Information Leak Detection and Prevention

Deepika. S. Patil[1], **Dr. Sujata Terdal**[2]

[1]Department of Computer Science & Engineering M.Tech.(CSE),
Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India

[2]Associate Professor , Department of Computer Science & Engineering
Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India

Abstract: ***Information leakage is one of the major challenges in end to end data exchange through Open Source light weight services like MQTT and Redis. These brokers are important part of information sharing in VPN and fog computing. Generally such brokers do not offer session management, access control and data security. The goal of this work is to extend the capabilities of such information brokers by implementing security layer on top of data service to prevent information leakage. This particular work consider Redis broker as a test case and offers security layer to prevent information leakage through Redis broker. We also design a UI system to demonstrate the capabilities of the system with and without the security. Our results shows that adding the security extension does not increase the overhead by much in terms of system resources and latency. We also extend the key-value based system to be able to store binary image data which is also stored in the encrypted pattern.***

Keywords: Redis, visual studio, Authentication Service, Encryption to Value data, Security to persistent data (one that is stored permanently),Security to blob data (multimedia data for images).

1. Introduction

Data loss, which means a loss of data that occur on any device that stores data. It is a problem for anyone that uses a computer. Data loss happens when data may be physically or logically removed from the organization either intentionally or unintentionally. The data loss has become a biggest problem in organization today where the organizations are in responsibility to overcome this problem. Data Leakage is an incident when the confidentiality of information has been compromised. It refers to an unauthorized transmission of data from within an organization to an external destination. The data that is leaked out can either be private in nature and are deemed confidential whereas Data Loss is loss of data due to deletion, system crash etc. Totally both the term can be referred as data breach, has been one of the biggest fears that organization face today. Data Loss/Leakage Prevention (DLP) is a computer security term which is used to identify, monitor, and protect data in use, data in motion, and data at rest [1]. DLP is sued to identify sensitive content by using deep content analysis to per inside files and with the use if network communications. DLP is mainly designed to protect information assets in minimal interference in business processes. It also enforces protective controls to prevent unwanted incidents. DLP can also be used to reduce risk, and to improve data management practices and even lower compliance cost. Systems are designed to detect and prevent unauthorized use and transmission of confidential information. Vendors refer to the term as Data Leak Prevention, Information Leak Detection and Prevention (ILDP), Information Leak Prevention (ILP), Content Monitoring and Filtering(CMF), Information Protection and Control (IPC) or Extrusion Prevention System by analogy to Intrusion-prevention system.

2. Related Work

”Data Leakage Detection In Networks”[1], the field of business, the owner of any organization, company or business firm having some crucial data may need to share it with third-parties. These trusted third-parties may use this data for their own benefit causing reputational and monetary damage to the owner‟s company. ” Data Leakage Detection” [2], this propose contains the results of implementation of Data Leakage Detection Model. Currently watermarking technology is being used for the data protection. But this technology doesn‟t provide the complete security against date leakage.“PDT-BI: Proactive Detection Technology based on the Biometric Information for Preventing Internal Information Leakage “[3], In this paper, he propose the proactive detection technique that can detect malicious behaviors of aninsider by biometric information. Additionally, we compare proposed technique with other techniques that use biometric information in the security field. “Information Leakage Detection in Distributed Systems using Software Agents”[4], this paper, propose a mobile agent-based approach to automate the process of detecting and coloring receptive hosts‟ filesystems and monitoring the colored ,filesystem for instances of potential information leakage. Implementation details and execution results are included to illustrate the merits of the proposed approach.“Data Leakage Detection”[5], They propose data allocation strategies (across the agents) that improve the probability of identifying leakages. These methods do not rely on alterations of the released data (e.g., watermarks). In some cases we can also inject “realistic but fake” data records to further improve our chances of detecting leakage and identifying the guilty party.“Data Leakage Detection: A Survey”[6], This paper includes brief idea about data leakage detection and a methodology to detect the data leakage persons.“Development of Data leakage Detection Using Data Allocation Strategies”[7], they implement and analyze a guilt model that detects the agents using allocation strategies without modifying the original data. ”Data Leakage Detection with security"[8], Perturbation is a very useful technique where the data is modified and made „less sensitive´ before being handed to agents. For example, one can add random noise to certain attributes, or one can replace exact values by ranges. However, in some cases it is important not to alter the original distributor‟s data."Data Leakage Detection Using Encrypted Fake Objects"[9], In this paper, we present unobtrusive techniques for detecting data leakage and assessing the "guilt" of agents."Data Allocation Strategies In Data Leakage Detection"[10], In this paper, they implement and analyze a guilt model that detects the agents using allocation strategies without modifying the original data. The guilty agent is one who leaks a portion of distributed data.

