



# PRIVACY-PRESERVING FOR DATA LEAKAGE DETECTION

Sethupathi.c – Paavai Engineering College

Guided by – Mr.M.SivaGanesh., M.E., – Paavai Engineering College

Edited by : Sethupathi.C

## ABSTRACT

The data from investigate organizations, safety firms and government organizations demonstrate that the numbers of data leak instances have developed rapidly. Among different data leak cases, major causes of data loss are one of the human being mistakes. Present live solutions detect unintentional responsive data leaks caused by human being mistakes and to provide alerts intended for organization. In this system, there a privacy preserving data leak detection solution where a particular set of receptive data digests is used in discovery. Privacy-preserving data-leak detection representation intended for preventing inadvertent data leak in scheme traffic. Such a demonstration yields a powerful and delegatable data-leak detection framework. The cloud computing surroundings the cloud source can perform data leak detection as add on service to its customers. The assessment results show that the detection method can supports accurate detection through very miniature number of false alarms under various data leak scenarios. The benefit of this method is that it enables the data owner to safely hand over the detection operation without enlightening the sensitive data.

## INTRODUCTION

Typical approaches to preventing data leak are under two categories host-based solutions and network based solutions. Host-based approaches may include encrypting data when not used detecting stealthy malware with antivirus scanning or monitoring the host and enforcing policies to restrict the transfer of sensitive data. These approaches are complementary and can be deployed simultaneously. It presents a network-based data-leak detection (DLD) solution that complements host-based methods. Network-based data-leak detection focuses on analyzing unencrypted outbound network through deep packet inspection or information theoretic analysis. In addition to provide the regular networking, computing, or storage services, network or cloud providers may introduce additional security protection for their customers. For their customers, these add-on security services such as data-leak detection are attractive, as they may have a lower cost compared to building inhouse security management of their own. Thus, one may outsource the data-leak detection to a DLD provider. However, the data owner may not allow the DLD provider to access the sensitive data.

To overcome existing issues, proposed system offer feature like security constraint before processing incoming request, Complex encryption system before store and decryption after retrieval information form back end server. Proposed system has implementation of AES, and domain ads on service, policy based request filter system. Those are performing individual action on every stage of this outsource business. All incoming request are easily traced and direct access on hosted server are prohibited. The technical challenge is that the detection algorithm needs to provide guarantees on the secrecy of customers' sensitive data while still enabling the provider to identify signs of data leak. This problem of the lack of support for privacy-enhancing data leak detection

has not been systematically addressed in the security literature. In this paper proposed design, implement, and experimentally evaluate a client technique that enhances the data privacy during the data-leak detection operations. Our method is based on a fast and practical one-way computation and does not require any expensive cryptographic operations. In the academic section available academic studies have been clustered into various categories according to the nature of the leakage and the protection provided. Next, the main data leakage scenarios are described, each with the most relevant and applicable solution or approach that will mitigate and reduce the likelihood or impact of data leakage.

Data-leak detection solution which can be outsourced and be deployed in a fully honest and reliable detection environment. The design, implements, and evaluates our policy based request filter, base information encryption that enhances data privacy during data-leak detection operations. Data leak approach is based on a fast and practical one-way computation on the sensitive data. It enables the data owner to securely delegate the content inspection task to DLD providers without exposing the sensitive data. Using data detection method, the DLD provider, who is modeled as an honest-but-curious adversary, can only gain limited knowledge about the sensitive data from either the released digests, or the content being inspected. Using fuzzy techniques, an Internet service provider (ISP) can perform detection on its customers' traffic securely and provide data leak detection as an add-on service for its customers. In another scenario, individuals can mark their own sensitive data and ask the administrator of their local network to detect data leaks for them.

## LITERATURE SURVEY

Proposed that fingerprinting scheme is a method for implementing fingerprints using polynomials over a finite field. The basic idea is that the file system computes the cryptographic hash of each block in a file. To save on transfers between the client and server, they compare their checksums and only transfer blocks whose checksums differ. But one problem with this scheme is that a single insertion at the beginning of the file will cause every checksum to change if fixed-sized (e.g. 4 KB) blocks are used. So the idea is to select blocks not based on a specific offset but rather by some property of the block contents. Any hash function could be used to divide a long file into blocks (as long as a cryptographic hash function is then used to find the checksum of each block): but the Rabin fingerprint is an efficient rolling hash, since the computation of the Rabin fingerprint of region B can reuse some of the computation of the Rabin fingerprint of region A when regions A and overlap. The challenges are currently, the new pointer blocks always hold a fixed number of scores. When a new block is inserted in an existing hash tree is split in the new archive, it now changes all pointer blocks to right of it in the hash tree.

The classic analysis of hashing schemes often entails the assumption that the hash functions used are random. More precisely, the assumption is that keys belonging to a universe  $U$  are hashed into a table of size  $M$  by choosing a function  $h$  uniformly at random among all the functions. File access patterns should also be taken into consideration. Whole file content and fixed size blocking strategies present the disadvantage that file updates may lead to the recompilation of SHA-1 digests for large amounts of data. There are lots of string matching algorithms that are faster than  $O(n+m)$  It's practically as slow as brute force matching and it requires additional space. Rather simple, but currently the size of a sub tree is always stored in 8 bytes. This grows the size of one pointer from 20 bytes to 28 bytes. The lowest-level pointer blocks, which point to blocks of at most 32KB, will be the most numerous. Data security is a company imperative. Seeks to limit the distribution of personally identifiable information consistent with the nature and sensitivity of the information. It strives to accurately report information in its products.

proposed that a bloom filter is a simple-efficient randomized data structure for representing a set in order to support membership queries. bloom filters allow false positives but the space savings often out of weigh this drawback when the probability of an error is made sufficient low. Bloom introduced bloom filters in the 1970's and ever since they have been very popular in database applications. Expressed that produces accurate correction results with much less memory compared with previous solutions. The algorithm, named Bloom-filterbased Error correction Solution for high-throughput Sequencing reads (BLESS), uses a single minimum-sized Bloom filter, and is also able to tolerate a higher falsepositive rate, thus allowing us to correct errors with a  $40\times$  memory usage reduction on average compared with previous methods. With the aims of providing a unified mathematical and practical framework for them and simulating their use in future application.

Exposed that the move toward large-scale deployment of electronic commerce is hampered by privacy concerns of potential customers. To appease such concerns, enterprises publish privacy statements that promise fair information practices. These problems are amplified if personal data is used not only by the enterprise that collected the data, but also by secondary users such as partner organizations, and government agencies. These flows of data are complex. A privacy control model should reflect that there are three entities that influence authorizations. The security administrator protects the interest of the organization. The challenges are Data security is a company imperative. Seeks to limit the distribution of personally identifiable information consistent with the nature and sensitivity of the information. It strives to accurately report information in its products.

That traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. In this paper, for the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. Fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. In our solution, exploit edit distance to quantify keywords similarity and develop two advanced techniques on constructing fuzzy keyword sets. A server that communicates with a target application can unexpectedly change its behavior during test runs, which is beyond our control. During installation and initial use, an application may prompt for user preferences or for personal data such as name, email address, organization, gender, and zip code. In our solution, exploit edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads.

## SYSTEM SPECIFICATION

### 3.1 HARDWARE SPECIFICATION

SYSTEM	:	Pentium IV 2.4 GHz
HARD DISK	:	40 GB
FLOPPY DRIVE	:	1.44 MB
MONITOR	:	15 VGA color
MOUSE	:	Logitech.
RAM	:	256 MB
KEYBOARD	:	110 keys enhanced.

### 3.2 SOFTWARE SPECIFICATION

OPERATING SYSTEM	:	WINDOWS XP PROFESSIONAL
FRONT END	:	.NET
BACK END	:	My SQL

## SYSTEM STUDY AND ANALYSIS

### 4.1 EXISTING SYSTEM

Existing system lets to detect and prevent data leaks require a solutions for outsourcing environment, where all business information are outsourced in terms of application hosting. The environment have front end and back end mode, there single model available for data-leak detection, data confinement, stealthy malware detection, and policy enforcement. Network pocket inspection techniques provide Network data-leak detection (DLD) by performing deep packet inspection (DPI) and searches for any occurrences of sensitive data patterns. All kind of protocol based communication has the TCP/IP as base for the communication. So header information section will have much section to have Meta data about pocket information.

**DISADAVANTAGE**

- As the Internet grows and network bandwidth continues to increase, administrators are faced with the task of keeping confidential information from leaving their networks.
- In response, researchers have created data loss prevention systems that check outgoing traffic for known confidential information.
- These systems stop naive adversaries from leaking data, but are fundamentally unable to identify encrypted or obfuscated information leaks.

**4.2 PROPOSED SYSTEM**

Data-leak detection solution which can be applied on outsource provider place, where outsourced data can be deployed in a fully honest detection environment. Design, implement, and evaluate our policy based filter system, and base encryption system technique that enhances data privacy during data-leak detection operations. AES approach is based on a fast and practical two-way computation on the sensitive data .which means data stored in cipher text in base system, and incoming request and outgoing response are deeply investigated. It enables the data owner to securely delegate the content-inspection task to DLD providers without exposing the sensitive data. Using detection techniques, an Internet service provider (ISP) can perform detection on its customers' traffic securely and provide data-leak detection as an add-on service for its customers. Host server logs are incoming request resource identity and prefix the response content for the all request based on comparison on existing activity log. Rather than making privacy content with business information, the new system looks out into the request limit and response content.

**ADVANTAGES**

- New system describe own privacy-preserving data leak detection model for preventing inadvertent data leak in network traffic.
- Proposed model supports detection operation delegation and ISPs can provide data-leak detection as an add-on service to their customers using our model.
- Every business records are converted and stored in terms of cipher text in base information system.
- For this implementation AES (Advanced encryption system) technique applied.
- Data leak detection system and perform extensive experimental evaluation on internet surfing traffic of 100 users, and also N number simulated request is prohibit.

**4.3 FEASIBILITY STUDY**

A feasibility study is carried out to select the best system that meets performance requirements. The main aim of the feasibility study activity is to determine whether it would be financially and technically feasible to develop the product. The feasibility study activity involves the analysis of the problem and collection of all relevant information relating to the product such as the different data items which would be input to the system, the processing required to be carried out on these data, the output data required to be produced by the system as well as various constraints on the behavior of the system.

**4.4 TECHNICAL FEASIBILITY**

This is concerned with specifying equipment and software that will successfully satisfy the user requirement. The technical needs of the system may vary considerably, but might include:

- The facility to produce outputs in a given time.
- Response time under certain conditions.

- Ability to process a certain volume of transaction at a particular speed.
- Facility to communicate data to distant locations.

In examining technical feasibility, configuration of the system is given more importance than the actual make of hardware. The configuration should give the complete picture about the system's requirements: How many workstations are required, how these units are interconnected so that they could operate and communicate smoothly? And what speeds of input and output should be achieved at particular quality of printing.

#### 4.5 ECONOMIC FEASIBILITY

Economic analysis is the most frequently used technique for evaluating the effectiveness of a proposed system. More commonly known as Cost / Benefit analysis, the procedure is to determine the benefits and savings that are expected from a proposed system and compare them with costs. If benefits outweigh costs, a decision is taken to design and implement the system. Otherwise, further justification or alternative in the proposed system will have to be made if it is to have a chance of being approved. This is an outgoing effort that improves in accuracy at each phase of the system life cycle.

#### 4.6 OPERATIONAL FEASIBILITY

This is mainly related to human organizational and political aspects. The points to be considered are:

- What changes will be brought with the system?
- What organizational structure are disturbed?
- What new skills will be required? Do the existing staff members have these skills? If not, can they be trained in due course of time?

This feasibility study is carried out by a small group of people who are familiar with information system technique and are skilled in system analysis and design process. Proposed projects are beneficial only if they can be turned into information system that will meet the operating requirements of the organization. This test of feasibility asks if the system will work when it is developed and installed.

### SOFTWARE DESCRIPTION

#### 5.1 .NET

Visual Studio .Net is the fast application improvement device for BASIC. Visual Studio .Net offers complete mix with ASP.NET and empowers to move and customize server controls and outline Web Forms as they ought to show up when client sees them. A percentage of alternate points of interest of making BASIC applications in Visual Studio .Net are

- Visual Studio .Net is a Rapid Application (RAD) apparatus. Rather than adding every control to the Web Form automatically, it serves to include these controls by utilizing tool stash, sparing programming endeavors.
- Visual Studio .Net backings custom and composite controls. Can make custom controls that embody a typical usefulness that may need to use in various applications.
- Visual Studio .Net makes a glorious showing of rearranging the creation and utilization of Web Services. Much of the software engineer neighborly stuff (making all the XML based reports) happens consequently, without much exertion on the developer's side.
- A characteristic based writing computer program is an effective idea that empowers Visual Studio .Net to mechanize a considerable measure of software engineer unpleasant assignments.

## 5.2 .NET PROGRAMMING DIALECTS

The .NET Framework gives an arrangement of instruments that assistance to assemble code that works with the .NET Framework, Microsoft gives an arrangement of dialects that are as of now .NET perfect. Fundamental is one of those dialects.

### 5.3 ASP.NET ENVIRONMENT

Dynamic Server Pages were discharged by Microsoft to empower the formation of element pages taking into account client information and cooperation with a Web website. ASP.NET enhances the first ASP by giving code-behind. With ASP.NET and code-behind, the code and HTML can be isolated.

ASP.NET Web administrations are XML-construct benefits that are presented with respect to the Internet that can be gotten to by other Web administrations and Web administration customers.

