Slicing of Graphical Password for Privacy-Preserving in Cloud Computing

Ashwini G. Raut

Abstract— Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public 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 easiest and fastest storage and retrieval of data. The main issue 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 short, simple, 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 [1]. Also For providing more security we are using the slicing concept. In case of slicing we are slice the graphical password and stored it in database on different location.

Index Terms— Cloud Data protection, Data recovery, Graphical password, Pass points, Encryption, Keys, Slicing of Data.

1 INTRODUCTION

CLOUD Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.For providing security to their data user uses the password. But this password is the alphanumeric password. Alphanumeric password contains uppercase letters, lowercase letter, digits and special symbols. This Alphanumeric password can easily hacked by intruder. For preventing the data we use the graphical password concept. The Graphical passwords have many advantages as it is easy to remember, nobody can guises the password. Using a graphical password, users click on images rather than type alphanumeric characters. We have designed a new and more secure graphical password system, called PassPoints. Our scheme: (1) allows any image to be used and (2) does not need artificial predefined click regions with well-marked boundaries – a password can be any arbitrarily chosen sequence of points in the image [1]. In cryptography there are basically two types of keys they are public key and private key. A public key is known to everyone decryption of the specific data file. 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 [3]. Slicing partitions the data both horizontally and vertically. We show that slicing preserves better data utility than generalization and can be used for membership disclosure protection. Another important advantage of slicing is that it can handle high-dimensional data [4].

Ashwini G. Raut is currently pursuing masters degree program in computer engineering in Amravati University, India,PH-09503883819. E-mail: ashwini.raut.819@gmail.com

2 CLOUD COMPUTING CHALLENGES

Despite its growing influence, concerns regarding cloud computing still remain. In our opinion, the benefits outweigh the drawbacks and the model is worth exploring. Some common challenges are:

2.1 Data Protection

Data Security is a crucial element that warrants scrutiny. Enterprises are reluctant to buy an assurance of business data security from vendors. They fear losing data to competition and the data confidentiality 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 [5].

2.2 Data Recovery and Availability

All business applications have Service level agreements that are stringently followed. Operational teams play a key role in management of service level agreements and runtime governance of applications. In production environments, operational teams support.

1. Appropriate clustering and Fail over
2. Data Replication
3. System monitoring (Transactions monitoring, logs monitoring and others)
4. Maintenance (Runtime Governance)
5. Disaster recovery
6. Capacity and performance management

If, any of the above mentioned services is under-served by a cloud provider, the damage & impact could be severe [5].

2.3 Management Capablities

Despite there being multiple cloud providers, the management of platform and infrastructure is still in its infancy. Features like "Auto-scaling" for example, are a crucial requirement for many enterprises. There is huge potential to improve on the scalability and load balancing features provided today [5].

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 state or country. In order to meet such requirements, cloud providers need to setup a data center or a storage site exclusively within the country to comply with regulations. Having such an infrastructure may not always be feasible and is a big challenge for cloud providers.With cloud computing, the action moves to the interface that is, to the interface between service suppliers and multiple groups of service consumers. Cloud services will demand expertise in distributed services, procurement, risk assessment and service negotiation areas that many enterprises are only modestly equipped to handle [5].

3 GRAPHICAL PASSWORD

A number of authentication techniques have been proposed in the recent times that are based upon graphical methods. Text based passwords are most commonly used for authentication; however, they are highly vulnerable to several kinds of attacks. Graphical techniques are coming up as an attractive alternative to the conventional methods of authentication. In this paper we have proposed a graphical method of authentication that employs graphical coordinates along with a novel introduction of time interval between successive clicks. The user needs to recall the coordinates and the time interval of the successive clicks. This leads to the incorporation of the advantages of the recent graphical methods along with the added security achieved through the use of time interval. The proposed scheme has a much higher password space than the other contemporary graphical authentication schemes. The scheme is robust, secure and very convenient to use. Both the administrator and the users should undergo the graphical password text. In view of the shortcomings of the traditional approach to authentication, i.e. alphanumeric passwords, 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. E.g. 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 should enter the correct graphical password for access to the file. The graphical pass approach is also sometimes called graphical user authentication (GUA). A graphical password is easier to remember than complex text-based password [3].

3.1 Slicing

The basic idea of slicing is to break the association cross columns, but to preserve the association within each column. This reduces the dimensionality of the data and preserves better utility than generalization and bucketization. Slicing preserves utility because it groups highly-correlated attributes together, and preserves the correlations between such attributes. Slicing protects privacy because it breaks the associations between uncorrelated attributes, which are infrequent and thus identifying. The key intuition 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 ($1 \leq i \leq c$), 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 [7,10,8] 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 [6]. Generalization [11, 13, 12] 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 [9] 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.

3.1.1 Bucketization

The advantages of slicing over bucketization can be understood as follows: First, by partitioning attributes into more than two columns, slicing can be used to prevent membership leak. Our empirical 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 privacy models, such as k-anonymity, l-diversity, confidence bounding, and t-closeness,assume the adversary has only very limited background knowledge. Specifically, they assume that the adversary's background knowledge is limited to knowing the quasi-identifier. Yet, recent work has shown the importance of integrating an adversary's background knowledge in privacy quantification. A robust privacy notion has to take background knowledge into consideration. Since an adversary can easily learn background knowledge from various sources. [15]

3.1.2 Bottom-Up Generalization

Algorithm 1 describes 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. Let us look at this computation in more details.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). In fact, 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).Therefore, we can focus on "critical generalizations" as defined below.

Algorithm 1 the bottom-up generalization
1: **while** R does not satisfy the anonymity requirement **do**
2: **for all** generalization G **do**
3: compute IP(G);
4: **end for**;
5: find the best generalization Gbest;
6: generalize R by Gbest;
7: **end while**;
8: output R;[16]

4 Conclusion

This work focused on the usability of PassPoints, but its security is also an important issue. While graphical users always took more time to input their passwords than alphanumeric users, even so there was evidence that with continuous use graphical passwords can be entered quite quickly[1].This work motivates several directions for future research. First, in this paper, we consider slicing where each attribute is in exactly one column. An extension is the notion of overlapping slicing, which duplicates an attribute in more than one column. This could provide better data utility, but the privacy implications need to be carefully studied and understood. It is interesting to study the tradeoff between privacy and utility [4].

References

[1] Susan Wiedenbeck Jim Waters, Jeam-Camille Birget, Alex Brodskiy Nasir Memon, "Authentication Using Graphical Password: Basic Results," www.google.com.

[2] Dawn Song, Elaine Shi, and Ian Fischer, "Cloud Data Protection For The Masses" ,IEEE Computer Graphics and Applications magazine,published by the IEEE Computer Society.[0018-9162/12/$31.00@2012 IEEE January 2012]

[3] 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.

[4] 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.

[5] 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.

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

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

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

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

[10] 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.

[11] P. Samarati. Protecting respondent's privacy in microdata release. TKDE, 13(6):1010–1027, 2001.

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

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

[14] X. Xiao and Y. Tao. Anatomy: simple and effective privacy preservation. In VLDB, pages 139–150, 2006.

[15] 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.

[16] 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