A passive Digital footprint for Extranet environment

Received : 7/12/2017 **Accepted : 18/1/2018**

Harith A. Hussein
Tikrit University/ College of Computer and Math Sciences
Harith_abd1981@tu.edu.iq

Keywords: "Access log file, internal threats, Extranet security, Java Filter Technology, Artificial Neural Networks"

Abstract:

Within both Extranet and Intranet, security and privacy is the greatest importance and a breach could result a tragedy. Maintaining these two aspects are the primary challenge within organization. Bothe Extranet and Intranet are attractive to a lot of organizations, because they can significantly decrease the transaction and management charges among themselves and their customers, vendors, partners or suppliers. However, the same technology that brings benefits can also put an organization at risk if the proper security mechanisms are not in place. Meanwhile, Organization gives too much attention to avoid external threats, a little attention is being paid to internal threats. Moreover, Access control and authentication technology ensure that the user is who or she claims to be, but it is in the same time says nothing about cheated identity or the individual's access rights. Thus, propose or present an access log that can log many information that can be used with further studies as auditing or monitoring system in order to track the internal staff digital footprint and finally predict or prevent the internal threat.

The web logger, can log information's and track user's digital footprint passively around the sites, logged information include user access time and date, user IP address, username, user authentication type, accessed resources, how long they stay within this resources and finally, browser type.

1. Introduction

In our modern linked community, the boundaries between organization's Intranet, Extranet and the internet are muddling. Nearly, each organization has some necessity to spread on the online access to commercial partners, suppliers, clients and/or customers via Extranet or Intranet. From the organization orientation, the key to understanding information security involves recognizing the vital differences between Intranets Extranet and the Internet. The "difference between Intranets Extranet and Internet is not great in terms of technology; the transmission of information is completely different from the organizational point of view" [1]. "Intranets have been defined as internal communication systems of organizations based on the standards of the Internet and the World Wide Web (WWW)" [2]. Additional, an Extranet is similar to Intranet, the different is that Extranet spread more to provide access controlled to authorized customers, vendors, partners or suppliers, which means users outside the organization but technically they are consider as internal users. It is very helpful to breakdown the word for a good understand on how Extranet operates, Extra, for instance, refers to anything vital to our business, up till now exists outside our organization for example clients, vendors, and suppliers. "Extranets linking trading partners need to be secure, and need the following security services: access controls, integrity, availability, confidentiality, repudiation, and authentication" [3].

Within both Extranet and Intranet, security and privacy is the greatest importance and a breach could result a tragedy. Maintaining these two aspects are the primary challenge within organization. Bothe Extranet and Intranet are attractive a lot of organizations, because they can meaningfully decrease the transaction and management charges among themselves and their customers, vendors, partners or suppliers. Though, alike technology that brings competitive advantage can likewise put an organization at danger if you don't apply a good security mechanisms.

Organization gives too much attention to avoid external threats; a little attention is being paid to internal threats. Access control and authentication technology for example, passwords, smart cards and biometric devices ensure that the user is who or she claims to be, but it is in the same time says nothing about cheated identity or the individual's access rights.

"Digital footprint refers to one's unique set of traceable digital activities, actions, contributions and communications that are manifested on the Internet or on digital devices "[4]. The digital foot print can be defined as "The record of your interactions with the digital world and how the data that is left behind can be exploited"[5]. The digital footprints can be classified in to two ways according to the way of tracking: passive and active. "A passive digital footprint is created when data is collected without the owner knowing, whereas active digital footprints are created when personal data is released deliberately by a user for the purpose of sharing information about oneself by means of websites or social media" [6].

Pioneering researchers have ended with fairly good work on investigating the web log data to improve the security and privacy of both Extranet and Intranet. [7] Develop an access log generator for analyzing malicious website browsing behaviors; the objective of the study is to prevent the Denial of Service Attack (DSA). [8] propose an approach to analyze the access control configuration to identify the set of credentials needed to reach a certain location in a system, the writer consider the problem of the insider attacks where the attackers actions usually will be logged as permissible, standard action if they are logged in to the system. Even more novel ideas have been proposed by [1], they proposes an effectual security process that includes a new model to allows flexible web security access control in securing information over the Intranet. Even more attractive idea have done good work on discovering knowledge from www log data [9], other contribute to the field of web log to improve website design[10], understanding the web traffic to improve the proxy cache [11]. None of the literatures propose or present an access log that can log much information that can be used with further studies as an alternative auditing or monitoring technology in order to track the internal staff digital footprint and finally predict or prevent the internal threat.

