Analyzing Security and Performance Issue in Web Data Mining Technology

Md Nadeem Ahmed
Research Scholar
IFTM University, India

Mohd Hussain, Ph.D
Director, JIT
Lucknow, India

ABSTRACT
Since as the internet and web application emerges security is the most challenging issue which we are facing, leads possibility of being easily damaged. Currently we based application structure is designed only by considering little security but avoid Performance issue. After the detailed study of web services architecture it is analyzed that it is not suitable in counter-tracing the WS attack, an adaptive intrusion detection and prevention (ID/IP) framework to protect The WS against attacks related to WSDL/JSON/SQL is thus introduced. Through Explanation by examples, the framework Verified that by making use of agents that act as Sensors, data mining techniques such as clustering, association and sequential rule coupled with fuzzy logic to further analyze and identify anomalies, is able to exhibit the adaptive nature of capturing anomalies and avoiding false alarm. Also the log files which contain User Name, IP Address, Visiting Path, Time Stamp, Page last visited , number of Bytes Transferred, Result Status, URL which can effectively supervise the network attack. In this paper we will discuss elaborately about several security problem and performance issue related to web application and their possible solutions.

Keywords
Web data mining, Security, Design performance, web services

1. INTRODUCTION
Web technologies has various features like simplicity, portability, Robustness forced the developer to develop application such as e-Banking, e-learning and business oriented applications being deployed over the Application server. Unfortunately, these basic building blocks for Web Services technology such as XSD (XML schema definition), SOAP (Simple Object Access Protocol), UDDI (Universal Description, Discovery, and Integration), and WSDL (Web Services Description Language) lacks security issue, poses vulnerabilities and are the loose holes for attacks. Although WS-Security Standards [1] are important to address the issues of cryptography, authorization, and authentication they are not very much effective in removing attacks such as SQL/XSD injection related to web service-based applications. In the area of Intrusion Detection as well as prevention techniques various research is going on clearly mentioned in [2] like Countermeasure, Deterrence, Prevention, Preemption, Deflection and Detection. The Technique used in these does not matter; the main objective is to provide security to web service application. Network security such as Firewalls, data encryption, anti-virus software, and intrusion can provide security to the intranet by blocking the unexpected access requests outside, it cannot check whether the data streams passed it have malice codes but most illegal transaction are still occurring in application layer during business process ,financial transaction and in the Database management .these may be hacking the websites, illegal transfer of money from banks, stealing password through cookies, stealing the sensitive data etc. Therefore the purpose of this paper is to propose the idea to solve the security and performance related issue in designing web application over the internet.

