Testing Websites by P^3R^2 Model

Rajiv chopra[1] **Dr. Sushila Madan[2]**

**[1]Guru Teg Bahadur Institute of Technology, Delhi,
Department of Computer Science Engineering,
Delhi-64, India.**

**[2] Lady ShriRam (LSR) College, Delhi University,
Department of Computer Science,
University of Delhi, Delhi.**

Abstract

Cyber espionage at the government, corporate and individual levels has increased dramatically over the last few years. The cyberspace has become the latest domain of Chinese belligerence. There is a crying need for a cohesive global arrangement to address the emerging issues of Cyberspace. Apart from trying to cull out information on internet traffic and content usage, many are also trying to secure sensitive data. The primary target seems to be the data leaks and an access to monitor content. There are links between leaks and usability of websites. A more usable application will reduce leaks. Thus, web security is at risk. Terms like navigation, navigational structures, usability, complexity and security are all inter-related. A more complex web software will have a complex navigation, so poor usability and hence more security flaws. The need is to thoroughly test these websites. Various techniques of black-box, white-box and gray box testing of websites have been proposed in literature for extensive and thorough testing of websites. Still the errors escape out. This paper proposes a new PPPRR (P^3R^2) MODEL for website testing.

Keywords: Website testing, Web security, Page-Test-Trees, Reusability, Path testing.

1. Introduction

In recent years, cyber espionage has become the topic of primary discussion among security experts and software testers. The survey results carried out by M. Eric Johnson and Nicholas D. Willey of Dartmouth College shows that there were data leaks throughout the healthcare chain [1]. They found out that many recent security breaches weren't break-ins but rather inadvertent data leaks. Even OWASP- Open Web Application Security Project [2] lists data leakage as one of the top ten web vulnerabilities. This vulnerability is not a technical security hole in operating system or server software but rather depends on the way a website is tested. At the Hindustan Times Leadership Summit 2011, WikiLeaks founder Julian Assange reports that China was stealing confidential emails of Central Bureau of Investigation (CBI). In December 2010, CBI's website was hacked. More recent data from the NASA's research shows that the attackers compromised the pages of NASA's website. Pentagon (USA) receives 6 million hacking and security threats every year. About 60% of health organizations suffer a US $6 billion loss annually. The following losses have been reported recently and are shown in Table-1.

To prevent these types of attacks, the need is to test the websites extensively. This paper proposes a technique for thorough testing of websites though not exhaustive but with optimum number of resources (cost, time and manpower/effort).

Recent survey by Penome's group is shown in the graph of figure 1. It shows the rising trend of data losses due to insufficient web testing. It is seen that the maximum data loss is the information loss that is 40-42% in 2010 and 2011. Other losses like business disruption are only 28 percent.

Table-1 below shows various security threats due to insufficient testing techniques:-

Table 1: Major threats faced due to insufficient testing.

S. no.	Company name	Losses incurred	Year of loss
1	WebSense Security Report	75% of most popular websites hacked.	2008
2	NASA	Hackers compromised pages of NASA's website	2011
3	Sony	Personal information of 8500 people leaked	2011
4	Pentagon (USA)	6 million hacking and security threats	2011
5	Anti-phishing working group	More than 4,85,000 strains of scareware are active now.	2011
6	60% of health organizations	US $6 billion annually.	2011
7	Indian Computer Emergency Response Team reports to Indian Government.	117 govt. websites hacked in last 3 years and a total of 90,119,252 websites hacked.	2008-2011 (till June,2011)
8	Techgig.com reports (In USA and UK)	Cyber criminals launch a new malware to steal money from your bank and reaasure that your money is safe.	10th – January, 2012.