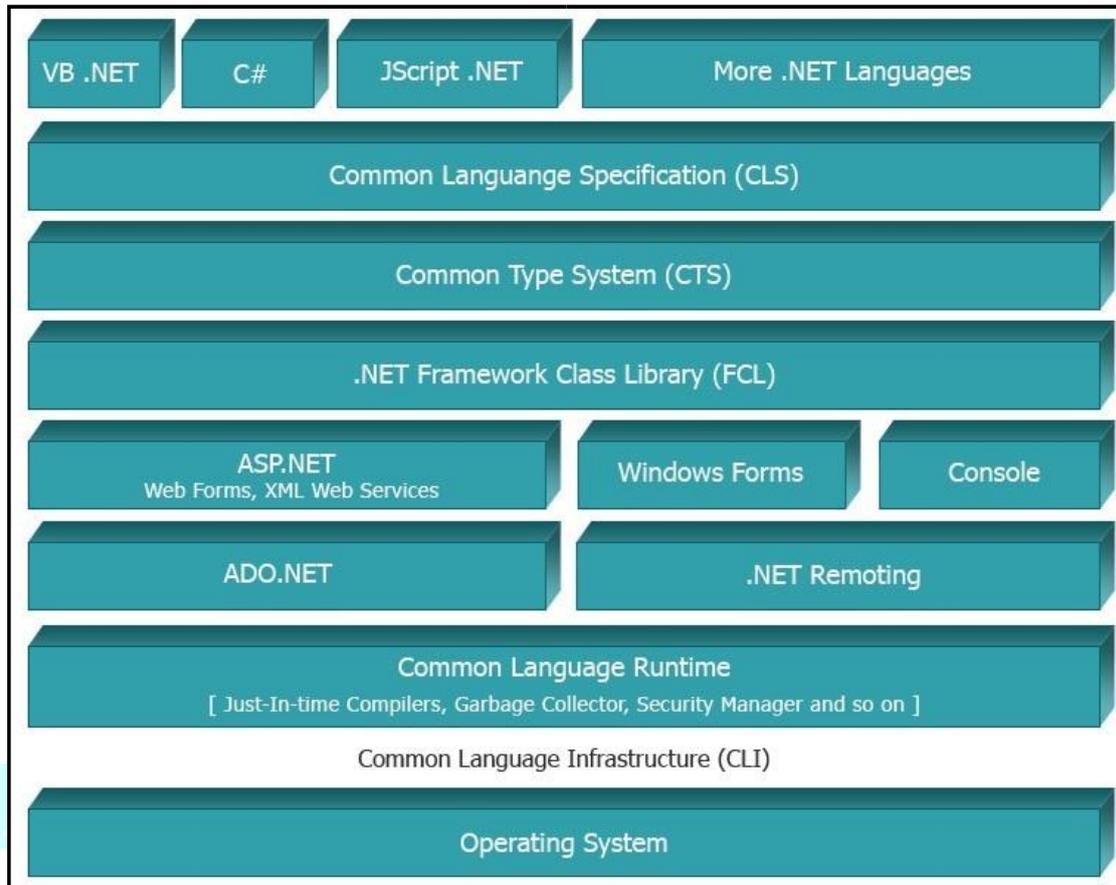
### 5.4 ASP.NET

ASP.NET is more than the following form of Active Server Pages (ASP); it is a brought together Web advancement stage that gives the administrations important to designers to fabricate undertaking class Web applications. While ASP.NET is to a great extent sentence structure perfect with ASP, it likewise gives another programming model and foundation for more secure, versatile, and stable applications.

ASP.NET is an assembled, .NET-based environment; you can creator applications in any .NET perfect dialect, including VisualBasic.NET, BASIC, and JScript.NET. Furthermore, the whole .NET Framework is accessible to any ASP.NET application. Engineers can undoubtedly get to the regale of these advances, which incorporate oversaw normal dialect runtime environment, sort wellbeing, legacy, et cetera.

ASP.NET has been intended to work consistently with WYSIWYG HTML editors and other programming instruments, including Microsoft Visual Studio .NET. Does this make Web improvement simpler, as well as gives every one of the advantages that these apparatuses bring to the table, including a GUI that designers can use to drop server controls onto a Web page and completely coordinated investigating backing. Engineers can browse the accompanying two elements when making a " ASP.NET application, Web Forms and Web administrations, or consolidate these in any capacity they see fit.

- Web Forms permits you to assemble intense structures based Web pages. At the point when building these pages, you can utilize ASP.NET server controls to make normal UI components, and system them for basic assignments. These controls permit you to quickly assemble a Web Form out of reusable implicit or custom segments, rearranging the code of a page.
- An XML Web administration gives the intends to get to server usefulness remotely.



## 5.5 FEATURES

- **INTUITIVE C++ BASED LANGUAGE**

Utilize a dialect displayed on C++ linguistic structure, instantly commonplace to C++ and Java designers, and also natural new dialect builds that incredibly streamline advancement errands.

- **RELIABLE INTEROPERABILITY**

Utilize code to call local Windows APIs, use pre-constructed COM parts, and influence existing ActiveX controls to flawlessly coordinate existing applications and segments.

- **ADVANCED, COMPONENT-ORIENTED LANGUAGE**

Exploit inborn backing for properties, indexers, delegates, single and multidimensional clusters, propelled legacy, traits, forming, and XML remarks.

- **CAPABLE DEBUGGING AND TESTING TOOLS**

ASP .NET incorporates a capable remote and multi-dialect debugger, empowering engineers to test applications and fabricate solid multi-level arrangements that compass process limits and are composed in different programming dialects.

## 5.6 NET FRAMEWORK CLASS LIBRARY

Addition experienced and capable, constructed in usefulness, including a rich arrangement of accumulation classes, systems administration bolster, multithreading bolster, string and customary expression classes, and wide backing for XML, XML patterns, XML namespaces, XSLT, XPath, and SOAP.

## 5.7 POWERFUL WEB DEVELOPMENT ENVIRONMENT

Make Web-based arrangements in C# utilizing the mutual Web Forms Designer and XML Designer. Engineers can likewise utilize IntelliSense elements and label finish or pick the WYSIWYG manager for move and customize creating to construct intelligent Web applications.

## 5.8 .NET FRAMEWORK

Microsoft planned VB from the beginning to exploit its new .NET Framework. The .NET Framework is comprised of four sections, the Common Language Runtime, an arrangement of class libraries, an arrangement of programming dialects, and the ASP.NET environment. The .NET Framework was composed on account of three objectives. In the first place, it was planned to make Windows applications considerably more solid, while likewise furnishing an application with more prominent level of security.

Second, it was proposed to improve the advancement of Web applications and administrations that work in the conventional sense, as well as on cell phones too. Finally, the structure was intended to give a solitary arrangement of libraries that would work with various dialects. The .NET Framework is the base for the new Microsoft .NET Platform. Furthermore, it is a typical situation for building, conveying, and running Web applications and Web Services. The .NET Framework contains a typical dialect runtime and basic class libraries - like ADO .NET, ASP .NET and Windows Forms - to give propelled standard administrations that can be coordinated into a mixed bag of PC frameworks. The .NET Framework gives a component rich application environment, streamlined improvement and simple mix between various diverse advancement dialects. The .NET Framework is dialect nonpartisan. At present it bolsters C++, C#, Visual Basic, and Jscript. Microsoft's Visual Studio.NET is a typical advancement environment for the new .NET Framework.