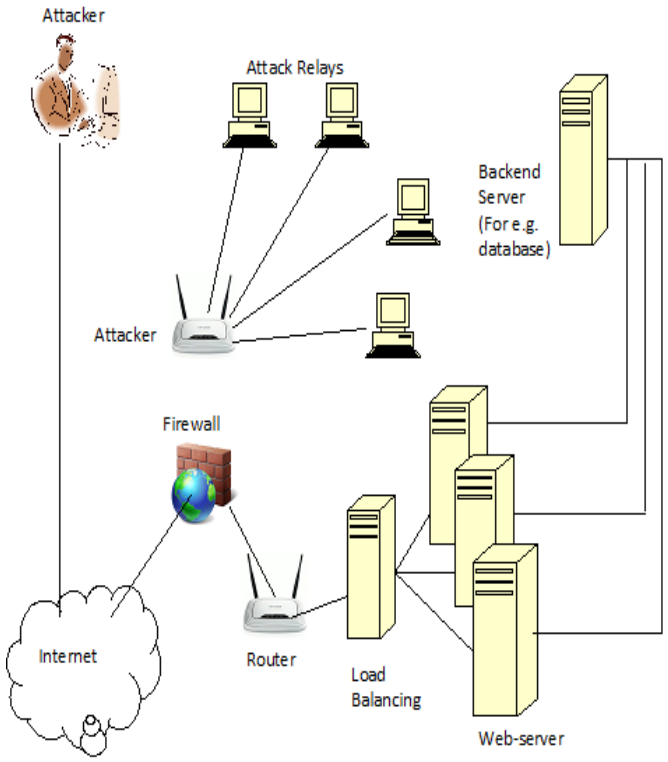

Figure 1: Web Services Attack Overview

2. CURRENT WEB APPLICATION ARCHITECTURE
The distribution of responsibilities between system components and the placement of the components on computers over the network is perhaps the most evident aspect of distributed system design. Application level designing is responsible for solving the security problem currently these are RBAC (Role-Based Access Control), Language-Based Systems, Commercial System, SWAP (Secure Web Applications Project). These will be discussed one by one. The SWAP proposed by David Scott is an interdisciplinary research which was initiated by the Laboratory for Communications Engineering and the Computer Laboratory of the University of Cambridge [5, 6]. The SWAP is attempted to address and fix the security problem of Web pages on the high-level application layer. Hence, the SWAP approach can shorten application development time and protect against attacks in the application layer. The key element of SWAP utilizes a specialized SPDL (Security Policy Description Language) language to design an application-level firewall

called security gateway on the system. Security policies written in SPDL are compiled for execution on the security gateway. The security gateway can dynamically analyze and transform HTTP requests and responses to execute the policy of detailed descriptions. Initially, the operational procedure of SWAP employs their existing framework to dynamically extend into secure Web applications, and selects a Spectre (Security Policy Enforcement through Runtime Checks) tool. Then, they have developed a secure IDL (Interface Definition Language) compiler. Under this situation, its source code is available, and also allows the statement of SPDL to directly map into Web-based applications. Figure 2 presents a typical Web-based application using the SWAP tool. The application contains source code written in three different languages, as well as some third-party component in which it is assumed that the source code is inaccessible. Spectre protects Language C or third-party component of the application by dynamically transforming HTTP requests and responses in accordance with the refined security policy of SPDL for this application. As you can see, Spectre does not protect the parts of the application written in Languages A and B. Instead, in secure IDL compiler has folded that the relevant parts of the SPDL policy are directly mapped into the source code. Also, the IDL compiler supports multiple languages, allowing us to convert the SPDL into Language A and Language B. In practice, an entity may act as different roles in different cases, which mean that any entity can cover many roles. In addition, each role can serve as different entities. The authors have proposed RBAC that entity and privilege have their indirect relationship as shown in Figure 3 [3, 4]. Also, they can attain to associate each other through their roles. The concept of the role represents the duty of enterprise organization in which means to have different tasks. The role can be deemed as a set of responsibilities something to do with particular task, which can express special duty assignment. The access authorization of this paper is appointed to the role. Each entity can be specialized to the accessing control of task, but do not need to be specified one by one. The entity can carry out all activities according to the authorized role, as long as an entity is assigned to one certain role.