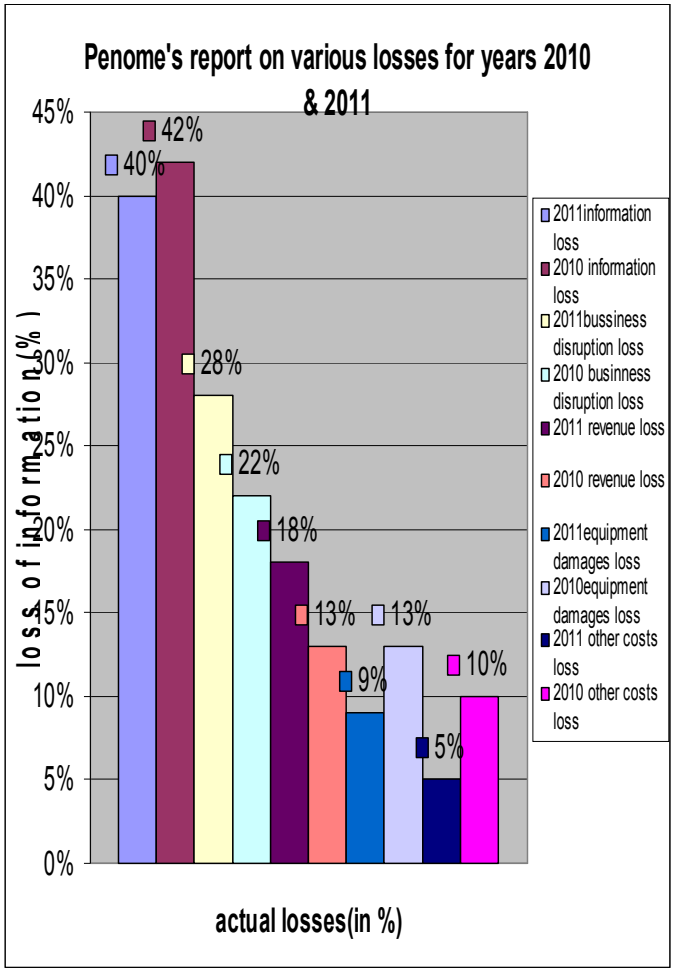

Fig. 1: Various losses given by Penome's annual report, 2011

In yet another recent survey of European Internet users [18] it is reported that-

- 45% respondents were worried about lack of online privacy.
- 60% say that they were concerned about business tracking, everything being done on internet.
- 52% predicted that cybercrime will continue to increase.
- 46% said that the world will become Internet-centric.
- 33% said that in next 9 years or so, all books, magazines, newspapers will disappear as the material is online available on websites.

Tim Burners Lee, the founder of World Wide Web, quotes that there are some straight forward things the industry

should be doing right now to fix things-security holes. The pie chart of figure2 shows what the application security overview looks if OWASP attacks occur. It has been created by Web Hacking Incident database, 2011.

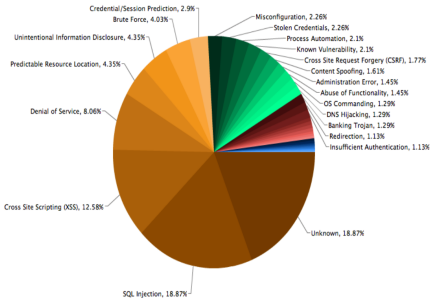

Figure 2: Created by Web Hacking Incident Database (2011)
Source:
www.techtimely.wordpress.com/2011/04/22/web-hacking-threats/

Whenever a user navigates from one web page to another, there is some small delay in between this navigation. There is a possibility that during this time interval an attacker will break into the system and can cause heavy data losses. Broadly speaking, there are four ways in which data can be leaked-

(1) Over an email or through an organizational websites or even over a telephone.
(2) Data lost accidentally with the computer or the media on which it is stored.
(3) Data may fall into unauthorized hands.
(4) Data may be stolen deliberately by insiders or by outsiders to the organization.

So, leakage of sensitive data should be limited or prevented whenever possible [3].
The main objective of this paper is to reuse path testing technique on website so as to test them thoroughly and extensively. Testing consumes 50% of the total software development time and we have chosen path testing technique because path testing technique can alone detect almost 65% of the errors in the program [17].

2. Related Work

Many researchers have proposed many testing model each of which have different origins and pursuing different test goals for dealing with the unique characteristics of websites. A few of related recent studies are stated below:

In 2000, B.M. Subraya and S.V. Subramanya [4] used a performance testing approach to decompose the behavior of a website into testable components.

In 2001, Fillipo Ricca and Paolo Tonella[5] stress that Web page is a central entity in any website. Web pages can be static or dynamic. A test case for a web application is a sequence of pages to be visited plus the input values to be provided to pages containing forms. They use white-box testing criteria like page testing, hyperlink testing, and definition use testing, all-uses testing and all-paths testing on websites. They use path expressions and node-reduction algorithms. Two tools Reweb and TestWeb have been developed. The all-path testing criterion is achieved, restricted to independent paths.

In 2002, M. Benedikt [6], built VeriWeb, a dynamic navigation testing tool for web applications. It explores sequences of links in web applications by non-deterministically searching action sequences, starting from given URL. VeriWeb's test is based on graphs where nodes are Web pages and edges are explicit HTML links.

In 2003, S. Elbaum[7] used user sessions data to generate test cases for web applications. It is the input data collected and remembered from previous user sessions. This data is captured from HTML forms. It can help produce effective test-suites at lesser costs.

In 2004, A. Andrews [8] describes a method deriving tests from FSMs. Web applications can be completely modeled using FSMs but even simple applications can suffer from problem of state space explosion.

In 2004, Eric Y. K. Chan and Y. T. Yu[9] used a classification tree method (CTM) as the black-box criterion and "same path" as white box criterion. CTM are hierarchical trees based on classes. No web pages are considered.

