International Journal of Computer Applications Technology and Research
Volume 3– Issue 7, 488 - 493, 2014

Security in Cloud Computing

Shweta Singh
DIT University ,Dehradun

Abstract: Cloud computing is a completely internet dependent technology where client data is stored and maintain in the data center of a cloud provider. Cloud computing is an architecture for providing computing service via the internet on demand and pay per use access to a pool of shared resources namely networks, storage, servers, services and applications, without physically acquiring them. So it saves managing cost and time for organizations. The security for Cloud Computing is emerging area for study and this paper provide security topic in terms of cloud computing based on analysis of Cloud Security treats and Technical Components of Cloud Computing.

Keywords: Cloud, Services, Cloud service user, Cloud service provider, Security Issues , License Risk, Data Availability

1. INTRODUCTION

The cloud computing is a new computing model which comes from grid computing, distributed computing, parallel computing, virtualization technology, utility computing and other computer technologies and it has more advantage characters such as large scale computation and data storage, virtualization, high expansibility, high reliability and low price service. The security problem of cloud computing is very important and it can prevent the rapid development of cloud computing. This paper introduces some cloud computing systems and analyzes cloud computing security problem and its strategy according to the cloud computing concepts and characters. The data privacy and service availability in cloud computing are the key security problem. Single security method cannot solve the cloud computing security problem and many traditional and new technologies and strategies must be used together for protecting the total cloud computing system.

We are conducting research on secure cloud computing .Due to the extensive complexity of the cloud , we contend that it will be difficult to provide a holistic solution to secure the cloud at present . Therefore our goal is to make increment enhancements to securing the cloud that will ultimately result in a secure cloud . I n particular , we are developing a secure cloud consisting of hardware ,software and data . Our cloud system will

(a) support efficient storage of encrypted sensitive data
(b) store, manage and query massive amounts of data
(c) support fine grained access control and
(d) support strong authentication.

2. ARCHITECTURE OF CLOUD COMPUTING

Cloud management system is divided into four layers, respectively the Resources & Network Layer, Services Layer, Access Layer, and User Layer. Each layer includes a set of functions[1]:

The Resources & Network Layer manages the physical and virtual resources.

- The Services Layer includes the main categories of cloud services, namely, NaaS, IaaS, PaaS, SaaS/CaaS, the service orchestration function and the cloud operational function.
- The Access Layer includes API termination function, and Inter-Cloud peering and federation function.
- The User Layer includes End-user function, Partner function and Administration function.
- The Cross layer includes Management, Security & Privacy, etc. are considered as that covers all the layers.

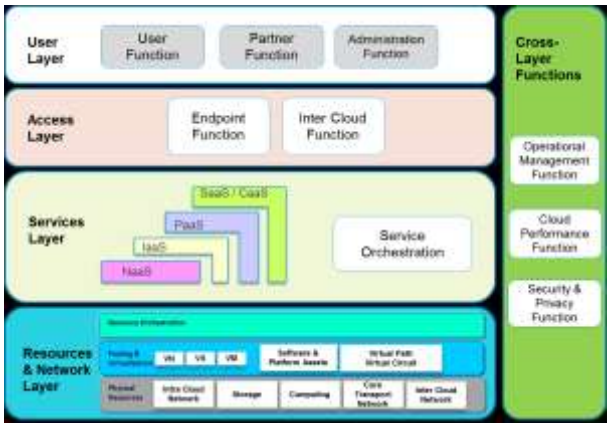


Figure. 1 The Cloud Computing Components

3. SECURITY SUBSYSTEM

The five functional security subsystems defined by IBM are as follows:

3.1 Audit and Compliance:

This subsystem addresses the data collection, analysis, and archival requirements in meeting standards of proof for an IT environment. It captures, analyzes, reports, archives, and retrieves records of events and conditions during the operation of the system .

3.2 Access Control:

This subsystem enforces security policies by gating access to processes and services within a computing solution via identification, authentication, and authorization[5]. In the context of cloud computing, all of these mechanisms must also be considered from the view of a federated access control system.

3.3 Flow Control:

This subsystem enforces security policies by gating information flow and visibility and ensuring information integrity within a computing solution .

3.4 Identity and Credential Management:

This subsystem creates and manages identity and permission objects that describe access rights information across networks and among the subsystems, platforms, and processes, in a computing solution [4]. It may be required to adhere to legal criteria for creation and maintenance of credential objects.

3.5 Solution Integrity:

This subsystem addresses the requirement for reliable and proper operation of a computing solution

4. SERVICES PROVIDED

Generally cloud services can be divided into three categories: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

4.1 Software-as-a-Service (SaaS):

SaaS can be described as a process by which Application Service Provider (ASP) provide different software applications over the Internet. This makes the customer to get rid of installing and operating the application on own computer and also eliminates the tremendous load of software maintenance; continuing operation, safeguarding and support . SaaS vendor advertently takes responsibility for deploying and managing the IT infrastructure (servers, operating system software, databases, data center space, network access, power and cooling, etc) and processes (infrastructure patches/upgrades, application patches/upgrades, backups, etc.) required to run and manage the full solution. SaaS features a complete application offered as a service on demand. Examples of SaaS includes: Salesforce.com, Google Apps.

4.2 Platform as a Service (PaaS):

PaaS is the delivery of a computing platform and solution stack as a service without software downloads or installation for developers, IT managers or end-users. It provides an infrastructure with a high level of integration in order to implement and test cloud applications.

The user does not manage the infrastructure (including network, servers, operating systems and storage), but he controls deployed applications and, possibly, their configurations. Examples of PaaS includes: Force.com, Google App Engine and Microsoft Azure.

4.3 Infrastructure as a Service (IaaS):

Infrastructure as a service (IaaS) refers to the sharing of hardware resources for executing services using Virtualization technology. Its main objective is to make resources such as servers, network and storage more readily accessible by applications and operating systems. Thus, it offers basic infrastructure on-demand services and using Application Programming Interface (API) for interactions with hosts, switches, and routers, and the capability of adding new equipment in a simple and transparent manner. In general, the user does not manage the underlying hardware in the cloud infrastructure, but he controls the operating systems, storage and deployed applications. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis. Examples of IaaS includes Amazon Elastic Cloud Computing (EC2), Amazon S3, GoGrid

5. CLOUD MODELS

Four different cloud deployment models namely Private cloud, Public cloud, Hybrid cloud and Community cloud.

5.1 Private cloud:

Private cloud can be owned or leased and managed by the organization or a third party and exist at on-premises or off-premises. It is more expensive and secure when compared to public cloud. In private cloud there are no additional security regulations, legal requirements or bandwidth limitations that can be present in a public cloud environment, by using a private cloud, the cloud service providers and the clients have optimized control of the infrastructure and improved security, since user's access and the networks used are restricted. One of the best examples of a private cloud is Eucalyptus Systems .

5.2 Public Cloud:

A cloud infrastructure is provided to many customers and is managed by a third party and exist beyond the company firewall. Multiple enterprises can work on the infrastructure provided, at the same time and users can dynamically provision resources. These clouds are fully hosted and managed by the cloud provider and fully responsibilities of installation, management, provisioning, and maintenance. Customers are only charged for the resources they use, so under-utilization is eliminated. Since consumers have little control over the infrastructure, processes requiring powerful security and regulatory compliance are not always

a good fit for public clouds. In this model, no access restrictions can be applied and no authorization and authentication techniques can be used. Public cloud providers such as Google or Amazon offer an access control to their clients. Examples of a public cloud includes Microsoft Azure, Google App Engine.

5.3 Hybrid Cloud:

A composition of two or more cloud deployment models, linked in a way that data transfer takes place between them without affecting each other. These clouds would typically be created by the enterprise and management responsibilities would be split between the enterprise and the cloud provider.

In this model, a company can outline the goals and needs of services . A well-constructed hybrid cloud can be useful for providing secure services such as receiving customer payments, as well as those that are secondary to the business, such as employee payroll processing. The major drawback to the hybrid cloud is the difficulty in effectively creating and governing such a solution. Services from different sources must be obtained and provisioned as if they originated from a single location, and interactions between private and public components can make the implementation even more complicated. These can be private, community or public clouds which are linked by a proprietary or standard technology that provides portability of data and applications among the composing clouds. An example of a Hybrid Cloud includes Amazon Web Services (AWS)[4].

5.4 Community Cloud:

Infrastructure shared by several organizations for a shared cause and may be managed by them or a third party service provider and rarely offered cloud model.

These clouds are normally based on an agreement between related business organizations such as banking or educational organizations. A cloud environment operating according to this model may exist locally or remotely

6.SECURITY GUIDANCE

General security guidance to deal with the above threats can be found in :

- Encryption and Key Management: Encryption provides data protection while key management enables access to protected data. It is strongly recommended to encrypt data in transit over networks, at rest, and on backup media. In particular, data encryption at rest (e.g., for long-term archival storage) can avoid the risk of malicious cloud service providers or malicious multi-tenants abuse. At the same time, secure key stores (including key backup and recoverability) and access to key stores must be securely implemented since improper (or access to) key storage could lead to the compromise of all encrypted data.
- Identity and Access Management: Secure management of identity and access control is a critical factor to prevent account and service hijacking. It is strongly recommended to prohibit sharing of account credentials, to leverage strong (multi-factor) authentication if possible, and to consider delegated authentication and managing trust across all types of cloud services.

7. THREATS FOR CLOUD SERVICE USER

7.1 Responsibility Ambiguity

Cloud service users consume delivered resources through service models. The customer-built IT system thus relies on the services. The lack of a clear definition of responsibility among cloud service users and Providers may evoke conceptual conflicts. Moreover, any contractual inconsistency of provided services could induce anomaly, or
incidents. However the problem of which entity is the data controller which on is the data processor stays open at an international scale (even if the international aspect is reduced to a minimal third party outside of the specific region like EU).

7.2 Loss of Governance

For an enterprise, migrating a part of its own IT system to a cloud infrastructure implies to partially give control to the cloud service providers. This loss of governance depends on the cloud service models. For instance, IaaS only delegates hardware and network management to the provider, while SaaS also delegates OS, application, and service integration in order to provide a turnkey service to the cloud service user.

7.3 Loss of Trust

It is sometime difficult for a cloud service user to recognize his provider's trust level due to the black-box feature of the cloud service. There is no measure how to get and share the provider's security level in formalized manner. Furthermore, the cloud service users have no abilities to evaluate security implementation level achieved by the provider. Such a lack of sharing security level in view of cloud service provider will become a serious security threat in use of cloud services for cloud service users.

7.4 Service Provider Lock-in

A consequence of the loss of governance could be a lack of freedom regarding how to replace a cloud provider by another. This could be the case if a cloud provider relies on non-standard hypervisors or virtual machine image format and does not provide tools to convert virtual machines to a standardized format.

7.5 Unsecure Cloud Service User Access

As most of the resource deliveries are through remote connection, non-protected APIs, (mostly management APIs and PaaS services is one of the easiest attack vector). Attack methods such as phishing, fraud, and exploitation of software vulnerabilities still achieve results. Credentials and passwords are often reused, which amplifies the impact of such attacks. Cloud solutions add a new threat to the landscape. If an attacker gains access to your credentials, they can eavesdrop on your activities and transactions, manipulate data, return falsified information, and redirect your clients to illegitimate sites. Your account or service instances may become a new base for the attacker. From here, they may leverage the power of your reputation to launch subsequent attacks.

7.6 Lack of Information/Asset Management

When applying to use Cloud Computing Services, the cloud service user will have serious concerns on lack of information/asset management by cloud service providers such as location of sensitive asset/information, lack of physical control for data storage, reliability of data backup (data retention issues), countermeasures for BCP and Disaster
Recovery and so on. Furthermore, the cloud service users also have important concerns on exposure of data to foreign government and on compliance with privacy law such as EU data protection directive.

7.7 Data loss and leakage

The loss of encryption key or privileged access code will bring serious problems to the cloud service users. Accordingly, lack of cryptographic management information
such as encryption keys, authentication codes and access privilege will heavily lead sensitive damages on data loss and unexpected leakage to outside. For example, insufficient authentication, authorization, and audit (AAA) controls; inconsistent use of encryption and/or authentication keys; operational failures; disposal problems; jurisdiction and political issues; data center reliability; and disaster recovery can be recognized as major behaviors in this threat category.

8. THREATS FOR CLOUD SERVICE PROVIDER

8.1 Responsibility Ambiguity

Different user roles, such as cloud service provider, cloud service user, client IT admin, data owner, may be defined and used in a cloud system. Ambiguity of such user roles and responsibilities definition related to data ownership, access control, infrastructure maintenance, etc, may induce business or legal dissention (Especially when dealing with third parties. The cloud service provider is somehow a cloud service user)[6].

8.2 Protection Inconsistency

Due to the decentralized architecture of a cloud infrastructure, its protection mechanisms are likely to be inconsistency among distributed security modules. For example, an access denied by one IAM module may be granted by another. This threat may be profited by a potential attacker which compromises both the confidentiality and integrity.

8.3 Evolutional Risks

One conceptual improvement of cloud computing is to postpone some choices from the design phase to the execution phase. This means, some dependent software components of a system may be selected and implemented when the system executes. However, conventional risk assessment methodology can no longer match such an evolution. A system which is assessed as secure during the design phase may exploit vulnerabilities during its execution due to the newly implemented software components.

8.4 Business Discontinuity

The "as a service" feature of cloud computing allocates resources and delivers them as a service. The whole cloud infrastructure together with its business workflows thus
relies on a large set of services, ranging from hardware to application. However, the discontinuity of service delivery, such as black out or delay, may bring out a severe impact related to the availability.

8.5 Supplier Lock-in

The platform of a service provider is built by some software and hardware components by suppliers. Some supplier-dependent modules or workflows are implemented for integration or functionality extension. However, due to the lack of standard APIs, the portability to migrate to another supplier is not obvious. The consequence of provider locked-in could be a lack of freedom regarding how to replace a supplier.

8.6 License Risks

Software licenses are usually based on the number of installations, or the numbers of users. Since created virtual machines will be used only a few times, the provider may have to acquire from more licenses than really needed at given time. The lack of a "clouded" license management scheme which allows to pay only for used licenses may cause software use conflicts.

9. SECURITY ISSUES...

- Virtual Machine Security
- Network Security
- Data Security
- Data Privacy
- Data Integrity
- Data Location
- Data Availability

9.1 Virtual Machine Security:

Virtualization is one of the main components of a cloud. Virtual machines are dynamic i.e it can quickly be reverted to previous instances, paused and restarted, relatively easily. Ensuring that different instances running on the same physical machine are isolated from each other is a major task of virtualization. They can also be readily cloned and seamlessly moved between physical servers. This dynamic nature and potential for VM sprawl makes it difficult to achieve and maintain consistent security. Vulnerabilities or configuration errors may be unknowingly propagated. Also, it is difficult to maintain an auditable record of the security state of a virtual machine at any given point in time. Full Virtualization and Para Virtualization are two kinds of virtualization in a cloud computing paradigm. In full virtualization, entire hardware architecture is replicated virtually. However, in para-virtualization, an operating system is modified so that it can be run concurrently with other operating systems. VMM (Virtual Machine Monitor), is a software layer that abstracts the physical resources used by the multiple virtual machines. The VMM provides a virtual processor and other virtualized versions of system devices such as I/O devices, storage, memory, etc. Many bugs have been found in all popular VMMs that allow escaping from Virtual machine. Vulnerability in Microsoft Virtual PC and Microsoft Virtual Server could allow a guest operating system user to run code on the host or another guest operating system. Vulnerability was found in VMware's shared folders mechanism that grants users of a guest system read and write access to any portion of the host's file system including the system folder and other security-sensitive files. Vulnerability in Xen can be exploited by "root" users of a guest domain to execute arbitrary commands. The other issue is the control of administrator on host and guest operating systems. Current VMMs (Virtual Machine Monitor) do not offer perfect isolation. Virtual machine monitor should be 'root secure', meaning that no privilege within the virtualized guest environment permits interference with the host system.

9.2 Network Security:

Networks are classified into many types like shared and non-shared, public or private, small area or large area networks and each of them have a number of security threats to deal with. Problems associated with the network level security comprise of DNS attacks, Sniffer attacks, issue of reused IP address, etc which are explained in details as follows.

A Domain Name Server (DNS) server performs the translation of a domain name to an IP address. Since the domain names are much easier to remember. Hence, the DNS servers are needed. But there are cases when having called the server by name, the user has been routed to some other evil cloud instead of the one he asked for and hence using IP address is not always feasible. Although using DNS security measures like: Domain Name System Security Extensions (DNSSEC) reduces the effects of DNS threats but still there are cases when these security measures prove to be inadequate when the path between a sender and a receiver gets rerouted through some evil connection. It may happen that even after all the DNS security measures are taken, still the route selected between the sender and receiver cause security problems[7].

Sniffer attacks are launched by applications that can capture packets flowing in a network and if the data that is being transferred through these packets is not encrypted, it can be read and there are chances that vital information flowing across the network can be traced or captured. A sniffer program, through the NIC (Network Interface Card) ensures that the data/traffic linked to other systems on the network also gets recorded. It can

be achieved by placing the NIC in promiscuous mode and in promiscuous mode it can track all data, flowing on the same network. A malicious sniffing detection platform based on ARP (address resolution protocol) and RTT (round trip time) can be used to detect a sniffing system running on a network .

Reused IP address issue have been a big network security concern. When a particular user moves out of a network then the IP-address associated with him (earlier) is assigned to a new user. This sometimes risks the security of the new user as there is a certain time lag between the change of an IP address in DNS and the clearing of that address in DNS caches. And hence, we can say that sometimes though the old IP address is being assigned to a new user still the chances of accessing the data by some other user is not negligible as the address still exists in the DNS cache and the data belonging to a particular user may become accessible to some other user violating the privacy of the original user .

9.3 Data security:

For general user, it is quite easy to find the possible storage on the side that offers the service of cloud computing. To achieve the service of cloud computing, the most common utilized communication protocol is Hypertext Transfer Protocol (HTTP). In order to assure the information security and data integrity, Hypertext Transfer Protocol Secure (HTTPS) and Secure Shell (SSH) are the most common adoption. In a traditional on-premise application deployment model, the sensitive data of each enterprise continues to reside within the enterprise boundary and is subject to its physical, logical and personnel security and access control policies. However, in cloud computing, the enterprise data is stored outside the enterprise boundary, at the Service provider end. Consequently, the service provider must adopt additional security checks to ensure data security and prevent breaches due to security vulnerabilities in the application or through malicious employees. This involves the use of strong encryption techniques for data security and fine-grained authorization to control access to data. Cloud service providers such as Amazon, the Elastic Compute Cloud (EC2) administrators do not have access to customer instances and cannot log into the Guest OS. EC2 Administrators with a business need are required to use their individual cryptographically strong Secure Shell (SSH) keys to gain access to a host. All such accesses are logged and routinely audited. While the data at rest in Simple Storage Service (S3) is not encrypted by default, users can encrypt their data before it is uploaded to Amazon S3, so that it is not accessed or tampered with by any unauthorized party[3] .

9.4 Data Privacy:

The data privacy is also one of the key concerns for Cloud computing. A privacy steering committee should also be created to help make decisions related to data privacy. Requirement: This will ensure that your organization is prepared to meet the data privacy demands of its customers and regulators. Data in the cloud is usually globally distributed which raises concerns about jurisdiction, data exposure and privacy. Organizations stand a risk of not complying with government policies as would be explained further while the cloud vendors who expose sensitive information risk legal liability. Virtual co-tenancy of sensitive and non-sensitive data on the same host also carries its own potential risks[2].

9.5 Data Integrity:

Data corruption can happen at any level of storage and with any type of media, So Integrity monitoring is essential in cloud storage which is critical for any data center. Data integrity is easily achieved in a standalone system with a single database. Data integrity in such a system is maintained via database constraints and transactions. Transactions should follow ACID (atomicity, consistency, isolation and durability) properties to ensure data integrity. Most databases support ACID transactions and can preserve data integrity. Data generated by cloud computing services are kept in the clouds. Keeping data in the clouds means users may lose control of their data and rely on cloud operators to enforce access control.

9.6 Data Location:

In general, cloud users are not aware of the exact location of the datacenter and also they do not have any control over the physical access mechanisms to that data. Most well-known cloud service providers have datacenters around the globe. In many a cases, this can be an issue. Due to compliance and data privacy laws in various countries, locality of data is of utmost importance in many enterprise architecture. For example, in many EU and South America countries, certain types of data cannot leave the country because of potentially sensitive information. In addition to the issue of local laws, there's also the question of whose jurisdiction the data falls under, when an investigation occurs. Next in the complexity chain are distributed systems. In a distributed system, there are multiple databases and multiple applications .

In order to maintain data integrity in a distributed system, transactions across multiple data sources need to be handled correctly in a fail safe manner. This can be done using a central global transaction manger. Each application in the distributed system should be able to participate in the global transaction via a resource manager.

9.7 Data Availability:

Data Availability is one of the prime concerns of mission and safety critical organizations. When keeping data at remote systems owned by others, data owners may suffer from system failures of the service provider. If the Cloud goes out of operation, data will become unavailable as the data depends on a single service provider. The Cloud application needs to ensure that enterprises are provided with service around the clock. This involves making architectural changes at the application and infrastructural levels to add scalability and high availability. A multi-tier architecture needs to be adopted, supported by a load-balanced farm of application instances, running on a variable number of servers. Resiliency to hardware/software failures, as well as to denial of service attacks, needs to be built from the ground up within the application. At the same time, an appropriate action plan for business continuity (BC) and disaster recovery (DR) needs to be considered for any unplanned emergencies.

10. CONCLUSION

Cloud service providers need to inform their customers on the level of security that they provide on their cloud. In this paper, we first discussed various models of cloud computing, security issues Data security is major issue for Cloud Computing. There are several other security challenges including security aspects of network and virtualization. New security techniques need to be developed and older security techniques needed to be radically tweaked to be able to work with the clouds architecture.

11. REFERNCES

[1] A. Kundu, C. D. Banerjee, P. Saha, "Introducing New Services in Cloud Computing Environment", International Journal of Digital Content Technology and its Applications, AICIT, Vol. 4, No. 5, pp. 143-152, 2010.