The main objective of this paper is to propose a customized access log generator based on Java filter technology that can used with further studies to prevent or predict any possible strange activity, the prototype capture the accessed users, authentication level, accessed resources and how long how long they stay within this resources and finally browser type. Strange activity should immediately be escalated so the issue can be investigated to see what is happening. Furthermore, the generated data can be used as source to any form of computer forensics.

The rest of this paper is ordered as follows: Section 2 discuss the influence of both internal, external threats and non-technical threat; in Section 3 we clarify the Extranet security risk and key considerations; Section 4, include the pioneer researches have done in this field of research; Section 5; comprise the methodology, which includes the passive digital footprint framework, the proposed http web logger, the technology of filter and filter chain, authentication and filtering data finally, logging the data. Section 6; discuss the result and lastly, Section 7; include the conclusion.

2. Extranet and Intranet internal threats:

There is a constant fact that "the use of Extranet and the way that technology is used comprises a new threat source, it's highly needed to not only consider the mathematical and physical properties of the system" [12]. As it mentioned early, these threats can be nosy trading partner, or it can be from a hateful members who has gain the access to the partner network in order to acquire the Extranet data and make use of organization trusts, or it can be any form of the non-technical attacks like social engineering attack.

The main occurrences of data theft are stated from activities executed by the organization's internal staff, venders or customers, for many reason seeking for revenge, making money or any other motive.

According to the FBI and the Computer Security Institute (CSI) 2015 survey, "twice as many respondents cited their Internet connection as a more frequent point of attack as those who said assaults came from within their internal systems." Concerning the same FBI/CSI report, "For the fifth year in a row, more respondents (74%) cited their Internet connection as a frequent point of attack than cited their internal systems as a frequent point of attack (33%)." [13].

Over and above, "Industry research indicates over 65% of all reported security breaches occur from within the network, such as password security leaks"[14].
Instead on above, Organization gives too much attention to avoid external threats, a little attention is being paid to internal threats. Trust is respectable but blind trust may carry lack of security gaps.

3. Extranet security risk and consideration:
With Extranets organizations must be extra careful because the lack of security very high, as Intranets and Extranets extensively installed, many security problems have arose. "While many organizations have deployed firewalls and access control technology to improve security, these technologies leave many security issues unaddressed, technology is only one aspect of the security picture" [13, 14].

TCP/IP considerations: The Transition Control Protocol/Internet Protocols (TCP/IP) and other connection technologies are designed to be open. Since, TCP/IP is a connectionless protocol; data is fragmented into packets that navigate over the network looking for the best possible way to reach the destination. Therefore, without good defenses are taken, many types of threat can be occurring, for example the Man in the Middle Attack (MIMA), in which the data can hack.

Misuse of authorized access: authorized user, such as a contractor or supplier, might gain access to unauthorized area of the company's computer system. For example, a contractor gains the access into the HR database to get private financial statistics or information. Passwords are mainly useless against internal threats; there are many technical or non-technical ways to gain the password. Simply, sometime passwords are easy to guess and can be discovered on a notes on employee's computers or it can be easily discovered via any type of social engineering techniques. It is very essential that the two Extranet partners confident that they are interactive with each other and not with fraud user seeking for security breach.

Expose of sensitive information: Users might share information between physically separated workplaces via Extranet, doing that from their home computer can expose private information to be exposed.
According to [15], there are four important strategies for securing Extranets need to be carefully considered: isolation, granular access controls, strong authentication and use of encryption.

Separation: the more opening the great possibility for the wrong people to damage the internal system and database**,** simply by protecting the Extranet from itself, businesses need to manage the firewall rolls very well. According to [15], "isolation can be three zones: a private network, a public network and a DMZ Demilitarized Zoon, DMZ is a physical or logical sub-network that separates an internal Local Area Network (LAN) from other untrusted networks".

Access control and strong authentication: If the organization decides to adopt Extranet, the principle of privilege and levelization must be enforcing. Obviously, the ideal scenario is isolation in a degree that Extranet users grant access to the exact resources they are authorized to access to. Alongside with, the traditional authentication method, it is recommended to add digital certificates to provide an additional level of confidence in the authentication process.

