Access-Control Architecture to Support E-CRM and Distributed Data Mining

Simon Fong, Zhuang Yan
Faculty of Science and Technology
Universidade de Macau
Av. Padre Tomás Pereira S.J., Taipa
Macau SAR
{*ccfong@umac.mo syz@umac.mo*}

Serena Chan
e-Business Development Team
Omron Asia Pacific
U.E. Square
Singapore
serenachan@omron.com.sg

***Abstract* – This paper presents an Access-Control (AC) architectural model for Distributed Data Mining (DDM), which is tailored to meet the needs of online Customer Relations Management (CRM) where the services are personalized and the web content delivery are targeted to specific segments of customers. This hybrid AC architecture integrates the client-server and the mobile agent technologies that can easily be upgraded from most existing web-based platforms. The AC model focuses on the online e-personalization (for CRM) and web intelligence issues (for decision support), which are particularly relevant in the context of making a website customer-service oriented while enabling top management to gain insights from the operation of the website. Technically, a software agent architecture that consists of functions of access-control, customer-profiling, dynamic web pages creation, data-mining and a rule-based system is employed. Also in this paper, we argue that not only we can provide richer business intelligence by profiling customers than by merely mining raw web-logs for the traffic statistics, a new paradigm of business insights and e-CRM functions can be obtained and deployed respectively, by combining the customers' profiles and traffic statistics under this AC architecture.**

I. INTRODUCTION

Nowadays, most companies have moved into e-business where they use Internet as a platform to interact with their customers. This has created more touch points than the traditional face-to-face visits, and therefore we have a new realm of approach called e-CRM to deal with. Similarly to the traditional CRM, e-CRM is a methodology used to build and maintain a loyal customer relationship online and to discover the most profitable way to build that relationship through e-mails, interactions at the website, and personalizing web contents for customers. It is an ever ongoing process of acquiring, retaining and growing profitable customers via the Internet.

Specifically, e-CRM adheres closely with three business areas: decision support, personalized service, and customer management. One must have a very good understanding about the operation of the business as well as the customers, so that sound decisions can be made by the management on how to run a successful e-CRM. Since e-Business is like any other business perhaps at a faster pace, its business information evolves and changes continuously. Business intelligence must be obtained timely in order to remain effective. A more specific kind called Web intelligence will be discussed later on.

In e-CRM, the two business areas that operate at the front-line are Personalized service and Customer management. Personalized service is about providing better customer service on the website. This is done by presenting the right information to the right customers. Users can easily access to the information that they want, receive the right piece of recommendation, etc. for example. So that would enhance the user's experience on the website, hoping that they will like to visit the same website again. The latter business area that e-CRM covers, is in general called Customer management that operates through the touch-points which the customers communicate with the company on the Internet. Instead of passively serving the customers by customizing the web content for them to view, Customer management proactively contact them on the web via emails, e-newsletters and or an interactive help-desk. Sales force automation may also fit into Customer management. It is however beyond the scope of this paper as we focus on e-business services. Together, personalized service and customer management help retain good customers, win their loyalty, and may attract new ones by earning word-of-mouth from existing satisfied customers.

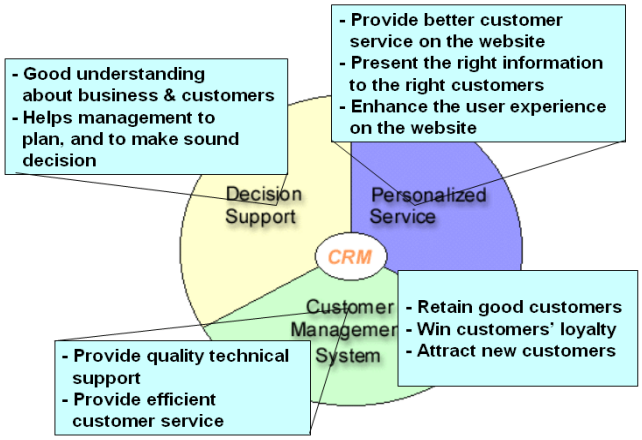

Figure 1. e-CRM covers 3 business areas and their functions

In general, for targeting the business areas of decision support and personalized service, we have identified three main technologies namely Data-warehousing (that includes data-mining modules and decision support modules), Web log analyzer [1] and Web spy [2]. Each one of them provides certain insight or understanding of one of the many aspects of a business respectively, as shown in Figure 2. In particular, we state that having a web log analyzer alone is not sufficient to provide a complete picture of business intelligence. Nevertheless, web-log analysis gives very good detailed reports on web site traffic statistics. Other information such as competitors, market positions, and most importantly the customers and their behaviors on the website needed to be known. In our case, e-Personalization that means personalizing web-content according to the customer's interest/needs will be guided by the knowledge that we have on him.