[2] Tim Mather, Subra Kumaraswamy, Shahed Latif, Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, O' Reilly Media, USA, 2009.

[3] Ronald L. Krutz, Russell Dean Vines "Cloud SecurityA Comprehensive Guide to Secure Cloud Computing", Wiley Publishing, Inc.,2010

[4] K. Vieira, A. Schulter, C. B. Westphall, and C. M. Westphall, "Intrusion detection techniques for Grid and Cloud Computing Environment," IT Professional, IEEE Computer Society, vol. 12, issue 4, pp. 38-43, 2010.

[5] Marios D. Dikaiakos, Dimitrios Katsaros, Pankaj Mehra, George Pallis, Athena Vakali, "Cloud Computing: Distributed Internet Computing for IT and Scientific Research," IEEE Internet Computing Journal, vol. 13, issue. 5, pp. 10-13, September 2009. DOI: 10.1109/MIC.2009.103
.

[6] A. Williamson, "Comparing cloud computing providers," Cloud Comp. J., vol. 2, no. 3, pp. 3–5, 2009.

[7] Aman Bakshi, Yogesh B. Dujodwala, "Securing cloud from DDoS Attacks using Intrusion Detection System in Virtual Machine," ICCSN '10 Proceeding of the 2010 Second International Conference on Communication Software and networks, pp. 260-264, 2010, IEEE Computer Society, USA, 2010. ISBN: 978-0-7695-3961-4.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 494 - 499, 2014

An Evolutionary Based Data Mining technique in Engineering Faculty Evaluation using weka

Lingaraj.K
Dept of CSE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Ranjana.V
Dept of ISE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Pruthvi.H.M
Dept of ISE
Rao Bahadur Y
Mahabaleswarappa
Engineering college.
Cantonment Bellary, India

Abstract: Data mining, the extraction of hidden knowledge from large amounts of data repositories. Data mining is used in a vast area and numerous commercial data mining applications including retail sales, e-commerce, remote sensing, bioinformatics etc. Education is an essential element for the progress of country. Mining in educational environment is called Educational Data Mining. Educational data mining is concerned with developing new methods to discover knowledge from educational database. Educational data mining is concerned with developing new methods to discover knowledge from educational database. The main goal of this paper is gathering manageable experiences with data mining and also using of these experiences at E learning system and traditional education according to teacher evaluation. In this paper are verified hidden patterns of teacher evaluation by students and is predicted that which teachers will be invited to faculty classes and which teachers will be refusing and education managers due to evaluation reasons will cut the education contract with these teachers in next semesters? And what's effect of some items for examples Evaluation's score, Teacher's degree, Degree's type, Teaching experience, Acceptation to next semesters on teacher's evaluation?

Keywords: Data mining, WEKA , Classification, Clustering, Association rule, Data mining, Web mining.

1. INTRODUCTION

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration [1].

Data Mining is a non-trivial process of identifying valid, novel, useful and ultimately understandable patterns in data. Alternative names for data mining are Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archaeology, data dredging, information harvesting, business intelligence, etc. Data mining can be used in various applications [12]:

Banking: loan/credit card approval, predict good customers based on old customers, view the debt and revenue changes by month, by region, by sector, and by other factors , access statistical information such as maximum, minimum, total, average, trend, etc.

Telecommunication industry: identify potentially fraudulent users and their atypical usage patterns, detect attempts to gain fraudulent entry to customer accounts, discover unusual patterns which may need special attention, find usage patterns for a set of communication services by customer group, by month, etc., promote the sales of specific services, improve the availability of particular services in a region.

Retail Industry: Identify customer buying behaviors, discover customer shopping patterns and trends, improve the quality of customer service, achieve better customer retention and satisfaction, enhance goods consumption ratios, design more effective goods transportation and distribution policies

DNA analysis: compare the frequently occurring patterns of each class (e.g., diseased and healthy), identify gene sequence patterns that play roles in various diseases

Now a day, large quantities of data is being accumulated. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Data could be large in two senses: in terms of size & in terms of dimensionality. Also there is a huge gap from the stored data to the knowledge that could be construed from the data. Manual data analysis has been around for some time now, but it creates a bottleneck for large data analysis. The transition won't occur automatically; in this case, there is a need for data mining. Data Mining could help in a more in-depth knowledge about the data [2].

2. METHODOLOGY

Data mining is relatively a new technique to the world of information sciences. Successful implementation of this technique requires a sound methodology built on best practices. In this research study, has followed a popular data mining methodology called Cross Industry Standard Process for Data Mining (CRISP-DM), which is a six-step process [8]:

Problem description: Involves understanding project goals with business perspective, transforming this information into data mining problem description and

making project plan to reach the related goals .

Understanding the data: Involves identifying the sources of data, obtaining an initial set of data to assess the information coverage of the data for the problem on hand.

Preparing the data: Involves pre-processing, cleaning, and transforming the relevant data into a form that can be used by data mining algorithms.

Creating the models: Involves developing a wide range of models using comparable analytical techniques (i.e., selecting the appropriate modelling technique and setting the parameters related to the model to optimal values).

Evaluating the models: Involves evaluating and assessing the validity and the utility of the models against each other and against the goals of the study.

Using the model: Involves in such activities as deploying the models for use in decision making processes (i.e., making it a part of the decision support system/process).

A graphical representation of the methodology used in this study is shown in Figure 1.

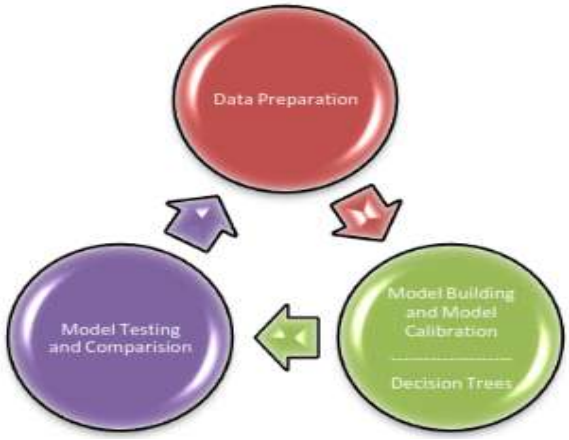


Fig.1.Agraphical illustration of the methodology employed in this study

2.1 DATA

In this study 104 records were used which is taken Sanandaj Daughter Vocational Faculty. Dataset have teacher' information such as Evaluation‘s score, Teacher’s degree, Degree’s type, Teaching experience, Acceptation.

TABLE 1: The List of Independent Variable used in the Study

Variable Name	Data Type	Description
Evaluation Score	Text	Evaluation’s Score
Teacher’s Degree	Text	Teacher’s Score
Degree type	Text	Teacher’s Degree type
Teaching Experience	Text	Teaching Experience of a Teacher
Acceptation	Text	Acceptation for a Teacher

Table 2. The list of independent variables and values used in this study

Variable Name	Data Type	Values
Evaluation Score	Nominal	{ Weak,Good,Excellent }
Teacher’s Degree	Nominal	{B.E,M.TECH,PHD}
Degree type	Nominal	{Full-Time, Part-Time}
Teaching Experience	Nominal	{ True, False}
Acceptation	Nominal	{ Yes, No }

Evaluation score of teachers which are studying in Sanandaj Daughter Vocational Faculty are represented by the word system. Score ranges of these words are shown in Table 3.

Table 3. The output variable (Evaluation score) used in the study

Raw-Score	Nominal Representation
Score<60	Weak
60<=Score<75	Good
75<=Score<100	Excellent

Table 4. The output variable (Teaching experience) used in the study

Raw-Years of Teaching	Nominal Representation
Years<3	False
Years>=3	True

2.2 Background

In this research are used WEKA and Data mining (Classification, clustering & Association Algorithm).The following subsections includes a brief overview of these topics

2.2.1 Weka

WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code [10]. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and Predictive modelling,

together with graphical user interfaces for easy access to this functionality [11].

It is freely available software. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any platform. WEKA has several standard data mining tasks, data pre-processing, clustering, classification, association, visualization, and feature selection.

2.2.2 Data Mining

Data mining is the process of discovering interesting knowledge from large amount of data stored in database, data warehouse or other information repositories. It includes various tasks such as classification, clustering, association rule etc.

2.2.3 Association Rule

Association rules are used to show the relationship between data items. Mining association rules allows finding rules of the form: If antecedent then (likely) consequent where antecedent and consequent are item sets which are sets of one or more items. Association rule generation consists of two separate steps: First, minimum support is applied to find all frequent item sets in a database. Second, these frequent item sets and the minimum confidence constraint are used to form rules [11]. Support & confidence are the normal method used to measure the quality of association rule. Support for the association rule X->Y is the percentage of transaction in the database that contains XUY [12]. Confidence for the association rule is X->Y is the ratio of the number of transaction that contains XUY to the number of transaction that contain X [7]. Association rule can be used in educational data mining and teacher's evaluation system for analyzing the learning data.

2.2.4 Classification

Classification is a data mining task that maps the data into predefined groups & classes. It is also called as supervised learning .It consists of two steps:

Model construction: It consists of set of predetermined classes. Each tuple /sample is assumed to belong to a predefined class. The set of tuple used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae.

Model usage: This model is used for classifying future or unknown objects. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur [11].

2.2.5 Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. Clustering can be considered the most important unsupervised learning technique. In educational data mining and teacher's evaluation system, clustering has been used to group the teachers according to their behavior e.g. clustering can be used to distinguish active teacher from non-active teacher according to their performance in activities.

3. ARCHTECTURE OF PROPOSED SYSTEM

In this paper, it is done a web base survey from 3000 students then it is prepared results of this survey for 201 teachers

3.1 The Explorer Interface of Weka

In WEKA application issue, this is probably the most confusing part of becoming familiar with WEKA because you are presented with quite a complex screen.

Initially "pre-process" will have been selected. This is the tab you select when you want to tell WEKA where to find the data set that you want to use.

WEKA processes data sets that are in its own ARFF format. Conveniently, the download will have set up a folder within the WEKA-3.6 folder called "data". This contains a selection of data files in ARFF format.

3.2 ARFF Format

You do not need to know about ARFF format unless you wish to convert data from other formats. However, it is useful to see the information that such files provide to WEKA.

```
@RELATION Engineering_Faculty

@ATTRIBUTE Evaluation_Score
      {Weak,Good,Excellent}

@ATTRIBUTE Teacher_Degree
      {BE,MTECH,PHD}

@ATTRIBUTE Degree_Type      {FT,PT}

@ATTRIBUTE Teaching_Experience
      {True,False}

@ATTRIBUTE Acceptation
      {Yes,No}

@DATA

Weak,BE,PT,False,No

Weak,BE,PT,True,No

Good,BE,PT,False,No

Good,BE,PT,True,No

Excellent,BE,PT,False,No

Excellent,BE,PT,True,No

Weak,BE,FT,False,No
```

```
Weak,BE,FT,True,No

Good,BE,FT,False,Yes

Good,BE,FT,True,Yes

Excellent,BE,FT,False,Yes

Excellent,BE,FT,True,Yes

Weak,MTECH,PT,False,No

Weak,MTECH,PT,True,No

Good,MTECH,PT,False,Yes

Good,MTECH,PT,True,Yes

Excellent,MTECH,PT,False,Yes

Excellent,MTECH,PT,True,Yes

Weak,MTECH,FT,False,No

Weak,MTECH,FT,True,No

Good,MTECH,FT,False,Yes

Good,MTECH,FT,True,Yes

Excellent,MTECH,FT,False,Yes

Excellent,MTECH,FT,True,Yes

Weak,PHD,PT,False,Yes

Weak,PHD,PT,True,Yes

Good,PHD,PT,False,Yes

Good,PHD,PT,True,Yes

Excellent,PHD,PT,False,Yes

Excellent,PHD,PT,True,Yes

Weak,PHD,FT,False,Yes

Weak,PHD,FT,True,Yes

Good,PHD,FT,False,Yes

Good,PHD,FT,True,Yes

Excellent,PHD,FT,False,Yes

Excellent,PHD,FT,True,Yes
```

Fig.2. ARFF file format for dataset in this paper .

It consists of three parts. The @relation line gives the dataset a name for use within Weak. The @attribute lines declare the attributes of the examples in the data set (Note that this will include the classification attribute). Each line specifies an attribute's name and the values it may take. In this paper the attributes have nominal values so these are listed explicitly. In other cases attributes might take numbers as values and in such cases this would be indicated as in the following example:

@attribute Teacher_degree numeric

The remainder of the file lists the actual examples, in comma separated format; the attribute values appear in the order in which they are declared above.

3.3 This Opening a DataSet

In the Explorer window, click on "Open file" and then use the browser to navigate to the 'data' folder within the WEKA-3.6 folder. Select the file called Teacher_evaluation.arff. (This is in fact the file listed above).

This is a 'Teacher evaluation' data set, like the ones used in class for demonstration purposes. In this case, the normal usage is to learn to predict the 'Acceptation' attribute from four others providing information about the Teacher evaluation.

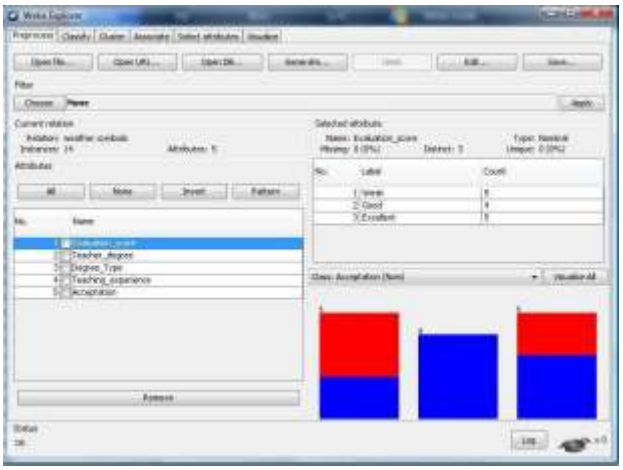


Fig3. Opening 'Teacher evaluation' Data Set in Weka

Most of the information it displays is self-explanatory: it is

a data set containing 14 examples (instances) each of which has 5 attributes. The 'Acceptation' attribute has been suggested as the class attribute (i.e. the one that will be predicted from the others).

Most of the right hand of the window gives you information about the attributes. Initially, it will give you information about the first attribute ('Evaluation Score').

This shows that it has 3 possible values tells you how many there are of each value. The bar chart in the lower right shows how the values of the suggested class variable are distributed across the possible values of the 'Evaluation Score'.

If you click on 'Teacher Degree' in the panel on the left, the information about the 'Evaluation Score' attribute will be replaced by the corresponding information about the Teacher Degree attribute.

3.4 Choosing a Classifier

Next it is necessary to select a machine learning procedure to apply to this data. The task is classification so click on the 'classify' tab near the top of the Explorer window.

The window should now look like this:

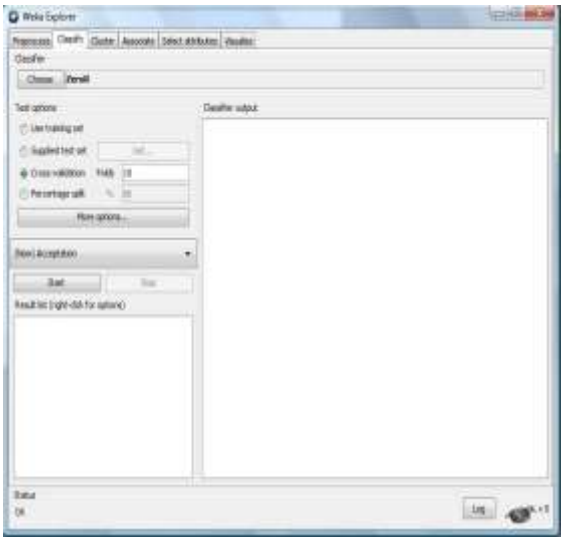


Fig4. Selecting Classifying algorithm in Weka for case study

By default, a classifier called ZeroR has been selected. A different classifier is desired so click on the Choose button. A hierarchical popup menu appears. Click to expand 'Trees', which appears at the end of this menu, then select J48 which is the decision tree program.

The Explorer window now looks like this indicating that J48 has been chosen

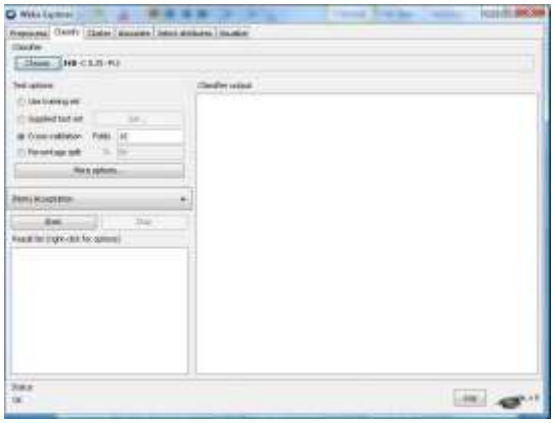


Fig5. Decision Tree with J48 tree with Weka.

The other information alongside J48 indicates the parameters that have been chosen for the program. This paper will ignore these.

3.5 Choosing the Experimental Procedure

The panel headed 'Test options' allows the user to choose the experimental procedure. This paper shall has more to say about this later in the course. For the present exercise click on 'Use training set'. (This will simply build a tree using all the examples in the data set).

The small panel half way down the left hand side indicates which attribute will be used as the classification attribute. It will currently be set to 'Acceptation'. (Note that this is what actually determines the classification attribute – the 'class' attribute on the pre-process screen is simply to allow you to see how a variable appears to depend on the values of other attributes)

3.6 Running the Decision Tree Program

Now, simply click the start button and the program will run.

The results will appear in the scrollable panel on the right of the Explorer window. Normally these will be of great interest but for present purposes all this paper needs to notice is that the resulting tree classified all 14 training examples correctly. The tree constructed is presented in indented format, a common method large for trees:

The panel on the lower left headed 'Result list (right-click for options)' provides access to more information about the results. Right clicking will produce a menu from which 'Visualize Tree' can be selected. This will display the decision tree in a more attractive format:

Note that this form of display is really only suitable for small trees. Comparing the two forms should make it clear how the indented format works.

```
J48 pruned tree
------------------

Teacher_Degree = BE
|   Degree_Type = FT
|   |   Evaluation_Score = Weak: No (2.0)
|   |   Evaluation_Score = Good: Yes (2.0)
|   |   Evaluation_Score = Excellent: Yes (2.0)
|   Degree_Type = PT: No (6.0)
Teacher_Degree = MTECH
|   Evaluation_Score = Weak: No (4.0)
|   Evaluation_Score = Good: Yes (4.0)
|   Evaluation_Score = Excellent: Yes (4.0)
Teacher_Degree = PHD: Yes (12.0)

Number of Leaves  :     8

Size of the tree :      12


Time taken to build model: 0.02 seconds
```

Fig6. Decision Tree with J48 tree with Weka.

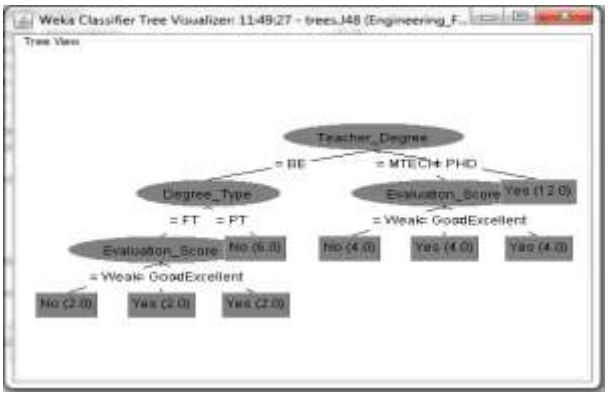


Fig7. Visualize Tree with J48 tree with Weka.

The panel on the lower left headed 'Result list (right-click for options)' provides access to more information about the results. Right clicking will produce a menu from which 'Cost/Benefit Analysis' can be selected. This will display the decision tree in a more attractive format

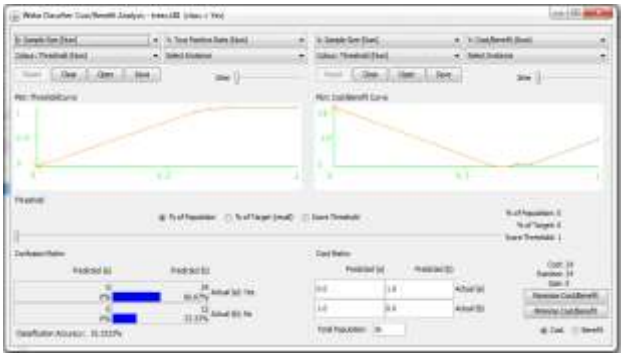


Fig8. Cost/Benefit Analysis with J48 tree with Weka

4. CONCLUSION

At teacher's evaluation, evaluation's score of students is very important factor that many universities gather this information on performance of teachers. New rules by using data mining and J48 tree as a decision tree in this paper are results that education managers could use these rules in future decisions to submit new teachers and continue with elected old teachers. Correctness of this rules depending variety of datasets and statistical instances can vary. But data mining tools such as WEKA as is showed in this paper can conclude variety results that help education managers in universities. These results will be used by managers in decision-making.

5. REFERENCES

[1] Han,J. and Kamber, M., "Data Mining: Concepts andTechniques", 2nd edition.

[2] Sunita B Aher, Mr. LOBO L.M.R.J. Data Mining in Educational System using WEKA, IJCA,2011

[3] A. P. Sanjeev ve J. M, Zytkow. "Discovering Enrollment Knowledge in University Databases," 1th Conference on KDD (Montreal.20-21 August 1995), 246.

[4] K. Becker, C. Ghedini ve E.L. Terra, "Using KDD to analyse the impact of curriculum revisions in a Brazilian university," SPIE 14th Annual International Conference (Orlando. April 2000), 412.

[5] J. Luan, "Data Mining, Knowledge Management in Higher Education, Potential Applications", 42nd Associate of Institutional Research International Conference (Toronto,Canada: 2002), 1.

[6] Ş.Erdoğan, M. Timor, "A Data Mining Application in a Student Database," Havacılık ve Uzay Dergisi. Cilt No 2,Sayı 2: 57-64, (July 2005), 57.

[7] M.Vranić, D. Pintar, Z.Skoćır, "The Use of Data Mining in Education Environment," ConTEL 2007 (Zagrep 13-15 June 2007), 243.

