IITM Journal of Information Technology

Annual Journal of Institute of Innovation in Technology & Management

Volume 2 **January-December, 2016**

CONTENTS

Research Papers & Articles

BLANK

Performance Parameters to Improve Efficiency of Wireless Sensor Network

Vishakha Sehdev*
Ankit Verma**

Abstract

In the current time and next decades, Wireless Sensor Networks (WSNs) represents a new category of ad hoc networks consisting of small nodes with three functions: sensing, computation, and wireless communications capabilities. Many routing, power management, and data dissemination protocols have been designed for WSNs where energy awareness is an essential design issue to improve the overall performance of WSN. There are many approaches and techniques explored for the optimization of energy usage in wireless sensor networks. Routing represents one of these areas in which attempts for efficient utilization of energy have been made. In this paper, we report on the current state of the research on optimizing the performance of WSN using various advanced approaches. There are various directions to enhance and optimize the performance as: avoiding congestion and keep it within certain controlled value, selecting the optimum routing approach, reducing the level of power consumption to increase the life time of the sensor node and others. So, the major objective of this paper is to investigate the various techniques used in improving and enhancing the performance of WSN to let it be more reliable in various applications like: health care and biomedical treatment, environment monitoring, military survival lance, target tracking and greenhouse monitoring.

Key Words: wireless sensor networks, performance parameters, WSN protocols

I. Introduction

The wireless sensor network is some type of an ad-hoc network. Mainly it consists of small light weighted wireless nodes called sensor nodes, deployed in physical or environmental condition. It measure the physical parameters such as sound, pressure, temperature, and humidity. These sensor nodes deployed in large or thousand numbers and collaborate to form an ad hoc network capable of reporting to data collection sink (base station). Wireless sensor network have various applications like habitat monitoring, building monitoring, health monitoring, military survival lance and target tracking. However wireless sensor network is a resource constraint if we talk about energy, computation, memory and limited communication capabilities. A typical wireless sensor network is comprised of tens, hundreds, or even thousands of sensor nodes. Typically each sensor node is composed of a microcontroller, a radio transceiver, one or more micro sensors, power source and other components. The microcontroller samples the micro sensors, send the data, either with or without processing, through radio links to the locations where the information is needed. Due to the limited radio range and the relatively larger target areas, in many cases a multiple hop ad hoc wireless network is formed for the information transmission. The devices that gather the information from the wireless sensor networks are defined as base stations. There may be one or more base stations for a wireless sensor network. The base stations may be static or mobile. However, for many

Vishakha Sehdev*
Department of Computer Science & Engineering
GITM, Gurgaon, Haryana

Ankit Verma**
Assistant Professor,
Department of IT, IITM, Janakpuri, New Delhi

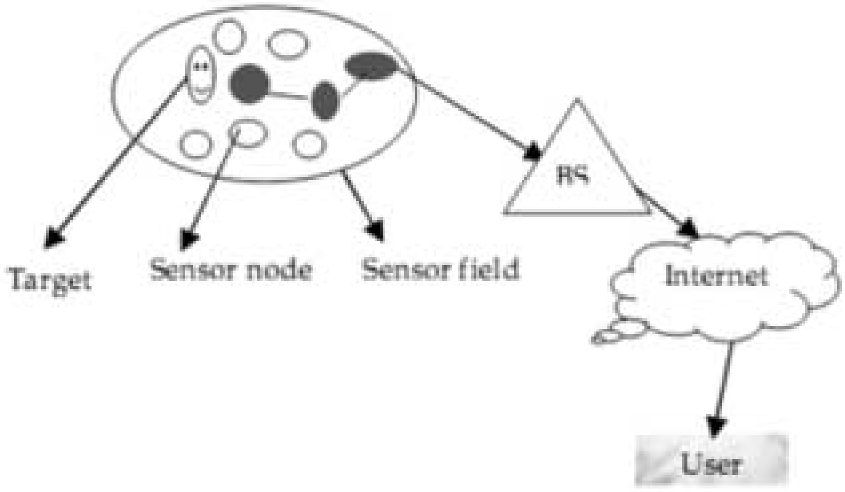


Fig. 1: Architecture of the Sensor network

applications, the sensor nodes themselves are not moving, either due to the scenario requirements, or technical or economical hindrance. All sensor nodes in the wireless sensor network are interacting with each other or by intermediate sensor nodes [1]. A sensor nodes that generates data, based on its sensing mechanisms observation and transmit sensed data packet to the base station (sink). This process basically direct transmission since the base station may locate very far away from sensor nodes needs (see Fig. 1). More energy to transmit data over long distances so that a better technique is to have fewer nodes sends data to the base station. These nodes called aggregator nodes and processes called data aggregation in wireless sensor network.

In a wireless sensor network, sensing nodes with limited power, computation, communication and storage resources cooperate to fulfill monitoring and tracking functionalities. Compared with conventional sensors, wireless sensor networks have the following advantages: Since the relevant technologies have become technologically and economically feasible, people want to gather much more information from more places in the physical world, which was either impossible due to technological difficulties or formidable due to high cost, in terms of money and human power Decreasing form factors and costs of micro sensors make deployment of hundreds even thousands of sensors much more feasible than conventional sensors that are too cumbersome and expensive [2]. Many conventional sensors send data to data acquisition (DAQ) modules in personal computers or workstations, and typically the sensors, personal computers and workstations are interconnected in a wired manner for data collection, aggregation and processing. When the number of sensors deployed in a certain area increases exponentially, this approach obsoletes. A more appropriate approach is to connect the sensors with low cost microcontrollers and to send and receive information through wireless links. This approach is made feasible by decrease in costs, form factors and power consumption of microcontrollers and radio transceivers. Based on the sensing ranges of quite a few commonly used types of sensors, a deployment density of one node per one hundred square meters is typical [2]. One of the design optimization strategies applied in WSN is to deterministically place the sensor nodes in order to meet the desired performance goals. In such case, the coverage of the monitored region can be ensured through careful planning of node densities and fields of view and thus the network topology can be established at setup time. However, in many WSNs applications sensors deployment is random and little control can be exerted in order to ensure coverage and yield uniform node density while achieving strongly connected network topology. Therefore, controlled placement is often pursued for only a selected subset of the employed nodes with the goal of structuring the network topology in a way that achieves the desired application requirements. In addition to coverage, the nodes positions affect numerous network performance metrics such as energy consumption, delay and throughput. For example, large distances between nodes weaken the communication links, lower the throughput and increase energy consumption.

II. Wireless Sensor Network

This section review the prior work on improving the congestion control over wireless sensor networks as one approach for enhancing the performance of WSN. In [3], a cross-layer TDMA-based protocol that guarantees collision-free communication by scheduling slots for each node and results in significant energy savings was presented. This technique has the capability of determining the collision-free slots that are to be assigned to wireless nodes in a multiple-hop network. In [4], another approach was proposed in which TRAMA that organizes time into frames and uses a distributed election scheme based on traffic information at each node to determine which node can transmit at a particular slot. TRAMA uses a distributed hash function to determine a collision-free slot assignment and builds a scheduling scheme when a node has data to send. This random scheduling scheme increases the queuing delays. Another technique called Queue based Congestion Control Protocol with priority support, using the queue length as an indication of congestion degree was presented in [5]. In this approach, the rate assignment to each traffic source is based on its priority index as well as its current congestion degree. A node priority-based congestion control protocol for wireless sensor networks was proposed in [6]. In this technique, the node priority index is introduced to reflect the importance of each node and uses packet inter-arrival time along with packet service time to measure a parameter defined as congestion degree and imposes hop-by-hop control based measurement as well as node priority index. In [7], it was proposed an energy efficient congestion control scheme for sensor networks called Enhanced Congestion Detection and Avoidance which comprises of three mechanisms. First, the approach uses buffer and weighted buffer difference for congestion detection. Secondly, proposed a bottleneck-node-based source data sending rate control scheme and finally uses a flexible queue scheduler for packets transfer. A new and more recent approach was proposed in [8] called a cluster head method to allow parallel transmission of data packets to form a schedule by arranging data transfer at each round. The cluster head accepts request for data transfer and assigns a slot for each node wishing to transmit. Each node of data transfer is divided into contention, data transmission and idle period. In WSN the single point of failure is eliminated by providing a decentralized control and nodes that have no data to send waste time slots in the contention period where idle listening and overhearing occurs. In [9] a suggested approach called an adaptive rate control for congestion avoidance in WBANs was presented. The scheme performs rate control dynamically each node based on a predication model which uses rate function including congestion risk degree and valuation function, without requiring congestion detection and congestion notification steps. There is another advanced approach presented in [10] based on a distributed and scalable algorithm that eliminates congestion within a sensor network, and ensures the fair delivery of packets to a central node or a base station. This routing structures often results in the sensors closer to the base station experiencing congestion, which inevitably cause packets originating from sensors to have a higher probability of being dropped. The problem of single path upstream congestion control in wireless sensor networks through the traffic control was investigated in [11], where authors of this work proposed a multi-agent system based approach to control the traffic in the upstream congestion. The traffic generated in a wireless sensor node is of two types named, source traffic and transit traffic. The source traffic is generated from each wireless sensor node and the transit traffic is generated from other wireless sensor nodes. A Reusable Task-based System of Intelligent Networked Agents (RETSINA) is a cooperative multi-agent system that consists of three classes of agents: interface agents, task agents and information agents. RETSINA provides a

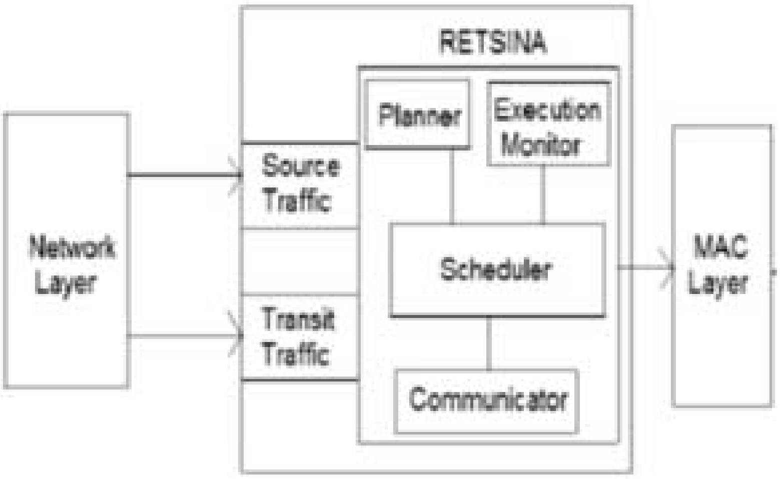


Fig. 2: Congestion control model

domain-independent, componentized, and reusable substratum to (a) allow heterogeneous agents to coordinate in a variety of ways and (b) enable a single agent to be part of a multi-agent infrastructure. RETSINA [12] provides facilities for reuse and a combination of different existing low-level infrastructure components, and it also defines and implements higher level agent services and components that are reconfigurable and reusable. Paper [11] proposed an upstream congestion control model by using RETSINA multiagent named Agent-based Congestion Control Protocol (ACCP). ACCP reduce the packet loss by its intelligent scheduling schemes. Fig.2 illustrates the proposed congestion control model in a wireless sensor node. ACCP consists of four components: Execution Monitor, Communicator, Planner, and Scheduler [13]. The execution monitor identifies the congestion based on the packet arrival time (ta) and packet service time (ts) at the Medium Access Control (MAC) layer. The packet arrival time (ta) is the time interval between two subsequent packets arrived from any source and the packet service time (ts) is the time interval between arrival of packets at the MAC and its successful transmission. These two parameters are monitored at each node by the execution monitor on a packet-by packet manner.

From this, a congestion index (Cx) is calculated and it is defined as the ratio of average packet service time over average packet arrival time at each wireless sensor node. The congestion index at node i is given [11] by:

$$Cx(i) = ts / ta \quad (1)$$

The execution monitor also takes the agents next intended action and prepares, monitors, and completes its execution. The communicator module communicates all the notifications at each wireless sensor node in the packet header to be forwarded. From the congestion index the communicator module computes a global congestion priority index by summating source congestion priority index and the global congestion priority index of the lower level wireless senor nodes. The planner receives goals through communication message packets and finds alternative ways to fulfill them. The planning component is reusable and capable of accepting different planning algorithms in an intelligent way. The scheduler has two queues for the source traffic and the transit traffic. By adjusting the scheduling rate the congestion can be reduced. The scheduling algorithm uses the earliest deadline-first heuristic. A list of all actions is scheduled and the action with the earliest deadline is chosen for execution. When a periodic action is chosen for execution, it is reinstated into the schedule with a deadline equal to the current time plus the actions period. The four modules of RETSINA multi-agent are implemented for the upstream congestion control as autonomous threads of control to allow concurrent planning and scheduling actions, and execution in an efficient way. Furthermore, all modules are executed as separate threads and are able to execute concurrently. So almost all the packets are forwarded to the next wireless sensor node without any loses.

III. Performance Parameters of WSN

A typical sensor network operates in five phases: the planning phase, deployment phase, post-deployment phase, operation phase and post-operation phase. In the planning phase, a site survey is conducted to evaluate deployment environment and conditions, and then to select a suitable deployment mechanism. In the deployment phase, sensors are randomly deployed over a target region. In the post deployment phase, the sensor networks operators need to identify or estimate the location of sensors and to access coverage. The operation phase involves the normal operation of monitoring tasks where sensors observe the environment and generate data. The post-operation phase involves shutting down and preserving the sensors by settings the sensors to sleep mode for future operations or destroying the sensor network. In a WSN setup, the nodes may be deployed in an ad-hoc manner with no predefined topology. The nodes automatically setup a network by communicating with one another in a multi hop fashion. New nodes can malfunction, be added or removed from the network at any time. Newly added nodes must integrate into the network seamlessly and the network must detect and react quickly when nodes are removed to avoid affecting the reliability of message delivery services. The timely detection, processing, and delivery of information are indispensable requirements in a real-time WSN application. In SPEED there are two types of communication associated with data delivery.

Unicast (a specific node will receive the packet) area-multicast (where a copy of the packet is send to every node inside the specified area)

Area-unicast (copy of the packet is sent to at least one node inside the specified area) for efficient communication both the route discovery cost and resulting route length are important. Unlike wired networks, where the delay is independent of the route length, in multi hop wireless sensor networks, the end-to-end delay depends on not only single hop delay, but also on the distance a packet travels. Any real time protocol should satisfy three design objectives: stateless nodes, load balanced routes and congestion control mechanism. The architectures of WSNs emerged from the experience gained from devising architectures for self-organizing, mobile, ad hoc networks. The latter show emphasis on the need for decentralized, distributed form of organization and this is a shared characteristic with WSNs. They benefit from the evolutions in real-time computing, peer-to-peer computing, active networks and mobile agents/swarm intelligence. Besides the networking and computing concepts just mentioned, many other factors play a significant role when devising architectures for a WSN. The critical factors that distinguished between different WSNs architectures are listed as follows:

- **Fault tolerance:** WSNs are mainly monitoring important phenomena. Therefore, it is essential for a WSN to sustain its functionality without disruptions, even if some nodes malfunction or die. Usually, WSNs are deployed in hostile environments where nodes may be damaged, due to environmental interference, or eventually die due to the impracticality of recharging or replacing their batteries. Nodes in a WSN are prone to failures and this may result in sever situations like partitioning the network. The design of a WSN should guarantee that its functionality and services are never degraded by these failures.

- **Scalability:** Sensor nodes are deployed densely to form a WSN. This huge number of nodes has a direct impact on the design of schemes and protocols at different layers. For example, a MAC protocol (data-link layer) should be able to grant, in a fair fashion, each node access to the medium while minimizing or preventing collisions, which is very difficult given the huge number of available nodes. Also, a routing protocol (network layer) that depends on exchanging routing tables among nodes may not be efficient since there will be excessive control traffic that underutilizes the bandwidth of the medium.

- **Production cost:** The cost of a single sensor node should be minimized since it determines the overall cost of the network under design.
- **Network topology:** The fact that Wireless Sensor Networks are constituted by a huge number of nodes raises the challenge of network topology maintenance and modification. The challenge occurs starting at the early stage of nodes deployment. Sensor nodes can be either thrown in a mass (e.g., from a plane) or manually placed one by one (e.g., by a human or a robot) in the field. Also, after nodes deployment, topology may change due to failures in some nodes, changes in nodes locations, lack of reachability (due to jamming for instance), and huge reductions in power resources at some nodes (which affect their transmission power levels to the limit that they vanish from the vicinity of neighboring nodes). The WSN should be able to adapt to these sudden changes to avoid any degradations in its functionality.
- **Security:** In the environment of deployment, sensor nodes are either deployed very close to the phenomenon of interest or directly inside it. As a result, we can see that WSNs are usually not supervised (especially in remote geographic areas). This means that WSNs may be targeted by intruders to exploit any security vulnerability.
- **QoS support:** Time-sensitive applications (especially in military) require support for real-time communication that provisions guarantees on maximum delay, minimum bandwidth, or other QoS parameters.
- **Power consumption:** this is a primary design factor for any WSN. Power consumption should be made minimal in order to prolong the lifetime of the network. In fact, “power conservation” is a distinguishing factor between designing a WSN and designing other classes of wireless networks. The latter may consider QoS parameters (like, delay, throughput, fairness, etc.) as key design requirements. Based on this observation, research activities target the development of power-aware protocols and algorithms for sensor networks. That is, power-awareness should be incorporated in every stage of designing a WSN. In fact, power-awareness imposes constraints on the size and complexity of a sensor nodes platform. In this context, hardware of sensor nodes should be designed to be power-efficient.

IV. Conclusion

In the current time, there is a new era of ubiquitous computing. One type of such ubiquitous is wireless sensor technologies which is characterized with a great potential in opening a world of sensing applications. This paper provides the different approaches used in enhancing the performance of such WSN with great focus on three important factors: congestion control, optimal routing approaches and reducing the consumption power. Consumption power was touched in deep since Wireless sensor networks are battery powered, therefore prolonging the network lifetime through a power aware node organization is highly desirable. An efficient method for energy saving is to schedule the sensor node activity such that every sensor alternates between sleep and active state. One solution is to organize the sensor nodes in disjoint covers, such that every cover completely monitors all the targets. These covers are activated in turn, in a round-robin fashion, such that at a specific time only one sensor set is responsible for sensing the targets, while all other sensors are in a low-energy, sleep state.

References

1. Heinzelman W B, Chandrakasan A P, Balakrishnan H. An application-speciûc protocol architecture for wireless micro- sensor networks. IEEE Trans Wirel Commun, 2002, 1: 660–670.
2. Ben-Othman J, Yahya B. Energy eûcient and QoS based routing protocol for wireless sensor networks. J Parallel Distrib Comput, 2010, 70: 849–857.
3. Lin F Y S, Yen H H, Lin S P. Delay QoS and MAC aware energy-eûcient data-aggregation routing in wireless sensor networks. Sensors, 2009, 9: 7711–7732.

4. Wang H P, Zhang X B, Na¨ýt-Abdesselam F, et al. Cross-layer optimized MAC to support multihop QoS routing for wireless sensor networks. IEEE Trans Veh Technol, 2010, 59: 2556–2563.
5. R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," IEEE Transactions on Evolutionary Computation, vol. 6, no. 4, pp. 321–332, 2002.
6. K. M. Sim and W. H. Sun, "Ant colony optimization for routing and load-balancing: Survey and new directions," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2003.
7. V. A. Cicirello and S. F. Smith, "Ant colony control for autonomous decentralized shop ûoor routing," in Proceedings of the International Symposium on Autonomous Decentralized Systems. IEEE Computer Society Press, 2001, pp. 383–390.
8. Kamat, P.; Zhang, Y.; Trappe, W.; Ozturk, C. Enhancing Source-Location Privacy in Sensor Network Routing. In Proceedings of the 25th IEEE Int. Conference on Distributed Computing Systems (ICDCS05), 2005; pp. 599-608.
9. Al-Karaki, J.N.; Kamal, A.E. Routing Techniques in Wireless Sensor Networks: a Survey. IEEE Wirel. Commun. 2004, 11, 6-28.
10. Iyengar, S.; Wu, H.-C.; Balakrishnan, N.; Chang, S.Y. Biologically Inspired Cooperative Routing for Wireless Mobile Sensor Networks. IEEE Syst. J. 2007, 1, 29–37.
11. Niannian D.; Liu, P.X.; Chao H. Data Gathering Communication in Wireless Sensor Networks Using Ant Colony Optimization. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005; pp. 697–702.
12. Al-Karaki JN, Kamal AE (2004). "Routing Techniques in Wireless Sensor Networks: a Survey". Wireless Communications, IEEE, 11: 628.
13. Ding N, Liu XP (2005). A Centralized Approach to Energy-Efficient Protocols for Wireless Sensor Networks, IEEE International Conference on Mechatronics and Automation, Niagara Falls, Ont., Canada, 3: 1636-1641.

A Study on Linux Live CD for Performance Enhancement

Ajay Kumar Phogat*

Abstract

CD-boot Linux is live Linux environment, which is easy to use because it is not installed in hard disk, but simply boots directly from a CD. This helps in protecting the sensitive information because clean environment can be prepared at boot time. To insure this environment protects sensitive information, we adapted the trusted Computing Technology to define trustworthy environment. Generally when we boot a system, the files are stored on a hard disk and then goes to RAM for further execution but in live CD or live distro, the files are not stored on hard disk and directly goes to RAM for execution. A Live CD is a bootable Compact Disk (CD) that runs a standalone operating system (O/S). Generally, Live CDs utilize an Open Source operating system such as Linux. An Individual Live CD may run either from the CD or from RAM. It is the way of Linux distribution. Any PC user can create a CD that runs win XP without being installed, and use it to troubleshoot and recover crashed machines. Live CDs let you turn any computer into a temporary Linux box. In this, the basic principle is to reduce the time needed to fix a hard- ware/software failure. Moreover the aim is to show the use of a bootable (live) device-here, a CD, can- improve both the reliability and the efficiency of the system.

Key Words: Live CD, PUPPY, NTFS, FAT, Unionfs.

I. Introduction

The term "live" derives from the fact that these CDs each contain a complete, functioning and operational operating system on the distribution medium.[1] A live CD or live distro is a computer operating system that is executed upon boot, without installation to a hard disk drive. A live CD contains an OS configured to run directly off the CD, i.e., it need not be installed on the system. Just pop the CD into the system's drive, boot from it and the OS will be up and running, without tampering with the OS loaded on the system's hard drive. Typically, the live distro is named after the bootable medium it is stored on, such as a CD-ROM or DVD (live CD/DVD) or a USB flash drive (live USB). A live distro does not alter the operating system or files already installed on the computer hard drive unless instructed to do so. Live distro often include mechanisms and utilities for more permanent installation, including disk partitioning tools. The default option, however, is to allow the user to return the computer to its previous state when the live distro is ejected and the computer is rebooted. It is able to run without permanent installation by placing the files that typically are stored on a hard drive into RAM, typically in a RAM disk.[1] However, this does cut down on the RAM available to applications, reducing performance somewhat. As of 2007, certain live distro run a graphical user interface in as little as 32MB RAM. In some live distro, the user can optionally install the OS from the removable media to the hard disk drive (they are called installable live distro).The first Linux-based live CD was Yggdrasil Linux in 1995, though in practice it did not function well due to the low throughput of then-current CD-ROM drives. The Debian-derived Linux distribution Knoppix was released in 2003, and found popularity as both a rescue disk system and as a primary distribution in its own right. Knoppix, or KNOPPIX, is a Linux operating

Ajay Kumar Phogat*
Assistant Professor, Department of IT
IITM, Janakpuri, New Delhi

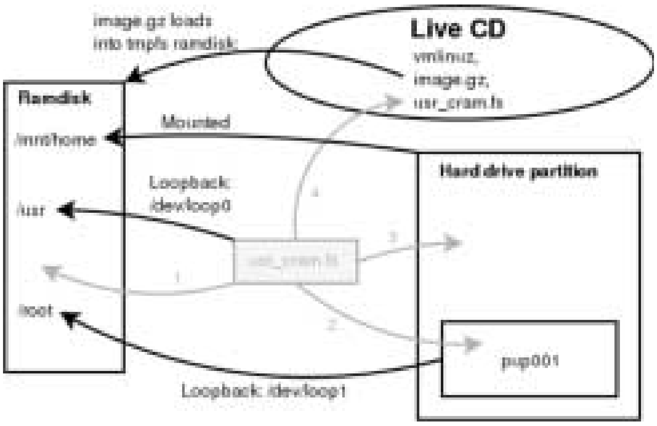


Fig. 1. How Puppy Linux Works

system based on Debian designed to be run directly from a CD / DVD. Even though Live CDs don't require a hard drive, they still enable students to work with Linux utilities, install applications, save data, and save configurations. While students may save to a hard drive, they may also utilize other storage devices[2]. For example, Puppy Linux and the Knoppix Live CD will automatically recognize, mount, and utilize a USB memory stick. These capabilities simultaneously empower students and minimize support requirements. Certain live CDs run a graphical user interface in as little as 32MB RAM.

II. Origin of Live CD

Although early Linux developers and users were able to take advantage of cheap optical disks and rapidly declining prices of CD drives for personal computers, the Linux distribution CDs or "distros" were generally treated as a collection of installation packages that must first be permanently installed to hard disks on the target machine.[2][3] However in the case of Linux, the free operating system was meeting resistance in the consumer market because of the perceived difficulty, effort, and risk involved in installing an additional partition on the hard disk, particularly the ext2 file system. The term "live CD" was coined because after typical PC RAM was large enough and 52x speed CD drives and CD burners were widespread among PC owners, it finally became convenient and practical to boot the kernel, run X11, a window manager and GUI applications directly from a CD without disturbing the OS (generally Windows on FAT32 or NTFS) on the hard disk. This was a new and different situation for Linux than other OSes, because the updates/

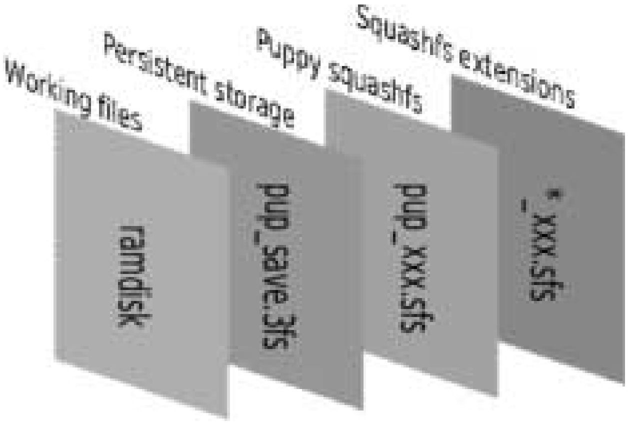


Fig.2. Layer architecture of unionfs

Table 1: Description of each layer

ram disk	This is the tmpfs file system running in RAM, with new and changed files.
pup_save.3fs	This is the persistent storage, where all your data, settings, email, installed packages, etc., get saved permanently. The ".3fs" means that the file contains a ext3 file system.
pup_xxx.sfs	This is Puppy. The built in applications, window manager, scripts, everything. The ".sfs" means the file contains a squashfs compressed file system. The "xxx" is the Puppy version number without the dots, for example "200".
*_xxx.sfs	These are additional squashfs files. The "*" can be anything. For example, devx_xxx.sfs is the complete environment for compiling C/C++ applications.

upgrades were being released so quickly, different distributions and versions were being offered online, and especially because users were burning their own CDs.

