ISSN 2395-1621

Data Mining In E-Commerce Security

[#1]Kiran More, [#2]Bhagyashri Nanaware, [#3]Seema Misal, [#4]Prof. M. J. Arote

[1]morek688@gmail.com
[2]bhagyashri2293@gmail.com
[3]sseemamisal@gmail.com
[4]mohiniarote@gmail.com

[#123]Student, BE Computer, Pune University
[#4]Prof, Computer Department
JSPM's
Bhivarabai Sawant Institute of Technology & Research,
Wagholi, Pune – 412207

ABSTRACT

E-commerce web sites how to use web mining technology for providing security on e-commerce web sites. The connection between web mining ,security and ecommerce analysed based on user behaviour on web .Different web mining algorithms and security algorithm are used to provided security on e-commerce web sites. Based on customer behaviour different web mining algorithms like page rank algorithm and trust rank algorithm is used for developing web mining framework in e-commerce web sites. We have developed false hit database algorithm and nearest neighbour algorithm to provide security on e-commerce web site.

***Keywords*— data mining ;e-commerce;web mining;security issue.**

ARTICLE INFO

Article History

Received: 9th October 2015

Received in revised form :

9th October 2015

Accepted : 12th October 2015

Published online :

13th October 2015

I. INTRODUCTION

Electronic Commerce may include any computer mediated business process, but a common usage is to use it to describe commerce taking place using the World Wide Web as an enabling transport. For many reasons, including our areas of expertise and experience, we will concentrate on this definition of E-Commerce. The web is the way to do business for many reasons. Thin, ubiquitous clients, the wide availability of access, and consistent interfaces to many different platforms are among the reasons to choose web solutions for many problems. In addition, the limited nature of the HTTP protocol makes security issues simpler. However, any transaction taking place across the public Internet is open to a wide variety of security problems.

E-commerce is defined as the buying and selling of products or services over electronic systems such as the Internet and to a lesser extent, other computer networks. It is generally regarded as the sales and commercial function of e-Business. There has been a massive increase in the level of trade conducted electronically since the widespread penetration of the Internet. A wide variety of commerce is conducted via e-Commerce, including electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems.

II. EXISTING SYSTEM

There are many e-commerce sites available which allows to register themselves. Registration and login is what they take into consideration as a part of security measures. Validation of inner details and their entries are not been included in it. E- Commerce sites like olx, ebay works on the same model. There is no way we can analyze our users in these sites. Their goal is to improve Web service and performance through the improvement of Web sites, including their contents, structure, presentation, and delivery. They focus on the mining of server side data. Their data sources are almost exclusively server logs, sometimes with site structure and/or page contents. They target groups of users instead of individual users. It is overwhelming for a Web site to deal with users on an individual basis.

III. PROPOSED SYSTEM

We are considering an ecommerce site, where every user will be registered and will be able to use the application only after their login. All the entries of the users are also analysed so that the security is taken into consideration and decision of valid or invalid user is taken.

© 2015, IERJ All Rights Reserved

Algorithm Explanation for PGP Encryption:

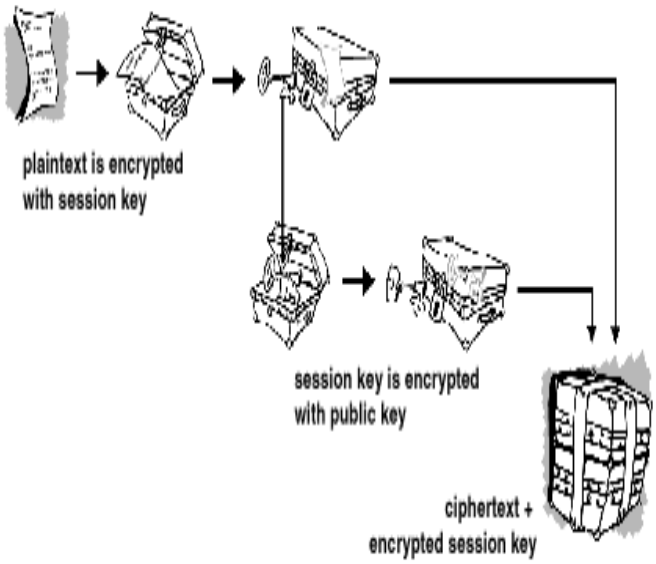

Fig 1: Encryption Process.

