operations of ECC but RSA/DSA need additional on chip hardware to avoid long processing delays [11].	Binary Field	Reduces the cost of Smart Card.	Intel 8051 microcontroller
2	ECC implementation, that relies on JAVA card technology and portable solution capable of running on PC and Smart card [12].	Binary Field	Implemented Efficient algorithms on low-resource smart cards.	Bull Odyssey I
3	Efficient implementation of the elliptic curve Digital Signature using optimized point addition and doubling algorithms when a crypto coprocessor for modular arithmetic is available [13].	Prime Field	ECDSA implementation is efficient compare with RSA and it has investigated curves over GF(p) because GF(2^m) field is used for efficient hardware implementation.	Motorola M-smart card Jupiter

a) Comparing ECC with other PKC Schemes

The majority of public key systems in use today use 1024-bit parameters for RSA and Diffie-Hellman. The US National Institute for Standards and Technology [NIST] has recommended that these 1024-bit systems are sufficient for use until 2010. Table 3 shows NIST guidelines on choosing computationally equivalent symmetric and public-key sizes [10].

Table 3 : Comparing ECC with other PKC schemes

Security(bits)	RSA key Length (bits)	ECC key Length (bits)	DSA/DH (bits)	Key Size Ratio of RSA and ECC	MIPS years to attack	Protection attack
80	1024	160-223	1024	1:6	10^{12}	Until 2010
112	2048	224-255	2048	1:9	10^{24}	Until 2030
128	3072	256-383	3072	1:12	10^{28}	Bey-ond 2031
192	7860	384-511	7860	1:20	10^{47}	
256	15360	512+	15360	1:30	10^{60}	

ECC is the best suited in constrained environments. The advantages like speed and smaller keys or certificates are especially important in environments where at least one of the following resources is limited [9]: processing power, storage space, band width, or power consumption. This advantage is because its inverse operation gets harder, faster, against increasing key length than do the inverse operations in Diffie Hellman and RSA.

Table4: Measured performance of public-key algorithms

	ECC-160	RSA-1024	ECC-192	RSA-1536	ECC-224	RSA-2048
Ops/sec	271.3	114	268.5	36.4	195.5	17.8
Performance ratio	2.4 : 1		7.4 :1		11.4 : 1	
Key-size ratio	1 : 6.4		1 : 8		1 : 9.1	

Table 4 shows a comparison of the RSA and ECC cryptographic operations performed by an SSL server. Open SSL speed program is used to measure RSA decryption and ECDH operation for different key sizes (a minor enhancement was made for collecting RSA-1536 numbers). These micro-benchmarks highlight ECC's performance advantage over RSA for different security levels. ECC's performance advantage increases even faster than its key-size advantage as security needs increase [10].

VIII. Conclusion

The smart card market has experienced a spectacular growth over the past few years. Along with their growing popularity there has been a corresponding growth of interest in their security. With respect to end-to-end security no other security solutions nearly as good and affordable as smart cards exist. Elliptic curve cryptography has been emerged as a vast field of interest for application specific security requirements. The elliptic curve discrete logarithm problem makes ECC most efficient compared to earlier RSA/DSA algorithms.

References References Referencias

1. Christopher Steel, Ramesh Nagappan and Ray Lai "Core Security Patterns: Best Practices and Strategies for J2EE™, Web Services, and Identity Management", Pp. 651-652.
2. Efficient and provably-secure identity-based signatures and signcryption from bilinear maps. In Proc. Of ASIACRYPT'05, volume 3778 of LNCS, pages 515–532, 2005.
3. Emmanuel Opara, Mohammad Rob and Vance Etnyre, "Biometric and Systems Security: An Overview of End-To-End Security System", Communications of the IIMA, 2006 Volume 6 Issue 2, PP. 53-58.
4. L. A Mohammed, Abdul Rahman Ramli, V. Prakash and Mohamed B. Daud, "Smart Card Technology: Past, Present, and Future", International Journal of The Computer, the Internet and Management Vol. 12#1 (January – April, 2004) pp 12 – 22.
5. Smart Card Alliance white paper, "Smart Cards and Biometrics" - PAC-11002, March 2011, pp. 8-9.
6. Darrel Hankerson, Alfred Menezes and Scott Vanstone,"Guide to Elliptic Curve Cryptography".
7. N. Koblitz, "EllipticCurve Cryptosystems", Mathematics of Computation, 48, 1987, pp. 203-209.
8. V.Miller, "Uses of Elliptic Curve in Cryptography", Advances in Cryptography, Proceedings of Crypto'85, Lectures notes on Computer Sciences, 218, Springer-Verlag, 1986, pp. 417-426.
9. Vivek Katiyar, Kamlesh Dutta and Syona Gupta, "A Survey on Elliptic Curve Cryptography for Pervasive Computing Environment", IJCA , Volume 11– No.10, December 2010.
10. Vipul Gupta, Douglas Stebila, Stephen Fung, Sheueling Chang, Nils Gura and Hans Eberle, "Speeding up Secure Web Transactions using Elliptic Curve Cryptography (ECC)", Link : http://www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Gupta.pdf.
11. Adam D. Woodbury, Daniel V. Bailey and Christof Paar, "ECC on smart cards without coprocessors", The Fourth Smart Card Research and Advanced Applications(CARDIS 2000) Conference, September 20-22, 2000.
12. Istvan Zsolt BERTA and Zoltan Adam MANN, "Implimentating ECC on PC and smart card", 2002 Link: http://www.crysys.hu/publications/files/BertaM2002pp.pdf
13. Yvonne Hitchcock, Edward Dawson, Andrew Clark and Paul Montague, "Implementing an efficient elliptic curve cryptosystem over GF(p) on a smart card", ANZIAM J. 44 (E) ppC354–C377, 2003.
14. Ahmad Khaled M. Al-Kayali, "Elliptic Curve Cryptography and Smart Cards", SANS Instutute 2004.
15. Aqeel Khalique, Kuldip Singh and Sandeep Sood, "A Password-Authenticated Key Agreement Scheme Based on ECC Using Smart Cards", IJCA, Volume 2 – No.3, May 2010.
16. Abhilasha, Anna Squicciarini and Elisa Bertino, "Privacy Preserving Multi-Factor Authentication with Biometrics", DIM'06, November 3, 2006.
17. D. Jena, S. K. Jena, D. Mohanty and S. K. Panigrahy, "A Novel Remote User Authentication Scheme Using Smart Card based on ECDLP", IEEE Proceeding of International Conference on Advanced Computer Control, 2008.
18. Debasish Jena, Saroj Kumar Panigrahy, Sanjay Kumar Jena and Subhendu Kumar Pani, "Modified Remote User Authentication Scheme using Smart Card based on ECDLP", ICIIS 2009, 28 - 31 December 2009, Sri Lanka.
19. K K Goyal and M S Chahar, "A Novel Remote User Authentication Scheme using Smart Card with Biometric Based on ECDLP" , International Journal of Information Technology and Knowledge Management, July-December 2011, Volume 4, No. 2, pp. 649-651.
20. Hoon K and Ronnie D. Caytiles, "A Review of Smartcard Security Issues", Journal of Security Engineering, Jun-2011, pp. 359-370.
    Branovic, R. Giorgi, E. Martinelli, "A Workload Characterization of Elliptic Curve Cryptography Methods in Embedded Environments", ACM, Vol. 32, No. 3, June- 2004.
21. Wasim A Al-Hamdani, "Elliptic Curve for Data protection", Information Security Curriculum conference Oct-2011, 7-9.

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

LH-Cipher: A Linear Hierarchical Cipher approach for Data

By M Sadanandam
Kakatiya University, Warangal

Abstract - Dividing the data into blocks and there by arranging the blocks in hierarchal order is termed as a linear hierarchical cipher approach for data .The encryption code, a access id code for each level based on the propagation code is generated in this technique. However the level propagation code and the access-id code of previous hierarchy level are matching with upper hierarchical level .The access-id code is set based on the time sensing key and a time seed, and the time seed updates with respect to the encryption pulse. By simply modifying and inversing the data security it is possible to decrease the number of key volume.

GJCST Classification: E.4

Strictly as per the compliance and regulations of:

LH-Cipher: A Linear Hierarchical Cipher approach for Data

M Sadanandam

Abstract - Dividing the data into blocks and there by arranging the blocks in hierarchal order is termed as a linear hierarchical cipher approach for data .The encryption code, a access id code for each level based on the propagation code is generated in this technique. However the level propagation code and the access-id code of previous hierarchy level are matching with upper hierarchical level .The access-id code is set based on the time sensing key and a time seed, and the time seed updates with respect to the encryption pulse. By simply modifying and inversing the data security it is possible to decrease the number of key volume.

I. Introduction

With the growth in communication technology, there have been lot many positives but to counterfeit a lot many techniques to misuse this technology have also been growing. It is important to make data protected from all such malpractice [2]. Of all the methods of protecting the data, encryption is the most effective one. In this technique the data is simply hidden and later it is recovered by de-encryption.[1].to encrypt a data, a specific process or pattern is followed which may include mathematical operations, shifting and substitute techniques .after the data is encrypted it is termed as ciphertext [3].with the help of key based algorithms it is possible to encrypt the data and this key based encryption technique is classified as symmetric and asymmetric. The former uses only a single key for both encryption and de-encryption where as the later uses two different keys each for encryption and de-encryption. There are number of key based Encryption techniques viz. DES, RSA, Elliptic curve, and several other mathematical methods [4, 5]. The wireless communication systems have seen a rapid development in recent times as such Wireless Sensor Network, Bluetooth, zigbee are the most recent ones. The WSN finds its application in monitoring systems especially security concerns.

The WSN constantly sends the information about the state of the object being monitored to the control room that enables collection of related information.

II. Related Work

Multilevel cryptosystems saw a steady growth in recent times. The following are some of the proposed multilevel encryption methods explored in table 1.

The models [1, 2] provide multi level ciphering but the final result so obtained is not generic and databases specific .Linear hierarchical cipher based data encryption and decryption is generic and considers the heterogynous in each level to overcome the drawbacks of the proposed AES and elliptical curve method [3,4,5].

III. Linear Hierarchical Cipher Approach

To ensure data protection in wireless communication encryption is the best technique but today we have lot many users accessing different levels of data. A multi user system has an access to different data levels. For each level we have an encrypted key which is used to de-encrypt and access it. However as the number of levels increases it get difficult to manage with the multiple keys .Hence forth managing the encryption with key technique is termed difficult. It is important to know that the keys are changed every time and hence data security is still ensured .Since the keys used are to be changed each time both level-based keys and time-based keys are to be altered each time.

To resolve the above problem of managing both the time based and level based keys at a time a data encryption technique is explained in detail. As stated earlier the data is initially partitioned into different level. While encrypting the data a specific level we also consider the encryption code of the previous level .Thus the user can de-encrypt the data easily from the already de-encrypted data levels however the data security still holds good. This technique of managing data at various levels is called a linear hierarchical cipher based encryption technique.

Author : Assistant Professor of CSE, KU College Of Engineering, Kakatiya University, Warangal. Telephone: 919440448790,
E-mail : sadanb4u@yahoo.co.in

Method	Proposed by	Special features
Multi level encryption	Zhou Yuping et al [1]	Encrypt the data system, table level and field level of objects
Multi level secondary storage	Chaitanya et al[2]	Flexible performance against security trade-offs
Multi level crpto disk(MLCD)		For generic storage devices
Multi level secure architecture	Sathiaseelan et al[3]	Integrated web services especially for academic institutions
Parallel AES algorithm	Deguang Le et al[4]	Fast Data Encryption on GPU to overcome the draw backs of CPU resource consumption.
ElGamal encryption and transmission scheme	Fu Minfeng et al[5]	It is based on elliptical based cryptosystem that aimed to improve ElGamal algorithm, ECC ElGamal encryption algorithm

Table 1 : Current State of the art in multi level encryption model

Firstly the data is divided into different levels and each level is related to at least one user. Then we encrypt each level by using the encryption key of each level based on the level propagation code and a access-id code of each level. However the level propagation code and the access-id code of one level are generated based on the level propagation code and the access-id code of previous level (the access-id code is produced based on the time propagation code and a time seed). The time seed has to be periodically altered. Then the encrypted data is transferred to the user. This method also includes the generation of encryption code for each level and also other authorized levels based on the level propagation and access id codes and then again decrypting the data at respective levels.

In the paper we elicit a new concept of linear hierarchical cipher based encryption considering a data storage and one encryption module. The data storage generates levels based on the different user approaching for the data access. This also produces the time propagation code, a time seed and a level propagation code based on the propagation codes of the previous levels. With the help of encryption key ,the encryption module encrypts the data with the help of time propagation code, the time seed, and the level propagation code of each level, and thus generates the access-id key (based on the accessed code of previous level) according to the time propagation key and the time seed. All the encrypted data is stored by the data storage. The decryption module finds the related encrypted data block in the related authorized level and then produces the encryption codes to access that level data block with the help of level propagation codes and access id codes .It then decrypts the encrypted data blocks with the help of the propagation level and access id codes which could be generated with the help of previous level codes The data storage also considers the variation in the encryption code with the time with the help of time seed and thus generates the access-id code for each level considering the encryption periods .

IV. Encryption Approach

- Divide data into multitude data blocks
- Generate access-id key of the highest level of the hierarchy
- Sequentially generate the access-id keys of the other levels of the hierarchy using FIPS-180-1 hash standard.
- Generate encryption key of the each hierarchical level based on level propagation and access-id keys of each level
- Encrypt the data block each level using corresponding encryption key
- Send encrypted data-block to data storage

V. Decryption Approach

- Authenticate the user according to user key and find user position in hierarchy
- if authentication succeed then
  - generate an access-id key for the hierarchy level of the user
  - send encrypted data blocks to the authorized levels
  - Encryption time and access-id key to the

decryption module.

- decryption module generates propagation keys for current level and other authorized levels of the hierarchy
- decryption module generates access-id keys of the authorized levels of the hierarchy by using access-id key of the current level
- Respectively decrypts the corresponding encrypted data-blocks according to the level propagation keys and access-id keys of the authorized levels by decryption module

VI. Emblematic Model of LH-Cipher

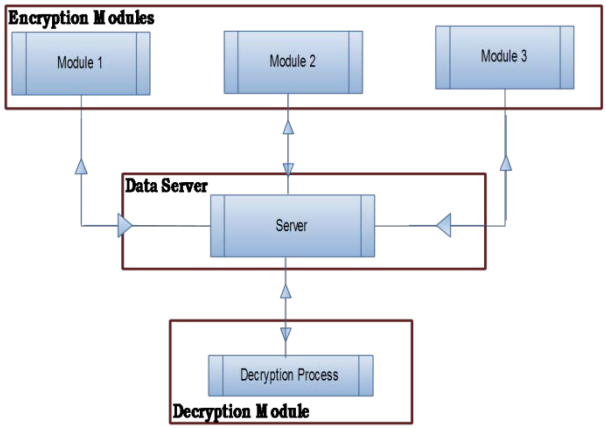


Figure 1 : schematic representation of LH-Cipher

Let us consider an Emblematic Model represented as fig 1, the linear hierarchical cipher system that includes three encryption modules, and data storage. In the selected emblematic model, the LH-Cipher implemented based on an ad hoc network. The considered Emblematic Model consist three nodes with encryption modules those serves as data collectors, and a control device that serves as data storage. In the Ad hoc network, the three nodes collect related data such as images around where they are disposed, and respectively encrypt the data through the first second and the third encryption modules and transmit the encrypted data to the control device that acts as data storage. Whenever required an eligible user can connect to the data storage through the network to read the data recorded therein. The operation of the Ad hoc Network is not in the context of proposal and therefore will not be discussed here. However, it should be understood that the LH-Cipher is not limited only to an ad hoc network; rather it may also be deployed in other constrained device communication environments, such as IEEE802.11 family standard device based networks.

The control device accepts the egress data of the encrypted modules as input and stores in data storage. A generic data access device with a database can be data storage.

As a part of its functionality the data storage controller groups the authorized users into multiple levels so as to manage these users and the data to be accessed by these users based on the users level. It is precise that a user of a level can access more data than a user of its lower levels. In other words, the users in upper level groups will have high end rights when compared to users in lower level groups. An upper level user can decrypt and access the data assigned to his group level and all lower group levels but it is not true in vice versa.

In order to manage the rights of preceding users of diverse levels, the data storage generates a level propagation key for each of the levels to encrypt the data. In particular, the level propagation key of a level is generated according to the level propagation key of its upper level so that the data can be managed based on the user levels.

Let consider that in the selected emblematic model, the data storage groups three different users u1, u2 and u3 into three levels, wherein the u1 belongs to the first level which has the highest right, the u2 belongs to a second level which has the second highest right, and the u3 belongs to a third level which has the lowest right.

The data storage randomly generate a group key $K_{(g)1}$ for the users of the first level that also considered as top level and then sequentially generates group keys $K_{(g)2}$, $K_{(g)3}$ for the other two levels through a fips-180-1 standard hashing technique, as shown below:

$$K_{(g)i} = F(h)^{l-1}(K_{(g)1}),$$

Where in $F(h)$ is a hash function and l represents the level (i.e. l=1..3). The Eq1 represents the FIPS-180-1 standard hashing function that is using to generate the group keys $K_{(g)2}$ and $K_{(g)3}$ of the other two level.

Next, the data storage respectively generates a level propagation key {$PK_{(m,l)}$, 'm' is node identification code and 'l' is level id.} for each level according to the group keys {$K_{(g)1}$,$K_{(g)2}$,$K_{(g)3}$ …$K_{(g)n}$} of the levels and an identification code of the encryption module through the fallowing function.

$$PK_{(m,l)} = f_{(e)}(K_l, m),$$

Wherein $f_{(e)}$ is the encryption function, and l represents the level id.

In the selected emblematic model, the encryption function is any standard encryption function of choice, as a part experimental results we opt to the advanced encryption standard (AES).

It should be mentioned that in the selected emblematic model, the node identification codes are used as one of the factors for generating the level propagation keys because a different level propagation key is provided to each of the node in selected network. However, if the situation of multiple nodes is not

considered or every node uses the same level propagation then as an alternative the group key can be used as the level propagation key.

The data storage also generates a access-id $key(A_k)$ and a time seed besides the level propagation keys. The access-id $key(A_k)$ and the access time as seed are used for generating an access identification $key(AI_k)$ for each encryption period. In the selected emblematic model, a different access identification $key(AI_k)$ is used during each encryption period so that the data to be encrypted can have forward and backward data security. Therefore, a user with expired authorization unable to use his original key to access the data, and can avoid a new authoritative user from accessing data that encrypted in past.

For example, the data storage generates the access-id $key\{A_{(k)m}$, m is device id$\}$ by using a primary key $K_{(p)}$ and an identification code of the wireless sensor through a sixth function. In the selected emblematic model, the encryption function used as shown below:

$$A_{(k)m} = f_{(e)}(K_{(p)}, m),$$

Wherein $f_{(e)}$ is the encryption function. In the selected emblematic model, the encryption function is an standard model of our choice.

Similarly, in the selected emblematic model, the identification codes of the nodes involved are used as one of the factors for generating the access-id $key(A_k)$ because a different access-id $key(A_k)$ is provided to each node involved. However, if the situation of multiple nodes is not considered or each of the nodes uses the same access-id $key(A_k)$, the primary key $K_{(p)}$ can be directly used as the access-id key.

The data storage generates a user key for each of the users and assigns the user key $K_{(u)m}$ to the user while assigning the group key to the user. This user key will be generated with the help of fallowing equation represents an encryption function.

$K_{(u)m} = f_{(e)}(K_{(p)}, m)$, wherein m is user identification.

$K_{(u)m}$ is user key for user identified by m.

$f_{(e)}$ is any encryption function of choice

$K_{(p)}$ is primary key

m is user identification

The primary key $K_{(p)}$ of the data storage is generated randomly. Besides, the data storage generates a different access identification seed S_T corresponding to different encryption periods T. In the selected emblematic model, the S_T corresponding to the current encryption period is generated according to the $K_{(p)}$ and the other parameter of choice such as current date or timestamp.

As described above, all encryption modules are used for encrypting the data to be transmitted by corresponding nodes. The process of encryption fallows.

The first encryption module receives the access-id $key(A_k)$, the time seed S_T, and the level propagation key $\{K_{(L)l}$, l is level identifier$\}$ of each level from the data storage, wherein l represents the level. In the selected emblematic model, the data storage broadcasts a new time seed S_T at certain intervals to the all encryption modules to allow them to generate the access identification keys of the current encryption period T according to the new time seed and the access-id key. Access Identification Key can be generated using the fallowing function.

$AI_{k(m,T)} = f_{(h)} (A_{K(m)}, S_T)$, wherein $f_{(h)}$ is the hash function.

The process of AI_k generation is sequential, that is the first encryption module generates the access identification key of the second level according to the access identification key of the first level and finally generates the access identification key of the third level according to the access identification $key(AI_k)$ of the second level.

The first encryption module divides a data to be transmitted into multitude of sub-data blocks corresponding to different user levels. In addition, the first encryption module generates an encryption key for each level according to the received level propagation key of the level and the access identification key(AIk) generated based on a new seed. Encryption key will be generated by using the fallowing function.

$$K_{(E)\,(m,l,t)} = f_{(h)} (K_{L(m,l)}, f_{(h)}^{l-1} (AI_{k\,(m,t)})),$$

Wherein $f_{(h)}$ is the hash function, and l represents the level and m represents the node id.

The first encryption module uses the encryption key $K_{(E)(1,L,T)}$ (wherein 1 is first node id and L=1..3) of each level for respectively encrypting the sub-data blocks.

If the first encryption module does not receive the new time seed but generates the access identification $key(AI_k)$ by using the old time seed and encrypts the sub- data blocks by using the encryption key generated by using the old access identification key, the data storage determines the time seed after it receives the encrypted sub-data blocks and records the sub-data blocks which are encrypted by using the incorrect time seed as reference for subsequent data decryption. In addition, the data storage broadcasts the current time seed to the first encryption module again if the first encryption module does not use the correct time seed to encrypt the data.

After the encryption modules encrypt the sub-data blocks and the encrypted data is sent to the data storage, the respective users can read the encrypted sub-data blocks stored in the data storage through the decryption module allotted to respective end user device. In the selected emblematic model, the end-user

device is connected to the control device with the choice of network model; here we consider a wired connectivity.

The decryption module reads the encrypted sub-data blocks corresponding to the level of a user and other authorized levels of the user and corresponding to the encryption period from the data storage. To be specific, in the selected emblematic model, a user having higher right can read the data assigned to users having lower rights but a user having lower right cannot read the data assigned to users having higher rights. Thus, the data storage provides the corresponding authorized data to a user according to the level of the user after it authenticates the user according to a user key of the user.

In the selected emblematic model, the data storage generates a access identification $\mathbf{key(AI_k)}$ corresponding to the level of the user and sends $\mathbf{AI_k}$ together with the encrypted sub-data blocks to the decryption module of the end-user device.

The decryption module generates the encryption cipher keys for the authorized levels (i.e., the second level and the third level) of the user according to the level propagation keys and the access identification keys of the authorized levels and decrypts the encrypted sub-data blocks by using the encryption keys. In particular, the decryption module generates the level propagation key and $\mathbf{AI_k}$ of a lower level according to the level propagation key and the access identification $\mathbf{key(AIk)}$ of an upper level.

VII. Conclusion and Future Work

The proposed linear hierarchical ciphering model is robust and scalable where data is encrypted corresponding to multiple levels so that a user having higher right can access the data assigned to users having lower rights but a user having lower right cannot access data assigned to users having higher rights. In addition, in the present invention, an access identification key updated by using a access seeds generated based on access time is adopted to ensure the encrypted data to have forward and backward security and that no synchronous process is required. Thus, the encryption process relaxed from computational complexity. Moreover, the level propagation key and the access identification $\mathbf{key(AI_k)}$ of a lower level are generated according to the level propagation key and the access identification $\mathbf{key(AI_k)}$ of one level up in hierarchy. Thereby, the number of keys to be managed by an end-user device is reduced and accordingly the calculation load of the end-user device is also reduced. In future this solution can be extended to achieve the key generations without considering the sequence in levels of hierarchy.

References References Referencias

1. Freeman J., Neely R., and Megalo L. “Developing Secure Systems: Issues and Solutions”. IEEE Journal of Computer and Communication, Vol. 89, PP. 36-45. 1998.
2. Agnew G. B., Mullin R. C., Onyszchuk I. M., and Vqanstone S. A. “An Implementation for a Fast Public-Key Cryptosystems”. Journal of Cryptology, Vol.3, No 2, PP. 63-79. 1995.
3. Beth T. and Gollmann D. “Algorithm Engineering for Public Key Algorithms”. IEEE Journal on Selected Areas in Communications; Vol. 7, No 4, PP. 458-466. 1989.
4. IBM. “The Data Encryption Standard (DES) and its strength against attacks”. IBM Journal of Research and Development, Vol. 38, PP. 243-250. 1994.
5. IBM. “The Data Encryption Standard (DES) and its strength against attacks”. IBM Journal of Research and Development, Vol. 38, PP. 243-250. 1994
6. Zhou Yuping; Wu Xinghui; , "Research and realization of multi-level encryption method for database," Advanced Computer Control (ICACC), 2010 2nd International Conference on , vol.3, no., pp.1-4, 27-29 March 2010.
7. Chaitanya, S.; Urgaonkar, B.; Sivasubramaniam, A.; , "Multi-level Crypto Disk: Secondary Storage with Flexible Performance Versus Security Trade-offs," Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2010 IEEE International Symposium on , vol., no., pp.434-436, 17-19 Aug. 2010.
8. Sathiaseelan, J.G.R.; Rabara, S.A.; Martin, J.R.; , "Multi-Level Secure Architecture for distributed integrated Web services," Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on , vol.8, no., pp.180-184, 9-11 July 2010.
9. Deguang Le; Jinyi Chang; Xingdou Gou; Ankang Zhang; Conglan Lu; , "Parallel AES algorithm for fast Data Encryption on GPU," Computer Engineering and Technology (ICCET), 2010 2nd International Conference on , vol.6, no., pp.V6-1-V6-6, 16-18 April 2010.
10. Fu Minfeng; Chen Wei; , "Elliptic curve cryptosystem ElGamal encryption and transmission scheme," Computer Application and System Modeling (ICCASM), 2010 International Conference on , vol.6, no., pp.V6-51-V6-53, 22-24 Oct. 2010.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Performance Analysis of Stock Price Prediction using Artificial Neural Network

By K.K.Sureshkumar, Dr.N.M.Elango
Kongu Arts and Science College / Bharathiar University

Abstract - Stock market predictions are one of the challenging tasks for financial investors across the globe. This challenge is due to the uncertainty and volatility of the stock prices in the market. Due to technology and globalization of business and financial markets it is important to predict the stock prices more quickly and accurately. Last few years there has been much improvement in the field of Neural Network (NN) applications in business and financial markets. Artificial Neural Network (ANN) methods are mostly implemented and play a vital role in decision making for stock market predictions. Multi Layer Perceptron (MLP) architecture with back propagation algorithm has the ability to predict with greater accuracy than other neural network algorithms. In this research, neural works predict tools are used to predict the future stock prices and their performance statistics will be evaluated. This would help the investor to analyze better in business decisions such as buy or sell a stock.

Keywords : *Artificial Neural Network (ANN), Multi Layer Perceptron (MLP), National Stock Exchange (NSE), Stock Prediction, Performance Measures.*

GJCST Classification: *I.2.6*

Strictly as per the compliance and regulations of:

Performance Analysis of Stock Price Prediction using Artificial Neural Network

K.K.Sureshkumar [α], Dr.N.M.Elango [Ω]

Abstract - Stock market predictions are one of the challenging tasks for financial investors across the globe. This challenge is due to the uncertainty and volatility of the stock prices in the market. Due to technology and globalization of business and financial markets it is important to predict the stock prices more quickly and accurately. Last few years there has been much improvement in the field of Neural Network (NN) applications in business and financial markets. Artificial Neural Network (ANN) methods are mostly implemented and play a vital role in decision making for stock market predictions. Multi Layer Perceptron (MLP) architecture with back propagation algorithm has the ability to predict with greater accuracy than other neural network algorithms. In this research, neural works predict tools are used to predict the future stock prices and their performance statistics will be evaluated. This would help the investor to analyze better in business decisions such as buy or sell a stock.

Keywords : Artificial Neural Network (ANN), Multi Layer Perceptron (MLP), National Stock Exchange (NSE), Stock Prediction, Performance Measures.

I. Introduction

Stock price prediction is a heated topic in prediction study of financial area. The use of ANN in business environments has been increasing over the last few years. Excellent algorithm has been applied to predict stock price or index. Interest in neural networks has led to a considerable surge in research activities in the past decade. Artificial neural network models are based on the neural structure of the brain. The brain learns from experience and so do artificial neural networks. As a useful analytical tool, ANN is widely applied in analyzing the business data stored in database or data warehouse. Identifying customer behavior patterns and predicting stock price are emerging areas of neural network research and its application. Most of the companies have created new methods of evaluating financial data and investment decisions. Artificial Neural Networks are being used by most companies for improved forecasting capabilities in analysis of stock market. So, artificial neural network suits better than other models in predicting the stock market.