III. Live CD Initialization

During **live CD** initialization, a user typically may resort to using one or more cheat codes to change the booting behavior. These vary from distribution to distribution but can most often be accessed upon first boot screen by one of the function keys.

Puppy is revolutionary. From a user's point of view, it is still the same Puppy, apparently unchanged. From the developer's point of view, the basic ideas, such as running in RAM, are still there — just some implementation details have changed. Puppy is actually simpler.[4] The startup and shutdown scripts are simpler. Options required in the isolinux.cfg or syslinux.cfg files are simpler.

Puppy2 is better for several reasons:

Works with any size Flash drive (minimum 128M).

Saves RAM disk (your working files) to Flash drive every 30 minutes, so extending lifetime of flash media (by restricting the number of writes).

Works on PCs with very little RAM, probably 32MB and boot very fast.

The entire file system, that is, "/", is writable and is saved.

Much simpler structure.

Simplified boot params. Ex: "PMEDIA=USB flash" is all that is needed to boot from USB pen drive.

"image.gz size problem" eliminated.

Iso file can now grow from 60-65M to 65-70M.

Improved security.

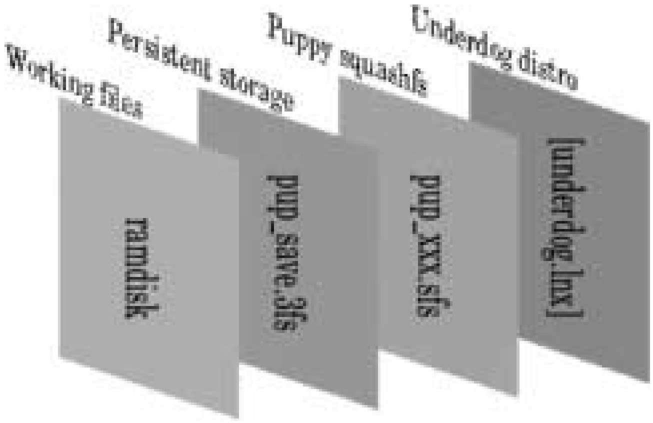


Fig.3. Linux distro as the bottom layer

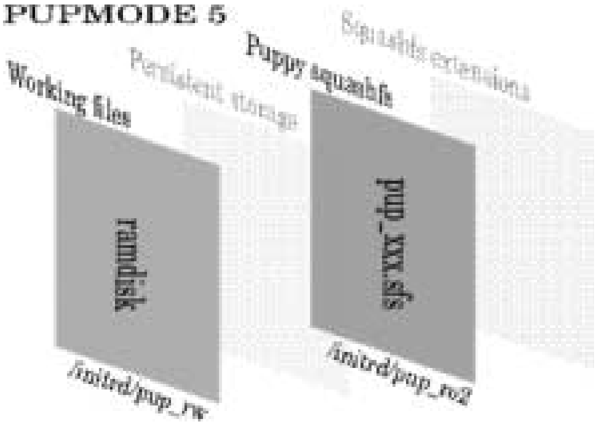


Fig.4. Pupmode 5 Layers

Simplified, more reliable multisession CD/DVD management.

One iso for normal and multisession.

In summary, puppy2 is faster at boot up and shutdown, works better on older hardware, more flexible and expandable in the future, more standardized, and considerably simpler.

IV. Architecture Overview

The diagram is to view each of those layers as a complete file system, that is, a complete directory hierarchy from "/" down. These layers are laid one on top of the other, which is achieved by the unionfs file system. The way it works is very simple: say that the "off-pink" layer has a file /usr/lib/libgdkxft.so. This file will be visible at the top layer. If the "off-blue" layer has the same file, it will not be visible, as it is overlaid by the same file on a higher layer. Anyway, here is a description of each layer:

While running Puppy, all you see is one file system, which is the top layer. Thus we see /usr/lib/libgdkxft.so and we don't care what layer it is actually on.

An exciting alternative to the squashfs extensions is to use an existing installed Linux distro as the bottom layer:

What the above diagram is intended to convey is that the bottom layer is a partition, not the "underdog.lnx" file itself. File underdog.lnx is just a text file, containing the name of a partition, for example "hda1". At boot

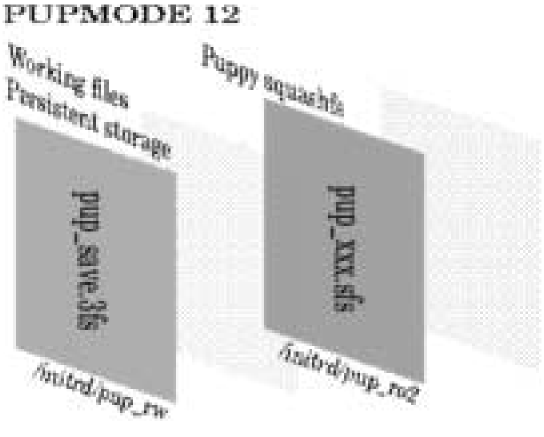


Fig.6. PUPMODE 12 Layers

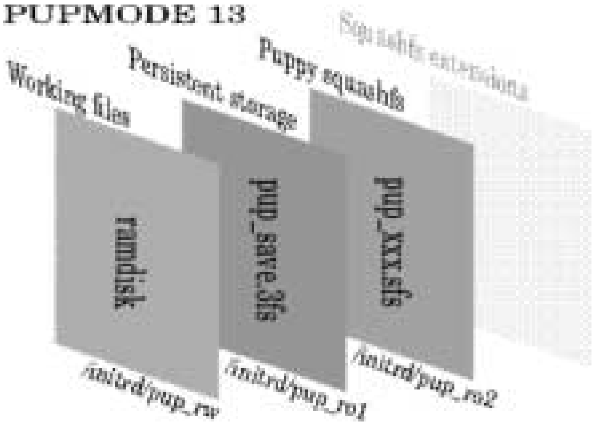


Fig. 7. PUPMODE 13 Layers

up Puppy will read underdog.lnx and will mount the partition as the bottom layer. If that partition happens to have a Linux distro installed in it, then the entire distro file system will "show through" on the top layer of Puppy's unionfs.[5]

It will look like a normal Puppy, running JWM window manager or whatever, same desktop, but everything in the underlying distro is available to be executed. All the applications, compile environment, package manager, etc.

There are other variations, and Puppy has a "state variable" named PUPMODE that shows what state (configuration of layers) Puppy is currently using. There is a file, /etc/rc.d/PUPSTATE that has the PUPMODE variable defined in it, for example, "12", and this is the current layer-configuration. Each PUPMODE value needs its own description.

Pupmode 5

This is the configuration that Puppy is booted from live-CD or USB Flash drive.

The first time that we plug in the live-CD and boot up, there is no persistent storage, and the "union" consists of only two layers, the top "working files" and the pup_xxx.sfs squashfs file system that has all the Puppy files. These two layers appear overlaid at "/", however they can be viewed individually, at their respective mount points. The tmpfs ram disk is mounted read-write on /initrd/pup_rw and pup_xxx.sfs is mounted read-only on /initrd/pup_ro2.

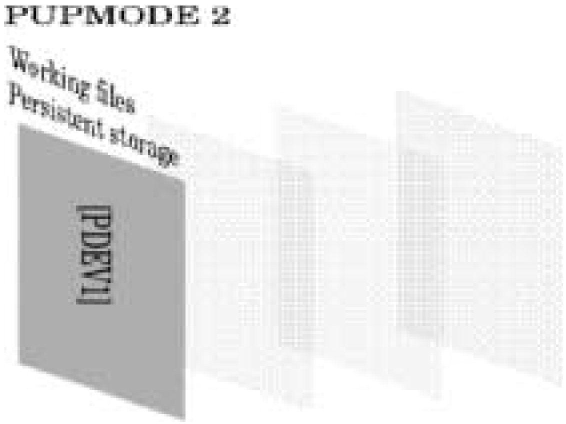


Fig. 8. PUPMODE 2 Layers

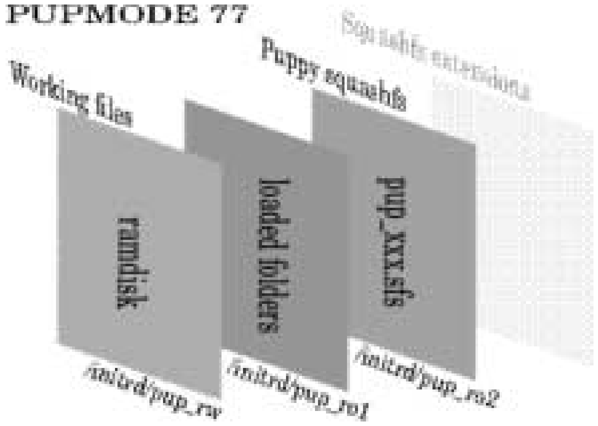


Fig. 9 PUPMODE 77 Layers

So, you are using Puppy but not touching the hard drive at all. You can run applications, configure, download, install packages, but it is all happening in the tmpfs ram disk, so not getting saved. The amount of space you have in the ram disk depends on how much RAM is in the PC (and whether there is a Linux swap partition).

When the shutdown script, which is actually /etc/rc.d/ rc.shutdown, will execute and will bring up a dialog window asking you where you want to save the session to. That is, whatever directories and files that have been created in the ram disk can now be saved.

If this is a live-CD that you have booted off, you would choose to save to a hard drive partition normally, but you could also choose a floppy drive, a USB drive, or even back to the CD/DVD (as long as it was burned "open" or "multisession") then, a GUI dialog window offers a list of partition.

Not only does the shutdown dialog offer a wide range of devices to save to, but it also gives a choice of saving to a file or to a partition. That is, you can have a file named pup_save.3fs with a ext3 file system inside it, nominally 512M in size (actually, the size is also selectable), or Puppy can take over an entire partition in which to save. The choice of file or partition depends on whether a partition is a Linux file system — if only a FAT file system, and then only a pup_save.3fs file can be created.

Pupmode 12

The second time that you boot your Puppy live-CD, assuming that you chose to save the previous session to a pup_save.3fs file in a Linux, FAT or NTFS partition, Puppy will be in PUPMODE 12.

At boot up Puppy found the pup_save.3fs file is on a fast hard drive partition, so decided to mount it directly on the top layer. Thus, there is no tmpfs ram disk intermediary. File pup_save.3fs is directly read and written to.

This scenario is very good for RAM-challenged PCs. The initial RAM disk, that is, the file initrd.gz that gets loaded first when Puppy boots, is still there in ram, mounted at /initrd, however it only uses about 1.9M.

Installing Puppy to a USB Flash drive

Pupmode 13

If you install Puppy to a USB Flash drive, perhaps by using the Puppy Universal Installer program, or manually, you will have a bootable drive with the files vmlinuz (the Linux kernel), initrd.gz (the initial ram disk), pup_xxx.sfs (squashfs file system with all the Puppy files) and syslinux.cfg (Syslinux config file). The situation is just like booting from a live-CD — on first boot Puppy will be in PUPMODE 5, as no persistent storage has yet been created. On first shutdown, as described in the PUPMODE 5 section above, you will create a persistent storage — either a pup_save.3fs file or an entire partition.

In the above diagram, the top layer is a tmpfs ramdisk, into which all new and modified directories go. It is the working area, and has the potential limitation of the amount of RAM available. But of course, if the hard drive has a Linux swap partition that will be used to increase the effective size of the ramdisk.

In the case of the persistent storage on Flash memory, this is the second layer (off-orange color), Puppy will save everything from the top layer to the second layer every 30 minutes. From the "unionfs" point of view, the second layer is mounted read-only; it is only the top layer that is written to; however Puppy is able to "flush" the top layer down to the next layer at periodic intervals. Save everything down to the persistent storage layer, and then the top layer would be empty thus giving you a completely empty ram disk. However, when I tried to do that, unionfs crashed. So, the compromise is that the contents of the top layer are copied down to the next layer — that is officially supported by unionfs.[5] It does mean however, that the ramdisk never gets really flushed, so if you download lots of stuff, or install a big package, you can fill it up. To get around this problem, Puppy has a background program (a daemon) that will warn if RAM space is getting low. Perchance you do run low on RAM, the cure is simple: reboot.

It is the persistent storage, but in the PUPMODE 5, at the first shutdown you are offered to create persistent storage in either a file (pup_save.3fs) or a partition. In the latter case, the session can be saved to an entire partition, but only if it is a Linux partition. In that case, what gets mounted on the second layer (off-orange layer) will be a partition, not a file.

In the case of booting off a USB Flash device, to use the entire partition for personal storage is real nice. The pup_save.3fs file is a limited size; nominally 512M or smaller if the partition does not have enough room. There are also reports on the Puppy forum that it can only be increased to 750M - 1G. ..And is efficient for most USB Flash drives! If we save sessions directly to the USB partition, it has to be a Linux partition, that is, have a ext2, ext3 or reiserfs file system. Whereas, saving to a pup_save.3fs file, the USB partition can be anything — its factory setting is a FAT16 file system, which is fine.

Pupmode 2

Many Puppy users will know that there are two ways to install Puppy to hard drive: what we called "option1" or "option2" installations.

Option1 is the less-invasive choice, which just copies the files vmlinuz, initrd.gz and pup_xxx.sfs to a chosen partition. The Puppy Universal Installer offers to create a floppy boot disk to boot Puppy installed in this way (but currently the boot floppy will only boot Puppy installed on a FAT partition). It is also possible to configure Grub or Lilo to boot Puppy. The advantage of this option is that it doesn't impact on the partition in any way and whatever is already in the partition, like indows or another Linux distro, is unaffected.

For an option1 h.d. installation, Puppy will boot up in PUPMODE 12.

Option2 is to install Puppy to an entire partition, which has to be a Linux partition (ext2, ext3, or reiserfs). This is recommended for developers or anyone wanting to compile applications, as it gives you lots of room.

This is the absolute simplest configuration. There is no ram disk; the partition itself is mounted directly on the top "layer".

PDEV1 is actually a variable in the file /etc/rc.d/PUPSTATE that contains the name of a partition for example "hda1", which is the partition to mount. The partition is mounted directly on "/".

How to know what PDEV1 is at boot up? Saying it is in /etc/rc.d/PUPSTATE in the partition it is putting the egg before the chicken. That file is really just for scripts, so that they know what Puppy has booted off.

The way Puppy is booted in this case is from a floppy disk or USB Flash drive "boot disk", or via Grub or Lilo. The Puppy Universal Installer allows you to create any or all of these three: floppy, USB, Grub.

Pupmode 77

This is for the multisession CD or DVD. In this case, the persistent storage is composed of folders on the CD/DVD. At each shutdown, the session is saved to a folder, so the CD/DVD accumulates folders. At boot up, these folders are read in reverse order and loaded into ramdisk.

At boot up, the folders are read from the CD/DVD and loaded into the second layer (off-orange). During a session, new and changed directories and files will be written to the top layer (off-green). At shutdown, the top layer is written as a new folder on the CD/DVD.

V. Live CD Architecture

Live CD Architecture is conceptually divided into two parts:

One is devoted to the connection and the control of the physical processes (server side), whereas the other is related to the communication with the user (client side). In this paper, the server side will be taken into account. The main issue in connecting physical processes to the remote laboratory is equipping the PCs that are devoted to the processes with special hardware and software, such as, for instance, DAQs and their drivers.[9] The proposed approach is to use a bootable (live) CD on the server side of a remote laboratory. A bootable CD is a CD-ROM that contains an OS along with other software, which can be run directly from the CD drive on system startup, without installing into permanent memory. Since the bootable CD does not require a hard disk to work, it can be used on PCs without hard disks, or if a hard disk is present, this CD does not alter the data that are stored in the device, unless specifically requested. Using a bootable CD in a remote laboratory requires that all the software (and the OS) that are needed by the server must be stored on the CD. The main motivations for using a bootable CD are reported here.

● *Process update.* Adding a new process to a remote laboratory is a task that requires two main steps. The first is devoted to the physical process and regards hardware, i.e., connecting all hardware devices along with safety mechanisms. The second regards software installation and configuration. For the reasons that were previously explained, the time that is needed by the software installation can be greatly reduced by using a bootable CD (which contains all the needed software) and a configuration file[7] That is stored in a floppy disk or USB pen drive.

● *Software update.* In addition to increasing the number of available processes, improvements on a remote laboratory can lead to the development of new functionalities, fixing of bugs, performance improvements, etc. In this case, one only needs to remaster the CD with the new software version and reboot the servers from CD. The updated version of the laboratory may then work on every process with minimal effort.

● *Failure restoring time.* In a remote laboratory, it is common that some software and/or hardware failures occur in some servers/processes. Regarding process failures (e.g., component breakdowns), it is, of course, impossible to find a standard way to repair them since they strongly depend on the specific process. However, other kinds of hardware failures concerning the PC that is connected to the process can be avoided or easily and quickly fixed by using a bootable CD. For example, in the case of a hard disk failure, the process usually goes offline until a new hard disk is found, and all the software that is needed by the server is reinstalled and properly reconfigured. Since the number of software applications that are usually required, this operation can take much time. Instead, by using a bootable CD that contains all the needed software, it is possible to connect processes to PCs with no hard disks, preventing such kinds of problems. Moreover, failures of other PC components can also be easily solved; in fact, it is only needed to temporarily replace the broken PC with any other (containing a DAQ) and to boot it from CD. Since the software in the CD does not need to read/write the hard disk, the original content of the hard disk is preserved. So, it is possible to use this PC without any problem until a new PC is ready to replace it.

In case of software failures, it is just needed to reboot the system to have the server fully operating. Moreover, the reboot procedure is not dangerous since the working data is written in memory random access memory (RAM) disk and not on permanent devices.

● *Increased reliability.* The reliability of a remote laboratory is increased by using a live CD for several reasons. The lack of a hard disk obviously prevents certain problems, such as, e.g., mechanical failures. Such kinds of failures, which can appear rare, may instead arise quite often, particularly, in view of the fact that such devices work continuously for 24 h a day for several years. One more reason for improved reliability is that the software installation is correct and all needed applications are been installed with the proper version. Finally, since the CD is a read-only storage device, it is free from virus corruption or other hacker attacks.[11]

Notice that the CD does not need to continuously run 24 h a day. In fact, the CD loads the useful data to RAM only during the boot process, after which it stops. From this moment, all the applications run from memory, guaranteeing preservation of the CD drive as well as a fast execution time.e.g, LINUX LIVE CD ROUTER NETWORK DIAGRAM:

Linux Live CD Router allows you to share, firewall, and balance and optimizes your broadband connection. You can use ADSL, Cable Modem, USB 3G Cards, T1, Dial-Up, WiFi.

VI. Conclusion

Live CD shows how such a facility can increase both the reliability and the efficiency. Since a live CD, which contains an OS (e.g., Linux) and other useful software, is a relatively new tool, it is the authors' opinion that such a tool can be used with success also in other frameworks of

The use of a Knoppix-based CD has increased the reliability of the whole system and has greatly reduced the time that is needed to add new processes, allowing maintaining and increase the number of available processes for remote control. Although such changes are not directly visible to users, they are essential in providing a more reliable system whereby practical experiments can be performed.

References

1. Marci Casini, Domenico Prattichizzo and Antonio Vicino,"*Operating Remote laboratories through a bootable device"*, IEEE transaction on Industrial electronics, vol... 54, no. 6, December 2007.
2. Ed Crowley, *"Open Source, live CD based, and security lab design: tutorial presentation"*, in Consortium for Computing Sciences in Colleges, USA, Volume 21, issue 4(April 2006), pages: 278-279.
3. Ed Crowley, *"Developing "Hands-on "Security activities with open source Software& live CDs"*, in Consortium for Computing Sciences in Colleges, USA, Volume 21, issue 4(April 2006), pages: 139-145.
4. Daniel Barlow, *"building your own live CD",* in Specialized Systems Consultants, Inc. Seattle, WA, USA, Volume 2005, issue 132(April 2005), Page: 2.
5. Mick Bauer, *"Paranoid Penguin: Customizing Linux live CDs, part I"*, in Specialized Systems Consultants, Inc. Seattle, WA, USA, Volume 2008, Issue 169, Article no. 12, May 2008.
6. Mick Bauer, *"Paranoid Penguin: Customizing Linux live CDs, part II"*, in Specialized Systems Consultants, Inc. Seattle, WA, USA, Volume 2008, Issue 170, Article no. 8, June 2008.
7. Mick Bauer," *Paranoid Penguin: Customizing Linux live CDs, part III"*, in Specialized Systems Consultants, Inc. Seattle, WA, USA, Volume 2008, July1st, 2008.
8. F. Cohen, *"Bootable CDs"*, Network security, vol...2001, no. 8, pp: 17-19, 2001.
9. Sanjay Majumdar,"*Live CDs"*, PC Quest, Article 12, Aug 2004.
10. K. Knopper, *Knoppix Linux live CD*. [Online], Available: http://www.knoppix.org
11. E-fense, *The Helix live CD Page*. [Online], Available: http://www.e-fense.com/helix/
12. N.Brand, frozentech's *live CD list*. [Online], Available: http://www.livecdlist.com.
13. M. Casini, D. Prattichizzo, and A. Vicino, "The automatic control telelab: A user-friendly interface for distance learning," *IEEE Trans. Educ.*, vol. 46, no. 2, pp. 252–257, May 2003

Introduction to GIS and its Application in Real World

Manjusha Singh*
Sushma Malik**

Abstract

Geographic Information System (GIS) is a computer based information system used to digitally represented and analyze the geographic features present on the earth surface and the events that takes place on it. The meaning of digitally is to convert the analog data into the digital form. A GIS is a system of hardware, software, data, and people for collecting, storing, analyzing and disseminating information about areas on the earth. The handling of spatial data usually involves processes of data acquisition, storage and maintenance, analysis and output. For many years, this has been done using analogue data sources and manual processing. The introduction of modern technologies has led to an increased use of computers and information technology in all aspects of spatial data handling. The software technology used in this domain is Geographic Information Systems (GIS). GIS is being used by various disciplines as tools for spatial data handling in a geographic environment. This article deals with history of GIS and fundamentals of GIS such as elements of GIS; data models; data structures & data base; and elementary spatial analysis.

Key Words: GIS, Vector data, Raster data, GIS tools, Application of GIS

I. Introduction

A **geographic information system** or **geographical information system** (**GIS**) is a system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data. In a general sense, the term describes any information system that integrates stores, edits, analyzes, shares, and displays geographic information. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations [1].

A geographic information system (GIS) is a computer-based tool for mapping and analyzing things that exist and events that happen on earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. Whether setting a new business, finding the best soil for growing bananas, or figuring out the best route for an emergency vehicle, local problems also have a geographical component GIS will give you the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems. Mapmaking and geographic analysis are not new, but GIS performs these tasks better and faster than do the old manual methods. And, before GIS technology, only a few people had the skills necessary to use geographic information to help with decision making and problem solving [2].

II. Components of a GIS:

A working GIS integrates five key components: hardware, software, data, people, and methods.

Manjusha Singh*
JRA (Junior Research Assistant)
NECTAR, IIT campus, New Delhi

Sushma Malik**
Assistant Professor, IT Department
IITM, Janak Puri, New Delhi

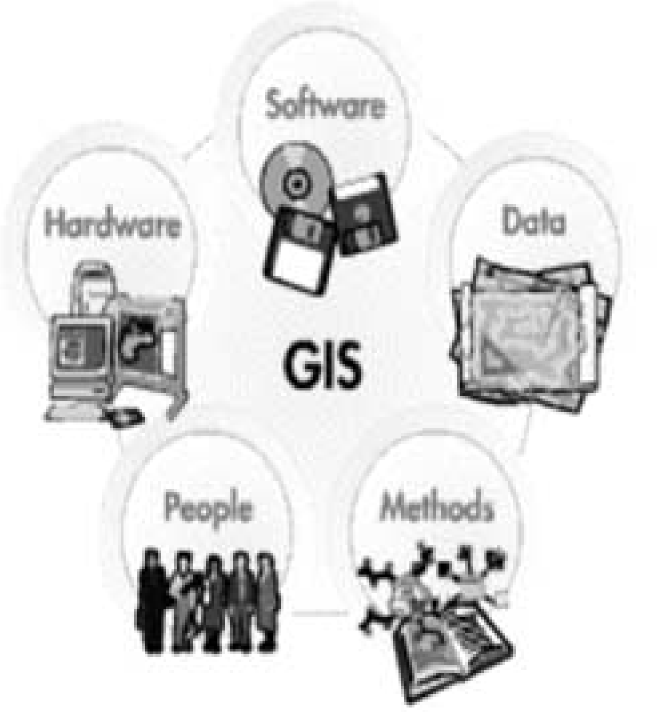


Figure 1: Components of GIS [2]

Hardware

Hardware is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are

- Tools for the input and manipulation of geographic information
- A database management system (DBMS)
- Tools that support geographic query, analysis, and visualization
- A graphical user interface (GUI) for easy access to tools

Data

Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organizations to organize and maintain their data, to manage spatial data.

People

GIS technology is of limited value without the people who manage the system and develop plans for applying it to real-world problems. GIS uses range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work.

Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

III. How GIS Works:

GIS processes information about the world as a collection of thematic layers that can be linked

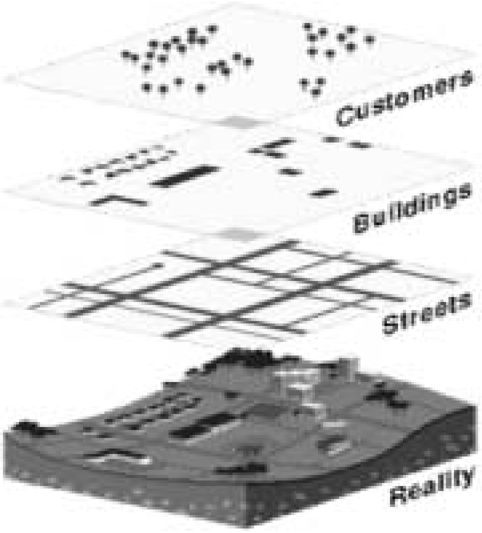


Figure 2: Layers in GIS [2]

together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation.