In 2008, Zhongsheng Qian, Huaikou Miao, Hongwei Zeng [10], used page flow graphs and page test trees to derive the test paths for black box navigational testing. The path expressions are used to generate all test paths and translate

them into a test specification which is an input to the test engine. The test engine generates test cases, executes them and finally produces the test report. No consideration of white box testing is considered.

In 2008, Jingxian Gu, Lei Xu, Baowen Xu, Hongji Yang[11], applied traditional Module-to-Module path based strategy to component based web application. So, they extended Module-to-Module path testing method. They stressed that normally, Module-to-Module path always starts from the main function and ends at the main function but paths in component-based web application can start from one page and end at another page without return paths. But this testing is not very intelligent.

In 2009, Bo Song and Huaikou Miao[12] developed navigation model to generate test. But it is a directed graph and graphs have problems like cycle. So, extended FSM (EFSM) is difficult to use. It will generate a set of test cases with redundancy. So, a FSM test tree (FSM-TT) is proposed. Shorter test sequences can be generated without loss of states. A FSM-TT is a spanning tree constructed from EFSMs. But this sequence of paths has not been utilized further except for generation of black-box paths for testing.

In 2009, Wenhua Wang, Yu Lei, Sreedevi Sampath, Raghu Kacker, Rick Kuhn and James Lawerence [13], modeled the navigational structure of a dynamic web application. They presented a combinatorial approach to building a navigation graph. They used abstract URLs to control the page explosion problem. Pages are abstracted based on their URLs instead of their contents. They devised an algorithm to build web navigation graph. Also a prototype tool, called Tansuo was implemented in Java to explore complete navigation structures.

In 2010, Hui-Zhong Shi, Bo Chen and Ling Yu[14] integrated black box and white-box testing to propose a generic framework of web security evaluation. No consideration on Basis Path testing is considered.

In 2010, Nicha Kosindrdecha, JiraPun Daengdej[15] aims to derive tests from state chart diagrams. This is a UML based diagram from which test sequences are generated to test a web application.

In 2011, Alejandra Garrido, Gustavo Rossi [16], did refactorings intended to improve usability in web applications. They define navigation model refactoring as a change to the navigation model of a web application that preserves the set of operations made available by all the nodes in the model and the reachability of each operation through a navigation path from the home node.

On the basis of the above mentioned related works, we identify a new testing framework for path based reusability testing of websites.

2. Research Methodology

In its research design, we test the navigation between web pages. In testing, the set of WebPages within the website that are of interest is first considered, then the testing scope of the website is limited to any web page that belongs to the set or is directly linked by any web page belonging to the set. The testing process consists of following a path down a page test tree (PTT) until the end. Our work here proposes to reuse the same set of test paths obtained from path expressions of page test trees for white box testing now.

3. Proposed Work

Various studies as stated in literature focus on either black-box or white-box testing of websites or even a combination of both. In this paper, it is proposed to reuse the same set of basis paths as obtained from Page-Test-Trees to do white-box testing that is, testing code at these nodes only. This reuses the existing test paths, test cases and test suites, thereby saving a lot of time, cost and manpower (effort).
The proposed work will follow the following PPPRR (P^3R^2) MODEL-

- Construct a Page Flow Diagram (PFD) of website under test (WUT).
- From the PFD, derive its Page-Test-Tree (PTT) using algorithm in [10].
- Get black-box paths from the PTT above which should be a set of basis independent paths.
- Reuse these paths again (reusability of paths) to conduct white-box testing (source code testing) at these specific nodes only.
- According to vulnerabilities, generate test cases, test suites and test reports.

The proposed PPPRR (P^3R^2) MODEL has many significant benefits as doing only black-box or only white-box testing of websites cannot detect those many errors as is possible with our proposed technique. The proposed technique saves time, cost as the test cases and test suites may be reused, newer test cases may be added or deleted as needed. A lot of manpower (effort) is also reduced as the test paths are already available from PTTs and we are reusing them again for white box testing.

4. Conclusion

Web Application's vulnerability losses can be reduced if the proposed methodology is used. Website's usability, complexity, security and hence testing are all inter-related. A complex web structure means high cyclomatic complexity, V (G), which means that the system will be more insecure. This is so because then website testing also becomes difficult. Our technique may not be 100% exhaustive but still will make WUT more secure. To solve these problems, we propose a new PPPRR (P^3R^2) MODEL Our future work will focus on generating test cases for these paths, using Ajax based applications for web testing, minimizing test cases and hence tests paths, using and developing tools for automated testing for both Ajax and non-Ajax based applications.

References

[1] M. Eric Johnson and Nicholas D. Willey, "Usability Failures and Healthcare Data Hemorrhages", IEEE Journal , March-April 2011.