To predict stock prices there are so many conventional techniques can be used, in which fundamental and technical analysis one among them (Atiya, A. F, El-Shoura et al, 1999). Fundamental analysis involves various macro-economic factors, results of the company, financial conditions and other related attributes are used to measure the value of the company with reflect to stock price changes. Technical analysis, on the other hand, involves analyzing statistics generated by market activity, such as past prices and volume (Kai Keng Ang and Chai Quek, 2006). Recent development in soft computing has set a new dimension in the field of financial forecasting. Tools based on ANN have gained more popularity due to their inherent capabilities to approximate any non linear function to a high degree of accuracy.

The idea of forecasting using neural network is to find an approximation of mapping between the input and output data through training. The trained neural network is then used to predict the values for the future (Abhyankar, A. et al, 1997). This research work presents the use of artificial neural network as a forecasting tool for predicting the stock market price.

The remainder of the paper is organized as follows. Section II reviews the background study of the stock market prediction by Artificial Neural Network. Section III focuses on the objectives of the study. Section IV discusses about the basic of Artificial Neural Network; benefits and limitations of ANN were presented. Section V discusses about the Indian Stock market. Section VI explains about data and methodology of using NeuralWorks Predict to predict the stock prices and calculating result performance. Section VII concludes the research.

II. Background Study

In the last two decades lot of research has been done on models based on intelligent soft computing. In general, the approaches to predict stock market could be classified into two classes, fundamental analysis and technical analysis (Kai Keng Ang and Chai Quek, 2006). Fundamental analysis is based on macroeconomic data and the basic financial status of companies like money supply, interest rate, inflationary rates, dividend yields, earnings yield, cash flow yield, book to market ratio, price-earnings ratio, lagged returns (Fama and French, 1988; Lakonishok, 1994). Technical analysis is based on the rationale that history will repeat itself and that and the correlation between price and volume reveals market behavior. Prediction is made by exploiting implications

Author α : Assistant Professor, Department of MCA, Kongu Arts and Science College, Erode -638 107, Tamilnadu, INDIA. Telephone: +91 9842765456, E-mail : kksuresh_oda@yahoo.com
Author Ω : Professor & Head, Department of MCA, RMK Engineering College, Chennai - 601 206, INDIA. Telephone: +91 9600679283 E-mail : nmeoxford@yahoo.com

hidden in past trading activities and by analyzing patterns and trends shown in price and volume charts (Smirlock and Starks, 1985; Brush 1986).

According to (Refenes, Zapranis and Franchis, 1994) "neural networks are capable of making better prediction in capturing the structural relationship between a stock's performance and its determinant factors more accurately than MLR models". (Kryzanowski, Galler and Wright, 1993) using Boltzmann machine trained an artificial neural network with 149 test cases of positive (rise in the stock price) and negative (fall in the stock price) returns for the years 1987-1989 and compared this to training the network with positive, neutral (unchanged stock price), and negative returns for the same 149 test cases for the years 1987-1989. The network predicted 72% correct results with positive and negative returns. However the network predicted only 46% correct results with positive, neutral, and negative returns.

Using neural networks to predict financial markets has been an active research area in both fundamental and technical analysis, since the late 1980s (White, 1988; Fishman, Barr and Loick, 1991; Shih, 1991; Utans and Moody, 1991; Katz, 1992; Kean, 1992; Swales and Yoon, 1992; Wong, 1992; Azoff,1994; Rogers and Vemuri, 1994; Ruggerio, 1994; Baestaens, Van Den Breg and Vaudrey, 1995; Ward and Sherald, 1995; Gately, 1996; Refenes Abu-Mostafa and Moody, 1996; Murphy, 1999; Qi, 1999; Virili and Reisleben, 2000; Yao and Tan, 2001; Pan, 2003a; Pan 2003b).

Fujitsu a Japanese technology company and Nikko Securities - an investment company joined together to develop a stock market prediction system for TOPIX (Tokyo based stock index). The emergence of artificial intelligence techniques has seen their enormous application to financial forecasting, such as expert systems (Tsaih Yenshan Hsu, and Charles Lai, 1998), fuzzy logic (Hiemstra, 1994), and neural networks (Kryzanowski, Galler and Wright, 1993). Among them, neural networks are the most popular and successful tools. There is extensive literature about the application of neural networks in financial forecasting (Azoff, 1994; Goonatilake and Treleaven, 1995; Wong and Selvi, 1998). One of the most popular Journals published on the application of neural networks in finance is the Journal of Computational Intelligence in Finance (Bhagirathi Nayak, et al, 2011).

Also, all of the researches using neural network applications in prediction of stock market trend are mainly based on the assumption that the basic laws in a certain stock market is consistent through the time of experiment data.

III. Objectives of the Study

The main objective of this study is to use NeuralWorks Predict tool to obtain more accurate stock prediction price and to evaluate them with some performance measures. This study can be used to reduce the error proportion in predicting the future stock prices. It increases the chances for the investors to predict the prices more accurately by reducing error percentage and thus gain benefits in share markets.

IV. Artificial Neural Network

Artificial Neural Network (ANN) is an information processing system where the elements called neurons, process the information. The signals are transmitted by means of connection links. The links possess an associated weight, which is multiplied along with the incoming signal (net input) for any typical neural network. The output signal is obtained by applying activations to the net input. The network consists of a set of sensory units that constitute the input layer and one or more hidden layer of computation modes. The input signal passes through the network in the forward direction. This type of network is called as multilayer perceptron (MLP) (Sivanandam, S.N. et al, 2006). The multilayer perceptron are used with supervised learning and have to lead the successful back propagation algorithm where logistic sigmoid function is widely used. The MLP network has hidden neurons and this will make the network more active for complex tasks. The layers of network are connected by synaptic weights and have a high computational efficiency.

a) Benefits of Using Artificial Neural Network

Neural networks often lead to significant results, e.g. in weather forecasting, a rule of weather change is less probable than a steady weather pattern. According to (Schoneburg, 1990), this is also true for stock prices.

A key aspect to successful forecasting lies in the ability to merge data available in diverse formats (Steven H. Kim and Se Hak Chun, 1998). The data analysis performed by neural networks tolerates a considerable amount of imprecise and incomplete input data due to the distributed mode of information processing. Neural network lie in their ability to predict accurately even in situations with uncertain data, and the possible combinations with other methods. Despite the benefits of artificial neural networks, there are still some limitations to neural networks that are discussed below.

b) Limitations of Artificial Neural Network

Some methods are executed with insufficient reliability tests, data design and with inability to identify the optimal topology for a specific problem domain.

There is no known method of designing an optimal neural network, but the best network is highly dependent on the data and application (Carlos Cinca. 1996).

Some of the limitations are mentioned below:

1. NN require very large number of previous data.
2. The best NN architecture topology is still unknown.
3. For complex networks the result and accuracy may

decrease.
4. Statistical relevance of the result is needed.
5. More careful data design is needed and systematically analyzed.

In order to improve the NN applications, there are some other limitations, concerning the problems of evaluation and implementation of NN that should be discussed. Large number of research is done and implemented by companies that are not published in scientific indexes.

V. Indian Stock Market

Investors are mostly preferred the stock market investments because it has the opportunity of highest return over other schemes. For companies, stock market is one of the key sources to raise money through initial public offer (IPO). This allows businesses to be publicly traded, or raise additional capital for expansion by selling shares of ownership of the company in a public market. Indian stock market is mainly consists of two major stock indices, Bombay Stock Exchange (BSE) and National Stock Exchange (NSE). The benchmark for these two exchanges are Sensex (30 Stocks) and Nifty (50 Stocks).

BSE was the first stock exchange in the country and approved under the Securities Contract Regulation Act, 1956. Sensex is an index of 30 stocks with 12 major sectors. In the year 1993, National Stock Exchange of India has been the frontier of Indian securities market. NSE is located at Mumbai, India referred as Nifty. Nifty is a well diversified index consisting of 50 major stocks from 21 sectors of the economy (Refer NSE, 2010). It is the largest stock exchange in India in terms of daily turnover and number of trades, for both equities and derivative trading. Trading on both exchanges is carried out in dematerialized form.

Securities and Exchange Board of India (SEBI) is the regulatory authority and have the rights to monitor all the stock markets in India established by Government of India in the year 1988. The main goal of the board is to protect the investors in securities and regulate the stock market. There are 23 stock exchanges in India, out of that only 18 stock exchanges are currently in the operative mode. Among 18 exchanges BSE and NSE are considered to be the primary exchanges of India.

VI. Data And Methodology

The actual problem discussed in this paper is to forecast the stock price of National Stock Exchange in India. For this purpose we have used available daily stock data of TCS (i.e., bhavcopy) from the National Stock Exchange beginning from 01-November-2009 to 12-December-2011 (Refer NSE, 2011).

For this study, we select 508 day's NSE stock data of TCS Company. The data field used in this research consists of previous close, open price, high price, low price and close price. In order to predict the stock price, past data is necessary and it has been collected for the trading days from 01-November-2009 to 12-December-2011. The historical data set is available on the National Stock Exchange website.

The main task is to predict the stock price of TCS will be up or down for tomorrow by using the historical values of the company stock. In this research, NeuralWorks Predict version 3.24 packages are applied to predict the future stock price of TCS The historic data of previous close, open price, high price, low price and closing price data is used. NeuralWorks Predict 3.24 tool is used throughout the process, this research choose 5 important attributes including previous close, open price, high price, low price and closing price. The performance of the neural network largely depends on the architecture of the neural network. Issues critical to the neural network modeling like selection of input variables, data pre-processing technique, network architecture design and performance measuring statistics should be considered carefully.

a) Methodology of Building a Predict Model

General steps of building and predicting the value by using Multi Layer Perceptron model in the NeuralWorks Predict.

1. Building a Predict Model: To make predictions from data if target outputs can be any value in a continuous range of numeric values or a discrete ordered range of numeric values.
2. Selection of model: Multi Layer Perceptron (MLP) model is selected to predict the stock value.
3. MLP Input training data
4. MLP Output training data
5. MLP Training data characteristics
6. MLP Network parameters
7. Reviewing parameters and training the model
8. Saving the model
9. Training statistics
10. Testing a predict model
11. Specifying data sets for testing
12. Interpreting test results
13. Running a MLP predict model

b) Results and Performance Statistics

Performance statistics that are computed for prediction model train and test sets are shown in Table1.

Close Price	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9971	0.995268	9.872307	49.1059	12.68418	1	24.72697	508
Train	0.997092	0.995172	9.816971	49.1059	12.68724	1	24.76381	355
Test	0.99713	0.995504	10.0007	37.47821	12.67708	1	24.87982	153

Table.1 : Train and Test Results of TCS

In Table 1, R Correlation (R) is the linear correlation between predicted outputs and target

outputs, in problem domain units. Average Absolute Error (Avg Abs) denotes the average absolute difference between predicted output values and target output values. Maximum Absolute Error (Max Abs) is the maximum absolute difference between a predicted output value and a target output value. The Root Mean Square Error (RMS) is the error between the predicted outputs and the target outputs. Accuracy is the percent of predicted output values that lie within 20% of their corresponding target output values. Confidence Intervals (Conf Interval) 95% of the model predictions lie within the range around target output values bounded by the confidence intervals and number of records processed. Finally, records indicate the number of records processed during training or testing.

The close correlation between the market value predicted by the neural network and the true value suggests that such networks may indeed become very powerful tools in financial applications. In this study, a real world output range is calculated whose limits are the minimum and maximum of all real world targets and real world model outputs. This range is used in several of the analysis results as shown in the Table 2.

Close Price	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9971	0.995268	9.872307	49.1059	12.68418	1	24.72697	508
Primary	0.9971	0.995268	9.872307	49.1059	12.68418	1	24.72697	508
Secondary	0.9971	0.995268	9.872307	49.1059	12.68418	1	24.72697	508
Train	0.997092	0.995172	9.816971	49.1059	12.68724	1	24.76381	355
Test	0.99713	0.995504	10.0007	37.47821	12.67708	1	24.87982	153
Valid	0.9971	0.995268	9.872307	49.1059	12.68418	1	24.72697	508

Table.2 : Results Interpretation and Performance Statistics

c) Output Summary

The result of the predicted value has been shown in the Fig. 1. In Fig. 1, the actual close price of TCS have compared with the predicted price. Here, days refer to 508 values for each day starting from (19-November-2008 to 14-December-2010).

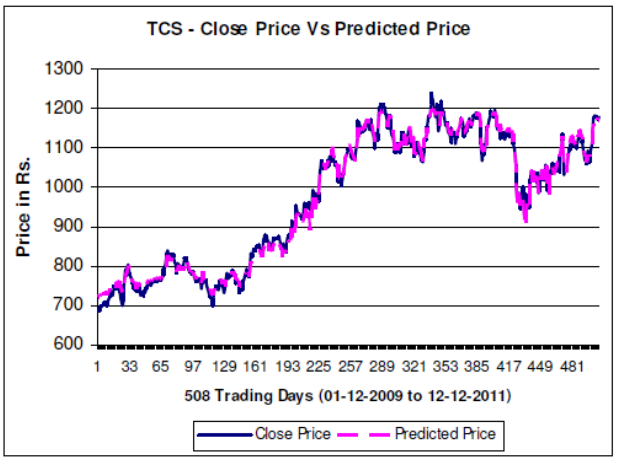


Fig.1 : Comparison of Actual Vs Predicted Price

The error percentage rate of the actual close price and predicted price of TCS as shown in Fig. 2.

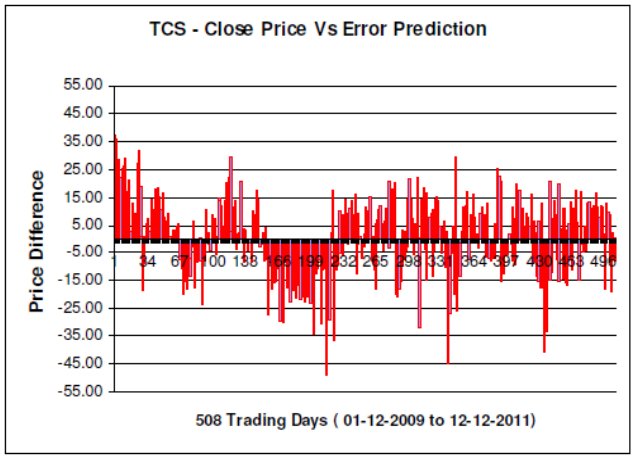


Fig.2 : Error Percentage rate of Actual Close Price Vs Predicted Price

There are many different methods to measure performance of systems. In order to evaluate the net performance of the stock value some of the following indicators have to be considered.

The indicators are R, Net-R, Average Absolute, RMS, Accuracy measures and Confidence limits. Net-R measure is the linear correlation between the real world target output and the raw neural net output. RMSE is a basic measure is used to find out the difference between values predicted by a model and the value actually observed.

We have used NeuralWorks Predict package tool for training, testing and predicting the stock prices. It is found that the percentage of correct prediction has been made and the result of this analysis is shown in the Table 2.

The train and test data sets are selected from the primary and secondary working sets, which will preliminary, trim the data sets. The following are the outcomes of all test and train set data.

All	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.9971	0.995268	9.872307	49.1059	12.68418	24.72697
2	0.93295	0.933051	20.20458	37.47821	21.68696	43.78338
3	0.920938	0.920919	5.275382	18.66345	6.604043	13.04181
4	0.932548	0.932497	10.21521	27.56531	11.92409	24.05034
5	0.839133	0.839492	16.69528	30.37744	17.85807	36.75226
6	0.788178	0.787327	13.77196	49.1059	17.27659	35.55556
7	0.805987	0.808338	11.81858	40.94519	15.69253	32.45215
8	0.851505	0.853173	9.347101	33.25348	11.24513	22.4312
9	0.904238	0.904655	9.318429	25.60205	11.07237	21.91745
10	0.85448	0.859078	6.983322	31.7677	9.281714	18.3149
11	0.613367	0.615451	11.67487	44.85999	15.94526	32.97479

Table.3 : Test and Train of all Working Set Data

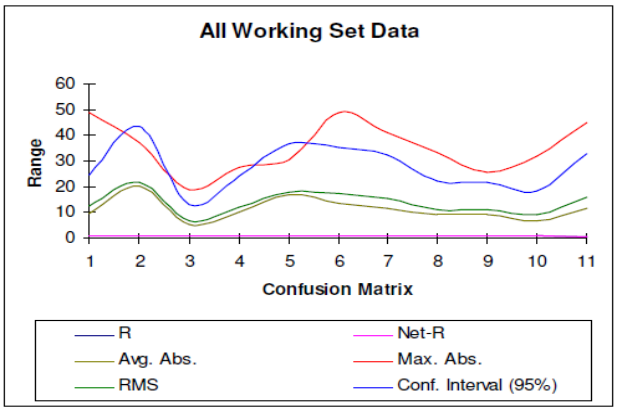


Fig.3 : Graph of all Working set Data

Primary	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.9971	0.995268	9.872307	49.1059	12.68418	24.72697
2	0.93295	0.933051	20.20458	37.47821	21.68696	43.78338
3	0.920938	0.920919	5.275382	18.66345	6.604043	13.04181
4	0.932548	0.932497	10.21521	27.56531	11.92409	24.05034
5	0.839133	0.839492	16.69528	30.37744	17.85807	36.75226
6	0.788178	0.787327	13.77196	49.1059	17.27659	35.55556
7	0.805987	0.808338	11.81858	40.94519	15.69253	32.45215
8	0.851505	0.853173	9.347101	33.25348	11.24513	22.4312
9	0.904238	0.904655	9.318429	25.60205	11.07237	21.91745
10	0.85448	0.859078	6.983322	31.7677	9.281714	18.3149
11	0.613367	0.615451	11.67487	44.85999	15.94526	32.97479

Table.4 : Interpretation of Primary Working Set Data

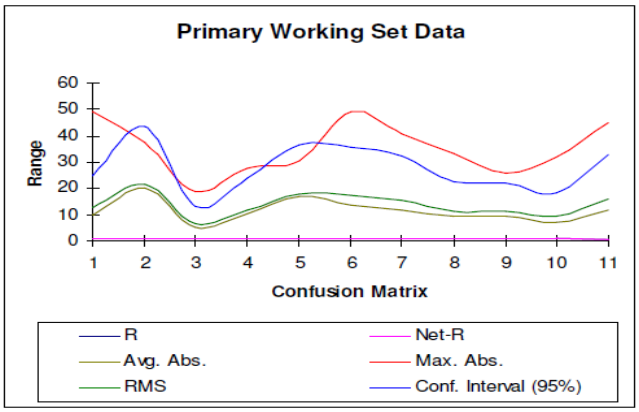


Fig.4 : Graph of Primary Working Set Data

Secondary	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.9971	0.995268	9.872307	49.1059	12.68418	24.72697
2	0.939141	0.93923	20.14999	36.18195	21.49552	44.14341
3	0.916674	0.91667	5.283045	18.66345	6.705967	13.33403
4	0.947125	0.947079	9.806209	23.55493	11.26785	23.0938
5	0.818945	0.819175	17.88929	30.37744	19.00811	40.15015
6	0.808768	0.809089	12.9965	49.1059	16.87607	35.487
7	0.780828	0.783043	13.4198	40.94519	17.92326	38.2675
8	0.883901	0.886189	9.083319	33.25348	10.99159	22.16953
9	0.899728	0.89992	9.392587	25.60205	11.10217	22.14604
10	0.874985	0.878135	6.668418	21.67969	8.546409	16.97648
11	0.730967	0.732311	11.01076	44.85999	15.92176	33.80093

Table.5 : Interpretation of Secondary Working Set Data

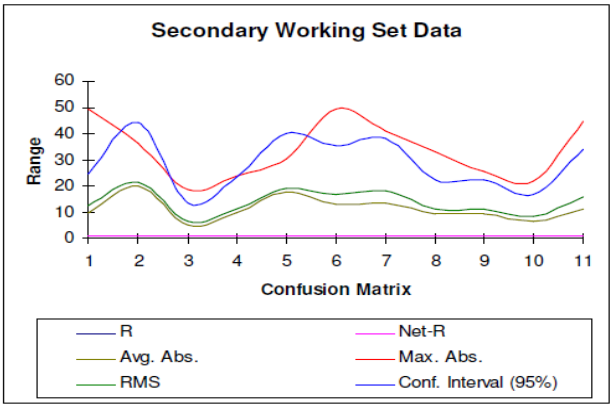


Fig.5 : Graph of Secondary Working Set Data

Train	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.997092	0.995172	9.816971	49.1059	12.68724	24.76381
2	0.921331	0.921463	20.32742	37.47821	22.11164	48.79345
3	0.932825	0.932779	5.257592	13.69232	6.361141	13.03735
4	0.898039	0.897986	11.16954	27.56531	13.33025	29.41568
5	0.909343	0.909957	14.00878	23.06158	14.95047	35.41038
6	0.752458	0.749573	15.87678	34.11078	18.31965	44.84638
7	0.935137	0.936032	8.616142	13.91046	9.812246	23.24043
8	0.762268	0.761667	9.967763	22.94531	11.82026	25.09368
9	0.91631	0.917143	9.142303	20.4364	11.00127	22.75063
10	0.821826	0.828276	7.697107	31.7677	10.76402	21.98249
11	0.437907	0.440396	13.28768	26.85999	16.00216	39.1732

Table.6 : Interpretation of Training Set Data

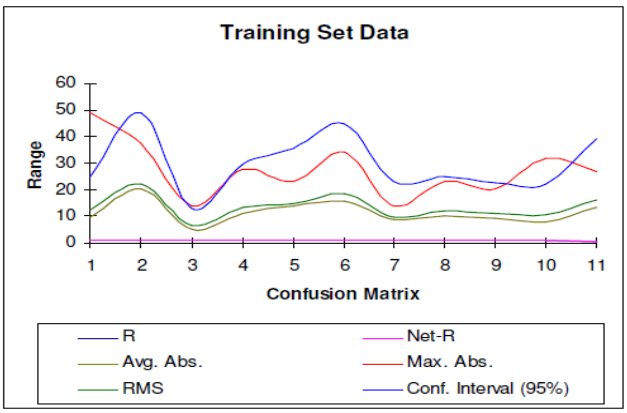


Fig.6 : Graph Training Set Data

Test	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.99713	0.995504	10.0007	37.47821	12.67708	24.87982
2	0.93295	0.933051	20.20458	37.47821	21.68696	43.78338
3	0.920938	0.920919	5.275382	18.66345	6.604043	13.04181
4	0.932548	0.932497	10.21521	27.56531	11.92409	24.05034
5	0.839133	0.839492	16.69528	30.37744	17.85807	36.75226
6	0.788178	0.787327	13.77196	49.1059	17.27659	35.55556
7	0.805987	0.808338	11.81858	40.94519	15.69253	32.45215
8	0.851505	0.853173	9.347101	33.25348	11.24513	22.4312
9	0.904238	0.904655	9.318429	25.60205	11.07237	21.91745
10	0.85448	0.859078	6.983322	31.7677	9.281714	18.3149
11	0.613367	0.615451	11.67487	44.85999	15.94526	32.97479

Table.7 : Interpretation of Test Set Data

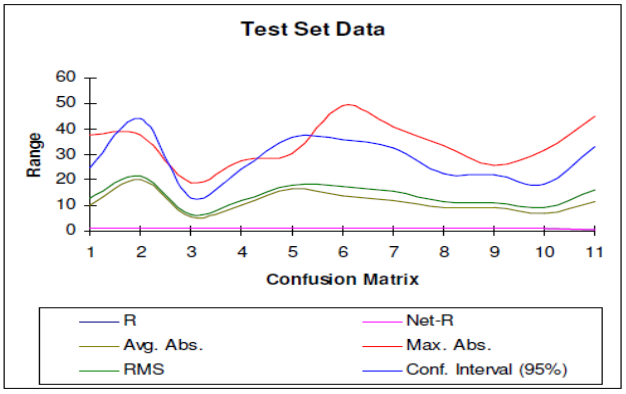


Fig.7 : Graph of Test Set Data

Close Price	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Conf. Interval (95%)
1	0.93295	0.933051	20.20458	37.47821	21.68696	43.78338
2	0.920938	0.920919	5.275382	18.66345	6.604043	13.04181
3	0.932548	0.932497	10.21521	27.56531	11.92409	24.05034
4	0.839133	0.839492	16.69528	30.37744	17.85807	36.75226
5	0.788178	0.787327	13.77196	49.1059	17.27659	35.55556
6	0.805987	0.808338	11.81858	40.94519	15.69253	32.45215
7	0.851505	0.853173	9.347101	33.25348	11.24513	22.4312
8	0.904238	0.904655	9.318429	25.60205	11.07237	21.91745
9	0.85448	0.859078	6.983322	31.7677	9.281714	18.3149
10	0.613367	0.615451	11.67487	44.85999	15.94526	32.97479

Table.8 : Interpretation of Predicted Price

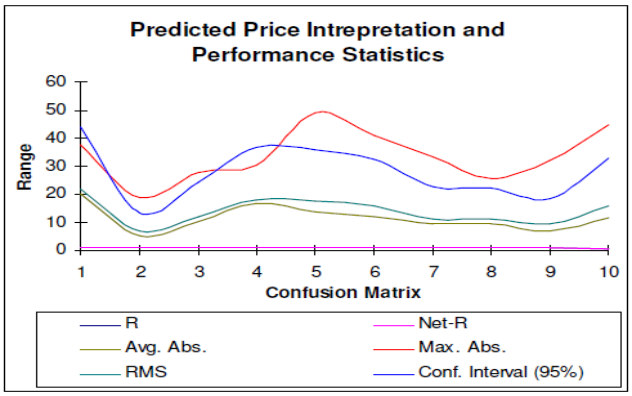


Fig.8 : Interpretation of Predicted Price and Performance Statistics

VIII. Conclusion

In this research, we examined and applied multilayer perceptron model by using the NeuralWorks Predict tool. The results from analysis shows that NeuralWorks Predict offer the ability to predict the stock prices more accurately than the other existing tools and techniques. The accuracy of the predicted output values that lie within 20% of their corresponding target output value. By using this tool one can have the ability to forecast the stock price of NSE more accurately. This analysis can be used to reduce the error percentage in predicting the future stock prices. It increases the chances for the investors to predict the prices more accurately by reducing the error percentage and hence increase their profit in share markets. Utilizing neural network models together with other forecasting tools and techniques can be considered yet another valuable advancement in the age of technology.