IV. Vector and Raster Models:

Geographic information systems work with two fundamentally different types of geographic models—the "vector" model and the "raster" model. In the vector model, information about points, lines, and polygons is encoded and stored as a collection of *x,y* coordinates. The location of a point feature, such as a bore hole, can be described by a single *x,y* coordinate. Linear features, such as roads and rivers, can be stored as a collection of point coordinates. Polygonal features, such as sales territories and river catchments, can be stored as a closed loop of coordinates.

The vector model is extremely useful for describing discrete features, but less useful for describing continuously varying features such as soil type or

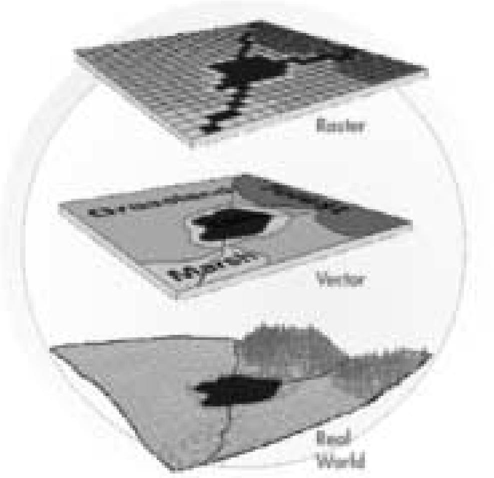


Figure 3: Vector and Raster model of GIS [2]

Table 1: Advantages and Disadvantages of Vector Data [3]

Advantages	Disadvantages
Data can be represented at its original resolution and form without generalization.	The location of each vertex needs to be stored explicitly.
Graphic output is usually more aesthetically pleasing (traditional cartographic representation);	For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.
Since most data e.g. hard copy maps, is in vector from no data conversion is required. Accurate geographic location of data is maintained.	Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.	Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.
	Spatial analysis and filtering within polygons is impossible

Table 2: Advantages And Disadvantages Of Raster Data [3]

Advantages	Disadvantages
The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.	The cell size determines the resolution at which the data is represented.
Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.	It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.	Processing of associated attribute data may be cumbersome if large amounts of data exist. Raster maps inherently reflect only one attribute or characteristic for an area.
Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.	Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization and choice of inappropriate cell size.
Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.	Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

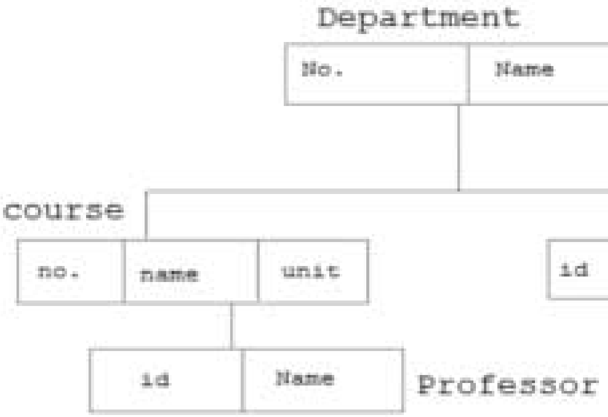


Figure 4: Hierarchical Model

accessibility costs for hospitals. The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture.

Both the vector and raster models for storing geographic data have unique advantages and disadvantages. Modern GISs are able to handle both models.

Data Models:

A separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software, or may be reflected in external commercial Database Management Software (DBMS). A variety of different data models exist for the storage and management of attribute data.

Data models define how the logical structure of a database is modeled. Data Models are fundamental entities to introduce abstraction in a DBMS. Data models define how data is connected to each other and how they are processed and stored inside the system.

The most common are:

1. Tabular
2. Hierarchical
3. Network
4. Relational
5. Object Oriented

The tabular model is the manner in which most early GIS software packages stored their attribute data. The next three models are those most commonly implemented in database management systems (DBMS). The object oriented is newer but rapidly gaining in popularity for some applications.

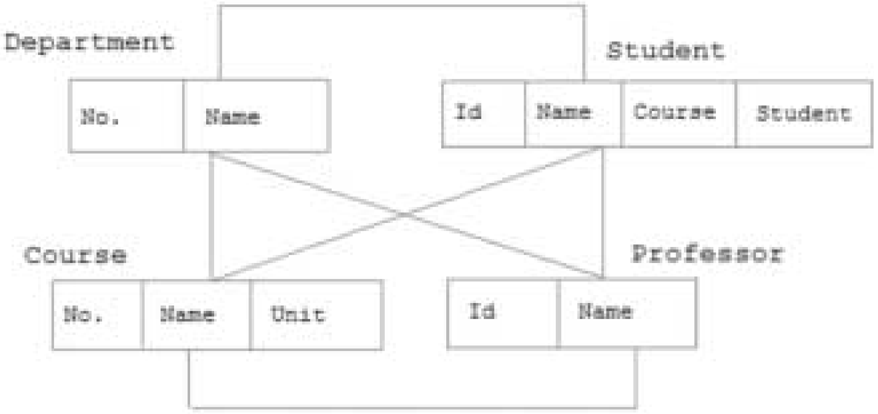


Figure 5: Network Model

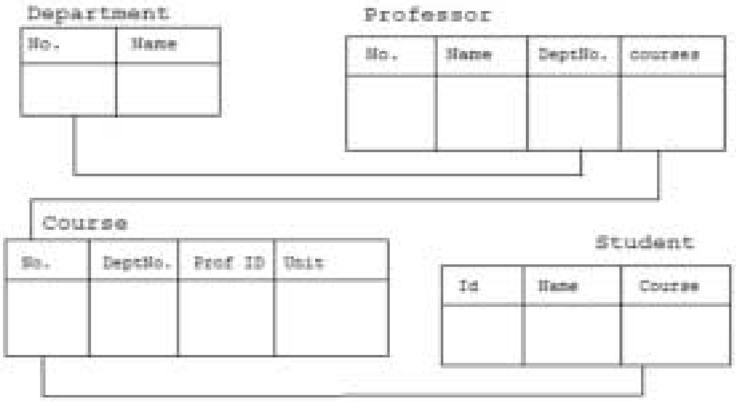


Figure 6: Relational Model

Tabular Model:

The simple tabular model stores attribute data as sequential data files with fixed formats (or comma delimited for ASCII data), for the location of attribute values in a predefined record structure. This type of data model is outdated in the GIS arena. It lacks any method of checking data integrity, as well as being inefficient with respect to data storage, e.g. limited indexing capability for attributes or records, etc.

Hierarchical Model:

In this model each entity has only one parent but can have several children. At the top of hierarchy there is only one entity which is called **Root**. The hierarchical database organizes data in a *tree* structure. Data is structured downward in a *hierarchy* of tables. Any level in the hierarchy can have unlimited *children*, but any *child* can have only one *parent*.

Hierarchical DBMS have not gained any noticeable acceptance for use within GIS. They are oriented for data sets that are very stable, where primary relationships among the data change infrequently or never at all [4].

Network Model:

In the network model, entities are organized in a graph, in which some entities can be accessed through several paths. The network database organizes data in a network structure. This model allows for children to have more than one parent.

Network DBMS have not found much more acceptance in GIS than the hierarchical DBMS because they have the same flexibility limitations as hierarchical databases; however, the more powerful structure for representing data relationships allows a more realistic modeling of geographic phenomenon. However, network databases tend to become overly complex too easily. In this regard it is easy to lose control and understanding of the relationships between elements [.4]

Relational Model:

In this model, data is organized in two-dimensional tables called **relations**. The tables or relation are related to each other. The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* and *columns*. Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g. cover group, tree height. Rows represent one record in the table. In a GIS each row is usually linked to a separate spatial feature, e.g. a forestry stand. Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature[4].

Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually the common column is an identification number for a selected geographic feature, e.g. a forestry stand polygon number. This identification number acts as the *primary key* for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product. Unlike the other previously discussed database types, relationships are implicit in the character of the data as opposed to explicit characteristics of the database set up. The relational database model is the most widely accepted for managing the attributes of geographic data. There are many different designs of DBMSs, but in GIS the relational design has been the most useful. In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link those together [4].

The relational DBMS is attractive because of it's:

1. Simplicity in organization and data modeling.
2. Flexibility - data can be manipulated in an ad hoc manner by joining tables.
3. Efficiency of storage - by the proper design of data tables redundant data can be minimized; and
4. The non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.

Object-Oriented Model:

Object –oriented models define a database as a collection of objects with features and methods. The object-oriented database model manages data through *objects*. An object is a collection of data elements and operations that together are considered a single entity. The object-oriented database is a relatively new model. This approach has the attraction that querying is very natural, as features can be bundled together with attributes at the database administrator's discretion. Today, only a few GIS packages are promoting the use of this attribute data model [4].

Application Areas of GIS:

1. **GIS in Mapping:** Mapping is a central function of Geographic Information System, which provides a visual interpretation of data. GIS store data in database and then represent it visually in a mapped format. People from different professions use map to communicate. It is not necessary to be a skilled cartographer to create maps. Google map, Bing map, Yahoo map are the best example for web based GIS mapping solution [5].
2. **Urban Planning:** GIS technology is used to analyze the urban growth and its direction of expansion, and to find suitable sites for further urban development. In order to identify the sites suitable for the urban growth, certain factors have to consider which is: land should have proper accessibility, land should be more or less flat, land should be vacant or having low usage value presently and it should have good supply of water [5].
3. **Transportation Planning:** GIS can be used in managing transportation and logistical problems. If transport department is planning for a new railway or a road route then this can be performed by adding environmental and topographical data into the GIS platform. This will easily output the best route for the transportation based on the criteria like flattest route, least damage to habitats and least disturbance from local people. GIS can also help in monitoring rail systems and road conditions [5].
4. **Agricultural Applications:** GIS can be used to create more effective and efficient farming techniques. It can also analyze soil data and to determine: what are the best crop to plant?, where they should go? how to maintain nutrition levels to best benefit crop to plant?. It is fully integrated and widely accepted for helping government agencies to manage programs that support farmers and protect the environment. This could increase food production in different parts of the world so the world food crisis could be avoided [5].
5. **Disaster Management and Mitigation:** Today well-developed GIS systems are used to protect the environment. It has become an integrated, well developed and successful tool in disaster management and mitigation. GIS can help with risk management and analysis by displaying which

areas are likely to be prone to natural or man-made disasters. When such disasters are identified, preventive measures can be developed [5].

6. **Navigation (routing and scheduling):** Web-based navigation maps encourage safe navigation in waterway. Ferry paths and shipping routes are identified for the better routing. ArcGIS supports safe navigation system and provides accurate topographic and hydrographic data. Recently DNR, s Coastal Resources Division began the task of locating, documenting, and cataloging these no historic wrecks with GIS. This division is providing public information that makes citizens awareness of these vessel locations through web map. The web map will be regularly updated to keep the boating public informed of these coastal hazards to minimize risk of collision and injury [5].
7. **Soil Mapping:** Soil mapping provides resource information about an area. It helps in understanding soil suitability for various land use activities. It is essential for preventing environmental deterioration associated with misuse of land. GIS Helps to identify soil types in an area and to delineate soil boundaries. It is used for the identification and classification of soil. Soil map is widely used by the farmers in developed countries to retain soil nutrients and earn maximum yield [5].
8. **Tourism Information System:** GIS provides a valuable toolbox of techniques and technologies of wide applicability to the achievement of sustainable tourism development. This provides an ideal platform tools required to generate a better understanding, and can serve the needs of tourists. They will get all the information on click, measure distance, find hotels, restaurant and even navigate to their respective links. Information plays a vital role to tourists in planning their travel from one place to another, and success of tourism industry. This can bring many advantages for both tourist and tourism department [5].
9. **Reservoir Site Selection:** GIS is used to find a suitable site for the dam. GIS tries to find best location that respect to natural hazards like earthquake and volcanic eruption. For the finding of dam site selection the factors include economic factors, social considerations, engineering factors and environmental problems. This all information are layered in the GIS [5].
10. **Forest Fire Hazard Zone Mapping:** Forest is one of the important elements of the nature. It plays an important role in the local climate. Forest fires caused extensive damage to our communities and environmental resource base. GIS can effectively use for the forest fire hazard zone mapping and also for the loss estimation. GIS also help to capture real time monitoring of fire prone areas. This is achieved by the help of GNSS and satellite Remote Sensing [5].
11. **Deforestation:** Nowadays forest area is decreasing every year, due to different activities. GIS is used to indicate the degree of deforestation and vital causes for the deforestation process. GIS is used to monitor deforestation [5].
12. **Pipeline Route Selection:** Pipeline route planning and selection is usually a complex task. GIS technology is faster, better and more efficient in this complex task. Accurate pipeline route selection brings about risk and cost reduction as well as better decision making process. GIS least cost path analysis have been effectively used to determine suitable oil and gas pipeline routes. An optimal route will minimize reduce economic loss and negative socio-environmental impacts [5].
13. **Geologic Mapping:** GIS is an effective tool in geological mapping. It becomes easy for surveyors to create 3D maps of any area with precise and desired scaling. The results provide accurate measurements, which helps in several fields where geological map is required. This is cost effective and offers more accurate data, there by easing the scaling process when studying geologic mapping [5].
14. **Locating Underground Pipes and Cables:** Pipe line and cable location is essential for leak detection. It can be used to understand your water network, conducting repairs and adjustments, locating leaks

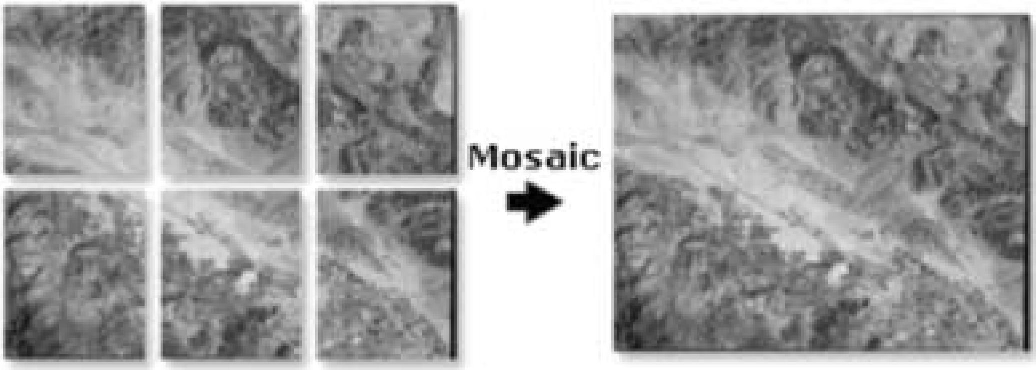


Figure 7: Mosaic Tool [6]

known distance for correlating etc. Pipelines are continually monitored, check for leak detection and avoid the problem of geo hazards [5].

Tools Used in GIS:

ArcMap has been the primary application used in ArcGIS for Desktop for mapping, editing, analysis, and data management. Few tools used to process raster file in GIS are:

1) Calculating statistics allows ArcGIS applications to properly stretch and symbolize raster data for display.
2) Building pyramids improves the display performance of raster datasets.
3) Build raster attribute: Create or update a table with information about the classes in your raster datasets
4) Extraction tool: This tool can be used to get the AOI from a vast dataset. One can extract the data using shapefile of a particular area or known points of a polygon.
5) Mosaic tool: Merges multiple existing raster datasets into an existing raster dataset. Some hydrological tools are present such as fill, flow direction, flow accumulation and many more to provide the hydrological parameters to the river or streams under consideration. Here the streams are made from the DEM(Digital Elevation Model) and using ArcGIS we could make the watershed for the same and then analyse it for different situations.
6) Many math related operations can also be performed on raster such as if we want only the integer value of the raster dataset, we can use int function of Arcmap.

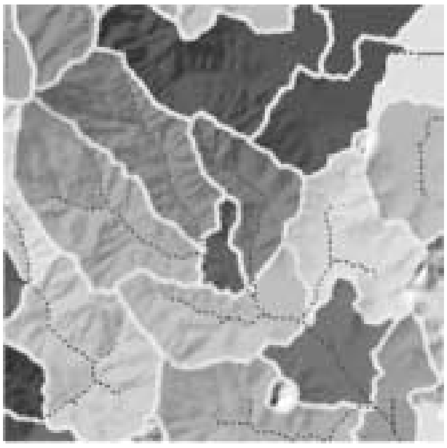


Figure 9: Hydrological Tool [8]

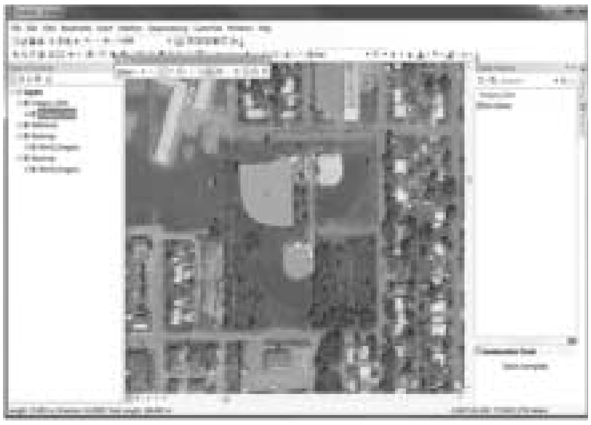


Figure 10: Editing Tool [7]

7) Various editing tools are also present to edit the vector shapefiles. Legends and title for the map can also be given.

   As could be seen if we want to draw the entire agricultural field in a particular area, we can do it using editing tools. Points, lines and polygons can be made or edited using this tool.

8) To see the data set in proper coordinate system with the respect to the earth various projection systems are also provided to project our data in proper reference to earth or globe. In other words, proper coordinates in this case latitude and longitude can be given to the particular area of study. This can enhance our results and real time monitoring and forecasting can be made.

V. Conclusion and Future Work:

GIS combines all kinds of information and applications with a geographic component into one system. The main objective of this paper was to describe GIS and several applications were discussed. The paper summarized the main features of GIS and its functions. In addition, Geographic Information System (GIS) is used as tools for spatial data handling in a geographical environment. GIS is one of the essential tools for decision making in problems related to geo-information. Its ability to integrates and analyze all spatial data to support a decision-making process. Basic elements of GIS consist of hardware, software, data and live ware.

GIS is very useful in case of natural hazards. Flood mapping, relief work and disaster management is done mainly using GIS. Images obtained from various satellites help us to know the actual ground situation and using GIS we can help to analyze the damage or other parameters such as agricultural productivity and urban area development. There is lot of scope GIS in this area.

References:

1. https://en.wikipedia.org/wiki/Geographic_information_system
2. http://www.rst2.edu/ties/gentools/what_gis.html
3. http://planet.botany.uwc.ac.za/nisl/GIS/GIS_primer/page_19.htm
4. http://planet.botany.uwc.ac.za/nisl/GIS/GIS_primer/page_20.htm
5. http://grindgis.com/blog/gis-applications-uses
6. http://pro.arcgis.com/en/pro-app/tool-reference/data-management/mosaic.htm
7. http://www.esri.com/esri-news/arcuser/fall-2014/use-arcgis-online-imagery-to-digitize-analyze-contribute
8. http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-watershed-works.htm

Testable Web Design: A Centralized Approach

Sonal Anand*

Abstract

This paper proposes a new web application design which is proved to be less complex and more testable than the design of existing web applications. OOD (Object Oriented Design) metrics including NOC (Number of Children) and CBC (Coupling between Classes) have been used to demonstrate that the proposed web application design is more testable.

Key Words: CBC, NOC, OOD, Testability, Web design

I. Introduction

Web application testing is a specialization of software testing. It focuses on identification of errors in a web application to verify whether the output meets the specification. Web applications are dynamic and interactive, and work on autonomous systems, as compared to traditional applications. Therefore, traditional testing techniques cannot successfully be incorporated in testing web applications.

ISO has defined testability of software or a web application as the attributes that bear on the effort needed to validate the product [1].Testability is the ease with which it can be tested or the degree to which it supports testing. In other words, it is the degree to which software and web components are designed and implemented, to make defects more discoverable [2].The object oriented design metrics have been used to quantify the relationship between complexity and testability. High complexity leads to reduced testability. A lower degree of testability results in an increased effort.

In this paper, we distinguish between two categories of web application design. Current websites are of decentralized nature. A decentralized web application is the one where every web page contains links to every other page, without requiring the need for a central web page, as shown below. We have coined the term "**decentralized design**" to compare it with the "**centralized design**", a new design concept that has been proposed by us and proved to be much lesser complex and thus relatively more testable. A centralized web application is the one which has one central page that stores the links to every web page. All other web pages are directly connected to this central page. So to traverse from one web page to other requires redirecting to this central page.

The organization of this paper is as follows. Section II provides the literature review. In Section III, elaborates the current web design, and apply the OOD metrics to give a quantitative measure of NOC and CBC values, and hence testability. In Section IV, we propose a new web design called centralized design and measure NOC and CBC values and hence, testability. In Section V, we compare the results of OOD metric values of both the conventional (decentralized) and the proposed (centralized) designs and prove that the proposed design is more testable.

II. Literature Review

In this section, a review of the literature studied has been discussed.

R.A. Khan and K.Mustafa have proposed "A metric based testability model for object oriented design", [3] which provides adequate object oriented metrics to determine the testability of a system. The metrics include ENM(encapsulation metrics) which counts the number of all methods defined in the class, REM(Reuse –inheritance metric) which counts the depth of inheritance in a class, and CPM(Coupling

Sonal Anand*
Assistant Professor,
Department of IT, IITM, Janakpuri, New Delhi

metric), which counts the number of classes related to a class. The emphasis is to predict class testability in initial phase of development. It also provides the empirical analysis of proposed metrics and implements it on industrial projects.

Aymen Kout, Fadal Toure and Mourad Badri have presented "An Empirical Analysis of a Testability Model for Object Oriented Programs", [4] which aims at exploring the empirical relationship between testability and classes in object oriented systems. The data has been collected from two open source java software systems. A metrics has been presented to quantify the testing effort required for a class by evaluating the corresponding JUnit test classes.

Sadaf Khan, Saima Zehra and Fahim Arif have presented "Analysis of Object Oriented Complexity and Testability Using Object Oriented Design Metrics" [5], wherein object design metrics have been modified to analyze in detail the relationship between complexity, testability and different attributes of object oriented design. The extended metrics include AHF (Attribute hiding factor) , MHF(Method hiding factor), DIT(Depth of inheritance tree), NOC(number of children) and CBC(coupling between classes). The proposed metrics has been mapped on several software projects, and a clear relationship between testability, complexity and object oriented design attributes has been depicted.

Magiel Bruntik has proposed "Testability of Object Oriented systems: A metric based approach" [6]. The metrics has been evaluated on two Java software systems. The goal is to access the testability of classes in Java systems. An extended set of metrics has been included, viz. FOUT (Fan out), LCOM (Lack of cohesion of methods), LOCC (lines of code per class), NOF (Number of fields), NOM (Number of methods), RFC (Response for class), WMC (Weighted methods per class). Bruntink and Deursen have presented a method of predicting class testability using OOD metrics [7].

III. Description of Application to be Tested

Here, we discuss an example of a web application. Both centralized and decentralized designs of the same application have been compared for testability.

The goal of the application is to develop a web portal that will help to provide technical support to people. The portal must give maximum flexibility to the users so that they are able to convey their problems and get solutions for the same. The portal should have features including: 1) **Online tutorials**, 2) an interface for **asking questions to experts**, and 3) links to authenticated **software download**. A user after logging in is directed to **user main** page to carry out further traversing.

A. Decentralized Design

In Fig 1, we describe the decentralized design of the example applications, and apply the OOD metrics to give a quantitative measure of NOC and CBC values, and hence testability. Following is the decentralized design for the application discussed. Every page is linked to every other web page, giving it a full mesh structure.

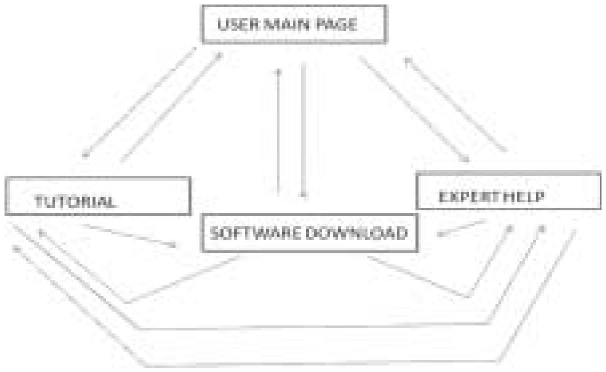


Fig. 1: Decentralized Design

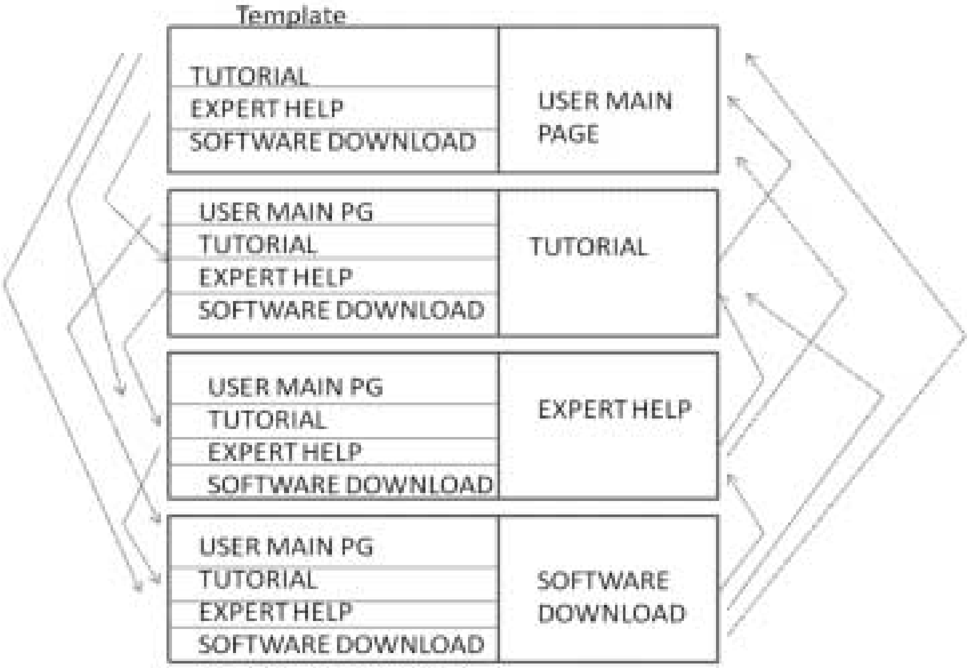


Fig. 2: Inherited Template: Decentralized Design

B. Measure of NOC Value (Number of Children) of Decentralized Design

As per the metric NOC (Number of Children) defined in [5], higher the number of inherited classes, lower is the testability. Considering each web page as a separate class, and the template as the inherited attribute, this template has been inherited from the base class, the user main page, by the three subsequent classes namely tutorial, expert help and software download, thereby increasing the NOC and decreasing the testability. Considering the user main page as the base class, there are three subclasses inheriting the 'web link' template form the base class, thus making the value of NOC for base class to be 3.the inheritance of links has been depicted in the Figure 2.

