INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

TECHNIQUES TO SLICE THE PASSWORD FOR PRESERVING THE PRIVACY IN CLOUD COMPUTING

MR. FAIZAN I KHANDWANI[1], PROF. ASHOK P KANKALE[2]

Department of Computer Science and Engineering,, Rajarshi Shahu College of Engineering, Buldana.

Accepted Date: 05/03/2015; Published Date: 01/05/2015

Abstract: Cloud computing is a computing prototype, where a large pool of systems are connected in public or private networks, to provide dynamically scalable infrastructure for application, data and file storage. A pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption. Cloud computing is an efficient solution for the fastest and easiest storage and retrieval of data. The main concern in cloud computing is security. Access to computer systems is most often based on the use of alphanumeric passwords. However, users have difficulty remembering a password that is long and random-appearing. Instead, they create small, easy, and insecure passwords. Graphical passwords have been designed to try to make passwords more memorable and easier for people to use and, therefore, more secure.

Keywords: Encryption, Keys, Slicing of Data, Cloud Data protection, Data recovery, Graphical password, Pass points.

Corresponding Author: MR. FAIZAN I KHANDWANI

Access Online On:

www.ijpret.com

How to Cite This Article:

Faizan I. Khandwani, IJPRET, 2015; Volume 3 (9): 449-455

PAPER-QR CODE

INTRODUCTION

Cloud computing is a model for enabling suitable, ubiquitous, on-demand network access to a common pool of configurable computing resources that can be rapidly provisioned and released with least management effort or service provider interaction. For providing security to their data user uses the password, but it is alphanumeric password. It contains uppercase and lowercase letters, numbers and special symbols. It can be easily hacked by intruder. For preventing the data we can use the graphical password concept. It has many advantages as it is easy to remember, nobody can guess the password. Using a graphical password, user needs to click on images in place of alphanumeric characters. Graphical password system has the Pass Points. It: (i) allows any image to be used and (ii) does not need artificial predefined click regions with well-marked boundaries – a password can be any randomly selected series of points in the image [5].

In cryptography there are two types of keys as public and private keys. Keyword based search is one of the popular ways to selectively identify and retrieve data files instead of retrieving all the files. Keywords are parts of file name or phrases used in the file which will help us to find the exact data file at the time of retrieval if you don't remember the exact keyword. There are many keyword searching methods [6]. Slicing partitions the data both horizontally and vertically. It preserves better data utility than generalization and can be used for membership disclosure protection. An additional significant benefit of slicing is that it can handle high-dimensional data [11].

I. 2. Challenges IN Cloud Computing

Despite of its rising influence, concerns regarding cloud computing still remains. But the benefits overshadow the drawbacks and the model is worth exploring. Some common challenges are as follows:

A. *2.1 Data Security*

Data Security is a vital element that warrants scrutiny. Enterprises are hesitant to purchase an assurance of business data security from the vendors. They fear losing data to competition and the data privacy of consumers. In many instances, the actual storage location is not disclosed, adding onto the security concerns of enterprises. In the existing models, firewalls across data centers (owned by enterprises) protect this sensitive information. In the cloud model, Service providers are responsible for maintaining data security and enterprises would have to rely on them [12].

B. *2.2 Data Recovery and Availability*

All business applications have Service level agreements that are rigorously followed. Operational teams play a vital role in the management of service level agreements and runtime governance of applications. In production environments, operational teams support the data replication, system recovery and maintenance, disaster recovery, performance management.

If, any of the mentioned services is under-served by a cloud provider, the damage & impact could be severe [12].

C. *2.3 Management Capabilities*

Despite of multiple cloud providers, the management of platform and infrastructure is still in its early days. Features like “Auto-scaling” are a crucial necessity for many enterprises. There is huge potential to improve on the scalability and load balancing features provided today [12].

2.4 Regulatory and Compliances Restrictions

In some of the European countries, Government regulations do not allow customer's personal information and other sensitive information to be physically located outside the country. In order to meet such requirements, cloud providers need to setup a data center or a storage site solely within the country to fulfill the regulations. Having such an infrastructure may not always be realistic and is a big challenge for cloud providers. Cloud computing interfaces service suppliers with multiple groups of service consumers. Cloud services demands expertise in distributed services, procurement, risk assessment and service negotiation areas that many enterprises are only reasonably equipped to handle [12].