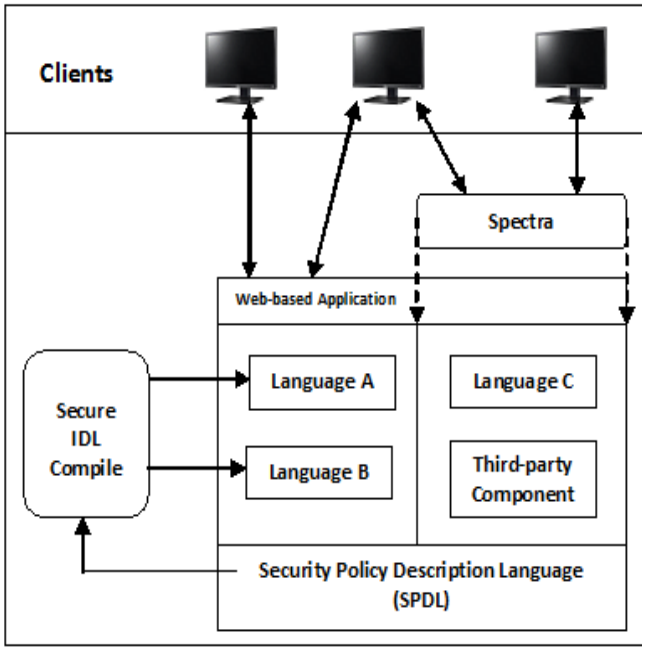

Figure 2: SWAP Architecture

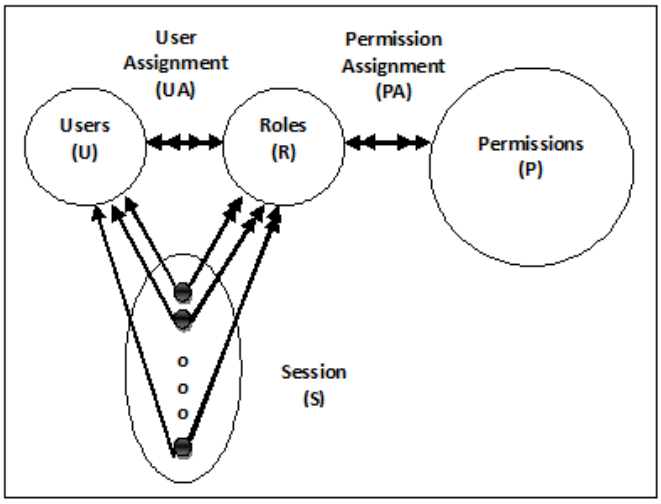

Figure 3: RBAC Architecture

2.1 Design Level Security Issue

Most of web development industries are not following the industry standard in developing and hosting the web applications the user is unaware of the differences between a secured website and unsecured website. So it is necessary to develop the trust path intermediaries building algorithm, false hit database algorithm and nearest neighbor algorithm to provide security on online business application. Multi-step processing is used for nearest neighbor and similarity search in application involving web data and/or costly distance computations. CAMNC- to reduce the size of False hit database. In order to secure Web application against the attacks, various mechanisms do exist. For example, as mentioned in [5], WSDL gives all the sensitive data required to insecure a web application and effort to defend this is just a manually review WSDL to look for dangerous functions and then resolve it out. Scholars in [6] have tell the problem with semantic Web Services which can publish the information about their functional and non-functional properties through the UDDI Business Registry thus opening up targets for attacks, especially distributed attacks. Therefore it is suggested that distributed firewalls and intrusion detection system should share a common vocabulary based on ontology so that different F/IDS would be able to interact and cooperate with each other. While these researchers claimed that it is rather easy to implement and deploy this ontology approach, the requirement for the involvement of all the F/IDS vendors in support of this approach already seemed complex. The default technological nature of SOAP/JSON has improved the chances to attacks such as recursive payload attacks, oversized payload attacks and replay attacks which in turn may lead to Denial of service attacks. To trace these attacks, [7] makes use of such techniques as format and syntax inspection and validation to first validate XML documents that conform to the rules and protocols governing the XML/JSON and SOAP specifications. This is then followed by a deeper inspection of the content to look for policy Contravention such as heavy documents, unsuitable or unpredicted values in fields and so on. Similarly in [8], techniques such as syntax parsing is performed on SOAP messages to check the structure of XML/JSON for syntax errors, filtering policy to check and restrict the size of the SOAP message is used to prevent payload attacks, XML schema validation is used to validate SOAP messages against a predefined XML schema and message monitoring with timestamp. In Web Application development, there are some important design concerns for implementing the system.

However, we will propose following design issues according to two main aspects: security and performance. The popularity of the Internet with lots of opportunities to connect computers anywhere in the world has significantly increased the potential threat of the organization's Web-based applications. Hence, the Web-based applications are required to concern with security of accessing and storing information. In particular, following design issues are to solve the security problems in a Web-based application:

- SQL injection: Stream SQL statement to attack database or records in database.
- Session hacking: trace session Id of Web user and enter user's Session.
- Data revealing: Let user look the detail information while occurs exception.
- Hidden-field unauthorized alterations: Update contents of hidden-field in the Web pages
- Access control: Different users and roles can have different interfaces or permissions.

2.2 Design issue for performance

Besides security, the second most important characteristics of Web-based applications is performance. In general, following performance solutions are concerned with programming in a Web-based application:-

- ❖ Accessing time for web page: It will suffer from the effect of programming structure.
- ❖ Database Communication Time: Time required to perform CRUD operation
- Data Retrieval: Time for retrieving data from database.
- Data Insertion: Time for inserting data into database.
- Data Update: Time for updating data from database.
- Data Delete: Time for deleting from database.