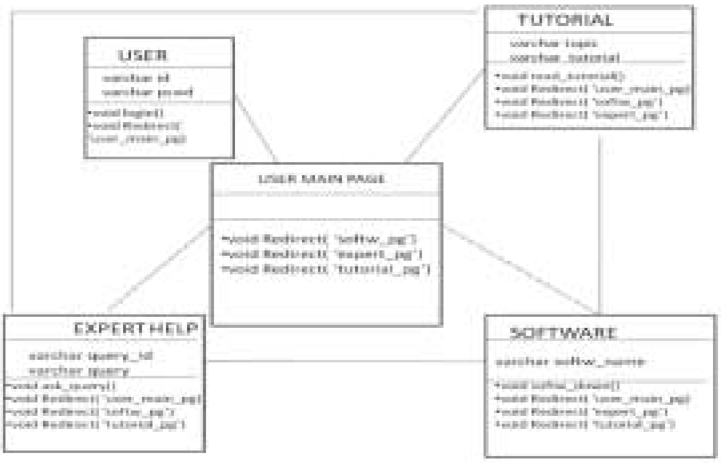

TUTORIAL
varchar topic
varchar tutorial
+void read_tutorial()
+void Redirect('user_main_pg')
+void Redirect('softw_pg')
+void Redirect('expert_pg')
USER
varchar id
varchar pswd
+void login()
+void Redirect(
'user_main_pg)
USER MAIN PAGE
+void Redirect('softw_pg')
+void Redirect('expert_pg')
+void Redirect('tutorial_pg')
EXPERT HELP
varchar query_id
varchar query
+void ask_query()
+void Redirect('user_main_pg)
+void Redirect('softw_pg')
+void Redirect('tutorial_pg')
SOFTWARE
varchar softw_name
+void softw_down()
+void Redirect('user_main_pg)
+void Redirect('expert_pg')
+void Redirect('tutorial_pg')

Fig. 3: Decentralized: High Coupling

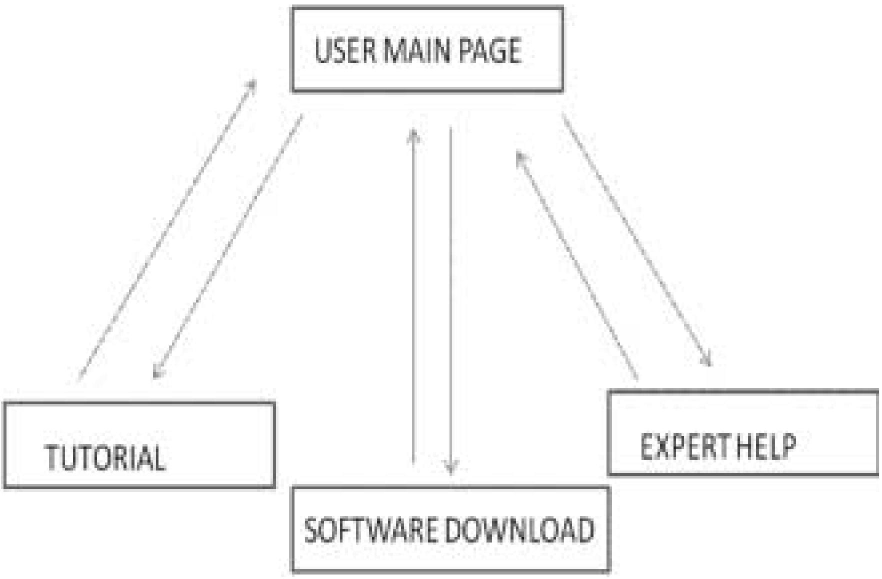


Fig. 4: Centralized Design

C. Measure of CBC Value (Coupling Between Clases) of Decentralied Design

CBC refers to coupling between classes, i.e. the number of classes related to a given class. According to [18], higher the value of CBC, lower is the testability.

Class diagram for **decentralized** website is depicted in Fig 3. The user main page class contains methods that redirect the user to tutorial, expert help and software download. Since the classes tutorial, expert help and software download have inherited templates from user main page class, these three classes also contain links to each other, thus making the total CBC value to be summation of coupling links for all classes, i.e. 7.

IV. Proposed Design: Centralized Design

Here, in Fig 4, the user main page acts as the central page to connect all other pages. So to traverse from one page to other, the user needs to traverse via the user main page.

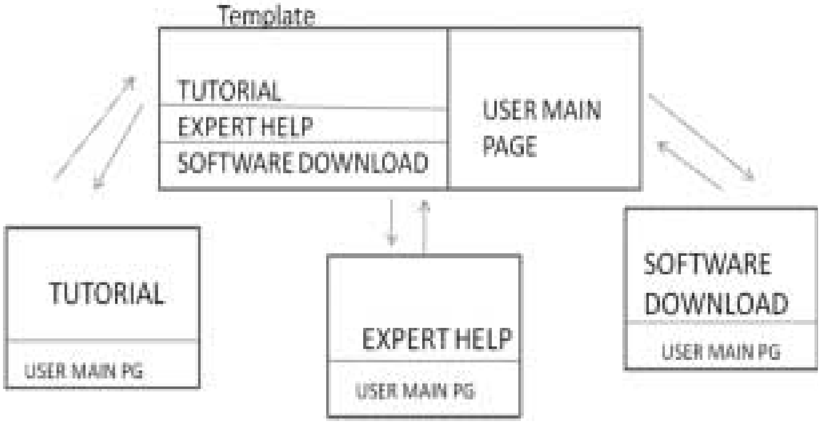


Fig. 5: NOC of Centralized Design

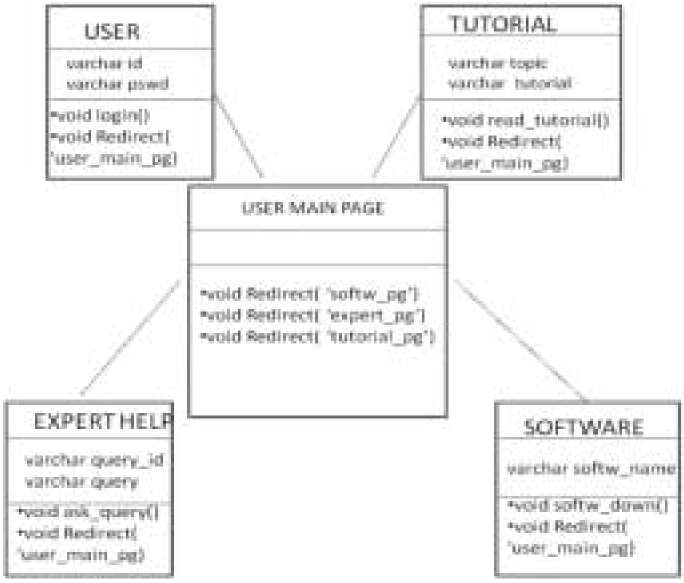


Fig. 6: CBC of Centralized Design

A. Measure of NOC of Centralized Design

As Fig 5 depicts, there is no inheritance of links among web pages, as their central page connecting all other pages, therefore value of NOC is zero.

B. Measure of CBC of Centralized Design

For **centralized** website, class diagram is shown in Fig 6. Since, there is no inheritance of templates; the classes expert help, software download and tutorial are not directly linked to each other. Therefore coupling here is lesser, and CBC value is 4. And hence, testability is higher as compared to decentralized websites.

V. Results

By comparing the results of NOC and CBC values of both designs, we conclude that the metric values for our proposed design called centralized design are lesser than the conventional decentralized web design. As per [5], values of CBC and NOC are directly proportional to complexity of the product, and a high value of complexity is responsible for low testability. Thus, our proposed design is lesser complex and more testable than the current web design.

References

1. ISO/IEC 9126: Software Engineering Product Quality, ISO Press, 1991
2. Vishal Chowdhary," Practicing Testability In The Real World', International Conference On Software Testing And Verification, 2009
3. R.A. Khan and K.Mustafa, "A metric based testability model for object oriented design", ACM SIGSOFT Software Engineering Notes, volume 34 number 2
4. Aymen Kout, Fadal Toure and Mourad Badri, "An Empirical Analysis of a Testability Model for Object Oriented Programs", ACM SIGSOFT Software Engineering Notes, July 2011 volume 36 number 4
5. Sadaf Khan, Saima Zehra and Fahim Arif "Analysis of Object Oriented Complexity and Testability Using Object Oriented Design Metrics", NSEC '10 Proceedings of the 2010, National Software Engineering Conference , Article 4[6] Magiel Bruntik "Testability of Object Oriented systems: A metric based approach" Universiteit Van Amsterdam
6. Magiel Bruntink, Arie Van Deursen,"Predicting Class Testability Using Object Oriented Metrics," Fourth IEEE International Workshop on Source Code Analysis and Manipulation, 2004.

Security Infrastructure of Cloud Computing

Sahaj Arora*
Dr. Geetali Banerji**

Abstract

Security Infrastructure is all about giving the security to the network layer which is most attacked by the "hackers", although the central attraction for security threats is increasingly shifting towards the application layer. This paper, discuss about the attacks by the attackers and how to give security to our applications in different ways.

Key Words: Cloud Security, Cyber Security, Security Infrastructure

I. Introduction

Security infrastructure, in context of cyber security, relates to the security provided to the particular organization [3]. Now-a-days, everything is now on internet whether it is playing games, chatting, sharing of information even while searching for any information, everything needs security. Think what will happen if you put no locks on your house and there is no member in there, your house will be at high risk of getting mugged. In reality thieves are the one who attacks to the house (In example) but in security world 'Hackers' do this job.

Hacker is the person who attacks over a system to take out important information. Hacker can attack in many ways and that's why we need to be aware of the security to the systems. Security infrastructure lets you know about:

- How the security internally is working.
- Why do we need to provide security to our system?
- In how many ways we can secure our systems?
- Places where high security is needed, etc.

II. Cloud Security

There is no such definition made on cloud computing but it refers to the safety of the cloud itself for running applications, storing data and processing transactions. This is a concern of more companies as they try to leverage the low-cost advantages of cloud security solutions without compromising corporate or customer information. Cloud computing refers to a network of computers, connected through internet, sharing the resources given by cloud providers catering to its user's needs like scalability, usability, resource requirements. The USA National Institute of Standards and Technology (NIST) defines it as follows: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." Cloud computing allows users to access software applications and computing services. They might be stored off-site at locations rather than at local data centre or the user's computer. In general cloud providers offer three types of services i.e. Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). There are various reasons for organizations to move towards IT solutions that include cloud computing as they are just required to pay for the resources on consumption basis. In addition, organizations can easily meet the needs of rapidly changing markets to ensure that they are always on the leading edge for their consumers.The cloud or the network specially the public one, raises many security concerns for companies that are accustomed to hosting their data and applications within their own

Sahaj Arora*
Student of BCA
IITM, Janakpuri, New Delhi

Dr. Geetali Banerji**
Professor, Department of IT
IITM, Janakpuri, New Delhi

four walls. Within a traditional internal IT security infrastructure, it is comparatively easy to ensure proper security mechanisms such as authentication, authorization, non repudiation, privacy, confidentiality. These mechanisms must be accompanied by proper security policies and processes that are followed by employees [4]-[5].

III. Cloud Security Models

A. SaaS

This particular model is focused on managing access to applications. For example, policy controls may dictate that a sales person can only download particular information from sales CRM applications. For example, they are only permitted to download certain leads, within certain geographies or during local office working hours. In effect, the security officer needs to focus on establishing controls regarding users' access to applications.

B. PaaS

The primary focus of this model is on protecting data. This is especially important in the case of storage as a service. An important element to consider within PaaS is the ability to plan against the possibility of an outage from a Cloud provider. The security operation needs to consider providing for the ability to load balance across providers to ensure fail over of services in the event of an outage. Another key consideration should be the ability to encrypt the data whilst stored on a third-party platform and to be aware of the regulatory issues that may apply to data availability in different geographies.

C. IaaS

In this model the focus is on managing virtual machines. The CSOs priority is to overlay a governance framework to enable the organization to put controls in place regarding how virtual machines are created and spun down thus avoiding uncontrolled access and potential costly wastage.

IV. Cloud Security Checklist

The following list a checklist of cloud security[1]:

- Continuous web application scanning to detect vulnerabilities. Boot and data volume encryption with external key management to protect data at rest and keep control of the keys.
- SSL certificates to protect data-in-motion with encryption.
- Intrusion Prevention with virtual patching to protect against vulnerabilities even before you patches.
- Host-based bi-directional firewall to prevent unauthorized outbound communication – with logging and alerting capabilities to make it easier to manage.
- File integrity monitoring to catch unauthorized system component changes.
- Anti-malware with web reputation to protect against viruses and malicious URLs.

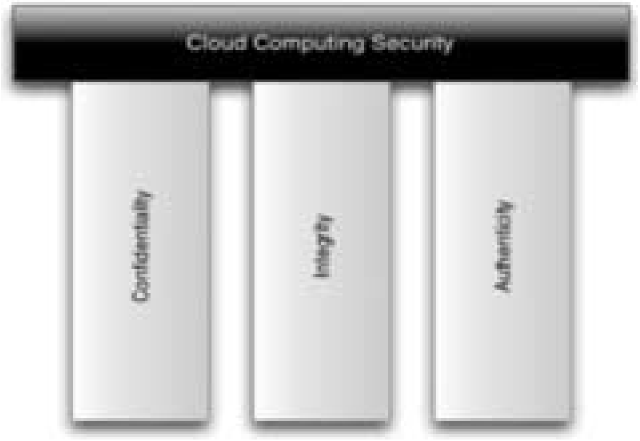


Figure 1: Cloud computing Security Issues

Table-I. Advantages and Challenges of security

Advantages	Challenges
Dedicated Security Team	Data dispersal and international privacy laws
Greater Investment in Security Infrastructure	Quality of service guarantees
Fault Tolerance and Reliability	Dependence on secure hypervisors
Greater Resiliency	Attraction to hackers (high value target)
Hypervisor Protection Against Network Attacks	Security of virtual OSs in the cloud
Reduction of Assessment and Authorization Activities (FedRAMP)	Possibility for massive outages
Simplification of Compliance Analysis	Encryption needs for cloud computing
Data Held by Unbiased Party (cloud vendor assertion)	Encrypting access to the cloud resource control interface
Low-Cost Disaster Recovery and Data Storage Solutions	Encrypting administrative access to OS instances
On-Demand Security Controls	Encrypting access to applications
Real-Time Detection of System Tampering	Encrypting application data at rest
Rapid Re-Constitution of Services	Public cloud vs internal cloud security
	Data ownership issues
	Need for isolation management
	Multi-tenancy
	Logging challenges
	Data retention issues
	Exposure of data to foreign government and data subpoenas

V. Cloud Security Issues

The three issues of cloud computing security are[2]:

A. Availability

Availability is the attestation that data will be available to the user in a perpetual manner irrespective of location of the user. It is ensured by: fault tolerance, network security and authentication.

B. Integrity

Integrity is the assurance that the data sent is same as the message received and it is not altered in between. Integrity is infringed if the transmitted message is not same as received one. It is ensured by: Firewalls and intrusion detection.

C. Confidentiality

Cloud computing can offer many advantages when it comes to security. Because security is such a 'hot button' issue, in many cases cloud computing solutions over compensate for security risks, sometimes dedicating entire security teams to monitor the system. Security elements are also often able to be delivered in real time. Furthermore, data is monitored in real time due to the nature of cloud computing.

Below is a list of the security advantages and challenges associated with cloud computing:

VI. Conclusions

Security Infrastructure of Cloud Computing has become one of the top most concerns in Cyber World. It becomes very important for cyber citizen to be aware as well to be secure. This paper throws a light on cloud computing, its models, measures to be taken for securing oneself in cyber world. It also discusses about the issues, advantages and challenges to face for getting secure infrastructure.

References

1. Security Guidelines for Critical Areas of Focus in Cloud Computing. Cloud Security Alliance.http://www.cloudsecurityalliance.org/guidance/csaguide.pdf 13.
2. Vouk, M. A. (2008). "Cloud Computing – Issues, Research and Implementations". In Proceedings of the 30th International Conference on Information Technology Interfaces (ITI'08), pp. 31-40, Cavtat, Croatia, June 2008
3. Why Cloud Computing Needs Security. Describes the challenges that consolidated cloud computing sites pose for security. http://gigaom.com/2008/06/10/the-amazon-outage-fortresses-in-the-clouds/17
4. Alexa Huth, James Cebula, " The Basics of Cloud Computing, United State Computer Emergency Readiness Team (US-CERT)
5. Grace Walker, Cloud Computing Fundamentals – A different way to deliver computer resources, IBM.

Bitcoin: Currency of Internet

Shashank Mahajan*
Simranjeet Kaur**

Abstract

Bitcoin is online currency used for making/receiving payments directly from one person to person without paying bank fees. Bitcoin works on Peer to Peer connection which gives privacy to the customer and as there is no intermediate so transaction is without any delay. It also saves the transaction fees and it is a worldwide currency so there is no need to convert in traditional currency.

Key Words: Bitcoin, Block Chain, Deep web, Mining, Peer to Peer

I. Introduction

Bitcoin is entirely Digital currency. Users communicate with each other using the bitcoin protocol, primarily via the Internet. Bitcoin is open-source and bitcoin stack can easily work on laptops, smart phones and many devices. Using bitcoin as a medium to exchange money is much secured than the traditional money because of it's encryption and digital signatures on every bitcoin.

*A bitcoin address is the most important part of this currency which looks like '***1HD9GtNrVH3KGxg PQfpChUM6cNGEPrjjiB***'. Bitcoin address will always start from 1.*

The address is the only information which is needed to send money to users. It may be different characters or in QR code.

The users have software called as 'bitcoin wallet'. Wallet can be online or it can be stored on any of the user's computer. There are no physical bitcoins officially but some firms for example "Casascius Coins" are making physical bitcoin. They are often featured in mainstream media coverage of bitcoin. Each coin has its own bitcoin address and it also contain a private key which is required to spend it. It is hidden underneath the hologram.

Bitcoin is the modern currency which is entirely virtual. It not just like our traditional currency. Bitcoin users don't need to wait for banks to accept their request. Bitcoin protocol included built-in algorithms that are needed to solve   the transaction. For the solving algorithm, bitcoin has a separate community which is called as Bitcoin miners. Here 'Mining' represents for solving the mathematical problems. Miners get bitcoins as their reward.

Bitcoin is more than just normal money. Bitcoin makes it possible to transfer money without trusting the bank. It is a decentralized currency, which means the exchange directly depends upon the demand and supply of bitcoins.

II. History

In 2007, Satoshi Nakamoto published a paper on the cryptography mailing at metzdowd.com. In 2009, Nakamoto releases the first bitcoin software and units of the bitcoin crypto currency. One of the greatest strength of bitcoin is it's open source software. With open source software anyone can read the source code. So Nakamoto continued to collaborate with other developers, one of the senior most developer was Gavin Andresen since Satoshi's disappearance.

Shashank Mahajan*
Student of BCA
IITM, Janakpuri, New Delhi

Simranjeet Kaur**
Assistant Professor, Department of IT
IITM, Janakpuri, New Delhi

III. How Bitcoin Work

Every Transaction is recorded in a block chain that acts as a ledger that is encoded on the bitcoin itself.

This prevents the using of same bitcoins more than once. As bitcoin doesn't work from any server it uses the P2P (Peer to peer) network for the process.

It is the number associated which can be stored on the phone or hard drive.

The user can change other currency in terms of bitcoins through many services, for example, Bitinstant, Coinbase or user can mine their own bitcoins.

Bitcoin contains inbuilt complex mathematical algorithm which is needed to solve in order to complete the transaction of other bitcoin users. After user solves the algorithm, the reward is given in return for solving the problem. This is how new bitcoin is generated in the economy. As there is no decentralized authority Satoshi Nakamoto set a maximum limit of bitcoin which is 21 million.

It's like gold, the more people found bitcoin, the lesser it gets so they become harder to find. With the little competition, the problems are easier to solve but more people join the search, the problems get difficult.

After verification of the transaction by miners, the payment is recorded in one central hash table which works as a ledger, in which all the transaction is being recorded and it is transparent to the public.

IV. Advantages and Disadvantages

Bitcoin is for everyone as sender just needs to scan the QR code of the receiver and enter the amount which user need to send.

As Satoshi's original paper states in his concluding paragraph, "We have proposed a system for electronic transactions without relying on trust."

Bitcoin is decentralized which makes the bitcoin more trustworthy. No one can change the value of the bitcoin. Everything depends upon demand and supply.

As no bank is needed in bitcoin transaction or maintenance so there will be no account closure. No freezing of account can be done.

In order to use the current money. We must have a significant amount of trust in multiple institutes:

A. ***Banks*** – Public can use banks for saving their money and public trust banks that they will be able to pay them back. Banks don't keep the money with themselves they invest the public money. It is not certain that public money is safe inside a bank.

B. ***Central Bank*** – In India, Reserve Bank of India, Federal Reserve in the United States, Japan has the Bank of Japan. Citizen trust that the central banks will create a balance in the market so that there is no excess money supply in the market or inflation can take place which will higher prices for everybody.

C. ***Payment Processors*** – We trust when we spend or accept money online, the payment processors ensure that there is no double spending of money.

V. Who Accepts Bitcoins as Payment

- **Wikipedia:** A Non-Profit Organization which content encyclopedia project supported by the Wikimedia Foundation which runs on donation and they accept bitcoin as a donation.
- **Overstock.com:** One of the leading in online direct retail sales. It sells big ticket items at lower prices due to overstocking.
- **WordPress.com:** One of the most popular companies for the creating blogs. It is free and open-source content management system based on PHP and MYSQL.
- **Lamborghini:** Buy cars with Bitcoin. In 2013, Lamborghini dealership sold a Tesla Model S for an unknown sum. Tesla S starts from £49,900 in the UK, which is around 81 Bitcoins.
- **Mel B:** Melanie Janine Brown, a British recording artist, actress, television personality, and model. She made an album available to purchase in bitcoin.
- **Bitcoin.Travel**: A travel site that provides accommodation, apartments, attractions, bars and beauty salons around the world.

- **Reddit:** You can buy premium features there with bitcoins.
- **Zynga:** Zynga is among a new generation of companies focused on social gaming.
- **Expedia.com:** At the time of payment the website re-direct to Coinbase's website, where the customer will see the total cost of booking in Bitcoin for their flights.
- **Tigerdirect:** Major electronic online retailer.

VI. Problems Caused by Bitcoins

People started using bitcoins for the illegal purposes on the dark web. All the transaction of goods like drugs from Silk Road. Silk Road was an online Black market and first darknet market. It was the Best place for selling illegal drugs. The website was launched in February in 2011 by Ross William Ulbricht the founder of the Silk Road. More than a billion dollars $ worth of illegal goods was pump through the website. In October 2013, the federal bureau of investigation (FBI) shut down the website.

On 6th November 2013, Silk Road 2.0 came online. It was shut down too on 6th November 2014.

Bitcoin gives the power to purchase and sell drugs anonymously. That was the major issues to government. Even the most powerful government in the world, at least when it comes to global finance, doesn't even agree that the bitcoin is a currency at all.

It was not just the Silk Road but dozens of 'hitman ' are available for hire through the deep Web or Tor. They are often paid through bitcoins. One boasts: 'I always do my best make it look like an accident or suicide'.

Hiring a hitman has never been easier. Nor has cocaine or heroin.

Hackers are available too for some bitcoins. For example, Facebook account hacking is for 5 bitcoins and website hacking is for 15.00 bitcoins And even some dark website are selling fake passports, driving license etc.

All these transactions are through bitcoins.

People are uploading horrific child pornography for bitcoins too.

As it was a part of the dark web it was operated on a special browser "tor". Tor is software that allows users to browse the Web anonymously.

Tor is short for "The Onion Router."

This refers both to the software that you install on your computer to run Tor and the network of computers that manages Tor connections. Put simply, Tor enables you to route web traffic through several other computers in the Tor network so that the party on the other end of the connection can't trace the traffic back to you.

This was initially developed for the army for the use of encrypted email services, With "tor".

VII. Security Issue

As there is no regulatory agency and it is all decentralized setting bitcoin value, it tends to fluctuate widely which can make the investment much riskier than other currencies. The rates of the bitcoin are very uncertain. We cannot estimate bitcoin's value; it can increase by 120% in 6hours and can also decrease by 80% which is very unhealthy for the economy.

The value of bitcoin in November 2011 crash to 2$ in the US and in November 2013 the value was estimated of 1000$ in the US.

Another bitcoin problem is security even through the algorithm of the bitcoin is remarkable but that doesn't stop hackers to rob any bitcoin wallets.

Again this is not the fault of the currency. We cannot say if the bank is robbed, we wouldn't blame the ₹ rupees/$dollars for the robbery.

It is quite possible to have a better algorithm of crypto currency than what will happen to the bitcoin.

References

1. Bitcoin: A Peer-to-Peer Electronic Cash System by Satoshi Nakamoto https://bitcoin.org/bitcoin.pdf
2. 'Mastering Bitcoin by Andreas M. Antonopoulos 2014'
3. A Step By Step Guide To Buying, Selling And Investing In Bitcoin - Patterson, Sam

Cyber Crime and its Types

Nalci*
Charul Nigam**

Abstract

As we all know that Cyber crime has been one of the common practices made by the computer expert. Cyber crime is that activities made by the people for destroying organization network, stealing others valuable data, documents, hacking bank account and transferring money to their own. This paper gives detailed information regarding Cyber crime, its types, including prevention to deal effectively with Cyber crime.

Key Words: Cyber crime, Drug Trafficking, Hacking

I. Introduction

Criminal activities carried out by means of computers or the Internet. Computer crime, or Cyber crime, refers to any crime that involves a computer and a network [1]. The computer may have been used in the commission of a crime, or it may be the target.Cyber crime is the latest and the most complicated problem in the cyber world. "Cyber crime is a big problem in which either the computer is an object or subject of the conduct constituting crime". "Any criminal activity that uses a computer either as an instrumentality or target it cause a problem called Cyber crime". Definition of Cyber crime may be "unlawful acts wherein the computer is either a tool or target or both"**.** The computer may be used as a tool in the following kinds of activity- financial crimes, sale of illegal articles, pornography, online gambling, intellectual property crime, e-mail spoofing, forgery, cyber defamation, cyber stalking. The computer may however be target for unlawful acts in the following cases- unauthorized access to computer and computer system and also computer networks, theft of information contained in the electronic form, e-mail bombing, logic bombs, Trojan attacks, internet time thefts, web jacking, theft of computer system, physically damaging the computer system. In Cyber crime when an attacker wants to hack data from a particular system across the internet network is called Cyber crime. The computer can be considered as the tool rather than the target. Cyber crime is a crime that involves a computer and a network. Example of crime is Scams, theft etc.