[8] C. Shearer, "The CRISP-DM model: The new blueprint for data mining" Journal of Data Warehousing, (2000). 5: 13-22.

[9] "Data Mining Introductory and Advanced Topics" by Margaret H. Dunham

[10] WEKA (2007). http://www.cs.waikato.ac.nz/ml/WEKA/.

[11] International Educational Data Mining, available at http://www.educationaldatamining.org/

[12] Sunita B Aher and Lobo L.M.R.J.. Data Mining in Educational System using WEKA. IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT) (3):20-25, 2011. Published by Foundation of Computer Science, New York, USA (ISBN: 978-93-80864-71-13)

[13] Fateh Ahmadi and M.E.Shiri Ahmad abadi "Data mining by teacher evaluation using Weka" IJCA ISSN:-0975-8887,page.No-14 to 18 Volume 63-No-10, Feb 2013

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 500 - 504, 2014

Smooth Ordering of Patches and its Application

Dhanya Pious
Dept. of Computer Science
GEC.Idukki,Kerala,India

Philumon Joseph
Dept. of Computer Science
GEC.Idukki,Kerala,India

Abstract: Here a new image processing technique is used. For a given corrupted image, we extract all patches with overlaps, refer to these as coordinates in high dimensional space, and order them such that they are chained in the "shortest possible path". The obtained ordering applied to the corrupted image implies a permutation of the image pixels to what should be a regular signal. This technique is used for image denoising,image inpainting and image deblurring. Here we include an encryption scheme using Bakers algorithm to enhance security of image under military or medical section.

Keywords:Inpainting,Denoising,Clustering.

1. INTRODUCTION

Image processing using local patches has become very popular and was shown to be highly effective and useful. The main idea behind these and many other algorithms are the same. When we are given the image to be processed, all possible patches with over-laps need to be extracted. Now these patches are typically very small when we are comparing it to the original image size (a typical patch size would be 88 pixels). The processing itself proceeds by operating on these patches and interrelations between them are made use. The modified patches (or sometimes only their centre pixels) are then put back into the original image canvas to form the resulting image.

There are numerous ways in which the relations between patches can be considered. Weighted averaging of pixels with similar surrounding patches, as in the case of NL-Means algorithm , clustering the patches into disjoint sets and treating each set differently, seeking a representative dictionary for the patches and using it to sparsely represent them, gathering groups of similar patches and applying a sparsifying transform on them. A common theme to many of these methods is the expectation that every patch taken from the image may find similar ones extracted elsewhere in the image.

Here the problem addressed is of reconstructing and enhancing an image given the noisy observations gathered by a digital camera sensor. There are several methods in which we view both denoising and demosaicking as image reconstruction problems, and propose a novel image model that combines two now classical techniques into a single framework.The non-local means approach to image restoration explicitly exploits self-similarities in natural images to average out the noise among similar patches, whereas sparse coding en-codes natural image statistics by decomposing each image patch into a linear combination of a few elements from a basis set called a dictionary. The two applications implemented here are image denoising and image inpainting. Image deblurring can be considered as another application in enhancement.

2. PREVIOUS WORK

Let's go through few of the previous methods that have remarkably removed noise form images and reconstructed them.

2.1. Non Local Means Filtering

The goal of image denoising methods is to recover the original image from a noisy image,

$$v(i)=u(i)+n(i)$$

where v(i)is the observed value, u(i) is the "true" value and n (i)is the noise perturbation at a pixel i. The best simple way to model the effect of noise on a digital image is to add a Gaussian white noise. In that case,n (i)are Gaussian values with zero mean and variance s^2 . Formally we define a denoising method D_h as decomposition

$$v=D_h v+n(D_h,v)$$

where v is the noisy image and h is a filtering parameter which usually depends on the standard deviation of the noise. Ideally, D_h v is smoother than v and n (D_h ,v) looks like the realization of a white noise. The de-noising methods

should not alter the original image u . Now, most denoising methods degrade or remove the fine details and textures of u .To overcome the defects non-local means approach was introduced.

2.2. Clustering-Based Denoising With Locally Learned Dictionaries

Here algorithm aims to erase the limitations like, the static nature of the dictionary, and the constancy of the approximation order across the image. We go about this task with a clustering-based[3] algorithm which consists of three stages the clustering step where the image is clustered using features that capture the local structure of the underlying image data (patches of pixels from the image), the dictionary selection stage where we form an optimized dictionary that adapts to the geometric structure of the image patches in each cluster; and, finally, the coefficient calculation stage where the coefficients for the linear combination of dictionary atoms are estimated, subject to the (steering) kernel weights. In the following sections, we describe and motivate each of the above stages in detail.

In the initial stage, our algorithm attempts to perform clustering to identify regions of similar structure in the image. To perform clustering we need to first identify informative features from the image. While the choice of features remains an open research problem, in many cases the features are directly computed from the input image. Commonly used low level features to identify similar pixels (or patches) are pixel intensities, gradient information etc., or a combination of these. The use of such features directly from the input image is not advisable for our denoising problem due to their instability in the presence of noise. However, it has been observed that the steering weights computed in a neighbourhood are robust to the presence of significant amounts of noise. These weights are roughly representative of the under-lying local data structure. Thus, clustering is performed using feature vectors of size N X 1 for each local steering kernel computed over a NX N window centered at pixel in the image. That is to say, every pixel of the image is mapped to a feature vector of size N X 1.At the end of this stage we expect the image to be divided into not necessarily contiguous (W_k) regions , each containing patches of similar structure. Hence, the entire noisy image can be thought to be composed of a union of such clusters Distance Metric: Before we proceed to perform clustering on the weights, we need to specify a metric to calculate the distance between two weight functions.Once the clusters are formed, we proceed to form a dictionary best suited to each cluster independently. For each cluster we intend to find a dictionary that best describes the structure of the under-lying data within that cluster. In other words, for each image patch in a cluster W_k we want to find an estimate $Y_i^{\ 0}$ which best approximates the input vectorized patch Y_i Coefficient Calculation Once the dictionary is formed for each cluster, we proceed to estimate the b_i parameters under a regression framework. We pose this as an optimization problem. The dictionary now is adapted to a specific class of image structure that is captured by each cluster. Furthermore, the number of principal components or dictionary atoms that will be needed to fit a prespecified percentage of data varies across the different clusters. The patches thus estimated are overlapping, so we should ideally optimally combine the overlapping regions somehow to form the final image.

2.3. Non-local Sparse Models for Image Restoration

Here both denoising and demosaicking as image reconstruction problems, and propose a novel image model that combines two now classical techniques into a single framework: The non-local means approach to image restoration explicitly[4],[5] exploits self-similarities in natural images to average out the noise among similar patches, whereas sparse coding encodes natural image statistics by decomposing each image patch into a linear combination of a few elements from a basis set called a dictionary. Although fixed dictionaries based on various types of wavelets have been used in this setting, sparse decompositions based on learned, possibly over complete, dictionaries adapted to specific images have been shown to provide better results in practice. We propose to extend and combine these two approaches by using simultaneous sparse coding to impose that similar patches share the same dictionary elements in their sparse decomposition. To the best of our knowledge, this is the first time that the corresponding models of image self-similarities are explicitly used in a common setting with learned dictionaries. Experiments with images corrupted by synthetic or real noise show that the proposed method outperforms the state of the art in both image denoising and image demosaicking tasks, making it possible to effectively restore raw images from digital cameras at a reasonable speed and memory cost.

2.4. From Patch Likelihoods to Patch Restoration

For many patch priors a[2] closed form of log likelihood, Bayesian Least Squares (BLS) and Maxi-mum A-Posteriori (MAP) estimates can be easily calculated. Given that, we start with a simple question: Do priors that give high likelihood for natural image patches also produce good results in a restoration task such as denoising, for many

popular MRF priors, neither the log likelihood nor the MAP estimate can be calculated exactly.Here we compare several popular priors, trained over 50,000 8 X 8 patches randomly sampled from the training set. We compare the log likelihood each model gives on a set of unseen natural image patches (sampled from the test set and the performance of each model in patch denoising using MAP estimates. The models we use here are: Independent pixels with learned marginals (Ind. Pixel), Multivariate Gaussian over pixels with learned covariance (MVG), Independent PCA with learned (non-Gaussian) marginals and ICA with learned marginals. The results for each of the models can be seen in Figure As can be seen, the higher the likelihood a model gives for a set of patches, the better it is in denoising them when they are corrupted.

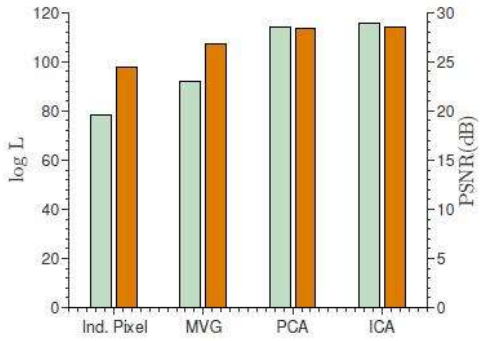


Figure 1. Likelihood of Models

3. IMPLEMENTATION

The smooth ordering of patches is done and this can be used for many applications. The applications are like the image denoising, image inpainting image de-blurring etc. The method for the smooth ordering of the image is reordering the patches of the image. Thus reconstructing the original image.For implementation, we have an image with us. Now we are adding some disturbances into it for making it a corrupted image. Now this is the image on which we have to apply the smooth reordering of patches and perform image denoising and image in-painting. Y is the original image and Z the image after adding impurities.Z could be having noise or it could have missing pixels. The corrupted image then satisfies

$$z = My + v$$

For reconstructing Y from z, a permutation matrix P is used. P is a N X N matrix, where N = N1 X N2.Now Y is of size N1 X N2. We assume that when P is applied to the target signal y, it produces a smooth signal

$$yp = Py$$

We will explain how such a matrix may be obtained using the image patches in Section II-B. We start by applying P to z and obtain

$$Z\ p = Pz$$

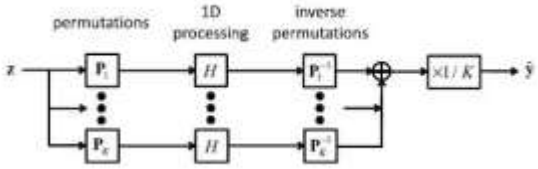


Figure 2. The basic image processing scheme

Next, we take advantage of our prior knowledge that yp should be smooth . and apply a ”simple” 1D smoothing Operator H on Zp, such as 1D interpolation or filtering. Finally, we apply P1 to the result, and obtain the re-constructed image.

3.1. Permutation matrix

To design a matrix P, that would produce a smooth signal when it is applied to the target image y is as follows.. When the image Y is known, the solution is to reorder it as a vector, and then apply a simple sort operation on the obtained vector.

3.2. Image Inpainting

The problem of image inpainting is of the recovery of missing pixels in the given image. Here v = 0, and M is a diagonal matrix of size N X N which contains ones and zeroes in its main diagonal corresponding to existing and missing pixels, correspondingly. Each patch may contain missing pixels, and we denote by Si the set of indices of non-missing pixels in the patch xi . We choose the distance measure between patches xi and x j to be the average of squared differences be-tween existing pixels that share the same location in both patches. First the matrix P is calculated. when a patch does not share pixels with any of the unvisited patches, the next patch in the path is chosen to be its nearest spatial neighbour. An operator H is used, which recovers the missing values using cubic spline interpolation. We apply the matrix P 1 on the resulting vectors and obtain the estimated subimages y j . The final estimate is obtained from these subimages. We improve our results by applying two additional iterations of a modified version of this inpainting scheme, where the only difference

is that we rebuild P using reconstructed (and thus full) patches.

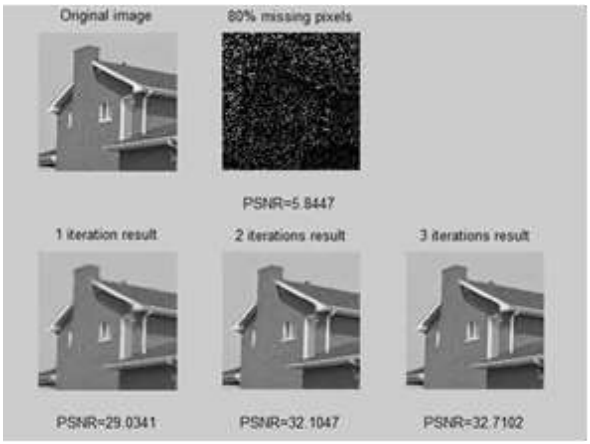


Figure 3. The Image Inpainting

The figure shows the original image, the corrupted one in the first row. Each iteration results of the permutation matrix are shown in the bottom row.

3.3 Image denoising

In image denoising, the recovery of an image from its noisy version is carried out. In that case M = I and the corrupted image satisfies z = y + v. The patches xi may contain noise, and we choose the distance measure between xi and x j to be the squared Euclidean distance divided by n. A 1D linear shift invariant filter, is used for this purpose. There are two filters to switch between based on the patch content. The smooth areas in the image are treated differently than areas with edges or texture. First patches are partitioned into those smooth Ss and those with edges and texture Se.

Next divide the sub images also into two signals. A vector of length |Ss| that contain the smooth patches and a vector of length |Se| .Now make use of the nearest neighbour search method and extract the sub images from both divisions. Now find the filters hs,he each of length NH. Now define a filter h of length 2Nh.The vector h stores the filter taps to be designed. We substitute and obtain the reconstructed image.

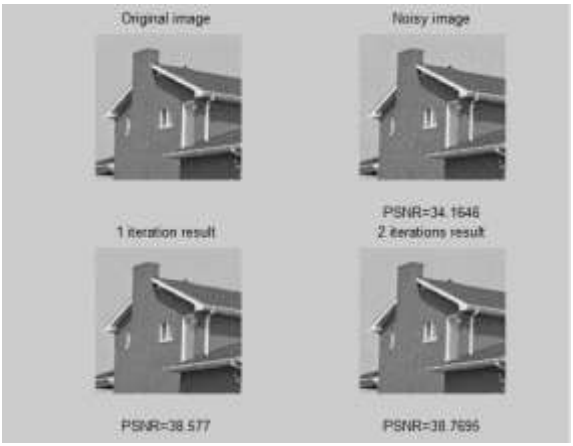


Figure 4. The Image Denoising

The figure shows the image denoising scheme. The top row shows the image initially, and after adding the noise from left. In second row we get to see the results of first and the second iterations of the denoising algorithm described above. We have also included an encryption scheme in this inorder to enhance the security of the image being used in the scheme.

3.4 Encryption

An Encryption Scheme known as Bakers Algorithm is implemented on the image after denoising and inpainting. This encryption makes these images free to be used for military purposes where the secrecy of recovered image is necessary. Encryption is implemented as follows.

1. The M X M square matrix is divided into rectangles of width v_i and number of elements M.
2. The elements in each rectangle are rearranged to a row in the permuted rectangle. Rectangles are taken from right to left beginning with upper rectangles, and then lower ones.
3. Inside each rectangle, the scan begins from the bottom left corner towards upper elements.

4. CONCLUSIONS

We have proposed a new image processing scheme which is based on smooth 1D ordering of the pixels in the given image. We have shown that using a carefully designed permutation matrices and simple and intuitive 1D operations such as linear filtering and interpolation, the proposed scheme can be used for image denoising and inpainting, where it achieves high quality results. The Bakers algorithm used for encryption enhances the security of images in military ,medical and many other fields where in security of image reconstructed from patches is of high importance.

There are several research directions to extend this work that we are currently considering. The first is to make use of the distances between the patches not only to find the ordering

matrices, but also in the re-construction process of the subimages. These distances carry additional information which might improve the obtained results. Improvements can also be made to the patch ordering scheme itself. We have seen in that this scheme performs poorly near the end of the found path, when only a small number of unvisited patches remain. A possible solution could be to develop a scheme which allows patches to be revisited more than once. A different direction is to develop new image processing algorithms which involve optimization problems in which the 1D image reordering act as regularizers. These may both improve the image denoising and inpainting results, and allow to tackle other applications such as image deblurring.
The Research highlights of this paper are

- New method for image denoising
- New fast method for image inpainting.
- The permutation matrix formed.
- Encryption provided for securing the image details

5. ACKNOWLEDGEMENTS

I thank the authors of all the previous implementation methodologies, for the fruitful discussions and advices, which helped in developing the presented work. I also thank the anonymous reviewers for their helpful comments.

6. REFERENCES

[1] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, ”Non-local sparse models for image restoration,” in Proc. IEEE 12th Int. Conf. Comput. Vis., Sep.Oct. 2009, pp. 2272-2279.

[2] D. Zoran and Y. Weiss, ”From learning models of natural image patches to whole image restoration,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 479-486.

[3] P.Chatterjee and P. Milanfar, ”Clustering-based denoising with locally learned dictionaries,” IEEE Trans. Image Process., vol. 18, no. 7, pp. 1438-1451, Jul. 2009.

[4] M. Elad and M. Aharon, ”Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[5] R. Zeyde, M. Elad, and M. Protter, ”On single image scale-up using sparse-representations,” in Proc. 7th Int. Conf. Curves Surf., 2012, pp. 711-730

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 505 - 509, 2014

How Mindreading Computer Work And How It Is Useful In Different Working Areas?

Komal Tomar
DIT University
Dehradun,India

Abstract:Mind reading is the ability to infer other people's mental state and use that to make sense of and predict their behavior and also to understand both oneself and other agents in terms of beliefs, desires and intentions[15]. A lack of or impairment in the theory of mind (mindblindness) is thought to be the primary inhibitor of emotion understanding and social intelligence in individuals with autism[6] .The goal in building mind reading machines is to enable computer technologies to understand and react to people's emotions and mental states. In this paper we present how to read minds using different techniques. The paper presents a mind-reading architecture based on an Eye Language Interpreter and Baron-Cohen's Mindreading System. This architecture enables the recognition and prediction of complex mental states, allowing for more natural man-machine interaction. This paper also presents where mind reading useful.

Keywords: mindblindness, autism, mind reading architecture, impairment, Eye language interpreter, man machine interaction.

1. INTRODUCTION

A theory of mind is a representational set of abilities that allows one to mind read. It is the ability to attribute mental states to others, and use that to understand the actions and expressions of others within an intentional or goal-directed framework (Dennett in [11] refers to that as the intentional stance).

Facial movements and eye language in particular play a crucial role in recognizing basic emotions (such as happiness, sadness, disgust and fear) as well as "cognitive" or complex mental states (such as distrust, recognize, scheme, admire, interest, thoughtfulness, etc.) [9]. Using a digital video camera, the mind-reading computer ppt system analyzes a person's facial expressions in real time and infers that person's underlying mental state, such as whether he or she is agreeing or disagreeing, interested or bored, thinking or confused.

Prior knowledge of how particular mental states are expressed in the face is combined with analysis of facial expressions and head gestures occurring in real time. The model represents these at different granularities, starting with face and head movements and building those in time and in space to form a clearer model of what mental state is being represented. Software from Nevenvision identifies 24 feature points on the face and tracks them in real time. Movement, shape and colour are then analyzed to identify gestures like a smile or eyebrows being raised. Combinations of these occurring over time indicate mental states. For example, a combination of a head nod, with a smile and eyebrows raised might mean interest. The relationship between observable head and facial displays and the corresponding hidden mental states over time is modeled using Dynamic Bayesian Networks.

Current projects in Cambridge are considering further inputs such as body posture and gestures to improve the inference. We can then use the same models to control the animation of cartoon avatars. We are also looking at the use of mind-reading to support on-line shopping and learning systems.

The mind-reading computer system may also be used to monitor and suggest improvements in human-human interaction.

2. HOW MIND READING IS DONE?

2.1. Futuristic head:

The mind reading actually involves measuring the volume and oxygen level of the blood around the subject's brain, using technology called functional near-infrared spectroscopy (fNIRS).

The user wears a sort of futuristic headband[1](see fig 1) that sends light in that spectrum into the tissues of the head where it is absorbed by active, blood-filled tissues. The headband then measures how much light was not absorbed, letting the computer gauge the metabolic demands that the brain is making.

[1] Headband is a type of headgear which a user can wear, fig 1 shows a user wearing head gear.

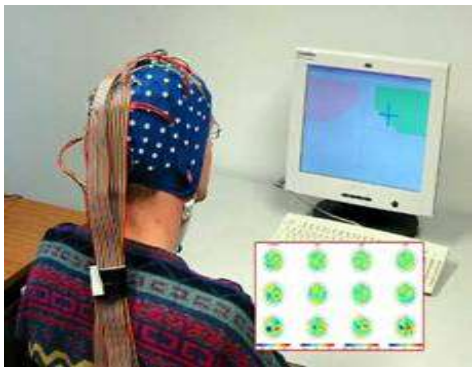


Figure 1

Figure 2

Future Headband

2.2. Brain Chip:

It is an independent processor linked to the neurocomputer built to house an artificial intelligence. The artificial intelligence program has access to the sensory data and information in the neurocomputer, and can "read" surface thoughts of the owner. Having a (or several) as advisor/secretary/partner is becoming more and more common, although most people realy on an external artificial intelligence system and a wireless neural connection[5]. It is not uncommon for users to get a motoric shunt to give the artificial intelligence the ability to control the body. Chips with monitoring artificial intelligence are sometimes used for or behavior correction in Landfall.

Figure 3

Scientists successfully implants chips that control Brain allowing thoughts, memory and behavior to be transferred from one brain to another brain. In a scene right out of a George Orwell novel, a team of scientists working in the fields of "neural engineering" and "Biomimetic MicroElectronic Systems" have successfully created a chip that controls the brain and can be used as a storage device for long-term memories. In studies the scientists have been able to record, download and transfer memories into other hosts with the same chip implanted. The advancement in technology brings the world one step closer to a global police state and the reality of absolute mind control.

3. MINDREADING ARCHITECTURE

While a number of researchers are building a theory of mind into humanoid robots [13] and [10] as a tool to test and evaluate developmental theories, we propose integrating a comparable model into mainstream interfaces as a novel approach to building social and emotion intelligence in interfaces. Seamless, non-obtrusive and implicit mind reading capabilities in mainstream interfaces could open new possibilities for man-machine interaction. In this section we introduce an architecture for developing Mindreading Interfaces. It integrates an automated eye language interpreter and a theory of mind model to enable the recognition and prediction of the user's mental state.