In summary, two demands as stated above are driving the need for designing Web-based applications. Therefore, ISPWAD framework has been developed to support security and performance features on the Web-based applications

3. RELATED WORK

In Sensor Web Agent Platform (SWAP) proposal, security strategy written in SPDL is compiled for implementation on the gateway security. The security gateway dynamically examines and converts HTTP requests and responses to impose the specified policy. But, this approach only secures interactions between Web pages. In contrast, Role-based access control (RBAC) model is only confined user to access Web pages, but unable to safeguard interactions between Web pages. Also, the Web-based application development process is a multiple step procedure consisting of: Requirement, Design, Coding, Testing, Development, and Deployment [9]. Normally, to provide security to a Web-based application is difficult task, but the authors have proposed to solve the security of the applications using Language-Based systems such like J2EE/J2ME programming tool [10]. Also some researchers have proposed the specific domain modeling and the service-oriented design frameworks for designing the secure Web-based applications [11]. So far as now, most of researches only focus on designing the secure Web-based applications. Although, a well developed Web-based applications design can use the framework to support both security and performance features. Current IP trace back schemes can be categorized into two main areas, proactive and reactive [12]. Reactive trace back systems are responses to an ongoing attack, thereby must remain active during the attack, otherwise they cannot react to a Distributed denial of attack (Ddos). This makes reactive systems, like Control flooding [13] and Input debugging [14], unsuitable for the internet but is best suited for controlled networks. One of the problems with reactive schemas is that they require ISP co-operation, which usually is not usually forthcoming due to a loss of competitive advantage. In contrast, proactive schemas actively record tracing information as packets transgress the network, in which the victim can reconstruct the path taken by the attack packets and subsequently identify the source of the attack. Some examples of proactive schemas include messaging [15][16], logging [17][18] and packet marking [19][20]. Intelligent Decision Prototype (IDP) can be used in most of these areas but the main focus of this paper is the packet marking area. Researchers in educational institutions and in companies such as Microsoft, Oracle, and IBM are devoting time and effort to develop and maintain security issues for a Service-Oriented Architecture (SOA). However, one of the most remarkable works in the industry has been introduced by IBM, who has recently announced its proposal for a complete security model of SOA applications, especially those within banking systems [21]. The suggested IBM model consists of three basic levels: Business Security Services, Security Policy Infrastructure and IT Security Service. The framework discusses most of the security issues involved in an SOA environment and it is primarily designed based on the Web Services Standards. However, the details of the proposed security services still need to be validated.

4. POSSIBLE SOLUTIONS AND FUTURE DIRECTION

XML/JSON became very much popular and significant by the evolution of the Web technology as mentioned in [22] , from Web 1.0 which is web browser or HTML based in which presentation logic, and business data are identical, to Web 2.0 which is XML/Web Services based and Application program interface managed that isolate application logic from tight coupling of presentation and business data logic pages and to now Web 3.0 or XML/RDF (Resource Description Framework) based Semantic Web, consisting of three distinct components basically include the Presentation (HTML and XHTML), Application logic (Web Services API), and Data (Data Models). As mentioned in [25] the problems with HTML and XML as explained below; Web 1.0 identified the complications arises in case of HTML which does not implement a way for presenting rich semantics and syntax, therefore obstruct in the exchange of data. But in case of XML, it provides a standard serialized syntax for defining the structure and semantics of data, thus minimize the space between Web 1.0. and Web 2.0. However, XML does not provide standard data structure and phraseology to define business processes. Thus RDF, a data model on top of XML to provide three elements: objects, values of properties applied to a certain object, and properties hence reducing the gap for Web 2.0 and evolved into now the Semantic Web or Web 3.0. The idea of Semantic Web as a layered approach is thought up by Tim Berners-Lee, the pioneer and founder of the web technology – World Wide Web(WWW). In Layered Approach to the Semantic Web and as explained in [23], the bottom most layer after the UNICODE and URI layer, is the XML (namespace and