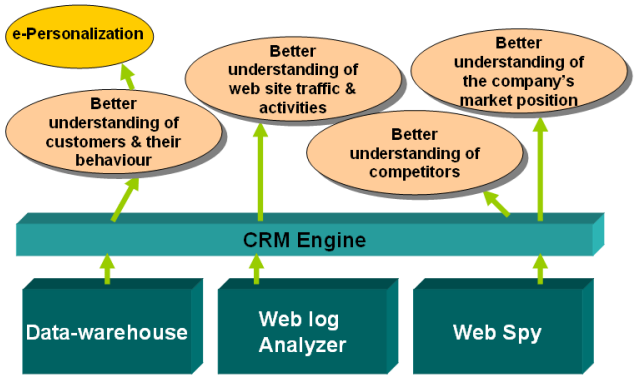

Figure 2. The conceptual view of the e-CRM architecture and the objectives

To realize the objectives as depicted in Figure 2, we propose an Access-control (AC) architecture that can progressively built upon a basic website. The objective is to distribute the functions of e-CRM and business intelligence using an agent-architecture and to monitor users' accesses on the website. Details of the AC architecture, functions of e-CRM and web intelligence are given in the subsequent sections of the paper. A Conclusion is drawn at the end

II. HYBRID AC ARCHITECTURE

We present in this section our hybrid model for e-CRM and data-mining for business intelligence, that is tailored to meet the requirements for easy integration and upgrade for some primitive web-server to operate in e-business environments. The distinguishing features of this architecture are the integration of the distributed software agents each takes a specific role for producing a particular result, and the users' access monitoring system on the website. It supports the ability for mining to be performed at remote sites using mobile agents provides insightful business intelligence reports, and it also feeds back the results to some rule-based systems residing at the server for doing e-Personalization. This helps it to be easily adopted by most existing web-sites who want to have complete functions of e-CRM and BI, but yet reluctant to replace the whole e-business platform.

The hybrid model operates on the principle of letting the website with only some slight modification to act as the portal for capturing the user's particular, their trails, and also for providing them a better online service via e-Personalization. Thus, it has the option of using the client-server model or the mobile-agent model or an integrated approach involving both. We suggest that a web business should evolve progressively by taking a phrasal approach. The difference in performance between the three models are how fast the BI can be obtained and how much real-time the e-CRM can take place.

The components of the hybrid AC architecture illustrated in Figure 3 are as follows:

e-Business website

This is a general term that consists of a web-server and a personalization server. At the simplest form, the web server hosts the website and handles HTTP requests from users. This is the most fundamental platform from which we can implement distributed data-mining and e-CRM by upgrading it to an agent-based architecture. The web server, however, is responsible for capturing users' profile, storing submitted online forms and web log which will then be transported into a data-warehouse. The e-Personalization server is responsible for generating dynamic web-pages for the users.

e-CRM server

This is the server that stores the operational data into a data mart. It serves as a data repository for data-mining agents to use.

OLAP agent

It produces multi-dimensional data-cubes for information browsing. The resultant data-cubes could be integrated on a separate Intranet or extranet server, so that authorized personnel could access to the business information in multi-dimensional views in real-time.

Web Alert agent

Its main role is to alert designated personnel in the company should exceptional events happen. For examples, if the values on any data channel exceeds or falls below a pre-defined threshold, some selected people in the company would be alerted by emails and/or short-messages on mobile phones. Other examples are when a web-server is down, a 404-error is detected, or a service is denied. This gives an option of providing a means for alarming urgent events.

Web Spy agent

As the name suggests, it spies on the online competitors' websites, retrieving information such as prices, updates, or new product news, as well as their market positions on the search engines.

Data-mining agent

It can be seen that with all the components working together, the three business areas are covered, as shown in the figure. The highlight of this paper is on about how a login and access control mechanism is implemented on the web site; thereafter enhanced business information as well as e-CRM can be realized.