A. History of Cyber Crime

When computers and networks came into being in the 1990s, hacking was done basically to get more information about the systems. Hackers even competed against one another to win the tag of the best hacker. As a result, many networks were affected, hacking started making networks and systems slow. As hackers became more skilful, they started using their knowledge and expertise to gain benefit by exploiting and victimizing others [7].

II. Types of Cyber Crime

When any crime is committed over the Internet it is referred to as a Cyber crime [4]. There are many types of Cyber crimes and the most common ones are explained below:

A. Hacking:

This is a type of crime wherein a person's computer is broken into so that his personal or sensitive information can be accessed. In hacking, the criminal uses software to enter a person's computer and the

Nalci*
Student of BCA
IITM, Janakpuri, New Delhi

Charul Nigam**
Assistant Professor, Department of IT
IITM, Janakpuri, New Delhi

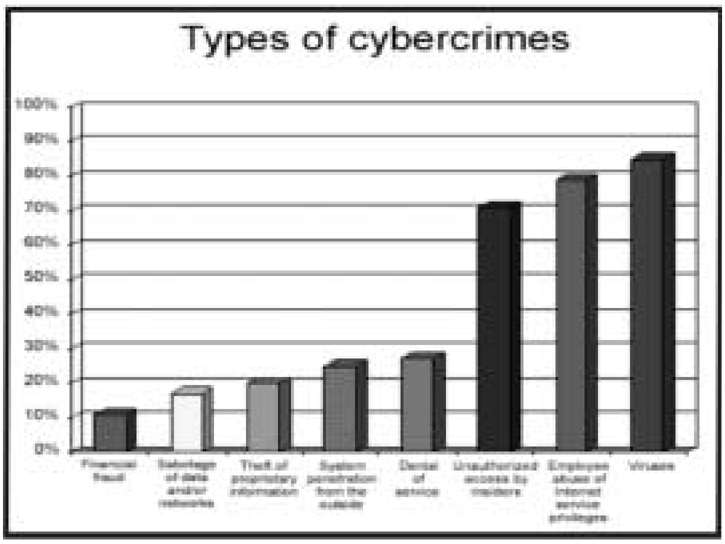


Figure 1: Types of Cyber crime Statistics

person may not be aware that his computer is being accessed from a person that is hacker. Hacker is a person or human being who tries to break into computer systems. Hacker is a clever programmer who might be able to access your personal details. He will be able to misuse your personal detail or any other secret information.

B. Theft:

Theft is the taking of another person's property without that person's permission with the objective to steal the rightful owner of it. This crime occurs when a person downloads music, movies, games and software.

C. Malicious Software:

These are Internet-based software or programs that are used to disrupt a network. The software is used to gain access to a system to steal sensitive information or data in the system.

D. Child Abuse:

This is also a type of Cyber crime; in this kind of crime criminals solicit minors via chat rooms for the purpose of child pornography.

E. Harassment:

Harassment is the Cyber crime most commonly encountered in chat rooms or through newsgroups comments directed towards a specific individual e.g. on gender, race, religion & nationality.

F. Drug Trafficking:

Drug traffickers use the Internet as a medium for trading their illegal substances or any useful information and data by sending out email & other Internet technology. Most of the drug traffickers can be found arranging their illegal deals at internet cafes.

G. Cyber Terrorism:

Due to the increase in cyber terrorism, hacker can crash the official websites and they can also steal the information from the government official website.

III. Impact of Cyber Crime

The impact of Cyber crime is financial losses, theft of intellectual property and loss of customer confidence and trust. The first step of Cyber crime is too aware and gives training to all the peoples, citizen, consumer and employee and students as well. They all should be aware of these attacks and they can take action to protect their own information and personal details.

IV. Conclusion and Future Scope

This paper discuss about the current scenario of Cyber crime. How it has become a problem in our country.

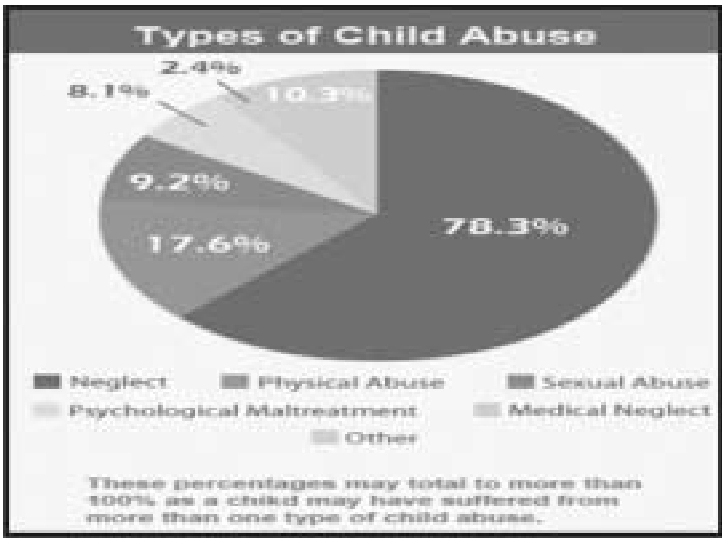


Figure 2: Types of Child Abuse Statistics

It also becomes very important to be aware of the security risks of Cyber crime that can be done on the internet side. We have to know about Cyber crimes that now become a problem to us. The internet can open the door to a world of entertainment and education, but you could be at risk if you don't know how to use it safely. "If you ever feel uncomfortable, tell your parents or a trusted adult."

References

1. searchsecurity.techtarget.com/definition/cybercrime
2. infosecawareness.in/cyber-crime-cells-in-india
3. www.cyberlawsindia.net/cases.html
4. www.naavi.org/pati/pati_cybercrimes_dec03.html
5. http://www.crossdomainsolutions.com/cyber-crime/
6. http://drmzz.blogspot.in/2013/07/abstract-of-cyber-crime.html
7. www.indiancybersecurity.com/.../8_history_of_cyber_law_in_india.html

Association Rule Mining Algorithms: A Comparative Review

Dr. Rashmi Jha*

Abstract

Association rule mining has become one of the core data mining tasks and has attracted tremendous interest among data mining researchers. It is the one of the most important technique of the data mining. Association rule mining aim is to extract interesting correlations, frequent patterns and association among set of items in the transaction database. In this paper, a review of four different association rule mining algorithms Apriori, AprioriTid, Apriori hybrid and tertius algorithms and their drawbacks which would be helpful to find new solution for the Problems found in these algorithms and also presents a comparison between different association mining algorithms.

Key Words: Data mining, Association rule algorithms, Apriori, AprioriTid, Apriori hybrid and Tertius algorithms

I. Introduction

Data Mining is a Science of generating novel, meaningful, simple, understandable, sustainable, and result oriented patterns which can help in understanding the various kinds of data as per different business or application context. The patterns should be novel in the sense they should provide new ideas to the users after applying data mining. Meaningful refers to the right generation of various patterns rather than just normal introduction of various kinds of patterns. Simple refers to simplicity in the implications so that could be derived business implications or application implications easily. Understandable refers to the reduction in various types of complexities which helps in making the patterns easier to understand and relate to. Sustainable refers to the long term solutions generated through the data mining process rather than just short term and volatile patterns. And the most important being result oriented, the patterns should be able to produce results for the users or else the whole concept of mining will not be useful.

II. Association Rule Mining

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [2]. Based on the concept of strong rules. A typical and widely-used example of association rule mining is Market Basket Analysis . The problem is to generate all association rules that have support and confidence greater than the user-specified minimum support and minimum confidence.

$$\begin{tikzcd} & Support = \frac{frq(X,Y)}{N} \\ Rule:\ X \Rightarrow Y \arrow[ur] \arrow[dr] & \\ & Confidence = \frac{frq(X,Y)}{frq(X)} \end{tikzcd}$$

A. Support (S)

Support(S) of an association rule is defined as the percentage/fraction of records that contain X*"Y to the total number of records in the database. Suppose the support of an item is 0.1%, it means only 0.1percent of the transaction contain purchasing of this item.

Support (XY) = Support count of (XY) / Total number of transaction in D

Dr. Rashmi Jha*
Assistant Professor, Department of IT
IITM, Janakpuri, New Delhi

B. Confidence (C)

Confidence(C) of an association rule is defined as the percentage/fraction of the number of transactions that contain X*"Y to the total number of records that contain X. Confidence is a measure of strength of the association rules, suppose the confidence of the association rule XÒ!Y is 80%, it means that 80% of the transactions that contain X also contain Y together.

Confidence (X|Y) = Support (XY) / Support (X)

C. Association Rules Goals

Find all sets of items (item-sets) that have support (number of transactions) greater than the minimum support (large item-sets).

Use the large item-sets to generate the desired rules that have confidence greater than the minimum confidence.

D. General AR Algorithm

In the first pass, the support of each individual item is counted, and the large ones are determined

In each subsequent pass, the large item-sets determined in the previous pass is used to generate new item-sets called candidate item-sets.

The support of each candidate item-set is counted, and the large ones are determined

This process continues until no new large item-sets are found.

III. Apriori Algorithm

Apriori Algorithm is mining scheme designed for frequent item set mining and association rule learning over transactional databases. It identifies the frequent individual items in the database and extends them to larger and larger item sets. The frequent item sets determined by Apriori which can be used to determine association rules which highlight general trends in the database. Apriori uses a "bottom up" approach, i.e. frequent subsets are extended one item at a time called candidate generation, and groups of candidates are tested against the data. "The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules." It makes use of the downward closure property. The algorithm is a bottom search, moving upward level-wise in the lattice. However, before reading the database at every level, it prunes many of the sets which are unlikely to be frequent sets, thus saving any extra efforts.

Candidate Generation: Given the set of all frequent (k-1) item-sets. We want to generate superset of the set of all frequent k-item-sets. The intuition behind the aprior candidates generation procedure is that if an item-set X has minimum support, so do all subsets of X. after all the (l+1)- candidate sequences have been generated, a new scan of the transactions is started (they are read one-by-one) and the support of these new candidates is determined.

IV. Aprioritid Algorithm

The database is not used at all for counting the support of candidate item-sets after the first pass.

- The candidate item-sets are generated the same way as in Apriori algorithm.
- Another set C' is generated of which each member has the TID of each transaction and the large item-sets present in this transaction. This set is used to count the support of each candidate item-sets [1].

Drawbacks

a) For small problems, AprioriTid did about as well as Apriori, but performance degraded to about twice as slow for large problems.

b) During the initial passes the candidate item sets generated are very large equivalent to the size of the database. Hence the time taken will be equal to that of Apriori. And also it might incur an additional cost if it cannot completely fit into the memory.

V. Apriori Hybrid Algorithm

Apriori performs better than AprioriTid in the initial passes but in the later passes AprioriTid has better performance than Apriori. Due to this reason we can use another algorithm called Apriori Hybrid algorithm [1].

In which Apriori is used in the initial passes but we switch to AprioriTid in the later passes. The switch takes time, but it is still better in most cases.

Estimate the size of C'

$$\sum_{\text{candidates } c \in C_k} \text{support}(c) + \text{number of transactions}$$

Drawbacks

a) An extra cost is incurred when shifting from Apriori to AprioriTid.

b) Suppose at the end of K th pass we decide to switch from Apriori to AprioriTid. Then in the (k+1) pass, after having generated the candidate sets we also have to add the Tids to C'k+1.

VI. Tertius Algorithm

This algorithm finds the rule according to the confirmation measures (P. A. Flach, N. Lachiche 2001). It uses first order logic representation. It includes various option like class Index, classification, confirmation Threshold, confirmation Values, frequency Threshold, horn Clauses, missing Values, negation, noise Threshold, number Literals, repeat Literals, roc Analysis, values Output etc[6].

Drawback

Tertius is its relatively long runtime, which is largely dependent on the number of literals in the rules. Increasing the allowed number of literals increases the runtime exponentially, so we want to keep the maximum to three. Even with an allowed maximum of three literals, the runtime is still quite long - running Tertius can take up to several hours for some of our larger tests.

VII. Conclusion

In this article provided an overview on four different association rule mining algorithms Apriori, AprioriTid, Apriori hybrid and tertius algorithms and their drawbacks which would be helpful to find new solution for the Problems found in these algorithms and also presents a comparison between different association mining algorithms.

References

1. R.Agarwal and R Srikant: "fast algorithms for mining association rules", proc of intl. Conf on VLDB, 1994.
2. Ming-syan chen, Jiawei Han, Philip S.Yu: "Data mining: a overview from Database perspective", IEEE transactions on knowledge and data engineering, vol8, No 6, December 1996.
3. Sergey Brin, Rajeev Motwani, Jeffrey D.Ullman, and Shalom Tsur, Dynamic Itemset Counting and Implication Rules for Market Basket Data, Proceedings of the ACM SIGMOD Conference, PP255-264, 1997.
4. J.S.Park, M.S.Chen, and P.S.Yu : "an effective Hash-Based Algorithm for mining association rules", proc. ACM SIGMOD intl conf. management of data, May 1995
5. Han J, Pei J and Yin Y 2000: " Mining frequent patterns without candidate generation" proc. ACM-Sigmod Intl conf. management of Data.
6. Usma Fayyad, G.P Shapiro, P. Smith, "The KDD process for extracting useful knowledge from volumes of data", communications of the ACM.
7. Ramakrishnan Srikant and Rakesh Agarwal, Mining Quantative Association Rules in large Relational Tables, Proceedings of the !996 ACM SIGMOD International Conference on Management of Data, PP 1-12 Montreal, Quebec, Canada, $-6 June 1996.
8. Ramakrishnan Srikant, Fast Algorithms for Mining Association Rules and sequential Patterns, Ph.D Dissertation, 1996, Universities of Wisconsin, Madison.
9. Jong Soo Park, Phillip S. Yu, Ming- Syan Chen: "Mining Association Rules with Adjustable Accuracy", IBM Research Report, 1997.
10. Ashoka Savasere, Edward Omiecinski, and Shamkant B. Navathe, An Efficient Algorithm for Mining Association Rules in Large databases, Proceedings of the 2nd International Conference on Very Large Databases, PP. 432-444, Zurich, Switzerland, 1995.
11. http://en.wikipedia.org/wiki/Association_ rule_learning
12. http://associationrule.blogspot.in/2008/09/apriori-aprioritid-and-apriori-hybrid.html

Cyber Crime, Culture and Security

Nidhi Goel*

Abstract

Oxford defines the Cyber crime as the criminal activities carried out by means of computer or internet. Cyber crime can be simple one such as hacking a facebook account and big as using that Id to do some unethical things or fraud. The Cyber crime came into light in the year 1820 by Joseph Marie Jacquard. The various types of cybercrime recorded are hacking, fraud, offensive content, harassment, threats, drug trafficking. Cyber law contains the legal issues and punishment which are associated with each of the cyber crimes. As all the people use the technology so there is need for security and hence the term "Cyber Security" came into picture.

Key Words: Cyber crime, cyber security.

I. Introduction

Cyber crime is the term which is used to define the crime which will happen on the internet. Cyber crime refers to the exploitation of the internet. . Dr. Debarati Halder and Dr. K. Jaishankar (2011) defines Cybercrimes as: "Offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm to the victim directly or indirectly, using modern telecommunication networks such as Internet (Chat rooms, emails, notice boards and groups) and mobile phones (SMS/MMS)". These types of crimes may threaten the nation's security and as well as affect the individual privacy such as their facebook or Gmail account. Issues surrounding these types of crimes have become high-profile, particularly those surrounding cracking, copyright infringement, child pornography, and child grooming. There are also problems of privacy when confidential information is lost or intercepted, lawfully or otherwise.

Cyber crime encompasses any criminal act dealing with computers and networks (called hacking). Cyber crime includes the crimes which are conducted through the Internet. For example; hate crimes, telemarketing and Internet fraud, identity theft, and credit card account thefts are considered to be cyber crimes when the illegal activities are committed through the use of a computer and the Internet.

Cyber crime means the criminal activities done with the help of computers, internet, cyberspace or worldwide web. The term cyber crime is not used by any institute such as Indian Penal Code even after its amendment by the Information Technology Act 2008. But "Cyber Security" is defined under Section (2) (b) means protecting information, equipment, devices computer, computer resource, communication device and information stored therein from unauthorized access, use, disclosure, disruption, modification or destruction.

Cyber terrorism is the use of Internet based attacks in terrorist activities, including acts of deliberate, large-scale disruption of computer networks, especially of personal computers attached to the Internet, by the means of tools such as computer viruses. Cyber Terrorism leads to affect the nation's security and functioning of its government. Let's take an example of the cyber crime from the bollywood, in 2012, a movie named "Players" come in which there is a group of the different kind of people who are masters in their works. The team includes one person who is an expert hacker of the world. The team plans to steal the gold worth Rs 10000 crores from a moving train of the

Nidhi Goel*
Assistant Professor,
Department of IT, IITM, Janakpuri, New Delhi

Russian government. So for this purpose the hacker hacks the Russian satellite system which is used to keep track of the train and the hacker is able to hack the system so that other people can do their work. In this way Cyber crime effects the security of the nation and can give rise to the thefts, frauds etc.

II. History

The evolution of the cybercrime happens in the year 1820 when Joseph Marie Jacquard, a textile manufacturer in France, produces the loom. This device allowed the repetition of a series of steps in the weaving of the special fabrics. This resulted in a fear amongst Jacquard's employees that their traditional employment and livelihood were being threatened. They committed act of sabotage to discourage Jacquard from further use of the new technology. This is the first recorded cyber crime.

III. Types of Cybercrime

There are various cyber crimes which can be done through the help of using the computers and network. These consist of the specific crime which is associated with the specific victim. Cyber terrorism focuses upon the use of the Internet by nonstate actors to affect a nation's economic and technological infrastructure. Since the September 11 attacks of 2001, public awareness of the threat of cyber terrorism has grown dramatically.

The crimes consist of:-

A. Fraud

Computer fraud is any dishonest misrepresentation of fact intended to let another to do or refrain from doing something which causes loss. In this context, the fraud will result in obtaining a benefit by:

- Altering in an unauthorized way. This requires little technical expertise and is not an uncommon form of theft by employees altering the data before entry or entering false data, or by entering unauthorized instructions or using unauthorized processes;
- Altering, destroying, suppressing, or stealing output, usually to conceal unauthorized transactions: this is difficult to detect;
- Altering or deleting stored data;
- Altering or misusing existing system tools or software packages, or altering or writing code for fraudulent purposes.

Other forms of fraud may be facilitated using computer systems, including bank fraud, identity theft, extortion, and theft of classified information.

B. Obscene or offensive content

The content of websites and other electronic communications may be distasteful, obscene or offensive for a variety of reasons. In some instances these communications may be illegal.

Over 25 jurisdictions within the USA place limits on certain speech and ban racist, blasphemous, politically subversive, libelous or slanderous, seditious, or inflammatory material that tends to incite hate crimes.

The extent to which these communications are unlawful varies greatly between countries, and even within nations. It is a sensitive area in which the courts can become involved in arbitrating between groups with strong beliefs.

One area of Internet pornography that has been the target of the strongest efforts at curtailment is child pornography.

C. Harassment

Whereas content may be offensive in a non-specific way, harassment directs obscenities and derogatory comments at specific individuals focusing for example on gender, race, religion, nationality, sexual orientation. This often occurs in chat rooms, through newsgroups, and by sending hate e-mail to interested parties (see cyber bullying, cyber stalking, hate crime, Online predator, and stalking). Any comment that may be found derogatory or offensive is considered harassment.

D. Threats

Although freedom of speech is protected by law in most democratic societies, it does not include all types of speech. In fact spoken or written "true threat" speech/text is criminalized because of "intent to harm or intimidate", that also applies for online or any type of network related threats in written text or speech. The US Supreme Court definition of “true threat” is “statements where the speaker means to communicate a serious expression of intent to commit an act of unlawful violence to a particular individual or group”.

E. Drug trafficking

Drug traffickers are increasingly taking advantage of the internet to sell their illegal substances through encrypted e-mail and other Internet Technology. Some drug traffickers arrange deals at internet cafes, use courier Web sites to track illegal packages of pills, and swap recipes for amphetamines in restricted-access chat rooms.

The rise in Internet drug trades could also be attributed to the lack of face-to-face communication. These virtual exchanges allow more intimidated individuals to more comfortably purchase illegal drugs. The sketchy effects that are often associated with drug trades are severely minimized and the filtering process that comes with physical interaction fades away.

IV. Cases of the Cyber Crime

A. Two arrested for cybercrime

Visakhapatnam: Two persons were arrested in two separate instances of causes related to cybercrime in the city on Monday. In the first case, 31-year-old Karasu Sampath Kumar was sending threatening and abusive messages to a woman residing at Kancharapalem, following which she lodged a complaint with the police station. The cybercrime arrested the accused at Seethammadhara and seized one cell phone. In another case, the cybercrime arrested 49-year-old KRPV Rama Krishna Rao residing in Sagarnagar.

B. Student arrested for posting obscene content

Aurangabad: The cybercrime branch of the police on Monday night arrested a 27-year-old computer science post-graduate student for allegedly creating a youth’s fake profile on a social networking site by using a cell phone and uploading obscene pictures of his sister.

The arrested accused has been identified as Sunil Sonuji Gaikwad, a resident of Barshi Naka in Beed. The accused has been remanded in police custody till January 24. The police have also seized the cell phone, which was used for uploading the obscene content.

Cybercrime branch Chief Gautam Patare said, “The complainant and the accused both hailing from Beed were friends but sometime back, the relations between the two soured and they parted ways. The victim shifted to Aurangabad for studies. He recently came to know that a fake profile on his name has been created on a social networking site and pictures of his sister with obscene messages are being uploaded.”

Disturbed by the content which was uploaded on the website, the youth approached the cybercrime branch, which launched a probe into the matter. “The IP address from which the obscene content was being uploaded led us to the accused, who was found using his high-end Smartphone for uploading the obscene content. We have seized the cell phone from his possession,” Patare said.

After gathering enough evidence against the accused, the cybercrime branch officials rounded him up on late Monday night and brought him to the city before arresting him. The accused has been booked under sections of Information Technology Act, 2000.

C. Net banking fraud most common cybercrime

Gurgaon: Higher internet connectivity in Millennium City has also given cyber crimes a bigger playing field. Of the total 759 cases of online crime recorded in the past year by the police, 248 were related to bank fraud, making it the most common among cybercrime cases registered by the police.

According to data released on Saturday, a total of 759 complaints in the past year have kept the cybercrime cell busy. Of these, 248 are related to net banking/credit/debit card fraud, followed by 72 cases of abuse on social networking sites, 68 cases of email ID hacking and 59 complaints of SMS and call abuse.

Increased number of complaints on abuse on the internet led the police to inaugurate a new cybercrime cell in the city earlier this year. Officials blame the alarming numbers on advanced technological expertise and ease of access to information. “Unlike in the preceding years, this year the number of net banking-related crimes have surpassed social media-related complaints. The vulnerability of credit and debit cards and net banking has made it easy for criminals who are just a click away from easy money. They are difficult to investigate and this is why some of the most trying cases this year belong to this category," an officer of the cybercrime cell said.

Yet, it was this year that Inspector Suresh Kumar, head of the cybercrime cell, was adjudged cyber cop of year for his work on cases of online banking fraud. Under his leadership, the department continues to spread awareness on how to beat web criminals.

The cell has approached the problem by taking students into the loop and conducting awareness drives, since they are some of the most frequent users of the Internet and easiest targets.

V. Preventive Measures

The beginning of the new year is the ideal time to make resolutions which should also include your internet habits. Listed below are a number of suggestions that can help prevent your email address from becoming a target to spammers.

- Never respond to spam. If you reply, even to request removing your email address from the mailing list, you are confirming that your e-mail address is valid and the spam has been successfully delivered to your inbox. Lists of confirmed email address are more valuable to spammers than unconfirmed lists, and are frequently bought and sold by spammers.
- Check to see your email address is visible to spammers by typing it on a web search engine. If your email address is posted on any websites or newsgroups, remove it if possible to help reduce how much spam you receive, disable in-line images or do not open spam messages. Frequently spam messages include "web beacons" enabling the spammer to determine how many, or which email addresses have received and opened this message. Most current email programs disable in line images by default to prevent this from occurring.
- Don't click on the links in spam messages, including unsubscribed links. These frequently contain a code that identifies the email-address of the receiptant, and can confirm the spam has been delivered and that you responded.
- When filling in web forms, check the site privacy policy to ensure it will not be sold or passed on to other companies. There may be a checkbox to opt out of third party mailings.

VI. Cyber Security

Cyber security is information security as applied to computers and computer networks.

Nowadays the cyber security is one of the important emerging topic rises in technology world. In the 21st century where each and every work in the office has been converted from the pen& paper to the computers, each organization use the internet technology for the proper functioning of their daily work like use of the emails by the company members to transfer the information from one source to the other, or the make use of the video conferencing in order to conduct meeting while at the remote place. But due to the various hackers available in the it might be possible that company's important information might be at risk, so for this purpose we need the security i.e. called "Cyber Security"

Governments, military, corporations, financial institutions, hospitals and other businesses collect, process and store a great deal of confidential information on computers and transmit that data across networks to other computers. With the growing volume and sophistication of cyber attacks, ongoing attention is required to protect sensitive business and personal information, as well as safeguard national security.

Cyber security is information security as applied to computers and computer networks.

The field covers all the processes and mechanisms by which computer-based equipment, information and services are protected from unintended or unauthorized access, change or destruction. Computer security also includes protection from unplanned events and natural disasters.

Cyber security, also referred to as information technology security, focuses on protecting computers, networks, programs and data from unintended or unauthorized access, change or destruction.

A cybersecurity plan is critical to highly sensitive company information, such as U.S. Department of Defense or associated federal agency data.

User cybersecurity may be employed in the following ways:

- Continuous antivirus software updates
- Strong passwords
- Never disclosing personal information

Many aspects of our lives rely on the Internet and computers, including communications (email, cell phones, texting), transportation (traffic control signals, car engine systems, airplane navigation), government (birth/death records, social security, licensing, tax records), finance (bank accounts, loans, electronic paychecks), medicine (equipment, medical records), and education (virtual classrooms, online report cards, research).

Cyber security involves protecting the information and systems we rely on every day- whether at home, work or school.

There are three core principles of cyber security:

Confidentiality, Integrity, and Availability.

- Confidentiality: Information which is sensitive or confidential must remain so and be shared only with appropriate users.
- Integrity: Information must retain its integrity and not be altered from its original state.
- Availability: Information and systems must be available to those who need it.

For example, your confidential medical records should be released only to those people or organizations (i.e. doctor, hospital, insurance, government agency, you) authorized to see it (confidentiality); the records should be well protected so that no one can change the information without authorization (integrity); and the records should be available and accessible to authorized users (availability).