schema), a language that lets one write structured Web documents with a user-defined vocabulary, particularly suitable for sending documents across the Web. Next comes the RDF data model layer which may also provide RDF schema for modeling primitives for organizing Web objects into layered structure. Other layers such as the Ontology Vocabulary layer is to allow the representations of more complex coupling between Web objects, the Logic layer is to allow the writing of application-specific declarative knowledge, while the Proof layer involves the actual deductive process as well as proof of validation, the final layer, Trust layer emerge through the use of digital signatures. The main focusing point here is that given in this layered approach and the example as shown in [24], security cannot be examine in segregation, but to provide in each layers to impart for end-to-end security. Let us explain with example, only securing XML is not sufficient to direct access to the portion of document for browsing, reading and modifications, but RDF security also required for elucidation and semantics is also required because under certain circumstance, portions of the document may be "unclassified" while other portions treated as "classified". Consequently, research work on Semantic Web security is further going and should include the area for protecting it against security threats mentioned above. The proposed adaptive ID/IP framework for WS come in role and can be expanded to enclose the security threats arises in Semantic Web particularly in the XML and RDF layers. Also, research for protecting other layers against attacks shall also continue in the future.

5. CONCLUSION

In this paper work, related to Web Services focusing in the various vulnerabilities and attacks in WS and reveal some of the existing available defending techniques which being not adaptive, are not sufficient in counter measuring those WS attacks. Then an adaptive ID/IP framework in which data mining techniques, fuzzy logic and agent technology are used to recognize and filter off the exceptions has been proposed. In this proposed framework, initially, normal user reaction, behaviors and service request/response are captured and profiled. Agents which act as sensors are then further used to identify the suspected items. Furthermore interpretation on these doubtful items is done by using association rule-based, clustering and sequential rule based techniques together with fuzzy logic. Index values and attack indicators are then attached to these items to indicate the level of seriousness or the high probability of them being real attacks. As explained with the examples that the lower the index value, the higher is the probability that the anomaly is a genuine attack and further preventive steps can then be taken. After putting forward this framework as the preliminary idea, further research and development towards conceptualizing it shall be on-going in the future. Areas of interested should be geared into two, first towards exploring the various data mining and fuzzy logic algorithms with the aim to optimize the performance of the framework, and second, towards securing the Semantic Web. Presently, there are still many researches about security solutions in Web-based applications. But, it can identify that previous researches do not have proposed a complete case for solving the security issue and improving the performance problem. The proposed ISPWAD of this paper is to integrate SWAP and RBAC approaches to provide a total solution for eliminating security risk and improving system throughput in designing Web-based applications. The purposes of ISPWAD approach are to fix the security gap and improve the processing performance during in designing Web-based applications. Also, it has been illustrate how to implement the secure Web-based applications with tuning performance. Then, we will utilize the ISPWAD to set up a practical case – EIP system, and also perform simulations and make results analysis. Finally, the experimental results indicate that the proposed ISPWAD can solve related security mismatches, obtain better processing performance, and reduce development time as well as cost. Especially, the ISPWAD can obtain a good balance-point between security and performance. In the future, the ISPWAD framework can provide a reference model for implementing online banking, Telecom, Healthcare and ERP based application.

6. REFERENCES

[1] A. Stamos and S. Stender, "Attacking Web Services: The Next Generation of Vulnerable Enterprise Apps", BlackHat2005, USA, 2005.

[2] A Murali M Rao, "A Survey on Intrusion Detection Approaches", Proceedings of the First International Conference on Information and Communication Technologies, ICICT 2005 IEEE, 27-28 August 2005 Pages(s):233-240.

[3] Prasanna H Bammigatti and Dr. P. R. Rao, "GenericWA-RBAC: Role Based Access Control Model for Web Applications," In Proceedings of 9th International Conference on Information Technology (ICIT'06), No. 6, pp. 237-240, December 2006.

[4] Ravi Sandhu, Edward J. Coyne, Hal L. Feinstein and C. E. Youman, "Role-based Access Control Models," IEEE Computer, Vol. 29, No. 2, pp. 38-47, February 1996.