## 5.9 COORDINATING WITH IIS

IIS is the web server is utilized here. IIS 5.0 or above is key for the ASP.NET for the earth. This arrival of ASP.NET uses IIS 5.0 as the priKim host environment. IIS dependably accept that an arrangement of accreditations maps to a Windows NT record and uses them to verify a client. There are three various types of validation accessible in IIS 5.0: BASIC, DIGEST, and INTEGRATED WINDOWS Authentication (NTLM or Kerberos). You can choose the kind of verification to use in the IIS regulatory administrations.

On the off chance that you ask for a URL containing an ASP.NET application, the solicitation and confirmation data are given off to the application. ASP.NET gives the two extra sorts of verification depicted in the accompanying table.

## 5.10 WEB SERVICE

Web administrations are ostensibly the most energizing and improve elements of Microsoft's. NET activity and they are liable to significantly influence the way business collaborate utilizing PC application. Rundown of conceivable Web administrations is as changes as the rundown of conceivable business opportunities. Web administration would normally perform a center business administration, for example, client confirmation, Visa approval, valuing a derivatives security, submitting a buy request for a stock or estimating a same-day shipment.

A web administration is a part that performs a capacity or administration. A segment is a bit of programming that has a very much characterized interface, shrouded internals, and the ability of being found. By "found" implies that you can figure out what the part manages without expecting to see the code inside of it. A segment is like a strategy since we can call it with contentions that fit an arrangement of parameters, and it has the ability of returning results.

A web administration might likewise return data to the guest. This administration dwells some place on the Web and can be gotten to from different areas on the Web. For this administration to be called, there are various components that must be set up. To start with, the guest must know how to call the administration. Second, the call must be made over the Web. At long last, the "web administration must know how to react".

## 5.11 SQL SERVER 8.0

SQL Server 2008 (formerly codenamed "Katmai")<sup>[12][13]</sup> was released on August 6, 2008, announced to the SQL Server Special Interest Group at the ESRI 2008 User's Conference on August 6, 2008 by Ed Katibah (Spatial Program Manager at Microsoft), and aims to make data management self-tuning, self organizing, and self maintaining with the development of SQL Server Always On technologies, to provide near-zero downtime. SQL Server 2008 also includes support for structured and semi-structured data, including digital media formats for pictures, audio, video and other multimedia data. In current versions, such multimedia data can be stored as BLOBs (binary large objects), but they are generic bitstreams. Intrinsic awareness of multimedia data will allow specialized functions to be performed on them. According to Paul Flessner, senior Vice President of Server Applications at Microsoft, SQL Server 2008 can be a data storage backend for different varieties of data:

XML, email, time/calendar, file, document, spatial, etc. as well as perform search, query, analysis, sharing, and synchronization across all data types. Other new data types include specialized date and time types and a spatial data type for location-dependent data. Better support for unstructured and semi-structured data is provided using the new file stream data type, which can be used to reference any file stored on the file system. Structured data and metadata about the file is stored in SQL Server database, whereas the unstructured component is stored in the file system. Such files can be accessed both via Win32 file handling APIs as well as via SQL Server using T-SQL; doing the latter accesses the file data as a BLOB. Backing up and restoring the database backs up or restores the referenced files as well.<sup>[17]</sup> SQL Server 2008 also natively supports hierarchical data, and includes TSQL constructs to directly deal with them, without using recursive queries. The full-text search functionality has been integrated with the database engine. According to a Microsoft technical article, this simplifies management and improves performance. Spatial data will be stored in two types. A "Flat Earth" (GEOMETRY or planar) data type represents geospatial data which has been projected from its native, spherical, coordinate system into a plane. A "Round Earth" data type (GEOGRAPHY) uses an ellipsoidal model in which the Earth is defined as a single continuous entity which does not suffer from the singularities such as the international dateline, poles, or map projection zone "edges". Approximately 70 methods are available to represent spatial operations for the Open Geospatial Consortium Simple Features for SQL, Version 1.1. SQL Server includes better compression features, which also helps in improving scalability. It enhanced the indexing algorithms and introduced the notion of filtered indexes. It also includes Resource Governor that allows reserving resources for certain users or workflows. It also includes capabilities for transparent encryption of data (TDE) as well as compression of backups.<sup>[15]</sup> SQL Server 2008 supports the ADO.NET Entity Framework and the reporting tools, replication, and data definition will be built around the Entity Data Model.<sup>[21]</sup> SQL Server Reporting Services will gain charting capabilities from the integration of the data visualization products from Dundas Data Visualization, Inc., which was acquired by Microsoft.<sup>[22]</sup> On the management side, SQL Server 2008 includes the Declarative Management Framework which allows configuring policies and constraints, on the entire database or certain tables, declaratively.<sup>[14]</sup> The version of SQL Server Management Studio included with SQL Server 2008 supports IntelliSense for SQL queries against a SQL Server 2008 Database Engine.<sup>[23]</sup> SQL Server 2008 also makes the databases available via Windows PowerShell providers and management functionality available as Cmdlets, so that the server and all the running instances can be managed from Windows

Power Shell

## THE MOST CRITICAL PARTS OF SQL SERVER 8 ARE

Social database frameworks are the most critical database frameworks utilized as a part of the product business today. A standout amongst the most remarkable frameworks is Microsoft SQL Server. SQL Server is a database administration framework created and showcased by Microsoft. It runs solely under Windows NT and Windows 95/98.

- SQL Server is anything but difficult to utilize.
- SQL Server scales from a portable tablet to symmetric multiprocessor frameworks.
- SQL Server gives information warehousing elements that as of recently have just been accessible in Oracle and other more costly DBMSs.

A database framework is a general gathering of distinctive database programming segments and databases containing the parts viz. Database application projects, Front-End segments, Database administration frameworks, and Databases.

## A DATABASE FRAMEWORK MUST GIVE THE ACCOMPANYING ELEMENTS

- A mixture of client interfaces
- Physical information autonomy
- Logical information autonomy
- Query advancement
- Data honesty
- Concurrency control
- Backup and recuperation
- Security and approval

SQL Server is a Relational Database Management System. The SQL Server social dialect is called Transact-SQL. SQL is resource arranged dialect. This implies that SQL can inquiry numerous lines from one or more tables utilizing only one announcement. This component permits the utilization of this dialect at a coherently larger amount than procedural dialects. Another vital property of SQL is its non-procedurally. SQL contains two sub dialects DDL and DML.

SQL Server functions as a characteristic augmentation of Windows NT and windows 95/98. SQL Server is generally simple to oversee through the utilization of a graphical registering environment for each undertaking of framework and database organization. SQL Server uses administrations of Windows NT to offer new or expanded database capacities, for example, sending and accepting messages and overseeing login security.