Eye Language Interpreter:

We pursue Baron-Cohen's research on a Language of the Eyes [9] to develop an automated eye language interpreter

that recognizes eye expressions off image sequences and interprets them as mental states. We are interested in being

able to recognize a limited set of complex mental states that go beyond the 6 basic emotions. We use the Facial Action Coding System (FACS) developed by Ekman & Friesen [12] to identify expressions of the eye. FACS is based on the enumeration of all action units of a face that causes facial movements. We identify head position and orientation, eye (and eye-brow) position and gaze actions units as the most relevant ones in addressing the problem of understanding eye expressions. We then develop a mapping between mental states and eye expressions.

The Mindreading System

We use Baron-Cohen's Mindreading System [6], which consists of four modular components for our implementation. We perform a number of projections specific to HCI to accommodate the differences in interaction format and modalities.

The Intentionality Detector (ID) interprets the self-propelled motion of stimuli in terms of primitive volitional mental states of goal or desire. It builds dyadic representations, which specify the relation (desire or goal) between an agent and an object (or another agent). We suggest using stimuli from the pointing device to build the dyadic representations of ID.

Eye-Direction Detector (EDD) detects the presence of eye-like stimuli in the visual field, computes whether the eyes are looking at it or at something else, and interprets gaze direction as a perceptual state. EDD codes dyadic representational states of the form [Agent-looking at-me] and [Agent-looking at-not me]. The self is mapped to elements of the user's computing environment.

The Shared Attention Mechanism (SAM) is held to be necessary for the development and production of joint-attention behaviors. SAM links ID to EDD by importing volitional terms from dyadic representations, into the relation slot of triadic representations, such that a person's goal or desire can be read from their eye-direction. SAM is also responsible for triggering ToMM.

The theory of Mind Mechanism or (ToMM) is responsible for our everyday ability to make sense of behavior in terms of mental states, and predicts an Agent's behavior on the basis of such states. It represents the full range of mental state concepts, and integrates mental state knowledge into a coherent and usable "theory" for humans to employ.

Enabling Technologies : Automated Facial Feature Analysis

In order to be able to fully implement the Eye Language Interpreter and the Mindreading System, we need the services of a number of enabling technologies. We identify automated facial feature analysis as one of those technologies, and describe ongoing work on the implementation of specific work in that area. We are developing an automated facial feature tracking methodology that tracks spontaneous expression over time. The module should detect a face, locate features within the face, and then track the motion of those features across frames of an image sequence. The methodology is used to implement the Eye Language Interpreter and the Eye Direction Detector.

4. RECOMMENDER METHOD

Mind reading allows us to make sense of other people's behavior, predict what they might do next, and how they might feel. A lack of or impairment in mind reading abilities are thought to be the primary inhibitor of emotion and social understanding in people diagnosed with autism (e.g. Baron-Cohen et. al [6]). People employ a variety of nonverbal communication cues to infer underlying mental states, including voice, posture and the face. The human face in particular provides one of the most powerful, versatile and natural means of communicating a wide array of mental states. One subset comprises cognitive mental states such as thinking, deciding and confused, which involve both an affective and intellectual component [9].

Mindreading can be used for shopping purpose using mental states such as thinking, deciding, voice, posture and face. Suppose when we go in the new mall, we don't know anything about it like where is the Madam showroom or where is the Dunking Donuts parlor. So due to this reason we have to roam all over the mall to find a particular place. This over all headache can be overpower using mindreading technique. This can be understand by taking an example. When we go in a mall or in a shop there is scanner before entering in the mall. We can use the concept of mindreading in this scanner. We can use headband during this scanning process.

Mind reading involves measuring the volume and oxygen level of the blood around the subject's brain.So during scanning process, the user wears a sort of futuristic headband when he/she walk through the scanner where that sends light in that spectrum into the tissues of the head where it is absorbed by active, blood-filled tissues. The headband then measures how much light was not absorbed, letting the computer gauge the metabolic demands that the brain is making. When computer know the demands that the brain is making then by analyzing this he make a list of the showrooms or where ever user want to go with the floor numbers and with proper directions. So in this way our time save. No more roaming on different floors of mall to find a showroom or anything else.

5. WHY MINDREADING?

Imagine a future where we are surrounded with mobile phones, cars and online services that can read our minds and react to our moods. How would that change our use of technology and our lives?

Mind-reading can also support on-line shopping and learning systems .There are many uses of minreading(See table 1)

Table 1 . Different working areas of mindreading

WORKING AREAS	USES
MILITARY AREAS	A science fiction fantasy – the "Thought Police" – where the government reads people's memories and thoughts and then rehabilitate them through torture before they ever even commit a crime based on a statistical computer.
MEDICAL AREAS	1. Mind-Reading Computer Gives Voice, Movement to the Paralyzed: New technology may help those who are locked inside their own bodies (like in locked-in-syndrome). 2. Mind reader computer can communicate with the patients in coma: Canadian researchers have developed an interesting Computer which can read mind and could be helpful to communicate with the people who are in coma. This can be called a Computer Mind reader.
GADGETS	Mindreading technique can also used in mobile phones, cars, keyboards and mouse etc.

5.1. Mind-Reading Computer Gives Voice, Movement to the Paralyzed:

New technology may help those who are locked inside their own bodies. In locked-in syndrome, a condition in which people with normal cognitive brain activity suffer severe paralysis, often from injuries or an illness such as Lou Gehrig's disease. Boston University neuroscientist Frank Guenther works with the National Science Foundation's Center of Excellence for Learning in Education, Science and Technology (CELEST). Its purpose is to synthesize the experimental modeling and technological approaches to research in order to understand how the brain learns as a whole system. In particular, Guenther's research is looking at how brain regions interact, with the hope of melding mind and machine, and ultimately making life much better for people with "locked-in syndrome".

5.2. Mind-reading program translates brain activity into words:

Algorithms translated the brain activity associated with hearing 'Waldo', 'structure', 'doubt' and 'property' into recognizable words. Scientists have picked up fragments of people's thoughts by decoding the brain activity caused by words that they hear.

The remarkable feat has given researchers fresh insight into how the brain processes language, and raises the tantalizing prospect of devices that can return speech to the speechless. Though in its infancy, the work paves the way for brain implants that could monitor a person's thoughts and speak words and sentences as they imagine them. Such devices could transform the lives of thousands of people who lose the ability to speak as a result of a stroke or other medical conditions.

Experiments on 15 patients in the US showed that a computer could decipher their brain activity and play back words they heard, though at times the words were difficult to recognize.

5.3. Mind reader computer can communicate with the patients in coma:

Canadian researchers have developed an interesting Computer which can read mind and could be helpful to communicate with the people who are in coma. This can be called a Computer Mind reader[14]. A team of researchers from the University of Western Ontario were using the neuroimaging technique to read human thoughts brain activity while giving the specific answer in Yes or No.

In their study they asked very simple questions to the participants and told to concentrate only on the response either Yes or No. They were asked questions like "Are you married?" , "Do you like ice cream?" and answers the question which were asked in scanner by only concentrating upon the word they want to speak. By analyzing the activity of their brain they were able to decode the answers perfectly for every single person and for this time window is attached for communication with the brain computer interfaces answers in between 3 min of scanning. This technique is used to talk with the patients who are not able to response because of unconscious state of their mind.

5.4. Mind reading computer can be used for police and military purpose:

A science fiction fantasy – the "Thought Police" – where the government reads people's memories and thoughts and then rehabilitate them through torture before they ever even commit a crime based on a statistical computer analysis showing people with certain types of thoughts are likely to commit a certain type of crime in the future. We already pre-emptively invade nations and torture alleged terrorist suspects with absolutely no due process of law, so the idea

of pre-emptively torturing a terrorist suspect before hand to prevent them from committing an act of terrorism in the future really isn't that far fetched of an idea.

6. CONCLUSION

Mindreading is the ability to infer other people's mental state and use that to make sense of and predict their behavior. A lack of or impairment in the theory of mind (mindblindness) is the primary inhibitor of emotion understanding and social intelligence in individuals with autism [7].In this paper we conclude how mind reading is done using some techniques. And different working areas of mindreading computer. A mind-reading architecture based on an Eye Language Interpreter and Baron-Cohen's Mindreading System is developed for the recognition and prediction of complex emotion states. And also proposed one more use of mind reading computer in real world. We believe that seamless, non- obtrusive and implicit mind-reading capabilities in mainstream interfaces will open new possibilities for intelligent and effective interfaces.

7. REFRENCES

1. Mindreading computers-Wikipedia-the free encyclopedia.
2. Kevin brown , 2011. IBM research.
3. Ian Sample,2012. Mindreading translate brain activity into words.
4. Baron-Cohen, S. 1993. From attention-goal psychology to belief-desire psychology: the development of a theory of mind and its dysfunction. Understanding other minds: perspectives from autism. Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D.J. Eds. Oxford University Press.
5. Mindtech Sweden, Mind control mind computer.
6. Baron-Cohen, S. 1994. How to build a baby that can read minds: Cognitive mechanisms in mindreading. Current Psychology of Cognition 13 (5): 513-552.
7. Baron-Cohen, S. 1995. Mindblindness: an essay on autism and theory of mind. MIT Press.
8. Julia Layton,2011.Have scientists found a way to read your mind?
9. Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. 1997. Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. Visual Cognition 4 (3): 311-331.
10. Deak, G., Fasel, I., and Movellan, J.R. 2001. The Emergence of Shared Attention: Using Robots to Test Developmental Theories. First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems.
11. Dennett, D. 1987. The Intentional Stance. MIT Press.
12. Ekman, P. and Friesen, W. 1978. Facial Action Coding System: A technique for the measurement of facial movement. Consulting Psychologists Press.
13. Scassellati, B. 2000. Models of Social Development Using a Humanoid Robot. Biorobotics. Webb, B. and Consi, T. Ed. MIT Press.
14. Mark prigg .Mind reading computer could communicate with the people in coma.
15. Paul bello,2012. Cognitive Foundations for a Computational Theory of Mindreading.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 510 - 514, 2014

Information Security in Cloud Computing

E.Kesavulu Reddy
Dept.of Computer Science,
S.V.University College of CM & CS,
Tirupati, Andhra Pradesh, India-517502.

Abstract :-The National Institute of Standards and Technology (NIST) defined cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or cloud provider interaction. Cloud Computing refers to the following concepts of Grid Computing, Utility Computing, software as a service, storage in the cloud and virtualization. These are termed as a client using a provider's service remotely, known as cloud. Cloud computing has the potential to change how organizations manage information technology and transform the economics of hardware and software at the same time. Cloud computing promised to bring a new set of entrepreneurs who could start their venture with zero investment on IT infrastructure. A principal goal of this paper is to identify privacy and security issues in the distributed environment and concern to cloud computing participants and users .

Keywords: Cloud computing, Security and Privacy, Information Technology, IT, Software as a service, Grid Computing, Utility Computing, Security.

1. INTRODUCTION

The Cloud Computing is a latest concept to become popular in computer industry. The basic idea of Cloud Computing is the sharing of computing resources among a community of users. At present cloud computing emerged as a web based technology computing that provides a freedom in the establishment of IT infrastructure[1]. Cloud is basically representing internet and web based applications. It basically works on user interactive software which is as simple as web browser. The various cloud vendors do not require their own infrastructure rather they can rent or use third party providers

2. DIFFERENT TYPES OF CLOUD COMPUTING

2.1. Cloud Computing

The Cloud Computing can be termed as internet based and are connected through the remote servers. Through this sharing of data processing tasks, online access to computer resources or services and centralized data storage. The best examples are electric station, in which consumer use power without having the knowledge of infrastructure to provide the service. In the same manner, the cloud vendors use the resources as a service and pay only for resources that they use. majority cloud computing infrastructures includes services delivered through common centers and build on servers.

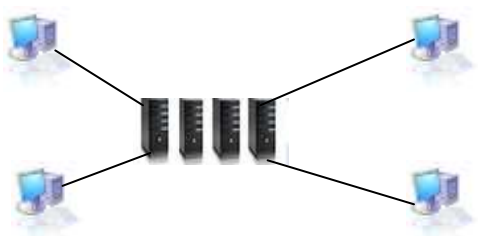


Fig.1.1. Cloud Containing Servers

2.2.. Grid Computing

Grid computing attaches computers from multiple administrative spheres to reach a common goal for solving a single task. The strategies used by Grid computing are to use middleware to divide the pieces of program among several computers. It includes computation in a distributed fashion. Grid computing is providing the resources of many computers in a network to a problem at the same time to a scientific or technical problem that needs large number of computers processing or ease to access large amount of data.

2.3..Utility Computing

Utility computing is the packaging of computation resources, such as computation, storage and service as a metered service. This model has the benefit of lesser cost to get hold of computer resources. Utility computing can be same to some extent which has the features of very large computations or a sudden height of demand which are supported by a huge number of computers. Utility computing is having some features of virtualization, so the large amount of storage or computing power is utilized at a single time sharing computers.

3. SERVICES OF CLOUD COMPUTING

Cloud computing provides both the software and hardware services through or over the internet. The services are mainly classified into three categories[3]:-

3.1. Software as a Service (SaaS)

The SaaS allows a user to use the software or application as service on demand using the Internet.

3.2. Infrastructure as a Service (IaaS):

It allows a user to use IT infrastructure such as hardware, storage and networking components as a service. The user can access the operating system, storage and application.

3.3. Platform as a Service (PaaS)

The provider provides a platform for their own use and user.

4. CLOUD COMPUTING MODEL

4.1. Private Cloud

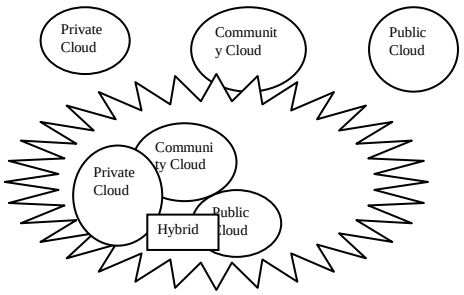
In this model, the infrastructure is used, maintained and operated for a specific company or organization.

4.2. Community Cloud

In this model, the infrastructure is shared among the various companies or organizations with similar areas of interests and requirements.

4.3. Public cloud

In this model, the infrastructure is available to the public for business purpose by various cloud service providers.



Flg.1.2. Cloud Computing Model

5. PRIVACY AND SECURITY ISSUES OF CLOUD COMPUTING

5.1. Privacy issues

1. Compelled Disclosure to the government Cloud can be subject to different levels of protection than on the information it contained
2. Data Security and Disclosure of Breaches: How does cloud provider protect customer's data how can customer ensure security compliance when storing information on the cloud?
3. Data Accessibility, Transfer and Retention: Can companies and consumers have access to data on cloud? [4]Can the data be destructed by cloud owners or should it be returned to customers?
4. Location of Data : The physical location of the server storing the data may have legal implications

5.2. Security issues

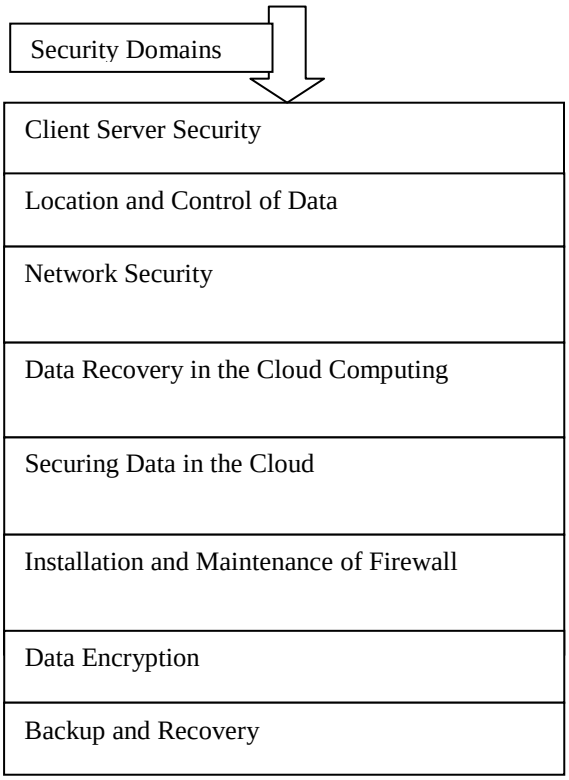


Fig.1.3. Security Domains

5.2.1. Client server security

Cloud computing encompasses a client and a server. To maintain secure client, organizations should review existing security practices and employ additional ones to ensure the security of its data. Clients must consider secure VPN to connect to the provider.

Web browsers are used in client side to access cloud computing services. Cloud providers usually provide the consumers with APIs which is used by the latter to control, monitor the cloud services. It is vital to ensure the security of these APIs to protect against both accidental and malicious attempts to evade the security. The various plug-ins and applications available in the web browsers also causes a serious threat to the client systems used to access the provider. Many of the web browsers do not allow automatic updates which will append to the security concerns. Cloud providers should also incorporate these measures to assure secure transaction among its customers

5.2.2. Location and control of data

In traditional data centers business had the privilege to know about the data flow, exact data location, precautions used to protect data from unauthorized access. The physical location' raises the question of legal governance over the data. Another impediment issue is incase of disputes arises between the provider and the customer.

Public cloud has the attraction of cost saving and low maintenance but the enticement comes with a drawback. The

infra structure has to be shared with unknown people. A cyber invader can act as a subscriber and can spread malicious viruses in the system. It is a responsibility of the provider to check the authenticity of the consumers. The vendor may grant some privileged third parties access to your stored data. The identity of such parties, if any, must be disclosed to the customer. Here, the third party could be a legal authority or even an internal employee. The customer should always be informed before the vendor allows third parties to access the stored data. Non cloud services also have security concerns but cloud has additional risk of external party involvement and exposure of critical and confidential data outside organizations control. Modifying security measures or introducing pristine Cloud provider stores the data in provider's side and maintenance is exclusively done by the providers, hence the clients have no means to check on the providers security practices, providers employees, their skills specializations etc.

5.2.3. Network security

Public cloud services are delivered over the internet, exposing the data which were previously secured in the internal firewalls. Applications which people used to access within organizations intranet are hence exposed to networking threats and internet vulnerabilities which includes distributed denial of service attacks, phishing, malwares and Trojan horses. If an attacker gains access to client credentials, they can eavesdrop on all activities and transactions, manipulate data, return falsified information, and redirect clients to illegitimate sites.

5.2.4. Data recovery in cloud computing

Usually cloud users do not know their data location and the vital query of data recovery in all circumstances may not be possible. The difficulty in retrieving data if there is a change in provider or a need to roll to different platform adds to the apprehension to embrace cloud computing.

5.2.5. Securing data in the cloud

A Proper implementation of security measures is mandatory in cloud computing. The fact that application is launched over the internet makes it susceptible for security risks. Cloud providers should think beyond the customary security practices like restricted user access, password protection etc. Physical location of stored data is also vital and it's the responsibility of the provider to choose the right location of storage.

5.2.6. Installation and maintenance of firewall

Installation of firewall and its maintenance is mandatory to ensure the protection. A firewall should be present in all external interfaces. Assessment of firewall policies and rule sets and reconfiguration of router should be done in regular intervals. Build and deploy a firewall that denies access from untrusted sources or applications, and adequately logs these events. Build and deploy a firewall that restricts access from systems that have direct external connection and those which contain confidential data or configuration data.

5.2.7. Data encryption

Data encryption is one common approach the providers to protect their clients data but the question is whether the data is getting stored in encrypted format or not. Many providers follow private/public key encryption to ensure data security. To store crucial data organizations can think of private or hybrid cloud where the data will be in secure corporate firewall.

5.2.8. Back up and recovery

In cloud computing data is stored in distributed location.. Backup software should include public cloud APIs, enabling simple backup and recovery across major cloud storage vendors, such as Amazon S3, Nirvanix Storage Delivery Network.

It is critical for the backup application to encrypt confidential data before sending it offsite to the cloud, protecting both data-in-transit over a WAN to a cloud storage vault and data-at-rest at the cloud storage site. Consumers need to verify that the cloud backup software they choose is certified and compliant with the Federal Information Processing Standards (FIPS) 140 requirements issued by the National Institute of Standards and Technology.

6. ENSURING SECURITY AGAINST THE VARIOUS TYPES OF ATTACKS

Problems associated with the network level security comprise of: DNS attacks, Sniffer attacks, issue of reused IP address, Denial of Service (DoS) and Distributed Denial of Service attacks (DDoS) etc.

6.1. DNS attacks

A Domain Name Server (DNS) server performs the translation of a domain name to an IP address. Although using DNS security measures like: Domain Name System Security Extensions (DNSSEC) reduces the effects of DNS threats but still there are cases when these security measures prove to be insufficient when the path between a sender and a receiver gets rerouted through some evil connection. It may happen that even after all the DNS security measures are taken, still the route selected between the sender and receiver cause security problems..

6.2. Sniffer attacks

A sniffer program, through the NIC (Network Interface Card) ensures that the data/traffic linked to other systems on the network also gets recorded. It can be achieved by placing the NIC in promiscuous mode and in promiscuous mode it can track all data, flowing on the same network. A malicious sniffing detection platform based on ARP (address resolution protocol) and RTT (round trip time) can be used to detect a sniffing system running on a network.

6.3. Issue of Reused IP Addresses

Each node of a network is provided an IP address. IP address is basically a finite quantity. A large number of cases related to re-used IP-address issue have been observed lately. When a particular user moves out of a network then the IP-address associated with him (earlier) is assigned to a new user. This sometimes risks the security of the new user as there is a certain time lag between the change of an IP address in DNS and the clearing of that address in DNS caches. We can say that sometimes though the old IP address is being assigned to a new user still the chances of accessing the data by some other user. It is not negligible as the address still exists in the DNS cache and the data belonging to a particular user may become accessible to some other user violating the privacy of the original user.