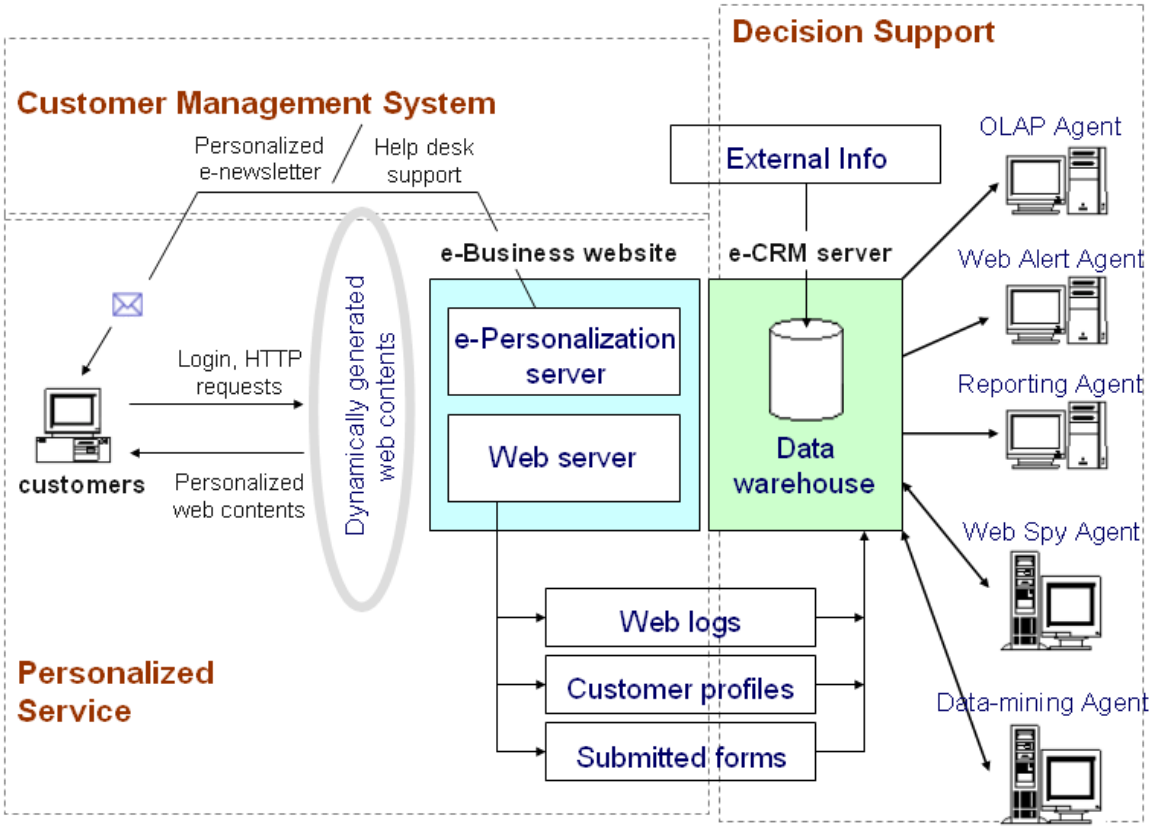

Figure 3. Agent-based Framework for BI and e-CRM.

III. ACCESS-CONTROL MECHANISM

AC mechanism is a prerequisite for implementing an e-CRM enabled website with customer-centric business intelligence. It works, basically by requiring a user to login, in order to identify himself before he is granted access to certain protected resources. It is assumed that a user would think it is fair that they would have to be bothered about logging in for getting some invaluable resources in return. At the first time, he would have to sign up a registration form with a user-name and a password of his choice.

On the website, the hypertext links to the protected resources such as a free-trail of software, a product's manual, etc., would have to be modified. Upon clicking on any one of these protected hypertext links, not only a user would be challenged with a password for checking access rights, the event of that "click" will be recorded in a database. Specifically, we record down who, at what time, on which website (in the case of a multi-regional website), have accessed to what resources from which web page. This recording event happens from when the user has logon (both success failure cases) to every time he clicks on a protected link. The idea of doing so is that we want to be able to keep track of which resources have been accessed by whom, when, and how often, on our website. Hence we can derive insights on what is happening on the website rather in terms of accesses than merely traffic hits per visit. The information obtained are more insightful than those from web-logs, since we beforehand have kept a full demographic details about the users when they registered, compare to the information obtained from web-logs which are only IP addresses. We will later discuss about how the records of user access will be analyzed for web business intelligence. The technical implementation flows for login procedure and thereafter access procedure are shown in Figures 4 and 5 respectively. For both cases, a record is generated in two different database tables namely login and access.