The SQL Server chairman's essential device for connecting with the framework is Enterprise Manager. The Enterprise Manager has two primary purposes: Administration of the database server and Management of database items.

- SQL Server Query Analyzer gives a graphical presentation of the execution arrangement of a question and a programmed segment that recommends which list ought to be utilized for a chose inquiry. This intelligent segment of SQL Server performs the assignments like:
- Generating and executing Transact-SQL explanations
- Putting away the produced Transact-SQL explanations in a document

- Analyzing execution gets ready for produced inquiries
- Graphically representing the execution arrangement for a chose question.

A put away method is an exceptional sort of clump written in Transact-SQL utilizing the SQL dialect and SQL augmentations. It is saved money on the database server to enhance the execution and consistency of monotonous undertakings. SQL Server backings put away methods and framework techniques. Put away techniques can be utilized for the accompanying purposes: to control access approval, to make a review trial of exercises in database tables, to discrete information definition & information control articulations concerning a database & every single comparing application.

The database article perspective can be utilized for:

- Restricting the utilization of specific sections and lines of tables - that is to control access to a specific piece of one or more tables,
- To shroud the points of interest of confounded inquiries, to limit embedded & redesigned qualities to certain extents.

The Query Optimizer is the piece of SQL Server that chooses how to best perform a question. It creates a few inquiry execution gets ready for the given question & chooses the arrangement with the most minimal expense.

SQL Server can work in one of two security modes:

- Windows NT
- Mixed

Windows NT security mode solely utilizes Windows NT client records to sign into the SQL Server framework. Blended mode permits clients to associate with SQL Server utilizing the Windows NT security framework or the SQL Server framework. Moreover it gives three security offices to controlling access to database objects:

- Transact-SQL explanations GRANT, DENY, and REVOKE.
- Views.
- Stored methodology

A Windows NT client record or a SQL server login name permits a client to sign into the SQL server framework. A client who hence needs to get to a database of the framework needs a database client record to work in the DB. In this manner clients must have a DB client represent each DB they need to utilize. In the event that there is no such record the client may be permitted to work in the DB under the visitor account."

Put away methods can likewise be utilized to limit information access. The confinement of information access utilizing put away methodology is based upon the property that the consent to execute a put away' strategy is free of any authorization for DB objects that are referenced by the put away system.

SQL server gives an instrument called a trigger for upholding procedural respectability requirements.

A DBMS handles 2 sorts of honesty requirements:

- Declarative Integrity limitations characterized utilizing CREATE& ALTER TABLE articulations.
- Procedural honesty requirements took care of by triggers.

A trigger is an instrument that is conjured when a specific activity happens on a specific table. Every trigger has 3 general parts:

- A name
- The activity
- The execution

SQL server keeps record of every change it makes to the db amid an exchange. This is essential in the event that a lapse happens amid the execution of the exchange. For this situation all already executed explanations inside of the exchange must be moved back. SQL server keeps every one of these records, specifically the previously, then after the fact values, in one or more documents called the exchange log. Each DB of the SQL server framework has its own particular exchange log. Concurrency in multi-client frameworks, for example, SQL Server has chosen impact of execution. At the point when access to the information is taken care of such that stand out project at once can utilize the information, preparing moderates significantly. SQL Server like all different DBMSs takes care of this issue utilizing exchanges. All announcements inside an exchange manufacture a nuclear unit. This implies that either all announcements are executed or for the situation of disappointment, all announcements are wiped out.

### 5.12 ELEMENTS OF SQL SERVER

Microsoft SQL Server bolsters a full arrangement of elements that outcome in the accompanying. SQL incorporates an arrangement of managerial and advancement instruments that enhance our capacity to introduce, convey, oversee and use SQL Server over a few locales.

- **ADAPTABILITY**

The same database motor can be utilized crosswise over stages going from smart phones Microsoft Windows95 to substantial; multiprocessor servers running Microsoft Windows NT, Enterprise Edition.

- **EASE IN BUILDING INFORMATION DISTRIBUTION CENTERS**

SQL Server incorporates instruments for removing and examining synopsis information for online investigative preparing (OLAP). SQL Server likewise incorporates apparatuses for outwardly planning databases and breaking down information utilizing English based inquiries.

### 5.13 SQL API (SQL APPLICATION PROGRAMMING INTRTFACE)

Implanted SQL applications utilize the DB-library DLL to get to SQL server. The SQL Server ODBC driver clients don't get to Microsoft SQL Server straightforwardly. They utilize an application kept in touch with access the information in SQL Server. SQL Server can likewise be gotten to through COM, Microsoft ActiveX, or Windows DNA (Windows Distributed Internet Applications Architecture) parts. Applications are composed to get to SQL Server through a database Application Programming Interface (API).

### 5.14 WEB CLIENTS

A Web customer comprises of two sections:

- Dynamic Web pages containing different sorts of markup dialect which are created by Web parts running in the Web level.
- Web program, which renders the pages got from the server.

A Web customer is now and again called a slim customer. Slim customers as a rule don't question databases, execute complex business guidelines, or associate with legacy applications.

## 5.15 HTML

HTML remains for Hyper Text Markup Language. It is a basic content designing dialect used to make hypertext records. It is a stage free dialect not at all like most other programming dialect. HTML is impartial and can be utilized on numerous stage or desktop. It is this component of HTML that makes it mainstream as standard on the WWW. This adaptable dialect permits the making of hypertext connections, otherwise called hyperlinks. These hyperlinks can be utilized to unite reports on diverse machine, on the same system or on an alternate system, or can even indicate purpose of content in the same record. HTML is utilized for making archives where the accentuation is on the presence of the record. It is likewise utilized for DTP. The records made utilizing HTML can have content with diverse sizes, weights and hues. It can also contain graphics to make the document more effective. Hyper Text Markup Language, commonly referred to as HTML, is the standard markup language used to create web pages. Along with CSS, and JavaScript, HTML is a cornerstone technology, used by most websites to create visually engaging web pages, user interfaces for web applications, and user interfaces for many mobile applications.<sup>[1]</sup> Web browsers can read HTML files and render them into visible or audible web pages. HTML describes the structure of a website semantically along with cues for presentation, making it a markup language, rather than a programming language. HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. The language is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>). Browsers do not display the HTML tags and scripts, but use them to interpret the content of the page. HTML can embed scripts written in languages such as JavaScript which affect the behavior of HTML web pages. Web browsers can also refer to Cascading Style Sheets (CSS) to define the look and layout of text and other material.

## 5.16 FEATURES OF HTML

- **Web Workers:** Certain web applications use heavy scripts to perform functions. Web Workers use separate background threads for processing and it does not affect the performance of a web page.
- **Video:** You can embed video without third-party proprietary plug-ins or codec. Video becomes as easy as embedding an image.
- **Canvas:** This feature allows a web developer to render graphics on the fly. As with video, there is no need for a plug in.
- **Application caches:** Web pages will start storing more and more information locally on the visitor's computer. It works like cookies, but where cookies are small, the new feature allows for much larger files. Google Gears is an excellent example of this in action.