Various risk associated:-

There are many risks, some more serious than others. Some examples of how your computer and systems could be affected by a cyber security incident — whether because of improper cyber security controls, manmade or natural disasters, or malicious users wreaking havoc—include the following:

- Denial-of- service: It refers to an attack that successfully prevents or impairs the authorized functionality of networks, systems or applications by exhausting resources. What impact could a denial-of-service have if it shut down a government agency's website, thereby preventing citizens from accessing information or completing transactions? What financial impact might a denial-of-service have on a business? What would the impact be on critical services such as emergency medical systems, police communications or air traffic control? Can some of these be unavailable for a week, a day, or even an hour?
- Malware, worms, and Trojan horses: These spread by email, instant messaging, malicious websites, and infected non-malicious websites. Some websites will automatically download the malware without the user's knowledge or intervention. This is known as a "drive-by download." Other methods will require the users to click on a link or button.
- Botnets and zombies: A botnet, short for robot network, is an aggregation of compromised computers that are connected to a central "controller." The compromised computers are often referred to as "zombies." These threats will continue to proliferate as the attack techniques evolve and become available to a broader audience, with less technical knowledge required to launch successful attacks. Botnets designed to steal data are improving their encryption capabilities and thus becoming more difficult to detect.
- "Scareware" –fake security software warnings:This type of scam can be particularly profitable for cyber criminals, as many users believe the pop-up warnings telling them their system is infected and are lured into downloading and paying for the special software to "protect" their system.

- Social Network Attacks: Social network attacks are major sources of attacks because of the volume of users and the amount of personal information that is posted. Users' inherent trust in their online friends is what makes these networks a prime target. For example, users may be prompted to follow a link on someone's page, which could bring users to a malicious website.

VII. Conclusions

This paper discuss about cyber crimes, its evolution, types of cyber crimes. It also states the various cases of cyber crime. The preventive measures to be taken. Finally it throws light on cyber security.

References

1. http://en.wikipedia.org/wiki/Computer_crime
2. http://www.brookings.edu/~/media/research/files/papers/2013/09/19%20cybersecurity%20and%20trade%20global%20local%20friedman/brookingscybersecuritynew.pdf
3. http://www.cspri.seas.gwu.edu/uploads/2/1/3/2/21324690/research_summary.pdf

Secure Swipe Machine with the Help of Biometric Security

Sushma Malik*

Abstract

In this modern scientific world, technologies are transforming rapidly but along with the effortlessness and comfort they also bring in a big concern for security. Security has become a great issue in these days for everyone. Security is a major issue in electronic data capture machine also known as swipe machine transaction with the wide spread utilization of electronic transactions. The primary aim of this research paper is to design an EDC that will improve the authentication of user using Electronic Data Capture machine. EDC machine use a magnetic card reader that is Credit Card or Debit card. The user is identifying by swiping a magnetic card with the EDC machine that contain unique information such as card number, card holder name, expiry date etc.

The user can pay shopping bills at malls, at petrol pumps and many more transaction through EDC by entering a PIN (Personal Identification Number). An invader can read the information of the magnetic strip of a credit or debit card by installing a skimming device on EDC (or swipe machine). Cases of card fraud are another problem once the user's bank card is missing and the password is stolen, or simply steal a customer's card& PIN the criminal will draw all cash with in a very short period of time, which will cause financial loss to customer and this type of fraud has increase globally. The primary aim of this paper is to design a system of EDC that improve the authentication of user by adding biometric with the traditional PIN. In this project, a fingerprint biometric technique is implanted with PIN (personal identification number) to authenticate a user and for enhancing the security of electronic fund transfer via EDC.

Key Words: Biometric, EDC, fingerprint, Electronic Data Capture Machine

I. Introduction

The way of transaction has changed due to rapid change of the banking system. For paying shopping bills, paying at petrol pumps, paying restaurants bills and hotel bills, user prefer to use credit and debit card instead to carry the cash for the payment. Credit and debit cards provide customers a quick and easy way to access their accounts and to carry out a financial transaction. [2] For these type of transaction EDC (Electronic data capture) machines (generally known as swipe machine) are used. Magnetic strip of credit or debit card is swipe through the EDC machine to complete the transaction. Magnetic strip contains all information about the customer. EDC machine reads the magnetic strip and authenticate the customer to complete the transaction after verifying the card number, expiration date, and other details of the customer. The only authentication is used in these type of transactions are PIN number. Progress of technology gives a way to hackers and criminals for fraud. [3] Intruder can use skimming device to read the information of the magnetic strip. The skimming device reads the card's magnetic strips before it enters into original EDC machine slot and customer PIN number is either observed by a person or a hidden camera which is pointed at the keypad. To overcome this problem in money transaction, I proposed the idea of using the finger biometric of the user into the

Sushma Malik*
Assistant Professor, IT Department
IITM, Janak Puri, New Delhi

EDC (Swipe machine) along with the traditional PIN number.

Biometric security gives the solution for that kind of frauds. Biometric authentication of a user example fingerprints, face recognition, retina scan, hand geometry, digital signature and voice recognition gives a way to authenticate the user which cannot be copied or stolen. Biometric security is more secure for user's financial transactions and provides more protection from the fraud.

Using biometric identity verification and authentication procedure we can get more and more monetary security and protection from all the thefts and frauds.

II. Literature Review:

According to Samir Nanavati [1], 80 percent of finger-scan technologies are based on minutiae matching but that pattern matching is a leading alternative. This technology bases its feature extraction and template generation on a series of ridges, as opposed to discrete points. The use of multiple ridges reduces dependence on minutiae points, which tend to be affected by wear and tear [2].

Bazen et al. [3] proposed a three step correlation based fingerprint verification system. In the first step small size templates are selected in the primary reference fingerprint. In the second step template matching is used to find the position in the secondary (query) fingerprint image at which the template match the best. Finally, the third step aims to compare the template position in both the fingerprints to make the decision regarding the authenticity.

Cavusoglu et al. [4] proposed a robust correlation based fingerprint matching algorithm. The proposed algorithm requires segmentation, ridge orientation, reference point detection and normalized operation before the application of correlation algorithm. During the enrollment stage starting from the selected reference (core) point of the template image a set of features is obtained with deferent radius (r) and angle è. For the authentication purpose the features of the input query image are obtained by rotating the query image with an incremental size of 1° in the given range of -15° to + 15°. For each rotation the normalized cross correlation values of both images are calculated. The maximum value of cross correlation in the given range, determine the similarity of the query and template image. This method is efficient in terms of storage since instead of template (reference image or part of reference image) the features of the template are stored but this method requires the accurate detection of core point which is a trivial task. Moreover, this method also fails in case the core point is not present in the fingerprint image.

A fingerprint is the pattern of ridges and valleys on the surface of a fingertip, the formation of which is determined during the first seven months of fetal development. Fingerprints of identical twins are different and so are the prints on each finger of the same person. One problem with the current fingerprint recognition systems is that they require a large amount of computational resources, especially when operating in the identification mode. Finally, fingerprints of a small fraction of the population may be unsuitable for automatic identification because of genetic factors, aging, environmental, or occupational reasons [5].

III. Biometric:

Biometric is the branch of Science to verifying the identity of an individual through psychological or Behavioral characteristics because the biometric characteristics are associated with the user throughout the life.

Biometrics offers several advantages over traditional security measures. These include:

1. **Identification is accurate:** Unlike other security systems that rely on passwords or smart cards, one of the greatest advantages of biometrics is the accuracy it provides. When the system is set up correctly, biological characteristics like fingerprints and retinal scans provide completely unique data sets that cannot be replicated easily. This makes it very difficult for anyone but an authorized user to gain access without permission.[6]

Table1: Comparison of Biometric and Password/Key based authentication [7]

Biometric Authentication	**Password/Key based authentication**
Based on physiological measurements or behavioral traits	Based on something that the user ’has’ or ’knows’
Authenticates the user	Authenticates the password/key
Is permanently associated with the user	Can be lent, lost or stolen
Utilizes probabilistic matching	Requires exact match for authentication

2. **Non-repudiation:** With token and password based approaches, the perpetrator can always deny committing the crime pleading that his/her password or ID was stolen or compromised even when confronted with an electronic audit trail. There is no way in which his claim can be verified effectively. This is known as the problem of deniability or of ’repudiation’. However, biometrics is indefinitely associated with a user and hence it cannot be lent or stolen making such repudiation infeasible [7].

3. **Screening:** In screening applications, we are interested in preventing the users from assuming multiple identities (e.g. a terrorist using multiple passports to enter a foreign country). This requires that we ensure a person has not already enrolled under another assumed identity before adding his new record into the database. Such screening is not possible using traditional authentication mechanisms and biometrics provides the only available solution [7].

4. **Biometrics methods offer convenience**: Some other advantages of biometrics products concern the efficiency and convenience they lend to access control. Passwords and pins are easily forgotten, can be written down and subsequently stolen, and are sometime hacked. Once obtained, they can be easily used by someone other than the authorized person. Smart cards and keys can similarly be lost or stolen, and can also be used by an imposter without detection. But with biometrics, something like fingerprints won’t be lost and can’t be easily obtained and replicated by someone trying to illicitly gain access [6].

5. **Systems are user friendly:** Once they’ve been installed and implemented, biometrics systems are able to identify people very rapidly, uniformly, and reliably. Typically, only minimum training is needed to get the system operational, and there’s no need for expensive password administrators. In addition, high-quality systems don’t tend to need a large amount of maintenance, further cutting costs.[6]

In our research we use the fingerprint as a biometric. These are the some advantages to choose the fingerprint as a biometric tool for accessing EDC [7]:

1) **High universality**: A large majority of the human population has legible fingerprints and can therefore be easily authenticated. This exceeds the extent of the population who possess passports, ID cards or any other form of tokens.
2) **High uniqueness**: Even identical twins who share the same DNA have been shown to have different fingerprints, since the ridge structure on the finger is not encoded in the genes of an individual.
3) **High permanence**: The ridge patterns on the surface of the finger are formed in the womb and remain invariant until death except in the case of severe burns or deep physical injuries.
4) **Easy collectability**: The process of collecting fingerprints has become very easy with the advent of online sensors. These sensors are capable of capturing high resolution images of the finger surface within a matter of seconds.
5) **High performance**: Fingerprints remain one of the most accurate biometric modalities available to date.
6) **Wide acceptability**: While a minority of the user population is reluctant to give their fingerprints

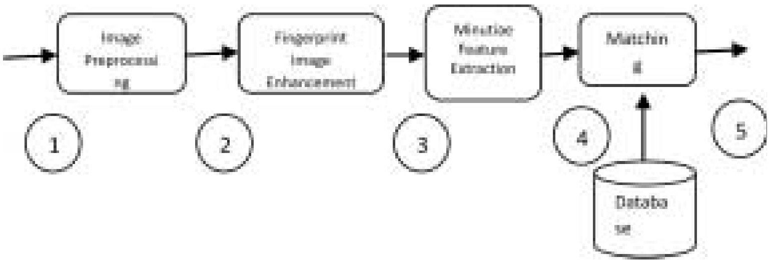


Figure 1: Architecture of Fingerprint Biometric

due to the association with criminal and forensic fingerprint databases, it is by far the most widely used modality for biometric authentication.

7) **Illiterate and old age person:** To sign a document is necessity for every person. But due to some reasons either illiterate or very old age person become incapable of sign a documents. In this case person can use biometric thumb impression to secure their identity.

8) **To identify the missing person:** Police and Several agencies which are working to help missing persons can use biometric thumb impression to identify the identity of missing person.

IV. Architecture of Fingerprint Biometric:

The various stages of a typical fingerprint recognition system are shown in Figure 1. The fingerprint image is acquired using the image reader [7]. It includes the following steps:

1) Read the image from the image reader
2) Good Quality image after image preprocessing.
3) Good quality fingerprint image after image enhancement.
4) Minute features of fingerprint are extract and store in the Database
5) Authentication of access after matching the pre stored minutes of fingerprint in the database

V. Proposed System

In this paper we mainly focus the new authentication approach of PIN with finger biometric through the Biometric device which is attached with the EDC.

It includes the two phases:

a) Enrollment
b) Authentication

Enrollment phase is basically used for the registration of the new user in the Database. In this phase read the fingerprint image from the biometric Scanner at the user side. Biometric Scanner device convert the scanned image into digital form. It stores the

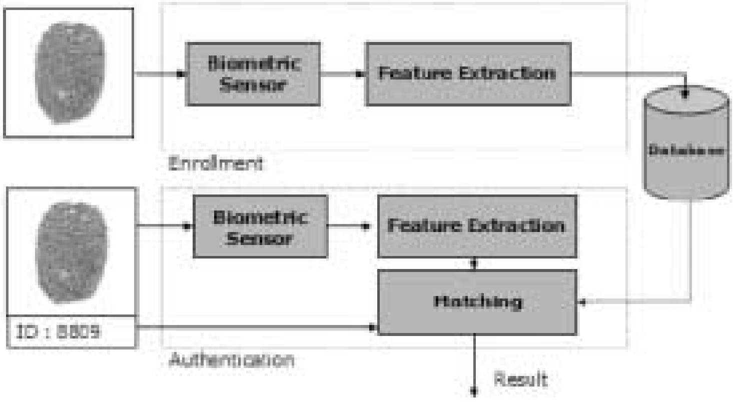


Figure 2: General architecture of a biometric system [7]

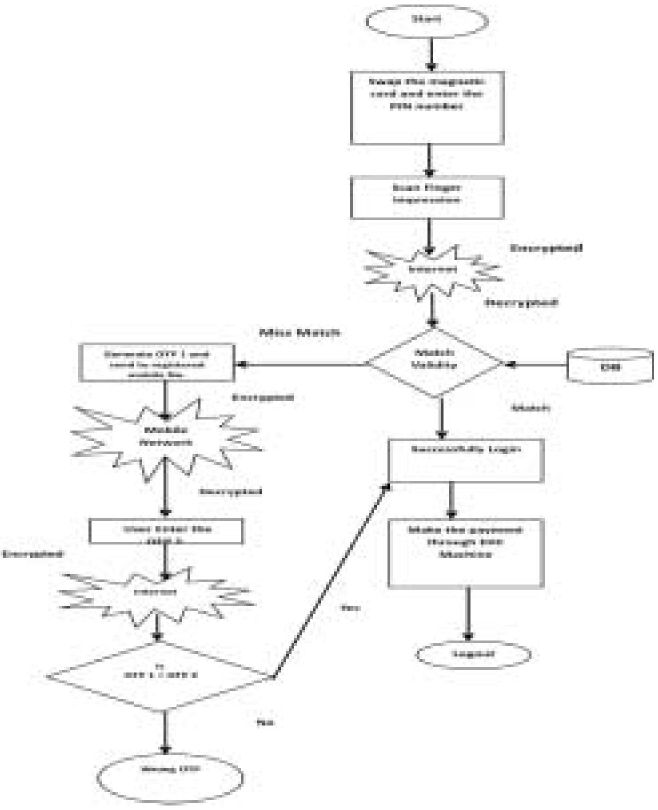


Figure 3: Flow Chart of Authentication

fingerprint features and user mobile number in encrypted form in the database.

During Authentication phase, two types of authentication is fingerprint match and another one is OTP (One Time Password) matching which is supported with the mobile Network.

Algorithm:

1) Swap the magnetic card (Credit or Debit) into swapping machine and enter the PIN number.
2) User finger impression read by the image scanner biometric device.
3) Data move through internet in encrypted form using Advanced Encryption Standard (AES) algorithm.
4) Check if the information read by scanner through magnetic card, PIN entered by user and finger impression match the pre stored credentials in the data base.
5) If true, set session and redirect user to the successfully payment through swipe machine.
6) If false, than generate an OTP and send to the registered mobile number of the user, and ask user to enter generated OTP.
7) Generated OTP move through mobile network in encrypted form by using OCRA (Online Authentication Challenge Replace Authentication) type of OTP type.
8) If entered OTP match the generated OTP, set session and redirect user to the payment through swipe machine.
9) If OTP not match, print a failure message and redirect on the same login page.
10) Again swipe the magnetic card, enter the PIN number and scan the finger to swipe machine for the payment.

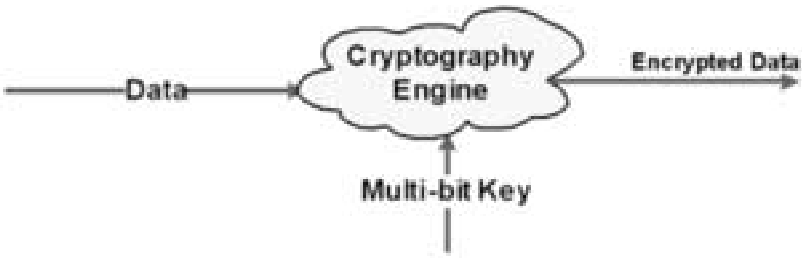


Figure 4: Multi-bit key to encrypt data using cryptographic algorithm

Encryption Algorithms for Image

There are many image encryption algorithms are available like:

1) **DATA ENCRYPTION STANDARD (DES):** It was developed by an IBM team around 1974 and adopted as a national standard in 1997. DES is a 64-bit block cipher under 56-bit key. The algorithm processes with an initial permutation, sixteen rounds block cipher and a final permutation. There has been considerable controversy over the design, particularly in the choice of a 56-bit key [12].

2) **TRIPLE DES (TDES):** The triple DES uses three round message This provides TDES as a strongest encryption algorithm since it is extremely hard to break 2^168 possible combinations. Another option is to use two different keys for the encryption algorithm. This reduces the memory requirement of keys in TDES. The disadvantage of this algorithm is that it is too time consuming [12].

3) **ADVANCED ENCRYPTION STANDARD (AES):** AES was developed by two scientists Joan and Vincent Rijmen in 2000. AES uses the Rijndael block cipher. Rijndael key and block length can be 128, 192 or 256-bits. If both the key-length and block length are 128-bit, Rijndael will perform 9 processing rounds. If the block or key is 192-bit, it performs 11 processing rounds. If either is 256-bit, Rijndael performs 13 processing rounds [12].

4) **BLOWFISH:** Bruce Schneier designed blowfish in 1993 as fast, free encryption algorithms. The Blowfish algorithm has many advantages. It is suitable and efficient for hardware implementation and no license is required. Blowfish is a 64 bit block cipher and is suggested as a replacement for DES. Blowfish is a fast algorithm and can encrypt data on 32-bit microprocessors [12].

In the world of embedded and computer security, one of the often debated topics is whether 128-bit symmetric key, used for AES (Advanced Encryption Standard) is computationally secure against brute-force attack. Governments and businesses place a great deal of faith in the belief that AES is so secure that its security key can never be broken, despite some of the inherent flaws in AES.

Any cryptographic algorithm requires multi-bit key to encrypt the data as shown in Figure 1.

The key length used in the encryption determines the practical feasibility of performing a brute-force attack,

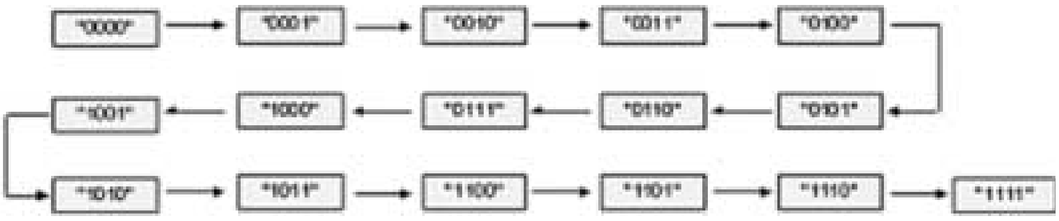
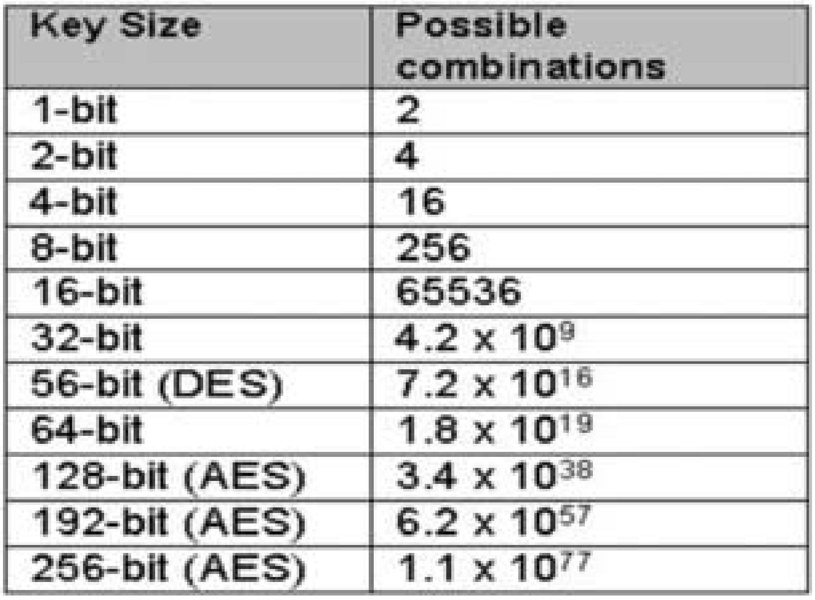


Figure 5: Brute Force attack on 4-bit key

Table 2: Key combinations versus Key size



Key Size	Possible combinations
1-bit	2
2-bit	4
4-bit	16
8-bit	256
16-bit	65536
32-bit	4.2 x 10^{9}
56-bit (DES)	7.2 x 10^{16}
64-bit	1.8 x 10^{19}
128-bit (AES)	3.4 x 10^{38}
192-bit (AES)	6.2 x 10^{57}
256-bit (AES)	1.1 x 10^{77}

with longer keys exponentially more difficult to crack than shorter ones.

Brute-force attack involves systematically checking all possible key combinations until the correct key is found and is one way to attack when it is not possible to take advantage of other weaknesses in an encryption system.

Here is an example of a brute force attack on a 4-bit key:

As shown, it will take a maximum 16 rounds to check every possible key combination starting with "0000." Given sufficient time, a brute force attack is capable of cracking any known algorithm.

The following table just shows the possible number of key combinations with respect to key size:

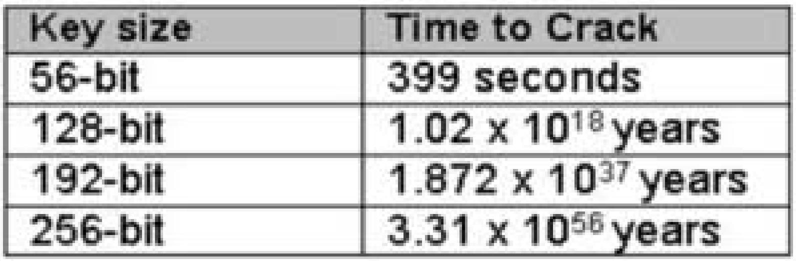
Notice the exponential increase in possible combinations as the key size increases. "DES" is part of a symmetric cryptographic algorithm with a key size of 56 bits that has been cracked in the past using brute force attack.

There is also a physical argument that a 128-bit symmetric key is computationally secure against brute-force attack. Just consider the following:

Faster supercomputer (as per Wikipedia): 10.51 Pentaflops = 10.51 x 1015 Flops [Flops = Floating point operations per second]

No. of Flops required per combination check: 1000 (very optimistic but just assume for now)

No. of combination checks per second = (10.51 x 1015) / 1000 = 10.51 x 1012



Key size	Time to Crack
56-bit	399 seconds
128-bit	1.02 x 10^{18} years
192-bit	1.872 x 10^{37} years
256-bit	3.31 x 10^{56} years

Table 3: Time to crack Cryptographic Key versus Key size

No. of seconds in one Year = 365 x 24 x 60 x 60 = 31536000

No. of Years to crack AES with 128-bit Key = (3.4 x 1038) / [(10.51 x 1012) x 31536000] =(0.323x1026)/ 31536000

$=1.02x10^{18}$

= 1 billion years

As shown above, even with a supercomputer, it would take 1 billion billion years to crack the 128-bit AES key using brute force attack. This is more than the age of the universe (13.75 billion years). If one were to assume that a computing system existed that could recover a DES key in a second, it would still take that same machine approximately 149 trillion years to crack a 128-bit AES key[12]

Algorithems for OTP (One Time Passward)

Access controls exist to prevent unauthorized access. Companies should ensure that unauthorized access is not allowed and also authorized users cannot make unnecessary modifications. The controls exist in a variety of forms, from Identification Badges and passwords to access authentication protocols and security measures. There are mainly two types of password:

- **Static password**
- **Dynamic Password**

Static password is the traditional password which is usually changed only when it is necessary: it is changed when the user has to reset the password, i.e., either the user has forgotten the password or the password has expired. Static passwords are highly susceptible to cracking, because passwords used will get cached on the hard drives.

To solve this we developed One Time Password Token. Unlike a static password, dynamic Password is a password which changes every time the user logs in. An OTP is a set of characters that can act as a form identity for one time only. Once the password is used, it is no longer used for any further authentication. Even if the attacker gets the password, it is most likely that it was already used once, as it was being transmitted, thus useless to the attacker. This reduces the vulnerability of the hacker sniffing network traffic, retrieving a password, and to successfully authenticate as an authorized user. This password is used only for that session and when the user logins next time, another password is generated dynamically [11].

Types of OTP [9]:

- HOTP: HMAC OTP, Hash Message Authentication OTP.
- TOTP: Time OTP.
- OCRA: OATH Challenge Response Authentication OTP.

HOTP and TOTP using the two basic things:

a) Shared Secret

b) Moving Factor(Counter)

In the HOTP, Hash is the moving factor, will be generated using the shared secret. This algorithm is event based, meaning that whenever a new OTP is generated, the moving factor will be incremented, and hence the subsequently generated password should be different each time. The generated password is valid for unknown amount of time [10].

TOTP is same as HOTP; however the moving factor works a bit differently. The moving factor constantly changes based on the time passed. TOTP passwords keep on changing and are valid for a short period of time [10].

OCRA is a generalization of HOTP with variable data input s not solely based on an incremented counter and secret key [9].

VI. Limitations:

The main limitations of the fingerprint technology are [8]:

1) Inability to enroll some users. About 2 % of the population has poor quality of fingerprint, especially the elder people and manual worker. For these cases one need to consider other biometrics or any other solution.

2) Association with forensic application. The fingerprint technology has been associating with forensic and this can cause discomfort to some

people. Specially, in the countries where it is not habitual the use of fingerprint.

3) Need to deploy specialized devices. The device needed for fingerprint capture is not yet present on Swiping machine, at sales point.

4) Some time others members of family are also carry the magnetic card (credit or debit) instant of the owner. They are the authenticate persons but due to biometric, now they are not able to use. The physically presence of the owner in necessary.