Encryption: Normally, Extranets include sharing sensitive organizational data over the internet. It is highly recommended for Extranet clients to consider the use any technology that provides robust encryption technique for data over the Extranet.

4. Methodology
In this section, we discuss the passive digital footprint then, we explain the http web logger in more details. What is java filter and how it's created, the authentication and captured data similarly will be debated. Finally, ways of collecting data out will be discussed.

4.1 passive digital footprints
The "passive digital footprints data can be stored in many ways depending on the situation, in web environment a footprint may be stored in data base, plain file as a hit" [16]. This footprint may track the user IP, when it

was created, and where they originally came from, the accessed resources and many other attributed depend on the system or organization requirements. These footprint later can be analyzed and escalated to highlighted and register any strange activity. This should be immediately evaluated to see what happen. The proposed solution to implement the passive digital footprint in this paper, we create an http web logger base on java classes. The http web logger will log information and track user's moves around the required web sites without their knowledge. The architecture of the proposed system structured is shown in the next Figure 4.1.

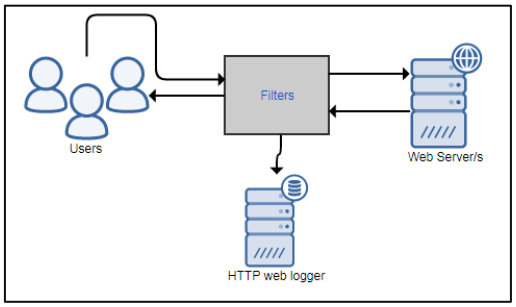

Figure (4.1): passive digital footprint framework

4.2 http web logger

The proposed web logger created using java programming language by using java class that implementing javax.servlet.Filter interface. The filter can completely control web sites access since we can fully works with both the requested object that's hold the data sent from the user and the response object sending data back to the user. As already mentioned, the http web logger can log information's and track user's digital footprint passively around the sites, by applying the filter, users have no way to know that the server side track them and logs information about them, such as user IP address, username, used authentication type, URL (Unified Resource Location) to the accede resources, how long they worked with this page, browser type. In fact, the web logger can furthermore block the access of any undesirable users.

The focal point of implement the filter that will generate web logger is that there is no need to change a web resource configuration with Extranet http resources, such as a JSP page, HTML page, or a servlet, to log users activates and who come to that resource or to restrict access to it. Moreover, no need to alter the data sent by user, or alter the data comes back from the system. All we need is chain of java servlet filters.

Meanwhile, it is difficult to install the logger with real Extranet environment duo to the organization security and privacy, the proposed web logger in a simulated manor installed on local Apache HTTP server in conjunction with Tomcat application server, Tomcat is used to deploy our Java Servlets filter.

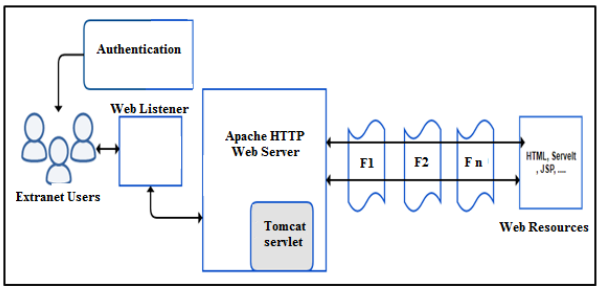

Figure (4.2): System Architecture

4.3 filter and filter chain

To create the filter, first of all we need to create Java class that implements the Filter interface as we declare previously , secondly, we need to includes the init, doFilter, and destroy methods. When the user accesses the resource that managed by filter, the doFilter method handle the control, gaining access to the request object sent by the user (ServletRequest) and the response object sends back to the user (ServletRequest).

In web servers, we can use filters in a filter chain via FilterChain object in doFilter, in which, control goes to the first filter, then the next, the next, and so on, all the way to the filtered resource, such as a JSP page, JSF page, servlet, or an HTML page.