References Références Referencias

1. Abhyankar, A., Copeland, L. S., & Wong, W. (1997). "Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100". *Journal of Business & Economic Statistics*, 15, 1–14.
2. Atiya, A. F, El-Shoura, S. M, Shaheen, S. I and El-Sherif, M. S. (1999). "A comparison between neural network forecasting techniques case study: river flow forecasting," in *IEEE Transaction Neural Networks* pp. 402-409, 10-2.
3. Azoff, E.M. (1994): *Neural Network Time Series Forecasting of Financial Markets*. Chichester; New York: Wiley.
4. Baestaens, D.E., Van Den Berg, W.M. and Vaudrey, H. (1995). "Money market headline news flashes, effective news and the DEM/USD swap rate: An intraday analysis in operational time". *Proc. 3rd Forecasting Financial Markets Conference*, London.
5. Bhagirathi Nayak, Dr. C. Nahak , Dr. Arun KR. Misra, (2011). "Forecasting of Financial Markets – Application of Fuzzy Association Rules", *International Journal of Research in Commerce IT & Management, Vol No. 1, Issue No. 5, ISSN 2231-5756.*
6. Brush, J. (1986), "Eight Relative Strength Models Compared," *Journal of Portfolio Management*, 21-28.
7. Carlos Serrano-Cinca. (1996). "Self organizing neural networks for financial diagnosis". *Decision Support Systems*, 17:227–238.
8. Fama, E.F. and K.R. French. (1988a). "Permanent and Temporary Components of Stock Prices." *Journal of Political Economics, vol. 96, no. 2, pp.* 264-273.
9. Fama, E.F., and K.R. French. (1988b). "Dividend Yields and Expected Stock returns," *Journal of Financial Economics, vol. 22, pp.* 3-25.
10. Fishman, M.B., Barr, D.S. and Loick, W.J. (1991). "Using neural nets in market analysis". *Technical Analysis of Stocks and Commodities* 9(4):18–22.
11. Gately, E. (1996). *Neural networks for financial forecasting*. New York7 John Wiley & Sons.
12. Goonatilake, S. & Treleaven, P. (ed.), (1995). *Intelligent systems for finance and business*, Wiley, New York.
13. Hiemstra, Y. (1994). "A stock market forecasting support system based on fuzzy logic", *Proceedings of the Twenty-Seventh Hawaii International*

*Conference on System Sciences, 1994. Vol. III: Information Systems: Decision Support and Knowledge-Based Systems, Volume 3, 4-7 Page(s):*281 – 287.
14. Kai Keng Ang and Chai Quek. (2006)."Stock Trading Using RSPOP: A Novel Rough Set-Based Neuro-Fuzzy Approach". *IEEE Transactions of Neural Networks*, 17(5):1301–1315.
15. Katz, J.O. (1992). "Developing neural network forecasters for trading". *Technical Analysis of Stocks and Commodities* 10(4).
16. Kean, J. (1992). "Using neural nets for intermarket analysis". *Technical Analysis of Stocks and Commodities* 10(11).
17. Kryzanowski, L., Galler, M., Wright, D.W. (1993). "Using Artificial Neural Networks to Pick Stocks", *Financial Analysts Journal, July-August 1993. pp.*21-27.
18. Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, (1994). "Contrarian investment, extrapolation, and risk", *Journal of Finance* 49, 1541–1578.
19. Murphy, J. (1999). "A Comprehensive Guide to Trading Methods and Applications". *Technical Analysis of the Financial Markets:* Prentice Hall Press.
20. National Stock Exchange Historical Data for TCS Stock. Retrieved December 12, 2011 from http://www.nseindia.com/content/equities/scripvol/datafiles/01-12-2009-TO-12-12-2011TCSALLN.csv
21. National Stock Exchange of India Fact Book (2010). Retrieved October 1, 2011, from NSE Website: http://www.nseindia.com/archives/us/fact/us_factbook2011.htm
22. Pan, H.P. (2003a). "Swingtum – A computational theory of fractal dynamic swings and physical cycles of stock market in a quantum price-time space". Ibid.
23. Pan, H.P. (2003b). "A joint review of technical and quantitative analysis of the financial markets towards a unified science of intelligent finance". *Proceedings of 2003 Hawaii International Conference on Statistics and Related Fields*, June 5–9, Hawaii, USA.
24. Qi, M. (1999). "Nonlinear predictability of stock returns using financial and economic variables". *Journal of Business and Economic Statistics* 17:419–429.
25. Refenes, A. P., Abu-Mostafa, Y., and Moody, J. (1996). "Neural Networks in Financial Engineering". *Proceedings of the 3rd International Conference on Neural Networks in the Capital Markets*. WEIGEND A. (eds). World Scientific.
26. Refenes, Zapranis, and Francis. (1994). "Stock Performance Modeling Using Neural Networks: A Comparative Study With Regression Models", *Journal of Neural Networks, Vol. 7, No. 2, 1994. pp.* 375-388.
27. Rogers, R. and Vemuri, V. (1994). "Artificial Neural Networks Forecasting Time Series". *IEEE Computer Society Press, Los Alamitos, CA.*
28. Ruggerio, M. (1994). "Training neural nets for intermarket analysis". *Futures*, Sep:56–58.
29. Schoneburg, E. (1990). "Stock prediction using neural networks: A project report". *Neurocomputing*, 2:17–27.
30. Shih, Y.L. (1991). "Neural nets in technical analysis". *Technical Analysis of Stocks and Commodities* 9(2):62–68.
31. Sivanandam, S.N., Sumathi, S., Deepa, S.N., (2006). *Introduction to Neural Networks using MATLAB 6.0.* New Delhi: Tata McGraw-Hill Publishing Company Limited.
32. Smirlock, M. and Laura Starks. (1985). "A Further Examination of Stock Price Changes and Transaction Volume". *Journal of Financial Research*, 8: 217-225.
33. Steven H. Kim and Se Hak Chun. (1998). "Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index". *International Journal of Forecasting*, 14:323–337.
34. Swales, G.S. and Yoon, Y. (1992). "Applying artificial neural networks to investment analysis". *Financial Analysts Journal* 48(5).
35. Tsaih, Yenshan Hsu, Charles C. Lai. (1998). "Forecasting S&P 500 stock index futures with a hybrid AI system", *Decision Support Systems 23, pages:* 161–174.
36. Utans, J. and Moody, J.E. (1991). "Selecting neural network architectures via the prediction risk: application to corporate bond rating prediction". *Proceedings of the 1st International. Conference on AI Applications on Wall Street*, IEEE Computer Society Press.
37. Virili, F. and Reisleben, B. (2000). "Nonstationarity and data preprocessing for neural network predictions of an economic time series". *Proceedings of the International. Joint Conference on Neural Networks 2000*, Como, 5: 129–136.
38. Ward, S. and Sherald, M. (1995). "The neural network financial wizards". *Technical Analysis of Stocks and Commodities*, Dec:50–55.
39. White, H. (1988). "Economic predictions using neural networks: the case of IBM daily stock returns". *Proceedings of IEEE International Conference on Neural Networks*, 2: 451–458.
40. Wong, F.S. (1992). "Fuzzy neural systems for stock selection". *Financial Analysts Journal*, 48:47–52.
41. Wong, B. and Y., Selvi. (1998). "Neural network applications in finance: A review and analysis of literature (1990-1996)", *Information & Management* 34, 129-139.
42. Yao, J.T. and Tan, C.L. (2001). "Guidelines for financial forecasting with neural networks". *Proceedings of the. International Conference on Neural Information Processing*, Shanghai, China, 757–761.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Intentional Software Product Line

By Sami Ouali, Naoufel Kraiem, Henda Ben Ghezala
National School of Computer Sciences University, Tunisia

Abstract - Software product line engineering optimizes the development of individual systems by leveraging their common characteristics and managing their differences in a systematic way. These differences are called variabilities. We argue that it is difficult for business people to fully benefit of the SPL if it remains at the software level. The paper proposes a move towards a description of software product line in intentional terms, i.e. intentions and strategies to achieve business goals. We present ISPL, the model to describe intentional Software Product Line. Thereafter, we propose our process to show how to use this model.

Keywords : *Software Product Line, variability, intentional level, comparison framework, features modeling and metamodels.*

GJCST Classification: *D.4.6,*

Strictly as per the compliance and regulations of:

Intentional Software Product Line

Sami Ouali [α], Naoufel Kraiem [Ω], Henda Ben Ghezala [β]

Abstract - Software product line engineering optimizes the development of individual systems by leveraging their common characteristics and managing their differences in a systematic way. These differences are called variabilities. We argue that it is difficult for business people to fully benefit of the SPL if it remains at the software level. The paper proposes a move towards a description of software product line in intentional terms, i.e. intentions and strategies to achieve business goals. We present ISPL, the model to describe intentional Software Product Line. Thereafter, we propose our process to show how to use this model.

Keywords : Software Product Line, variability, intentional level, comparison framework, features modeling and metamodels.

I. Introduction

Software product line engineering optimizes the development of individual systems by leveraging their common characteristics and managing their differences in a systematic way (Clements & Northrop, 2001). These differences are called variabilities. In software product line engineering, two kinds of variability can be distinguished: product line variability and Software variability. Software variability refers to the ability of a software system to be efficiently extended, changed, customized or configured for use in a particular context (Svahnberg et al., 2005). While product line variability describes the variation between the systems that belong to a product line (Coplien et al., 1998; Pohl et al., 2005; Kang et al., 2002) in terms of properties and qualities, like features that are provided or requirements that are fulfilled. Defining product line variability concerns the determination of what should vary between the systems in a product line. In SPLE, single system can be built rapidly from reusable assets, such as a set of components.

The framework analysis which we proposed in our previous work (Ouali et al., 2011) allows us to identify many drawbacks of existing SPL construction methods. In these methods, apart requirement approaches ones, the problem is the matching between users' needs and the product offered by developers. Many writers have observed that there is a "conceptual mismatch" (Woodfield, 1997; Kaabi, 2007). The position adopted in this paper is to suggest a move to intention-driven SPL to bridge the gap between high level users' goals and low level software product line obtained. We present in this paper a model for intentional SPL modeling.

Our process is based on goal modeling, feature modeling and metamodels. Goal models model stakeholder intentions to fulfill the system-to-be. Feature modeling allows us to model the common and variable properties of product-line members throughout all stages of product-line engineering. Metamodels allow the expression of common and variable characteristics of a set of applications. A metamodel represents the concepts, relationships, and semantics of a domain.

This paper is organized as follows. A brief description of different concept concerning software product line and variability is presented in the next section. Our previous work, which is the comparison framework, is described in section 3. An intentional software product line model is presented in section 4. In section 5 we present our proposed process. The section 6 concludes this work with our contribution and research perspectives.

II. Software Product Line and Variability Concepts

Software product lines are recognized as a successful approach to reuse in software development (Clements & Northrop, 2001; Bosch, 2000). The idea behind software product line is to economically exploit the commonalities between software products, but also to preserve the ability to vary the functionality between these products. These differences refer to the variability which is a key success factor in product lines and reuse. This approach is based on the undertaking of the development of a set of products as a single, coherent development activity. Indeed, products are built from a collection of artifacts from a core asset base that have been specifically designed for use. Core assets include not only the architecture and its documentation but also specifications, software components, tools…

Variability is the ability of a system to be efficiently extended, changed, customized or configured for use in a particular context (Van Grup, 2000). Another definition presents variability as the ability of a system, an asset, or a development environment to support the production of a set of artifacts that differ from each other in a preplanned fashion (Czarnecki & Eisenecker, 2000). In this definition variability means the ability of a core asset to adapt to usages in the different product contexts that are within the product line scope. Indeed,

Author [α] : National School of Computer Sciences University, Tunisia. Telephone: +21623695033 E-mail : samiouali@gmail.com
Author [Ω] : Higher Institute of Computer Science of El Manar and RIADI Labs, Tunisia. E-mail : Naoufel.kraiem@ ensi. rnu.tn
Author [β] : department of Informatics at the National School of Computer Sciences of Tunis and the director of RIADI Labs, Tunisia. E-mail : henda.bg@cck.rnu.tn.

variations in a product line context must be anticipated.

The purpose of Variability modeling is to present an overview of a product line's commonality and variability. Variability modeling terms concerns also commonality modeling. The content of a variability model serves as a basis for defining variability within the artifacts that make up the product-line infrastructure as well as for configuring individual product instances and deriving them from the infrastructure.

SPL engineering is defined (Czarnecki & Eisenecker, 2000) by distinguishing two levels of engineering: Domain Engineering and Application Engineering as presented in Fig. 1.

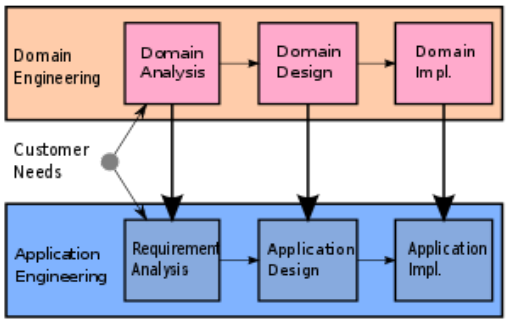


Fig.1 : SPL Engineering levels

Domain Engineering corresponds to the study of the area of product line, identifying commonalities and variabilities among products, the establishment of a generic software architecture and the implementation of this architecture. Indeed, the domain engineering consists on the construction of reusable components known as asset which will be reused for the products building.

Application Engineering is used to find the optimal use for the development of a new product from a product line by reducing costs and development time and improve the quality. At this level, the results of the domain engineering are used for the derivation of a particular product. This derivation corresponds to the decision-making towards the variation points.

In the literature, the majority of variability research concerns requirements and architecture. But some works deals with implementation, verification and validation, traceability and software product line management. The literature basically proposes methods or techniques that address only a specific portion of SPL development.

III. Comparison Framework

We have elaborated a framework to compare different approaches for the construction of SPL. The idea is to consider a central concept (SPL) on four different points of view. Defining a comparison framework has proved its effectiveness in improving the understanding of various engineering disciplines (process, requirements, information systems…) (Rolland, 1998; Jarke & Pohl, 1993). Therefore, it can be helpful for the better understanding of the field of engineering SPLs. As a result, our framework (Fig. 1) is presented in (Ouali et al., 2011).

The framework analysis allows us to identify the following main drawbacks of existing SPL construction methods. We realize that we have a lack of sufficient tool support for them and for their interactivity with their users. The SPL approaches themselves are not enough automated for deriving automatically a product from a SPL. In addition, these methods didn't cover all aspects of SPL engineering. Indeed, every method tries to focus on a particular part of SPL construction process. Finally, in these methods, apart requirement approaches ones, the problem is the matching between users' needs and the product offered by developers. Many writers have observed that there is a "conceptual mismatch" (Woodfield, 1997; Kaabi, 2007).

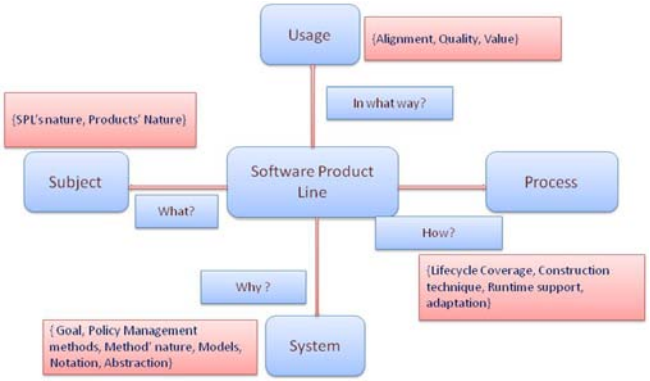


Fig.2 : Software Product Line comparison framework evolution

We try in the next section of this paper to resolve this last drawback by the proposal of a model for intentional SPL modeling. We try to establish the matching between users' needs and the product offered by developers by the expression of users' needs in an intentional way.

IV. Intentional Software Product Line Meta-Model

This section describes a meta-model synthesizing the different interesting points that we previously identified after a state-of-the-art (software product line, intention, feature…). We chose to transform this meta-model into a UML profile to facilitate the integration into UML models and to use it in our MDA approach.

a) Meta-model Description

As depicted in Fig. 3, a *product line* contains *features*. A *product* belongs to one *product line* and is composed of *features*. These *features* associated to a product must check some constraints (mutual exclusion

and require relation) throw the conflict and require relationships. The recommends relationship concerns another feature that could be pertinent.

An intentional software product line is a set of *features* captured at the business level, in business comprehensible terms and described in an intentional perspective. In this perspective, we focus on the *intention* it allows to achieve rather than on the functionality it performs. A *feature* is a set of related *requirements* that allows the user to satisfy an *intention*. We have two specializations of features which are *MandatoryFeature* and *VariantFeature*. Mandatory features are features which must be present in every configuration of a product from the product line.

A variant feature is modeled as a set of variation point. The metamodel allows atomic variation points *(Variant)* or composite ones *(Composite VariationPoint)* for a variant feature. We use the composite pattern to compose a variation point.

In our meta-model, we use a part of an existing meta-model map (Rolland et al., 1999c) which is a Process Model in which a non-deterministic ordering of intentions and strategies has been included. Map is a labeled directed graph with intentions as nodes and strategies as edges between intentions. A *map* consists of a number of sections. Each section is a triplet formed by a source *intention*, a target *intention* and a *strategy*. A *strategy* is a manner to achieve an *intention*.

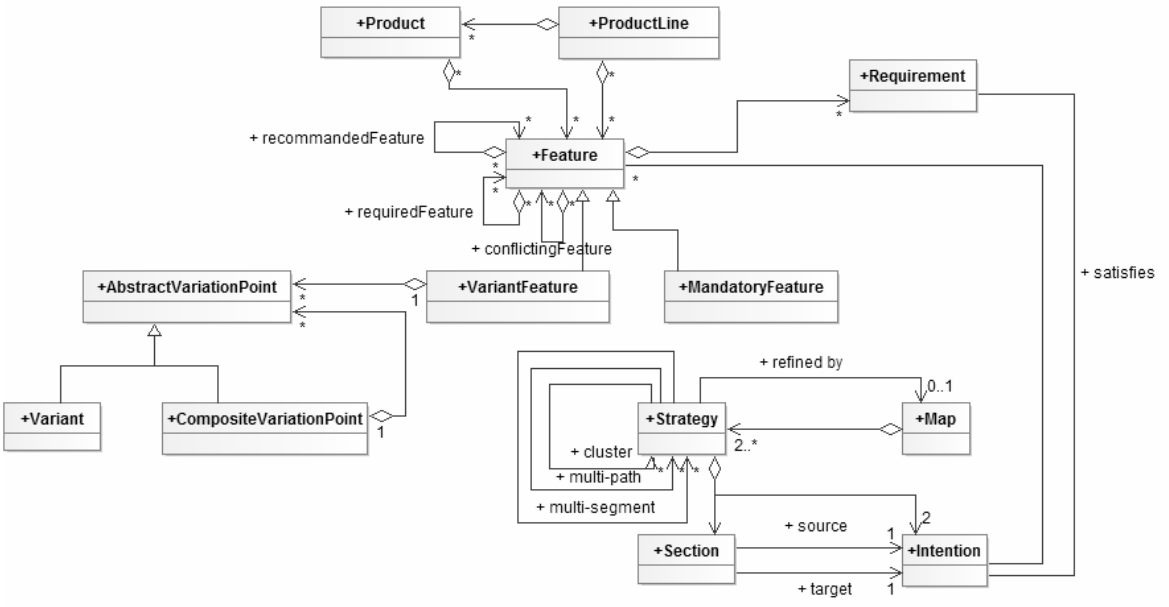


Fig.3 : Above is the example of single column image. Images must be of very high quality.

V. Proposed Process

To avoid the drawbacks of the existing methods, we try to propose a new process for the construction of SPL. This process is a flexible approach for automatically building SPL based on variability models. This process is based basically on goal modeling, features modeling, metamodels, constraints…

In our process, we try to cover domain engineering and application engineering. The domain engineering process involves the creation of core assets. In this process, our interest concerns the elicitation of intentions and strategies using the MAP for the design of users' requirements. A map is a process model expressed in a goal driven perspective which can provides a process representation system based on goals and strategies. The directed nature of the graph shows which goals can follow which one. MAP is considered as Intention-oriented process modeling which follows the human intention of achieving a goal as a force which drives the process (Soffer & Rolland, 2005). Having represented software product line features intentionality as maps, we will proceed in our process to determine features and their composition according to the Intentional Software Product line. This approach is presented in Fig. 4. Users' intentions are captured and modeled using Map Model to obtain an SPL Model. This model contains an intentional view. Variability in intentional software product line modelling is mandatory and due to the need to introduce flexibility in intention achievement. We use features diagrams to model variability in software product line. We try to capture commonality and variability of domain and to reuse it for the derivation of a specific requirement model in application Level.

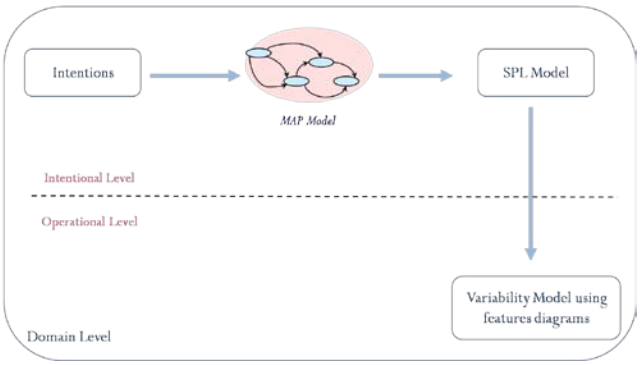


Fig.4 : Domain Level Engineering using Intentional Software Product Line Model

We try to manage variability in SPL construction process (functions, structures, behaviors, technologies). Our strategy follows feature modeling approach, MDA approach and the managing of the constraints. We base our work on the creation of features models representing the SPL structure. We use state machine to model the behavior in the SPL. This process is based on the automatic transformation of models until obtaining executable applications. The process is flexible because SPL developer has a lot of possibilities for the creation of SPL and its constraints. It permits the generation of a flexible SPL suitable to the users' requirements elicited in the beginning of the creation process and new ones.

VI. Conclusion

In this paper, our contribution was the proposal of a model combining software product line, variability, requirements and intentions. This suggested model clarifies the notion of an intentional software product line to model SPL in intentional context. It was build to respond to the following purpose: to focus on the intention it allows to achieve rather than on the functionality it performs. An intentional software product line is captured at the business level, in business comprehensible terms and described in an intentional perspective. This model will be useful to improve the method used for software product line construction by avoiding the conceptual mismatch. We try to establish the matching between users' needs and the product offered by developers by the expression of users' needs in an intentional way.

In this paper, we have presented a proposal to manage variability during the SPLs construction process using a MAP for goals modeling, features diagrams allows us to model the common and variable properties of product-line members throughout all stages of product-line engineering, metamodels allow the expression of common and variable characteristics of a set of applications.

Our future work will be the proposal of a tool support to improve interactivity with users and to cover the overall lifecycle of SPL. This tool support will be based on Eclipse plug-in for feature modeling using the Eclipse Modeling Framework (EMF), which significantly reduced our development effort. Our tool support is based on generative development for goal modeling, feature modeling and metamodels. Integrating goals modeling, feature modeling and metamodels as part of a development environment helps to optimally support modeling variability in different artifacts including implementation code, models, documentation, development process guidance...

References References Referencias

1. Clements, P. & Northrop, L. (2001). *Software Product Lines: Practices and Patterns*. Addison-Wesley, Boston, MA.
2. Svahnberg, M., Van Gurp, J., Bosch J. (2005). A taxonomy of variability realization techniques, In: *Software Practice & Experience*, Vol. 35, No. 8, pp. 705–754.
3. Coplien, J., Hoffman, D., Weiss, D. (1998). Commonality and variability in software engineering. In: *IEEE Software*, Vol. 15, No. 6, pp. 37 – 45.
4. Pohl, K., Böckle, G., Van der Linden, F. (2005). *Software Product Line Engineering: Foundations, Principles and Techniques*, Springer.
5. Kang, K. C., Lee, J., Donohoe, P. (2002). Feature-oriented project line engineering. In: *IEEE Software*, Vol. 19, No. 4, pp. 58–65.
6. Woodfield, S. N. (1997). *The Impedance Mismatch between Conceptual Models and Implementation Environments*, ER'97 Workshop on Behavioral Models and Design Transformations: Issues and Opportunities in Conceptual Modeling, UCLA, Los Angeles, California.
7. Kaabi, R. (2007). *Une Approche Méthodologique pour la Modélisation Intentionnelle des Services et leur Opérationnalisation* (Thèse de doctorat). Sorbonne: Université de Paris I.
8. Ouali, S., Kraïem, N. and Ben Ghezala, H. (2011). A Flexible Process for SPL construction. *Journal of Computer Science and Engineering (JCSE)*, Volume 8, Issue 1.
9. Rolland, C. (1998). *A Comprehensive View of Process Engineering*, Proceeding of the 10th International Conference CAiSE'98, LNCS 1413, Springer Verlag Pernici, C. Thanos (Eds), Pisa, Italie, p. 1-24.
10. Jarke, M. & Pohl, K. (1993). *Requirements Engineering: An Integrated View of Representation, Process and Domain*, in Procedings 4th Euro. Software Conf., Springer Verlag.
11. Bosch, J. (2000). Design & Use of Software Architectures, Adopting and Evolving a product-line approach, Addison-Wesley, ISBN 0-201-67494-7.
12. Van Grup, J. (2000). *Variability in Software Systems, the key to software reuse* (Thesis). Sweden: University of Groningem.

13. Czarnecki, K. and Eisenecker, W. (2000). Generative Programming: Methods, Tools, and Applications. Addison-Wesley.
14. Rolland, C., Prakash, N. and Benjamen, A. (1999c). A Multi-Model View of Process Modelling, *Requirements Engineering Journal*, 4:4, 169-187.
15. Soffer, P. & Rolland, C. (2005). *Combining Intention-Oriented and State-Based Process Modelling*, Proceedings of the International Conference on ER'05, LNCS, pp47-62.

Global Journal of Computer Science and Technology Volume XII Issue I Version I January 2012
32

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Towards full protection of web applications based on Aspect Oriented Programming

By Elinda Kajo Mece, Lorena Kodra
Polytechnic University of Tirana, Tirana, Albania

Abstract - Web application security is a critical issue. Security concerns are often scattered through different parts of the system. Aspect oriented programming is a programming paradigm that provides explicit mechanisms to modularize these concerns. In this paper we present a technique for detecting and preventing common attacks in web applications like Cross Site Scripting (XSS) and SQL Injection using an aspect oriented approach by analyzing and validating user input strings. We use an aspect to capture input strings and compare them to predefined patterns. The intrusion detection aspect is implemented in AspectJ and is woven into the target system. The resulting system has the ability to detect malicious user input and prevent SQL Injection and Cross Site Scripting. We present an experimental evaluation by applying it to an insecure web application. The results of our tests show that our technique was able to detect all the attempted attacks without generating any false positives.

Keywords : *symbolic information, artificial intelligence, Flow control, Architecture.*

GJCST Classification: *D.4.6, K.6.5,H.2.7*

Strictly as per the compliance and regulations of:

Towards full protection of web applications based on Aspect Oriented Programming

Elinda Kajo Mece[α], Lorena Kodra[Ω]

Abstract - Web application security is a critical issue. Security concerns are often scattered through different parts of the system. Aspect oriented programming is a programming paradigm that provides explicit mechanisms to modularize these concerns. In this paper we present a technique for detecting and preventing common attacks in web applications like Cross Site Scripting (XSS) and SQL Injection using an aspect oriented approach by analyzing and validating user input strings. We use an aspect to capture input strings and compare them to predefined patterns. The intrusion detection aspect is implemented in AspectJ and is woven into the target system. The resulting system has the ability to detect malicious user input and prevent SQL Injection and Cross Site Scripting. We present an experimental evaluation by applying it to an insecure web application. The results of our tests show that our technique was able to detect all the attempted attacks without generating any false positives.

Keywords : symbolic information, artificial intelligence, Flow control, Architecture.

I. Introduction

User and critically important company information is managed using web applications. For this reason, web applications serve as a door for attacks. The vulnerabilities present in the application can be exploited by an attacker. Even with the rapid development of Internet technologies, web applications have not achieved the desired security levels. As a result, web servers and web applications are popular attack targets.

Two common attacks on this type of systems are Cross Site Scripting (XSS) and SQL Injection. SQL Injection is a technique where an intruder injects SQL code into the user input field in order to modify the original structure of the query to post hidden data, or execute arbitrary queries in the database. Cross Site Scripting occurs when an intruder injects and executes scripts written in languages like JavaScript or VBScript.

Aspect Oriented Programming is a programming paradigm that provides explicit mechanisms to modularize crosscutting concerns (behavior that cuts across different divisions of the software) such as security. This makes it a good candidate for applying security to a system.

In this paper, we propose an Aspect Oriented protection system that detects and prevents attacks on web applications. This system analyzes and validates user input strings. We use an aspect to capture input strings and compare them to predefined patterns. The intrusion detection aspect is implemented in AspectJ and is woven into the target system. The resulting system has the ability to detect malicious user input and prevent SQL Injection and Cross Site Scripting. The advantage in using aspect oriented programming lies in separating the security code from application code. In this way it can be developed independently to adapt to new attacks.

The rest of the paper is organized as follows. Section 2 presents principles of SQL Injection, XSS and AOP. Section 3 presents related work in this area and our proposed solution. Section 4 describes in detail the architecture of our system and its integration with the web application. Section 5 describes the experimentation and evaluation results. Section 6 concludes and discusses some future work.

II. Background

a) SQL Injection

SQL Injection consists in inserting malicious SQL commands into a parameter that a web application sends to a database in order to execute a malicious query. As a result, database contents can be corrupted or destroyed. The most popular techniques used in SQL injection are tautology, union, and comments.

The general idea behind tautology is finding a disjunction in the WHERE clause of a SELECT or UPDATE statement and inserting malicious code into one or more conditional statements so that they always evaluate as true. Let us consider the case where the web application authenticates users by executing the following query:

```
SELECT * FROM users WHERE username = 'admin' and password = 'pass'
```

This query doesn't select any rows because the password is incorrect. Injecting `' OR 1=1` gives:

```
SELECT * FROM users WHERE username = 'admin' and password = '' OR 1=1'
```

This causes the WHERE clause to be true for every row and all table rows are returned.

The UNION clause allows the chaining of two separate SQL queries. An attacker can use this clause to manipulate an SQL statement into returning rows from another table. As an example, consider the following query that allows users to get the product name by inserting the product ID.

Author α Ω : Department of Computer Engineering, Polytechnic University of Tirana, Tirana, Albania.
E-mails : ekajo@fti.edu.al, lorena.kodra@gmail.com

```
SELECT productName FROM products WHERE productID = '5'
```

An attacker can use the UNION clause to modify the structure of this query to:

```
SELECT productName FROM products WHERE productID = '5' UNION SELECT username, password FROM users
```

As a result, this query will display the product name together with the usernames and passwords of the users table.

