ACMW: Access Control Model on Web Environment

Dr. Selma Elsheikh

Abstract—**A Web includes implementations of complex business processes. A key challenge in Web services security is to design access control models that can effectively meet the unique security challenges posed by the Web services environment. Although many access control models have been proposed, there is no much work has been done for finding efficient techniques for performing access control. This paper attempts to propose Web access control model which control the Web access based on the user access behavior by tracking the Web access history. The active user's access pattern is matched with user access data discovered from user access history, based on Web usage data in Web server log data using Association rules mining and prefix-projected sequential pattern mining algorithms.**

Keywords: Access control, Web Access Control, Web usage Mining, Web Mining

I. INTRODUCTION

Access control problems are a key problem for information processing in a World Wide [2], [3]. In this global environment, people often lose control of information, how it is used and to whom it is disclosed [1[, [10] The distribution and sharing of information via the Web require the definition and enforcement of security controls, ensuring that information will be accessible only to authorized entities [3], [7]. One of used access controls is password access control, which requires the user to enter his/her username and password [1], [8]. Although passwords are a simple and effective method of Web security, their security is limited by the combined problems of picking good passwords, password sniffing, and the ease with which people share them. In addition, password based access control is not fully effective and to be used with caution [5], [8]. Even if user's accessibility has been controlled, there are always people who can misuse other peoples' passwords who will create opportunities to obtain access even if they are not authorized or not familiar with using the Web [3], [11]. For example, a Medical research centre provides clinical information systems over the Web. Research projects are conducted and the research database is made available over the internet. These projects' content sensitive data is released only to specific authorized users under specific conditions. Most projects involve human health data collected from hospitals which must be made available to heath care institutions or related partners for research purposes (e.g. Universities) under specific conditions with restricted use.

II. RELATED WORK

Web mining research is active research relates to several research communities such as database, information retrieval and artificial intelligence. Web mining is defined as the discovery and analysis of useful information from WWW, including Web content mining (automatic search of information from on line resources), Web structure mining (discovery of how to model the underlying link structures of the Web), and Web usage mining (discovery of user access patterns from Web log files[9]. In recent years, Web usage mining attracted many research communities [6], [12]. Online mining where activities of new users are monitored online and reaction to misuse of privileges is generated automatically and immediately [9], [12]. ActiveWeb) XML- based active rules for deriving Web views and for defining access control by user access behaviors has been proposed [6]. The access right for a page is given to a user by his user ID and password, IP addresses, the limitation of ActiveWeb is the location or the IP address, which is continuously changing specially in mobile computing. To find out whether a user is exploiting the current server's data/services to publish similar data/services, both Web usage and Web content information were used [9].

The main goal of our paper is to develop a Web access control model based on user access patterns, through mining usage data by tracking The Web access history. The Mining Phases In ACMW goes through Web Usage preprocessing , pattern discovery through association rule mining and Sequential Pattern Mining (using Prefix-projected Sequential pattern mining **(PrefixSpan**) [5], and pattern analysis [4]. This paper begins with a brief description of the related work. Section two illustrates the ACMW security policy assumption . Section three presents ACMW Design Architecture. Section four presents and demonstrates the ACMW Mining Phases. The conclusion is included in the last section.

III. ACMW DESIGN ARCHITECTURE

In the proposed model the data sources are the server log data and the login data. The system utilizes these data through different phases as displays in figure 1. The first phase is preprocessing; the main goal of this phase is to create minable objects for knowledge discovery, through data transformation and integration. The output of this phase is preparing the data. The rule mining algorithms are identified for pattern discovery in phase two. The SQL rational engine is used as an access control enforcement to produce Web access patterns. Phase three

Dr. Selma Elsheikh is with the Faculty of Computing, Alghurair University (P.O. Box 37374, Dubai, UAE; Email: selma_111@yahoo.com)

includes the analysis and identification of the user access patterns. The goal of this phase is to eliminate the irrelative rules and to extract the interesting rules or patterns from the output of the previous phase. The system checks the user access right depending on the Web user access data and access query pattern. In the last phase, the access control decision is made and either allows access or denies it.