[2] http://www.owasp.org/documentation/topten.html.

[3] A White Hat Website Security Report, Spring 2010, 9th Edition.

[4] B. M. Subraya , S.V. Subrahmanya, "Object Driven Performance Testing of Web Applications", The first Asia-Pacific Conference on Quality software, HongKong, China, Oct. 2000.

[5] Fillippo Ricca and Paolo Tonella, "Analysis and Testing of Web Applications", ITC-irst, Centro per la Ricerca Scientifica e technologica , Italy, IEEE 2001.

[6] M. Benedikt, J. Freire and P. Godefroid, "VeriWeb: automatically testing Dynamic Websites", In Proceedings of 11th International WWW Conference, Honolulu, HI, May 2002.

[7] S. Elbaum, S. Karre and G. Rothermal, " Improving Web Application with User Session Data" in the proceedings of the 25th International Conference on Software Engineering, Portland, Oregon, May 2003, p-p 49-59.

[8] A. Andrews, J. Offut and R. Alexander, "Testing web Applications by Modeling with FSMs", Software and Systems Modeling, 2004.

[9] Eric Y. K. Chan and Y. T. Yu, "Evaluating several Path-based Partial Dynamic Analysis Methods for selecting Black-Box generated Test Cases", Fourth International Conference on Quality Software , IEEE 2004.

[10] Zhongsheng Qian, Huai kou Miao, Hongwei Zeng, "A Practical Web Testing Model for Web Application Testing.", Third International IEEE conference on Signal Image Technologies and Internet Based System, 2008.

[11] Jingxian Gu, Lei Xu, Baowen Xu, Hongji Yang, "An Extended MM-path Approach to Component-based Web Application Testing", 12th IEEE International Workshop on Future Trends of Distributed Computing Systems, IEEE 2008.

[12] Bo Song, HuaiKou Miao, "Modeling Web Applications and Generating Tests: A Combination and Interactions-Guided Approach", 3rd IEEE International Symposium on Theoretical Aspects of Software Engineering, IEEE 2009.

[13] Wenhua Wang, Sreedevi Sampath, Yu Lei, Raghu Kacker, "An Interaction based Test Sequence Generation Approach for Testing Web Applications", IEEE, High Assurance Systems Engineering Symposium, 2008, p-p 209-218.

[14] Hui-Zhong Shi, Bo Chen and Ling Yu, "Analysis of Web Security Comprehensive Evaluation Tools", 2nd International Conference on Network Security, Wireless Communications and Trusted Computing, IEEE 2010.

[15] Nicha K, JiraPun Daengdej, "A Black box Test Case Generation", International Journal of Computer Science and Information Security (IJCSIS), September 2010.

[16] Alejandra Garrido and Gustavo Rossi, "Refactoring for Usability in Web Applications", IEEE computer society, 2011.

[17] Sangeeta Sabharwal, Ritu Sibal, Chayanika Sharma, "A Genetic Algorithm based Approach for Prioritization of test case scenarios in static testing", International Conference on Computer and Communication Technology (ICCCT)-2011 (IEEE).
[18] Lee Garber, "CyberWarfare-Security and Privacy", IEEE Journal September/October 2011.

Er. Rajiv Chopra is an Assistant Professor in Department of Computer Science and Engineering at Guru Teg Bahadur Institute of Technology, Delhi. This institute is affiliated to Guru Gobind Singh Indraprastha University, GGSIPU Delhi. He is M. Tech (IT) from GGSIPU, Delhi, MIT from MAHE, Manipal University and pursuing Ph. D. in computer science and engineering. He has 15 years of teaching experience including 1 year in software industry. He has authored 14 books –Software engineering, Software testing, Software project management (3- books), Microprocessors (2-books), OOSE, OS, ACA, Computer Graphics, DBMS, AI, Requirements and estimation techniques. He has presented papers at National conferences and attended many National and International conferences. Also he is the life member (LM) of Indian Society for Technical Education (ISTE).

Dr. Sushila Madan is an Associate Professor at Lady Shri Ram College for Women, University of Delhi, Delhi. She is Ph.D. from Delhi University, M. Tech. (software systems) from BITS-PILANI and M. Sc. in Applied Mathematics from IIT Delhi. She has also submitted a project funded by UGC titled "Security Risk Management in E-Commerce". Dr. Madan has authored books on IT, Multimedia and web technology, Management Information and Control Systems, E-commerce and Essential PC tools. To her credit there are a number of research papers which have appeared in leading journals; some of which have been presented in conferences and included in conference proceedings; leading national level magazines and in-house journals of corporate sector. Moreover, white papers on the Internet are also available. Also she is a member of the Computer Society of India.