Reads data into byte of uploaded file for the purpose of encryption. This mechanism is using Bouncy castle package, which is java implementation for cryptography algorithm.
There are 2 keys-

1. Public key-> used in encryption
2. Private key-> used in decryption

For Encryption it uses public key. But before using public key it first encrypts this key which is keyless encryption (data compression performed by system) by using PGP Encrypted Data Generator class (available in BC API).The actual data is encrypted with the encrypted key which is in binary format.Now, it uses BASE64Encoder to store encrypted data in a legal format (that can be sequential string format). And this data will then use for decryption purpose and it will use BASE64Decoder and will follow reverse process.

IV. MODULE DESCRIPTION

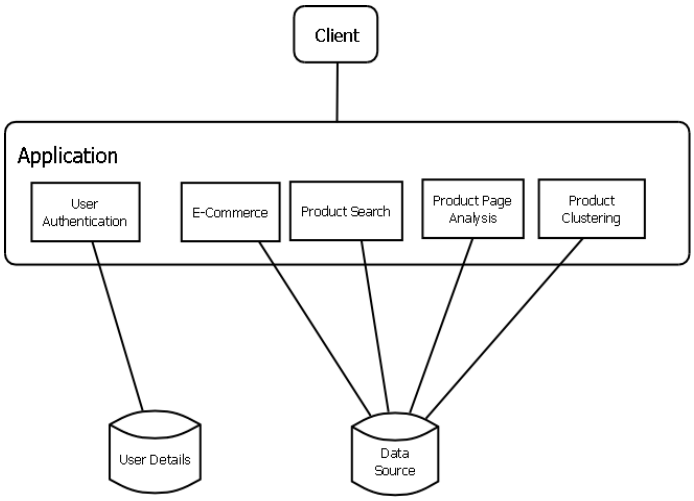

Fig 2: System architecture.

1. E-Commerce Module: This will facilitate users to add products and their features into the system in a particular format. It will also facilitate to view the valid product list populated. This will lead to generation of pages related to products which will be validated for its trust level based on the Trust Rank mechanism.

2. Product Search Module: This module will facilitate searching of the products present in the product list based on the keyword specified as well as the Trust Rank associated with the product. Thereby, the user will be facilitated to perform an accurate searching of products

3. Product Page Analysis: This module will facilitate the system to analyse various product pages and finalize the Trust Rank associated with them. Thereby, the product validity will be decided by the system based on the Rank associated with them. Thus, only valid products will be displayed by the application on the product page.

4. Product Clustering: This module will facilitate the system to categorize the products into various categories based on a selected set of characteristics. These clustered products will be utilized for performing the search process.

Security issue in e-commerce:

E-commerce security strategies deal with two issues: protecting the integrity of the business network and its internal systems, and transaction security between the customer and the business. The main tool for protecting internal net is the firewall. A firewall is a hardware and software system that allows only external users with determined characteristics access a protected network. But there are hacker tools that can successfully penetrated firewalled networks. In this situation, transaction security can be helpful; it depends on the organization's ability to providing privacy, authenticity, integrity and availability.

Privacy: The abuse of consumer privacy is becoming a concern at the consumer, business and government level. There will be resistance to participating in certain types of e-commerce transactions if the assurance of privacy is low or non-existent.

V. CLUSTERING ALGORITHM USED FOR SEARCHING

Step 1:
input=product name or any keyword of the feature
output = cluster set where the entered keyword is included
Step 2:
Assign frequency values to all the keywords,then we calculate the tdidf from these frequency values.
Where,
TF = (Number of times term t appears in a document) / (Total no. of terms in the document)
IDF= log e (Total number of documents / Number of documents with term t in it).
Step3:
Cosine Similarity : similarity(x, y)=cos(=x.y/||x||*||y||)
Where, x represent one record and y represent second record.