3. Problem Statement

Because Redis is a Key-Value pair based database, it is quite popular open source light weight information sharing system which is predominantly broker-agent model based.

The data communication involves entities to store data of any data type through a data agent to data broker(server). However such a system does not enforce any session or access control mechanism. Hence an intruder can easily get access to the data if he has the partial information about the data definition.

Any intruder acquiring a knowledge about the keys can get the values. Also as no authentication is supported in Redis system, anybody can acquire the knowledge of the data. In order to secure NSQL database and to provide highest data security, we design a Redis Client which provides Authentication, Authorization and Encryption services for both text and multimedia-binary data.

4. Proposed Work

In figure [1] Proposed work contains following major modules. In this subsection we present a brief overview of the modules which are:

a) **Authentication and Authorization:** We first create a universal key which can hold all username password pairs. Any user wanting to access the system needs to first register with the system.

b) **Command Restriction:** We restrict the commands to **Get, Set** commands and system does not support any other commands. This is done to show the capability of restricting command sets. Also we provide high level command for storing and retrieving images. The images are also secured.

c) **Encryption Services:** Every variable and their value is encrypted and is packed in a key which is derived from the username. Therefore every user has a single data node or key in the Redis database that contains all other keys in the encrypted form. Symmetric key cryptography with AES is used for encryption service. In the symmetric key cryptography, decryption is performed using the same key as that of encryption. The keys are generated from the salt of password created at the time of login. Images are first converted into binary data and then are encrypted using AES using block cipher AES technique. Encrypted images are stored as binary data in the Redis database. Images of a particular user are packed in an image key corresponding to that user. Therefore the database contains an image node related to every user.

d) **Network Security:** Encryption is always performed at the client side. Therefore the data that is propagated in the network is secured data. Therefore chances of the data being acquired by intruder by network packet hacking is minimized.

e) **Data Persistence and Persistent data security:** Persistent data is one which is saved in the Redis database. This data should be available at any time when user comes back. This is done by packing the user variable which contains variable of current and previous sessions and their values being packed in the user data node in the encrypted form where the encryption method adopted is as explained above.

4.1 Architecture

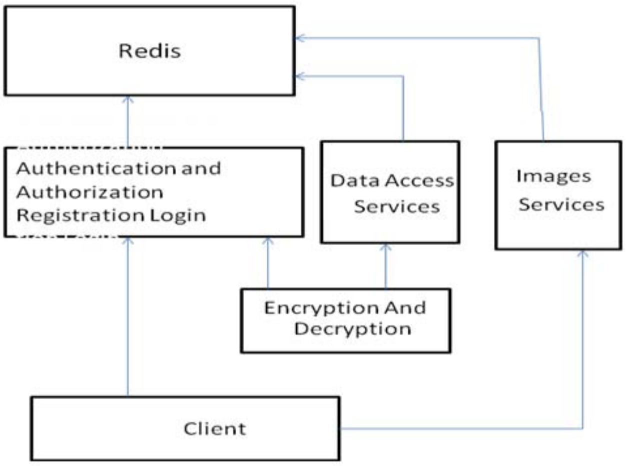

Redis
Authentication and Authorization Registration Login
Data Access Services
Images Services
Encryption And Decryption
Client

Figure 1: Architecture of proposed system.