4.4 Authentication and filtering data

The authentication system within Extranet organization in the real world normally, based on levelization, since it can includes wide range of users, business partners, suppliers, vendors, customers or internal staff. In the simulated prototype we edit the users file on the web server (in Tomcat, that's called tomcat-users.xml) two role where created basic and admin, see the next Figure 4.3. The key

component of a secure Extranet is the use of strong authentication techniques.

```
<?xml version='1.0' encoding='utf-8'?>
<tomcat-users>
  <role rolename="tomcat"/>
  <role rolename="role1"/>
  <role rolename="manager"/>
  <role rolename="admin"/>
  <role rolename="author"/>
  <user username="digital" password="footprint" roles="basic"/>
  <user username="harry" password="passiv" roles="admin"/>
  <user username="tomcat" password="tomcat" roles="tomcat"/>
  <user username="both" password="tomcat" roles="tomcat,role1"/>
  <user username="role1" password="tomcat" roles="role1"/>
  <user username="admin" password="" roles="admin,manager"/>
</tomcat-users>
```

Figure (4.3): authenticated users

The first filter capture user access time and date via date object that holds the current system date and time (system time in millisecond). In the next filter, we get the user's IP address with the requested object's getremoteAddr method, the resources that the user is trying to reach with the request object's getrequestURI method, at last, the name of the user's host with the request object's geTRemoteHost method, and determining the browser type by reading the user-Agent HTTP header.

Finally, we collect the user authentication, If the user has logged in using HTTP authentication, we can get the type of authentication (basic or admin in this prototype) using the HttpServletRequest object's getAuthType method, if the value is not null, we can get the user's name by getting a Principal object and using that object's getName method. Many interesting technologies and techniques can be deployed using Java filter.

4.5 logging the data

There are many ways to collect the data out. It can be directed to local text file or logged out directly to DBMS (Data Base Management System). In our proposed prototype we used the FilterConfig object sent to the filter's init method and object's getServletContext method to get the servlet context. The servlet context supports a log method, which can log text. Finally, we create a new FileWriter object, then using the write method and the close method to close the FileWriter. The logged data include username is (if they logged in to the required website, what their IP address is, how long they interact with the web page they accessed, and finally, what type of browser they are using.

5. Result and Discussion

The result of the proposed passive web logger is shown in the next Figure 5.1, each entry in the log file has seven important fields:

- Access time: represent the user access time registered according the system time in millisecond.
- Authentication type: represent the user level of authentication; as we already noted that the authentication system in Extranet environment based on level, since we have many types of users, partners, suppliers or others. In our proposed prototype we have only two roles of users, which is basic and admin as we already mention in section 4.4.
- User name: represent the ID username that is registered in the system, that's will be very helpful in many cases, can be used in conjunction with the accessed resources and IP address to discover any type of internal threat.
- Accessing: represent the accessed resource.
- Hosting: represent the IP address from which the request originated.
- Brower: represent the user browser type. A small piece of information, in case of attack that will be very helpful for computer forensic people.
- Millisecond used: represent how long they worked with the web page they accessed

```
User access at Mon Mar 07 15:47:53 EST 2017
Authentication type: BASIC
User name: digital
User IP: 127.0.0.1
Accessing: /logger/target.jsp
Host: 127.0.0.1
Browser: Internet Explorer
Milliseconds used: 109
```

Figure (5.1): the logged data via the proposed prototype

What does the user will see in case of browsing the filtered web site? Nothing at all, as shown in Figure 5.2 where the user is accessing a page called main-page.jsp. We tracked the users in a passive way as they move around the required sites. In fact, we can even block access if we want.

Figure (5.2): Accessed resources

Normally, the typical access log have web site or services most likely you have 'access' log files where we can see all the requests to our server are recorded with the client IP address, access date and time, HTTP codes, user agent information, etc.
Normally, the typical access log where we can see all the requests to our server recorded with the client IP address, access date and time, HTTP codes, user agent information, etc, looks like this. Figure 5.3.

```
Safari/537.36 [FB_IAB/FB4A;FBAV/134.0.0.25.91;]"91.106.50.141 - -
[31/Jul/2017:15:13:26 +0300] "GET /announcements/1548- HTTP/1.1" 200 6315
"http://m.facebook.com/" "Mozilla/5.0 (Linux; Android 6.0.1; SM-G900F Build
/MMB29M; wv)
```

Figure (5.3): typical apache access log

It looks very intimidating and boring by just looking at this text file. It is not a typical flat form and that makes it hard to see what information is even there. Even after doing data cleaning and transaction to import this text data into a tabular form or readable format, it still doesn't seem to give us much interesting information.
the contribution of this work can be divided into two parts; the first part is about readability, the developed access log file is readable, no need to the process of data cleaning and data transaction or any kind preprocessing process, which is required in the typical access log file. The second part about the data that can be captured, since our access log generator can capture user name, user privilege and how log each user spent in each accessed resources.
The data can be migrated in real time to any intelligent DBMS for further analysis or atomize the notification, a further work can be done to modify the proposed prototype to pipe that access log file in to customized DBMS.