6.4. DBGP Prefix Hijacking

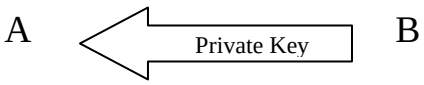
Prefix hijacking is a type of network attack in which a wrong announcement related to the IP addresses associated with an Autonomous system (AS) is made malicious parties get access to the untraceable IP addresses. On the internet, IP space is associated in blocks and remains under the control of AS's. An autonomous system can broadcast information of an IP contained in its regime to all its neighbours. These ASPs communicate using the Border Gateway Protocol (BGP) model. Sometimes, due to some error, a faulty AS may broadcast wrongly about the IPs associated with it[7]. In such case, the actual traffic gets routed to some IP other than the intended one. Hence, data is leaked or reaches to some other destination that it actually should not.

7. SECURITY AGAINST THE VARIOUS TYPES OF ATTACKS

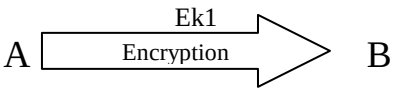
7.1. Symmetric Key Cryptography

It is equally important to secure the data in transit and security of transmitted data can be achieved through various encryption and decryption schemes. In such a scenario, even if the data gets into the hands of a hacker, he won't be able to make any unauthorized use until he knows how to decrypt it. A few of the encryption-decryption techniques include private and public key encryption. In a symmetric key (private key) encryption such as: DES, Triple DES, RC2, RC4 etc, the same key is used for encryption and decryption. Before the data is transferred, the key is shared between both the receiver and the sender. Sender then sends the data after having encrypted it using the key and the receiver decrypts it using the same key.

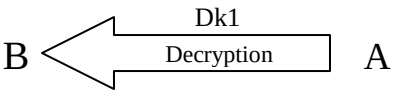
Step.1. Receiver sends its Private Key to sender



Step.2. Sender encrypts the Data using sender's Private key and sends it to Receiver



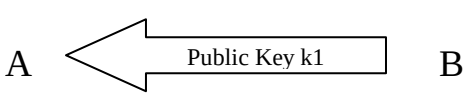
Step.3. Receiver using his Private Key and Decrypts the same data



7.2. Asymmetric Key Cryptography

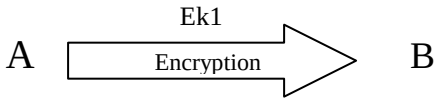
In case of Asymmetric key algorithm (RSA, DSA etc..) there are two types of keys known as Public Key and Private Key. Public key is common for both sender and receiver and the Private Key is used for decrypts the data from the sender

Step.1. Receiver sends its Public key to sender

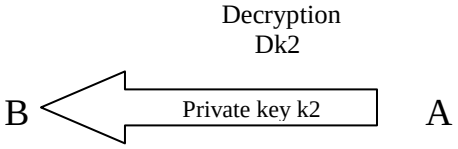


In Public key encryption bit processing time is more than private key encryption,. But the security is more concern rather than the speed, public-key encryption provides more secure data transmission in comparison to private-key encryption. Security issues in a virtualized environment wherein a malicious virtual machine tries to take control of the hypervisor and access the data belonging to other [8].

Step.2. Sender encrypts the Data using sender's Public Key and sends it to Receiver



Step.3. Receiver using his Private Key and Decrypts the same data



8. CHALLENGES OF CLOUD COMPUTING

8.1. Data Security

Security is the main area of concern. A cloud vendor watches the usage of the cloud and the data. The person who is using the cloud doesn't have the knowledge about the back-end data storage. The user doesn't have the fair idea where they are storing their data. This can be rectified if vendors can provide a good security[7] or strong firewall and if they adopt encryption facility.

8.2. Data Recovery and Availability

This challenge is faced by the vendors. The vendor should maintain a good recovery system and good maintenance management system.

8.3. Management Abilities

The management of platform and communication are in its starting phase. There is a huge requirement to improve on the scalability and load equal balancing features.

9. ADVANTAGES OF CLOUD COMPUTING

- Cost Benefits
- Flexibility
- Reliability
- Maintenance
- Mobile Accessibility

10. CONCLUSION

Cloud computing is artifact of highly advanced research done for virtualization, distributed computing with usages of software

and its related services and also networking. It completely opens a new advanced and secured world of occasions for businesses, but mixed with the offers and high level of security challenges that needs to be definitely considered when society using the advanced cloud computing concepts. We are presenting the various hidden security challenges to be precisely and closely monitor. In this paper we are also discussed the intrinsic use of virtual systems as a tool for implementing an improved and advanced cloud environment.

11. REFERENCES

[1] Anderson, D., Frivold, T. & Valdes, A (May, 1995). Next generation Intrusion Detection Expert System (NIDES):

[2] Pant Durgesh, Sharma M.K "Cloud Computing "CSICommunication-2009",Vol-32, pp10-13

[3] George Reese. Cloud Application Architectures "Building Applications and Infrastructure in the Cloud" (Theory in Practice)

[4] Esteves, R.M. and Chunming Rong (2010), Social Impact of Privacy in Cloud Computing In 2010 , IEEE Second International Conference on Cloud Computing Technology and Science Nov. 30-Dec. 3 ,2010, pp. 593-596.

[5] Mell, P. and Grance, T. (2011) The NIST Denition of Cloud Computing (Draft): Recommendations of the National Institute of Standards and Technology. Special publication 800-145 (draft), Gaithersburg (MD).

[6] In Gutwirth, S., Poullet, Y., de Hert, P., and Leenes, R.,editors, Computers, Privacy and Data Protection: an Element of Choice, Springer,pages 293–314.

[7] Lin, G., D. Fu, J. Zhu and G. Dasmalchi (2009). Cloud Computing: IT as a Service. IT

[8] Weinhardt C., A. Anandasivam, B. Blau and J. Stosser (2009), Business Models in the Service World, IT Professional, Vol. 11 No 2, pp. 28-33

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 515 - 518, 2014

Automatic Image Annotation using Image clustering in Multi – Agent Society

Abbas Akram Khorsheed
University of Mustansiriyah, College of science, Computer Science Department, Iraq

Hussein Chible
Doctoral School of science and Technology, Lebanese University, Beirut, Lebanon

Giorgio Giacinto
Cagliari Univesity, Electronic Engineering Department, Italy

Abstract: Image annotation is the natural language description for essential blobs within an Image. Many methodologies have recruited to reveal image semantics and represent it as annotation. Due to the exclusionary growth of number of images distributed over massive repositories, the task of manual annotation is tedious and over killing in term of times and efforts. This paper is presenting automatic annotation system based on the interaction between intelligent agents. Agent interaction is synonym to socialization behavior dominating Agent society. The presented system is exploiting knowledge evolution revenue due to the socialization to charge up the annotation process

Keywords: Social effect, Automatic annotation, intelligent agents

1. INTRODUCTION

Type Style and Fonts Image annotation is one of the most used methods to retrieve images from enterprise repositories by matching user text queries to these annotations. In general, annotation is represented as metadata or keywords assigned to digital images based on image contents [1][2].

More than 200 billion images are accessible online and the number is continuously growing [3] due to the numerous number of sources as digital cameras, mobile phones and other devices. This brings a great challenge in retrieving designated images which are identified by a unique number over the internet, i.e., the URI (Unique Resource Identifier), that is used to access each image over the web. Social decision theory [1] extends the theory of individual decisions to decisions made by the interaction of a group of agents.

Recent systems like Lable Me and Amazon mechanical turk distribute image annotation and evaluation tasks to Internet users. The volume of annotations generated from such crowd-sourcing techniques helps reduce the burden on experts without significantly sacrificing the quality of annotations. The annotators are provided with detailed instructions on how to best select labels that can be directly used for concept modeling. This ensures that relatively good quality annotations are generated for object detection, and relevance estimation tasks. It is shown that crowd-sourcing is a reasonable substitute for repetitive expert annotations, when there is high agreement among annotators.

Other sources of image annotations are collaborative games and social media sharing which undoubtedly represent the fastest growing labeled image collections in the world[3][4].

In this paper the multi Agent paradigm is proposed to simulate the social behavior of humans in developing knowledge regarding certain subjects. The JADE (Java Agent Development) environment has been used because it allows building multi agent platforms thanks to the utilities and wide spectrum classes provided by that environment.

2. RELATED WORKS

Many models were produced to characterize automatic image annotation frameworks. The exact specification of automatic image annotation is not yet established and researchers are investigating different approaches continually. Anyway, recent approaches can be categorized into two categories: one category is the semantic interpretation of image contents, while the other category is drawn from the epistemology field, where knowledge is revealed from the interaction among sources of knowledge. The society can be represented as a network of knowledge resources, and knowledge can be sustained or rejected upon the interaction among these resources. In this section we will focus upon previous efforts within the second category due to the orientation of this paper.

In [5] a model has been proposed to formulize the growth dynamics in social networks; in this model a great attention has been presented to the effect of node behavior, and how it affects the behavior of other nodes, and this eventually will affect the growth of the network. In term of knowledge evolution due to socialization; this model has a lot in common with our approach, though it has nothing to do with image retrieval system. The key similarities are:

1. The behavior-awareness where the interaction of node (i.e., the co-author s) with certain events (e.g., papers) is to be realized as a potential relationship among those nodes. In fact this approach develops knowledge at the network level, which helps increasing the growth factor of social network and, eventually, the productivity of such a network.

2. The clustering-coefficients where the tendency of grouping is related to the factor compose of these coefficients.

In [6] an ontological approach was presented to accomplish a computing model aimed to annotate images on two levels: Image Annotation and Annotation of Annotations; this model.

is focused on queries for annotations using the National Cancer Institute's Cancer Biomedical Informatics Grid's (caBIG) Annotation and Image Markup (AIM) project

The AIM project defines an ontology of annotations and image markup, a UML information model and provides the extensible markup language (XML) artifacts for creating them. A long-term vision of the AIM project is for large collections of annotations to be created in conjunction with the already large collections of clinical and research medical images. This will allow query of annotation, not only for the retrieval of relevant images, but also for the correlation of image observations and their characteristics with biomedical data including genomic expression.

In that paper many concepts are coherent with what we presented in our work in the area of retrieving images based on associated annotations, but this approach does not introduce autonomous annotation in any context, and it does not consider the behavior of image requesters; knowledge can't be developed to cluster images which is a crucial element in automatic image annotation strategy.

The model presented in [6] exploits annotations to build a semantic network among images, while our work provides autonomous annotation schema based on the behavioral interpretation of the user. The AIM project can be integrated with what we are presenting to provide consistent ontological environment for image retrieval and annotations. The same annotation context is presented by [7] and [4] but both depend on the retrieval and extraction of knowledge from the resources available on the global net.

In [8] a novel system is presented to exploit the format of multimedia sharing web sites in order to discover the underlying structure; this has been used to allow later, more sophisticated mining tasks for these sites to infer knowledge about certain images. Again, we have many features in common with these approaches, but still the effect of the behavioral responses of the users is absent.

In [9] a study for establishing a stable architecture for socialization is conducted and conclusion has been reached out along this study which is: in a society of agents there are three main parameters that enforce the stabilization of the architecture; these are: take on roles, play roles and locate in some society organization at all time. In our proposal, the society composed by agents is maintained stable by strict discipline through which roles are fairly distributed, and all agents are capable of playing these roles by accurate interpretation of client behavior. Furthermore, we adopt fixed organizational distribution of the agents which sustain the stability. In our proposal, the specification of the problem domain has different characterization due to the potential tendency toward clusterization on two different levels: the host level, and the network level. This approach has its roots back to [10] where a study addressed the fault assumption of regarding multi-agent systems as single learning system which is a wrong assumption due to the intuitive tendency to introduce social activity with neighbors rather than communicating with other far agents. This dual capabilities of an agent's referencing, i.e., self-referential, and social-referential, has been presented by [10] as a bi-referential model, in which each referencing capability is implemented by an evolutional computation method of classifier system.

In our referential model the evaluation function is global and updated on the fly by delivering knowledge to central a repository that holds the annotation for images. The annotations are revealed and referenced based on a confidence degree assigned to that annotation. In our referential model, the behavior of the evaluation function is dynamic due to the continuous change of confidence degree of annotation; this is due to the activities produced by the client clusterization behavior (i.e., the self-referential model).

Interactive query for images' content by semantic descriptors is an effort presented in [11]; this effort introduced a distributed content-based image query system (DCBIQ) based on the WWW. A model was proposed to integrate knowledge from image processing, semantic descriptor, multi-agent, and WWW navigation. Again in this model the image content plays the essential role in describing the image, thus low level extraction methodologies are more important than the opinion of the social communities which are using it.

In our proposal, the knowledge obtained by social interaction is more important than low level features like colors, textures or spatial relationships, and even semantic interpretation of image contents is not important as the social opinion about the image and its relation to other images or domains.

In [12] an attractive model is presented where a web-based image digital library is proposed; in this library agent system was used to traverse part of the web page looking for images that fit certain criteria. The methodology used by the agent is based on detecting URLs within web pages that refer to images, and when such URLs are encountered, then the text that is associated to that image is inferred for correlation with other features such as topic name, domain that this image falls in, or any other matching criteria. In our proposal the same ontology for allocating text accompanied the image is used as the following matched methodologies:

1-

$\forall$Image$\exists$tag$\exists$txt((presenting(image,tag) AND Asso(tag,txt)) $\longrightarrow$ Asso(image,txt) AND Select(txt))

2-

$\exists$paragraph $\exists$hyper ((hyper $\in$ paragraph) $\longrightarrow$

Select(Paragraph))

3-

$\exists$page$\exists$ time $\exists$!title((image $\in$ page) AND has(title,page) $\longrightarrow$ Associate(image,title)AND Select(title))

The main novelty of our approach is that we don't design a mining agent that is responsible on inferring web pages, but we exploit Google search APIs which are published over the web. The only web page we analyze is the results of the Google search APIs and don't investigate individual pages.

3. THE PROPOSED SYSTEM

This paper will focus on new category which is the dominant tags of the image as it is recognized by the society. Image repository (RTI) is a database holding labeled images (i.e. images tagged with annotation), thus it can be represented by eq.1

$$R_T^I = \sum_{i=1}^{k} (I_i, Anno_i) \quad \text{eq.1}$$

Where $Anno_i = \sum_1^m c_j$ collection of concepts (i.e., these concepts are keywords, tagged or labels), and let the query made by the user in order to request images is representing by the following equation:

$$Query_i = \sum_1^n q_j \quad \text{eq.2}$$

∀ indexed(⟦imgae⟧_i) ∃ c ∈ Anno(⟦image⟧_i) so that c ∉ Anno(⟦image⟧_j) for all j ≠i

SearchResult($Query_i$) = $\sum_{j=1}^{M} images_j$ iff

$Query_i \cap Anno(image_j) \neq \emptyset$

and OSF($Query_i$) $\neq c$

OR

SearchResult($Query_i$) = $image_i$ iff

OSF(Query) = c

Behavior	Weight
Saved	5
Selected, saved	4
Revisited, saved	3
Revisited	2
Highlighted	1

User interaction behavior with the resultant list of images is weighted according to above table.

Hypothesis 1: Automatic annotation member is broadcasting candidate annotation for queried images.

Prove:

Let f(query) be a mapping function that maps images from the huge repositories spread over the internet to the desired image list requested by the user, such that

$$f: Re \longrightarrow De$$

Where Re is the huge repository over the internet and De is the desire domain where resultant of f(query) satisfies client request. The input to this function is the query entered by requestor and the output is a scalar value represents confidence degree

$f(query) = 0$ Where 0: not desired and 1: desired, thus

0= De

$$f(query) = \begin{cases} v \\ < o < 1 \end{cases}$$

0 ≠ De

Where

$v = T.W_{Agent}^{image}$ Which is the total weight produced by the interaction between the requestor client and the resultant list of images. If v ≥threshold then agent will broadcast a data structure composed of the following fields (Image URI , query, f(query)).

Definition: dominant annotation is the candidate new annotation for image being queried by society of agents, where f(query) for I want here sigmoid function to be the decision function to decide that certain annotation is to be added to the image annotation list .

Hypothesis 2: social group add new annotations to image

Let :

$Query_A \cap Anno(image_j) = S_A$

and $= \sum_{k=1}^{K} weight(behavior)$

$Query_B \cap Anno(image_j) = S_B$

and $= \sum_{l=1}^{L} weight(behavior)$

⋮

$Query_n \cap Anno(image_j) = S_n$

and $T.W_{Agent\_n}^{image} = \sum_{m=1}^{M} weight(behavior)$

$T.W_{Agent_{ij}}^{image_j}$ is Total weight produced by $Agent_{ij}$ for $image_j$

Then

Added annotation set S = $S_A \cup S_B \ldots \cup S_n$ is a set of new valuable annotations to be added

to $image_j$ with a binding value $T.W_{Agent_{ij}}^{image_j}$, hence the resultant set is only a candidate

annotation, it has to be dominant to get corresponding image get indexed with.

4. SOCIAL EFFECT

Social effect over Automatic Annotation Society will be treated in this section. In social environment, members are investigating propositions based on total weight granted by trusted members of the society. In this proposal we assume that all members are trustful and other members of the society are considering their weight evenly. Let us first define a new function that describes the acceptance of the society for the candidate annotation to be a dominant annotation, and the corresponding image can be indexed with. In this paper we propose the social effectiveness function to be a sigmoid function, due to the properties of this function especially the continuity and flexibility, hence $f(query)$ is defined as

$$SocialAcc(Image) = \frac{1}{1 + e^{-Su(query)}}$$

Where $su(query)$ is a total weight gained from all agents involved in the automatic annotation system and it is represented as the following:

$$Su(query) = v_1 + v_2 + \ldots + v_n$$

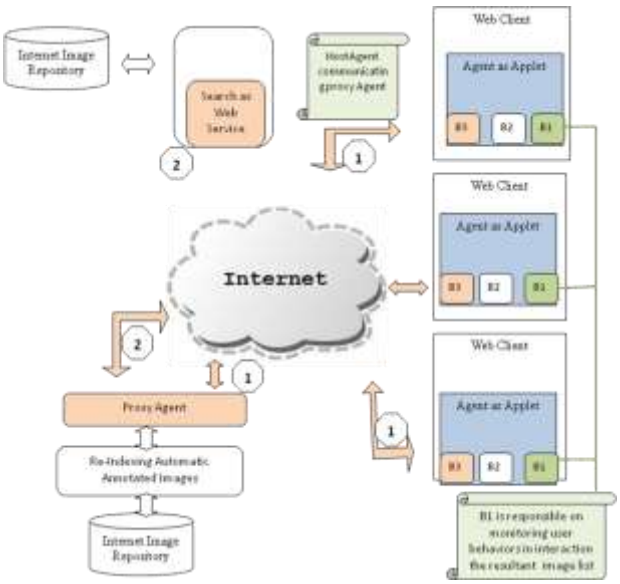


Figure 1: General scheme of social basic automatic annotation system

The role of Agent in this proposal is characterized by three behaviors, as presented in figure(2)and are briefed as the following:

Web Service Integration (B3): this behavior encapsulates web service integration functionalities by implementing SOAP based invocation to Google web service.

Socialization Behavior(B2): this behavior is responsible on socializing other agents within the platform to determine dominant tags for an image.

User behavior monitoring (B1): this module is responsible for monitoring selections made by the user after querying the Google search Web service. Highlighted images are grouped in clusters and socialization behavior is signaled. The following behaviors are considered: Highlighted images, selected images, revisited images and saved images.

Proxy Agent: this agent is responsible of initiating the communication session over the internet. Proxy Agent is a crucial element in grant multi-agent system the ability to communicate over the internet. This Agent resides at the server side.

Host Agent: this Agent is an instance constructed at the client side and monitor his/her behavior and report back to Proxy Agent his observations.

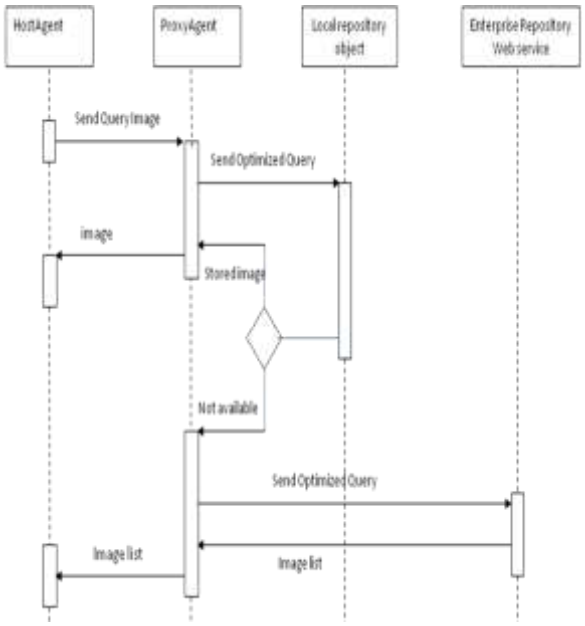


Figure 2: sequence diagram of complete 2-tier image annotation session

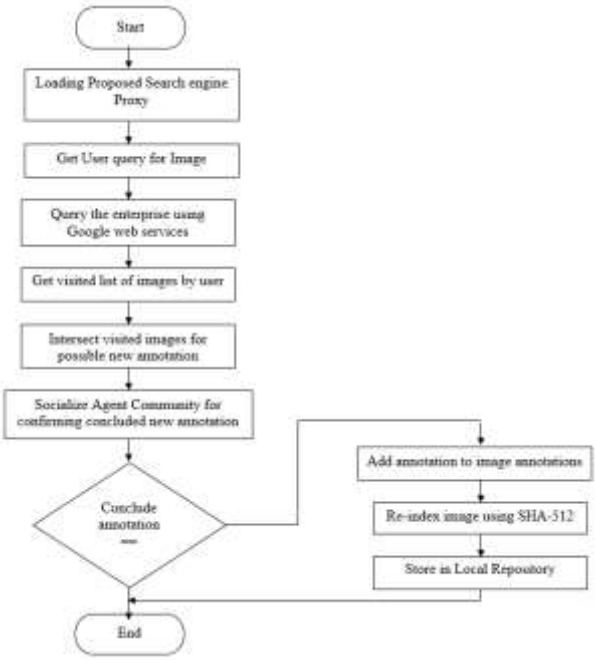


Figure 3: Automatic Image Annotation Procedure Using Multi-Agent socialization

.