In figure[2] first part execution starts with authentication. This data flow diagram shows the overall execution of proposed system. Using the user name and password have to register will give authentication service as admin is performed. Authentication services is connected the client and login without encryption decryption will enter the string ,variables.It flow the encrypted data then in remote server values will be packing and unpacking with redis server, Again login with encryption decryption will enter the string ,variables and also images ,It flow the encrypted data then in remote server values will be packing and unpacking with redis server. It security will happen with information leak detect and prevent the data.

4.2 System Data Flowchart Diagram

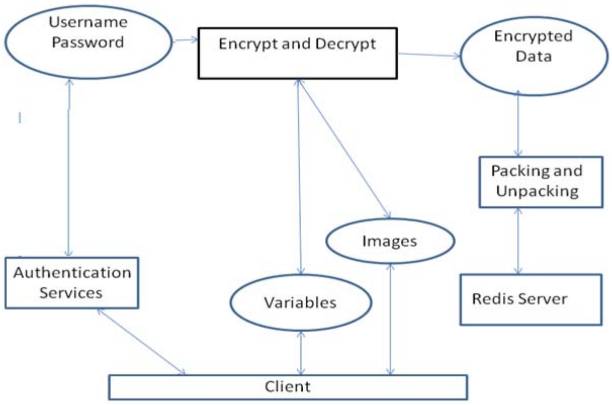
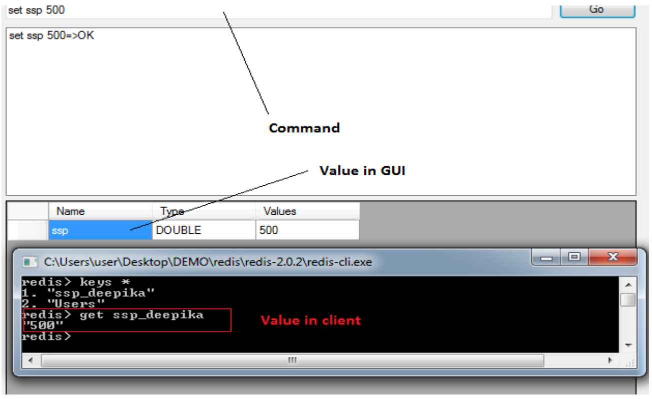


Figure 2: System Data Flowchart diagram

5. Results

It gives the suspected information has to be leaked in box highlighted with red color. It uses the get keywords that are stored in client system. We first evaluate the threat perception by logging into user‟s session without encryption and decryption. This mode offers Session management without any security.



set ssp 500

set ssp 500=>OK

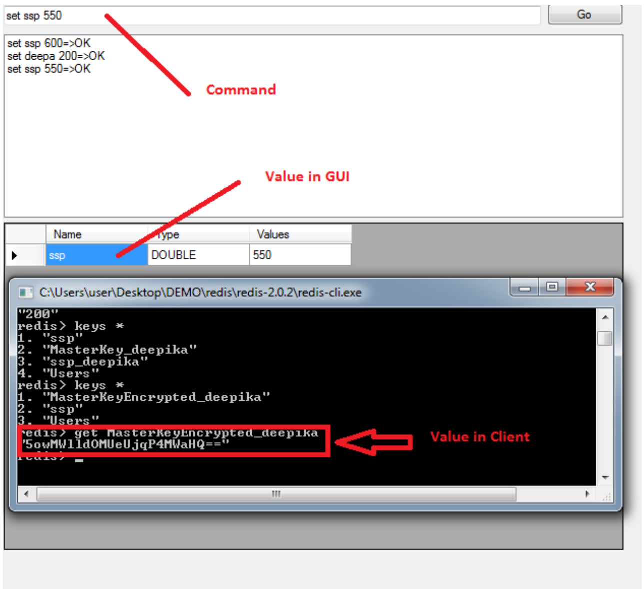
Name	Type	Values
ssp	DOUBLE	500

C:\Users\user\Desktop\DEMO\redis\redis-2.0.2\redis-cli.exe

```
redis> keys *
1. "ssp_deepika"
2. "Users"
redis> get ssp_deepika
"500"
redis>
```