VII. Conclusion and Future Work

The growth of financial transactions through the electronic fund transfer has result a greater demand for fast and secure user identification and authorization. To access the bank accounts for credit and debit card through traditional PIN number are not to secure. Instead of many warnings, people often choose easily guessing PIN numbers like date of birthdays, phone number digits, area code pin number etc. If the criminal steal a customer's card and PIN, than can be misused the cards and will draw all cash with in a very short period of time, which will cause financial loss to customer and this type of fraud has increase globally. Biometric means verifying the personal identity of the user by measuring and analyzing unique behavioral or physical characteristics like fingerprints, voice, iris etc. The conclusion of this paper is that finger Biometric which provides advance level of security to the critical systems where security is one of the major concerns. The combination of finger biometric with PIN in the swipe machine, become a strong multifactor authentication, which is used to fulfill the requirements of the user as a security level. Through biometric login the EDC become more secured as compare to the earlier login methods. The only authenticated user can be login into the account and easily paying the payment anywhere without any fear of the frauds.

References

1. Samir Nanavati, Michael Thieme, and Raj Nanavati, "Biometrics: Identity Verification in a Networked World", John Wiley & Sons, 2002.
2. Julian Ashbourn, "Biometrics: Advanced Identity Verification", Springer-Verlag, London, 2002
3. A. M. Bazen and S. H. Gerez, "Directional Field Computation for Fingerprints based on the Principal Component Analysis of Local Gradients", Proceedings of 11th Annual Workshopon Circuits, Systems and Signal Processing, Veldhoven, Netherlands, pp. 1-7, November, 2000.
4. A. Cavusoglu and S. Gorgunoglu, "A Robust Correlation based Fingerprint Matching Algorithm for Verification", Journal of Applied Sciences, vol. 7, no. 21, pp. 3286-3291, 2007.
5. Anil k. Jain, fellow, IEEE, Arun Ross, member, IEEE," Biometrics : A Tool for information security" IEEE Transactions on information forensics and security. VOL.1.No.2.June 2006.
6. https://enlightenme.com/advantages-of-biometrics
7. Kommu Ayyanna, "study of some fingerprint verification algorithms", National Institute of Technology, Rourkela
8. http://www.griaulebiometrics.com/enus/book/understanding-biometrics /types /strengthen
9. https://www.protectimus.com/blog/otp-generation-algorithms-and-token-types/
10. http://blogs.forgerock.org/petermajor/2014/02/one-time-passwords-hotp-and-totp/#top
11. Himika Parmar1, Nancy Nainan2 and Sumaiya Thaseen" Generation of secure one-time password based on image authentication" AIRCC Journal, volume 2
12. http://www.eetimes.com/document.asp?doc_id=1279619

Smart Security Protocol for Body Area Networks

M. Junaid Iqbal*
Noor ul Amin**
Nizamuddin***
Arif Iqbal Umar****

Abstract

Secure and authentic patient information communication is pivotal. This paper proposes an efficient and secure protocol. It is suitable for securing small amount of in terms of efficiency and memory consumption. It enhances the security requirement of the system without causing extra overload to the system.

Key Words: Body Sensor Network; memory consumption

I. Introduction

The graph of the popularity of Wireless communication has rise much higher due to their low weight, less cost and quick responsiveness. The major areas where WBAN are used along with healthcare is for assisted living environment and their use or in entertainment and games. A WBSN comprises of several components such as sensors, base stations and medical server.

II. Required Features of Wban.

Confidentiality of Data

It is always important to keep data secure in this regards many algorithms have been designed for the secure transmission of data in WBAN which are able to transmit data safely. Data is encrypted using cryptographic techniques to make the data confidential.

Data Integrity

It is also important to make it sure that original data is transmitted. Data integrity makes it sure that the attacks like data digest cannot be launched and data only from the authentic sender is sent and same reaches at the target. Node authentication and data authentication are necessary to confirm that the data is coming from the right source and not from any hacker.

Energy Saving

WBSN devices have low power low size devices, it is important to use the energy of those sensors in economic way, shortage of energy may result in loosing critical alarms on physical conditions in some crucial sign monitoring applications.

III. Related Work

Liu in [5] gave a hybrid security framework for WBAN. They proposed an algorithm by combining both asymmetric and symmetric algorithms. They gave a concept of tradeoff among security and resource constraints.

Liu et al [4] security is the key and proper security means are needed for body sensor networks to provide confidentiality of data authenticity of information, integrity and availability of resources. It is also required to have a mechanism of key exchange which must be

M. Junaid Iqbal*
sejunidiqbal@gmail.com

Noor ul Amin**
naminhu@gmail.com

Nizamuddin***
sahibzadanizam@yahoo.com

Arif Iqbal Umar****
arifiqbalumar@yahoo.com

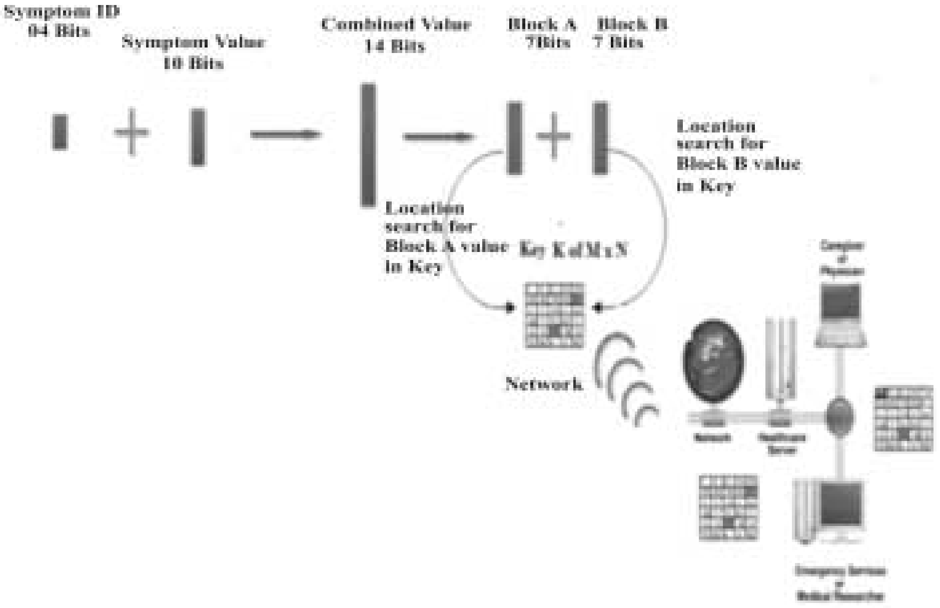


Figure1: Block model of proposed scheme

light weight. Symmetric cryptography can ensure light weight security function but cannot provide sender authentication, while asymmetric cryptography provides all security functions but

Uses resources extensively, so a tradeoff between the two types of cryptography is needed in body sensor networks

[2] used very small key which is easy to be cracked and the vital patient information is at risk. Key size is only 4bits.

Ayushi[8] proposed to design a new technique (that is not cost efficient) for the purpose of encrypting small amount of data our proposed technique can simulate the work as well.

IV. Proposed Solution

The paper focuses on security and proposed change in the algorithm which will not only secure the without additional burden in terms of resource utilization. The algorithm is one step forward towards providing security of the valuable patient data by adding a feature of key refreshing.

The data set consists of symptom id, symptom name and the value against each symptom. In proposed techniques 4 bits are reserved for symptom id , 10 bits for symptom value and a key in size m x n to securely encrypt our message and send it to the medical server. 4 bits of symptom id and 10 bits of the symptom value are concatenated together to form a 14 bit long string. Here a special consideration is made that one 1 symptom id is reserved to send key refresh message to th is receiver. Key of 1025 values is preloaded to both ends and key need not to be transferred to the medical server. Key contains multiple values each of which is < 128. 14 bit concatenated string is divided to equal substrings of 7 bits. Against each substring decimal value is calculated. That value is searched in the key.

The location address is calculated. The location number at which value resides is then transmitted the medical serve. The same way second value is transmitted after the same encryption procedure. At the receiver end reverse of above is performed to do the decryption of the message. When the symptom id is 0 it means it is special message for the key refresh and 2 values are generated and encrypted to be transmitted.

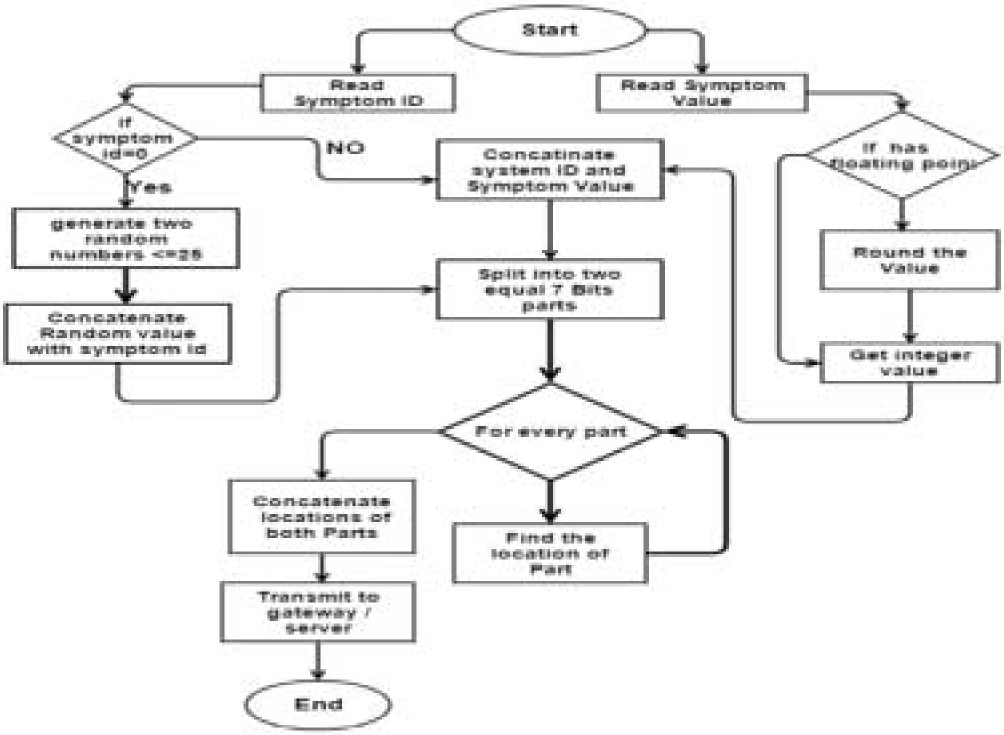

Start
Read Symptom ID
Read Symptom Value
if symptom id=0
NO
Yes
Concatinate system ID and Symptom Value
If has floating point
generate two random numbers <=25
Split into two equal 7 Bits parts
Round the Value
Concatenate Random value with symptom id
Get integer value
For every part
Concatenate locations of both Parts
Find the location of Part
Transmit to gateway / server
End

Figure 2: Process at the sender end.

Key Refresh Algorithm

Although the large size of the keys ensures the security in WBAN to some extent but still there persist some security it is not free from the threat of being hijacked. Therefore it is necessary that key must be refreshed after a specific time in order to protect the valuable patient information from attacker. If transmitted message has symptom id =0 this means the this is the message for key refresh at medical server end and symptom value part of the message contains those random values which have been generated by encryption algorithm this is the key refresh instruction. First random value received in the symptom value part means change first column with the column that comes after adding message part and perform the same operation to every column. Similarly the second random value is the instruction to change the rows in the same way.

a) Encryption Technique

```
Step 1: Start
Step 2: Get device ID/Symptom ID
Step 3: Get value of symptom
        if ( symptom id = = 0)
        Generate 2 random numbers $\leq 25$
        then concatenate symptom id with each random number
        else concatenate symptom id with symptom value
Step 5: Divide the concatenated string into equal parts
        i.e 7 bit each
Step 6: Get input key
        (key: predefined array of 2025 elements where the value of each element < 128)
Step 7: Find the location of message part in the key
        and get the locations
Step 8: Message part length => $\log_2$(Key Length)
Step 9: Represent the binary values of locations in
        message parts length
Step 10: Send binary values
Step 11: Stop
```

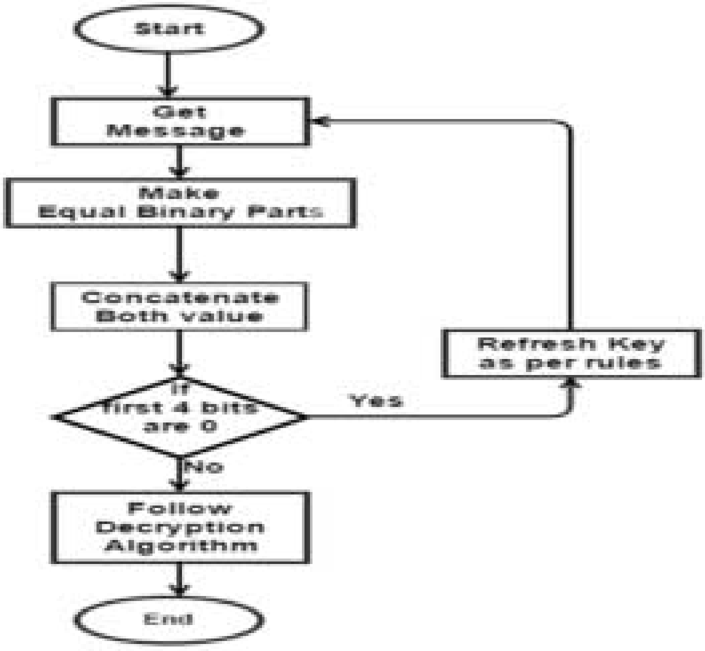


Figure 3: Process at the receiver end.

b) Decryption Technique

```
Step 1: Start
Step 2: Get message
Step 3: Find its binary value and represent it into
        7 bits
Step 4: Concatenate both values
        if ( first 4 bits are 0 )
        Refresh key according to the rules
else got to step 5
Step 5: Get the value from the locations
Step 6: Represent the value in decimal
Step 7: Stop
```

Table 1: Symptom parameters

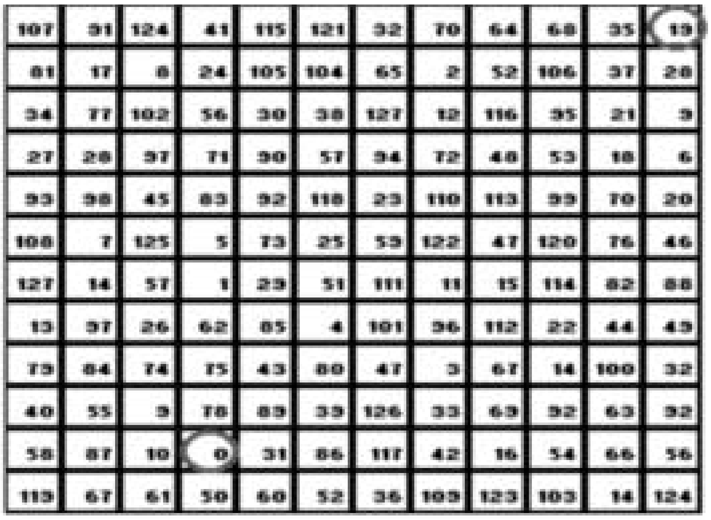
Symp ID	Symptom	SympValue
1	Blood Pressure Upper	120
2	Blood Pressure Lower	80
3	Heart Beat	75
4	Body Temperature	100
5	Pulse Oximetry	70
6	Arterial Gases	79
7	Blood Flow	99
9	Muscle Power	150
10	Respiratory Rate	30
0	Key Refresh	00

Table 2: Guide for notations

Notation	Description
K	Key
Symid	Symptom id
Rnd	Random number
Conc	Concatenate
Str	String
Substr	Substring
Loc 1	Location 1
Loc 2	Location 2
$\overline{K}$	Refreshed Key

At Sender end

```
key Ksympid == 0
Gen Rnd X1, X2
```



107	91	124	41	115	121	32	70	64	68	35	19
81	17	8	24	105	104	65	2	52	106	37	28
34	77	102	56	30	38	127	12	116	95	21	9
27	28	97	71	90	57	94	72	48	53	18	6
93	98	45	83	92	118	23	110	113	99	70	20
108	7	125	5	73	25	59	122	47	120	76	46
127	14	57	1	29	51	111	11	15	114	82	88
13	97	26	62	85	4	101	96	112	22	44	43
79	84	74	75	43	80	47	3	67	14	100	32
40	55	9	78	89	39	126	33	69	92	63	92
58	87	10	0	31	86	117	42	16	54	66	56
113	67	61	50	60	52	36	109	123	103	14	124

Figure 4: A portion of key (key size 45 x 45)

```
Conc symid , X1 = str1
Str1 = substr1, substr2 (Divide intoequal parts)
find decimal value of substr1, substr2
Locate substr1, substr2 in K i.e loc1, loc2
 send loc1, loc2
```

At the receiver end

```
Rec loc1, loc2
find value at loc1, loc2 = Y1, Y2
find binary of Y1, Y2
Conc Y1, Y2
first 4 bit, Symp id, last 10 bits rnd X1
ifsumpid=0
change Col with Y2 sequence
Now repeat same for X2
Rec loc1, loc2
find value at loc1, loc2 = Y1, Y2
find binary of Y1, Y2
Conc Y1, Y2
first 4 bit, Symp id, last 10 bits rnd X2
change row with Y2 sequence
```

We have new key *K* with columns changed with X1 sequence and rows changed with X2 sequence.

Symptom id is obtained from the shared table. Against the symptom value the medical data is sent the medical server. At the receiver end these values are fetched from the key and reverse of the above mentioned operation is performed to get the sent symptom id and symptom value. If the symptom id sent is 0 this is predefined that 0 is meant to change the key order as per sent instruction. During encryption if the symptom id is 0 it means the other part of the message contains instruction for the key refresh.

Example

Step 1: Get Symptom ID

0	0	0	0

Binary of symptom no (4)

Step 2: Generate two random numbers < 25

X1= 19 and X 2= 23

For X1

10 bit binary of 19

0	0	0	0	0	1	0	0	1	1

Step 3: Concatenate both

0	0	0	0	0	0	0	0	0	1	0	0	1	1

Symptom ID Symptom Value

04 Bits 10 Bits

Step 4: Divide 14 Bits in equal blocks of 7bits

0	0	0	0	0	0	0		0	0	1	0	0	1	1

7 Bits value = 0 7Bits value = 19

Step 4: Get input key of size m x n

Step 5: 0 is located at loc 47 and 19 is located at 133. Send both loc1 and loc 2 to the receiver.

At Receiver End

Step 1 : Get loc1 and loc 2

loc1 47

loc2 133

Step 2: Find values at loc1 and loc 2

0 is located at 47

19 is located at 133

Step 3: convert both loc1 and loc 2 into binary

loc 1

0	0	0	0	0	0	0

loc 2

0	0	1	0	0	1	1

Step 4: Concatenate loc 1 and loc 2

0	0	0	0	0	0	0		0	0	1	0	0	1	1

Step 5:

As the first 4 bits are so refresh the key according to rule predefined that

V. Performance Analysis

As compared to exiting algorithm for small amount of data our algorithm is much more efficient .The processing speed of our algorithm is faster and further more it shows more resistance against brute – force attack. Our block size is very small as compared to existing techniques [2] and key is not to be transferred thus saving 25% of computation cost, as most of the energy is consumed during transmission and we only need to send 28 bit which is very less as compared to previous techniques, and security is additional feature by having a refreshing key after specific amount of time with routine message.

VI. Conclusions

Our encryption and key refreshing technique provides security for transmission of small data. The length of the key enhances the security without being overload as key is pre shared and not needed to be transmitted through link. Only two instructions are enough to shuffle the key after specific interval of time. The binary conversion, concatenation and then division in other order increase toughness for any unauthorized user to access the data.

References

1. Amin, N. Asad, M. Nizamuddin, and Chaudhry, S. A. (2012) An Authenticated Key Agreement with Re keying for Secured Body Sensor Networks Based on Hybrid crypto system" 9th IEEE International Conference on Networking, Sensing and Control, pp: 118 121

2. Ayushi, "A Symmetric Key Cryptographic Algorithm," International Journal of Computer Applications Volume 1 – No. 15, 2010

3. Coppersmith, D, "The Data Encryption Standard (DES) and Its Strength against Attacks," IBM Journal of Research and Development May 1994.

4. E. Monto´n, J.F. Hernandez, J.M. Blasco, T. Herve´, J. Micallef, I. Grech, A. Brincat and V. Traver "Body area network for wireless patient monitoring" IET Commun., 2008.

5. Hu, Chunqiang, et al. "Body area network security: A fuzzy attribute-based signcryption scheme." *Selected Areas in Communications, IEEE Journal on* 31.9 (2013): 37-46.

6. He, Daojing, et al. "Secure and lightweight network admission and transmission protocol for body sensor networks." *Biomedical and Health Informatics, IEEE Journal of* 17.3 (2013): 664-674.

7. Kumar, R. Satheesh, et al. "A Novel Approach for Enciphering Data of Smaller Bytes." International Journal of Computer Theory and Engineering 2.4 (2010): 1793-8201.

8. Liu, Jingwei, and Kyung Sup Kwak. "Hybrid security mechanisms for wireless body area networks." *Ubiquitous and Future Networks (ICUFN), 2010 Second International Conference on.* IEEE, 2010.

9. Min Chen SergioGonzalez AthanasiosVasilakos HuasongCao Victor C. M. Leung "Body Area Networks: A Survey" August 2010

10. Satheesh Kumar, E.Pradeep, K.Naveen, R.Gunasekaran "Enhanced Cost Effective Symmetric Key Algorithm for Small Amount of Data," IEEE International Conference on Signal Acquisition and Processing 2010

11. Sarker, M.Z.H. Parvez, M.S. "A Cost Effective Symmetric Key Cryptographic Algorithm for Small Amount of Data," 9th IEEE INMIC, pp: 1-6 , 2005.

12. W. Soomro , Nizamuddin, Arif Iqbal Umar, Noorul Amin, "Secured Symmetric Key Cryptographic Algorithm for Small Amount of Data" ICCET 2013.

13. Rongxing, "SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency" MARCH 2013.

Categorical Study of Nature Inspired Algorithms in the Various Domains

Shruti Goel*

Abstract

Now a days there researchers have shifted there interest from the classical algorithm to new concept of Swarm Intelligence/ Nature Inspired Algorithms. Nature being the powerful paradigm such diversified providing the researcher scope to discover as many new algorithm such as Ant Colony Optimization based on the behaviour of ant to Pigeon Optimization based on the behaviour of Pigeon. But the issues coming in here is which algorithm can be used in a given problem definition. Authors in this paper have attempted to present an abstract view of the successful use of these algorithm in a given environment. This is done based on the research survey performed on the basic characters of each algorithm mentioned. Also, the author has attempt to classify these algorithms into Time-Space Variant and Time-Space Invariant.

Key Words: Swarm Algorithm, Time-Space Variant and Time-Space Invariant, Complexity.

I. Introduction

Computational Intelligence (CI) is the study of Population Based Stochastic Optimization Techniques, modelled on the behavioural aspects in decentralized system. It is the set of various nature inspired Computational methodologies and approaches to address complex real-world problems to which traditional approaches, i.e., first principles modelling or explicit statistical modelling, are ineffective or infeasible. Many real-life problems are not considered to be well-posed problems mathematically, but nature provides many counter examples of biological systems exhibiting the required function, practically. We are going to detailed algorithms which comes under the category of CI, will study their characteristics and discuss its paradigm of applicability.

I. Research Methodologies

Research in Computational Intelligence has gone far beyond neural networks (NN), fuzzy systems (FS) and evolutionary computation (EC), the current development in research such as Agent Behaviour shows that Computational Intelligence is entering the second golden phase. The aim of this paper is to bring together the study of the new upcoming algorithm which can help the academic institutions and R&D Research Labs, working in the related fields. These new algorithms in CI can be applied to various field such as image processing, computer vision, motion analysis and machine learning. Different advancements in these techniques have resulted in wide applicability of Algorithms in varied domains. Researches shows that many population based algorithms has the ability to solve complex problems specially NP hard ones in an extraordinary way. Hence, in order to benefit various industry and IT sector with such advancements study of application of these algorithms is must. Thus, the authors here highlights the applications of the popular algorithms and their characteristics such as their level of convergence based on the global optima and local optima.

II. Nature Inspired algorithms

Nature inspired algorithms are those which are inspired from the natural occurrence such as shifting of plate tectonics, behaviour of ants to work

Shruti Goel*
Shruti Goel Computer Science,
University School of California, USA

Table-I: Classification of Popularly used Nature Inspired Algorithm on various characteristics [7] [8] [9].

Algorithm	Proposed by	Based on	Conver- gence	Applied On	Time-Space Invariant/ Variant
Ant Colony Optimization 1992	Dorigo and Stützle	Food foraging behaviour of ants (Swarm Behavior)	No Premature Convergence	● Clustering ● TSP ● Optimal networks Routing ● Scheduling problems ● QAP ● Vehicle Routing ● Connection-Oriented Network Routing ● Sequential Ordering ● Graph Coloring ● Redundancy allocation ● Constraint satisfaction ● Multiple knapsack ● Generalized assignment	Time-Space Invariant
Particle Swarm Optimization 1995	Eberhart and Kennedy	Swarming behaviour of animals and human beings.	No Premature Convergence	● Clustering ● TSP ● Neural Network Applications ● Scheduling Problems ● Vehicle Routing ● Connection-Oriented Network Routing ● Sequential Ordering ● Graph Coloring ● Redundancy allocation ● Constraint satisfaction ● Multiple knapsack ● Damping the cavity-mode anti-resonances' peaks on a power plane	Space Variant
Honey Bees Optimization 2001	Abbass	Mating flights of honey bees (Swarm Behaviour)	No Premature Convergence	● 3-SAT problem optimization ● MAX-SAT problem optimisation ● Water resources management ● Stochastic dynamic programming ● stepped spillway optimum design ● Reconfiguration of multi-objective distribution feeder ● Optimization Problems	Space Variant

Algorithm	Proposed by	Based on	Conver-gence	Applied On	Time-Space Invariant/Variant
Artificial Fish School Algorithm 2001	Xiaolei Li	Swarming behaviour of fishes	No Premature Convergence	● Clustering ● Optimization Problems ● Knapsack Problem ● Color Quantization ● UCAV Path Planning ● Multi- Threshold Image Segmentation ● Probabilistic Causal-effect. Model ● Job Scheduling in Grid Computing ● Needle-in-haystack Problem ● Parameter Identification ● Electronic Tongue in Edible Oil Detection	Time –Space Invariant
Bat Algorithm 2005	Pham, Otri	Food foraging behaviour of Bat. Projects the quality of mimicking.	Premature Convergence	● Training of multi-layered perception neural networks ● Job shop scheduling optimisation ● Protein folding optimisation using the torsion angles model ● Optimisation of fuzzy logic controller parameters ● Peer-to-peer file sharing in mobile ad-hoc networks ● Interference suppression of linear antenna arrays ● Optimization Problems	Time-Space Variant
Cat Swarm Optimization 2006	SC Chu, PW Tsai and JS Pan	Common behaviour of cats.	Premature Convergence	● Optimal Contract Capacity ● Clustering ● Emotion Recognition ● Spectrum Sensing in Cognitive Radio ● Support Vector machine ● Enhancing the performance of Watermarking ● Reliability Task Orientation in Distributed Systems ● Optimization Problems	Space Variant
Firefly Algorithm 2008	Xin She Yang	Flashing behaviour of fireflies.		● Clustering ● Stochastic dynamic programming ● TSP ● QAP ● Vehicle Routing ● Graph Coloring ● Job Shop Scheduling ● Economic dispatch load problem ● Annual Crop Planning (ACP) ● Optimization Problems	Time-Space Invariant

Algorithm	Proposed by	Based on	Convergence	Applied On	Time-Space Invariant/Variant
Monkey Algorithm 2008	W. Yang	Mountain-climbing process of monkeys	Premature Convergence	● TSP ● Optimization Problems ● Multidimensional Assignment Problem ● Improving Lifetime of Heterogeneous Wireless Sensor Networks ● Minimizing Energy Consumption ● Transmission network expansion planning ● Optimal Sensor Placement ● Hybrid Power Systems Optimization	Space Variant
Cuckoo Search 2009	Yang and Suash Deb	Breeding behaviour of cuckoo. Solitary Intelligence	No Premature Convergence	● Clustering ● Feed forward neural network training ● Knapsack Problem ● Structural Optimization Problems ● Linear array synthesis ● Weighted sum optimization ● Annual Crop Planning (ACP) ● Damping the cavity-mode anti-resonances' peaks on a power plane ● Optimization Problems	Space Variant

collectively. These algorithms can be broadly classified into two categories, Swarm Behaviour which is the behaviour of collectiveness and Solitary Behaviour which is exhibited not under the influence of any other object/independent of environment.