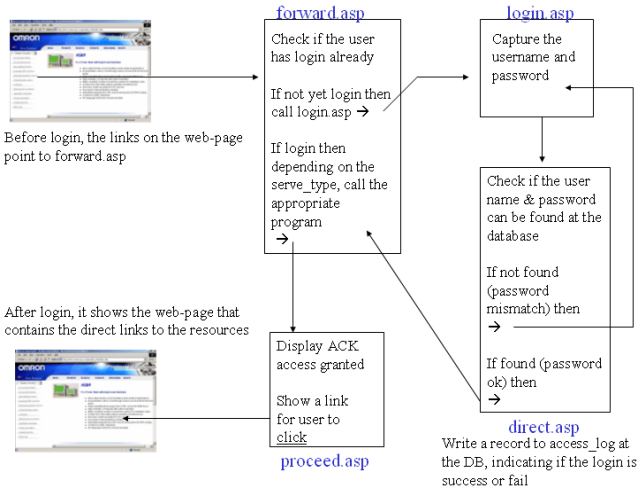

Figure 4. Logical flow for login procedure

The AC architecture has three distinguishing features:

(1) Instead of using web log data "GET" that is relative only to the name of a web page, we can more accurately pinpoint which "resources" on the page that we want to monitor. It can be a link, a JPG file, a PDF file, or a movie clip. Access to them requires the user to login. We have the flexibility to design which links require the users to login and therefore we can monitor the access patterns on them.

(2) Users can be uniquely identified by their login ID

(3) Before the analysis, we can have relatively clean user profiles because the data fields of an access record are carefully chosen.

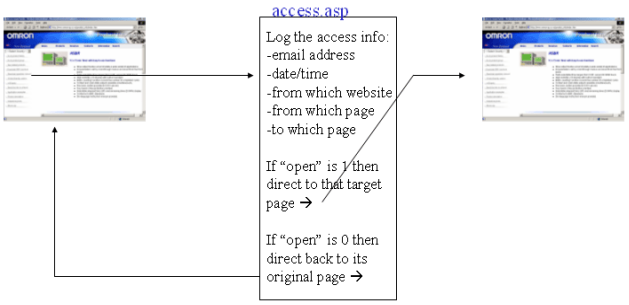

Figure 5. Logical flow for access procedure

IV. ACCESS-CONTROL MECHANISM

Web business intelligence refers to the ability to make better business decisions in running a website through intelligent use of collected data, mostly from the same website. It is about gathering, managing, and analyzing data. Then the data are transformed into useful, actionable information to run an information-driven e-Business.

With the AC mechanisms implemented, at the appropriate links at which the company wants to monitor about their popularity, a wealth of good business intelligence (BI) can be obtained. In essence, we would be able to obtain a new realm of BI with the information about when and what the users have access and login to the website. We suppose that the web business intelligence can be obtained at two levels; one with and the other one without the aid of data-mining.

At the first level, where simple statistics are used, we can show the following information as business intelligence about the products and the website:

(1) *Customer profiling* (demographics). When integrated with web traffic, one can tell the online customers behaviours. However, this requires some advanced techniques like web log mining. At the first level, for example, we can tell about the top 10 (or top 5, 100, etc) of the countries, the industries where the users come from, or their designations;

(2) *Customers' interests*. The top customers' interests on the products can be listed out. In a simple way, this information is obtained from the registration form when the users signed up. There is other more sophisticated technique to estimate the user's interest of a web page without directly asking the user. This method [3] is proposed for locating multi-word phrases in order to estimate interests on certain pages.

(3) *Product popularity*. The web pages of products that have been accessed most could be ranked and displayed as vital business information. This is a relatively fair approach when compare to measuring web page hits as each page may have different hit count per access

(4) *Technical knowledge base*. Over a certain period of time, the online inquiry forms posted by the customers have been collected, organized that form a searchable technical knowledge base. The base can made to be searchable that stores the past problem cases and the appropriate solutions. In a simpler form, it could be provided to the users as a set of frequently asked questions.

V. PATTERN DISCOVERY UNDER AC ARCHITECTURE

Web usage mining, also known as pattern discovery is a process of mining for user browsing and access patterns. With our AC architecture, mining is an important backend process for deriving business intelligence at the second level. There are a number of issues in pre-processing data for mining that must be addressed before the mining algorithms can be run.

5.1 Preprocessing Task

The task of data cleaning to clean a server log to eliminate irrelevant attributes is important to any analysis, especially data mining. This task is somewhat different from those used for cleaning web log data. For web log mining, it is necessary to determine which items are important accesses that we want to analyze, and eliminate those that are reluctant from a large pool of raw web log entries. Making sense out of them requires sorting out of filename suffixes, distinguishing IP addresses, machine names and temporal information to identify users. Mechanisms such as local caches and proxy servers can severely distort the overall picture of user transversals through a web site. Current methods to try overcoming this problem include the use of cookies and cache busting, checking repeated user-requests [4] and measuring the lengths of users' sessions [5]. Special techniques and a substantial amount of efforts are required. As detailed in [6], none of these methods are without serious drawbacks. Furthermore, a problem associated with proxy servers that of user identification. Use of a machine name to uniquely identify users can result in several users erroneously grouped together as one user. Perhaps using user registration can better identify a user. But yet the issue of linking user accesses to user identification remains a problem.