Figure 3: Without Encrypted Decryption

Following figure shows the enter command values then go, it shows the value in GUI with encrypted decrypted it will not leak the data, its has to be security system.



set ssp 550

set ssp 600=>OK
set deepa 200=>OK
set ssp 550=>OK

Name	Type	Values
ssp	DOUBLE	550

C:\Users\user\Desktop\DEMO\redis\redis-2.0.2\redis-cli.exe

```
"200"
redis> keys *
1. "ssp"
2. "MasterKey_deepika"
3. "ssp_deepika"
4. "Users"
redis> keys *
1. "MasterKeyEncrypted_deepika"
2. "ssp"
3. "Users"
redis> get MasterKeyEncrypted_deepika
"5owMWlldOMUeUjqP4MWaHQ=="
redis> _
```

Figure 4: With Encrypted Decryption

Statistical Manager extracts length and total length statistics of the packets as well as time bound standard deviation of the packets. It also gives the Time with encryption and without encryption, Storing Length and size.This figure as flows.

Time with Encryption	Time without Encryption	Length with Encryption	Length without Encryption
5.4546	1.7344	44	25
5.5322	4.5925	64	35
5.6451	4.347	88	45
5.7392	4.5045	88	55
5.7972	4.2636	108	65
5.496	4.2493	128	75
5.7883	2.277	152	85
3.3725	2.2882	172	95
5.6915	4.4144	192	105
5.5853	4.7723	192	116
2.7308	4.5652	216	127
3.4002	2.7625	236	138

Figure 5: Captured Statistic

The Figure[6] graph shows that for less amount of data, encryption and security increases time complexity. But when the number of variable increases, time complexity is neutralized. This proves that adding security to prevent information leakage does not add extra overhead as far as system and network resources are concerned.

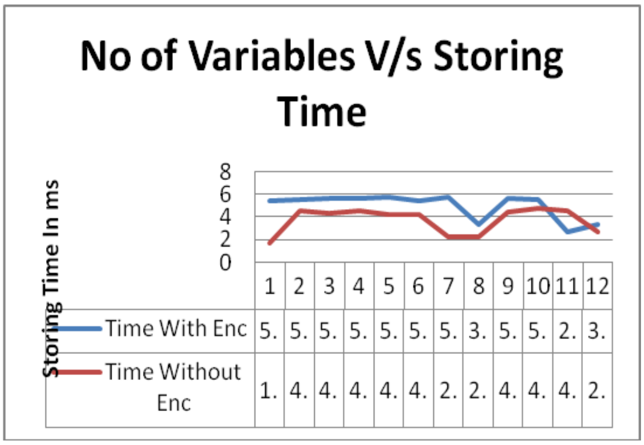
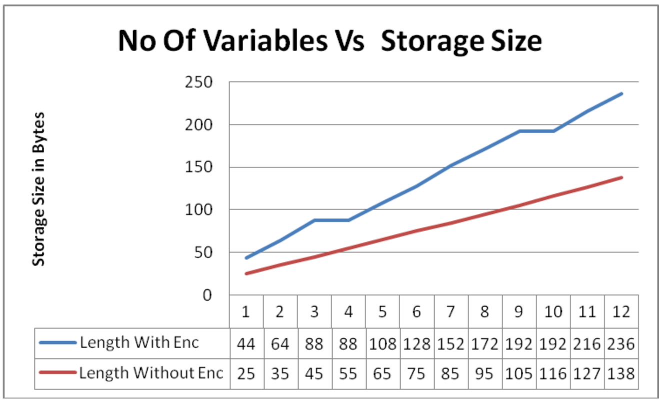


Figure 6: Network Analysis with Time

The figure[7] graph shows that space complexity for both secured information sharing and insecured information sharing is O(n) where n is the variable length. Space requirement for secured information is about 50% more than the insecured information. Therefore only sensitive data that needs to be protected from the intruders must be used with encryption. Other information could be protected with access and session control.