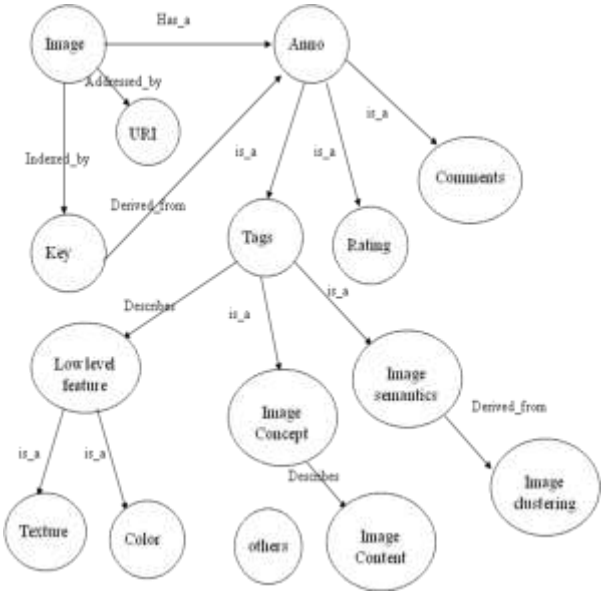


Figure 4: Ontology to conceptualize Image Annotation

5. RESULTS

Example:

Phase 1:

$client_1$ has posted the following query through the Chrome internet explorer

$$query_1 = \{car, rental, company, race\}$$

After posting that query, 37,212 images have been listed in the internet explorer.

$client_1$ has selected and saved the following image

The session manager agent, which has HTTP listener, captured the URL or URI corresponding that image.

The following is the URL

Imgurl:http://www.koopman-racing.nl/images/sd2_1559.jpg

This image will be indexed using $query_1$.

Phase 2:

$client_2$ and $client_3$ have posted queries as the following

$$query_2 = \{aspiring, driver, koopman, race\}$$

with max weight (5)

$$query_3 = \{race, rental, koopman, car\}$$

with weight (4)

$$query_1 \cup query_2 = \{aspiring, driver, kooman\}$$

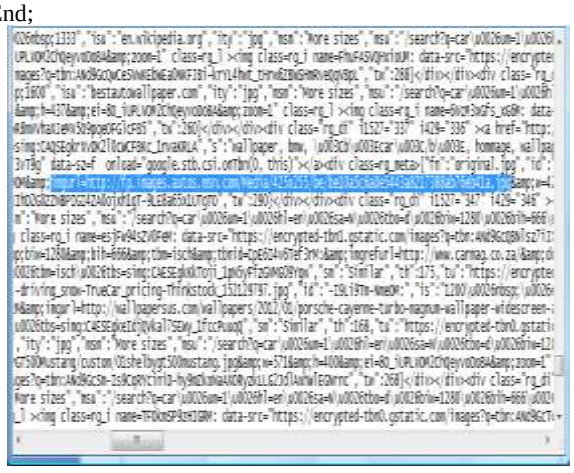
these annotations are to be more convenient to be used as indexing due to its weight factor and by socializing it to other clients like $client_3$ the highest effective annotations will be $\{aspiring, driver\}$, This is for the same URL.

https://secure.booking.com/confirmation.fr.htm
l?aid=350433;label=edr-xmlvswl-fr-
users;sid=05686c51355c9e5ba1e2a8d843e2c461;dcid=2;bn=6
08419725;hostname=hotels.edreams.fr;pincode=6604#print

Algorithm1:
Automate Image Search Using Google

- Procedure: Search Web
- Input: query As String
- Output: array of imgurls
- Begin
- Initialize user Query = query;
- Initialize GoSearchConnection as URLConnection to Google URL + user Query;
- Set GoSearchConnection Properties as
- Method = 'GET';
- Char-set = 'utf-8';
- User-Agent = 'Mozila-4.0';
- GoSearchConnection. Open;
- Get input Stream from GoSearchConnection to stream Reader;
- while stream Reader has imgurl do
- add current imgurl to imgurl_list;
- return imgurl_list;

End;



Algorithm 2:

Reveal Local Knowledge

- Input: Selected imageUrl_list
- Output: weighted imageUrl_list
- Begin
- For each image in image Url list Do
- begin
- Capture mouse and keyboard events
- Assign weight to image
- end
- End.

Algorithm 3

: Broadcast local knowledge

- Input: weighted imageUrl_list
- Begin
- Instantiate msg from ACLMessage;
- Set msg.receiver to be address of global Agent
- Set msg.content to be weighted imageUrl_list
- and the Query;
- Send msg;

End

Algorithm 4:

Intersect broadcast knowledge

- Input: msges[] as Array of Agents' Messages
- Begin
- Initial CommonVisited List as String Array
- Initial SumWeight as integer Array
- For all messages in msges[]
- tempMsg = nextMsg in msges[]
- for all messages in msges[] and NOT tempMsg do
- find shared imageUrl and add it to
- CommonVisitedList.

Sum total weight and add it to SumWeight in index manner.

- End.

After intersecting queries from different Agents, the following URI

- http://wallpapersget.com/wallpapers/2012/03/car-bmw-328-hommage-wallpaper-1080x1920.jpg

will be indexed using key {race, fancy, celebrity and sport .

6. CONCLUSIONS

After investigating a bunch of papers published within the same topic of our proposal, we found correlation in basic terminologies, but with distinct methodologies. Many models have been introduced to develop knowledge about retrieved images like what we introduced here and the significant features of Agent-based system are also exploited but the key differences between all these efforts and what we devised in our proposal can be summarized:

1-A Multi-Agent system has been deployed on two levels: host level and network level to develop knowledge regarding certain images, other approaches target mainly behavioral aspects of network interactions rather than host based.

2-Annotation is generated autonomously and a confidence value is assigned to each annotation; this value represents the acceptance of society for this annotation as a key index for associated image.

3-Third party web based tools has been included (i.e., the Google search engine APIs) while all other approaches tend to design custom search software modules. Google has a very massive repository of images, thus it is more convenient to address this repository rather than inferring other repositories or web pages. Furthermore, Google search engine receives millions of request for images in multiple subjects, thus this will assist, statistically, revealing more reliable annotations.

4-An image is annotated, in our proposal, not on the basis of the graphical objects in the image or the low level features, but on the basis of its relation to the environment, for example an image could have some planets and this image can be interpreted using low level features and semantic contents as to relate to planet science, flowers, garden or some of the like.

In our approach the planets image can be categorized into drug, medicine, health or so on; this is due to society opinion.

Other approaches Index this image based on its low level feature and its composed visual objects.

7. ACKNOWLEDGMENTS

My thanks to the University of Mustansiriyah – Iraq, that gives me the possibility to support and continue my PhD.

8. REFERENCES

[1] NursuriatiJamil and SitiAisyahSa'adan, Proceeding, "Visual Informatics: Bridging Research and Practice", visual informatics conference, IVIC 2009, Kuala Lumpur, Malaysia, 2009.

[2] Masashi Inoue, "On the Need for Annotation-Based Image Retrieval", National Institute of Informatics, Tokyo, Japan, 2006

[3] Aixin Sun, Sourav S. Bhowmick, Khanh Tran Nam Nguyen, "Tag-Based Social Image Retrieval: An Empirical Evaluation", published in American Society for Information Science and Technology, USA, 2011.

[4] NeelaSawant, Jia Li and James Z. Wang, "Automatic Image Semantic Interpretation Using Social Action and Tagging Data", survey, Pennsylvania State University, USA,2011.

[5] Baojun Qiu, Kristinka Ivanova, John Yen, and Peng Liu, "Behavior Evolution and Event-driven Growth Dynamics in Social Networks"

[6] David S. Channin MD, Pattanasak Mongkolwat, Vladimir Kleper, and Daniel L. Rubin, "Computing Human Image Annotation"

[7] Chen-Ue Lee, Von-Wun Soo, and Yi-Ting Fu, "How to Annotate an Image? The Need of an Image Annotation Guide Agent"

[8] Nikolaos Papadakis, Klimis Ntalianis, Anastasios Doulamis, and George Stamoulis, "An Automatic Multi-Agent Web Image and Associated Keywords Retrieval System"

[9] Jian Zhang, Guang-Zhou zeng, and Zhi-Feng Li, " The Study of a Sociality Agent Architecture Based On Role"

[10] Takayuki Shiose, Tetsuo Sawaragi, Osamu Katai, and Michio Okada, "Dynamics of Reciprocal Learning by Bi-Referential Model within Multiagent Systems"

[11] Weng Zu Mao, and David A. Bell," Interactive Query for Image Contents by Semantic Descriptors and Multi-Agent".

[12] Jesus Favela and Victoria Meza,"Image-retrieval agent: integrating image content and text".

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 519 - 528, 2014

Data mining Innovativeness of data give-and-take service station in Teradata client-server construction

K. Hepsiba
Department of CSE
Gokula Krishna College of Engineering,
Sullurpet, Nellore(Dt)
Andhra Pradesh, India.

Y. Madhusekhar
Department of CSE
Gokula Krishna College of Engineering,
Sullurpet, Nellore(Dt)
Andhra Pradesh, India.

Abstract: Teradata is a relational database management system that drives a company's data warehouse. Teradata provide the foundation to give a company the power to grow, to complete in today's dynamic marketplace, to achieve the goal of "transforming transactions into Relationships" and to evolve the business by getting answer to a new generation of questions. Teradata's scalability allows the system to grow as the business grows, from gigabytes to terabytes and beyond. Teradata's unique technology has been proven at customer sites across industries and around the world. Teradata is a large database server that accommodates multiple client applications making inquiries against it concurrently. Various client platforms access the database through a TCP-IP connection across an IBM mainframe channel connection. The ability to manage large amounts of data is accomplished using the concept of parallelism, where in many individual processors perform smaller tasks concurrently to accomplish an operation against a huge repository of data.

Keywords RDBMS, Data warehouse, Transformation, Scalability, Parallelism, shared-nothing, server-client architecture, leaner expansion.

1. INTRODUCTION

Teradata is a relational database management system which is especially designed for running very large commercial databases. Teradata uses the parallelism to manage terabytes of data. Teradata is a shared nothing architecture. Can start with teradata as small as gigabytes and grow large as volume of data increase. Teradata supports UNIX and Windows operating system. Teradata supports ANSI standard SQL. Teradata act as a database server for many client applications. Teradata supports the Network and mainframe connectivity. Fault tolerance at all levels of hardware and software. It has the data integrity and reliability.

1.1 Brief History

In 1979 Teradata corporation founded in Los Angeles, California Development begins on a massively parallel database computer.

In 1984 Teradata sells the first database computer DBC/1012 to wells Fargo Bank of California.

In 1989 Teradata and NCR partner on next generation of DBC.

In 1990 First Terabyte system installed and in production.

In 1991 NCR is acquired by AT&T.

In 1992 Teradata is merged into AT&T/NCR.

In 1995 Teradata version 2 for UNIX operating systems released.

1.2 Teradata

Teradata is a large database server that accommodates multiple client applications making inquiries against it concurrently. Various client platforms access the database through a TCP-IP connection across an IBM mainframe channel connection. The ability to manage large amounts of data is accomplished using the concept of parallelism, wherein many individual processors perform smaller tasks concurrently to accomplish an operation against a huge repository of data. To date, only parallel architectures can handle databases of this size.

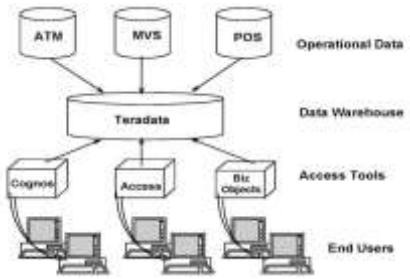


Figure:1.2 Tera data overview

• Designed to process large quantities of detail data.
• Ideal for data warehouse applications.
• Parallelism makes easy access to very large tables possible.
• Open architecture (system) – uses industry standard components.
• Performance increase is linear as components (nodes) are added.
• Runs as a database server to client applications.
• Runs on multiple hardware platforms.

2. TERA DATA MANAGEABILITY

One of the key benefits of Teradata is its manageability. The list of tasks that Teradata Database Administrators do not have to do is long, and illustrates why the Teradata system is so easy to manage and maintain compared to other databases. Things Teradata Database Administrators Never Have to Do Teradata DBAs never have to do the following tasks:

- Reorganize data or index space.
- Pre-allocate table/index space and format partitioning. While it is possible to have
- Partitioned indexes in Teradata, they are not required.
- Pre-prepare data for loading (convert, sort, split, etc.).
- Unload/reload data spaces due to expansion.

With Teradata, the data can be redistributed on the larger configuration with no offloading and reloading required. Write or run programs to split input source files into partitions for loading.

With Teradata, the workload for creating a table of 100 rows is the same as creating a table with 1,000,000,000 rows. Teradata DBAs know that if data doubles, the system can expand easily to accommodate it. Teradata provides huge cost advantages, especially when it comes to staffing Database Administrators. Customers tell us that their DBA staff requirements for administering non-Teradata databases are three to 10 times higher. How Other Databases Store Rows and Manage Data Even data distribution is not easy for most databases to do. Many databases use range distribution, which creates intensive maintenance tasks for the DBA. Others may use indexes as a way to select a small amount of data to return the answer to a query. They use them to avoid accessing the underlying tables if possible. The assumption is that the index will be smaller than the tables so they will take less time to read. Because they scan indexes and use only part of the data in the index to search for answers to a query, they can carry extra data in the indexes, duplicating data in the tables. This way they do not have to read the table at all in some cases. As you will see, this is not nearly as efficient as Teradata method of data storage and access.

Other DBAs have to ask themselves questions like:

- How should I partition the data?
- How large should I make the partitions?
- Where do I have data contention?
- How are the users accessing the data?

Many other databases require the DBAs to manually partition the data. They might place an entire table in a single partition. The disadvantage of this approach is it creates a bottleneck for all queries against that data. It is not the most efficient way to either store or access data rows. With other databases, adding, updating and deleting data affects manual data distribution schemes thereby reducing query performance and requiring reorganization. A Teradata system provides high performance because it distributes the data evenly across the AMPs for parallel processing. No partitioning or data re-organizations are needed. With Teradata, your DBA can spend more time with users developing strategic applications to beat your competition.

2.1 Scalability

"Linear scalability" means that as you add components to the system, the performance increase is linear. Adding components allows the system to accommodate increased workload without decreased throughput. Teradata was the first commercial database system to scale to and support a trillion bytes of data. The origin of the name Teradata is "tera-," which is derived from Greek and means "trillion".
The chart below lists the meaning of the prefixes: 10^3

Table 1:

Prefix	Exponent	Meaning
Kilo	10^3	1,000(thousand)

Mega	10^6	1,000,000(million)
Giga	10^9	1,000,000,000(billion)
Tera	10^{12}	1,000,000,000,000(trillion)
Peta	10^{15}	1,000,000,000,000,000(quadrillion)
Exa	10^{18}	1,000,000,000,000,000,000(quintillion)

Teradata's scalability provides investment protection for customer's growth and application development. Teradata is the only database that is truly scalable, and this extends to data loading with the use of parallel loading utilities. Teradata is scalable in multiple ways, including hardware, complexity, and concurrent users.

Hardware

Growth is a fundamental goal of business. A Teradata MPP system easily accommodates that growth whenever it happens. The Teradata Database runs on highly optimized NCR servers in the following configurations:

SMP - Symmetric multiprocessing platforms manage gigabytes of data to support an entry-level data warehousing system.

MPP - Massively parallel processing systems can manage hundreds of terabytes of data. You can start small with a couple of nodes, and later expand the system as your business grows. With Teradata, you can increase the size of your system without replacing:

Databases - When you expand your system, the data is automatically redistributed through the reconfiguration process, without manual interventions such as sorting, unloading and reloading, or partitioning.

Platforms - Teradata's modular structure allows you to add components to your existing system.

Data model - The physical and logical data models remain the same regardless of data volume.

Applications

Applications you develop for Teradata configurations will continue to work as the system grows, protecting your investment in application development.

Complexity

Teradata is adept at complex data models that satisfy the information needs throughout an enterprise. Teradata efficiently processes increasingly sophisticated business questions as users realize the value of the answers they are getting. It has the ability to perform large aggregations during query run time and can perform up to 64 joins in a single query.

Concurrent Users

As is proven in every benchmark Teradata performs, Teradata can handle the most concurrent users, who are often running multiple, complex queries. Teradata has the proven ability to handle from hundreds to thousands of users on the system simultaneously. Adding many concurrent users typically reduces system performance. However, adding more components can enable the system to accommodate the new users with equal or even better performance.

2.2 Unconditional Parallelism

Teradata provides exceptional performance using parallelism to achieve a single answer faster than a non- parallel system. Parallelism uses multiple processors working together to accomplish a task quickly. An example of parallelism can be seen at an amusement park, as guests stand in line for an attraction such as a roller coaster. As the line approaches the boarding platform, it typically will split into multiple, parallel lines. That way, groups of people can step into their seats simultaneously. The line moves faster than if the guests step onto the attraction one at a time. At the biggest amusement parks, the parallel loading of the rides becomes essential to their successful operation. Parallelism is evident throughout a Teradata system, from the architecture to data loading to complex request processing. Teradata processes requests in parallel without mandatory query tuning. Teradata's parallelism does not depend on limited data quantity, column range constraints, or specialized data models -- Teradata has "unconditional parallelism".

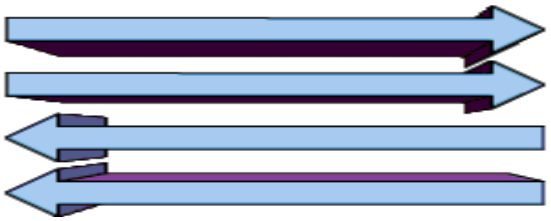


Figure2.2: unconditional parallelism

2.3 Ability To Model The Business

A data warehouse built on a business model contains information from across the enterprise. Individual departments can use their own assumptions and views of the data for analysis, yet these varying perspectives have a common basis for a "single version of the truth".With

Teradata's centrally located. logical architecture, companies can get a cohesive view of their operations across functional areas to:

- Find out which divisions share customers.
- Track products throughout the supply chain, from initial manufacture, to inventory, to sale, to delivery, to maintenance, to customer satisfaction. Analyze relationships between results of different departments.
- Determine if a customer on the phone has used the company's website.
- Vary levels of service based on a customer's profitability.

You get consistent answers from the different viewpoints above using a single business model, not functional models for different departments. In a functional model, data is organized according to what is done with it. But what happens if users later want to do some analysis that has never been done before? When a system is optimized for one department's function, the other departments' needs (and future needs) may not be met.

A Teradata system allows the data to represent a business model, with data organized according to what it represents, not how it is accessed, so it is easy to understand. The data model should be designed with regard to usage and be the same regardless of data volume. With Teradata as the enterprise data warehouse, users can ask new questions of the data that were never anticipated, throughout the business cycle and even through changes in the business environment. A key Teradata strength is its ability to model the customer's business.

Teradata's business models are truly normalized, avoiding the costly star schema and snowflake implementations that many other database vendors use. Teradata can do Star Schema and other types of relational modeling, but Third Normal Form is the methodology Teradata recommends to customers. Teradata's competitors typically implement Star Schema or Snowflake models either because they are implementing a set of known queries in a transaction processing environment, or because their architecture limits them to that type of model. Normalization is the process of reducing a complex data structure into a simple, stable one. Generally this process involves removing redundant attributes, keys, and relationships from the conceptual data model. Teradata supports normalized logical models because Teradata is able to perform 64 table joins and large aggregations during queries.

Figure 2.3 : Ability to model the business

3. TERADATA COMPONENTS

1) Parsing Engine 2) BYNET 3) AMP 4) VDISKS

3.1 Parsing Engine

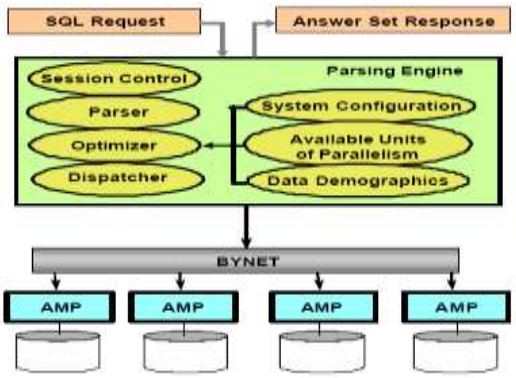


Fig 3.1. P.E Architecture

The Parsing Engine is responsible for:

• Managing individual sessions (up to 120 sessions per PE)

• Parsing and optimizing your SQL requests

• Building query plans with the parallel-aware, intelligent Optimizer

• Dispatching the optimized plan to the AMPs

• Sending the answer set response back to their questing client.

3.2 BYNET (Banyan Network) Architecture

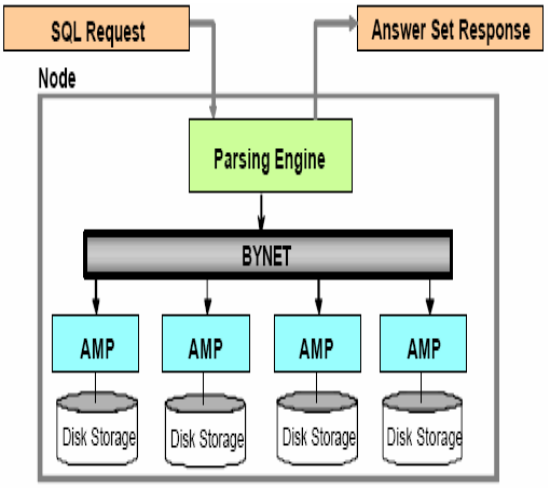


Fig.3.2 BYNET Architecture

• Automatic load balancing of message traffic.

• Automatic reconfiguration after fault detection.

The Bynet Connects All The Amps On the System

• Between nodes, the BYNET hardware carries broadcast and point-to-point Communications. (1 AMP….1AMP)

• On a node, BYNET software and PDE together control which AMPs receive a multicast communication (1….more, 1….many.)(1….more AMPs)

Bynet Features

- Enables multiple SMP nodes (MPP) to communicate.
- Automatic load balancing of message traffic.
- Automatic reconfiguration after fault detection.
- Fully operational dual BYNETs provide fault tolerance.
- Scalable bandwidth as nodes is added.

3.3 Access Module Processor (AMP)

Amp is called as Heart of teradata and every AMP will consist of its own virtual disk (VDISK). It retrieves data and updates the data on its own virtual disks.