In AC environment, we record only those user accesses taken from the links that we chose. It is assured that the user must have logged in, given us his unique user-id and then granted the access. So each access record stored in the database contains vital information of the user's identification, time of access, the website's identity, from which page he is making the access, and which "item" he is accessing. The item or the place of tracking can be a graphics, a movie clip, a PDF file or a hypertext link. Each user-id can be easily linked to the user registration table in the database; that implies a full set of information about that user is available. The problems of user identification as with web log processing no longer exist. We can take for granted that the user is readily identified as he has passed

the password challenged. The recording of his trail is made almost instantly by our program code embedded at the link as soon as he mouse clicked on it. Besides the flexibility of choosing whichever resources we want to monitor, the biggest advantage is that the effort for data pre-processing is minimal. This AC method is believed to be more reliable because there is no missing or ambiguous value, and no need to remove any reluctant information. However, this requires a good design in setting up the links that we want to trace in advance. A non-technical drawback however, would be the difficulty in luring the users to sign up, and then get them perform the login action at every first time he wants to access a resource during a session. The value of the resources that the user seeks must be able to justify the trouble of login.

5.2 Discovery Techniques on Web Transactions

There are several kinds of access pattern mining that can be performed on the Access records, such as path analysis, discovery of association and sequential patterns, and clustering and classification. They have different features and are suitable for generating different aspects of business intelligence, as shown in Table 1. Each one of these management approaches covers many principles. We only chose to discuss a few of them that are relevant to our AC architecture due to space constraint.

Sequencing	If they buy Product A, will they buy Product C or Product X and when?
Segmentation	What differentiates my most loyal clients from all the others?
Profiling	Who are my most profitable customers and how do I keep them?
Association	What relationship does visitor gender have to sales at my site?
Classification	How do I recognize high propensity to purchase visitors?
Clustering	What attributes describe high return visitors to my site?
Prediction	Who is likely to be my most loyal customer over time?
Optimization	How do I design my site to maximize my online sales?

Table 1. Discovery techniques for different analysis.

5.2.1 Path Analysis

For web business, the most common approach is path analysis. It could be used to determine the most frequently visited paths in a web site. A graph represents some relation defined on web pages or other objects. Usually a graph is used to represent the physical layout of a web site, with web pages as nodes and hypertext links between pages as directed edges. The graph in the following figure has generalized the web pages as nodes of the graph representing two main categories: Products and Services that have their own sub-nodes PG (Product Groups) and S (Services) respectively.

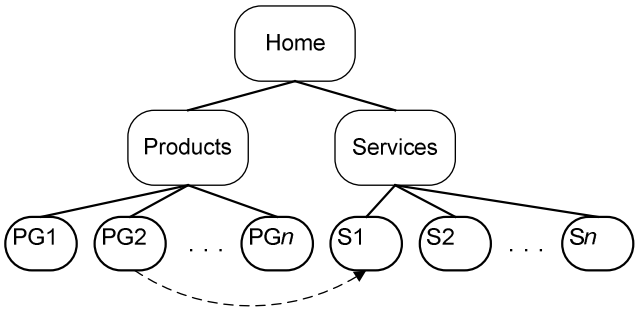

Figure 6. Tree graph representing the structure of a web site for monitoring users' navigation.

Some examples of information that can be discovered through path analysis are:

- 40% of users who accessed the site started from Home->P->PG3->S301 (/products/sensors/catalog.pdf);
- 30% of users who accessed PG4 (/products/cables) did so by starting at S2 (/news/promotion), and proceeding through S5 (/service/selection_guide) and PG3 (/products/sensors);
- 80% of users log out or left the site after five or less link accesses.

The first example shows nearly half of the users are clicking into product group sensors and looked for a catalog PDF file. It might be a good idea to add a shortcut link on the home page for users' convenience. The second example indicates that the user found some useful information at the promotion section, then made a route through some catalogue type of information, and finally found the product. Such pattern suggests a cross-link from the promotion directly to the product should be implemented. The last example represents an average length of visit for the users. Since many users don't browse further than five pages at the web site, it would be prudent to ensure that important information is contained within five pages of the common site entry points. For example, a separate page of directory type called e-Download that contains links to the important resources should be implemented.

5.2.2 Association

Association rule discovery techniques [7, 8] are usually applied to records that have a set of attributes. We are trying to discover the associations or the correlations among the AC data records where the presence of one set of attributes in a record implies the presence of other attributes with a certain level of confidence. The following examples show some correlations discovered by using association rule discovery techniques:

- 70% of users who accessed PG3 (/products/sensor), also accessed PG5 (/products/vision_software);
- 20% of users who accessed PG9 (/products/new_release), placed us an e-quotation on S8 (service/sales).