Geo location: Best known for use on mobile devices, geo location is coming with HTML5.

## PROJECT DESCRIPTION

### 6.1 PROBLEM DEFINITION

- Data leakage is defined as the accidental or unintentional distribution of private or sensitive data to an unauthorised entity .
- Data leakage poses a serious issue for companies as the number of incidents and the cost to those experiencing them continue to increase.
- Data leakage is enhanced by the fact that transmitted data including emails, instant messaging, website forms, and file transfers among others, are largely unregulated and unmonitored on their way to their destinations.
- The main scope of this module is provide complete information about the data/content that is accessed by the users within the website.

## 6.2 MODULES

- Network security privacy module
- Security goal threat module
- Experimental evaluation module
- Domain add-on module
- Policy filter module

### NETWORK SECURITY PRIVACY MODULE

Network-accessible resources may be deployed in a network as surveillance and early-warning tools, as the detection of attackers are not normally accessed for legitimate purposes. Techniques used by the attackers that attempt to compromise these decoy resources are studied during and after an attack to keep an eye on new exploitation techniques. Such analysis may be used to further tighten security of the actual network being protected by the data's. Data forwarding can also direct an attacker's attention away from legitimate servers. A user encourages attackers to spend their time and energy on the decoy server while distracting their attention from the data on the real server. Similar to a server, a user is a network set up with intentional vulnerabilities. Its purpose is also to invite attacks so that the attacker's methods can be studied and that information can be used to increase network security.

### SECURITY GOAL AND THREAT MODULE

DLD provider from gaining knowledge of sensitive data during the detection process, we need to set up a privacy goal that is complementary to the security goal above. We model the DLD provider as a semi-honest adversary, who follows our protocol to carry out the operations, but may attempt to gain knowledge about the sensitive data of the data owner. Our privacy goal is defined as follows. The DLD provider is given digests of sensitive data from the data owner and the content of network traffic to be examined. Data present a privacy-preserving DLD model with a new fuzzy fingerprint mechanism to improve the data protection against semi-honest DLD provider. We generate digests of sensitive data through a one-way function, and then hide the sensitive values among other non-sensitive values via function. The sensitive data is sent by a legitimate user intended for legitimate purposes. In this paper, we assume that the data owner is aware of legitimate data transfers and permits such transfers. So the data owner can tell whether a piece of sensitive data in the network traffic is a leak using legitimate data transfer policies.

### EXPERIMENTAL EVALUATION MODULE

Data-leak detection solution can be outsourced and be deployed in service provider's host that enhances data privacy during data-leak detection operations. Our approach is based on a fast and practical two-way computation on the sensitive data (SSN records, classified documents, sensitive emails, etc.), which are policy based filter and base encryption. It enables the data owner to securely delegate the content-inspection task to DLD providers without exposing the sensitive data. Using our detection method, the DLD provider, who is modeled as fully honest gain knowledge about the sensitive data from either the released digests, or the content being inspected. Using fuzzy techniques, an Internet service provider (ISP) can perform detection on its customers' traffic securely and provide data leak detection as an add-on service for its customers. We set up a networking environment in Virtual Box, and make a scenario where the sensitive data is leaked from a local network to the Internet. Multiple users' hosts are put in the local network, which connect to the Internet via a gateway. Multiple servers and an attacker-controlled host are put on the Internet side. The gateway dumps the network traffic and sends it to a DLD server/provider. Using the sensitive data identity defined by the users in the local network, the DLD server performs off-line data-leak detection.

### POLICY FILTER MODULE

Domain-guard Data Loss Prevention (DLP) solution, a comprehensive data leakage protection solution, helps organizations effectively manage the threats and risks to organizations' information, and safeguard what organizations value. It carries rich and matchless features to help organizations against data security threat, and prevent accidental and malicious data leakage. It effectively controls and protects sensitive data as it leaves, moves and transforms through your IT environment. With it, organizations can

centrally monitor how the data is being used, records data operations and prevents data leaking out via various channels such as removable storage devices, email, instant messaging, and printing. IP-guard enables you to get a grip on your critical information, avoid costly loss and safeguard intellectual property. Distributor data will contain some additional hidden information, which is added to the real info for law enforcements. (Key information) Read and write permission for folder and file level on webs server Data base columns are encrypt using AES (Advanced encryption system).Request and response filter are applied in lifecycle of the web services.

**DOMAIN ADD ON-MODULE**

For every individual domain able to add this filter option as Add on service. After enabling the Add on service, Agent cannot access the local web service related files and resource available under a domain will be accessed via web server request only. Distributor and other allowed end user will have separate UI for accessing web service resources. For all kind of data transfer the communication stream should offer confidential data on the servers, for new and previous domain. Policy filter entries are applied while initiate the service begins. General request and response are filtered based on IP identity / and preventing manual file access with the public place of resources of a domain.

**SYSTEM DESIGN**

**DATA FLOW DIAGRAM**

A two -dimensional diagram that explains how data is processed and transferred in a system. The graphical depiction identifies each source of data and how it interacts with other data sources to reach a common output. Individuals seeking to draft a data flow diagram must identify external inputs and outputs, determine how the inputs and outputs relate to each other, and explain with graphics how these connections relate and what they result in. This type of diagram helps business development and design teams visualize how data is processed and identify or improve certain aspects.

**LEVEL 0**

The Level 0 DFD shows how the system is divided into 'sub-systems' (processes), each of which deals with one or more of the data flows to or from an external agent, and which together provide all of the functionality of the system as a whole. It also identifies internal data stores that must be present in order for the system to do its job, and shows the flow of data between the various parts of the system.

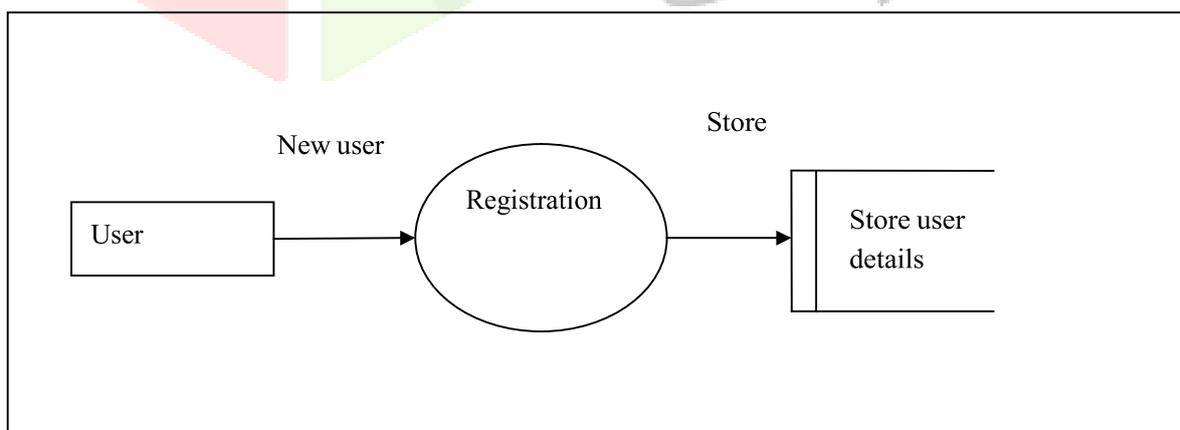


FIG : 7.1 Data Flow Diagram Level

**DFD LEVEL-1**

The next stage is to create the Level 1 Data Flow Diagram. This highlights the main functions carried out by the system. As a rule, to describe the system was using between two and seven functions - two being a simple system and seven being a complicated system. This enables us to keep the model manageable on screen or paper.