© 2015, IERJ All Rights Reserved

Step 4: Create two arrays and compare the cosine similarity values of one array to another.
Step5: Clusters created according to keywords.

Comparison between Cosine similarity algorithm used in e-commerce and k-means algorithm

1) In k-means tf-idf is calculated first and on that basis cosine similarity calculated likewise in our algorithm cosine similarity is calculated on the basis of vector Data same as tf-idf.

2) In k-means we taking user input for how many no of cluster user want and in our project we are directly clustering till end of last product gets clustered.

VI. PROTECT YOUR E-COMMERCE SITE

1. Choose a secure ecommerce platform

Put your ecommerce site on a platform that uses a sophisticated object-orientated programming language says Shawn Hess, software development manager. Our administration panel is inaccessible to attackers because it's only available on our internal network and completely removed from our public facing servers.

2. Require strong passwords.

While it is the responsibility of the retailer to keep customer information safe on the back-end, you can help customers help themselves by requiring a minimum number of characters and the use of symbols or numbers," says Sarah Grayson, senior marketing manager for the Web.

3. Layer your security.

One of the best ways to keep your business safe from cybercriminals is layering your security says Grayson. Start with firewalls, an essential aspect in stopping attackers before they can breach your network and gain access to your critical information. Next, she says add extra layers of security to the website and applications such as contact forms, login boxes and search queries.

4. Make sure you have a DDOS protection and mitigation service.

With DDOS [Distributed Denial of Service] attacks increasing in frequency, sophistication and range of targets, ecommerce sites should turn to cloud-based DDOS protection and managed DNS services to provide transactional capacity to handle proactive mitigation and eliminate the need for significant investments in equipment. In addition, a managed, cloud-based DNS hosting service can help deliver 100 precent DNS resolution, improving the availability of Internet-based systems that support online transactions and communications.

VII. CONCLUSION

E-commerce is dynamic and is throwing up new business models at a fast pace. Web usage mining techniques can be significant tools for fraud detection and finding unusual accesses to secure data and transactions in much electronic interaction. Clustering discovers rules allow grouping the items with similar attributes together. It applies to cluster of users, pages or sessions from web log file. User clustering is designed to find user groups pattern, therefore if we find out any abnormal or uncommon patterns in these groups, it may be a potential attack.

REFERENCE

1.Andad Sharma,"Web Usage Mining:Data Preprocessing,Pattern Discovery and Pattern Analysis on the RIT Web Data"MS Project Report, 2008.

2. Bhavani Thuraisingham, Latifur Khon, Murat Kantarcioglu "Semantic Web, Data Mining and Security", 1541-1672/10 2010 IEEE.

3. Eamonn O'Raghallaigh, "Major Security Issues in E-commerce"

4. F. Masseglia, D. Tanasa, B. Trousse, "web usage mining: Sequential pattern Extraction with a very low support."

5. Jadeep Srivastava,"accomplishments and Future Directions",University of Minnesota,USA.

6. L.K. Joshila Grace, V. Maheswari, Dhinaharan Nagamalai, " Analysis of web logs and web users in web mining",

7.J. Borges and M. Levene, Mining Association Rules in Hypertext Databases, Proc. 1998 Int'l Conf. on Data Mining and Knowledge Discovery (KDD'98), 149-153, 1998.

8. A. Buchner and M. Mulvenna, Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining, SIGMOD Record, 27(4), 1998.

9. A. Caglayan, M. Snorrason, J. Jacoby, J. Mazzu, R. Jones, K. Kumar, Learn Sesame -- a Learning Agent Engine, Applied Artificial Intelligence, 11:393--412, 1997.

10. L. Chen and K. Sycara, Webmate: A Personal Agent for Browsing and Searching, Proc. of Second International Conf. On Autonomous Agents (Agents 98), Minneapolis, MN, May, 1998.

© 2015, IERJ All Rights Reserved