Another type of SQL Injection uses comments to cut an SQL query and change its structure. The part of the SQL statement that comes after the comments will not be executed and the query will return the results that the attacker wanted. For example the following SQL statement:

```
SELECT * FROM users WHERE username = 'alice' and password = 'alice123'
```

can be transformed in the following way:

```
SELECT * FROM users WHERE username = 'admin' -- and password = ''
```

The query will return all the information about the admin user.

b) Cross Site Scripting

Cross Site Scripting (XSS) is an attack done towards the user's browser in order to attack the local machine, steal user information or to spoof the user identity. The attacker uses a web application to send malicious code usually in the form of a script. Together with the legitimate content, the users get the malicious script from the web application. This attack is successful in web applications that do not validate user input.

c) Aspect Oriented Programming and Security

Aspect Oriented Programming is a programming paradigm whose aim is to solve problems like code scattering and code tangling that cannot be solved by traditional programming methodologies. Code scattering means that the problem code is spread over multiple modules. This means that when developers want to fix a bug they have to modify several source files. Code tangling means that the problem code is mixed with other code. In the case of web applications, security code needs to be applied in different modules of the system. This process is error prone and difficult to deal with. AOP is a good candidate for applying security in web applications. The security code can be encapsulated into modules called aspects which can be maintained separately from the web application in order to adapt to new attacks.

III. Related Work and Proposed Solution

During recent years, different solutions have been proposed to address security issues in web applications. The most efficient way to protect against XSS and SQL Injection attacks is to inspect all the data the user inserts into the system, hence most of the work in this area treats user input.

Zhu and Zulkerine propose a model-based aspect-oriented framework for building intrusion-aware software systems [2]. They model attack scenarios and intrusion detection aspects using an aspect-oriented Unified Modeling Language (UML) profile. Based on the UML model, the intrusion detection aspects are implemented and woven into the target system. The resulting target system has the ability to detect the intrusions automatically.

Mitropoulos and Spinellis propose a method for preventing SQL Injection attacks by placing a database driver proxy between the application and its underlying relational database management system [1]. To detect an attack, the driver uses stripped-down SQL queries and stack traces to create SQL statement signatures that are later used to distinguish between injected and legitimate queries. The driver depends neither on the application nor on the RDBMS.

Hermosillo et al. present "AProSec" implemented in AspectJ and in the JBoss AOP framework, a security aspect for detecting SQL Injection and XSS [3]. They use the same aspect for dealing with SQL Injection and XSS. Their experiments show the advantage of runtime platforms such as JBoss AOP for changing security policies at runtime.

We propose a system that performs a two-step validation of user input. In the first step it is validated syntactically to check whether it contains dangerous characters that can be used in XSS and SQL Injection. In the second step, the input is validated by the SQL validator in the context of a query to check whether it contains always true statements, comments or combinations of SQL keywords. In contrast to the systems described above, our system analyzes directly user input before it is being used as part of an SQL query. This facilitates the analyzing process. Another advantage of our system is the fact that the SQL validator checks the presence of SQL keywords in the user input. This prevents attacks that do not contain comments or always true statements but contain SQL keywords that can modify the original structure of the SQL query. Our system does not generate false positives because it considers as attack the presence of a combination of SQL keywords and not the presence of a single SQL keyword such as "Union" that might be part of a legitimate user name.

IV. System Architecture

Our system consists of three parts. The first and the most important part is an aspect called WebAppInputFilter that contains the logic of the whole defense process. It defines the advices that control the validation process as well as the steps to be taken (code to be executed) based on the results of the

validation. The aspect also contains the pointcuts that define the vulnerable points of the web application and allow the weaving with the advice code. The second part consists of a validators class that validate against XSS and SQL Injection attacks the input defined in the advices. The third part consists of an encoder which encodes dangerous characters by converting them to their decimal equivalent, leaving them harmless.

The basic idea behind our technique is to capture user input and validate it by comparing it to predefined patterns. In the case of SQL Injection, in contrast with current solutions [1, 2, 3], the user input is validated before being used as part of a query. The final query is a combination of user input and a partial SQL statement defined by the developer. We consider as safe the part of the query that is defined by the developer, so there is no need to validate it and we only validate the user input part. This facilitates and speeds up the evaluation process.

The validation process happens in two steps. First the user input is validated to check whether it contains dangerous characters such as '<',' >', '=' and' –' that can be used to perform XSS and SQL Injection attacks. In the second step, the SQL Validator analyzes the input in the context of the query. This is done to check whether the query contains combined SQL keywords that can modify the original structure of the query or SQL code that can transform the original query in an SQL statement that results always true.

Figure 1 shows the flow of information within the defense system. The aspect captures the user input string and sends it to the first analyzer. If the string is not dangerous it is passed on to the second validation step. If the string is dangerous it is send to the encoder. It encodes the dangerous characters and the result is passed to the SQL Validator. If the string is not considered dangerous, it is passed on to the web application as a legitimate request. If it is considered dangerous, it is erased.

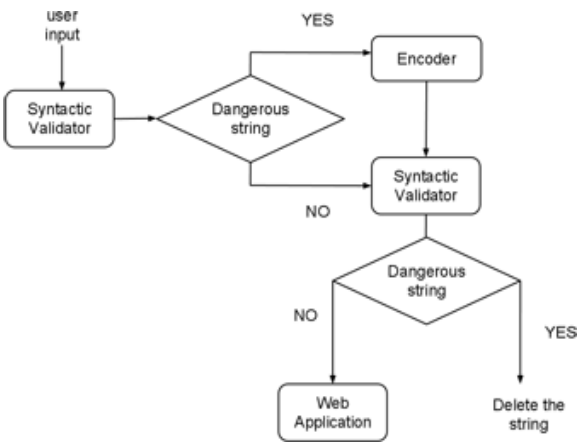


Fig. 1 : The flow of information within the defense system

a) The WebAppInputFilter Aspect

This aspect is implemented in AspectJ [7]. This is the most widely used language for aspect oriented programming. It represents the extension of Java for dealing with aspects. The aspect defines pointcuts in the vulnerable points of the web application. It monitors the traffic in servlets and captures some specific calls that implement the *ServletRequest* and *HttpServletRequest* interfaces. The pointcuts are:

```
pointcut pcGetParameter(): call(String
javax.servlet.http.HttpServletRequest.ge
tParameter (String))

pointcut
pcGetParameterValues():call(String []
javax.servlet.ServletRequest.getParamete
rValues(String))
```

b) The Validators

The validators class handles both XSS and SQL Validators. It uses regular expressions and pattern matching to validate user input against specific patterns.

The syntactic validator, analyzes separately each character of the user input string and acts as a filter that allows only characters 'a-z', 'A-Z', numbers '0-9', spaces and characters like "." and ",". The rest of the characters are considered dangerous and will be sent to the encoder.

The SQL Validator consists of several validation strings in the form of regular expressions that are matched against user input according to different possibilities of injecting SQL code into the user input field of the web application. The validation criteria include: always true comparisons (both string and numeric), presence of quotes or comments, keywords for executing stored procedures, combinations of SQL keywords like UNION, SELECT, DROP, INSERT, ALL, etc. As regards this least evaluation criterion, it protects in cases where no comments or always true statements are present in the query but it still may contain dangerous keywords that can execute arbitrary operations in the database. We would also like to emphasize that the SQL Validator doesn't simply detect the presence of SQL keywords, but the presence of combined SQL keywords that would potentially modify the original structure of the query. This means that input strings that simply contain SQL keywords (like UNION) will not be considered dangerous unless they contain some other SQL keyword that would create a risk for SQL Injection. This eliminates the false positive case of detection when a legitimate user has for example the word "Union" in their name.

V. Evaluation Results

We evaluated our system by using it against a vulnerable web application [8]. First we tried all sorts of SQL Injection and XSS injection attacks to see how the system behaved. Then we protected it using our system

but were unable to bypass the application's security.

For example, let's assume that an attacker tries to input the following script into the web application:

```
<script>alert(document.cookie)</script>
```

The system will detect the dangerous characters "<", ">", "(", ")" and "/" and encode them. In this way this input string will be considered as a simple string and not as a script and will not be interpreted by the browser. A wiser attack would be to encode the input string by using some encoding scheme (decimal, hexadecimal, octal, Unicode, etc) prior to inserting it into the web application. For example, the above string in hexadecimal format (\xNN) would be:

```
\x3c\x73\x63\x72\x69\x70\x74\x3e\x61\x6c
\x65\x72\x74\x28\x64\x6f\x63\x75\x6d\x65
\x6e\x74\x2e\x63\x6f\x6f\x6b\x69\x65\x29
\x3c\x2f\x73\x63\x72\x69\x70\x74\x3e
```

Even in this case the attack wouldn't be successful because the system detects the usage of "\" and encodes the string to make it harmless. We tested our defense system by using other encodings (decimal, octal and Unicode) and none of the attacks were successful.

In the case of SQL Injection, let's assume that an attacker tries to inject a query that contains a statement that is always true into the system:

```
SELECT * FROM user_data WHERE last_name
= 'Smith' OR '1'='1'
```

The SQL Validator will detect that there is a statement that is always true and will delete this string without passing it to the web application.

In order to evaluate the impact of the defense system in the performance of the web application we measured its response time using [9] under two scenarios. We measured the response time first in the absence of any defense and then in the presence of our defense system. We used a mix of input strings: harmless, XSS attack and SQL Injection attack strings. For every scenario we used 356 POST and 104 GET requests which make a total of 460 requests. We executed the series of requests 5 times and measured the average response time. Our defense system introduced an average overhead of 2.11%. We feel that this is an acceptable level of overhead for use in many production environments and it will not be noticeable by the user.

VI. Conclusions and Future Work

We have presented our approach for building a security system for a web application. This system detects XSS and SQL Injection attacks in requests. Our system was built separately and the initial code of the web application was not modified. This allows the separation of security concerns and allows the security system to be evolved independently from the web application to adapt to new attacks.

As an advantage to similar solutions, besides checking for comments and always true statements, our SQL Validator also checks for the presence of a combination of SQL keywords in the input string. This can protect in cases where comments or always true statements are not present in the query but it still may contain dangerous keywords that can execute arbitrary operations in the database. Our system does not simply check for SQL keywords but for a combination of them. This is considered as an advantage in eliminating false positives like in the case of having for example the word "Union" as part of a legitimate user name. Furthermore, in contrast to usual solutions, when protecting against SQL Injection our system analyzes directly the user input before being used as part of a query. There is no need to analyze the whole query because the other parts of it are defined by the developer and are considered safe. This has the advantage of facilitating and speeding up the evaluation process.

Our system can be improved in some directions. A possible improvement might be the implementation of defense against other form of attacks. Also new techniques like *machine learning* and *neural networks* can be used to detect more sophisticated attacks. Another direction of improvement might be the implementation of *runtime weaving* using the JBoss AOP Framework [10].

References Références Referencias

1. M. Dimitris, and Diomidis Spinellis, "SDriver: Location-specific signatures prevent SQL injection attacks", Computers & Security, Vol.28, No. 3-4, 2009, pp. 121-129.
2. J. Zh. Zhi, and Z. Mohammad, "A model-based aspect-oriented framework for building intrusion-aware software systems", Information and Software Technology, Vol.51, N.5, 2009, pp. 865-875.
3. H. Gabriel, G. Roberto, S. Lionel, and D. Laurence, "AProSec: An aspect for programming secure web applications", in Proceedings of the The Second International Conference on Availability, Reliability and Security, 2007, pp. 1026-1033.
4. K. Engin, J. Nenad, K. Christopher, and V. Giovanni, "Client-side cross-site scripting protection", Computers & Security, Vol.28, N. 7, 2009, pp 592-604.
5. M. Matias, L. Edward, W. Jacob, and c. Brian, "Watch What You Write: Preventing Cross-Site Scripting by Observing Program Output", in OWASP AppSec Conference, 2008.
6. J. Etienne, and Z. Pavol, "Preventing SQL Injections in Online Applications: Study, Recommendations and Java Solution Prototype Based on the SQL DOM", in OWASP AppSec Conference, 2008.

7. AspectJ, http://www.eclipse.org/aspectj/
WebGoat,
8. http://www.owasp.org/index.php/Category:OWASP_
WebGoat_Project
9. Apache Jmeter, http://jakarta.apache.org/jmeter/
10. JBoss AOP, http:// www.jboss.org/jbossaop

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Digital Watermarking: Digital Data Hiding techniques for BMP Images

By Tripat Deep Singh Dua
Guru Nanak Institute Of Management And Technology Model Town Ludhiana , Punjab India

Abstract - Purpose: This research evaluates the digital watermarking technology further for hide/retrieved data into the BMP file by manipulating the contents their pixel value using least significant bits (LSB) approach. Methodology: Various experiments have been applied on the pixel value of the BMP file to hide/store the maximum data. With a condition the size and the quality of the BMP file will not change. The trail and error methods have been used or applied to check the various sizes with various qualities. Findings: The study finds that the any digital data can be hiding into the BMP file by manipulating the contents of the Red Green Blue (RGB) value by applying least significant approach. Originality/Value: Due to the growing usage of multimedia content on the internet, serious issues have emerged. Counterfeiting, forgery fraud and pirating of this content are rising. The research is a mechanism which can help resolve the ownership issues for digital data.

Keywords : *Digital watermarking, Digital Data, 24 bit BMP Image, Red Green Blue (RGB) pixel Value, Least Significant Bit (LSB), lower order bits, Copy right, Tracking,*

GJCST Classification: *D.2.11*

Strictly as per the compliance and regulations of:

Digital Watermarking: Digital Data Hiding techniques for BMP Images

Tripat Deep Singh Dua

Abstract - Purpose: This research evaluates the digital watermarking technology further for hide/retrieved data into the BMP file by manipulating the contents their pixel value using least significant bits (LSB) approach.

Methodology: Various experiments have been applied on the pixel value of the BMP file to hide/store the maximum data. With a condition the size and the quality of the BMP file will not change. The trail and error methods have been used or applied to check the various sizes with various qualities.

Findings: The study finds that the any digital data can be hiding into the BMP file by manipulating the contents of the Red Green Blue (RGB) value by applying least significant approach.

Originality/Value: Due to the growing usage of multimedia content on the internet, serious issues have emerged. Counterfeiting, forgery fraud and pirating of this content are rising. The research is a mechanism which can help resolve the ownership issues for digital data.

*Keywords : Digital watermarking, **Digital Data**, 24 bit BMP Image, Red Green Blue (RGB) pixel Value, Least Significant Bit (LSB), lower order bits, Copy right, Tracking,*

I. Introduction

Due to the growing usage of multimedia content on the internet, serious issues have emerged. Counterfeiting, forgery fraud and pirating of this content are rising. Virtually anyone with a sound card, scanner, video frame grabbers or multimedia authoring systems allow them to incorporate copyrighted material into presentations, web designs and internet marketing campaigns. A simple search on any of the search engines returns hundreds and thousands of images which can be easily downloaded on to a personal computer. The desire for the availability of information and quick distribution has been a major factor in the development of new technology in the last decade. There is the increased use of multimedia across the internet. Multimedia distribution has become an important way to deliver services to people around the world. It is commonly applied in internet marketing campaigns and electronic commerce web sites.

Digital Watermarking describes methods and technologies that hide information on bmp images, for example a number, text, image, video in any digital media. The embedding takes place by manipulating the content of the digital data, which means the information is not embedded in the frame around the data. For images this means that the modifications of the pixel values have to be invisible. In other words it is a pattern of bits inserted into a digital image, audio or video file that identifies the file's copyright information (author, rights, etc.). The name comes from the faintly visible watermarks imprinted on stationery that identify the manufacturer of the stationery. The purpose of digital watermarks is to provide copyright protection for intellectual property that's in digital format. Digital watermarks on the images are designed to be completely invisible, moreover, the actual bits representing the watermark must be scattered throughout the file in such a way that they cannot be identified and manipulated.

The left size picture is true 24 BMP 2.25 MB Image file and right size picture is 2.25 KB image Jpg picture that is to be inserted in left size image (True 24 bit bmp file)

BMP file 2.25 MB

Jpg file 2.25 KB

Author : Assistant Professor, Guru Nanal Institute of Mgt & Tech, Gujarkhan Campus, Model Town Ludhiana.

The picture shows half of the right picture has been inserted into the left picture that is original BMP file. When we insert the jpg 2.25 KB file into 2.25 MB file the by manipulating the pixels of the BMP file the size will remains same 2.25 MB and the image quality will also remains same.

Digital Watermarking describes methods and technologies that hide information on bmp images, for example a number, text, image, video in any digital media. The embedding takes place by manipulating the content of the digital data, which means the information is not embedded in the frame around the data. For images this means that the modifications of the pixel values have to be invisible. The research will be able to water mark the BMP images with bmp, jpeg, video, audio or any other format files. Digital Watermarking, which is used to transfer or pass information in a manner that the very existence of the message is unknown. With this study we can hide any type of file with any format into a 24 Bit True BMP with password protection (password can be encrypted using any of the existing encryption technology). For example we can hide or insert a video file or an audio file into a given BMP file without changing the image or its size. 24 Bit BMP format has been chosen because of its large pixel data. More the number of pixels in the image more data we can embed in it. This manipulation neither changes the image nor its size. i.e. the Image quality and its original size is maintained.

The first part of this study depicts the overview of the digital watermarking and defined the problem. Second part discusses the objectives of the study. The third part review the findings of the scholars who studied the watermarking in the past. Fourth chapter discusses the methodology used for the research purpose. The fifth part shows the analysing of the data and experiments. The sixth part shows the findings and colclusion.

II. Objective of the Study

The study aims the following objectives

The objective of our study would be to develop a technique which would embed some kind of information into the digital multimedia content such that the information (which we call the watermark) can later be extracted or detected for a variety of purposes including copy prevention and control and will help us to address some of the challenges faced by the rapid proliferation of digital content.

- Watermark should remain invisible
- Use of any digital content as watermark
- Copy/Copyright protection

III. Review of Literature

There is no dearth of literature on watermarking. A number of scholars investigated the diffrents aspects of the topic. *Zhao and Koch (1995), Cox et al (1997), Hartung, Eisert, and Girod (1998), Hsien Fu (1998), Hsu and Wu (1998, 1999), Zhao et al. (1998), Unzign and Stirmark (1999), Fei et al (2001), Hasslacher (2004), Saryazdi & Hossein (2005) and Agrawal (2007)* evaluated the topic digital watermarking.

An interface has been defined in the watermark agent to the external watermark retrieval library.

This interface employs Java's native interface technology to allow Java objects to call watermark retrieval functions written in C. At present, SysCoP *(Zhao and Koch 1995)* is the only digital watermarking mechanism that has been supported in the watermark agent. However, the watermark agent can easily support any other watermarking system.

Cox et al (1997) describe a method for embedding a binary watermark sequence in the highest magnitude DCT coefficients.

Hsien Fu (1998) conducts a literature survey of digital watermarks used for images. It describes the previous work done on digital watermarks, including the analysis of various watermarking schemes and their results. Potential applications are discussed, and an implementation plan of the project is presented. Hsien Fu uncovers the fact that recent work has shown that digital watermarks can be fairly successful in achieving the desired properties mentioned in section 2. These watermarks, however, are not perfect, and more could be done to improve a watermark's robustness or accuracy in detection. Furthermore, the question of copyright infringement remains a legal issue. Courts need to determine which methods may or may not be used. Until these legal standards are set, the Internet continues to be unsafe for images.

Hartung, Eisert, and Girod (1998) studied the methods for digital watermarking of MPEG-4 facial animation parameter data sets. They used a model-based approach for the estimation of the facial parameters that combines a motion model of an explicit 3D textured wireframe with the optical flow constraint from the video data. This leads to a linear algorithm that is robustly solved in a hierarchical framework with low computational complexity. Experimental results confirm the applicability of the presented watermarking technique.

Hsu and Wu (1998, 1999) use the middle frequency coefficients of DCT/Wavelet transform to embed a binary watermark. These mentioned methods are robust against image processing. Their main drawback is requiring the original image to extract the watermark.

Digital watermark has found a multitude of potential applications other than the originally motivation for copyright protection *(Zhao et al. 1998).* Similarly, the various types of uses of the watermark agents create many spectrums and great business opportunities.

Zhao and Luo (1998) presents a complete

digital watermark agent system to effectively put the digital watermark technology into practice. This system enables an agency to dispatch digital watermark agents to agent servers and agent can perform various tasks on the server. Once all the actions have been taken, a report will be sent to an agency's database and an agent can continue to travel to another agent server.

Unzign and Stirmark (1999), integrate a variety of geometric attacks. Unzign introduces local pixel jittering and is very efficient in attacking spatial domain watermarking schemes. Stirmark introduces both global and local geometric distortions. We give a few more details about these attacks later in this paper. However, most recent watermarking methods survive these attacks due to the use of special synchronization techniques. Robustness to global geometric distortions often relies on the use of either a transform in variant domain (Fourier-Melline) or an additional template or of specially designed periodic watermarks whose auto covariance function (ACF) allows estimation of the geometric distortions.

Fei et al (2001) attempt to find a suitable transform domain to watermark images robust against JPEG compression attack. They show that the choice of the transform domain depends on the type of the embedded information. If the watermark is embedded by repetition coding, then the Hadamard transform gives the best results.

Hasslacher (2004) evaluates the watermarking and its uses. He reveals that the scaling factor is a critical system parameter. If ais too small. The image is not distorted but the robustness of the scheme is low. He also unearth that Modification of low-frequency coefficients distorts the image and Gives the hacker a clue about where the watermark is embedded.

Wang et al. (2004) describe a kind of blind watermarking based on relative modulation of the DCT coefficient value by referring to its estimated one. In their method, the DC values of a 3×3 neighborhood of 8×8 blocks are used to estimate the AC coefficients of central block. In each group of nine 8×8 blocks, five bits of watermark are embedded by modulating the first five DCT AC coefficients, in central block, with the following rule:

Set ACi □ AC'i □□ to embed bit "1"

Set ACi □ AC'i □□ to embed bit "0"

Where, ACi and AC'i are the real and estimated value of the AC coefficients, respectively. The watermark recovery is done by comparing ACi and its estimated value. If ACi □ AC'i , then the extracted bit is "1", otherwise, it is "0".

Saryazdi & Hossein (2005) propose a blind scheme for gray-level data embedding in Hadamard Domain. In the proposed algorithm, the host image is first divided into 4×4 non-overlapping blocks. Their embedding procedure contains two parts. The first part is estimating the first two Hadamard low frequency AC coefficients (i.e. H(0,2) and H(2,0)) in each block, using its neighbor blocks. We use the following equations, to estimate the low frequency AC Hadamard coefficients of a block using the DC values of its 3×3 neighbor blocks. *Saryazdi & Hossein (2005)* concludes that For most watermark application, it is desired to recover the embedded data without using host image. In this paper, such a watermarking scheme for embedding gray-level watermarks is presented. In the proposed method, the two first Hadamard AC coefficients are estimated by their neighbor blocks. Then, a number proportional to the gray-level watermark value is added to each estimated AC coefficient. The recovery procedure consists of comparing the estimated values with actual ones.

Agrawal (2007) propose a robust perceptual digital video watermarking procedure to embed a watermark image in digital video frames using the variable-temporal length 3-D DCT technique. He finds that in many existing video watermarking schemes, the raw video is needed for detection of watermark logo. This is refered to as non-blind method and is not convenient in many cases. In this thesis we propose a new blind watermark detection algorithm. The performance of the blind detection technique was evaluated for several types of video sequences. The watermarking is also done for color video samples in the YUV domain. We used only the luminance (Y-Component) to embed the watermark to make the watermarking scheme more robust since the chrominance (U and V) components is perceptually less sensitive to human visual system compared to the luminance (Y-Component).

Research scholars evaluated the different aspects of the digital watermarking and revealed a number of facts about the technology but not much research has been done on the watermarking on BMP files and on the method to hiding an BMP image in other without changing its view. This research will concentrate on the said topic.

IV. Research Methodology

Digital watermarking techniques can be used successfully with digital content in various forms like still images of bmp format using their least significant bit (LSB). In LSB substitution the lower order bits of selected pixels in the image are used to store watermarks. Techniques like flipping the lower order bits, replacing the lower order bits of each pixel with higher order bits of a different image (for e.g., a company logo), superimposing a watermark image over an area of image to be watermarked and adding some fixed intensity value are used to embed watermarks in spatial domain.

In Least Significant Bits substitution the lower order bits of selected pixels in the image are used to

store watermarks or the LSB's are replaced with the higher order bits of the data that is to be inserted in the image that will effect the slightest change in the colour value of the pixel but non noticeable in colour point of view. The 24 bit BMP data has been chosen because of large pixel value more the file size the more data can de hide into the file. Any logo, signature, company name, bmp image or nay image or any audio/ video file can be hide into BMP file. More over the size of the original image will remain same after embedding the other data into the image. Because the data that is to be hide into the BMP file that is not any extra data or not any embedded data but the original data that will be replaced with the original lower bits of the BMP file. The idea behind this technique is that modifying the LSB will not make much difference to the color of the pixel.

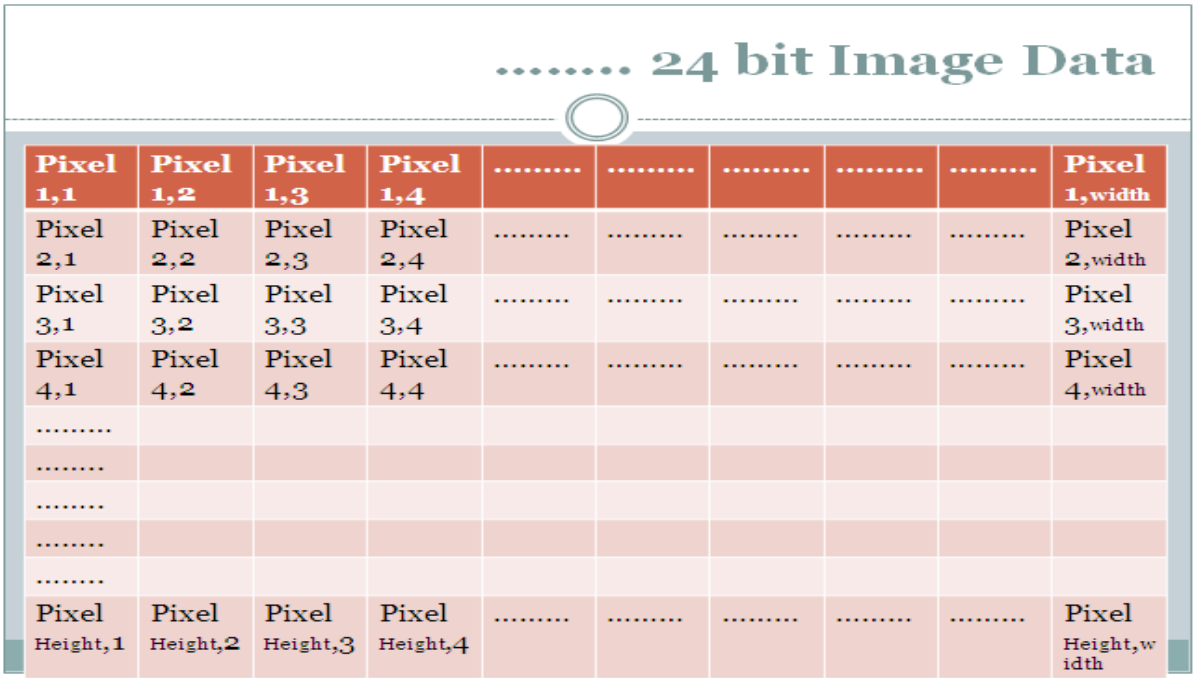
a) Methods used for the study

Experimentations refer to the used of the new techniques to innovate/implement new idea for experiment based. The research refers to the experiments to hide any digital data into the BMP file format so that it can be easily inserted into the 24 BMP file. The hidden data will be into the BMP file with its original form and will not loose its originality. On the other side the size of the BMP file will remains same because it will not increase in any case because the BMP has given the full space to the data that is to hide in the BMP file by replacing its contents or called pixels. The experiment is successful because the BMP file can store large data because of its large file size. And experiment is again successful when the quality/colour of the BMP file will remains same to normal human eye after modifying the contents of the pixel value in the BMP file.

First convert the file to be hidden into a binary stream and then read the BMP file pixel by pixel and substituting the LSB's of R, G, B component of each pixel with the bits from the binary stream until the entire binary stream had been substituted into the image. The binary stream that is substituted also has a format for easy and fast retrieval. We use a 12 Byte or 96 Bit headers, which is prefixed, to the Binary Stream before being substituted.

V. Experiments & Analysis

Figure 1 : shows the actual 24 bit BMP image data



The figure 1 shows how the pixels are stored in the form of BMP file. The data is stored in the form of the matrix of height and width. The RGB pixels are stored in the BMP file which describes the overall description of the image. The pixel value starts from (1,1) to until the size and the width of the picture.

