ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

REDUCING THE REDUNDANCY IN MONITORING OF CLOUD RESOURCES SUBSPACE CLUSTERING

M.Gayathri

Research scholar, Department of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

Dr.M.Prabakaran

Research Supervisor, Asst. Prof. of Computer Science, Govt. Arts College, Ariyalur, Tamilnadu, India.

ABSTRACT

A significant but uncertain dilemma in the literature of subspace clustering, which is referred to as "information overlapping-data coverage" dispute. Current solutions of subspace clustering usually invoke a grid-based Apriori-like procedure to identify dense regions and construct subspace clusters afterward. Due to the nature of monotonicity property in Apriori like procedures, it is inherent that if a region is identified as dense, all its projected regions are also identified as dense, causing overlapping/redundant clustering information to be inevitably reported to users when generating clusters from such highly correlated regions. However, naive methods to filter redundant clusters will incur a challenging problem in the other side of the dilemma, called the "data coverage" issue. In this paper, therefore, we further propose an innovative algorithm, called "Non Redundant Subspace Cluster mining" (NORSC), to efficiently discover a succinct collection of subspace clusters while also maintaining the required degree of data coverage. NORSC not only avoids generating the redundant clusters with most of the contained data covered by higher dimensional clusters to resolve the information overlapping problem but also limits the information loss to cope with the data coverage problem. As shown by our experimental results, NORSC is very effective in identifying a concise and small set of subspace clusters, while incurring time complexity in orders of magnitude better than that of previous works.

KEYWORDS: Data Mining, Subspace Clustering, Redundancy Filtering

INTRODUCTION

Nowadays there is a growing dependency on web applications, ranging from individuals to huge organizations. Almost the lot is stored, available or traded on the web. Web applications can be personal websites, blogs, news, social networks, web mails, bank agencies, forums, e-commerce applications, etc. The omnipresence of web applications in our approach of life and in our economy is so significant that it makes them a natural aim for malicious minds that want to exploit this new streak. The safety inspiration of web application developers and administrators should reflect the magnitude and importance of the assets they are supposed to protect. Although there is a rising concern about security (often being subject to regulations from governments and corporations, there are significant factors that create securing web applications a difficult task to achieve:

1. The web application market is growing fast, resulting in a huge proliferation of web applications, based on different languages, frameworks, and protocols, largely fueled by the (apparent) simplicity one can develop and maintain such applications.

2. Web applications are highly exposed to attacks from anywhere in the world, which can be conducted by using widely available and simple tools like a web browser.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

3. It is common to find web application developers, administrators and power users without the required knowledge or experience in the area of security. Web applications provide the means to access valuable enterprise assets. Many times they are the main interface to the information stored in backend databases; other times they are the path to the inside of the enterprise network and computers. Not surprisingly, the overall situation of web application security is quite favorable to attacks. In fact, estimations point to a very large number of web applications with security vulnerabilities and, consequently, there are numerous reports of successful security breaches and exploitations.

The proposed methodology is based on the thought that we can assess diverse attributes of existing web application security mechanisms by injecting realistic vulnerabilities in a web application and attacking them routinely. This follows a method inspired on the fault injection system that has been used for decades in the dependability area. In our case, the set of "vulnerability" þ "attack" represents the space of the "faults" injected in a web application, and the "intrusion" is the effect of the victorious "attack" of a "vulnerability" causing the application to enter in an "error" state. In practice, security "vulnerability" is a weakness (an internal "fault") that may be oppressed to cause harm, but its occurrence does not cause harm by itself. Conceptually, the attack injection consists of the opening of sensible vulnerabilities that are afterwards by design subjugated (attacked).

RELATED WORK

In [1] Kevin Y. Yip, David W. Cheung, and Michael K. Ng et al presents In high-dimensional data, clusters can exist in subspaces that conceal themselves from traditional clustering methods. A numeral of algorithms has been proposed to identify such predictable clusters, but most of them rely on some user parameters to guide the clustering procedure. The clustering accuracy can be seriously corrupted if incorrect values are used. Unfortunately, in actual situations, it is rarely possible for users to supply the constraint values accurately, which causes practical difficulties in applying these algorithms to real data. In this article, we observe the major challenges of predictable clustering and propose why these algorithms need to depend heavily on user parameters. Based on the analysis, we propose a new algorithm that exploits the clustering status to adjust the internal thresholds dynamically without the assistance of user parameters. According to the results of widespread experiments on real and artificial data, the new method has outstanding accuracy and usability.