AMPs are responsible for:

• Storing and retrieving rows to and from disks

• Lock management (lock/unlock)

• Sorting rows and aggregating columns

• Join processing

• Output conversion and formatting (ANSI, ASCII, EBCDIC)

• Creating answer sets for clients

• Disk space management and Accounting

• Recovery processing (ARC, LOCKS, JOURNAL, FALLBACK...)

4. FAULT TOLERANCES

Fallback:

A fallback table is a duplicate copy of a primary table. Each row in a fallback table is stored on an AMP different from the one to which the primary row hashes. This reduces the likelihood of loss of data due to simultaneous losses of the 2 AMPs or their associated disk storage.

AMP Clusters:

Clustering is a means of logically grouping AMPs to minimize (or eliminate) data loss that might occur from losing an AMP. Note that AMP clusters are used only for fallback data.

4.1 Cliques

The clique is a feature of multimode systems that physically groups nodes together by multiport access to common disk array units. A clique is the mechanism that supports the migration of vprocs under PDE following a node failure. If a node in a clique fails, then AMP and PE vprocs migrate to other nodes in the clique and continue to operate while recovery occurs on their home node. PEs for channel-attached hardware cannot migrate because they are dependent on the hardware that is physically attached to the node to which they are assigned. PEs for LAN-attached connections do migrate when a node failure occurs, as do all AMPs.

4.2 Hot Standby Nodes

The Hot Standby Node feature allows spare nodes to be incorporated into the production environment so that the Teradata Database can take advantage of the presence of the spare nodes to improve availability. A hot standby node is a node that:

- Is a member of a clique
- Does not normally participate in the production
- Can be brought into the production to compensate for the loss of a node in the clique

Configuring a hot standby node can eliminate the system-wide performance degradation associated with the loss of a single node in a single clique. When a node fails, the Hot Standby

Node feature migrates all AMP and PE vprocs on the failed node to other nodes in the system, including the node that you have designated as the hot standby. The hot standby node becomes a production node. When the failed node returns to service, it becomes the new hot standby node.

5. RAID PROTECTION

RAID 1 (Mirroring)

- Each physical disk in the array has an exact copy in the same array.
- The array controller can read from either disk and write to both.
- When one disk of the pair fails, there is no change in performance.
- Mirroring reduces available disk space by 50%.
- Array controller reconstructs failed disks quickly.
- good performance with disk failures
- higher cost in terms of disk space

RAID 5 (Parity)

- Data and parity are striped across a rank of 4 disks.
- If a disk fails, any missing block may be reconstructed using the other three disks.
- Parity reduces available disk space by 25% in a 4-disk rank.
- Array controller reconstruction of failed disks takes lesser than RAID1.
- Reduced performance with disk failure
- Lower cost in terms of disk space

6. TEMPORARY TABLES

Global Temporary Tables:

Global temporary tables are tables that exist only for the duration of the SQL session in which they are used. The contents of these tables are private to the session, and the system automatically drops the table at the end of that session. However, the system saves the global temporary table definition permanently in the data dictionary. In addition, global temporary tables allow the database administrator to define a template in the schema, which a user can reference for their exclusive use during a session.

Volatile Temporary Tables:

A volatile temporary table resides in memory but does not survive across a system restart. If a user needs a temporary table for a single use only, they should define a volatile temporary table. Using volatile temporary tables improves performance even more than using global temporary tables, since the system does not store the definitions of volatile temporary tables in the data dictionary. Moreover, users require no privilege to access volatile temporary tables.

The maximum number of temporary tables is provided in the following table:

global temporary tables (materialized) 2000 per session.
volatile tables 1000 per session.

6.1 Primary Keys And Primary Indexes

Indexes are conceptually different from keys:
- A *PK* is a relational modeling convention which uniquely identifies each row.
- A *PI* is a Teradata convention which determines how the rows are stored and accessed.

Primary Key	Primary Index
Logical concept of data modeling	Physical mechanism for access and storage
Teradata doesn't need to recognize	Each table must have exactly one
No limit on column numbers	64-column limit
Documented in data model (Optional in CREATE TABLE)	Defined in CREATE TABLE statement
Must be unique	May be unique or non-unique
Uniquely identifies each row	Used to place and locate each row on an AMP
Values should not change	Values may be changed (Del+ Ins)
May not be NULL—requires a value	May be NULL
Does not imply an access path	Defines most efficient access path
Chosen for logical correctness	Chosen for physical performance

- A significant percentage of tables may use the same columns for both the PK and the PI.
- A well-designed database will use a PI that is different from the PK for some tables.

Fig6.1. Difference between pk and Primary indiex

7. TERADATA WAREHOUSE MINER

Fig 7. Warehouse miner main screen

There are three windows on the main screen, the largest of which is for viewing and editing analysis forms. On the right

is the Project explorer window where open projects and the analyses they contain are displayed in tree view. Underneath both of these areas is the execution status window. Directly over the analysis work area is a toolbar with icons for primary functions and over that is a series of menus topics, including file, view, project, tools, window and help. In the sample screen above, the open connection icon has been selected to connect to data source DBC twm, and the add new analysis icon has been selected to select Data Explorer from the Descriptive Statistics category. Now looking at the data Explorer input from covering most of the main screen, selectors can be seen on the left side of the form for selecting databases, tables and columns, and on the right area to drag selected columns into.(the arrow buttons in the middle can also be used to select and de-select columns.)

Over the selectors are tabs for INPUT, OUTPUT and RESULTS, with sub-tabs that depend on the type of analysis. After the parameters for an analysis have been specified, the analysis can be executed by clicking the run button above, by right clicking on the project or analysis in the project work area and selecting run, or by pressing the F5 key on the keyboard. The status of the execution will be displayed in the execution status window below. When execution is complete, the results tab will be enabled, and upon selection, the resulting data, graphs and generated SQL can be viewed.

7.1 Exploring Data with a Data Explorer Analysis

Parameterized a Data Explorer analysis as follows

- Input source: MultiTable
- Available Databases: the databases where the demonstration data was installed.
- Available Tables:
  - TWM_CHECKING_ACCT
  - TWM_CREDIT_ACCT
  - TWM_CUSTOMER
  - TWM_SAVINGS_ACCT
- Analyses to Perform
  - Values: Enabled
  - Compute unique values: Enabled
  - Statistics: Enabled
  - Frequency :Enabled
  - Histogram : Enabled

Output Values analyses output table : twm_values

Statistics analyses output table : twm_stats

Frequency analyses output table : twm_freq

Histogram analyses output table : twm_hist

Run the analysis, and when it completes, click on the results tab.

Data

By clicking on data and then load, each of the four tables produced can be viewed by selecting the desired table in the pull-down selector.

Fig 7.1 Data Explorer

Graph

The following is a snapshot of the icon displayed when the graph tab is selected.

Fig 7.2 : Graph menu

By clicking anywhere in this picture the subsequent display of the actual graph object is displayed.

Fig 7.3: Graph

Clicking on the city_name thumbnail graph leads to the following display, while clicking on the bar for san Diego adds the drill down box to the displayed. By clicking on the drill down button the customers in san Diego can be displayed.

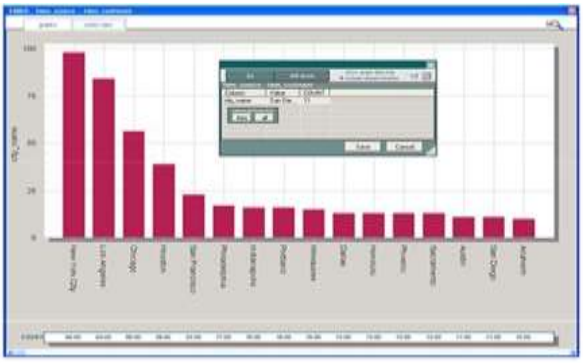


Fig 7.4.city name thumbnail graph

Creating an Analytic Data Set the following depicts an example of creating an analytic data set using the variable creation analysis. Following this depiction are step-by-step instruction for defining the variables creating in this example

Fig 7.5: variable creation

Parameterize the above variable creation analysis as follows

1. Select TWM_CUSTOMER as the available Table.

2. Create seven variables by double-clicking on the following columns.

TWM_CUSTOMER.cust_id
TWM_CUSTOMER.inncome
TWM_CUSTOMER.age
WM_CUSTOMER.years_with_blank
TWM_CUSTOMER.nbr_children
TWM_CUSTOMER.gender
TWM_CUSTOMER.marital_status

3. Select TWM_CREDIT_TRAIN as the Available Table.

4. Create a variable by clicking on the new button and build up an expression as follows.

5. Drag an Add SQL Element over the variable, and then drag the following two columns over the empty arguments.

TWM_CREDIT_TRAN.insert_amt
TWM_CREDIT_TRAN.principal_amt

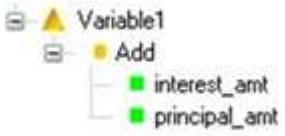


Fig 7.6 : Add(arthimetic)

Because there may be negative values, drag and drop an Absolute value (Arithmetic) SQL Element over both interest_amt and principal_amt.

6. Take the average of this expression, by dragging and dropping an average (Aggregation) on top of the Add.

7. Because this analysis may generate many NULL values by joining TWM_CUSTOMER to TWM_CREDIT_TRAN, drag a coalesce (case) on top of the Average.

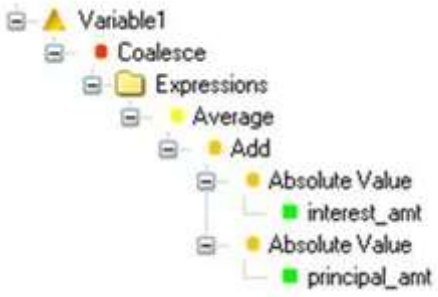


Fig 7.7 : Coalesce (case)

8. Drag and drop a number (Literal) 0 into the expression folder and rename it from variable to avg_cc_tran_amt to complete the variable.

9. Goto INPUT anchor Table and select TWM_CUSTOMER as the anchor table as seen below

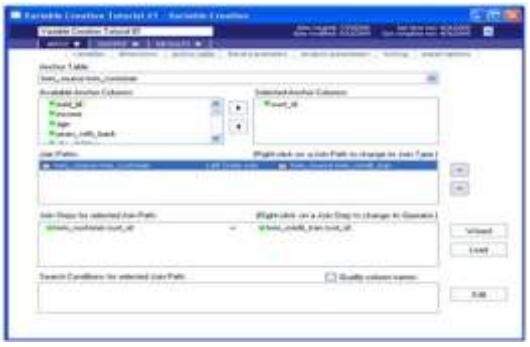


Fig 7.8: INPUT>Anchor Table : select TWM_CUSTOMER

10. Specify the join path from TWM_CUSTOMER to TWM_CREDIT_TRAIN by clicking on the Wizard button and specifying that they be joined on the column “cust_id”.

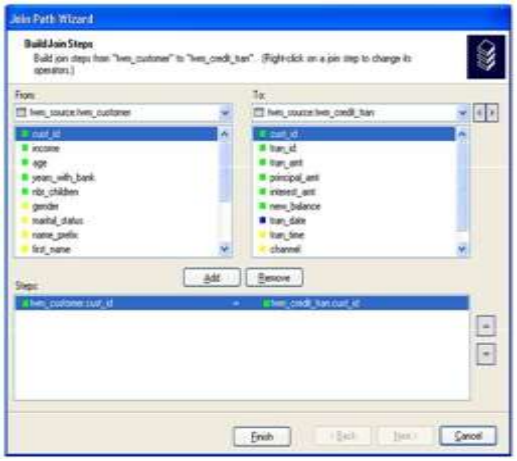


Fig 7.9 : Join Path Wizard

11. Go to OUTPUT storage, and select Store the tabular output of this analysis in the database.

Specify that a table should be created named twm_tutorials_vcl.

Creating and Scoring a Decision Tree model

Building a Decision Tree Model

The following depicts a tutorial example of creating a decision tree model. In this example a standard Gain Ratio tree was built to predict card ownership, based on 20 numeric and categorical input variables. Notice theat the tree initially built contained 100 nodes but was pruned back to only 11, couning the root node. This yielded not only a relatively simple tree structure, but also Model Accuracy of 95.72% on this training data.

Parameterize aDecision Tree as follows.

- Available Tables : twm_customer_analysis
- Dependent variable : ccacct
- Independent variables:
  - Income,age
  - Years_with_bank, nbr_children
  - Gender, marital_status
  - City_name, state_code
  - Female, single
  - Married, separated
  - Avg_ck_bal,avg_ck_tran_cnt
  - Avg_sv_tran_amt,avg_sv_tran_cnt
- Tree splitting : Gain Ratio
- Minimum Split count : 2
- Maximum Nodes :1000
- Maximum Depth :10
- Bin numeric variables :Disabled
- Pruning Method : Gain Ratio
- Include Lift Table : Enabled
- Response value : 1

Run the analysis and click onresults when it completes. For this example, the decision tree analysis generate the following pages.

Decision Tree Report

Table 1: Decision tree Report

Total observations	747
Nodes before pruning	33
Nodes after pruning	11
Model accuracy	95.72 %

Variables

Table 2: dependent variables

Dependent variables
ccacct

Table 3:Independent variables

Independent variable
Income
Ckacct
Avg_sv_bal
Avg_sv_tran_cnt

Confusion matrix

Table 4: confusion matrix

	Actual non-response	Actual response	Correct	Incorrect
Predict 0	340 / 45.52 %	0 / 0.00%	340 / 45.52%	0 / 0.00%
Predict 1	32 / 4.28%	375 / 50.20%	375 /50.20%	32 /4.28%

Cumulative lift table

Table 5: cumulative lift table

Graphs

By default the tree browser is displayed as follows.

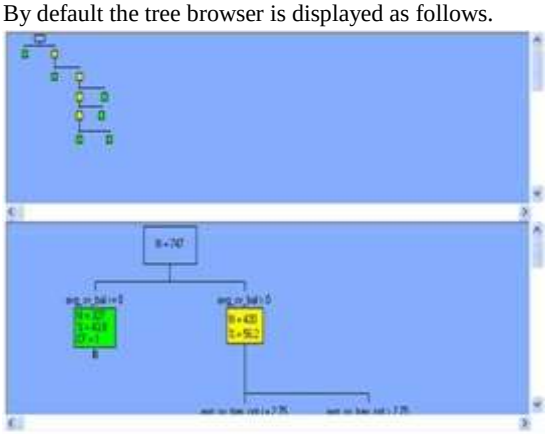


Fig 8: Tree Browser

Select the text tree tab to view the rules in textual format.

Fig 9 : Text tree tab

Additionally, you can click on lift chart to view the lift table graphically.

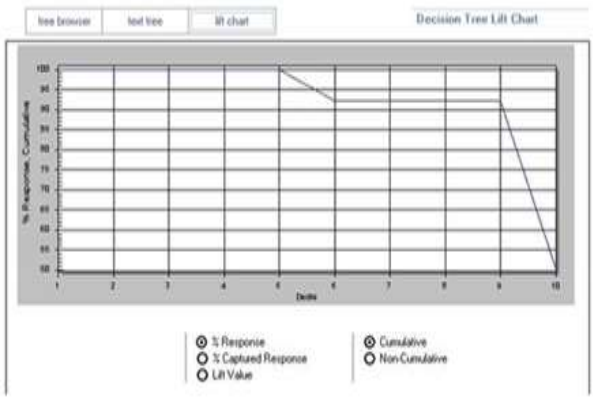


Fig 10: lift chart tab

8. CONCLUSION

Teradata is the forthcoming of the Data Mining. In upcoming the whole world we start using Teradata Database. Now it is expensive, works are successful on to reduce its cost. So, it will reach to small business people also. It is firm occupied environment will change organization goals.

9. ACKNOWLEDGMENTS

Our thanks to K Vasanth Kumar, Assistant Professor in NBKR Institute of Science and Technology, Vidyanagar, P RAJESH KUMAR, Assistant Professor in Siddhartha institute of Engineering and Technology College, Puttur, Andhra Pradesh for his guidance in regards of this paper and T GOPINATH Pursing Master of Computer Applications in JNTU, ANTHAPUR for his support in completing this paper.

10. REFERENCES

[1] https://www.teradata.com/.../TeradataData-Mining-Services-eb1719/
[2] http://www.teradatatech.com/?p=103
[3] http://www.teradata.com/businessneeds/data-mining-and- analytics/
[4] http://www.teradata.com/products-andservices/teradata-warehouseminer/?ICID=Ptwm
[5] http://decisionfirst.files.wordpress.com/2013/09/sap-acquires-kxen.pdf
[6] www.teradata.com/brochures/TeradataRapid-Insight-Service-eb6161

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 529 - 532, 2014

Resource Allocation in Computational Grids environment Using Improved Particle Swarm Optimization Algorithm

Mehdi Effatparvar
ECE Department, Ardabil Branch, Islamic Azad University, Ardabil, Iran

Shojaat Hoseinpour
Department of Computer Science and Research Branch
Islamic Azad University
Germi, Iran

Vahid Asadzadeh
Department of Computer Science and Research Branch
Islamic Azad University
Germi, Iran

Abstract: Resource allocation in computational grids is considered as a NP-Complete problem due to resources heterogeneity. Grid resources are related to various management areas exerting different management policies. Nowadays, enhancing grid efficiency is regarded as a problem requiring proper and effective schedule. Unfortunately, grid resources dynamic nature, in addition to the variety of users' requests has intensified grid resource allocation. The present paper offers a new heuristic method based on Particle Swarm Optimization (PSO) algorithm for resource allocation in grid environment. The proposed method creates an optimal scheduling in task completion with minimum flowtime and makespane.

Keywords: scheduling, computational grid, heuristic algorithm, resource allocation, optimization algorithm, particle swarm

1. INTRODUCTION

Grid computation make feasible sharing and highly-wide integrating of the distributed resources including super computers, data storage systems, and data resources as well as special instruments accessible to organizations; in addition to finding possible solutions to complex problems in science, engineering, and commerce. The main idea of grid computations emerged when there were no resources within management of a scientific problem solution requiring large amount of calculations or data. Almost any grid system possesses particular software for task scheduling. When an applied grid program is presented by a user, the software designates suitable machine (or machines) over which the program is to be performed. In the simplest case, this task is completely done blindly with a Round Robin algorithm. For instance, totally 4 machines were assumed while the requests of performing various applied programs are gradually issued by different users, through separating applied programs into tasks i.e. the first task is dedicated to the first machine, the second to the second, the third to the third, and so on. It must be mentioned that even in case of using a Round Robin algorithm; task performance is not so simple [5].To put it more precisely, performing the next task request goes to the first following machine essentially is capable of performing the task based on resources (hardware and software).The major problem in each grid is optimal and secure resource management based on resource owners' predetermined accessibility policies to resources. Once a user transfers an application to grid server, it is initially required to have a complete list of grid all accessible computational resources in order to know which ones based on what policies can be applied [1]. Researches show heuristic optimization methods', inspired from nature, higher efficiency compared to other methods. Most of these methods tried to minimize makespan tasks. Swarm Intelligence is a type of Artificial Intelligence Method based on swarm behaviors. There are many heuristic swarm intelligence algorithms including Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Firefly Algorithm (FA) proposed for optimization. Of these, PSO algorithm has been proven to be the best heuristic method; the efficiency of which can be improved through combining to other methods since properties such as high velocity convergence, error tolerance, flexibility, and insensitivity to initial values.

2. LITERATURE

Min-Min scheduling function is a rather acceptable efficiency heuristic method. It is initiated with a group of unallocated tasks in 2 steps: In the first step, a series of tasks is calculated with minimum completion time. In the second, the task with the minimum completion tome is selected and allocated to the resources. Then, the allocated tasks are removed from the list of unallocated tasks and the same procedure is repeated for the other unallocated tasks [2] [3]. The function of Max-Min scheduling is similar to Min-Min method consisting 2 steps. In the first step, a set of tasks is computed with the minimum time of completion. In the second step, the task with the maximum completion time is allocated to the resources. In most cases, Max-Min efficiency and load balancing are better than those of Min-Min in the grid resources [2] [4]. Genetic algorithm is an evolutionary technique to search a wide area General procedure of genetic algorithm search is as follows: First, creating the initial population including a set of chromosomes indicating a possible solution. This solution is mapped between the tasks and grid resources. The next step assesses chromosome where a value is attributed to each. This value represents the delay time of chromosome tasks. The purpose of genetic searching is to find chromosome optimal values. The third step incorporates combination and mutation operations. Combination is a process in which particular chromosomes sequences are removed. Mutation is a process changing particular chromosomes' sequences through using multiple task mappings (new for the present population). This

process is repeated until the criteria of the end of the process are satisfied [8]. The algorithms are regarded as the most popular natural heuristic methods in optimization as their simplicity. Simulated Annealing Algorithm is a local search method to find the general optimal solution of a complicated problem. The primary idea of this method was posed in 1953 [7]. In this approach, an object is first heated up to high temperature; then, cooled gradually down so that the system will be maintained at a thermodynamic balance at any time. If the temperature sufficiently decreases, the object reaches a thermal balance called an optimal condition. If, compared with grid system, thermal balance is, indeed, mapping between tasks and grid resources considered as the goal of optimization. Temperature is, then, the sum of one mapping completion times. Cost function and temperature changes cause mapping to be changed. If the next temperature is higher, means that it has not been a good mapping; so, the next case will be most probably accepted since accepting unsuitable conditions provides avoiding local optimality often occurs in local optimum. The system initial temperature is in fact makespan initial mapping implemented as follows: The first mapping is created through a uniform random distribution while the new makespan is evaluated. If the new makespan is greater, then a uniform random number in the range of z ϵ [0, 1] will be selected; z and y are obtained by the relation (1).

$$(1)\ y = \frac{1}{1+e^{\left(\frac{old\ makespan - new\ makespan}{temperture}\right)}}$$

If z>y, then the new mapping is maintained; otherwise rejected and the previous mapping will be maintained, too. Therefore, system temperature is cooled down, while most of unsuitable solutions would be hardly maintained. This may be more reasonable as the probability of finding a better solution than an unsuitable solution goes less when the temperature declines more. Following any changes, system temperature is reduced to 90% of its current level (cooling rate); and one stage of algorithm retrieve is completed. The algorithm will be stopped when there is no more changes in makespan as some iterations or system temperature get close to 0.