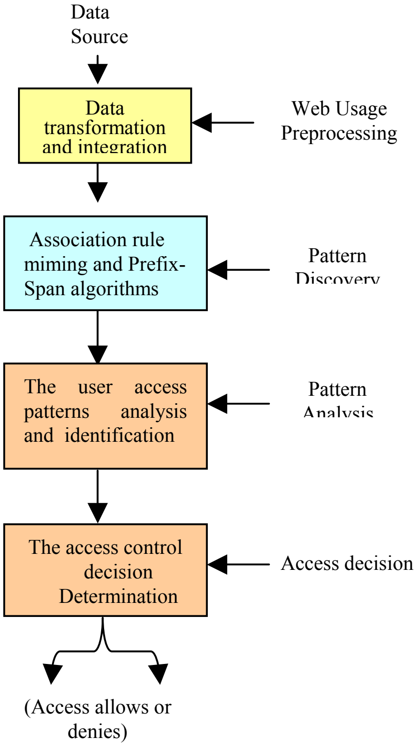
Data Source
Data transformation and integration
Web Usage Preprocessing
Association rule miming and Prefix-Span algorithms
Pattern Discovery
The user access patterns analysis and identification
Pattern Analysis
The access control decision Determination
Access decision
(Access allows or denies)

Figure 1: ACMW Design Architecture

IV. III. SECURITY POLICY ASSUMPTIONS

In ACMW The system administrator sets the system security policy and configures the access permission. The access rights are associated with the objects and the users must ensure that they have sufficient access rights before gaining access to the objects (access the Web). The new user provides his/her login data or registration data. The system administrator after given the user authority to access the specific Web page or pages, make sure the user used his/her access right and browsing the page or pages at lest once.

V. MINING PHASES IN ACMW

Mining Phases In ACMW are Web usage Preprocessing, pattern discovery, pattern analysis, and access decision.

A. *Web usage Preprocessing*

Before any knowledge discovery take place, the Web usage data goes through a preprocessing phase to clean the data from irrelevant or redundant entries. Then the data is formatted appropriately according to application (association Rules mining sequential pattern require the input data to be in different form). The Web usage data which was captured by Web server log file , contains a complete history of file access by clients, include user'IP address, access date and time, requested method ("get", "post", etc), url of the page accessed, data transmission protocol (http), etc.

Some data in the Web server log files are noisy, that may include a variety of image, sound, and video files, etc, can't be used for pattern discovery directly, so they are processed and filtered to transform into more meaningful representation data. The server log files (the access raw data) are converted and integrated with the user registration and then inserted in a database file. In this model four server log file data are selected which are:

User Identification: User password

Date (Date of last request): the time duration from when the user accessed last, e.g. today, yesterday, the day before yesterday, etc.

Page visited: URL of the page visited

Status action: The status field is set by the Web server and indicates the action taken in response to a request. For example (Codes from 200 through 299 indicate success, 300 through 399 indicate some form of redirection, 400 through 499 indicate an error serving the particular request and 500 through 599 indicate a problem in the Web server).

B. *4.2 Pattern Discovery*

In this phase, association rule mining are applied to the formatted data of Web user access transactions entries, and the sequential pattern is used to find the maximal sequences among this data.

1) *Association Rules Mining*

In ACMW the application of association rule mining is to discover the associations and correlations among Web user access transaction entries {user identification, date of the request, page visited (URL), and status action}, where the presence of one set of Web access entries pattern in the transaction implies the presence of others with 100% confidence, and minimum support ≥ 2. Association rule mining is a two-step process:

Step 1: Find all frequent patterns. Each of these patterns will occur at least as frequently as a pre-determined minimum support count (the minimum support = 2).

Step 2: Generate strong association rules from the frequent patterns. These rules must satisfy the minimum support ≥ 2 and minimum confidence =100%.