Discovery of such information is important in the development of effective marketing strategies. In accordance to the strategies, the website would have to be organized in a logical and easy-to-use way.

5.2.3 Sequential Patterns

Discovering sequential patterns is about to find the inter-access patterns such that the presence if a set of attributes is followed by another attribute in the time-stamp ordered access set. In AC access records, an access made by a user is collected over a period of time. By analyzing sequential patterns, we can determine temporal relationships among data sets such as the following:

- 70% of users who visited PG3 (/products/sensors), after received our email newsletter within one week;
- 80% of users who downloaded our PDF catalog on PG7 (/products/switches); also download our PDF catalog on PG4 (/products/timers) within 2 days.

The techniques described in [9, 10] would also be applicable on our AC records, using the temporal characteristics of the data. These techniques can find common characteristics of all users that accessed a particular file within some period of time, that can be a day, a week, etc.

One feature that most website organizations concern about is the ability to detect change of interest from the association and sequential patterns. Change of interest could be revealed as patterns that have strong temporal dependency among some product groups. In order to facilitate this kind of detection, the tree that represents the physical layout of the website needs to be modified. As shown in Figure 7, the graph has been arranged in such a way that the product groups become the main nodes from which the service sub-nodes are stringed. Here we assume the detection is solely on the interests of users on products; so that the interests on services become secondary. The nodes for the same types of services are moved under each product group because we assume that the same services provided would be covering different products. For example, an online inquiry service provided would be available for the users to inquire about any model of products.

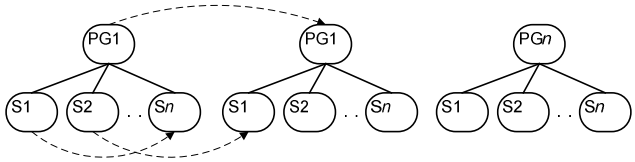

Figure 7. Modified tree graph for detecting change of users' interest on products.

So for the detection, it can be observed that the dashed lines represent frequent access from one product to another. We can also easily tell if the relationship between the PG though the direct links are with their attached Service nodes. The tree is now having the main nodes representing a product group. If most of the edges are going within the same group, then we assume that the interests of the user remain unchanged. Otherwise, there is a deviation of interest (with a confidence level) when the many edges are branching out to the other nodes. We can also use this method to find related product groups by detecting any strong linkages among the nodes

5.2.4 Classification

There are techniques as described in [11, 12] for sorting a set of items from the AC records and/or the registration profiles into a particular group according to their common attributes. Usually for web business, classification techniques enable one to develop a profile of users who access certain files/resources based on demographic information available on those users, or based on their access patterns. Some example relationships that can be discovered are as follow:

- Users from PCB manufacturing industry who visit the web site tend to be interested in the product group 17 (computer/pcb_inspection_sys);
- 70% of users who have downloaded trial software from PG9 are engineers and work in Malaysia.

5.3 A Generalized Data-mining Model with AC

The analysis mentioned above show different strength in different aspects. With our AC model, we attempt to generalize the discovery attributed to the following format:

$$who \Rightarrow when \Rightarrow how \Rightarrow visited\ which\ parts\ of\ the\ website \Rightarrow for\ what \Rightarrow why$$

Obviously, the format contains pivots of natural clauses such as who, when, what, etc. These pivots would be easily derived from the attributes or fields of the AC records, except the last two: for what and why.

Who: attribute of the user-id in AC record. Knowing the user-id, the rest of the demographic information about that user can be obtained from the user registration records.

When: time stamp information in AC record. A hierarchical time dimension model will need to be created later on, e.g. year, quarter, month, week, etc. prior to mining.

How: this information can either be a case in the form of patterns obtained through the analysis discussed in the previous section, or a simple path formed by the "From" and "To" fields in the AC records. Other abstract forms of taxonomical representation could be "through a promotion", "from a link in a newsletter" or "from a search engine".

Which parts: this is usually the destination information in the AC records plus the preceding trails if any. The destination information could be a single resource file or a page that is obtained from the "To" field of the access table. The preceding trails would have to be obtained from Path Analysis. Each one of the preceding trails is a paths a user during one visit has traversed within the physical web site layout.

For what: this is tacit information inferred from the results of Association and Clustering analysis. Alternatively, an analyst may have a conjecture for the

groups and explicitly set a value for this variable with his subjective judgment. In an information-only website, the ultimate purposes for the users may be to obtain certain information in the form of a report, a product brochure, some free software download or to post a technical question. However, in addition to those purposes, the users may want to purchase something in the case of an online shopping website.