Figure 2 : of BMP Image Data in the form of array

Image Data PixelArray [x,y]					
Pixel[0,h-1]	Pixel[1,h-1]	Pixel[2,h-1]	...	Pixel[w-1,h-1]	Padding
Pixel[0,h-2]	Pixel[1,h-2]	Pixel[2,h-2]	...	Pixel[w-1,h-2]	Padding
			⋮		
Pixel[0,9]	Pixel[1,9]	Pixel[2,9]	...	Pixel[w-1,9]	Padding
Pixel[0,8]	Pixel[1,8]	Pixel[2,8]	...	Pixel[w-1,8]	Padding
Pixel[0,7]	Pixel[1,7]	Pixel[2,7]	...	Pixel[w-1,7]	Padding
Pixel[0,6]	Pixel[1,6]	Pixel[2,6]	...	Pixel[w-1,6]	Padding
Pixel[0,5]	Pixel[1,5]	Pixel[2,5]	...	Pixel[w-1,5]	Padding
Pixel[0,4]	Pixel[1,4]	Pixel[2,4]	...	Pixel[w-1,4]	Padding
Pixel[0,3]	Pixel[1,3]	Pixel[2,3]	...	Pixel[w-1,3]	Padding
Pixel[0,2]	Pixel[1,2]	Pixel[2,2]	...	Pixel[w-1,2]	Padding
Pixel[0,1]	Pixel[1,1]	Pixel[2,1]	...	Pixel[w-1,1]	Padding
Pixel[0,0]	Pixel[1,0]	Pixel[2,0]	...	Pixel[w-1,0]	Padding

Figure 2 shows the 24 bit BMP Image pixels are stored in the form of Pixel Array or Matrix. The height and the width of the pixels are adjusted according to the size of the image. The lowermost left pixel of the image describes the starting point or the starting pixel values of the height and the width of the pixel. For example in the image 2 it is clearly shown that the lower most value of in the pixel array is pixel (0, 0). As the size of the picture grows the values in the array grows according to the size of the image.

Figure 3

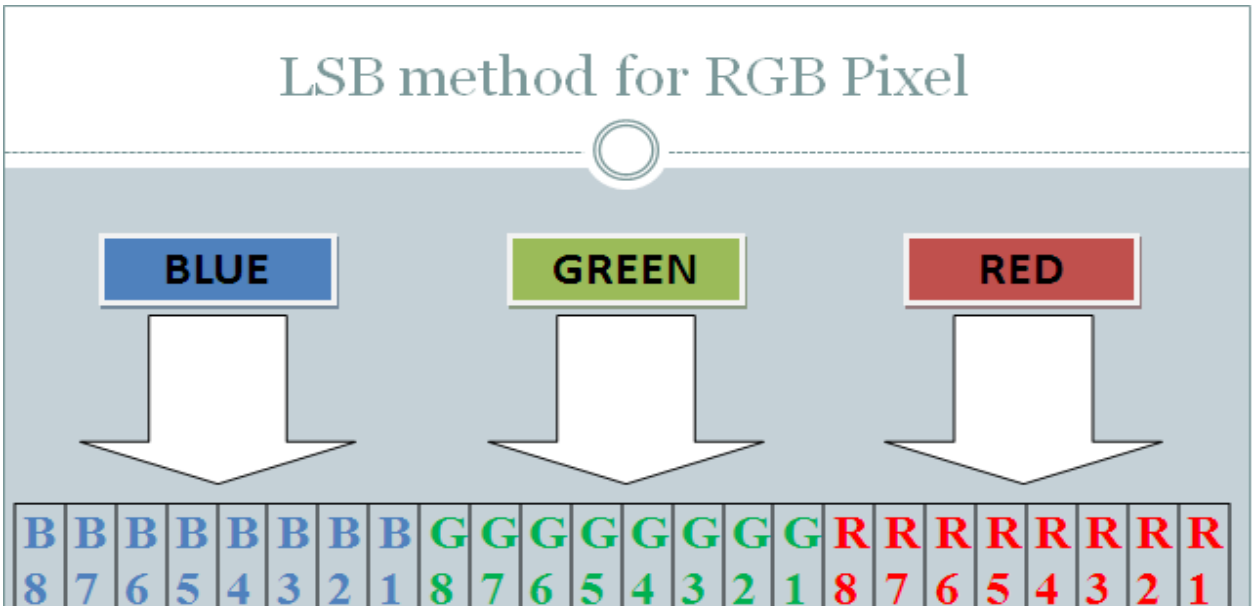


Figure 3 shows **pixel** Method of a 24 bit BMP file, the data is stored in the form 3 pixel RGB. The format of BMP file id 24 bit and each pixel has been assigned a colour value 8 for RGB. This is the colour combination of each pixel where all the colours in the colour palette designed by the combination of RGB.

Figure 4

The figure 4 shows that the left picture is true BMP format picture with a size of 2.25 MB and the right side picture Jpg picture with a size of 225 KB that is to be inserted in BMP file. The insertion takes place with the BMP file.

Figure 5 : shows to hide some text and a ring tone of a song into 2.25 MB file.

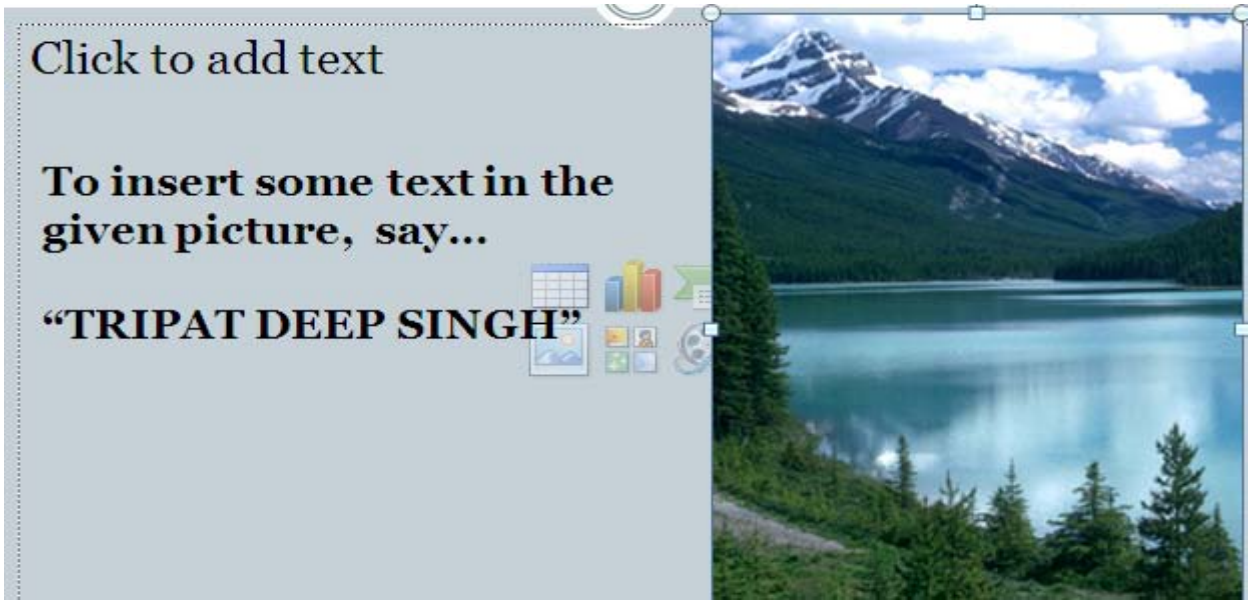
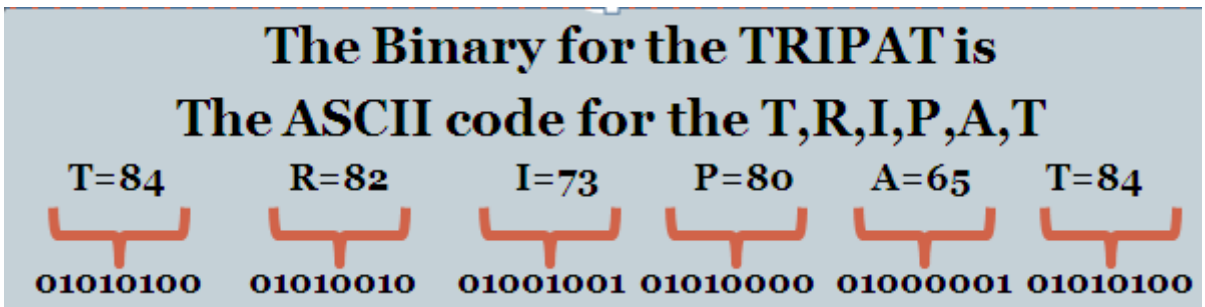


Figure 6



To hide something into BMP file, first covert the Binary for the TRIPAT into ASCII CODE, Than 24 bit RGB pixel value for the sample colour for the starting image is

VI. Findings and Conclusion

Digital watermarks for BMP images on the images are designed to be completely invisible, moreover, the actual bits representing the watermark must be scattered throughout the file in such a way that they cannot be identified and manipulated.

The development will be able to water mark the BMP images with bmp, jpeg, video, audio or any other format files. The research is based on a Technique known as Digital Watermarking, which is used to transfer or pass information in a manner that the very existence of the message is unknown. With this study we can hide any type of file with any format into a 24 Bit True BMP with password protection (password can be encrypted using any of the existing encryption technology). For example we can hide or insert a video file or an audio file into a given BMP file without changing the image or its size. 24 Bit BMP format has been chosen because of its large pixel data. More the number of pixels in the image more data we can embed in it. This manipulation neither changes the image nor its size. i.e. the Image quality and its original size is maintained. Our study is based on a Digital Watermarking Technique known as Least Significant Bit (LSB's)

In this technique the LSB's of the Pixels are modified to store the information. The idea behind this technique is that modifying the LSB will not make much difference to the colour of the pixel. Multimedia distribution has become an important way to deliver services to people around the world. It is commonly applied in internet marketing campaigns and electronic commerce web sites. Due to the growing usage of multimedia content on the internet, serious issues have emerged. Counterfeiting, forgery fraud and pirating of this content are rising. Virtually anyone with a sound card, scanner, video frame grabbers or multimedia authoring systems allow them to incorporate copyrighted material into presentations, web designs and internet marketing campaigns. A simple search on any of the search engines returns hundreds and thousands of images which can be easily downloaded on to a personal computer.

An important point that arises in these applications is the protection of ownership rights. Anybody and everybody can download digital content from the internet and can reuse or redistribute that as his own thus depriving the rightful owner of royalty or recognition for his/her work. Hence the need for developing new copy deterrence and protection mechanisms for digital content is felt.

We need to have a mechanism which can help resolve the ownership issues for digital content. The owner should be able to mark his work in some way which should later help in resolving the ownership in case of dispute. Moreover the mark should not affect the quality or the meaning of the image or should not change it. This process on hard copy of images is known as watermarking and when applied to digital content is known as digital watermarking. Consequently, copyright abuse is rampant among multimedia users who are rarely caught. This copyright abuse is the motivating factor for this study.

References References Referencias

1. Frank Hartung, Peter Eisert, and Bernd Girod (1998) *"Digital Watermarking of MPEG-4 Facial Animation Parameters"* Computers & Graphics, Vol. 22, No. 3
2. Alexander Hasslacher (2004) *"Digital Watermarking"*, EMT-Institut, JKU-Linz
3. Agrawal vinod (2007) "*Perceptual Watermarking Of Digital Video Using The Variable Temporal Length 3d-Dct"* Thesis M-tech, IIT Kanpur.
4. Fu Hsien (1998) *"Literature Survey on Digital Image Watermarking",* Multidimensional Signal Processing
5. Saryazdi Saeid, Nezamabadi-pour. Hossein (2005) *"A Blind Digital Watermark in Hadamard Domain"* World Academy of Science, Engineering and Technology 3.
6. Cox, I. J., Kilian, J., Leighton, F. T., Shamoon, T., "Secure Spread Spectrum Watermarking for Multimedia", IEEE Trans. On Image Processing, No. 6, Vol. 12,pp 1673-1687, 1997.
7. Wang, Y., Pearmain. A., "Blind Image Data Hiding Based on Self Reference", Pattern Recognition Letters, Vol. 25, Issue 15, pp. 1681-1689, November 2004.
8. Fei, C., Kundur, D., Kwong, R. H., "The Choice of Watermark Domain in the Presence of Compression", Proc. Of IEEE Int. Conf. On Information Technology: Coding & Computing, pp 79-84, Las Vegas, Nevada, April 2001.
9. Hsu, C. T., Wu, J. L., " Multi-resolution Watermarking for Digital Images", IEEE Trans. On Circuits & Systems: Analog & Digital Signal Processing, Vol. 45, No. 8, 1998.
10. Hsu, C. T., Wu, J. L., " Hidden Digital Watermarks in Images", IEEE Trans. On Image Processing, Vol.8, No. 1, 1999.
11. M. Kutter and F. Petitcolas, "A fair benchmark for image watermarking systems," *Electronic Imaging 199: Security and Watermarking of Multimedia Content*, Vol. **3657** of SPIE Proceedings, San Jose, California USA, 25-27 January 1999.
12. Zhao Jian & Luo Chenghui *"Digital Watermark Mobile Agents"* Fraunhofer Center for Research in Computer Graphics, Inc.
13. Zhao, J. and Koch, E. (1995). *Embedding Robust Labels Into Images For Copyright Protection*. In: Proc. of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, Austria, August 21-25, 1995.
14. Zhao, J., Koch, E. and Luo, C. (1998). *Digital Watermarking In Business Today and Tomorrow.* In: Communications of ACM, pp. 67-72, Vol. 41, No. 7, July 1998.

Global Journal of Computer Science and Technology Volume XII Issue I Version I January 2012
46

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Hotspot Identification System for identification of core residues in Diabetic Proteins

By P.V.S.L. Jagadamba, M.S.Prasadbabu, Allam Apparao
Professor, Dept of CS&SE, Andhra University, Visakhapatnam

Abstract - Data on genome structural and functional features for various organisms are being accumulated and analyzed in laboratories all over the world. The data are stored and analyzed on a large variety of expert systems. The public access to most of these data offers to scientists around the world an unprecedented chance to data mine and explores in depth this extraordinary information repository, trying to convert data into knowledge. The DNA and RNA molecules are symbolic sequences of amino acids in the corresponding proteins has definite advantages in what concerns storage, search, and retrieval of genomic information. In this study an attempt is made to develop an algorithm for aligning multiple DNA / protein sequences. In this process hotspots are located in a protein sequence using the multiple sequence alignment.

Keywords : *Symbolic sequences, DNA, RNA, Protein sequence, Multiple Sequence alignment.*

GJCST Classification: *Optional, DDC/LCC/UDC/Global Journals/NLMC/FOR/MSC Classifications Accepted*

Strictly as per the compliance and regulations of:

Hotspot Identification System for identification of core residues in Diabetic Proteins

P.V.S.L. Jagadamba[α], M.S.Prasadbabu[Ω], Allam Apparao[β]

Abstract - Data on genome structural and functional features for various organisms are being accumulated and analyzed in laboratories all over the world. The data are stored and analyzed on a large variety of expert systems. The public access to most of these data offers to scientists around the world an unprecedented chance to data mine and explores in depth this extraordinary information repository, trying to convert data into knowledge. The DNA and RNA molecules are symbolic sequences of amino acids in the corresponding proteins has definite advantages in what concerns storage, search, and retrieval of genomic information. In this study an attempt is made to develop an algorithm for aligning multiple DNA / protein sequences. In this process hotspots are located in a protein sequence using the multiple sequence alignment

Keywords : Symbolic sequences, DNA, RNA, Protein sequence, Multiple Sequence alignment.

I. Introduction

In Bioinformatics, sequence alignment is a prominent method of arranging the sequences of DNA, RNA or protein to identify regions of similarity. Similarity may be functional, structural or evolutionary relationships between the sequences. Aligned sequences of nucleotide, amino acid residues are represented in a row form of a matrix. Identical or similar characters are aligned in successive columns by inserting gaps between the residues. There is a storm of revolution in the areas of Genomics and Bioinformatics in recent years. Bioinformatics is widely used for computational usage and processing of molecular and genetic data. The biologists considered Bioinformatics for the use of computational methods and tools to handle large amounts of data and make the data more understandable and useful. On the other hand, others view Bioinformatics as an area of developing algorithms and tools and to use mathematical and computational approaches to address theoretical and experimental questions in biology. As genomic data is rapidly exposed to increasing research, knowledge based expert system is becoming indispensable for the emerging studies in Bioinformatics. Hence validation and analysis of mass experimental and predicted data to identify relevant biological patterns and to extract the hidden knowledge are becoming important.

```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                     ****: .***:  * *:** * :****.:* *******..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:***********:***:**.: .****************    :  *.: :
```

In recent years, semantic web based methods are introduced and are designed in such a way that meaning is added to the raw data by using formal descriptions of concepts, terms and relationships encoded within the data. To analyze and understand the data, today's information rich environment developed and designed a number of software tools. These tools provide powerful computational platforms for performing Insilco experiments (8). As there is much complexity and diversity in the analysis of tools, the need is for an intelligent computer system for automated processing. Present researches in Bioinformatics need the use of integrated expert systems to extract more efficient knowledge. In the biological process proteins undergo some interactions. These protein-protein interactions are mediated molecular mechanisms. During this interaction, a small set of residues play a critical role. These residues are called hot spots. The ability to identify the hot spots from sequence accurately and efficiently as expert system that enables and analysis of protein-protein interaction hot spots. This analysis may benefit function prediction and drug development. At present there is a strong need for methods to obtain an accurate description of protein interfaces. Many scientists try to extract protein interaction information from protein data bank.

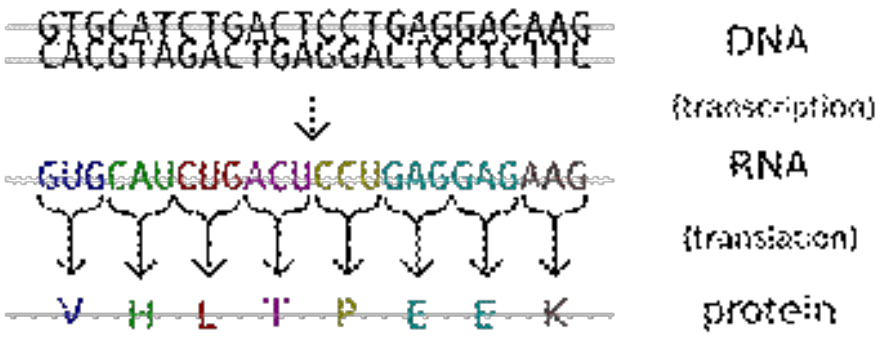
Alignment Methods Used: In general the hot spots are identified as active sites in protein structures as binding is done using structures. The researcher tried

Author α : Principle Investigator, Women Scientist Scheme (WOS-A), DST Project, JNTUK, Kakinada, AP, India.
Author Ω : Professor, Dept of CS&SE, Andhra University, Visakhapatnam
Author β : Vice Chancellor, JNTUK, Kakinada

to find the hotspots in protein sequence rather than structure. In this process, taking into consideration the evolutionary history, the families of sequences are aligned using multiple sequence alignment.

In the process of alignment two methods are used Standard method using dynamic programming and A proposed alternative- MSAPSO (Multiple Sequence alignment using Particle Swarm Optimization) method in which alignment is performed using PSO technique. A comparison of these two methods also made. If the sequences are very short or similar they can be aligned by hand. But lengthy and highly variable numerous sequences cannot be aligned manually. To produce high quality sequence alignments, construction of algorithms and application of human knowledge are necessary. Computational approaches to sequence alignments are of two types- Global alignments and local alignments. Global alignment is the alignment to span the entire length of sequences whereas local alignments identify regions of similarity within the long sequences.

1. Particle Swarm Optimization: Particle Swarm Optimization (PSO) is based on stochastic optimization technique. It is one of the machine learning algorithms. It has been considered to be an effective optimization tool in many areas. The interesting point in PSO is that each particle with potential solution searches through the problem by updating itself with its own memory and also the social information gathers from other particles. Multiple Sequence Alignment: When three or more biological sequences namely protein, DNA or RNA are generally aligned, it is called multiple sequence alignment. As it is difficult and also time consuming to align by hand, computational algorithms are used to analyze and produce such biological sequences. Most multiple sequence alignment programs use heuristic methods as the order of the sequences to align plays a vital role. Development of MSA algorithm is now an active are of research. MSA alignments are an essential tool for protein structure and function prediction, phylogeny inference and other common tasks in sequence analysis.
2. Pair wise Sequence Alignment: If two sequences are arranged for an alignment it is known as pair wise sequence alignment. The degree of relationship between the sequences is predicted computationally or statistically based on weights assigned to the elements aligned between sequences. The standard algorithm to align a pair of sequences is Needleman Wunch algorithm. This algorithm uses dynamic programming. In this study an algorithm PSAPSO (Pair wise Sequence alignment using Particle Swarm Optimization) is proposed and is also compared with the standard algorithm to know the accuracy of the results. A gene encoded in the genetic code defines the amino acid sequence in a protein. An amino acid residue is the combination of three nucleotides. Each three-nucleotide set is a codon. The set of codons forms a genetic code. For example AUG stands for methonine M. In this AUG is a codon, M is an amino acid and the residues A, U, G are nucleotides. Genes encoded in DNA are first transcribed into pre-messenger RNA (mRNA) known as primary transcript. Then pre-mRNA process to mature mRNA using various forms of modifications of posttranscriptional modifications. Then mature mRNA is used as a template for protein synthesis, which is known as translation onto a ribosome. Then read three nucleotides at a time by matching each codon to its base pairing anticodon to form transfer RNA (tRNA). Then tRNA recognizes the amino acid corresponding to the codon. The sequence thus obtained is protein sequence.



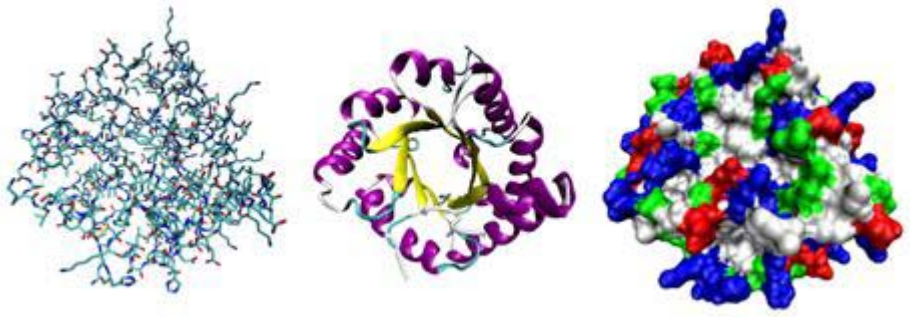
The amino acids in a protein sequence are shown in the following table.

Table 1

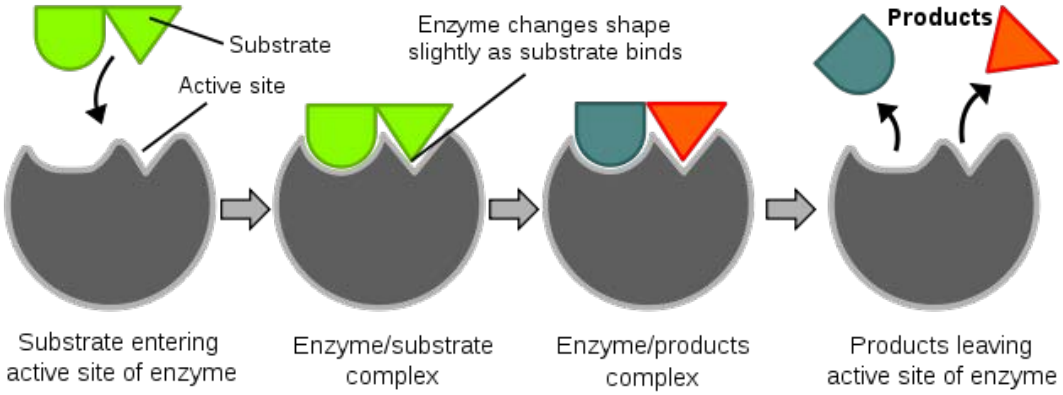
One Letter	Three Letter	Full Name	One Letter	Three Letter	Full Name
G	GLY	Glycine	W	TRP	Tryptopham
A	ALA	Alanine	Y	TYR	Threonine
V	VAL	Valine	N	ASN	Asparagine
L	LEU	Leucine	Q	GLN	Glutamine
I	ILE	Lsoleucnie	D	ASP	Asparatic Acid
F	PHE	Phenylalanine	E	GLU	Glutamic Acid
P	PRO	Proline	K	LYS	Lysine
S	SER	Serine	R	ARG	Arginine
T	THR	Threonine	H	HIS	Histidine
C	CYS	Cyctenie	M	MET	Methinine

The overall structure and function of a protein is determined by the amino sequence. Most proteins fold into 3-dimensional structures and its shape is known as its native state. There are four levels in a protein structure.

- Primary Structure: Primary structure is nothing but an amino acid sequence. Secondary Structure: Secondary structures are regularly repeating local structures and are stabilized by hydrogen bonds. As they are local in nature different secondary structures can be present in the same protein molecule. Example alpha helix, beta sheet and turns.
- Tertiary Structure: Tertiary structure is the special relationship of the secondary structures to one another and is generally stabilized by the formation of the hydrophobic core, a non-local interaction. Salt bridges, hydrogen bonds; disulphide bonds and even post-transnational modifications also stabilize it. It mainly controls the basic function of the protein.
- Quaternary Structure: This structure is formed by several protein molecules i.e. poly peptide chains and it functions as a single protein complex.



- Enzymes: Enzyme is one of the functions of the protein which carries out most of the reactions involved in metabolic activities. Enzymes are proteins that increase the rate of chemical reaction. Adding or participation of the substance called catalyst does the change in the rate of chemical reaction. Catalysts that speed the reaction are called positive catalysts. Substances that interact with catalysts to slow the reaction are called inhibitors (or negative catalysts). Substances that increase the activity of catalysts are called promoters, and substances that deactivate catalysts are called catalytic poisons.
- Active Sites in Proteins: An Active site is a part of an enzyme where substrates bind and undergo a chemical reaction. The substrate which is a molecule binds with the enzyme active site and then an enzyme-substrate complex is formed. It is then transformed into one or more products, which are released from the active site. The active site is now free to accept another substrate molecule. In the case of more than one substrate, these may bind in a particular order to the active site, before reacting together to produce products. A product is something "manufactured" by an enzyme from substrate. For example the products of its Lactase are Galactose and Glucose, which are produced from the substrate Lactose.



Two models- the lock and key model and induced fit model are the two models proposed to describe how the enzymes work. In the lock and key model the active site perfectly fits for a specific substrate. If once the substrate binds to the enzyme no further modification is necessary. On the other hand in the induced fit model, an active site is more flexible and the presence of certain residues (amino acids) of the active site the enzyme is encouraged to locate the correct substrate. Once the substrate is gone conformational changes may occur. Hot spots are a set of residues recognized or bound in the process of interacting with other proteins. These are the residues in the active site.

II. Results & Discussion

Insulin is one of the important protein sequences which cause diabetes. So we tried to identify the hotspots in this protein sequence using the following methodology.

- The protein structures are retrieved from protein data bank by mapping with insulin protein sequence accession p01038 shown in the following table.

SNO	PDB Code	Chain	First PDB residue	Last PDB residue	First P01308 (INS_Human) residue	Last P01308 (INS_Human) residue
1	1a7f	A	1	21	90	110
2	1a7f	B	1	29	25	53
3	1ai0	A	1	21	90	110
4	1ai0	B	1	30	25	53
5	1ai0	C	1	21	90	110
6	1ai0	D	1	30	25	54
7	1ai0	E	1	21	90	110
8	1ai0	F	1	30	25	54
9	1ai0	G	1	21	90	110
10	1ai0	H	1	30	25	54
11	1ai0	I	1	21	90	110
12	1ai0	J	1	30	25	54

12	1ai0	J	1	30	25	54
13	1ai0	K	1	21	90	110
14	1ai0	L	1	30	25	54
15	1aiy	A	1	21	90	110
16	1aiy	B	1	30	25	53
17	1aiy	C	1	21	90	110
18	1aiy	D	1	30	25	54
19	1aiy	E	1	21	90	110
20	1aiy	F	1	30	25	54
21	1aiy	G	1	21	90	110
22	1aiy	H	1	30	25	54
23	1aiy	I	1	21	90	110
24	1aiy	J	1	30	25	54
25	1aiy	K	1	21	90	110

- Then identify the protein-protein interactions for each of these protein structures shown in the following table.

SNO	PDB Code	Chain	Chain
1	1a7f	A	B
2	1ai0	A	B
3	1ai0	B	D
4	1ai0	C	D
5	1ai0	E	F
6	1ai0	F	H
7	1ai0	G	H
8	1ai0	I	J
9	1ai0	J	L
10	1ai0	K	L
11	1aiy	A	B
12	1aiy	B	D
13	1aiy	C	D
14	1aiy	E	F
15	1aiy	F	H
16	1aiy	G	H
17	1aiy	I	J
18	1aiy	J	L
19	1aiy	K	L
20	1b9e	A	B
21	1b9e	B	D
22	1b9e	C	D
23	1guj	A	B
24	1guj	B	D
25	1guj	C	D

Identification of Hotspot: The hot spots are identified using these interfaces and the hot spots in the protein sequence p01308 are

MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTR

REAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

III. Conclusion

Hot spots are of residues comprising only a small fraction of interfaces of the binding energy. We present a new and efficient method to determine computational hot spots based on pair wiser technique using potentials and solvent accessibility of interface residues. The conservation does not have significant effect in hot spot prediction as a single feature. Residue occlusions from solvent and pair wise potentials are found to be the main discriminative features in hot spot prediction. The predicted hotspots are observed to match with the experimental hot spots with an accuracy of 70%. The solvent is a necessary factor to define a hot spot, but not sufficient itself. This is also compared our methods and other hot spot prediction methods. Our method outperforms them with its high performance expert system.