Example1 (using the above definition):Let X, Y, Z, M $\subseteq$ U where X = user password, Y= Page_requsted , Z =

Date of request , and M = Status action . Using Table 1 (examples of Users Transaction Entries) the confidence means that . If X = Sudan and Y= /.../Research Subject Review Board. asp and Z = 7/30/2007 Then Status action = 200, with confidence is = 100%.

In ACMW the application of association rule mining is to discover the associations and correlations among Web user access transaction entries {user identification, date of the request, page visited (URL), and status action}, where the presence of one set of Web access entries pattern in the transaction implies the presence of others with 100% confidence, and minimum support ≥2. Table 2 shows the sample output of of Running Association Rules (Sample

Table 1: The Output of Running Association Rules (Sample)

Association rules	Conf. (%)	Support (Frequency
Password= samia_338 ⇒ date of last request = 7/30/2007	50%	2
Password= samia_338 And date of last request = 7/30/2007 ⇒ page requested = /../medicalhome/clinical_data.asp	75%	2
Password= samia_338 and date of last request = 7/30/2007 ⇒ page requested = userpages/user_pages/default.asp	75%	2
Password= samia_338 and date of last request = 7/30/2007 page requested = userpages/user_pages/default.asp ⇒ status = 200	100%	2

C. *Sequential Pattern Mining*

In ACMW, Sequential pattern mining for intra-transaction patterns such that the presence of user access transaction entries patterns followed by another in user access transaction. Prefix-projected Sequential pattern mining **(PrefixSpan**) approach was used to solve the problem [6]. These approach uses prefix projection to mine the set of frequent user access transaction entries patterns.

- **PrefixSpan Algorithm**

The algorithms of PrefixSpan are:

```
1. S = Frequent_sequence (D) // (is a sequence of database D of user access transactions
      Entry patterns. ; Minimum support ≥ 2.
2. L= length of S; //
3. $D|_s$ = S-projected database, if S ≠ <>, otherwise = the
      sequence of Database D.
4. The projected database $S|_L$ is created from all sequence
      S
5. Scan $D|_s$ for once to find intra-sequence b if not then
      appended. b to S to form the sequence.
7. For each b frequent sequence, extend prefix b
      accordingly //b can
    Assembled to the last element of D to form a
      sequence;
8. For each frequent item b, append it to S to form S', and
      output S'
9. For each S', construct S'-projected database $D|_{S'}$,
      recursively.
10. The process ends when no new frequent sequence can
      be generated.
11. The complete set of sequential patterns.
```

PrefixSpan (Prefix-projected Sequential pattern mining) algorithm starts by finding all frequent events in the input data (Web user access transaction entries). The search space is then divided by partitioning the sequential patterns into subsets having distinct frequent events as prefixes. These results in the same number of subsets as there are frequent events, the patterns in each subset starting with corresponding event. The subsets are mined by constructing corresponding projected databases and mined each recursively.

Example 1: (using the above algorithm): Based on data in Table 1, the Web access transaction entries sequence for a single user are < samia_338, /userpages/index/ ESRG_project_database.asp, 7/30/2007, 200 >, <.. /userpages/index/ ESRG_project_database.asp 7/30/2007, 200> < 7/30/2007,200> <(200 >. The output prefixes of the subsequence = < samia_338>, <../userpages/index/ ESRG_project_database.asp>, <7/30/2007 200>, and <200>. Figure 3 illustrates the process for sequence selection in the selected example, there are 4 frequent items have different prefixes, and projection continues based on those prefixes.

D. *Pattern Analysis and Access Decision*

Pattern analysis tools are used to convert discovered rules and patterns into knowledge. Structure Query Language SQL eliminates the irrelevant rules or patterns and provides the user control over the data mining process and allow the user to extract only relevant and useful patterns.

The example below presents a complete representation of the SQL rule // Select Password, date-last-request, $page_{visitited}$, status //

```
SELECT Password, date-last-request, $page_{visitited}$, status.
        If login password= user password,
    Then
        If date-of- last- request = date_of _last- request
then
          If $page_{requested}$=$page_{visitited}$ Then
             If status action = successful access
               Activate the page link ( page access
   permitted ).
          With minimum support=2 and confidence of the
   rule is 100%.
```

The ACMW make access decision based on the algorithms and procedures developed in the previous phases. Figure 3 illustrated ACMW flow chart decision steps. The input is the user access transaction data and the

Web server log data. The output is the user access control decision.