1. Graphical password:-

Lots of authentication techniques have been proposed in the recent times that are based upon graphical methods. Text based passwords are most usually used for authentication; however, they are highly prone to several kinds of attacks. Graphical techniques are coming up as a smart substitute to the conventional methods of authentication. Here we have proposed a graphical method of authentication that employs graphical coordinates along with the time interval between successive clicks. The user needs to recall the coordinates and the time interval of the successive clicks. This leads to the inclusion of the advantages of the recent graphical methods along with the extra security achieved through the use of time interval. The proposed method has a much higher password space than the other existing graphical authentication schemes. The scheme is robust, secure and very convenient to use. In view of the shortcomings of the traditional approach to authentication, Graphical techniques are gaining importance. A graphical password is an authentication system in which the user has to work with images, either selecting them or creating them. Ex. the user may select some points from the image which is stored as the graphical password in the database. If someone needs to store the file or retrieve the file stored in the system he/she should enter the correct graphical password for access to the file. A graphical password is easier to remember than complex text-based password [6].

D. *3.1 Slicing*

The concept of slicing is to break the association cross columns, but to maintain the association within each column. This reduces the dimensionality of the data and preserves better utility than generalization and bucketization. Slicing protects privacy because it breaks the associations between uncorrelated attributes, which are infrequent and thus identifying. Slicing preserves utility because it groups highly-correlated attributes together, and preserves the correlations between such attributes. The key function that slicing provides privacy protection is that the slicing process ensures that for any tuple, there are generally multiple matching buckets. Given a tuple t =<v1, v2, . . . , vc>, where c is the number of columns, a bucket is a matching bucket for t if and only if for each i (1 ≤ i ≤ c), vi appears at least once in the ith column of the bucket. Any bucket that contains the original tuple is a matching bucket. At the same time, a matching bucket can be due to containing other tuples each of which contains some but not all vi's. Two popular anonymization techniques are generalization and bucketization. Bucketization [1,4,8] first partitions tuples in the table into buckets and then separates the quasi-identifiers with the sensitive attribute by randomly permuting the sensitive attribute values in each bucket. The anonymized data consists of a set of buckets with permuted sensitive attribute values. In particular, bucketization has been used for anonymizing high-dimensional data [9]. Generalization [2,10,14] replaces a value with a "less-specific but semantically consistent" value. Three types of encoding schemes have been proposed for generalization: global recoding, regional recoding, and local recoding. Global recoding has the property that multiple occurrences of the same value are always replaced by the same generalized value. Regional record [13] is also called Multi-dimensional recoding (the Mondrian algorithm) which partitions the domain space into non-intersect regions and data points in the same region are represented by the region they are in. Local recoding does not have the above constraints and allows different occurrences of the same value to be generalized differently. By using the concept of slicing we protect our micro data.

1.1.1 Bucketization

Slicing has many advantages over bucketization. First, by partitioning attributes into more than two columns, slicing can be used to avoid membership disclosure. Our experimental evaluation on a real data set shows that bucketization does not prevent membership disclosure. Second, unlike bucketization, which requires a clear separation of QI attributes and the sensitive attribute, slicing can be used without such a separation. For data set such as the census data, one often cannot clearly separate QIs from SAs because there is no single external public database that one can use to determine which attributes the adversary already knows. Slicing can be useful for such data. Finally, by allowing a column to contain both some QI attributes and the sensitive attribute, attribute correlations between the sensitive attribute and the QI attributes are preserved. For example Zip code and Disease form one column, enabling inferences about their correlations. Attribute correlations are important utility in data publishing. For workloads that consider attributes in olation, one can simply publish two tables, one containing all QI attributes and one containing the sensitive attribute. Modeling adversary's background knowledge is most important privacy model, such as k-anonymity, l-diversity, confidence bounding, and t-closeness, assuming the adversary has very limited background knowledge. They assume that the adversary's background knowledge is limited to knowing the quasi-identifier. However, recent work has shown the importance of integrating an adversary's background knowledge in privacy quantification. A robust privacy notion must take background knowledge into consideration, since an adversary can easily learn background knowledge from various sources. [3]