[5] K. Spett, "Blind SQL Injection:Are Your Web Applications Vulnerable?". SPI Dynamics, 2005

[6] A. Vorobier and J. Han, "Security Attack Ontology for Web Services", Proceedings of the Second International Conference on Semantics, Knowledge, and Grid (SKG'06) IEEE.

[7] P. Lindstrom, "Attacking and Defending Web Services", A Spire Research Report, January 2004.

[8] Y. S. Loh, W. C. Yau, C. T. Wong and W. C. Ho, "Design and Implementation of an XML Firewall", Proceedings of the 2006 International Conference on Computational Intelligence and Security (CIS2006), Guangzhou, China, Nov. 3-6, 2006, pp. 1147-1150.

[9] Izhar Bar-Gad, Amit Klein and Sanctum Inc. "Developing Secure Web Applications," White Paper, June 2002.

[10] Andre N. Klingsheim , Veborn Moen and Kjell J. Hole, "Challenges in Securing Networked J2ME Applications," IEEE Computer, Vol. 40, No. 2, pp. 24-30, February 2007.

[11] Hiroshi Wada, Junichi Suzuki, "A Domain Specific Modeling Framework for Secure Network Applications," In Proceedings of 30th Annual International Computer Software and Applications Conference (COMPSAC'06), pp. 353-355, September 2006.

[12] Aljifri, M., (2003), 'IP Traceback: A New Denial-of-Service Deterrent?' Published By The Ieee Computer Society 1540-7993/03 2003

[13] Stone, R, (2000) "CenterTrack: An IP Overlay Network for Tracking DoS Floods," Proc. 9th Usenix Security Symp.,Usenix Assoc., 2000

[14] Burch, H., and Cheswick, B., "Tracing Anonymous Packets to Their Approximate Source," Proc. 14th

Conf.Systems Administration, Usenix Assoc., 2000, pp.313– 322.

[15] Bellovin, S., Leech, M., and Taylor, T., (2003), 'ICMP Traceback Messages,' Internet Draft, Internet Eng. Task Force, 2003; work in progress.

[16] Mankin, A., Massey, D., Wu, C.L., Wu S.F and Zhang, L., (2001), "On Design and Evaluation of 'Intention- Driven' ICMP Traceback," Proc. IEEE Int'l Conf. Computer Comm. and Networks, IEEE CS Press, 2001. pp. 159–165.

[17] Snoeren, A.C., et al., (2002), "Single-Packet IP Traceback," IEEE/ACM Trans. Networking, vol. 10, no. 6, 2002, pp. 721–734.

[18] Baba, T., and Matsuda, S., (2002). "Tracing Network Attacks to Their Sources," IEEE Internet Computing, vol. 6, no. 3, 2002

[19] Adler, M, (2002),'Tradeoffs in Probabilistic Packet Marking for IP Traceback,' Proc. 34th ACM Symp. Theory of Computing, ACM Press, 2002, pp. 407–418.

[20] Peng,T., Leckie, C., and Kotagiri, R., (2002), "Adjusted Probabilistic Packet Marking for IP Traceback", Networking 2002.

[21] "Understanding SOA Security Design and Implementation", http://www.redbooks.ibm.com/abstracts/SG247310.htm

[22] N. Reed, "Security Guards for the Future Web", The MITRE Corporation,2004.http://www.mitre.org/news/events/tech04/briefings/726 .pdf . Retrieved August 1, 2007.

[23] G. Antoniou, and F. V. Harmelen, "The Semantic Web Vision" in A Semantic Web Primer, The MIT Press, April 2004.http://mitpress.mit.edu/books/chapters/0262012103chap1.pdf Retrieved July 27, 2007.

[24] B. Thuraisingham, "Security Issues for the Semantic Web", Proceedings of the 27th Annual International0 Computer Software and Applications Conference (COMPSAC'03), 2003.

[25] D. Fensel, J. A. Hendler, H. Lieberman, and W. Wahlster, "Introduction" in Spinning the Semantic Web", The MIT Press, March2005. http://mitpress.mit.edu/books/chapters/0262062321intro1.pdf, Retrieved July 27, 2007.