The descriptions of the steps are as follows:

Step 1: starts the user access transaction
Step 2: page access request (Web access)
Step 3: input user password
Step 4: get the password from user profiling database
Step 5: compare and verify the password
Step 6: check user access transaction entries pattern sequence
Step 7: get date of last page visited URL, status action and match the patterns
Step 8: if pattern matching then step 9 or else step 3
Step 9: evaluation and make access decision
Step 10: if access permitted then display the page
else go to step 1
Step 11: end

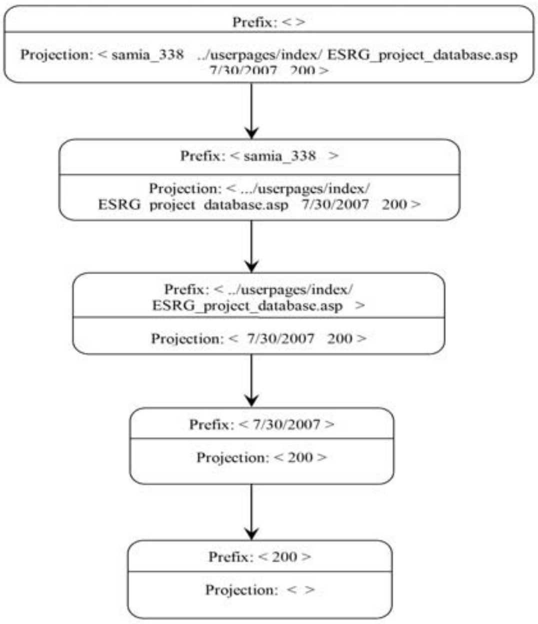
Prefix: < >
Projection: < samia_338 ../userpages/index/ ESRG_project_database.asp 7/30/2007 200 >
Prefix: < samia_338 >
Projection: < .../userpages/index/ ESRG project database.asp 7/30/2007 200 >
Prefix: < ../userpages/index/ ESRG_project_database.asp >
Projection: < 7/30/2007 200 >
Prefix: < 7/30/2007 >
Projection: < 200 >
Prefix: < 200 >
Projection: < >

Figure 2: The Sequence Selection Process
in the Selected Example

VI. Conclusion

Association rule mining and Prefix-projected sequential pattern mining (PrefixSpan) algorithm was used to discover frequent subsequences as patterns in a sequence Web user access transaction entries to find the Web user access transaction entries pattern. In the paper, we have attempted to develop a Web access control technique that has the ability to determine user access permission (either allow or deny) based on an identity plus a set of attributes associated with user access behavior by tracking his/her access history. Much work should be done on the problem of reliability of the Web server log data, which can be used as a knowledge discovery to identify the user access behavior in a fine- grained way.