Why: this is the concluding remark that an analyst draws for the patterns. This is a continuous review process with largely human intervention since the data-mining system would not be able to explain why such has happened. It would only reveal the hidden patterns from a large pool of data. The explanation of the outcomes would have to be first answered by a human expert. Subsequently, prediction using a decision tree will be possible as an extension to this model.

So examples according to our experience by using the generalized data-mining model with AC are as follow:

Mr Lee, at around noon, every weekday, by going using a link from yahoo portal, visited the latest product news section, to read the latest development on photoelectric sensors

Mr. Ong, after he has purchased a programmable terminal, by going directly to into the download section, downloaded the latest driver for WindowsXP, because he needs it

When the statements are fit into the generalized format, they could be displayed using OLAP, data-cube information model and pivot-table reports [13]. The format expresses out orderly the conditions in the form of who, what, when, etc. The conditions or the constraints would allow us to build a querying system to work on the knowledge that has been extracted by the mining process. An SQL-like querying mechanism can be adopted in our AC model.

As a result, we can reveal how many cases are similar to the examples shown above; see how easy people come to find that information; see what they need recently; and see why they need it. In any case, if there is something not operating smoothly in terms of usability, we can do something to rectify so. For example if we see there are too many cases that people are clicking through a long way in order to find a same piece of information or to obtain a same resource, we may then consider putting an URL shortcut to that hot resource on the front-page. This will add convenience to the users, which in effect will yield user satisfaction.

VI. ONLINE E-CRM

The web mining techniques as described above are essential for maintaining e-CRM strategies: These techniques discover hidden pattern and relationships within web data for the three main marketing actions [14]:

1. Discovering association rules for customer attraction,
2. Discovering sequential patterns for customer retention
3. Discovering classification rules and data clusters for cross-selling. These three techniques however have to rely on e-Personalization that is discussed as follow.

6.1 Customer attraction with association

The two essential parts of attraction are the selection new prospective customers and the acquisition of the selected potential candidates. The e-CRM strategy to perform this task is to find common characteristics in already existing visitors' information and behavior for the classes of profitable and non-profitable customers. These groups (e.g. "no customers", for browsers who have logged in, inquired a lot, but never purchase, "visitor once" and "visitor regular") are then used as results for a classifier to discover online marketing patterns, which are applied online on new site visitors. Depending on the outcome, a dynamically created page is displayed, with contents depending on found associations between browser information and offered products/services.

6.2 Customer retention with sequential patterns

Customer retention is the step of attempting to keep the online customer as loyal as possible. The scheme is similar to that of acquisition, which is dynamically creating web offers based on association but by considering across time (sequential patterns). The discovered sequence can then be used to dynamically display special promotions or personalize message after a certain page sequence has been visited.

6.3 Cross-sales

The goal of cross-sales is to horizontally and/or vertically diversify selling activities to an existing customer base. In order to increase sales of other products along with existing deals with existing customers, the technique of detection of interest (as mentioned above) and association on products are used.

6.4 Proactive Customers Assistance

As the login mechanism is enforced in the AC environment, users would have to be challenged by password at the first time he accesses a resource. Quite often the users may have problems logging into the website because he failed to remember the password. Most websites would provide a link called "forget password?" that allows users to contact the webmaster for the forgotten password to be mailed back to them. However, according to our experience, only a small portion of the users would do this. Therefore we propose a proactive method for customers help such that the user who had problems in logging in will be contacted proactively by our help desk. The scheme is to detect the users who have problems logging in from the AC records. The detection algorithm will filter out only the problem records for the customer service officers to take actions.

The logic of this detection basically consists of these two sequential checks: check if the login of the same user is failed consecutively x times, and then check for the absence

of some records of this particular user that have successfully logged in the next y days. The x and y variables can be adjusted accordingly by the webmaster.

6.5 e-Personalization

The objectives of performing e-Personalization are very much consistent with e-CRM strategies. In essence, it is about turning visitors into customers and generating customer loyalty; and allow customers get exactly what they want so to try retaining them with their satisfaction. Technically, it is all about presenting the customized user interface for different users or different groups of users. Customized interface aspects that can be personalized include:

1. The name of the customer uses; a personalized e-CRM system can recognize users and call them by name wherever possible on the website. As soon as the user logs in, we can find his record and could have all of his details.
2. The content available; content can be customized according to products the user has or as specified by the customer, so customers have the same apparent channel for support for any and all products they purchase
3. The look-and-feel and branding; create different interfaces for different targeted demographic groups
4. Easy form fill-up; customer's particulars can be automatically pulled from the database and put into most fields of the form after he has logged in.