Swarm Intelligence is a collective behaviour of decentralized, self-organized natural or artificial systems. The concept was introduced by Gerardo Beni and Jing Wang in 1989[1]. The collective behavior of the organisms like ants, bees, fish etc. inspires artificial intelligence to simulate and solve real world problems. In such systems, there are a number of simple agents which follow simple and fixed rules that determine their possible behavior during interaction among themselves and the surroundings. Although the individual particles are ignorant of it, their collective behavior leads to an intelligent global behavior. An optimization problem [2] is a real world problem where the objective function is not differentiable and its values can be acquired by simulation. These problems are either single-objective or multi-objective. In single objective optimization problems only one optimum solution with a single solution space exists [3]. Optimization problems based on swarm intelligence are known as meta-heuristic algorithms [4] which have features like self-organization, no central control, derivative free and easy to implement. These features lead to an emergent behaviour that overcome the main limitations of conventional methods and can be conveniently applied to various optimization problems. These met heuristic algorithms have two major components namely exploitation and exploration. The exploration ability ensures that the algorithm can search the whole space and escape from local optima while the Exploitation ability guarantees the algorithm can search carefully and converge to the optimal point.

III. Analysis

The applicability of an algorithm is divided based on the following characteristics [6].

a) Convergence

Supposing that we have (A1, A2,…) where A are real-valued random variables (Species) with the distribution functions (F1,F2,…) and F, respectively. We say that the distribution of An converges to the distribution of A as n → ∞ i
Fn(a) → F(a) as n → ∞
for all a at which F is continuous

b) Space Time Invariant

The behaviour of the species which have strict time and a space interval, such as Ant Colony. Their behaviour is similar to the problem constraints of Travelling Salesman Problem i.e. to tour on such a path which cover all the cities exactly at once in the least possible time interval. This is called space time Invariant.

c) Space Time variant

In this case the space and time can vary, this is the situation when direction matters such as Particle Swarm Optimization. In these situation application which are direction oriented not space and time restriction are known as Space Time Variant Algorithms.

IV. Conclusions

Hence this detailed survey can help us in using the specific algorithm in a given problem definition. This has been verified according to the successful results of each of the algorithm in that domain.

V. Future Scope

This comparative study can be made vast by studying the space time invariant, computational efforts vs. accuracy, linear vs. non linear problems, global vs. local maxima as these terms stands as the underlying challenges in an application domain.

References

1. R.S. Parpinelli, H.S. Lopes, "New inspirations in swarm intelligence: a survey", Int. J. Bio-Inspired Computation, Vol. 3, No. 1, 2011, pp. 1-16.
2. Ausiello, Giorgio; et al. (2003), Complexity and Approximation (Corrected ed.), Springer, ISBN 978-3-540-65431-5.
3. MA Sasa, Xue Jia, Fang Xingqiao, Liu Dongqing, "Research on Continuous Function Optimization Algorithm Based on Swarm Intelligence", 5th International Conference on Computation, 2009, pg no. 61-65.
4. Yang, X. S. (2008), Nature-Inspired Metaheuristic Algorithms,Frome: Luniver Press.
5. Xin-She Yang:Nature-Inspired Metaheuristic Algorithms, Second Edition,Luniver press,2011, p. 160
6. http://www.ieee.org/
7. X. S. Yang, "Firefly algorithm for multimodal optimization." In: Stochastic Algorithms: foundations and applications, archive/macros/latex/contrib/supported/IEEEtran/
8. Dan Simon, "Biogeography Based Optimization" IEEE Transaction on Evolution Computation, Vol 12., No. 6, Dec 2008.[9] Suruchi Sinha et. Al 'Classification of Mixed Pixel Using Hybridization of Ant Colony Optimization and Bio geography based Optimization' IEEE World Congress on Computational Intelligence. June 10-15, 2012.

Improved Chaos based Video Encryption using Harmony Search for Key Selection

Manas Gaur*

Abstract

To prevent video from unauthorized access and fabrication, the video should be encrypted before sending them on the unsecure network like internet. Traditional methods of encryption like AES, DES are not suitable for heavy multimedia objects like videos of conferences, due to their high computational complexity, restricted size of the key and time complexity of the algorithm. We propose a methodology of selective encryption of video using keys generated by harmony search, a meta-heuristic artificial intelligence technique. In our methodology, we perform selective encryption based on point of interest in the image, to reduce encryption time and also it results in encryption of potential confidential information. We compared the efficiency of the algorithm using entropy and found that proposed technique has achieved state of the art.

Key Words: MATLAB, Harmony Search, Selective encryption, Entropy

I. Introduction

In today's world there is a huge growth seen in multimedia highway. Large size images are sent over the network. Some of the images contain potential confidential information. Since an image has a fixed resolution of 1024X680 and small in size so it is very easy to decrypt an image. If a sender tries to send multiple images, he has to encrypt them separately which might generate similarity in pattern for eavesdropper. If a sequence of images differs in small position then it is better to convert the sequence of images into videos and then encrypt [2]. The images sent over network are not only grayscale, binary but coloured which change their dimension from 2D to 3D. The video encryption finds its importance in military communication, confidential video conferences, remote sensing videos which contain potential information about the landscape of the country. Traditional methods exist for video encryption which can be categorised as Block Cipher cryptographic algorithm like AES, DES and artificial intelligence cryptographic algorithm which are not limited by the length of the key like Ant Colony cryptosystem [9], Genetic Algorithm cryptosystem which are heuristics based. Falling into the category of artificial intelligence is Harmony search algorithm and more similar to genetic algorithm but much efficient than other.

Artificial Intelligence (AI) techniques. Harmony search is a relatively new meta-heuristic algorithm for continuous optimization, in which its concept initiates the process of music improvisation. We proposed the process of selective encryption because of the explosion of networks and the huge amount of content transmitted along therefore it increase the size of the data sent, our main goal is the object of consideration in the image or video, hence encryption the object is important irrespective of the environment [7]. The technique of selective encryption not only reduces the size of the encrypted data to be sent but also reduces the time to perform encryption of the video irrespective of its size [4]. The remainder of the article is organized as follows. In Section 2 a brief introduction of all the research done in field of encryption of images, videos. In Section 3 a brief introduction to harmony search with its algorithm modified for video encryption and a brief introduction to selective encryption. In Section 4, we discuss about

Manas Gaur*
University of New York, Albany USA

the simulation environment and results. In section 5, we provide the conclusion of our research.

II. Literature Survey

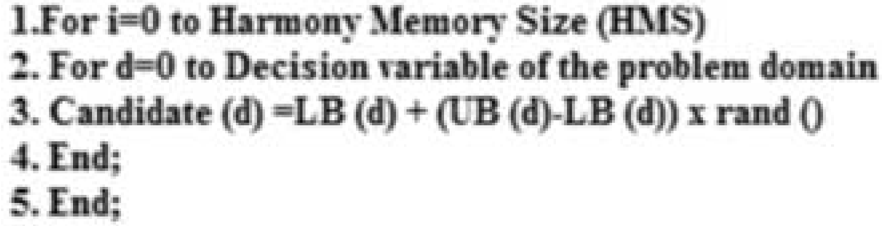
Adnan M. Alattar [7], proposed selective encryption methods for MPEG-II, MPEG-IV, H.261 and H.263+. He stated that encrypted I-macro block of video requires half the processing time. He stated that the method showed 60-82% reduction in processing time.Wenjun Zeng [9] proposed the method of joint encryption and compression of the video in the frequency domain by selecting bits in the frequency domain. This method is suitable for relative small scale video, but not suitable for remote sensed images where the images details are very minute. Yong Wang[10] proposed a new chaos-based fast image encryption where the image is partitioned into blocks, shuffled and encrypted using pseudorandom numbers from spatial-temporal chaos. A. Nag [3] divided the image into 2X2pixels blocks and each block is encrypted using XOR operation by 64 bit key which might be strong enough for small scale images and low processing system but when it comes to AVI files and high detailed images we can't create a clean demarcation in the image. Gaurav Bhatnagar [2] proposed a scheme to scramble pixels positions using Saw-tooth space filling curve followed by selection of pixels of interest and the diffusion is created using chaotic map.

III. Background

This section provides a brief introduction of every methodology used to carry out the research work. A brief overview of the technique is essential for analysing rest of the research.

A. Harmony Search

Geem [1] proposed and implemented Harmony Search, a meta-heuristic algorithm that utilized musical process concept for searching a perfect state of harmony. Harmony search keeps the possible candidate solution, which are initialized randomly within the search space as follows:



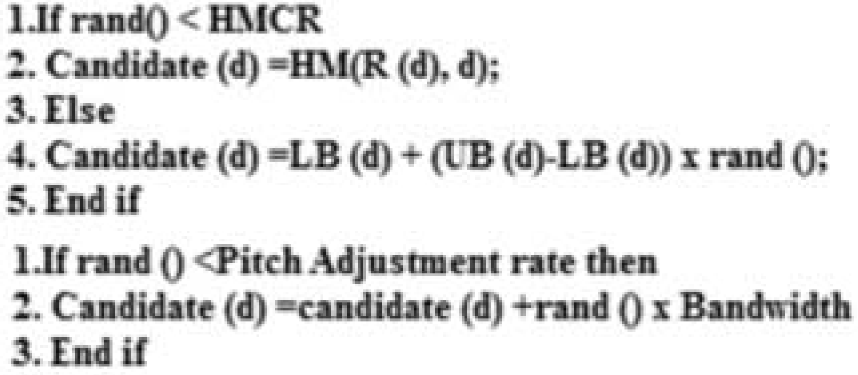
```
For i=0 to Harmony Memory Size (HMS)
    For d=0 to Decision variable of the problem domain
        Candidate (d) =LB (d) + (UB (d)-LB (d)) x rand ()
    End;
End;
```

Figure 1: Harmony search improvisation based on heuristic

Where LB (d) and UB (d) are the lower an upper bounds in the search space and rand () function delivers a random value between 0 and 1. This algorithm is an improvisation process of pitch value to bring search to optimal solutions. There are three main parameters controlling the improvisation process, that are harmony consideration rate (HMCR), pitch adjustment rate (PAR) and the bandwidth (bw). In each iteration rand () value is tested against HMCR, if less than candidate solution is generated with memory consideration otherwise by random selection. The candidate generated by memory consideration is further adjusted by pitch adjustment rate. The final candidate solution generated after d iterations is tested for fitness. If the fitness value of the candidate is higher than the previous candidate then new candidate is taken as a solution for further candidate improvement in HMS iterations. The process of harmony search can be related to genetic algorithm process of finding the best candidate solution.

1. Harmony Search Crossover

Genetic algorithm employs crossover based on crossover probability, with similar analogy harmony search auger the next element of the candidate vector based on HMCR. The code line for harmony search crossover is shown in figure 2(a).



```
If rand() < HMCR
Candidate (d) =HM(R (d), d);
Else
Candidate (d) =LB (d) + (UB (d)-LB (d)) x rand ();
End if
```

```
If rand () <Pitch Adjustment rate then
Candidate (d) =candidate (d) +rand () x Bandwidth
End if
```

Figure 2. a. Harmony search crossover. b. Harmony search mutation

The process compares the HMCR value with random value generator before proceeding with crossover in the algorithm. HM(R (d), d) selects the next candidate vector from the array of 240 x 320 pixel values generated from the image frame of the video. The value R (d) is selection of a vector of size 320 pixels using poisson probability. The essence of poisson probability if the consideration of non-negative integer values and use of mean and variance in the distribution. Due to large size of array the variability of numbers in the array is also large. Hence poisson probability distribution is in utility for our harmony search algorithm.

$$P(X|\lambda) = \frac{\lambda^x}{x!} e^{\lambda} \qquad (1)$$

The poisson distribution take ë for mean parameters can be vector and x is the number of different distribution id est. vectors.

2. Harmony Search Mutation

In genetic algorithm mutation occurs based on some mutation probability threshold. In the process of mutation changes are made in the candidate solution to make it more fit, analogous to it the pitch adjustment rate behave as mutation probability and the candidate solution is mutated using bandwidth variable. The steps in mutation is explained in the above figure 1(b).

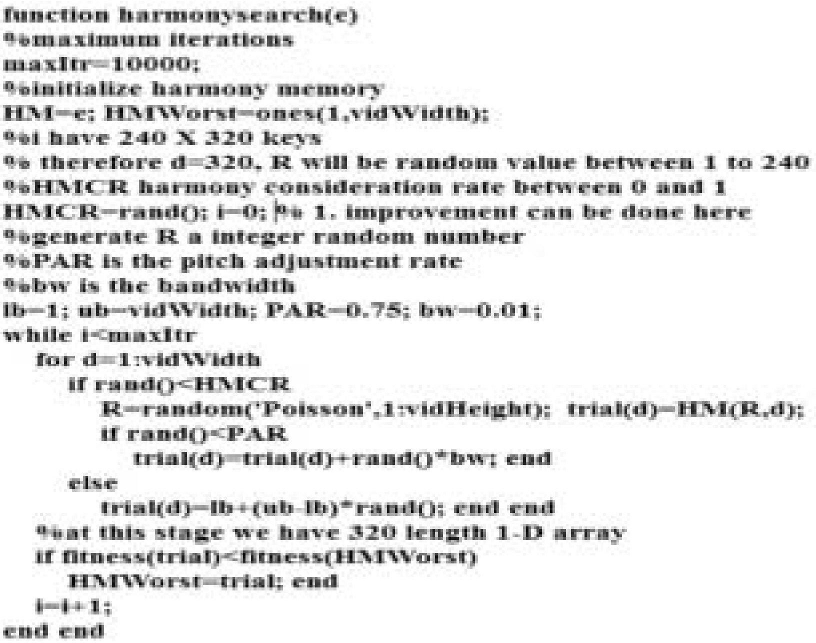
B. Selective Encryption

With large amount of data on the network highway requires greater amount of security and bandwidth. Same is true when we send our encrypted file on the insecure network. Encrypting the complete file generates patterns in the encrypted file and also encrypting entire video takes time. Here we are using a technique that encrypted the confidential information in the video and the time taken for encrypting is reduced [5]. The aim of selective encryption is to reduce time while at the same time preserving a sufficient level of security [6]. Selective encryption finds its application in real time networking, high definition delivery, mobile communication etc.

IV. Simulation and Results

A. Simulation

Selective encryption of video using harmony search was carried out using MATLAB R2011a with video in audio visual format (.avi). The methodology followed during the course of research was generation of pseudorandom random vectors of the length equivalent to Y dimension of the 2D image. The pseudorandom vectors are passed to harmony search routine to select one vector which we call a best candidate solution. The iterations for Harmony search are kept to 10000. The key generated by the harmony search routine is passed as input to the encryption routine of the video. During the process of encryption of video each pixel of the video in encrypted by the key that is a pixel at position (x, y, z) is encrypted by the key for 240X320 iteration (image resolution of our example video) followed by encryption of pixel at position (x,z,y), at position (y,x,z), at position (y,z,x),



```
function harmonysearch(e)
%maximum iterations
maxItr=10000;
%initialize harmony memory
HM=e; HMWorst=ones(1,vidWidth);
%i have 240 X 320 keys
% therefore d=320, R will be random value between 1 to 240
%HMCR harmony consideration rate between 0 and 1
HMCR=rand(); i=0; % 1. improvement can be done here
%generate R a integer random number
%PAR is the pitch adjustment rate
%bw is the bandwidth
lb=1; ub=vidWidth; PAR=0.75; bw=0.01;
while i<maxItr
   for d=1:vidWidth
      if rand()<HMCR
         R=random('Poisson',1:vidHeight);  trial(d)=HM(R,d);
         if rand()<PAR
            trial(d)=trial(d)+rand()*bw; end
      else
         trial(d)=lb+(ub-lb)*rand(); end end
   %at this stage we have 320 length 1-D array
   if fitness(trial)<fitness(HMWorst)
      HMWorst=trial; end
   i=i+1;
end end
```

Figure 3: Harmony Search Algorithm

at position (z,y,x) and at position (z,x,y). The resulting parts of the video are combined to demonstrate selective encryption where the environment show little change whereas the object of consideration blanks out. The activity diagram shown in the figure 2 shows the chronology of states arrived during the working of the algorithm. The elliptical boxes shows the transit state whereas the boxes with round corners shows stable (visible results) state achieved. The creation of activity diagram in rational rose was carried out abut with execution of events on MATLAB.

B. Result and Analysis

The main aim of encryption algorithm used for multimedia object is the decrease the correlation between the pixels in the image and increase the entropy of the pixel value so that decryption of the multimedia object is difficult. Entropy of a multimedia object is calculated as

$$H = -\sum_{k=0}^{G-1} P(k)\, log_2(P(k)) \qquad (2)$$

Where:H: entropy, G: grey value of input image (0…255). P (k): is the probability of the occurrence of the symbol k. The comparison between entropy values of different image frames of original video and encrypted vide is shown in table 2.

Table 1: Simulation environment

S.No.	Property	Value
1.	Video file	.AVI
2.	Image resolution	240X320
3.	Maximum Iteration for harmony search	10000
4.	HMWorst	{1……..320 times}
5.	Fitness Function	Permutations and Combination

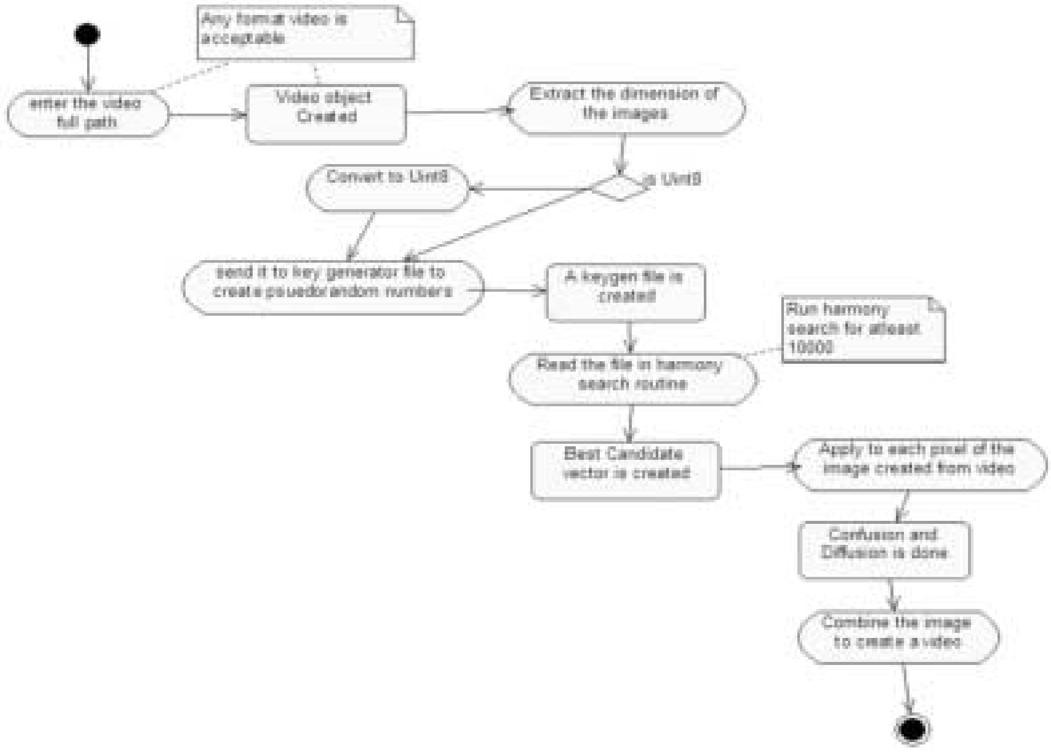

Any format video is acceptable
enter the video full path
Video object Created
Extract the dimension of the images
is Uint8
Convert to Uint8
send it to key generator file to create psuedorandom numbers
A keygen file is created
Run harmony search for atleast 10000
Read the file in harmony search routine
Best Candidate vector is created
Apply to each pixel of the image created from video
Confusion and Diffusion is done
Combine the image to create a video

Figure 4: Activity Diagram of Harmony search application on Selective encryption

Table 2: Entropy of original video and encrypted video

S.No.	Image Frame in Original Video	Image Frame in Encrypted video
1.	0.1663	0.8361
2.	0.1875	0.8475
3.	0.1712	0.8420
4.	0.1512	0.7802
5	0.1339	0.7732
6.	0.1176	0.7004
7.	0.0944	0.6922
8.	0.0691	0.6701
9.	0.0462	0.6254
10.	0.0392	0.5920
11.	0.0298	0.5213
12.	0.0261	0.4840
Avg.	0.1027	0.7270

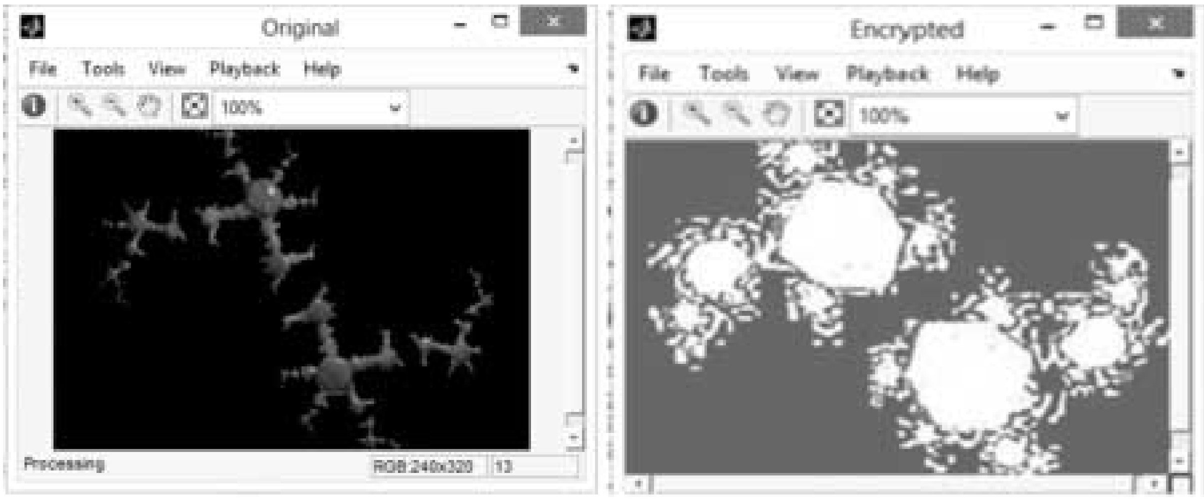


Figure 5: Showing the original video and selectively encrypted video

The average entropy of the encrypted video clearly shows strong encryption of the video using harmony search generated key. The original video has 240 X 320 resolution 13 images which were selectively encrypted with no change in resolution and combined to produce the encrypted video. On viewing the image in the figure 5 we observe that the object was selectively encrypted with just high contrast in the background due to confusion and diffusion. The process of encryption was tested against time, that is execution of the technique for 1st time took 2.35 minutes and subsequently, technique took 1.85, 1.82, 1.78…….0.32 minutes at 50th iteration of the program. At 51st iteration the program produced the encrypted video in 0.3125 minutes. This shows much improvement over traditional methods of encryption and genetic algorithm used in past for image encryption.

V. Conclusion

This paper presents a novel methodology of selective encryption of video using harmony search. This methodology provide a valuable tool for secure video transfer over the network. As shown in the result, this technique tremendously increase the entropy of the image frames of the video which makes it difficult decrypt and more secure. This technique has been testedusing MPEG format video, AVI format video and MPEG II format video and result was fascinating.

Harmony search is a versatile meta-heuristic algorithm that can be applied in various field. Changing the HMCR, PAR can further improve the performance of the algorithm. Videos of higher pixels will develop stronger key as the length will be high, large number of permutation and hence more strong encryption. In the process of generation of pseudorandom keys if initial key to start is high more robust encryption can be performed as strongest key will be selected by the harmony search. The field of encryption has wider prospects and involvement of artificial intelligence has led to generation of enormously long keys.

References

ChukiatWorasucheep, “A Harmony Search with Adaptive Pitch Adjustment for Continuous Optimization”, International Journal of Hybrid Information Technology, 2011.

Gaurav Bhatnagar, Q.M Jonathan Wu, “Selective Image Encryption based on pixels ofInterest and Singular value decomposition”, Digital Signal Processing, 2012.

A. Nag, J P Singh, “Image Encryption using Affine Transform and XOR Operation”, ICSCCN, 2011.

GuanrongChenb*et al*, "A New Chaos Based Fast Image Encryption Algorithm", Applied Soft Computing, 2011.

BholaNath Roy, JayantKushwaha, "Secure Image Data by Double Encryption", International Journal of Computer Application, 2010.

Zhen Chen, "A Lightweight Encryption Algorithm for Images", Advances in Intelligentand Soft Computing, 2012.

Adnan M. Alattar, "Improved Selective Encryption Techniques for Secure Transmission ofMpeg Video Bit-Streams", ICIP, 1999.

Mohammed AF. Al-Husainy, "Image Encryption Using Genetic Algorithm", InformationTechnology Journal Asian Network for Scientific Information, 2006.

Wenjun Zeng et al, "A multi-layer key stream based approach for joint encryption and Compression of H.264 video", IEEE International Conference on Multimedia and Expo, 2011.

Yong Wang et.al, "A new chaos based fast image encryption algorithm", Applied SoftComputing, 2011.