References References Referencias

1. Chao-Yie Yng and Shaomeng Wang, "Computational Analysis of Protein Hotspots",ACS Medicinal Chemistry Letters, 2010,1 (3) pp 125-129.
2. Hajduk, P.J et al (2005) "Druggability induces for protein targets derived from NMR-based Screening Data", J Med. Chem. 48, 2518-2525.
3. Dobson CM. (2000). The nature and significance of protein folding. In Mechanisms of Protein Folding 2nd ed. Ed. RH Pain. Frontiers in Molecular Biology series. OxfordUniversity Press: New York, NY.
4. Hintze Miller B.(1988), "Expert System An Introduction" PC AI where Intelligent technology meets the real world, 2(3), 26.
5. Robert S. Engelmopre, Edward Feigenbaum, 1993, "Expert Sstems and Artificial Intelligence", WTEC Hyper Librarian.
6. Mohamed Radhouene Aniba and Julie D. Thompson, "Knowledge Based Expert Systems in Bioinformatics", published in Expert Systems, Book edited by: Petrică Vizureanu, ISBN 978-953-307-032-2, pp. 181-192, 2010.
7. Roos DS. Computational biology. Bioinformatics – trying to swim in a seaof data. Science. 2001;291:1 260—1.

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Survey on Software Protection Techniques against Various Attacks

By N.Sasirekha, Dr.M.Hemalatha
University, Coimbatore, Tamilnadu, India

Abstract - Software security and protection plays an important role in software engineering. Considerable attempts have been made to enhance the security of the computer systems because of various available software piracy and virus attacks. Preventing attacks of software will have a huge influence on economic development. Thus, it is very vital to develop approaches that protect software from threats. There are various threats such as piracy, reverse engineering, tampering etc., exploits critical and poorly protected software. Thus, thorough threat analysis and new software protection schemes, needed to protect software from analysis and tampering attacks becomes very necessary. Various techniques are available in the literature for software protection from various attacks. This paper analyses the various techniques available in the literature for software protection. The functionalities and the characteristic features are various software protection techniques have been analyzed in this paper. The main goal of this paper is to analyze the existing software protection techniques and develop an efficient approach which would overcome the drawbacks of the existing techniques.

Keywords : *Software Security, Software Tampering, Tampering Attacks, Encryption, Cryptography, Decryption.*

GJCST Classification: *K.6.5*

Strictly as per the compliance and regulations of:

A Survey on Software Protection Techniques against Various Attacks

N.Sasirekha [α], Dr.M.Hemalatha [Ω]

Abstract - Software security and protection plays an important role in software engineering. Considerable attempts have been made to enhance the security of the computer systems because of various available software piracy and virus attacks. Preventing attacks of software will have a huge influence on economic development. Thus, it is very vital to develop approaches that protect software from threats. There are various threats such as piracy, reverse engineering, tampering etc., exploits critical and poorly protected software. Thus, thorough threat analysis and new software protection schemes, needed to protect software from analysis and tampering attacks becomes very necessary. Various techniques are available in the literature for software protection from various attacks. This paper analyses the various techniques available in the literature for software protection. The functionalities and the characteristic features are various software protection techniques have been analyzed in this paper. The main goal of this paper is to analyze the existing software protection techniques and develop an efficient approach which would overcome the drawbacks of the existing techniques.

Keywords : Software Security, Software Tampering, Tampering Attacks, Encryption, Cryptography, Decryption.

I. Introduction

Software protection has become one of the attractive domains with high commercial interest, from major software vendors to content providers which also comprises of the movie and music recording industries. The digital data of the software is especially at tremendous risk.

Confidentiality and data authenticity are two important concepts in security. Confidentiality provides data secrecy of a message and data authenticity protects the integrity of the message. Software protection falls between the domains of security, cryptography [30] and engineering among other disciplines.

The software protection technique mainly concentrates on protecting software from various attacks such as reverse engineering by obfuscation, modification by software tamper resistance, program-based attacks by software diversity, and BORE – break-once run everywhere – attacks by architectural design [2].

Protecting content needs protecting the software which processes the content. Copy protection is another form of software protection to the level that it needs several same protections against reverse engineering and software tampering.

Protecting code from attacks such as reverse engineering [32], analysis and tampering attacks is one of the main concerns for software providers. If a competitor succeeds in obtaining and reusing a algorithm, it would result in major issue. Moreover, secret keys, confidential data or security related code are not planned to be examined, extracted, stolen or corrupted. Even if legal actions such as patenting and cyber crime laws are in place, these techniques remain a significant threat to software developers and security expert.

This paper provides a survey on software protection and related areas which would encourage further research. This paper also provides a number of viewpoints, discuss challenges and suggest future directions.

II. Literature Survey

Piracy, reverse engineering and tampering have been the major software threats. Collberg et al. [1] provided a compact outline of the approaches to protect against these threats. Software watermarking for instance focuses on protecting software reactively against piracy. It usually implants hidden, distinctive data into an application in such a way that it can be guaranteed that a particular software instance belongs to a particular individual or company. When this data is distinctive for each example, one can mark out copied software to the source unless the watermark is smashed. The second group, code obfuscation, protects the software from reverse engineering attacks. This approach comprises of one or more program alterations that alter a program in such a way that its functionality remains identical but analyzing the internals of the program becomes very tough. A third group of approaches focuses to make software "tamper-proof", also called tamper-resistant.

Protecting the reliability of software platforms, particularly in unmanaged customer computing systems is a tough task. Attackers may try to carry out buffer

Author α : Doctoral Research Scholar, Karpagam University, Coimbatore, Tamilnadu, India,
E-mail : sasirekha.research@gmail.com, Telephone: number here
E-mail : here@here.com
Author Ω : Head, Department of Software Systems, Karpagam University, Coimbatore, Tamilnadu, India.
E-mail : hema.bioinf@gmail.com

overflow attacks to look for the right of entry to systems, steal secrets and patch on the available binaries to hide detection. Every binary has intrinsic weakness that attackers may make use of. In this paper Srinivasan et al., [3] proposed three orthogonal techniques; each of which offers a level of guarantee against malware attacks beyond virus detectors. The techniques can be incorporated on top of normal defenses and can be integrated for tailoring the level of desired protection. The author tries to identify alternating solutions to the issue of malware resistance. The techniques used are adding diversity or randomization to data address spaces, hiding significant data to avoid data theft and the utilization of distant evidence to detect tampering with executable code.

This paper focuses on the protection of a software program and the content that the program protects. There have been billions of dollars spent each year by the industries especially for software piracy and digital media piracy. The achievement of the content/software security in a huge segment is based on the ability of protecting software code against tampering and identifying the attackers who issue the pirate copies. In this paper, Hongxia Jin et al., [4] concentrates on the attacker identification and forensic examination. The author discussed about a proactive detection approach for defeating an on-going attack before the cooperation has occurred. The author also describes another detection approach for post-compromise attacker identification. Especially, the author takes into account the real world scenarios where the application programs connect with their vendors every so often, and where a discovery of attacking can bar a hacker user from further business.

Code obfuscation focuses to protect code against both static and dynamic study and there exists another approach to protect against code analysis, namely self-modifying code. This approach provides the opportunity to create code at runtime, rather than changing it statically. Practically, self-modifying code is highly restricted to the monarchy of viruses and malware. Yet, some publications regard self-modifying code as an approach to protect against static and dynamic analysis. Madou et al., [5] for instance regard dynamic code generation. The author proposed an approach where functions are generated earlier to their first call at runtime. Moreover, clustering is presented in such a way that a general template can be utilized to generate each function in a cluster, carrying out a least amount of alterations. In order to protect the constant `edits' against dynamic analysis, the authors suggested the usage of a Pseudo Random Number Generator (PRNG). The decryption at runtime technique is equal with code generation, apart from the fact that the decryption key can depend on other code, rather than on a PRNG. Moreover, it lessens re-encryption the viability of code during execution, while Madou et al. do not clearly protect a function template after the function executed.

Protecting code against tampering can be regarded as the issue of data authenticity, where 'data' refers to the program code. Aucsmith [6] explained an approach to implement tamper resistant software. The approach protects against analysis and tampering. The author utilizes small, armored code segments, also called Integrity Verification Kernels (IVKs), to validate code integrity. These IVKs are protected via encryption and digital signatures in such a way that it is tough to modify them. Morover, these IVKs can communicate with each other and across applications via an integrity verification protocol.

Chang et al. [7] proposed an approach depending on software guards. The protection technique of the author is chiefly based on a composite network of software guards which mutually validate each other's consistency and that of the program's critical sections. A software guard is a small segment of code carrying out particulars tasks, e.g. check summing or repairing. When check summing code discovers a modification, repair code is capable to undo this malevolent tamper challenge. The security of the approach depends partly on hiding the obfuscated guard code and the complexity of the guard network.

Horne et al. [8] described on the same idea of Chang et al. [7] and proposed `testers', small hashing functions that validate the program at runtime. These testers can be integrated with embedded software watermarks to result in a unique, watermarked, self-checking program. Other related research is unconscious hashing [9] which interweaves hashing instructions with program instructions and which is capable of proving whether a program is operated correctly. Recently, Ge et al. [10] presented a research work on control flow based obfuscation. Although the authors contributed to obfuscation, the control flow data is protected with an Aucsmith-like tamper resistance approach.

Buffer overflow utilization is a one of the notable threat to software security. In order to lessen the threat, Visual studio C/C++compiler facilitates to randomize the addresses of the compiled program in initialization time and to implant security stack guards by the compiled program in run time. Yongdong Wu [11] upgrades the compiler by raising the compiled program's abilities in the following features:

i. Protects a frame pointer from tampering without additional cost;
ii. Defeats the attack which tampers 1-2 bytes of a protected region at a very low cost;
iii. Checks the indirect function call against the prologue pattern so as to lessen the probability of software crash in case of being attacked.

The experiments demonstrated the enhancement on Microsoft Visual Studio in generating secure and robust software.

Cappaert et al., [12] presented a partial encryption approach depending on a code encryption approach [12], [13]. In order to utilize the partial encryption approach, binary codes are partitioned into small segments and encrypted. The encrypted binary codes are decrypted at runtime by users. Thus, the partial encryption overcomes the faults of illuminating all of the binary code at once as only the essential segments of the code are decrypted at runtime.

Jung et al., [14] presented a code block encryption approach to protect software using a key chain. Jung's approach uses a unit block, that is, a fixed-size block, rather than a basic block, which is a variable-size block. Basic blocks refer to the segments of codes that are partitioned by control transformation operations, such as "jump" and "branch" commands, in assembly code [12], [13]. Jung's approach is very similar to Cappaert's scheme. Jung's approach tries to solve the issue of Cappaert's approach. If a block is invoked by more than two preceding blocks, the invoked block is duplicated.

Unauthorized reverse-engineering of algorithms is a major issue for the software industry. Reverse-engineers look for security holes in the program to make use of competitors' vital approaches. In order to discourage reverse-engineering, developers use a wide range of static software protections to obfuscate their programs. Metamorphic software protections include another layer of protection to conventional static obfuscation approaches, forcing reverse-engineers to alter their attacks as the protection changes. Program fragmentation incorporates two obfuscation approaches, over viewing and obfuscated jump tables, into a novel, metamorphic protection. Segments of code are eliminated from the chief program flow and placed throughout memory, minimizing the locality of the program. These fragments move and are called using obfuscated jump tables which makes program execution hard. This research by Birrer et al., [15] evaluates the performance overhead of a program fragmentation engine and offers examination of its efficiency against reverse-engineering approaches. The experimental results show that program fragmentation has low overhead and is an effective approach to obscure disassembly of programs through two common disassembler/debugger tools.

Song-kyoo Kim [16] deals with the stochastic maintenance approach for the software protection through the closed queueing system with the untrustworthy backups. The technique shows the theoretical software protection approach in the security viewpoint. If software application modules are denoted as backups under proposed structural design, the system can be overcome through the stochastic maintenance model with chief untrustworthy and random auxiliary spare resources with replacement strategies. Additionally, the practical approach of technology improvement in software engineering through the technology innovation tool called TRIZ.

Zeng Min et al., [17] considered the supple manufacturing venture networks data security and software protection and proposed an enterprise classified data security and software protection solution, to describe the enterprise data storage, transmission and application software installation authorization, license and so on, presented a time and machine code depending on MD5, AES encryption algorithm dynamic secret key the encryption approach, to protect the enterprise data confidentiality, integrity and availability, to attain the software installation restrictions and using restrictions.

Kent [18] proposed a software protection technique which deals with the security needs of software vendors like protection from software copying and modification (e.g. physical attacks by users, or program-based attacks). Techniques proposed to handle these requirements include physical Tamper-Resistant Modules (TRMs) and cryptographic techniques. One approach comprises of using encrypted programs, with instructions decrypted immediately preceding to execution. Kent also observed the dual of this issue like user needs that externally-supplied software be confined in its access to local resources.

Gosler's software protection survey [19] investigates circa-1985 protection technologies which comprise of hardware security tools (e.g. dongles), floppy disc signatures (magnetic and physical), analysis denial approaches (e.g. anti-debug approaches, checksums, encrypted code) and slowing down interactive dynamic analysis. The main goal is on software copy prevention, but Gosler observed that the potency of resisting copying should be balanced by the potency of resisting software analysis (e.g. reverse engineering to study where to alter software and for protecting proprietary approaches) and that of software modification (to bypass security checks). Useful tampering is generally headed by reverse engineering.

Gosler also described that one should anticipate that an opponent can execute dynamic analysis on the target software without discovery (e.g. using in-circuit emulators and simulators) and that in such scenario, due to repeated experiments, one should anticipate the opponent to win. Thus, the main goal of practical resistance is to construct such experiments "enormously arduous". Another proposal [19] is cycling software (e.g. through some forced obsolescence) at a rate faster than an opponent can break it; this expects the model of forced software renewal (Jakobsson and Reiter [20]), who suggested hopeless pirates via forced updates and software aging). This technique is suitable

where protection from attacks for a restricted time period suffices.

Herzberg and Pinter [21] focused on the issue of software copy protection and presented a solution needing CPU encryption support (which was far less possible when presented almost 20 years ago, circa 1984-85). Cohen's research [22] on software diversity and obfuscation is directly concentrated to software protection and offers a wide range of algorithms.

The subsequent practical tamper resistance system of Aucsmith [23] handled similar problems by an integration of just-in-time instruction decryption, and re-arranging instruction blocks at run-time to vigorously change the deals with the executing statements during program execution.

Several researchers have proposed techniques on software obfuscation using automated tools and code transformations [24, 25]. One idea would be to employ language-based tools to transform a program (most easily from source code) to a functionally equivalent program which presents greater reverse engineering barriers. If implemented in the form of a pre-compiler, the usual portability issues can be handled by the back-end of standard compilers.

Collberg et al. [26] provides more information regarding software obfuscation which includes descriptions about:

- Categorizing code transformations (e.g. control flow obfuscation, data obfuscation, layout obfuscation, preventive transformations)
- Identification of control flow changes using opaque predicates (expressions not easy for an attacker to predict, but whose worth is recognized at compilation or obfuscation time)
- Preliminary suggestions on metrics for code transformations
- Program slicing tools
- The usage of (de)aggregation offlow control or data

Essential suggestions in software protection are done by Aucsmith [6], in combination with Graunke [23] at Intel. Aucsmith provides tamper prevention software which prevents inspection and change, and it is highly dependent to work accurately in unfriendly situations. Architecture is suggested according to an Integrity Verification Kernel (IVK) that checks the reliability of vital code segments. The IVK architecture is self-decrypting and includes self adjustment code.

Software tampering prevention using self-checking code was described by Horne et al. [27]. The integrity of segments of code is tested using some code known as testers. This can be a linear hash function and a predictable hash value. If the integrity condition is not satisfied, suitable actions will be carried out so as to make the integrity condition satisfied. The attackers can be confused and it is difficult for them to hack the testers if more number of testers is used.

Chang and Atallah [28] presented a technique with fairly extensive capacity containing a set of guards that can be programmed to perform arbitrary processes. An illustration for this is the check sum code segments for integrity checking which provides resistance against software tamper. An additional described guard function is repairing code (e.g. if a spoiled code segment is identified, downloading and installing a new version of the code section). The author also presents a technique for automatically keeping protections within code.

Chen et al. [29] put forth oblivious hashing that engages compile-time code alterations which outcomes in the calculation of a running trace of the execution history of a complete code. In this approach a trace are considered as increasing hash values of a subset of expressions that happens inside the usual program execution.

Gutmann [30] put forth an apparent conversation of the security concerns facing cryptographic usage in software under general-purpose operating systems, and analyzes the design difficulties in nullifying these concerns faced by using secure cryptographic co-processors.

Approaches	Functionalities
[1]	Outline of the approaches to protect against these threats. Software watermarking for instance focuses on protecting software reactively against piracy
[2]	Proposed three orthogonal techniques; each of which offers a level of guarantee against malware attacks beyond virus detectors.
[4]	Concentrates on the attacker identification and forensic examination. The author discussed about a proactive detection approach for defeating an on-going attack before the cooperation has occurred
[5]	an approach in which functions are generated earlier to their first call at runtime
[6]	The author utilizes small, armored code segments, also called Integrity
	Verification Kernels (IVKs), to validate code integrity
[7]	The protection technique of the author is chiefly based on a composite network of software guards which mutually validate each other's consistency and that of the program's critical sections.
[12]	Presented a partial encryption approach depending on a code encryption approach

[14]	Presented a code block encryption approach to protect software using a key chain
[16]	Deals with the stochastic maintenance approach for the software protection through the closed queueing system with the untrustworthy backups
[12]	Focused on the issue of software copy protection and presented a solution needing CPU encryption support
[27]	Software tampering prevention using self-checking code

III. Problems and Directions

The theoretical results to date on software obfuscation provide software protection of considerable practical value. The impracticality of constructing a program to find out whether other software is malicious does not preclude highly valuable computer virus detection technologies, and a feasible, anti-virus industry. It is still early in the history of research in the domains of software protection and obfuscation and that several discoveries and innovations lie ahead particularly in the domains of software diversity (which are utilized are less in the present scenario), and software tamper resistance. Increased number of secure techniques for software protection is very much needed which involves public scrutiny and peer evaluation. Cappaert proposed a tamper-resistant code encryption scheme, and Jung proposed a key-chain-based code encryption scheme. However, Cappaert's scheme did not meet the security requirements for code encryption schemes, and Jung's scheme had an efficiency problem. Moreover, time cost and space cost should also be taken into consideration. To improve efficiency, support from the compiler and operating system is needed [19].

More open discussion of particular approaches is very much needed. Cryptography is observed to be the technique that can be incorporated in the software protection technique for improved protection. Past trends of proprietary, undislosed techniques of software obfuscation approaches similar to the early days in cryptography have to be altered.

For decades encryption has provided the means to hide information. In this research, the self-encrypting code is used as a means of software protection. In this research work, the concept of efficient code encryption techniques, which offers confidentiality and a method to create code dependencies that implicitly protect integrity need to be established. Moreover, several dependency schemes based on a static call graph which allow runtime code decryption simultaneous with code verification can also be used. If code is modified statically or dynamically, it will result in incorrect decryption of other code, producing a corrupted executable. Better and efficient cryptographic techniques can be integrated for better results. This research uses the encryption technique to secure software static analysis and tampering attacks.

IV. Conclusion

This paper presented and discussed a survey on the protection of software because of various attacks. Several software protection techniques available in the literature are analyzed and examined. The characteristic features of the existing algorithms are thoroughly investigated in this paper. This study would facilitate in development of efficient software protection techniques. Encryption techniques can be incorporated with the existing software protection techniques to improve the overall security of the software. Code encryption schemes for protecting software against various attacks like reverse engineering and modification. Therefore, novel and efficient code encryption scheme have to be established based on an indexed table to guarantee secure key management and efficiency.

References Références Referencias

1. Collberg, C.S.; Thomborson, C.; "Watermarking, tamper-proofing, and obfuscation - tools for software protection", IEEE Transactions on Software Engineering, Volume: 28 , Issue: 8, Page(s): 735 – 746, 2002.
2. T. Ogiso, U. Sakabe, M. Soshi, A. Miyaji, "Software Tamper Resistance Based on the Difficulty of Interprocedural Analysis", 3rd Workshop on Information Security Applications (WISA 2002), Korea, August 2002.
3. Srinivasan, R.; Dasgupta, P.; Iyer, V.; Kanitkar, A.; Sanjeev, S.; Lodhia, J.; "A Multi-factor Approach to Securing Software on Client Computing Platforms", 2010 IEEE Second International Conference on Social Computing (SocialCom), Page(s): 993 – 998, 2010.
4. Hongxia Jin; Lotspiech, J.; "Forensic analysis for tamper resistant software", 14th International Symposium on Software Reliability Engineering, 2003. ISSRE 2003.
5. M. Madou, B. Anckaert, P. Moseley, S. Debray, B. De Sutter, and K. De Bosschere. Software protection through dynamic code mutation
6. D. Aucsmith. Tamper resistant software: an implementation. Information Hiding, Lecture Notes in Computer Science, 1174:317-333, 1996.
7. H. Chang and M. J. Atallah. Protecting software codes by guards. ACM Workshop on Digital Rights Managment (DRM 2001), LNCS 2320:160-175, 2001.
8. B. Horne, L. R. Matheson, C. Sheehan, and R. E. Tarjan. Dynamic Self-Checking Techniques for Improved Tamper Resistance. In Proceedings of Workshop on Security and Privacy in Digital Rights

Management 2001, pages 141-159, 2001.
9. Y. Chen, R. Venkatesan, M. Cary, R. Pang, S. Sinha, and M. Jakubowski. Oblivious hashing: a stealthy software integrity veri_cation primitive. In Information Hiding, 2002.
10. J. Ge, S. Chaudhuri, and A. Tyagi. Control flow based obfuscation. In DRM '05: Proceedings of the 5th ACM workshop on Digital rights management, pages 83{92, 2005.
11. Yongdong Wu; "Enhancing Security Check in Visual Studio C/C++ Compiler", Software Engineering, 2009. WRI World Congress on WCSE '09. Volume: 4 , Page(s): 109 – 113, 2009.
12. J. Cappaert et al., "Toward Tamper Resistant Code Encryption: Practice and Experience," LNCS, vol. 4991, 2008, pp. 86-100.
13. J. Cappaert et al., "Self-Encrypting Code to Protect Against Analysis and Tampering," 1st Benelux Workshop Inf. Syst. Security, 2006.
14. D.W Jung, H.S Kim, and J.G. Park, "A Code Block Cipher Method to Protect Application Programs From Reverse Engineering,"J. Korea Inst. Inf. Security Cryptology, vol. 18, no. 2, 2008, pp. 85-96 (in Korean)
15. Birrer, B.D.; Raines, R.A.; Baldwin, R.O.; Mullins, B.E.; Bennington, R.W. Program Fragmentation as a Metamorphic Software Protection, Third International Symposium on Information Assurance and Security, 2007 , Page(s): 369 – 374, 2007. IAS 2007.
16. Song-kyoo Kim; "Design of enhanced software protection architecture by using theory of inventive problem solving", IEEE International Conference on Industrial Engineering and Engineering Management, 2009. IEEM 2009.
17. Zeng Min; Liu Qiong-mei; Wang Cheng; Practices of agile manufacturing enterprise data security and software protection, 2010 2nd International Conference on Industrial Mechatronics and Automation (ICIMA).
18. S. Kent, Protecting Externally Supplied Software in Small Computers, Ph.D. thesis, M.I.T., September 1980.
19. J. Gosler, "Software Protection: Myth or Reality?", Advances in Cryptology – CRYPTO'85, Springer-Verlag LNCS 218, pp.140–157 (1985)
20. M. Jakobsson, M.K. Reiter, "Discouraging Software Piracy Using Software Aging", Proc. 1st ACM Workshop on Digital Rights Management (DRM 2001), Springer LNCS 2320, pp.1–12 (2002).
21. A. Herzberg, S.S. Pinter, "Public Protection of Software", pp.371–393, ACM Trans. Computer Systems, vol.5 no.4 (Nov. 1987). Earlier version in Crypto'85.
22. F. Cohen, "Operating System Protection Through Program Evolution", Computers and Security 12(6), 1 Oct. 1993, pp. 565–584.
23. D. Aucsmith, G. Graunke, Tamper Resistant Methods and Apparatus, U.S. Patent 5,892,899 (filed June 13 1996; issued Apr.6 1999).
24. C. Collberg, C. Thomborson, D. Low, "Manufacturing Cheap, Resilient, and Stealthy Opaque Constructs", Proc. Symp. Principles of Programming Languages (POPL'98), Jan. 1998.
25. C. Collberg, C. Thomborson, D. Low, "Breaking Abstractions and Unstructuring Data Structures", IEEE International Conf. Computer Languages (ICCL'98), May 1998.
26. C. Collberg, C. Thomborson, D. Low, "A Taxonomy of Obfuscating Transformations", Technical Report 148, Dept. Computer Science, University of Auckland (July 1997).
27. B. Horne, L. Matheson, C. Sheehan, R. Tarjan, "Dynamic Self-Checking Techniques for Improved Tamper Resistance", Proc. 1st ACM Workshop on Digital Rights Management (DRM 2001), Springer LNCS 2320, pp.141–159 (2002).
28. H. Chang, M. Atallah, "Protecting Software Code by Guards", Proc. 1st ACM Workshop on Digital Rights Management (DRM 2001), Springer LNCS 2320, pp.160–175 (2002).
29. Y. Chen, R. Venkatesan, M. Cary, R. Pang, S. Sinha, M. Jakubowski, "Oblivious Hashing: A Stealthy Software Integrity Verification Primitive", Proc. 5th Information Hiding Workshop (IHW), Netherlands (October 2002), Springer LNCS 2578, pp.400–414.
30. P. Gutmann, "An Open-source Cryptographic Co-processor", Proc. 2000 USENIX Security Symposium.
31. E. Eilam, Reversing: Secrets of Reverse Engineering, Wiley Publishing, Inc., 2005.

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Broadcasting methods in mobile ad hoc networks: Taxonomy and current state of the art

By Shaik Jaffar, Dr. M.V. Subramanyam
Madina Engineering college

Abstract - Flooding also known as broadcasting is one of the most primitive methodologies that focus on investigating searches concerning mobile ad hoc networking due to poorer network procedures which is a main feature in the concept of broadcasting which provides implications to superior applications that includes routing. Broadcasting means in conventional ways transmitting messages from a given branch to all other branches present in a network. The whole grid of the network is manned to ensure that the transmitted data is uniformly ported to the remaining nodes in a decentralized type of network setup. The two issues that renders nodes out of reach all the time are limited radio range and their immovability which assists in concluding that te issue of data transmission covering all networks is assumed to be a multi-objective issue that aims at increasing the count of number of nodules and also decreasing the time taken to reach the specified nodules and also reducing the network overhead which is a crucial characteristic because of the fact that this may direct to congestion also known as broadcast storm issue. This article aims at giving an insight of the taxonomy of transmitting methodologies in MANETS and current state of the art.

GJCST Classification: C.2.2, C.2.4,C.2.6

Strictly as per the compliance and regulations of:

Broadcasting methods in mobile ad hoc networks: Taxonomy and current state of the art

Shaik Jaffar [α], Dr. M.V. Subramanyam [Ω]

Abstract - Flooding also known as broadcasting is one of the most primitive methodologies that focus on investigating searches concerning mobile ad hoc networking due to poorer network procedures which is a main feature in the concept of broadcasting which provides implications to superior applications that includes routing. Broadcasting means in conventional ways transmitting messages from a given branch to all other branches present in a network. The whole grid of the network is manned to ensure that the transmitted data is uniformly ported to the remaining nodes in a decentralized type of network setup. The two issues that renders nodes out of reach all the time are limited radio range and their immovability which assists in concluding that te issue of data transmission covering all networks is assumed to be a multi-objective issue that aims at increasing the count of number of nodules and also decreasing the time taken to reach the specified nodules and also reducing the network overhead which is a crucial characteristic because of the fact that this may direct to congestion also known as broadcast storm issue. This article aims at giving an insight of the taxonomy of transmitting methodologies in MANETS and current state of the art.

I. Introduction

A mobile ad hoc network ensures building a provisional network sans the involvement of a recognized transportation or an integrated administration. MANETs are usually used for the common usage to emergency situations in warfields, rescue sites etc.

Every node present in MANET can be considered a router. The source node utilizes the intermediate nodes to transmit the message towards the destination node if a source node fails to transmit a message unswervingly to its destination node. MANET networks propose reliability, bandwidth and battery power and have erratic traits like topology. Strength signal and transmission routes. Transmission algorithms and procedures are supposed to be very light to save energy and bandwidth in computation and storage necessities [1, 2, 3, 4, and 5].