	1	2	3	4	5	6	7	8	9	10	11	12
Length With Enc	44	64	88	88	108	128	152	172	192	192	216	236
Length Without Enc	25	35	45	55	65	75	85	95	105	116	127	138

Figure 7: Network Analysis with Size

6. Conclusion

Information exchange in network is insecured, especially if the communication involves open source protocols like NoSql or MQTT. In such cases, data is released in a shared buffer. Therefore an intruder who is aware of the data definition and the meta data can easily get the access of the data. In order to prevent such undesirable information leakage, various sches are being suggested. Many solution improves the inherent drawback by using public key cryptography. In public key cryptography, end entities exchange a key sacredly and the key pair is used to encrypt and decrypt data. But due to extra overhead in number of packets, it is not suitable for open stack information sharing like NoSQL. Therefore in this work we have used Redis which is a popular open source light weight network data structure storage and enhanced the security by offering access control, session management, authorization and authentication and symmetric key cryptography. We also show that the proposed system prevents the leakage of data. Results shows that the system does not add communication or time overhead for the security enhancement. However symmetric key cryptography adds extra space complexity. We have also extended the framework by offering an image storage in Redis database which does not support multimedia storage. By adding both image encryption and decryption, we have also extended the security to multimedia data.

7. Future Scope

Our framework can be utilized as a future scope to build digital steganography based system where the information is embedded in image and encrypted image is saved in the database instead of the plain image. Such a technique will improve the information security to greater degree.

References

[1] " Data Leakage Detection In Networks" Ms. Aishwarya Potdar1, Ms. Rutuja Phalke2, Ms. Monica Adsul3, Ms.Prachi Gholap4 B.E, Department of Computer Engineering, KJCOEMR, Pune University, Pune, India1,2,3,4.

[2] "Data Leakage Detection" Sandip A. Kale1, Prof. S.V.Kulkarni2 Department Of CSE, MIT College of Engg, Aurangabad, Dr.B.A.M.University, Aurangabad (M.S), India1,2

[3] "PDT-BI: Proactive Detection Technology based on the Biometric Information for Preventing Internal Information Leakage "Seung-Hyun Lee1, Min-Woo Park1, Jung-Ho Eom2* and Tai-Myoung Chung3 1Dept. Electrical and Computer Engineering, Sungkyunkwan Univ, Suwon, Republic of Korea.

[4] "Information Leakage Detection in Distributed Systems using Software Agents" Yung-Chuan Lee, Stephen Bishop, Hamed Okhravi and Shahram Rahimi.

[5] "Data Leakage Detection" *NIKHIL CHAWARE 1, PRACHI BAPAT 2, RITUJA KAD 3, ARCHANA JADHAV 4, PROF.S.M.SANGVE 5 1,2,3,4Student Computer Department,ZES's DCOER,Pune 5Assistant Professor and HOD Computer Department,ZES's DCOER,Pune.*

[6] "Data Leakage Detection: A Survey" Sandip A. Kale C1, Prof.S.V. Kulkarni C21(Department Of Computer Sci. & Engg,MITCOEngg,Dr.B.A.M.University, Aurangabad(M.S),India).

[7] "Development of Data leakage Detection Using Data Allocation Strategies" Rudragouda G Patil Dept of CSE, The Oxford College of Engg, Bangalore.

[8] "Data Leakage Detection**"** Rekha Jadhav G.H.Raisoni Institute of Engg. And Technology.

[9] "Data Leakage Detection Using Encrypted Fake Objects" Anusha.Koneru#1 , G.Siva Nageswara Rao#2, J.Venkata Rao#3 #1M.Tech(CSE) Student Department of CSE, K L University Guntur, Andhra Pradesh, India.

[10] "Data Allocation Strategies In Data Leakage Detection" Unnati Kavali Tejal Abhang Mr. Vaibhav Narawade Student of P.V.P.P.C.O.E. Student of P.V.P.P.C.O.E. Associate Professor P.V.P.P.C.O.E., Mumbai.

[11] **"redis***.io/topics/***security"**

[12] *www.tutorialspoint.com/***redis/redis_security***.htm.*