In [2] Man Lung Yiu and Nikos Mamoulis et al presents Irrelevant attributes insert noise to high-dimensional clusters and render established clustering techniques inappropriate. Recently, numerous algorithms that discover expected clusters and their associated subspaces have been projected. In this paper, we understand the analogy between mining recurrent item sets and discovering dense projected clusters around random points. Based on this, we suggest a method that improves the competence of a predictable clustering algorithm (DOC). Our technique is an optimized adaptation of the frequent pattern tree growth method used for mining frequent item sets. We propose numerous techniques that employ the branch and bound paradigm to resourcefully discover the projected clusters. An experimental study with synthetic and real data demonstrates that our performance significantly improves on the accuracy and speed of previous techniques.

In [3] Cecilia M. Procopiuc, Michael Jonest, Pankaj K. Agarwal et al presents A geometric formulation for the analysis of greatest projective cluster, initial from natural requirements on the concentration of points in subspaces. This allows us to develop a Monte Carlo algorithm for iteratively computing projective clusters. We prove that the computed clusters are good with high probability. We implemented a customized version of the algorithm, using heuristics to speed up computation. Our extensive experiments demonstrate that our technique is significantly more accurate than previous approaches. In particular, we use our techniques to build a classifier for detecting rotated human faces in cluttered images. Recognizing the require for augmented flexibility in dropping the data

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

dimensionality, recent database investigate has proposed computing projective clusters, in which points that are directly connected in several subspace are grouped jointly

In [4] Le Lu, Ren´e Vidal et al presents Central and subspace clustering methods are at the core of lots of segmentation problems in computer vision. However, both methods not succeed to provide the accurate segmentation in numerous realistic scenarios, e.g., when data points are close to the intersection of two subspaces or when two cluster centers in different subspaces are spatially close. In this paper, we address these challenges by considering the problem of clustering a set of points lying in a union of subspaces and distributed around multiple cluster centers inside each subspace. We propose a simplification of K means and K subspaces that clusters the statistics by minimizing a cost purpose that combines both middle and subspace distances. Experiments on artificial data estimate our algorithm constructively beside four other clustering methods. We also analysis our algorithm on computer vision troubles such as face clustering with changeable clarification and video attempt segmentation of dynamic scenes

In [5] Karin Kailing, Hans-Peter Kriegel, Peer Kr□oger et al presents numerous application domains such as molecular biology and characteristics make a wonderful amount of data which can no longer be managed without the assist of competent and effectual data mining methods. One of the main data mining tasks is clustering. However, conventional clustering algorithms often fail to notice significant clusters because most real-world data sets are characterized by a high dimensional, inherently sparse data space. Nevertheless, the data sets often include interesting clusters which are hidden in a range of subspaces of the unique characteristic space. Therefore, the concept of subspace clustering has recently been addressed, which aims at automatically identifying subspaces of the feature space in which clusters exist. In this paper, we introduce SUBCLU (density-connected Subspace Clustering), an effective and efficient approach to the subspace clustering problem. Using the concept of density-connectivity underlying the algorithm DBSCAN, SUBCLU is based on a formal clustering concept.

PROPOSED SYSTEM

The method proposed was implemented in a existing Vulnerability & Attack Injector Tool (VAIT) for web applications. The tool was hardened on top of broadly used applications in two scenarios. The first to appraise the effectiveness of the VAIT in generating a large number of realistic vulnerabilities for the offline review of security tools, in exacting web application vulnerability scanners. The second to illustrate how it can exploit injected vulnerabilities to launch attacks, allowing the online evaluation of the effectiveness of the counter measure mechanisms installed in the objective system, in exacting an intrusion detection system.

MODULE SPECIFICATION

- Preparation Stage
- Attack Injection Stage
- Load Generation Stage
- Intruder Process stage
- Evaluate Security

PREPARATION STAGE

In this preparation stage, the web application is interacted (crawled) executing all the functionalities that need to be tested. Meanwhile, both HTTP and SQL communications are captured by the two probes and processed for later use. The interaction with the web application is always done from the client's point of view (the web browser). The outcome of this stage is the correlation of the input values, the HTTP variables that carry them and their respective source code files, and its use in the structure of the database queries sent to the back-end database (for SQLi) or displayed back to the web browser .

VULNERABILITY INJECTION STAGE

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

It is in this vulnerability injection period that vulnerabilities are injected into the web application. For this reason, it needs in sequence about which input variables carry applicable information that can be used to perform attacks to the web application. This point starts by analyzing the source code of the web application files searching for locations where vulnerabilities can be injected. The Vulnerability Operators are built upon a pair of attributes: the Location Pattern and the Vulnerability Code Change. The Location Pattern defines the conditions that a specific vulnerability type must comply with and the Vulnerability Code Change specifies the actions that must be performed to inject this vulnerability, depending on Vulnerability.