3.1.2 Bottom-Up Generalization

Algorithm A represents our bottom-up generalization process. In the ith iteration, we generalize R by the "best" generalization Gbest according to the IP metric. This algorithm makes no claim on efficiency because Line 2 and 3 requires computing IP (G) for all candidate generalizations G. Consider a candidate generalization G: fcg! p in an iteration. jRcj and freq(Rc; cls) can be maintained after each iteration. jRpj and freq(Rp; cls) can be obtained by aggregating jRcj and freq(Rc; cls). Therefore, I(G) can be easily computed, i.e., without accessing vids. In fact, any metric on a single attribute (plus the class label) can be computed this way. A (V ID) is available as a result of applying the previous generalization. Computing AG (V ID), however, depends on the "effect" of G, which is only available after applying G, and requires accessing vids. This is a new challenge to scalability. Our insight is that most generalizations G do not affect A(V ID), therefore, AG(V ID) = A(V ID). If a generalization G fails to generalize all anonymity vids, G will not affect A (V ID). For such G, P (G) = 0 and IP (G) = 1, and our metric does not need AG (V ID). Thus, we can concentrate on “critical generalizations”.

Algorithm A the bottom-up generalization

```
while R does not satisfy the anonymity requirement do
    for all generalization G do
        compute IP(G);
    end for;
    find the best generalization Gbest;
    generalize R by Gbest;
end while;
output R;[16]
```

II. 4. CONCLUSION

This work is focused on the usability of Pass Points, but its security is also of equal importance. Although graphical users always take more time to input their passwords than alphanumeric users, still there are evidences that with continuous use, graphical passwords can be entered more quickly[5].This work has several directions for future research. Here, we consider slicing where each attribute is in exactly one column. An expansion would be of overlapping slicing, which duplicates an attribute in more than one column. It might provide better data utility, but the privacy implications need to be carefully studied and understand. The tradeoff between privacy and utility is another interesting point to be considered [11].

REFERENCES

1. N. Koudas, D. Srivastava, T. Yu, and Q. Zhang.Aggregate query answering on anonymized tables. In ICDE, pages 116–125, 2007.

2. L.Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. Int. J.Uncertain Fuzz, 10(6):571–588, 2002.

3. Yedukondalu, SK.Mohiddin,”A Novel Approach For Data Publishing In Mining”, International Journal of Research in Computer and Communication Technology, Vol 2, Issue 7, July-2013.

4. K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In ICDE, page 25, 2006.

5. Susan Wiedenbeck Jim Waters, Jeam-Camille Birget, Alex Brodskiy Nasir Memon, “Authentication Using Graphical Password: Basic Results”.

6. Sunumol Cherian, Kavitha Murukezhan, “Providing Data Protection as a Service in Cloud Computing,” International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 ISSN 2250-3153.

7. Ke Wang,Philip S. Yu,Sourav Chakraborty,”Bottom-Up Generalization: A Data Mining Solution to Privacy Protection”, http://www.cs.sfu.ca/~wangk/pub/icdm04.pdf

8. D. J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Y. Halpern. Worst-case background knowledge for privacy-preserving data publishing. In ICDE, pages 126–135, 2007.

9. G. Ghinita, Y. Tao, and P. Kalnis. On the anonymization of sparse high-dimensional data. In ICDE, pages 715–724, 2008.

10. P. Samarati. Protecting respondent’s privacy in microdata release. TKDE, 13(6):1010–1027, 2001.

11. Tiancheng Li, Ninghui Li, Jian Zhang, Ian Molloy, “Slicing: A new Approach to Privacy Preserving Data Publishing,” IEEE 2012 Transactions on Knowledge and Data Engineering, volume: 24, Issue: 3.

12. R. Nicole, “P. Maniatis et al., “Do You Know Where Your Data Are? Secure Data Capsules for Deployable Data Protection,”Proc. 13th Usenix Conf. Hot Topics in Operating Systems (HotOS11), Usenix, 2011;www.usenix.org/events/hotos11/tech/final_files/ManiatisAkhawe.pdf.

13. T. Li and N. Li. On the tradeoff between privacy and utility in data publishing. In KDD, pages 517–526, 2009.

14. L. Sweeney. K-anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzz. 10(5):557–570, 2002.