Routing information discovery is crucial for all MANET networks using standards such as dynamic source routing (DSR), ad hoc on demand distance vector (AODV), zone routing protocol (ZRP) and location aided routing (LAR) employ the procedure of transmission to launch routes, which can be achieved through the process of data transmission where sender sends a data packet to rest all branches present in MANET. Node mobility and limited system reserves pose serious issues in broadcasting MANETs as compared to wired networks.

II. Related Work

Transmission standards have been categorized into 4 groups namely simple flooding, probability-based methods, area-based methods and neighbor knowledge based methods, in accordance with the fact that the branches should be in order so as to be implemented by Williams and Camp. Simple flooding involves forwarding received data packets by branches one at a time resulting in jamming of network. Probability related methodologies are typically appraised time and again whenever a packet reaches the destination node which happens with some specific probability. When some extra topographical locations are covered due to some emission process, then re-transmission can be expected where area and location related methods explain whether the facades or the span of the projected area is low or not and if yes, then the message is not resent. GPS or estimation by the triangulation procedure or calculation of power of radio signals hold basis for provision of network information. SBA, Flooding With Self-Pruning (FSWP), AHBP, Multi-point Relaying, etc are few neighbor knowledge related procedures which are essential to procure information of neighborhood neighbors like FWSP uses 1-hop neighbors, SBA, Multipoint Relaying and AHBP uses 2-hops neighbors etc. They comprise the last category of transmission methodologies which are again segregated into 2 sub-divisions: neighbor-designating and self-pruning methods whose standard procedure themselves take a decision whether to retransmit the message or not while the former functions by activating its fellows wich are ready to relay a packet.

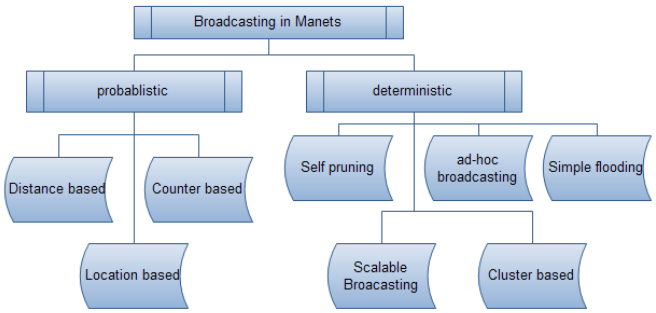
Stojmenovic and Wu introduced some classifications for transmitting standards which are dependent on their algorithmic nature or the data that is essential for its implementation (network information, “hello” messages content, broadcast messages content). When a transmitting algorithm works assuringly on a specified input, a confirmed output is always projected most of which are deterministic and is assumed to be secure only when all the noes [resent in

Author α : Associate professor, Madina Engineering college, kadapa, India. Telephone: 919441112010, E-mail : sj3j@rediffmail.com
Author Ω: Dr.M.V Subramanyam Ph.D, Principal, Santhi ram Engineering College, Nandyal, Kurnool Dist, A.P, India. Telephone: 919440352909 E-mail : mvsraj@yahoo.com

the network are taken into consideration. Probabilistic schemes and area-based methods are almost always risky to rely upon due to the fact that they usually fail in terms of randomness and heuristics, respectively.

Wu and Lou introduced a concept stating the quantity of data needed for transmission and also classified standards based on whether they depend on areas including global, quasi-global, local or quasi-local knowledge of the prevalent network wherein global and quasi-global transmitting algorithms are known as centralized standards whose main disadvantage is they are not scalable and hence can be utilized in MANETs. There are few localized standards whose examples are 1 and 2-hops neighborhood standards whose network status information and its topology are exchanged between various branches which is transmitted either by some random "hello" message or transmitted messages whose data content lays down a grave collision on the network throughput ultimately.

a) Taxonomy of broadcasting methods in mobile ad hoc networks



Making utmost use of the IEEE 802.11 MAC specifications, transmitting methodologies have been classified into four groups.

b) Statistical and geometrical model based broadcasting methods

Retransmitting of data packets ensue which involves every node in the simple flooding technique.

- Messages are distributed to all the neighboring nodes by a source node in MANET, the nodes will scan and check whether they have already seen the transmitted message and if yes, the packet is discarded and if not, t will again be re-initiated to all the potential nodes until the message reaches to every node present in the network. This methodology poses the issue of network jamming and weakening of battery power due to the presence of low concentration of nodes and high mobile power. If the messages constitute a polynomial number whose magnitude is (n2), it is of size n and is portrayed in the above diagram.

The topology of the prevalent network designs options for retransmitting of nodes based on probability standards.

- Probability Based Approach: This concept helps to identify and rectify the issues created due to the application of simple flooding methodology. A fixed probability pi for retransmission is assigned for each node I 2 N which involves lessening of the jamming circumstance and avoiding collisions. In situations when pi=1, then this concept turns itself towards simple flooding concept. There is a sufficient decrease in pi if there exists efficient transmission because of the increase and reduction in the count of neighbor density nodes.
- Counter-Based Scheme Approach: The random assessment delay (RAD) is posted, a threshold K is resolved and a counter k >= 1 is fixed on the basis of the count of the received transmitted message which in due course of RAD is increased considerably by one for every acknowledged message. The message is declined when RAD terminates and k >= 1. Few nodes won't be permitted to re-transmit in an opaque MANET while in a less intense MANET, all nodes will retransmit the messages.

An area related common broadcasting span is presumed and a node retransmits the message if there is a provision of adequate coverage location. Span and area based approaches are included in the methodologies mentioned below which can be explained as follows:

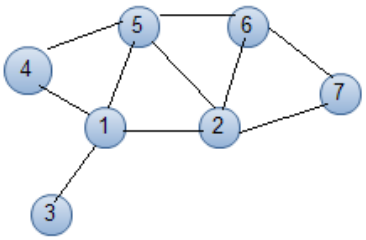
Distance Based Approach: The counter is made use of in the counter based approach to decline or retransmit a message wherein here, the span concerning the source and the destination node will be chosen by them both, say suppose the span is d. if the value of d is small, the retransmitting coverage span is less and if d is large enough, then the coverage will also be large but if d=0, then the coverage value is 0. The threshold span D is established by a receiving nodule and then RAD is preset where superfluous messages will be preserved until RAD is terminated. Now if d<D, then received transmitted messages will be declined else they'll be retransmitted again. It has been proposed by Ni et al that signal strengths are made use of to estimate the span starting from the source nodule. Span is capable enough to restore signal power by handling the signal threshold.

Location Based Approach: This is concerned with every nodule covering the basic need of instituting self sites for calculating extra coverage more clearly which is based on the global positioning system (GPS). Every node in MANET takes care in attaching its self area to the header part of every message that it is sending or retransmitting. The location of the sender is first deemed and extra coverage span is estimated where the message is declined in case RAD terminates when span area is considerably less than the given threshold. The price of estimating extra coverage spans, also considering estimations of intersections among

circles is a setback for this approach which can consume the inadequate energy currently available.

Neighborhood based: Status that is prevalent in the neighborhood is managed by the same method wherein matter received from the fellow nodes is employed for retransmission.

Self Pruning: Every node present in this feature is supposed to be aware of who its neighbors are, which can be attained by episodic messages. The receiving nodule evaluates with the source's list as to who all its neighbors are and may retransmit if extra nodes are within reach, else they will al be declined. Figure below depicts retransmission of message from node 2 to node 1 which retransmits them to node 3 and 4 respectively as they are extra nodes and so does node 5 with node 4. Idleness is rampant here even under such circumstances.



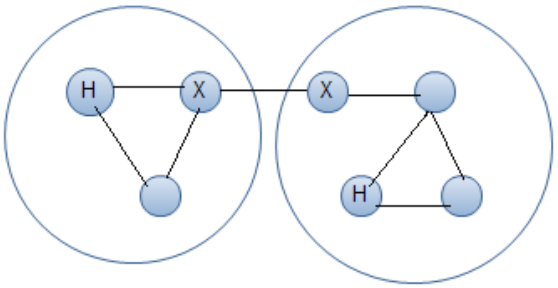
- Scalable Broadcasting Approach: Self pruning is enhanced here as there exist higher opportunities for message broadcasting later and also is supposed to make sure that all nodes present in the network are aware of their neighbors until the 2 hop span as each and every node in this methodology consists of a two-hop topology information, which is instituted by "Hello" messages.
- Ad Hoc Broadcasting Approach: This methodology permits only nodules chosen as gateway nodules and a transmission message controller to retransmit the message. It can be explained in detail as follows:

1. Choose one hop fellow nodule as gateway from amongst all the two hop neighboring nodes that can be accessible to a one hop fellow node.
2. Estimate cover set which receives message from the presently in use gateway set.
3. Chose and fix a one hop fellow node as gateway which can cover almost every two hp neighbor that's not present in the cover set.
4. Continue the above processes 2 and 3 till all two hop fellow nodules are enclosed.
5. The nodule which receives a message and is known to be a gateway, decides on which of its fellow nodules have already accepted delivery of the message, they are then presumed to be covered and is declined by the fellow nodule to chose the subsequent hop gateways.

c) Cluster Based Broadcasting Methods

Data traffic organization schemas, routing severities, fault tolerance problems are few topics of concern for which clustering approach is considered.

For enlightening its existence, every nodule sends timely "Hello" messages and possesses an exclusive ID. A group of nodes is collectively known as a cluster which can be fashioned as follows: A nodule possessing a restricted negligible IT will nominate self as the cluster in-charge within where a gateway is utilized for transmission between two members of different clusters. In case the two in-charge heads come together, the nodule with the bigger ID status sacrifices its head position. Cluster configuration can be portrayed as follows:



X is a gateway here while Y is the in-charge head.

d) Tree Based broadcasting Methods

This methodology is usually not preferable and termed as unsuitable for MANETs even though transmission with the help of tree techniques in wired networks is a famous and technique that is in use too often as they portray a drastic and a powerful transformation in network topologies.

e) Technical challenges

This segment focuses on few issues of importance that are to be tackled during the tenure of outlining transmission standards.

f) Hidden and exposed node problems

This issue is of a major concern during transmission as it renders the transport of nodes to all locations unfeasible concerning a given phase of network subsets. This is supposed to be rectified by the usage of acknowledgement packets (ACKs) but this would be against the rule of transmission that specifies to reduce the count of data packet production. Subsequently, the main principle of transmission within a static ad hoc network is to avail to as many nodes that are within reach. A transmitting standard may be unsuccessful even at the slightest hint of non-availability of acknowledgement packets.

g) Mobility and Partitioning

Mobility is another important issue of concern, the transmitting standards are supposed to face.

Transmission of messages on the basis of spanning trees endures problems posed by the mobility factor, which can now be dealt with expertly because of the availability of many equipped algorithms that are competent enough. These methodologies also are not specific on which application they want to work with. Hence, considering transmission tends to give an idea as to how the corresponding spanning tree is supposed to be built, which is done by exploring group of relevant nodes that have already been recipients of the sender.

This issue in static ad hoc networks can also unwillingly pave way to separated networks which is another issue of major concern, taking broadcasting in particular. A transmitting standard is supposed to be searching evidently for a methodology which can permit a transmitting message to leap to various subsets, so as to cover as many nodules as it is permitted. Epidemic dispersive replicas can be correlated with this concern.

h) Frequently cited broadcasting methods literature

The cluster that materializes in MANET is supposed to be preserved on a regular basis as said by Ni et al by the fundamental cluster algorithm. All the remaining nodules in a cluster other than the controlling nodule can be enveloped by the retransmission done by the controller. Gateway nodes are extensively utilized for retransmitting message to other nodules in various clusters and hence there is no extreme necessity of a non-gateway nodule to retransmit the same message. Even though there may be presence of many gateway nodules in various clusters, te specified gateways may probably utilize and employ different transmission concepts mentioned earlier to ascertain whether to retransmit or not.

Spanning trees are widely built though which transmission of messages is done by promoting messages only to the fellow nodules present in the tree which is basically acyclic in nature. Hence, every transmitted message is acquired only once at a time by the prevailing nodes. Many algorithms are available for building and upholding trees like the bridged Ethernet network's spanning tree algorithm which are built to prove suitable for working in stable networks rather than the ever-changing topology of a MANET.

Multicast trees and their uses have already been specified and explained in detail by many authors but what is nagging is the fact their algorithms are not qualified and efficient enough for dealing with the topology related modifications. Few algorithms render their services useless for handling the tree in an ever-changing topology even though they may involve a phase of building a spanning tree.

The logic of one-to-one broadcasting is proved feasible by the tree related technique as compared to other methods as many disadvantages of limited transmissions do not influence the algorithm thankfully, it is secure enough for transmissions.

A noteworthy point to mention is there is minimum or no effort required to be at par with the network states as it has been proposed from the start to reduce signaling traffic.

i) Current state of the art

A multiple channel medium access control (MAC) standard was recommended by Jenhui Chen et al which was named as ad hoc multichannel negotiation protocol (AMNP) used mainly for transmitting messages across multiple channels in a uniform manner and also referred to the problem of distributed scheme allowance for multihop MANETs but in the presence of one transceiver. Augmentation of description of AMNP known as AMNP with channel development was brought into existence.

The replication results prove worthy to make its stand that the throughput is comparatively large in comparison to its single path equivalent. Only a single transceiver is deployed by the recommended AMNP but with a specified constraint of suspending right to admission for a specific time period while getting swapped to a chosen data route.

It has been suggested by Chien-Chung Shen et al, a diagram-prospect related directional to curve percolation and also for omni-directional transmission for spot percolation and also gives a detailed explanation about the compilation of directional transmitter related transmission methods for static ad hoc networks. The author squabbles to support the stipulation of suggested copy, that countless transmission designs have been recommended almost all of which presumed the practice of omni-directional transmitters and transmission overhead is taken into consideration advancing number of dispatching nodules. Directional transmitters possess tapered emissions and can gradually diminish transmission overheard with respect to number of acknowledged packets to the count of nodules that expects transmission packets.

Observation: It has been suggested that diagram-prospect related directional to curve percolation and also for omni-directional transmission for spot percolation and also gives a detailed explanation about the compilation of directional transmitter related transmission methods for static ad hoc networks. A specific quantity of battery power is safeguarded by decreasing the count of replica of data packets that is acknowledged. For utilizing the longer range characteristics of directional transmitters for deceasing the delay, it is but essential to scrutinize the recommended ideas.

Transmission storm issue would be a troubling factor if the accelerating nodules are not cautiously allocated in the transmission procedure in static ad hoc networks (MANETs) said Wei Lou et al. The main idea behind diminishing transmission idleness is a foremost

issue of concern in MANETs and so an easy transmission algorithm has been recommended known as double-covered broadcast (DCB), which benefits from the transmission idleness state for enhancing delivery ratio in a high broadcasting error rate surrounding subset.

Few chosen promoting nodules rebroadcast the transmitted message from among the 1-hop fellow nodules pertaining to the sender. The above mentioned nodules are chosen as follows:

1) 2-hop fellow nodules of the source initiator are swathed and
2) The source initiator's 1-hop fellow nodules are either advancing nodules or non advancing ones that are enveloped by a minimum of two promoting fellow nodules.

The source initiator acquires hold of the rebroadcasts of the promoting nodules as authentication of reaction of the data packet. The non promoting 1-hop fellow nodules of the source initiator fail to admit the response reaction of their transmission which provokes the initiator to retransmit the data packet in case of it failing to recognize all the promoting nodules broadcast till large count of retries has been reached.

Observation: Suppose in DCB, a nodule v promotes a data packet choosing division of the prevalent 1-hop fellow nodules as promoting nodules depending on the greedy algorithm for the issue pertaining to Set Cover with certain restrictions which are as follows: (1) All the 2-hop fellow nodules of node v are supposed to be addressed by the chosen advancing nodes and (2) the 1-hop fellow nodules available in node v can either be chosen as an advancing nodule r enveloped by a minimum of two promoting nodes. Then, the IDs of the chosen promoting nodules to the relevant data packet is appended to the node v and then the same packet is transmitted. It is assumed beforehand by a 1-hop fellow node that the about-to-be received data packet is of a promoting nature and transmission procedure continues as it was before concerning node v. Another noteworthy characteristic of DCB is that whenever a nodule tends to broadcast a packet secure connections are ensured. Node v delays time to eavesdrop on transmission from all its chosen promoting nodules and if it falls short for the same, then rebroadcasting is permitted until all nodules are swathed and maximum tries are attained.

j) Limits and obstacles observed in Existing Broadcasting Methods

The disadvantages construed from itemizing relative studies are as follows:

1. There is a need for many retransmissions concerning with the count of the rebroadcasting nodules for all methodologies explained except for the neighbor related techniques.
2. RAD implemented techniques drowned in high density MANETs acclimatize RAD nodules concerning its surrounding conduct is cultivated.
3. The ad hoc transmitting technique faces discrepancies in a typically high static MANET network because of the fact that it fails to make use of local data to confirm whether to retransmit the packet or not.

On the basis of the wide proportional research on already existing transmitting techniques, it has been observed that every transmitting technique failed to work in wide ranging MANET surroundings.

Scalable transmission based line of attack has provided with noteworthy and promising results as compared to the non-adaptive tactics.

There is an urgent need to cultivate new competent data transmission tactics with the main intention of preserving the existing meager reserves in MANETs.

VII. Conclusion

Transmission is an indispensible feature for any MANET network, so it is vital to exploit the most proficient transmitting techniques that can make sure that a secure network is provisioned. This paper also has presented a brief synopsis on all chief transmission methodologies available in the prose, mainly concentrating on the intricacies of their roles and also the threats posed by them and also, recommending upgrading for few of the enumerated techniques. There is not one most favorable algorithm in existence for all the concerning techniques in the present circumstances even though a vigorous change is pertinently visible in the MANET topology and its rarely obtainable reserves.

References Références Referencias

1. Park and S. Corson. A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks. INFOCOM 97, 1407-1415, 1997.
2. Z. Haas. A New Routing Protocol for Reconfigurable Wireless Networks. ICUPC 97, 562-566, 1997.
3. C. Perkins and E. Royer Ad hoc on Demand Distance Vector Routing. 2nd IEEE Workshop on Mobile Computing Systems and Applications, 3-12, 1999.
4. P. Sinha, R. Sivakumar and V. Bharghavan. CEDAR: A Core Extraction Distributed Ad hoc Routing Algorithm. INFOCOM 99, 202-209, 1999.
5. D. Johnson and D. Maltz. Dynamic Source Routing in Ad hoc Wireless Networks. Mobile Computing, Academic Publishers, 153-181, 1996.
6. D. Johnson, D. Maltz and Y. Hu. The Dynamic Source Routing Protocol for Mobile Ad hoc Networks. Internet Draft: draft-ietf-manet-dsr-09.txt, 2003.

7. C. Perkins, E. Beldig-Royer and S. Das. Ad hoc on Demand Distance Vector (AODV) Routing. Request for Comments 3561, July 2003.
8. Z. Haas and M. Pearlman. The Performance of Query Control Schemes for the Zone Routing Protocol. IEEE/ACM Transactions on Networking, 9(4):427–438, 2001.
9. Z. Haas and B. Liang. Ad hoc mobility management with randomized database groups. Proceedings of the IEEE International Conference on Communications, 1756-1762, 1999.
10. Y. Ko and N. Vaidya. Location-aided Routing (LAR) in Mobile Ad hoc Networks. Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM),66-75, 1998.
11. I. S. Committee. Wireless LAN Medium ACCESS CONTROL (MAC) and Physical Layer Specifications. IEEE 802.11 Standard. IEEE, New York, ISBN 1559379359, 1997.
12. B. Williams and T. Camp. Comparison of Broadcasting Techniques for Mobile Ad hoc Networks. In Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC), 194–205, 2002
13. C. Ho, K. Obraczka, G. Tsudik and K. Viswanath. Flooding for Reliable Multicast in Multi-hop Ad hoc Networks. International Workshop in Discrete Algorithms and Methods for Mobile Computing and Communication, 64–71, 1999.
14. J. Jetcheva, Y. Hu, D. Maltz and D. Johnson. A Simple Protocol for Multicast and Broadcast in Mobile Ad hoc Networks. Internet Draft, draftietf-manet-simple-mbcast-01.txt, 2001.
15. S. Ni, Y. Tseng, Y. Chen and J. Sheu. The Broadcast Storm Problem in a Mobile Ad hoc Network. International Workshop on Mibile Computing and Networks, 151–162, 1999.
16. B. Williams and T. Camp. Comparison of broadcasting techniques for mobile ad hoc networks. In Proc. of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC), pages 194–205, 2002.
17. Sze-Yao Ni, Yu-Chee Tseng, Yuh-Shyan Chen, and Jang-Ping Sheu. The broadcast storm problem in a mobile ad hoc network. Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, pages 151–162, 1999.
18. Stojmenovic and J. Wu. Broadcasting and activity scheduling in ad hoc networks. S. Basagni, M. Conti, S. Giordano, I. Stojmenovic, Eds., Mobile Ad Hoc Networking, pages 205–229, 2004.
19. J. Wu and W. Lou. Forward-node-set-based broadcast in clustered mobile ad hoc. Wireless Communications and Mobile Computing, 3:155–173, 2003.
20. A. Pelc. Broadcasting in wireless networks. Handbook of Wireless Networks and Mobile Computing, pages 509–528, 2002.
21. E. D. Kaplan. Understanding GPS: Principles and Applications. Artech House, Boston, MA, 1996.
22. A. Casteigts. Model driven capabilities of the da-grs model. In International Conference on Autonomic and Autonomous Systems (ICAS'06), San Francisco, USA, 2006. IEEE.
23. M. Gerla and J. T. Tsai. Multicluster, Mobile, Multimedia Radio Network. ACM-Baltzer Journal of Wireless Networks, 1(3):255–265, 1995.
24. S. Alagar, S. Venkatesan, and J. Cleveland. Reliable Broadcast in Mobile Wireless Networks. In Military Communications Conference, MILCOM Conference Record, 1:236-240, 1995.
25. Alpr Jttner and dm Magi. Tree Based Broadcast in Ad hoc Networks. MONET Special Issue on WLAN Optimization at the MAC and Network Levels, 2004. to appear.
26. IEEE Specification 820.1d MAC Bridges D9, July 14, 1989.
27. A. Ballardie, P. Francis and J. Crowcroft. Core Based Trees (CBT) an Architecture for Scalable Interdomain Multicast Routing. SIGCOMM93, San Francisco, 85–95, 1993.
28. D. Estrin. Protocol Independent MulticastSparse Mode (PIM-SM). Protocol Specification RFC 2362, June 1998.
29. A. Adams, J. Nicholas and W. Siadak. Protocol Independent Multicast- Dense Mode (PIM-DM). Protocol Specification (Revised), IETF Internet-Draft, draft-ietf-pim-dm-new-v2-02.txt, October 2002.
30. D. Waitzman, C. Partridge and S. Deering. Distance Vector Multicast Routing Protocol. RFC 1075, November 1988.
31. J. Moy. Multicast Routing Extensions for OSPF. CACM, 37:61–66, 1994.
32. C. Perkins, E. Royer and S. Das. Ad hoc On-demand Distance Vector (AODV) Routing. IETF Internet-Draft, draft-ietf-manet-aodv-11.txt, Aug 2002.

Protocol Specification (Revised), IETF Internet-Draft, draft-ietf-pim-dm-new-v2-02.txt, October 2002.
30. D. Waitzman, C. Partridge and S. Deering. Distance Vector Multicast Routing Protocol. RFC 1075, November 1988.
31. J. Moy. Multicast Routing Extensions for OSPF. CACM, 37:61–66, 1994.
32. C. Perkins, E. Royer and S. Das. Ad hoc On-demand Distance Vector (AODV) Routing. IETF Internet-Draft, draft-ietf-manet-aodv-11.txt, Aug 2002.

33. Chen, J.;, "AMNP: ad hoc multichannel negotiation protocol with broadcast solutions for multi-hop

mobile wireless networks," Communications, IET , vol.4, no.5, pp.521-531, March 26 2010; doi: 10.1049/iet-com.2009.0318

34. Chien-Chung Shen; Zhuochuan Huang; Chaiporn Jaikaeo; , "Directional broadcast for mobile ad hoc networks with percolation theory," Mobile Computing, IEEE Transactions on , vol.5, no.4, pp. 317- 332, April 2006; doi: 10.1109/TMC.2006.1599402
35. Wei Lou; Jie Wu; , "Toward Broadcast Reliability in Mobile Ad Hoc Networks with Double Coverage," Mobile Computing, IEEE Transactions on , vol.6, no.2, pp.148-163, Feb. 2007; doi: 10.1109/TMC.2007.31
36. Chun-Yuan Chiu; Wu, E.H.-K.; Gen-Huey Chen; , "A Reliable and Efficient MAC Layer Broadcast Protocol for Mobile Ad Hoc Networks," Vehicular Technology, IEEE Transactions on , vol.56, no.4, pp.2296-2305, July 2007; doi: 10.1109/TVT.2007.897654
37. Ge-Ming Chiu; Cheng-Ru Young; , "Exploiting In-Zone Broadcasts for Cache Sharing in Mobile Ad Hoc Networks," Mobile Computing, IEEE Transactions on , vol.8, no.3, pp.384-397, March 2009; doi: 10.1109/TMC.2008.127
38. Xie, J.; Das, A.; Nandi, S.; Gupta, A.K.; , "Improving the reliability of IEEE 802.11 broadcast scheme for multicasting in mobile ad hoc networks," Communications, IEE Proceedings- , vol.153, no.2, pp. 207- 212, 1 April 2006; doi: 10.1049/ip-com:20045271
39. Wolf, B.J.; Hammond, J.L.; Noneaker, D.L.; Russell, H.B.; , "A protocol for construction of broadcast transmission schedules in mobile ad hoc networks," Wireless Communications, IEEE Transactions on , vol.6, no.1, pp.74-schedules in mobile ad hoc networks," Wireless Communications, IEEE Transactions on , vol.6, no.1, pp.74-78, Jan. 2007; doi: 10.1109/TWC.2007.05210 .

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 1 Version 1.0 January 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Enhanced Cuckoo Search for Optimization of Bloom Filter in Spam Filtering

By Arulanand Natarajan, Subramanian S, Premalatha K
Anna University of Technology, Coimbatore

Abstract - Bloom Filter (BF) is a simple but powerful data structure that can check membership to a static set. The trade-off to use Bloom filter is a certain configurable risk of false positives. The odds of a false positive can be made very low if the hash bitmap is sufficiently large. Spam is an irrelevant or inappropriate message sent on the internet to a large number of newsgroups or users. A spam word is a list of well-known words that often appear in spam mails. The proposed system of Bin Bloom Filter (BBF) groups the words into number of bins with different false positive rates based on the weights of the spam words. An Enhanced Cuckoo Search (ECS) algorithm is employed to minimize the total membership invalidation cost of the BFs by finding the optimal false positive rates and number of elements stored in every bin. The experimental results have demonstrated for CS and ECS for various numbers of bins.

Keywords : *Bin Bloom Filter, Bloom Filter, Cuckoo Search, Enhanced Cuckoo Search, False positive rate, Hash function, Spam word.*

GJCST Classification: *D.3.4 , G.1.6, B.1.4*

Strictly as per the compliance and regulations of:

An Enhanced Cuckoo Search for Optimization of Bloom Filter in Spam Filtering

Arulanand Natarajan [α], Subramanian S [Ω], Premalatha K [β]

Abstract - Bloom Filter (BF) is a simple but powerful data structure that can check membership to a static set. The trade-off to use Bloom filter is a certain configurable risk of false positives. The odds of a false positive can be made very low if the hash bitmap is sufficiently large. Spam is an irrelevant or inappropriate message sent on the internet to a large number of newsgroups or users. A spam word is a list of well-known words that often appear in spam mails. The proposed system of Bin Bloom Filter (BBF) groups the words into number of bins with different false positive rates based on the weights of the spam words. An Enhanced Cuckoo Search (ECS) algorithm is employed to minimize the total membership invalidation cost of the BFs by finding the optimal false positive rates and number of elements stored in every bin. The experimental results have demonstrated for CS and ECS for various numbers of bins.

Keywords : Bin Bloom Filter, Bloom Filter, Cuckoo Search, Enhanced Cuckoo Search, False positive rate, Hash function, Spam word.

I. Introduction

A spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting into user's inbox. A spam filter looks for certain criteria on which it stands decisions. For example, it can be set to look for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not effective, because often it is omitting perfectly legitimate messages and letting actual spam through. The strategies used to block spam are diverse and includes many promising techniques. Some of the strategies like black list filter, white list /verification filters rule based ranking and naïve bayesian filtering are used to identify the spam.

A BF presents a very attractive option for string matching (Bloom 1970). It is a space efficient randomized data structure that stores a set of signatures efficiently by computing multiple hash functions on each member of the set.

It queries a database of strings to verify for the membership of a particular string. The answer to this query can be a false positive but never be a false negative. The computation time required for performing the query is independent of the number of signatures in the database and the amount of memory required by a BF for each signature is independent of its length (Feng et al 2002).