ATTACK LOAD GENERATION STAGE:

After having the locate of copies of the web application source code files with vulnerabilities injected we require to produce the collection of malicious interactions (attack loads) that will be used to attack each vulnerability. This is done in the attack load generation stage. The attack load is the malicious activity data desirable to attack a given vulnerability. This statistics is built roughly the interface patterns derived from the preparation stage, by alteration the input values of the vulnerable variables.

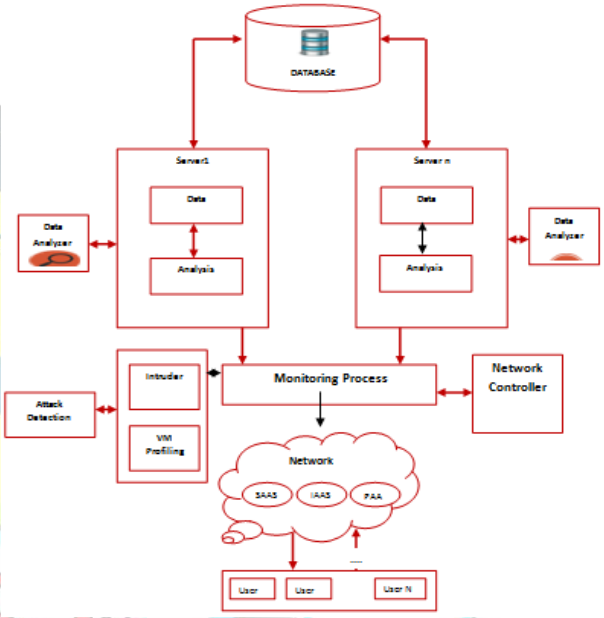
INTRUDER PROCESS STAGE

In the intruder stage, the web application is, once again, interacted. However, this time it is a "malicious" interaction since it consists of a collection of attack payloads in order to exploit the vulnerabilities injected. The attack intends to alter the SQL query sent to the database server of the web application or the HTML data sent back to the user.

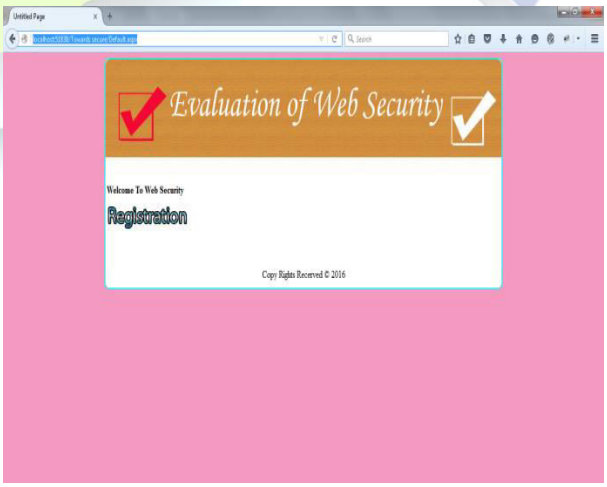
EVALUATE SECURITY

To evaluate security mechanisms like the IDS, providing at the same time indications of what could be improved. By injecting vulnerabilities and attacking them automatically the VAIT could find weaknesses in the IDS. These results were very important in developing bug fixes (that are already applied to the IDS software helping in delivering a better product).The VAIT was also used to evaluate two commercial and widely used web application vulnerability scanners concerning their ability to detect SQLi vulnerabilities in web applications.