3. Resource allocation on grid

Resource Allocation is independent tasks incorporates N tasks and M machines. Each task must be processed by each M machine so that scheduling would ultimately reach its minimum duration. The proposed algorithm considered service quality parameters, makespan, flowtime, and task performance cost, respectively. Each task can only be implemented on one of the resources continuing until the performance is completed. The proposed algorithm utilized ETC matrix model [1]. Since the scheduling algorithm is static, it can be assumed that the expected makespan over each resource of i for each task of j is predetermined inserted in matrix ETC [i, j] [9].

Completion_Time [i, j] is equal to task j completion time in resource i calculated as follows:

(2)Completion_Time[i,j]= ETC[i,j]

Makespan is the maximum completion time of Completion_Time [i, j] computed by the following equation:

(3)Makespan= Max (Completion_Time[i,j]) $1 \le j \le N$, $1 \le i \le M$

Flowtime is sum of tasks completion of Completion_Time [i, j] in all resources calculated by the following equation:

$$(4)\ \text{flowtime} = \sum_{i=1}^{m} \sum \text{completion\_Time}[i,j]$$

Scheduling of the proposed algorithm is focused on reducing makespan, flowtime, and task completion cost in sending tasks to resources.

4. The proposed Recource allocation algorithm

PSO algorithm, first stated by Kennedy and Russell Eberhart in 1995, is inspired by birds and other animal's migrations. In fact, more experienced bird flies ahead at migration to find food. In other words, it can be stated that particles or birds cooperate to find food. In recent years, this algorithm has been widely used in solving different problems including optimization. Moreover, Hass assessment demonstrated the power of this algorithm in solving such problems. The particle Xi possesses a position vector, velocity vector, and its own fitness amount. In each algorithm retrieval, positions and velocity's amounts change through (5) and (6), respectively:

(5) $v_{id}(t+1)= wv_{id}(t)+c_1r_1(pBest_{id}- x_{id}(t))+ c_2r_2(gBest_d-x_{id}(t))$

(6) $x_{id}(t+1)=x_{id}(t)+v_{id}(t+1))$

In above equation, w is the inertia weighted factor, is the best previous position of the particle, , the best previous positions of all previous steps particles, is the velocity of ith particle in repetition of t, is position of the ith particle in iterations of t, r1 and r2 are two random numbers, and C1 and C2 are two constant coefficients.

4.1 Particle representation

Of the main issues in applying PSO algorithm to solve resource allocation problem is how to turn a scheduling problem into a solution or actually how to create mapping between problem solutions and particles in the PSO algorithm. Within scheduling PSO algorithm, each particle is considered as a possible solution to resource allocation in such a way that each particle vector has a length of N, where N is the total number of input tasks. Each element inside the particle vector is a random integer between 1 and M (total number of resources). For instance, the 2nd particle of task 2, T2, performed on the resource 3, R3is illustrated in Fig 2-2.

Recourses/Tasks	T1	T2	T3	T4
Particle1	R1	R2	R3	R2
Particle 2	R1	R1	R3	R3
Particle 3	R1	R3	R1	R2

Figure 1. Displaying typical particles

4.2 Generating initial population

In a standard copy, initial population of PSO algorithm is randomly created so that a random number between 1 and M, representing the resource number on which the favorable task is performed, is generated. In the proposed method, part of initial population (schedulers) is created by Max-Min method. This leads to intelligently creation of initial population as well as improving their qualities and characteristics in order to reach an optimal or near-to-optimal answer as soon as

possible. Part of initial population is randomly created in the suggested method to keep the population diversity; hence, there would be no more in advance algorithm convergence.

4.3 Creating initial velocity vector

A new notion of velocity updating in turbulent PSO algorithm has been introduced based on the minimum velocity constraints [6]. One of the main reasons of PSO algorithm premature convergence is the particles static condition and lack of global search in the problem area. In this model, a procedure is introduced to move low-mobility particles while they are allowed to discover better solutions. If the particles' velocity is reduced (under the threshold) a new velocity will be attributed using following equation. Thus, the turbulent PSO algorithm is provided by the following new velocity equation:

In the above equation, is a random number uniformly selected within the range of [0, 1] and is the scale index to control particular fluctuation ranges based on . is the minimum velocity threshold. is the velocity achieved by the equations (7).

$$\hat{v} = \begin{cases} v_{ij} & if \quad |v_{ij}| \geq v_c \\ u(-1,1)v_{max/\rho} & if \quad |v_{ij}| < v_c \end{cases} \quad (7)$$

4.4 Calculating fitness function in proposed algorithm

The main goal of resource allocation with help of PSO algorithm is to minimize makespan and flowtime. The particle having this feature is more suitable for proposed algorithm.

$$x_i(t+1) = x_i(t) + a(rand - 1/2) \quad (8)$$

4.5 Termination conditions

To finish of swarm Intelligence algorithms such as pso, it must be mentioned the termination conditions. This algorithm will be terminated after reaching maximum iteration.

5. Evaluation the proposed algorithm

The proposed algorithm together with other scheduling algorithms were tested condition table 2-5, where all the models were considered identical to properly evaluate task length. Parameter of the limits of task length is indicative of the limits of uniform distribution of task length. The numbers of iterations show totally 20 retrievals have been implemented using the present algorithms to achieve the program makespan Time; and then, the amounts averages were evaluated.

Table 1. Recourses allocation values

Number of Resource	Number of Task	Typical
64	351	YAR-64-110
64	340	YAR-64-100
64	293	YAR-64-90

The proposed algorithm in condition of table 1 was compared with standard PSO algorithms, simulated annealing algorithms, and genetic algorithms. Before assessing results, it was necessary to determine the initial values of the parameters used in the algorithms. These values are shown in table 2.

Table 2. Initial values of the parameters of scheduling algorithms

Algorithm	Parameter	Value
TPSO	Population size	40
	Self-consciousness study factor C_1	1.49
	Swarm consciousness study factor C_2	1.49
	Inertia factor	0.9
GA	Size of the population	20
	Probability of crossover	0.8
	Probability of mutation	0.02
	Scale for mutations	0.1
SA	Number operations before temperature adjustment	20
	Number of cycles	10
	Temperature reduction factor	0.85
	Vector for control step of length adjustment	2
	Initial temperature	50

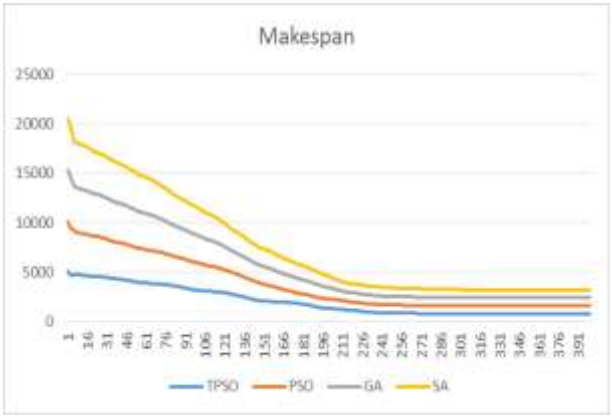


Figure 2. Diagram of makespan

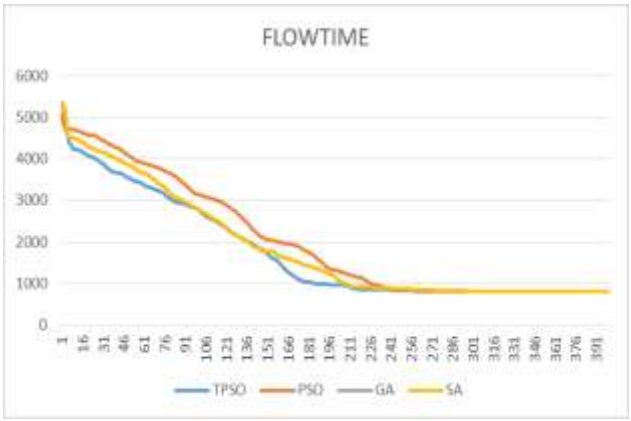


Figure 3. Diagram of flowtime

6. Conclusion

The computational grids provide reliable available to other computational resources. These resources are as Heterogeneous and distributed and are used shared. In additional, resources in grid are belonged to various organizations that have specific management policy and used for different users at different times.. In this complicated media management it cannot be used traditional methods for resources management that try to optimize the efficiency rate at the system level. In this paper, proposed method was presented for scheduling jobs in computational grid. In the proposed method combination of the Turbulent Pso was used.. For this purpose we used multi purposes function. Simultaneously two parameters makespan and flowtime evaluate service quality and minimum sum of the three mentioned parameters. In the proposed method, we have improved the mentioned parameters of service quality such as time of jobs implementation. Mentioned parameters are simulated carefully. The results show the superiority of the proposed method than the compared methods.

8. REFERENCES

[1] Yuanyuan, Z.,Wei, S., Yasushi, I. (2008). "Predict task running time in grid environments based on CPU load predictions", Future Generation Computer Systems 24 (6), pp. 489–497.

[2] Braun, T.D., et al. (2001). "A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems", Journal of Parallel and Distributed computing. 61(6), p.p 810-837.

[3] Chauhan, S.S. and R. Joshi. (2010). "A weighted mean time min-min max-min selective scheduling strategy for independent tasks on grid" , Advance Computing Conference (IACC), 2010 IEEE 2nd International Patiala, pp. 4-9

[4] He, X.S., Sun, X.H. and G. Von L. (2003). "QoS guided min-min heuristic for grid task scheduling" , Journal of Computer Science and Technology, 18(4): p.p 442-451..

[5] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E., Chem J., 1953, "Equation of State Calculations by Fast Computing Machines", 21, 1087-1092.

[6] Liu H., Abraham A., Zhang W.,(2005). " A fuzzy adaptive turbulent particle swarm Optimization", Vol 1, No. 1.

[7] Fidanova, S., (2006). " Simulated Annealing for Grid Scheduling Problem", Modern Computing,. JVA '06. IEEE John Vincent Atanasoff 2006 International Symposium on Sofia. p.p 41-45.

[8] Yang, G., Hongqiang, R., Joshua Z.H., (2005), "Adaptive grid job scheduling with genetic algorithms", Future Generation Computer Systems, p.p 151-161.

[9] Sajjad, A. Ch., Seyed Naser, R., Ali H.,(2014), "Job Scheduling on the Grid Environment using Max-Min Firefly Algorithm", International Journal of Computer Applications Technology and Research, Vol3. Issue 1, p.p 63-67

International Journal of Computer Applications Technology and Research
Volume 3– Issue 8, 535 - 535, 2014

Role of Bisection Method

Chitra Solanki
DIT University
Dehradun, India

Pragati Thapliyal
DIT University
Dehradun, India

Komal Tomar
DIT University
Dehradun, India

Abstract-: The bisection method is the basic method of finding a root. As iterations are conducted, the interval gets halved. So method is guaranteed to converge to a root of "f" if "f" is a continuous function at an interval [a,b] and f(a) and f(b) should have opposite sign. In this paper we have explained the role of bisection method in computer science research. we also introduced a new method which is a combination of bisection and other methods to prove that with the help of bisection method we can also develop new methods. It is observed that scientists and engineers are often faced with the task of finding out the roots of equations and the basic method is bisection method but it is comparatively slow. We can use this new method to solve these problems and to improve the speed.

Key words: continous, absolute error, Iteration, convergence, Newton-Raphson method, Regular- Falsi method

1. Introduction

Traditional iterative schemes such as Newton's method and related classes of algorithms [3,4] often fail to converge to a specific periodic orbit since their convergence is almost independent of the initial guess. Moreover, these methods are affected by the imprecision the mapping evaluations. It may also happen that these methods fail due to the nonexistence of derivatives or poorly behaved partial derivatives [3,4]. Recently, this method has been applied successfully to various difficult problems; see, for example, [7–11]. One of the first numerical methods developed to find the root of a nonlinear equation $f(x)=0$ was the bisection method (also called *binary-search* method)[1]. Since the method is based on finding the root between two points, the method falls under the category of bracketing methods. Since the root is bracketed between two points, x_ℓ and x_u, one can find the mid-point, x_m between x_ℓ and x_u. This gives us two new intervals

2. THE GRAPHICAL DISCRIPTION-:

What is the bisection method and what is it based on?
One of the first numerical methods developed to find the root of a nonlinear equation $f(x)=0$ was the bisection method (also called *binary-search* method). The method is based on the following theorem. [1]

What is the use of bisection method :

- It is used in computer science research to analyze safeguard zero finding methods
- It is simplest of other all methods
- We can safeguard bisection to detect cases where we don't have any roots

Theorem

An equation $f(x)=0$, where $f(x)$ is a real continuous function, has at least one root between x_ℓ and x_u if $f(x_\ell)f(x_u)<0$ (See Figure 1).

Note that if $f(x_\ell)f(x_u)>0$, there may or may not be any root between x_ℓ and x_u (Figures 2 and 3). If $f(x_\ell)f(x_u)<0$, then there may be more than one root between x_ℓ and x_u (Figure 4). So the theorem only guarantees one root between x_ℓ and x_u.

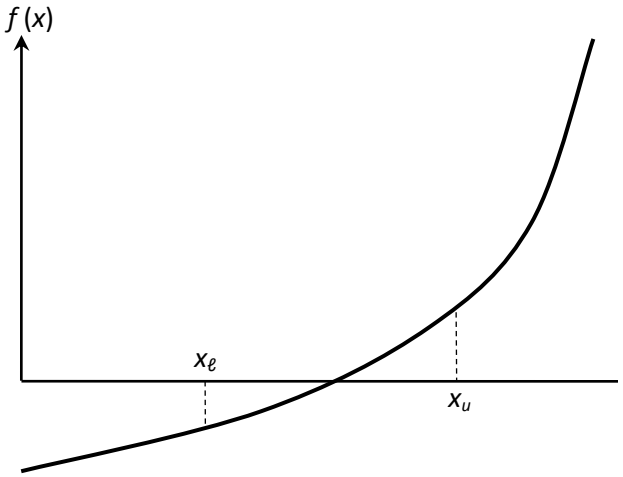


Figure 1 At least one root exists between the two points if the function is real, continuous, and changes sign.

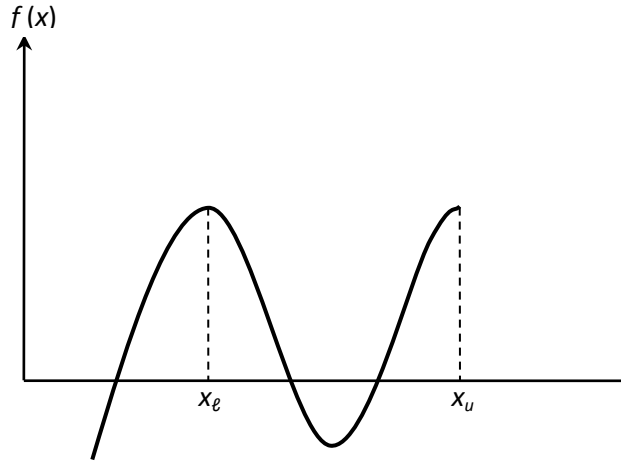


Figure 2 If the function $f(x)$ does not change sign between the two points, roots of the equation $f(x)=0$ may still exist between the two points.

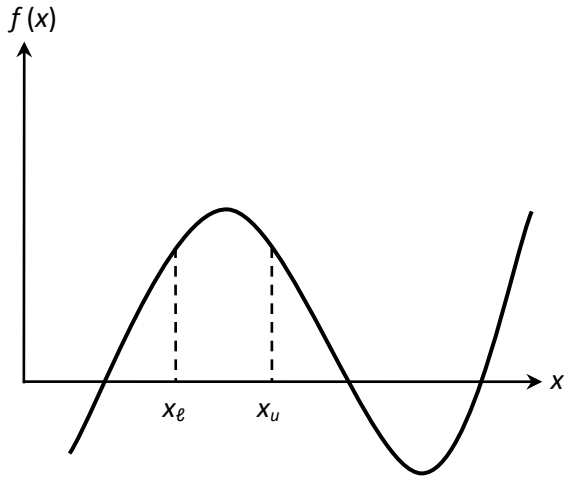


Figure 3 If the function $f(x)$ does not change sign between two points, there may not be any roots for the equation $f(x)=0$ between the two points.

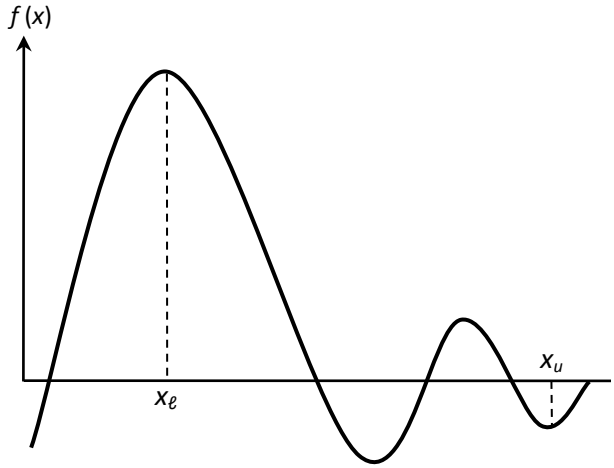


Figure 4 If the function $f(x)$ changes sign between the two points, more than one root for the equation $f(x)=0$ may exist between the two points.

3. PROBLEM DESCRIPTION:-

The bisection method guarantees a root (or singularity) and is used to limit the changes in position estimated by the Newton-Raphson method when the linear assumption is poor. However, Newton-Raphson steps are taken in the nearly linear regime to speed convergence.

In other words, if we know that we have a root bracketed between our two bounding points, we first consider the Newton-Raphson step. If that would predict a next point that is outside of our bracketed range, then we do a bisection step instead by choosing the midpoint of the range to be the next point. We then evaluate the function at the next point and, depending on the sign of that evaluation, replace one of the bounding points with the new point. This keeps the root bracketed, while allowing us to benefit from the speed of Newton-Raphson. Wrong assumption of Newton-Raphson method can increase no. of iterations.

An improved root finding scheme is to combine the BISECTION and REGULAR-FALSI methods.It is relatively faster then bisection method.

4. RELATED WORK:-

we first analyzed some of the conventional root finding methods and their limitations. Bisection always converges but is slow. Newton has quadratic convergence but may fail in some of the cases. Secant is a good alternative to Newton but it oscillates in some of the cases and fails to converge.

- It is explained that it is important that we safeguard bisection to detect cases where

we don't have any roots. The question of guessing the bound is more intuitive.

- The other method like Newton's method have a disadvantage that higher order roots can cause convergence to be slow,and the sequence may take undesirable jumps between roots or take a very large step upon encountering an reflection point. One case where it fails is when derivative of function f(x) is either zero or infinite then it fails to converge.
- We have proposed a new method by combining Bisection method with other methods. So, that we can find roots as well as the method can be fast in solving.
- The multidimensional bisection method allows to solve constrained minimization problem when the feasible region is n-dimensional simplex. This method does not require a differentiability of function and is guaranteed to converge to the minimize for the class of strictly unimodal function[12]

5. PROPOSED-METHOD

$x_{i+1=}3x_{i-1}f(x_i)-x_{i-1}f(x_{i-1})+x_if(x_i)-3x_if(x_{i-1})\ /4[f(x_i)-f(x_{i-1})]$

Algorithm for this new method:

The steps to apply the new method to find the root of the equation Choose x_{i-1} and x_i as two guesses for the root such that $f(x_i)\ f(x_{i-1})<0$, or in other words, $f(x)$ changes sign between x_{i-1} and x_i.

I. Estimate the root lies between x_{i-1} and x_i.

II. $x_{i+1=}3x_{i-1}f(x_i)-x_{i-1}f(x_{i-1})+x_if(x_i)-3x_if(x_{i-1})$ $/4[f(x_i)-f(x_{i-1})]$

III. Now check the following

a) If $f(x_{i+1})<0$; then $x_{i-1=}x_{i+1}$ and the root lies between x_{i+1} and x_i.

b) If $f(x_{i+1})<0$; then $x_{i-1=}x_{i+1}$ and the root lies between x_i and x_{i+1}.

c) If,

new x_{i-1} and x_i are same then previous one then stop and the solution will be $x_{i-1} + x_i / 2$

else

goto step I.

Comparisons table of new method with existing methods:-

For a given problem $\mathbf{f(x)=x^3 - 7}$.**[2]**The comparison is done between four methods the new method is as faster as Newton-Raphson method and Regular -Falsi method and also accurate as we don't take any guess

Table 1 Comparison

S.No.	Method name	No. of iterations
1	BISECTION METHOD	14
2	REGULAR-FALSI METHOD	5
3	NEWTON RAPHSON METHOD	5
4	NEW METHOD	6

6. CONCLUSION:

Bisection method is the safest and it always converges. The bisection method is the simplest of all other methods and is guaranteed to converge for a continuous function. It is always possible to find the number of steps required for a given accuracy.and the new methods can also be developed from bisection method and bisection method plays a very crucial role in computer science research.

7. REFERENCES :

[1] Chapter 03.03 Bisection Method of Solving a Nonlinear Equation

[2] Book numerical based analysis from DITU library

[3] J.M. Ortega, W.C. Rheinbolt, Iterative solution of nonlinear equations in several (1970)

[4] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, SIAM, Philadelphia, 1996

[5] L. Drossos, O. Ragos, M.N. Vrahatis, T.C. Bountis, Phys. Rev. E 53 (1996) 1206.

[6] M.N. Vrahatis, T.C. Bountis, M. Kollmann, Inter. J. Bifurc. Chaos 6 (1996) 1425.

[7] M.N. Vrahatis, H. Isliker, T.C. Bountis, Inter. J. Bifurc. Chaos 7 (1997) 2707.

[8] H. Waalkens, J. Wiersig, H.R. Dullin, Ann. Phys. 260 (1997) 50.

[9] N. Buri´c, M. Mudrini´c, J. Phys. A: Math. Gen. 31 (1998) 1875.

[10] N. Buri´c, M. Mudrini´c, Todorovi´c, J. Phys. A: Math. Gen. 31 (1998) 7847.

[11] V.S. Kalantonis, E.A. Perdios, A.E. Perdiou, M.N. Vrahatis, Celest. Mech. Dynam. Astron. (2001), in press.

[12] A multidimensional bisection method for unconstrained minimization problem