6. Conclusion:

The main objective of this paper is to propose a customized access log generator based on Java filter technology that can used with further studies to prevent or predict any possible strange activity. This paper does not focus on the technologies form to secure Extranet, this paper focuses on how to prevent or avoid the predictive non-traditional internal threats.
The major finding of this work can be divided into two parts; the first part is about access log readability, since the developed access log file (figure 5.1) is more readable, no need to any kind preprocessing process, which is required in the typical access log file (figure 5.3). The second part about the data that can be captured, since our access log generator can capture more attributes like user name, user privilege and how log each user spent in each accessed resources, which is not exist in the typical access log file.
The current version of this paper propose java-based http web access logger, can log information's and track user's activities around the sites. In the future work, the customized log file can be used as source to audit the users account and privileges. Strange activity should immediately be escalated so the issue can be investigated to see what is happening.

7. References:

1- J. Kajava and T. Remes. "Intranet Security from Organizational Point of View" in *Proc. IRIS 23 Laboratories for Interaction Technology, University of Trollhattan Uddevalla*, 2000, pp. 112-123.

2- Pal, A., Ring, K. and Downes, V. *Intranets for Business Applications; User and Supplier Opportunities*. Ovum Ltd, 1996.

3- Phaltankar, K. M. *Practical Guide for Implementing Secure Intranets and Extranets*. Norwood, MA: Artech House Inc, 2000.

4- Garfinkel, Simson; Cox, David. "Finding and Archiving the Internet Footprint" Presented at the first Digital Lives Research Conference. London, England. February, 2009.

5- Fish, T. "Definition of a digital footprint. My Digital Footprint. Internet:

http://blog.mydigitalfootprint.com/definition-of-digital-footprint-again , 2010, [January 16, 2017]

6- Stephen D. Weaver and Mark Gahegan. "Constructing, Visualizing, and Analyzing a Digital Footprint" *American Geographical Society*, Vol. 97, No. 3, pp. 324-350, Jul. 2007.

7- C. Lin, J. Liu and C. Chen. "Access Log Generator for Analyzing Malicious Website Browsing Behaviors" *Fifth international conference on information Assurance and security*, Xi'An China, China, 2009.

8- C. W. Probst and R. R. Hansen. "Analysing Access Control Specifications" *Fourth International IEEE Workshop on Systematic Approaches to Digital Forensic Engineering* DOI 10.1109/SADFE.2009.

9- F. Tao and F. Murtagh. "Towards Knowledge Discovery from WWW log Data" *International conference on Information Technology: coding and Computing*, Las Vegas, NV, USA, 2000.

10- M. Perkowitz and O. Etzioni. "Automatically Learning from User Accesses Patterns"

11- T. Sullivan. "Reading Reader Reaction: A proposal for Inferential Analysis of Web Server Log Files" In Proc. 3rd conference of Human Factors and the web, Denver, Colorado, June 1997.

12- E. Cohen and H. Kaplan. "Exploiting Regulations in Web Traffic Pattern for Cache Replacement" 31th STOC, 1999.

13- k. a. Korow. "Security Consideration for Extranets" SANS institute, December 18, 2001.

14- Versin "Guide to Securing Intranet and Extranet Servers. Internet: https://www.msctrustgate.com/pdf/secextint.pdf , 2000, [August 29, 2017].

15- M. Chapple. "Extranet security strategy considerations. Internet: http://searchsecurity.techtarget.com/tip/Extranet-security-strategy-considerations , Oct. 25, 2016 [August 29, 2017].

16-SANS "Extranets: The Weakest Link & Security. Internet: https://www.sans.org/reading-room/whitepapers/.../extranets-weakest-link-security-43, 2001, [September 12,2017].

17- Eke, Helen Nneka "Creating a digital footprint as a means of optimizing the personal branding of librarians in the digital society." *Webology*, Vol. 9, No.2, 2012