ARCHITECTURE DIAGRAM



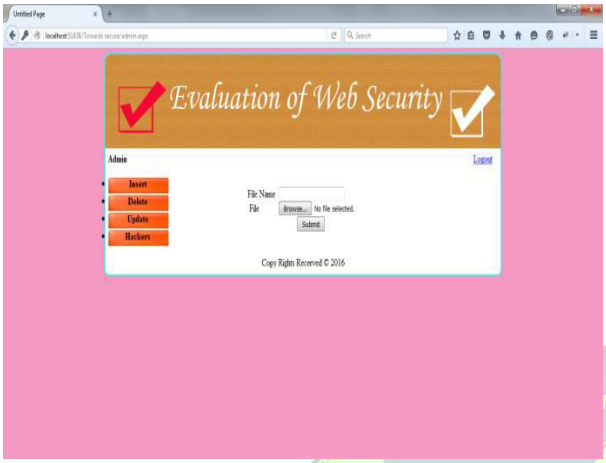
SIMULATION RESULT

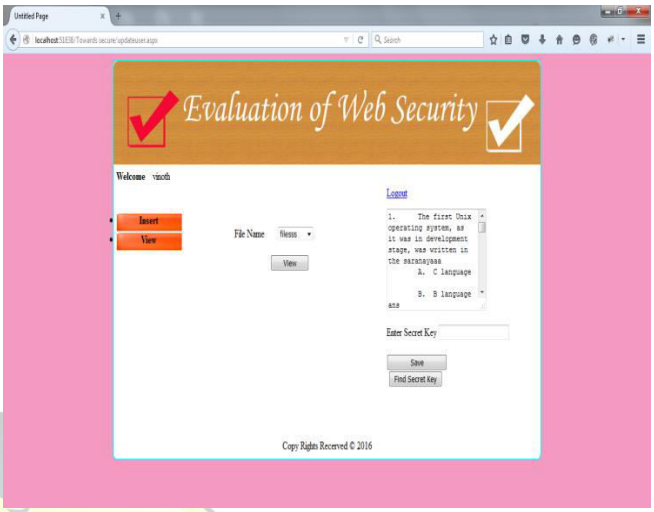


ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 5, Special Issue 12, April 2018

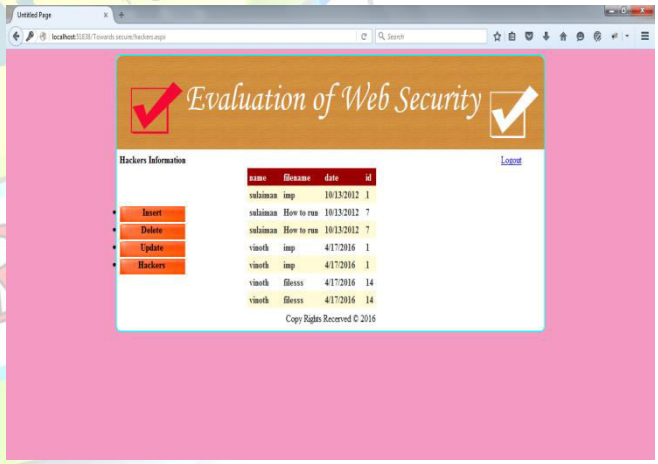
ADMIN HOME





FILE UPLOAD PROCESS

VIEW THE HACKER INFO



MODIFY THE CONTENT

CONCLUSION

An imperative dilemma in the journalism of subspace clustering, called "information overlapping-data coverage" challenge. Naive extensions of preceding works cannot guide to a good cut line to balance these two issues. To mixture this, we propose the NORSC algorithm to routinely discover a succinct collection of subspace clusters while also maintaining the required degree of data coverage. NORSC does not produce the clusters with most of the restricted data covered by higher dimensional clusters to evade the information overlapping problem. In addition, NORSC limits the

information loss in the recognized redundant clusters to cope with the data coverage problem. Our algorithm leverages the maximal dense units to generate nonredundant clusters. As established by our new results, NORSC can discover a concise and small collection of subspace clusters, and the time efficiency of NORSC outperforms the addition of previous works.

REFERENCE

[1] C.C. Aggarwal, J. Han, J. Wang, and P.S. Yu, "A Framework for Projected Clustering of High Dimensional Data Streams," Proc. 30th Int'l Conf. Very Large Data Bases (VLDB), 2004.

[2] C.C. Aggarwal, A. Hinneburg, and D. Keim, "On the Surprising Behavior of Distance Metrics in High Dimensional Space," Proc. Eighth Int'l Conf. Database Theory (ICDT), 2001.

[3] C.C. Aggarwal and C. Procopiuc, "Fast Algorithms for Projected Clustering," Proc. ACM SIGMOD, 1999.

[4] C.C. Aggarwal and P.S. Yu, "Finding Generalized Projected Clusters in High Dimensional Spaces," Proc. ACM SIGMOD, 2000.

[5] C.C. Aggarwal and P.S. Yu, "The IGrid Index: Reversing the Dimensionality Curse for Similarity Indexing in High Dimensional Space," Proc. ACM SIGKDD, 2000.

[6] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications," Proc. ACM SIGMOD, 1998.

[7] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB), 1994.

[8] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When Is Nearest Neighbors Meaningful?" Proc. Seventh Int'l Conf. Database Theory (ICDT), 1999.

[9] M.-S. Chen, J. Han, and P.S. Yu, "Data Mining: An Overview from Database Perspective," IEEE Trans. Knowledge and Data Eng., 1996.

[10] C.H. Cheng, A.W. Fu, and Y. Zhang, "Entropy-Based Subspace Clustering for Mining Numerical Data," Proc. ACM SIGKDD, 1999.