This paper presents a BBF which allocates different false positive rates to different strings depending on the significance of spam words and gives a solution to make the total membership invalidation cost minimum. BBF groups strings into different bins via smoothing by bin means technique. The number of strings to be grouped and false positive rate of each bin is identified through GA which minimizes the total membership invalidation cost. This paper examines different number of bins for given set of strings, their false positive rates and number of strings in every bin to minimize the total membership invalidation cost.

The organization of this paper is as follows. Section 2 deals with the standard BF. Section 3 presents the CS technique. Section 4 reports optimized BBF using ECS. Performance evaluation of CS and ECS for the BBF is discussed in section 5.1

II. Bloom Filter

Bloom filters (Bloom 1970) are compact data structures for probabilistic representation of a set in order to support membership queries. This compact representation is the payoff for allowing a small rate of false positives in membership queries which might incorrectly recognize an element as member of the set.

Given a string S the BF computes k hash functions on it producing k hash values and sets k bits in an m-bit long vector at the addresses corresponding to the k hash values. The value of k ranges from 1 to m. The same procedure is repeated for all the members of the set. This process is called programming of the filter. The query process is similar to programming, where a string whose membership is to be verified is input to the filter. The bits in the m-bit long vector at the locations corresponding to the k hash values are looked up. If at least one of these k bits is not found in the set then the string is declared to be a nonmember of the set. If all the bits are found to be set then the string is said to belong to the set with a certain probability. This uncertainty in the membership comes from the fact that those k bits in the m-bit vector can be set by any other n-1 members. Thus finding a bit set does not necessarily imply that it was set by the particular string being queried. However, finding a bit not set certainly implies that the string does

Author α : Anna University of Technology, Coimbatore.
E-mail : arulnat@yahoo.com
Author Ω : Sri Krishna College of Engineering and Technology, Coimbatore. E-mail : dsraju49@gmail.com
Author β : Bannari Amman Institute of Technology, Erode.
E-mail : kpl_barath@yahoo.co.in

not belong to the set. In order to store a given element into the bit array, each hash function must be applied to it and, based on the return value r of each function (r1, r2, … , rk), the bit with the offset r is set to 1. Since there are k hash functions, up to k bits in the bit array are set to 1 (it might be less because several hash functions might return the same value). Figure 1 is an example where m=16, k=4 and e is the element to be stored in the bit array.

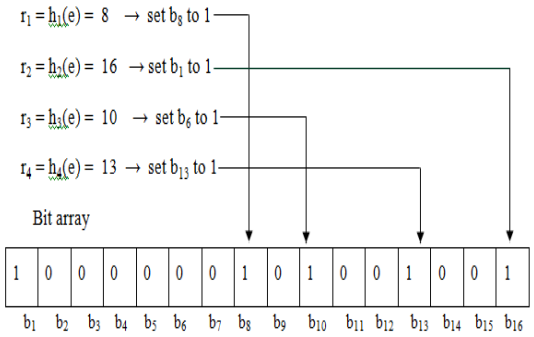


Fig.1 : Bloom Filter

One important feature of BF is that there is a clear tradeoff between the size of the filter and the rate of false positives. The false positive rate of BF is

$$f = \left(1 - e^{-kn/m}\right)^k = \exp(k \ln(1 - e^{-kn/m})) \quad (1)$$

Let $g = k \ln(1 - e^{-kn/m})$. Minimizing the false positive probability f is equivalent to minimizing with respect to k.

$$\frac{dg}{dk} = \ln\left(1 - e^{-\frac{kn}{m}}\right) + \frac{kn}{m} \frac{e^{-kn/m}}{1 - e^{-kn/m}} \quad (2)$$

The derivative equals 0 when kmin=(1n2)(m/n). In this case the false positive probability f is:

$$f(k_{min}) = (1 - p)^{k_{min}} = \left(\frac{1}{2}\right)^{k_{min}} = (0.6185)^{m/n} \quad (3)$$

of course k should be an integer, so k is $\lceil \ln 2 \,.\, (m/n) \rceil$

The BF has been widely used in many database applications (Mullin 1990; Mackert and Lohman, 1986). It is applied in networking literature (Brooder and Mitzenmacher, 2005). A BF can be used as a summarizing technique to aid global collaboration in peer-to-peer networks (Kubiatowicz et al., 2000 ; Li et al, 2002 ; Cuena-Acuna et al, 2003). It supports probabilistic algorithms for routing and locating resources (Rhea and Kubiatowicz 2004; Hodes et al,2002 ; Reynolds and Vahdat, 2003; Bauer et al, 2004) and share Web cache information (Fan et al,2000). BFs have great potential for representing a set in main memory (Peter and Panagiotis, 2004) in stand-alone applications. BFs have been used to provide a probabilistic approach for explicit state model checking of finite-state transition systems (Peter and Panagiotis, 2004). It is used to summarize the contents of stream data in memory (Jin et al,2004; Deng and Rafiei,2006), to store the states of flows in the on-chip memory at networking devices (Bonomi et al,2006), and to store the statistical values of tokens to speed up the statistical-based Bayesian filters (Li and Zhong,2006). The variations of BFs are compressed Bloom filters (Mitzenmacher,2002),counting Bloom filters (Fan et al,2000), distance-sensitive Bloom filters (Kirsch and Mitzenmacher,2006), Bloom filters with two hash functions (Kirsch and Mitzenmacher,2006), spacecode Bloom filters (Kumar et al,2004), spectral Bloom filters (Cohen and Matias,2003), generalized Bloom filters (Laufer et al,2005), Bloomier filters (Chazelle et al,2004), and Bloom filters based on partitioned hashing (Hao et al,2007).

III. Cuckoo Search

Cuckoo search is an optimization algorithm inspired by the brood parasitism of cuckoo species by laying their eggs in the nests of other host birds proposed by Yang and Deb (2009). If a host bird discovers the eggs are not their own, it will either throw these foreign eggs away or simply abandon its nest and build a new nest elsewhere. Each egg in a nest represents a solution, and a cuckoo egg represents a new solution. The better new solution (cuckoo) is replaced with a solution which is not so good in the nest. In the simplest form, each nest has one egg. When generating a new solution Levy flight is performed. The rules for CS are described as follows:

- Each cuckoo lays one egg at a time, and dumps it in a randomly chosen nest
- The best nests with high quality of eggs will carry over to the next generations;
- The number of available host nests is fixed, and a host can discover an foreign egg with a probability $p_a \in [0, 1]$. In this case, the host bird can either throw the egg away or abandon the nest so as to build a completely new nest in a new location

The algorithm for CS is given below:

Generate an initial population of n host nests;

while (t<MaxGeneration) or (stop criterion)

Get a cuckoo randomly (say, i) and replace its solution by performing Levy flights;

Evaluate its fitness Fi

Choose a nest among n (say, j) randomly;

if (Fi > Fj),[for maximization]

Replace j by the new solution;

end if

A fraction (pa) of the worse nests is abandoned and new ones are built;

Keep the best solutions/nests;

Rank the solutions/nests and find the current best;

Pass the current best to the next generation;

end while

IV. Enhanced Cuckoo Search for Bloom Filter Optimization

a) Bin Bloom Filter (BBF)

A BBF is a date structure considering weight for spam word. It groups spam words into different bins depending on their weight. It incorporates the information on the spam word weights and the membership likelihood of the spam words into its optimal design. In BBF a high cost bin lower false positive probability and a low cost bin has higher false positive probability. The false positive rate and number of strings to be stored is identified through optimization technique GA which minimize the total membership invalidation cost. Figure 2 shows Bin BF with its tuple <n,f,w> configuration.

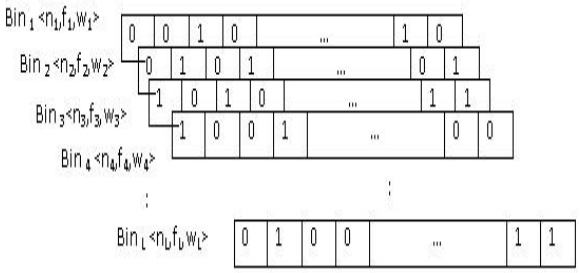


Fig.2 : Bin Bloom Filter

b) Problem Definition

Consider a standard supervised learning problem with a set of training data D = {<Y1,Z1 >,..., <Yi, Zi>, … ,< Yr ,Zr >} , where Yi is an instance represented as a single feature vector, Zi = C(Yi) is the target value of Yi , where C is the target function. Where Y1, Y2, … , Yr set of text document collection C is a class label to classify into spam or legitimate (non-spam). The collection is represented into feature vector by the text documents are converted to normalized case, and tokenized them, splitting on non-letters. The stop words are eliminated. The spam weights for words are calculated from the set. This weight value indicates its probable belongings to spam or legitimate. The weight values are discretized and assigned for different Bins. The tuple to describe the Bin BF is, {{n1, n2,, …, nL}, {w1, w2,…, wL}, m, {k1, k2, …, kL}, {f1, f2, …, fL}}. It is an optimization problem to find the value of n and f that to minimize the total membership invalidation cost. For membership testing the total cost of the set is the sum of the invalidation cost of each subset. The total membership invalidation cost (Xie et al., 2005) is given as,

$$F= n1f1w1 + n2f2w2 + \ldots\ldots + nLfLwL$$

The total membership invalidation cost

$$F(L) = \sum_{i=1}^{L} n_i w_i f_i \qquad (4)$$

to be minimized.

Where $$\sum_{i=1}^{L} n_i = N$$

N- Total number of Strings in a spam set.

$$f_i = \left(\frac{1}{2}\right)^{\ln 2 \times \left(r_i m \Big/ \sum_{j=1}^{i} n_j r_j \right)}$$

$$r_i = \ln(f_i) \quad (1 \le i \le L)$$

The objective function f(L) taken as standard for the problem of minimization is

$$f(L) = \begin{cases} C_{max} - F(L) & \text{if } F(L) < C_{max} \\ 0 & \text{if } F(L) \ge C_{max} \end{cases} \qquad (5)$$

where Cmax is a large constant.

c) ECS for Optimization of BF

The CS is extended to an ECS in which each nest has multiple eggs representing a set of solutions. Generate an initial population of n host nests with m eggs;

while (t<MaxGeneration) or (stop criterion)

Get a cuckoo randomly (say, i) by Levy flights using the best egg in the chosen nest;

Evaluate its fitness Fi

Choose a nest among n and choose an egg with the worst solution in the nest (say, j);

if (Fi > Fj),[for maximization]

Replace j by the new solution i;

end if

Find the best solution (among m) in each nest;

Rank the nests based on the best solution;

Abandon a fraction (pa) of the nests which have worse solutions and built new ones;

Keep the best solutions/nests;

Rank the solutions/nests and find the current best;

end while

When generating new solutions x(t+1) for a cuckoo i, a Levy flight is performed using the following equation (6)

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus \text{Levy}(\lambda) \quad (6)$$

The symbol ⊕ is an entry-wise multiplication. Basically Levy flights provide a random walk while their random steps are drawn from a Levy distribution for large steps

$$\text{Levy} \sim u = t^{-\lambda} \quad (7)$$

which has an infinite variance with an infinite mean. Here the consecutive jumps of a cuckoo essentially form a random walk process which obeys a power-law step- length distribution with a heavy tail. The representation of egg (solution) is given in figure 3.

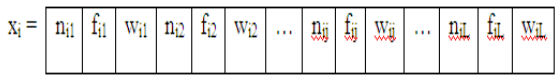


Fig.3 : Egg representations for Bin Bloom Filter

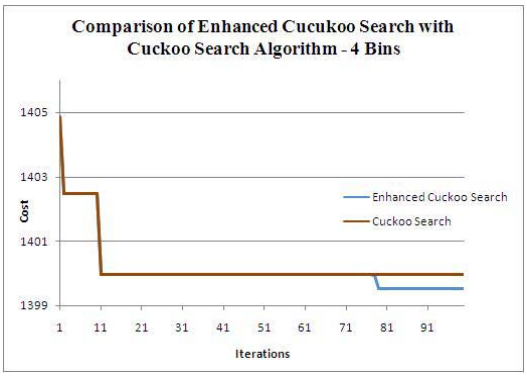
where nij, fij and wij refer respectively the number of words, false positive rate of and the weight of the jth bin of ith egg. The triplet <n,f,w> encodes a single bin. The false positive rate fij can be obtained from equation (1) where nij is drawn from the ith egg in the nest, m is known in advance and k is calculated from equation (3). One egg in the nest represents one possible solution for assigning the triples <n, f, w>. At the initial stage, each egg randomly chooses different <n, f, w> for L Bins based on the given constraints. The fitness function for each egg can be calculated based on the equation (5).

VII. Experimental Results

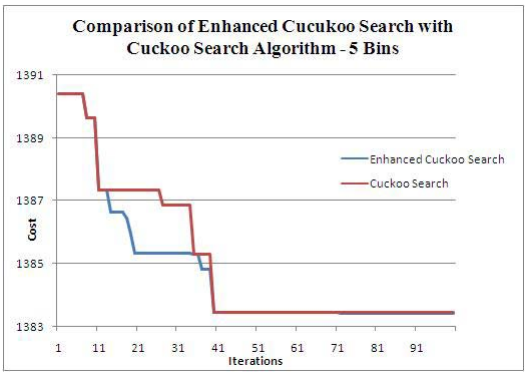
Cuckoo Search employs Levy flight for finding new solutions from equation (7). CS and ECS consider 10 nests and 50 iterations. The parameters pa,α and λ are set as 0.3, 1 and 1.5 respectively. The total number of strings taken for testing is 250, 500, and 1000. The string weights are varying from 0.0005 to 5. The size of the BF is 1024. This experimental setup is applied for number of bins from 4 to 7.

Figures 4a, 4b, 4c and 4d correspondingly show the total membership invalidation cost obtained from BBF for bin sizes from 4 to 7 for 1000 strings using CS and ECS algorithm. In this experimental setup the ECS performs better than CS. Figures 5a, 5b, 5c and 5d show the total membership invalidation cost obtained from BBF for bin sizes from 4 to 7 respectively for 500 strings. Figures 6a, 6b, 6c and 6d show the cost of BBF from bin sizes 4 to 7 for 250 strings. For all the string sizes the ECS outperforms CS.

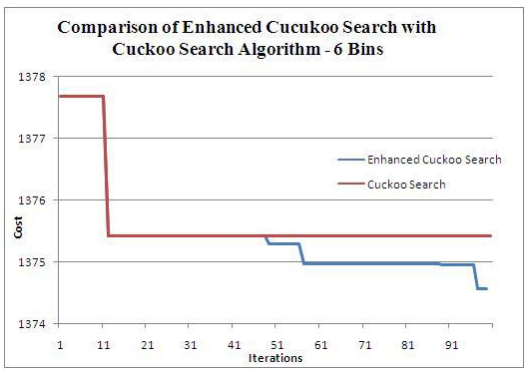
In CS, 10 nests which equals to number of nests in ECS and 40 nests which equals to number of eggs in ECS are taken to find the total membership invalidation cost for 1000 strings. Figure 7 shows the total membership invalidation cost obtained from BBF for the bin sizes ranging from 4 to 10 using CS and ECS. It shows that the cost is decreased when the numbers of bins are increased. The results obtained from ECS outperform CS for all bin sizes from 4 to 10.



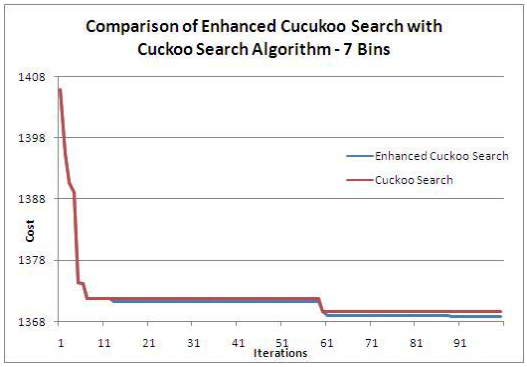
(A)



(B)

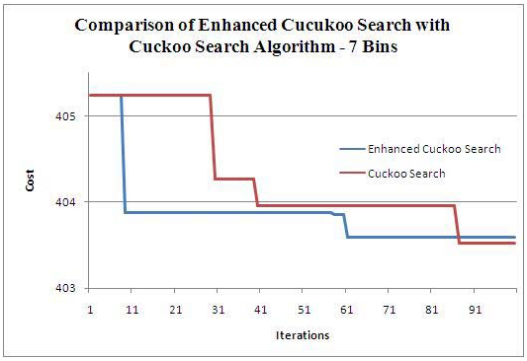


(C)



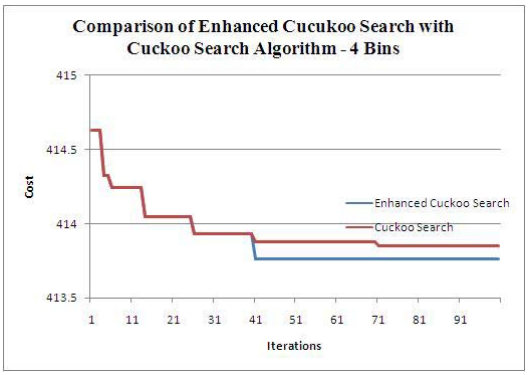
(D)

Fig.4 : Values obtained for 1000 Strings

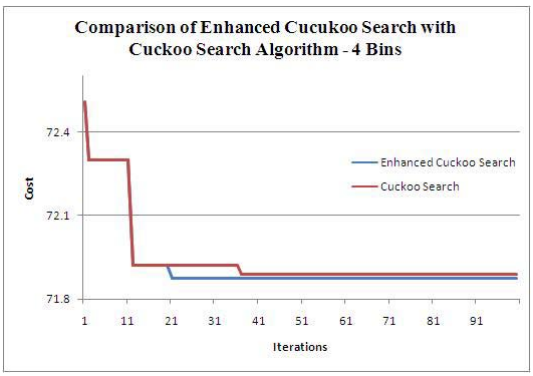


(D)

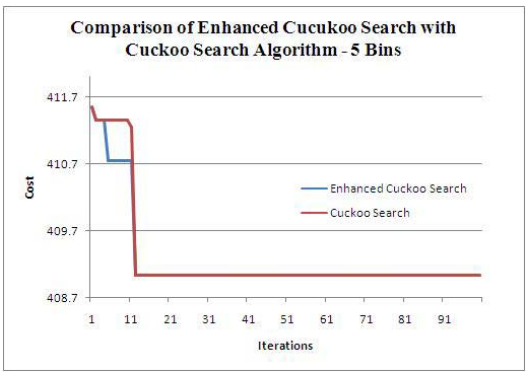
Fig.5 : Values obtained for 500 Strings



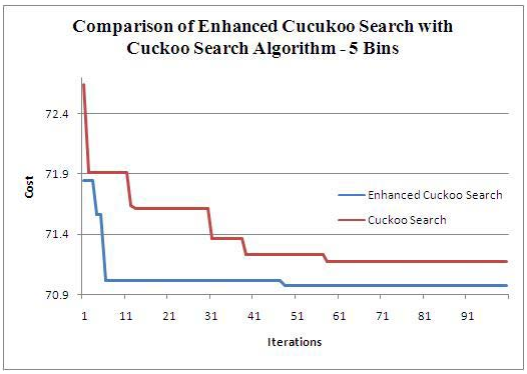
(A)



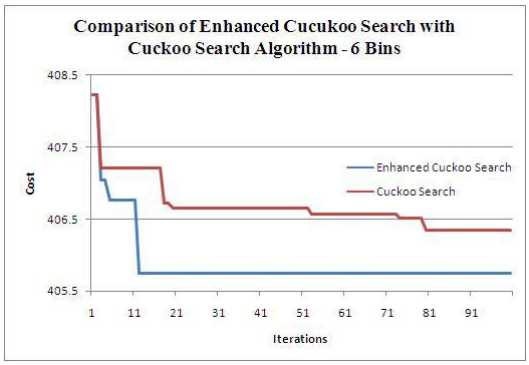
(A)



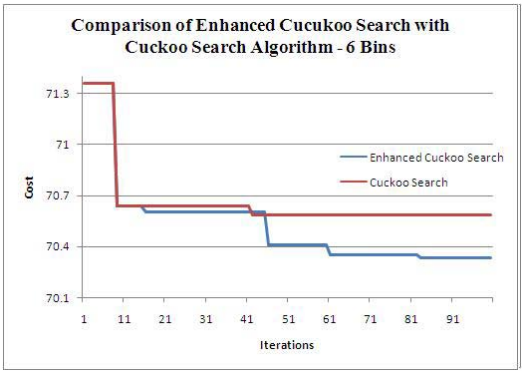
(B)



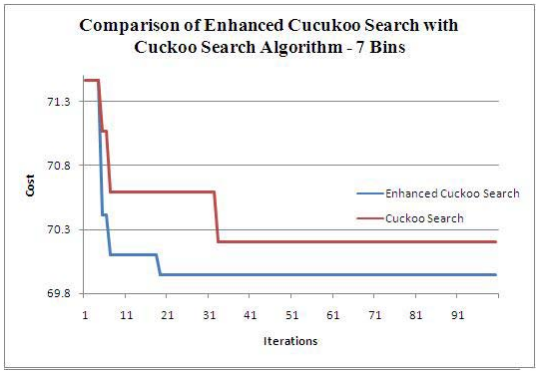
(B)



(C)



(C)



(D)

Fig.6 : Values obtained for 250 Strings

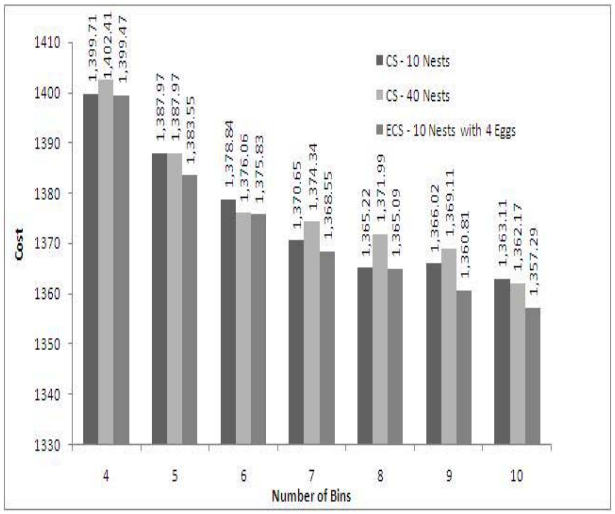


Fig.7 : Total Membership invalidation cost for CS and ECS

VIII. Conclusion

BFs are simple randomized data structures that are useful in practice. The BBF is an extension of BF, and inherits the best feature of BF such as time and space saving. The BBF treats strings in a set in a different way depending on their significance, groups the strings into bins and allocates different false positive rate to different bins. Important spam words have lower false positive rate than less significant words. In this work, we have applied CS and ECS for optimization of BF. The proposed system ECS outperforms CS.

References Références Referencias

1. Bloom B, Space/time tradeoffs in hash coding with allowable errors, Communications of the ACM, 13, 1970, 422–426.
2. Feng W,.Shin K.G, Kandlur D.D. & D.Saha, "The BLUE active queue management algorithms", IEEE/ACM Transactions on Networking, 10, 2002, 513 – 528.
3. Mullin J.K, Optimal Semijoins for Distributed Database Systems, IEEE Trans. Software Eng., 16, 1990, 558-560.
4. Mackert L.F. and Lohman G.M., Optimizer Validation and Performance Evaluation for Distributed Queries, Proc. 12th Int'l Conf. Very Large Data Bases (VLDB), 1986, 149-159.
5. Broder A and Mitzenmacher M. Network Applications of Bloom Filters: A Survey, Internet Math., 1(4), 2005, 485-509.
6. Kubiatowicz J Bindel D, Chen, Y Czerwinski S, Eaton P, and Geels D, Oceanstore: An Architecture for Global-Scale Persistent Storage," ACM SIGPLAN Notices, 35(11), 2000, 190-201.
7. Li J, Taylor J, Serban L, and Seltzer M, Self-Organization in Peer-to-Peer System, Proc. ACM SIGOPS, 2002.
8. Cuena-Acuna F.M, Peery C,Martin R.P, and Nguyen T.D, PlantP: Using Gossiping to Build Content Addressable Peer-to-Peer Information Sharing Communities, Proc. 12th IEEE Int'lSymp. High Performance Distributed Computing, 2003, 236-249
9. Rhea S.C and Kubiatowicz J, Probabilistic Location and Routing, Proc. IEEE INFOCOM, 2004, 1248-1257.
10. Hodes T.D, Czerwinski S.E, and Zhao B.Y, An Architecture for Secure Wide Area Service Discovery, Wireless Networks, vol. 8, nos. 2/3, 2002, 213-230.
11. Reynolds P and Vahdat A, Efficient Peer-to-Peer Keyword Searching, Proc. ACM Int'l Middleware Conf., 2003, 21-40.
12. Bauer D, Hurley P, Pletka R, and Waldvogel M, Bringing Efficient Advanced Queries to Distributed Hash Tables, Proc. IEEE Conf. Local Computer Networks, 2004, 6-14
13. Fan L, Cao P, Almeida J, and Broder A, Summary Cache: A Scalable Wide Area Web Cache Sharing Protocol, IEEE/ACM Trans. Networking, 8(3), 2000, 281-293.
14. Peter C.D and Panagiotis M, Bloom Filters in Probabilistic Verification, Proc. Fifth Int'l Conf. Formal Methods in Computer- Aided Design, 2004, 367-381.
15. Jin C, Qian W, and Zhou A, Analysis and Management of Streaming Data: A Survey, J. Software, 15(8), 2004, 1172-1181.
16. Deng F and Rafiei D, "Approximately Detecting Duplicates for Streaming Data Using Stable Bloom Filters," Proc. 25th ACMSIGMOD, 2006, 25-36.
17. Bonomi F, Mitzenmacher M, Panigrahy R, Singh S, andVarghese G, Beyond Bloom Filters: From Approximate Membership Checks to Approximate State Machines, Proc. ACM SIGCOMM, 2006 , 315-326.
18. Li K and Zhong Z, Fast Statistical Spam Filter by

Approximate Classifications, Proc. Joint Int'l Conf. Measurement and Modeling of Computer Systems, SIGMETRICS/Performance, 2006, 347-358.
19. Mitzenmacher M, Compressed Bloom Filters, IEEE/ACM Trans.Networking, 10(5) 2002, 604-612.
20. Kirsch A and Mitzenmacher M, Building a Better Bloom Filter, Technical Report tr-02-05.pdf, Dept. of Computer Science, Harvard Univ,2006.
21. Kirsch A and Mitzenmacher M, Distance-Sensitive Bloom Filters, Proc. Eighth Workshop Algorithm Eng. and Experiments (ALENEX '06), 2006.
22. Kumar A, Xu J, Wang J, Spatschek O, and Li L, Space-Code Bloom Filter for Efficient Per-Flow Traffic Measurement, Proc. 23rd IEEE INFOCOM, 2004, 1762-1773.
23. Cohen S and Matias Y, Spectral Bloom Filters, Proc. 22nd ACM SIGMOD, 2003, 241-252.
24. Laufer R.P, Velloso P.B, and Duarte O.C.M.B, GeneralizedBloom Filters, Technical Report Research Report GTA-05-43, Univ. of California, Los Angeles (UCLA), 2005.
25. Chazelle B, Kilian J, Rubinfeld R, and Tal A, The Bloomier Filter: An Efficient Data Structure for Static Support Lookup Tables, Proc. Fifth Ann. ACM-SIAM Symp. Discrete Algorithms (SODA), 2004, 30-39.
26. Hao F, Kodialam M, and Lakshman T.V, Building High Accuracy Bloom Filters Using Partitioned Hashing, Proc. SIGMETRICS/Performance, 2007, 277-287.
27. Yang X.S., Deb S. "Cuckoo search via Lévy flights". World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). IEEE Publications. 2009, 210–214.
28. Xie K., Min Y., Zhang D., Wen J., Xie G. & Wen J, Basket Bloom Filters for Membership Queries, Proceedings of IEEE Tencon'05,2005, 1-6.

Global Journals Inc. (US) Guidelines Handbook 2012

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

- 'FARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSC" can be added to name in the following manner. eg. **Dr. John E. Hall, Ph.D., FARSC or William Walldroff Ph. D., M.S., FARSC**
- Being FARSC is a respectful honor. It authenticates your research activities. After becoming FARSC, you can use 'FARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSC will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSC. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSC member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSC will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSC member's bank account after certain threshold balance. There is no time limit for collection. FARSC member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

- 'MARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSC" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSC or William Walldroff Ph. D., M.S., MARSC
- Being MARSC is a respectful honor. It authenticates your research activities. After becoming MARSC, you can use 'MARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSC will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- MARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSC member's bank account after certain threshold balance. There is no time limit for collection. MARSC member can decide its price and we can help in decision.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJMBR for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5.STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be

sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page

- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to

shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic

principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.

- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation)
by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

K

L

M

N

O

P

R

S

T

U

V

W

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350