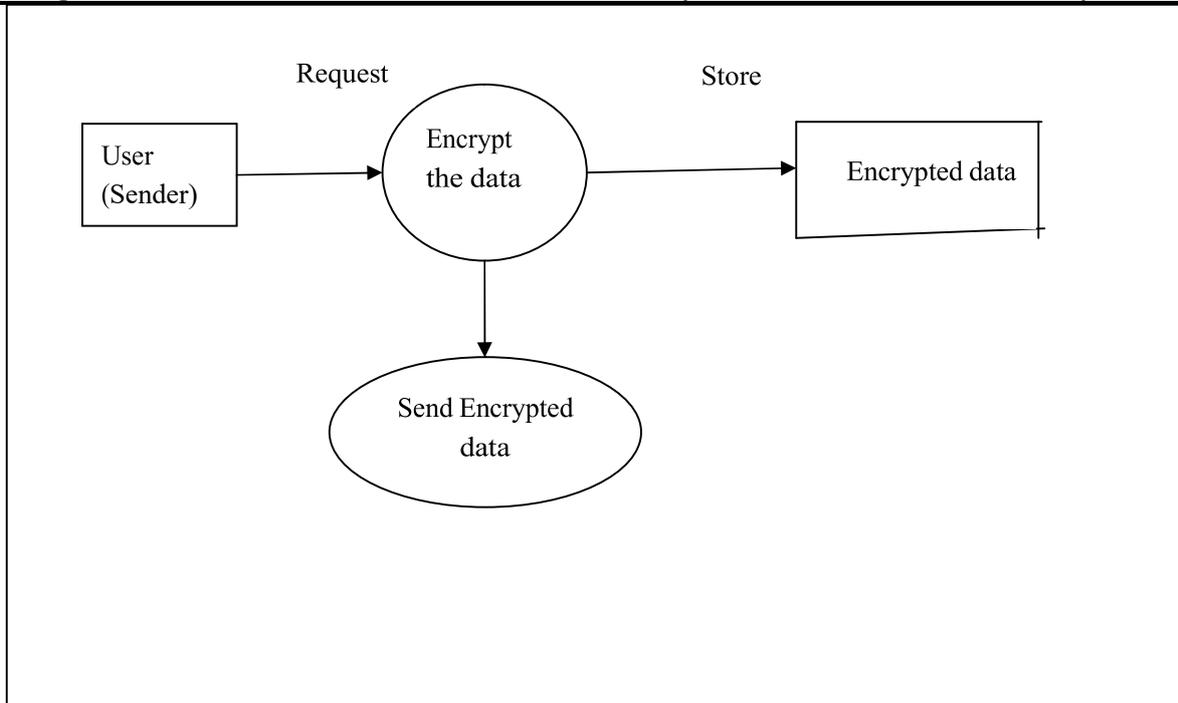


FIG : 7.2 Data Flow Diagram Level 1

**DFD LEVEL-2**

A Data Flow Diagram (DFD) tracks processes and their data paths within the business or system boundary under investigation. A DFD defines each domain boundary and illustrates the logical movement and transformation of data within the defined boundary. The diagram shows 'what' input data enters the domain, 'what' logical processes the domain applies to that data, and 'what' output data leaves the domain. Essentially, a DFD is a tool for process modeling and one of the oldest

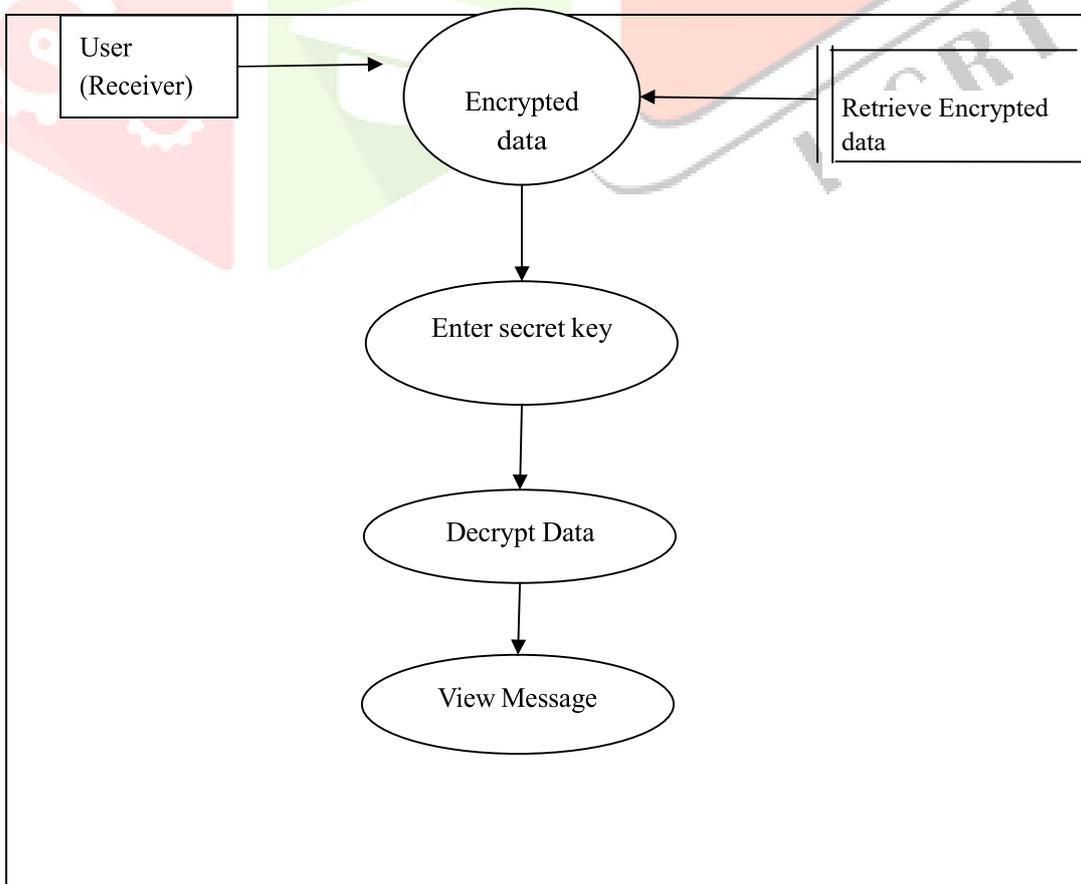


FIG : 7.3 Data Flow Diagram Level 2

TOTAL TABLE: 5

Table Name: admin

Field	Type	Description
username	varchar(30)	Username
password	varchar(50)	Password

