Chapter 37
A Web Backtracking Technique for Fraud Detection in Financial Applications

Tasawar Hussain
Mohammad Ali Jinnah University, Pakistan

Sohail Asghar
COMSATS Institute of Information Technology, Pakistan

ABSTRACT

The web based applications are maturing and gaining the confidence of their users gradually, however, www still lacks the mechanism to stop the hackers. The implementing the adhesive security measures such as intrusion deduction systems and firewalls, are no more useful breaker for online frauds. The Web Backtracking Technique (WBT) is proposed for fraud detection in online financial applications by applying the hierarchical sessionization technique on the web log file. The web log Hierarchical Sessionization enhances the focused groups of users from web log and paves the path for in-depth visualization for knowledge discovery. User clicks are compared with user profiles for change in previous user click records. Those transactions which do not conform to business rules are stopped from business activities. The WBT analyzes suspicious behavior and will produce reports for security and risk mitigation purposes Furthermore, suspicious transactions are mined for the up-gradation of business rules from hierarchical sessionization. The proposed WBT is validated against the university web log data.

INTRODUCTION

In 1990, internet was made available to public. Since then it has revolutionized the world and made it a global village. In the last two decades, the numbers of websites have grown from few hundreds to 650 million websites and thousands of new web pages are being added to this mega stream per day. Moreover, today we have more than 3 billion active users of the internet. This exponential growth of World Wide Web (www) has become the single largest knowledge repository with the world. Thousands of transactions are being carried out on the daily basis to execute the various web based businesses.

DOI: 10.4018/978-1-5225-3422-8.ch037

From sea to space, the web is working as a knowledge backbone to provide the basic level information to advance research. The web is the most powerful and cost effective media to deliver services to its users. Consequently, business community prefers internet for their services and users feel free to avail the web services (Mohammad Pourzarandi & Tamimi, 2013). The web is providing its services in all most every walk and department of life irrespective of geographical boundaries (Hussain & Asghar, 2013). The internet is simple in nature to deliver the services and motivated the organizations (Hawwash & Nasraoui, 2010) for online e-business for more competitive environment and challenges.

As the websites and users are growing day by day, millions of user clicks are being recorded per second. The websites lack the user feedback mechanisms (Hussain & Asghar, 2013) to improve the websites and provide to the point web knowledge to its users. The web mining tools are effectively supporting the web to study and analyze the websites. The web mining is the application of data mining techniques (Bari & Chawan, 2013) and web mining itself is divided into three broad categories, namely *web structure mining (wsm)*; *web content mining(scm)*; and *web usage mining(wum)*.

The website consists of web pages in a tree structure and are linked with each other via hyperlink. Each web page consists of web objects that is the core of the knowledge for web users. Website, web pages and web objects provide the website structure. The application of data mining techniques on web structure is commonly known as web structure mining. The information retrieval is a one of the major challenge of the World Wide Web and wsm is playing a pivotal role to provide the structural knowledge about the web pages for linked analysis.

Web contents are very important and play the key role to deliver the web services to the end users. Through wcm, we mine the web page contents for knowledge and information retrieval and extract the useful patterns. The internet is like an ocean of knowledge and to the point knowledge retrieval is a difficult task. By applying the data mining techniques we can mine the web contents for efficient search query results (Chaniara and Sherasiya, 2014). For structural knowledge about the web contents, the hybrid web techniques based on wsm and wcm are applied.

In web usage mining, data mining techniques are applied on the users' data available in the form of web log files (Sharma, 2013). These log files are maintained on the web server (the server hosting the banks' website). When client interact with the website to avail the desired service, log file captures the users' surfing on a website and each activity of the user is recorded. These log files contain the hidden knowledge about the users' traversal during the surfing. The primary objective of web log files is for the maintenance of the server not for the data mining. To extract the hidden knowledge from the log file is complex data mining exercise. Without deploying the proper data mining system, the accurate and precise hidden knowledge can't be mined. The analysis of web log file is providing various benefits to the website owner and website developers such as performance of web server; smartness of website; user click stream history; user profiles; user sessions; predictions; and fraud detections.

The unrestrained growth of the internet has not only opened the new competitive markets for business, however, has also given free hand to hackers to play fraudulent activities. The web-based applications are the most vulnerable to security threats and attacks (Garg & Singh, 2013; S.Mirdula & D.Manivannan, 2013). The hackers are not only devastating the confidence of clients, however, are posing serious threats to online business (Meyer, 2008). According to National Fraud Authority UK, in 2012, cyber attackers are plundering money around £1.1 billion per annum (Harrison, 2012) and the ratio of online fraud losses in the USA is in million dollars annually, while in India the losses are around 22.90 billion rupees (Jassal & Sehgal, 2013). The online fraud data is not available in Pakistan and other developing countries as these countries have no online fraud monitoring and gauging systems.