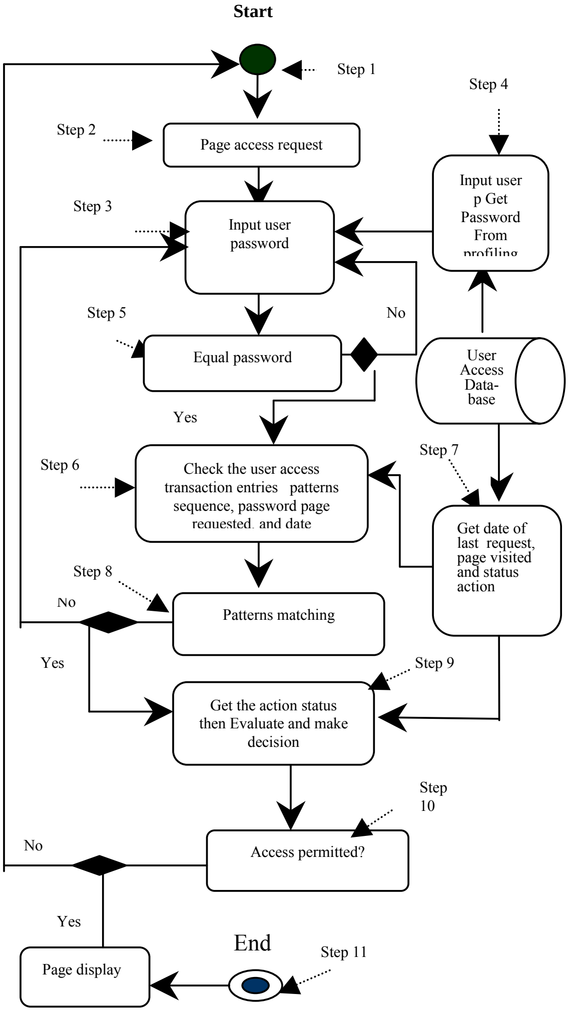
Start
Step 1
Step 4
Step 2
Page access request
Input user p Get Password From profiling
Step 3
Input user password
Step 5
No
Equal password
User Access Data-base
Yes
Step 6
Check the user access transaction entries patterns sequence, password page requested, and date
Step 7
Get date of last request, page visited and status action
Step 8
No
Patterns matching
Yes
Step 9
Get the action status then Evaluate and make decision
Step 10
No
Access permitted?
Yes
End
Step 11
Page display

Figure 3: ACMW flow chart for access decision.

VII. References

[1] A. Gal, V Atluri, "An Authorization Model for temporal Data," ACM Transactions on Information and System Security , vol. 5, no. 1, Feb. 2002.

[[2] B. Atkinson, G. Della-Libera, S. Hada, M. Hondo, P. Hallam-Baker, C. Kaler, J. Klein. LaMacchia, P. Manferdelli, and J. H Maruyama. (2002, April 5). Web Services Security (WS-Security) Version 1.0, Available: http://msdn.microsoft.com/webservices/default.aspx?pull=/library/en-us/dnglobspec/html/ws-security.asp.

[3] E. A. Selma, "Access Control Scheme for Web Services", Proceeding of the International Conference on Computer and Communication Engineering

(ICCCE08), Kuala Lumpur, Malaysia, 13-15, May, 2008.

[4] J. Han, and M. Kamber. "Data Mining: Concepts and Techniques," San Francisco, Morgan Kaufmann Publishers, 2 001.

[5] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, and Q. Chen."Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach," *Journal of Transactions On Knowledge And Data Engineering 10(16),* 2003.

[6] H. Kiyomitsu, A. Takeuchi, and K. Tanaka, "ActiveWeb: XML-Based Active Rules for Web View Derivations and Access Control". Pages 31-39. Proceedings of IEEE
Conference on Internet and Computer Security (ICSC '01), Las Vegas, USA 2001.

[7] L. Bauer, S. Edward, and , W. Felton. "A General and Flexible Access-Control System for the Web, Secure Internet Programming Laboratory," *in Proc. 11th USENIX Security Symposium, Department of Computer Science. San Francisco, USA,* 2002.

[8] E. A. Selma Mohamed Daud, Mohd Zohadie Bardaie and V Barkash, "Web Access Control: New Paradigm, " Brunei Darussalam Journal of Technology and Commerce, Vol 12(5) Nov.-December, 2005.

[9] M. Mahoui, B. Bhargava and M. Mohania, "Data Mining For Web Security: UserWatcher," Proceedings of the IC'2001 Conference, Las Vegas, USA 2001.

[10]P. Bonatti, P. Samarati, "A unified framework for regulating access and information release on the web", Journal of Computer Security, Vol. 10 No.3, pp.241-72, (2002).

[11] R. Bhatti, J. B. Joshi, E. Bertino, and A. Ghafoor. " Access Control in Dynamic XML-based Web-Services with XRBAC ," in Proc. 1th International Conf. Web Services, Las Vegas, USA, June 23-26, 2003.

[12] R. Kosala, H. Blockee. Web Mining Research: A Survey. SIGKDD Explorations, July 2000.

.