Table Name: enquiry

Field	Type	Description
id	int(11)	ID
name	varchar(25)	Name
username	varchar(30)	Username
email	varchar(50)	E-mail
comments	Text	Comments
post_date	Date	Posted Date

Table Name: tblencrypt

Field	Type	Description
Id	int(11)	ID
Uname	varchar(30)	Username
enc_txt	longtext	Encrypted Text
Secret	varchar(50)	Secret Key
created_date	Date	Created Date

Table Name: text palette

Field	Type	Description
id	int(11)	ID
txt_char	varchar(5)	Character
img1	varchar(30)	Image1
img2	varchar(30)	Image2
img3	varchar(30)	Image3

Table Name: user details

Field	Type	Description
id	int(11)	ID
name	varchar(25)	Name
email	varchar(50)	E-mail
contact	bigint(20)	Contact No.
username	varchar(50)	Username
password	varchar(50)	Password

## 7.1 SYSTEM TESTING

The development process involves various types of testing. Each test type addresses a specific testing requirement. The most common types of testing involved in the development process are:

- Unit Test.
- System Test
- Integration Test
- Functional Test
- Performance Test
- While box Test
- Black box Test.

## UNIT TESTING

The first test in the development process is the unit test. The source code is normally divided into modules, which in turn are divided into smaller units called units. These units have specific behavior. The test done on these units of code is called unit test. Unit test depends upon the language on which the project is developed. Unit tests ensure that each unique path of the project performs accurately to the documented specifications and contains clearly defined inputs and expected results. Functional and reliability testing in an Engineering environment. Producing tests for the behavior of components (nodes and vertices) of a product to ensure their correct behavior prior to system integration.

## SYSTEM TESTING

Several modules constitute a project. If the project is long-term project, several developers write the modules. Once all the modules are integrated, several errors may arise. The testing done at this stage is called system test. System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points. Testing a specific hardware/software installation. This is typically performed on a COTS (commercial off the shelf) system or any other system comprised of disparate parts where custom configurations and/or unique installations are the norm.

## FUNCTIONAL TESTING

Functional test can be defined as testing two or more modules together with the intent of finding defects, demonstrating that defects are not present, verifying that the module performs its intended functions as stated in the specification and establishing confidence that a program does what it is supposed to do.

## INTEGRATION TESTING

Testing in which modules are combined and tested as a group. Modules are typically code modules, individual applications, source and destination applications on a network, etc. Integration Testing follows unit testing and precedes system testing. Testing after the product is code complete. Betas are often widely distributed or even distributed to the public at large in hopes that they will buy the final product when it is released.

## WHITE BOX TESTING

Testing based on an analysis of internal workings and structure of a piece of software. This testing can be done using the percentage value of load and energy. The tester should know what exactly is done in the internal program. Includes techniques such as Branch Testing and Path Testing. Also known as Structural Testing and Glass Box Testing.

## BLACK BOX TESTING

Testing without knowledge of the internal workings of the item being tested. Tests are usually functional. This testing can be done by the user who has no knowledge of how the shortest path is found.

## CONCLUSION

The proposed system of the data leakage system is improved in two ways to prevent information leakages. In the new system probability of leakage prevention rather than detection is increased due to high security of hosting environment. As an Add on component for a domain, the new implementation encapsulates high security in terms of privacy of information and leakages prevention. Hypertext based request are intercepted and filtered then forwarded to right resource for processing the request. All the response content retired from the database are encrypted using Attribute based encryption, which is implemented in Base encryption system, exists as one of major component at outsource provider. As the result of the new system, probability of information leakage is reduced and data identification is increased.

### FUTURE ENHANCEMENT

Our future work includes the investigation of agent guilt models that capture leakage scenarios that are not studied in this project. For example, what is the appropriate model for cases where agents can collude and identify fake tuples ? A preliminary discussion of such a model is available in. Another open problem is the extension of our allocation strategies so that they can handle agent requests in an online fashion (the presented strategies assume that there is a fixed set of agents with requests known in advance).Any application does not end with a single version. It can be improved to include new features. Our application is no different from this. The future enhancements that can be made to Data Leakage Detection are:

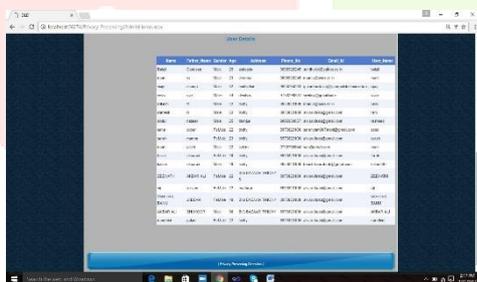
- Providing support for other file formats.
- Creation of a web based UI for execution of the application.
- Improving the detection process based on user requirements.

### 10.2 SCREEN SHOT

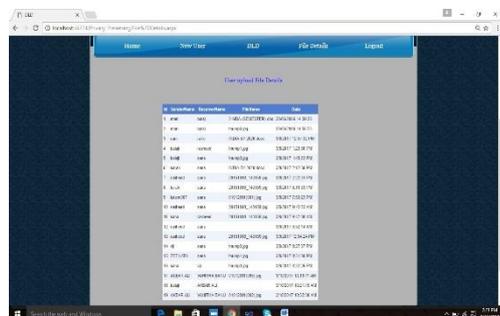
- ADMIN HOME



- USER DETAILS



- DLD DETAILS



- USER LOGIN PAGE

## REFERENCE

- Advanced .NET Remoting 2nd Edition (Ingo Rammer and Mario Szpuszta, Apress, March 2005)
- .NET to .NET Migration Handbook (Christian Nagel et al, Wrox, January 2003)
- OF .NET (Christian Nagel et al, Wrox, September 2001)
- Data-Centric .NET Programming (Christian Nagel et al, Wrox, December 2001)
- Professional .NET Network Programming 2nd Edition (Christian Nagel et al, Wrox, September 2004)
- Professional .NET (Christian Nagel et al, Wrox, March 2002)
- Professional .NET Web Services (Christian Nagel et al, Wrox, December 2001)
- The Joy of .NET Programming: A Beginner's Guide – by Alan Forbes
- .NET & MySQL Novice to Ninja – by Kevin Yank
- Head First .NET & MySQL – by Lynn Beighley& Michael Morrison
- .NET: A Beginner's Guide – by VikramVaswani
- Murach's .NET & MySQL – by Joel Murach& Ray Harris
- Learning .NET, MySQL, JavaScript, and CSS: A Step-by-Step Guide to Creating
- Dynamic Websites – by Robin Nixon
- Programming .NET – by Kevin Tatroe, Peter MacIntyre&RasmusLerdorf
- "Foreword By: Michael Bourque"
- .NET & MySQL Web Development – by Luke Welling & Laura Thompson
- .NET & MySQL: The Missing Manual – by Brett McLaughlin.