Coupling e-Personalization and data-mining is a vast topic. Other e-CRM strategies such as e-Forum, e-Consultation, and e-Marketing are also employed in our website. Their details are described on another paper due to space constraint. The following table summaries the coverage of e-CRM strategies by some data mining techniques.

Market basket analysis	In large sales transaction databases, find regularities in customer purchase behaviour in order to optimize store layout, promotional spending, etc.
Cross-selling	Among your current customer base, determine those people who are most likely to have latent interest in a cross-selling offer.
User segmentation	Find homogeneous groups of customers who possess similar characteristics in terms of a number of given criteria.
Response modelling	Involves building a model to predict which customers (who?) are most likely to buy a specific product (what?) at a specific moment in time (when?).
Target selling	Involves finding out what products your customer is most likely interested in.
Risk assessment	Determine the risk profile of a customer so that all necessary precautions can be taken in order to make the relationship as profitable as possible.
Churn management	Based on customer satisfaction data, find customers who have a high probability of defection in the near future and build a dissatisfaction management campaign in order to retain those customers.
Product assortment	Optimize product assortments and improve category management.

Table 2. Application of data-mining techniques

VII. CONCLUSION

For companies who are running e-Business, determining and enhancing the effectiveness of their sites is a vital business function, and the ability to accurately measure the traffic to a site and identify visitors' activities is critical to this function. Our Access Control (AC) technique is based on this key innovation: obtaining data from the resource links at the web server instead of web-server log files. It is a password-based forcibly method for capturing customers' trails. It is based on the assumption that the users are willing to login in exchange of some valuable resource from the website. By virtue of this innovation, AC architecture provides better information than log-file analysis, gives site owner more reliable business intelligence. Web log tells mostly about traffic statistics and treating each user anonymous (because it is difficult to identify a user by its IP address). Access patterns with unique user identification and exact navigated locations, would be more suitable for data-mining for customer-centric information and for providing e-CRM services. In essence, there are three highlights for AC architecture: (1) upgrade a conventional website to AC-enabled architecture as shown in Figure 3, (2) strategically plan and set login-protection on the links/resources on the website, (3) perform data-mining to order to aid e-CRM. An agent-based model has been proposed to support AC website architecture

VIII. REFERENCES

[1] WebTrend Weblog Analyser, "www.webtrend.com"

[2] S. Fong, A. Sun, K. Wong, "Price Watcher Agent for E-Commerce", Int. Conference on Intelligent Agent Technology, World Scientific, pp.294-300, 2001.

[3] P. Chan, "A Non-invasive Learning Approach to Building Web User Profiles", KDD-99 Workshop on Web Usage Analysis and User Profiling, pp.7-12, 1999.

[4] P. Pirolli, J. Pitkow, R. Rao, "Silk from a Sow's Ear: Extracting Usable Structures from the Web", Proc. of 1996 Conference on Human Factors in Computing Systems (CHI-96), Vanoouver, Canada, 1996.

[5] R. Cooley, B. Mobasher, J. Srivastava, "Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns", Technical Report TR97-021, University of Minnesota, Dept. of Computer Science, Minneapolis, 1997.

[6] J. Pitkow, "In Search of Reliable Usage Data on the WWW", 6th Int. World Wide Web Conference, Santa Clara, CA, pp.451-463, 1997.

[7] R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules", Proc. of the 20th VLDB Conference, Santiago, Chile, pp.487-499, 1994.

[8] A. Savasere, E. Omiecinski, S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases", Proc. of the 21th VLDB Conference, Zurich, Switzerland, pp.432-443, 1995.

[9] H. Mannila, H. Toivonen, A. Verkamo, "Discovering Frequent Episodes in Sequences", Proc. of the 1st Int. Conference on Knowledge Discovery and Data Mining, Montreal, Quebec, pp.210-215, 1995.

[10] R. Weiss, B. Velez, M. Sheldon, et al, "Hypursuit: A Hierarchical Network Search Engine that Exploits Content-link Hypertext Clustering", Hypertext '96: The 7th ACM Conference on Hypertext, 1996.

[11] J. Han, Y. Cai, N. Ceroone, "Data-driven Discovery of Quantitative Rules in Relational Databases", IEEE Trans. on Knowledge and Data Eng., Vol.5, pp.29-40, 1993

[12] R. Wiess, C. Kulikowski, "Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems", Morgan Kaufmann, San Mateo, CA, 1991.

[13] J. Gray, A. Bosworth, A. Layman, H. Piraheash, "Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-tab, and Sub-total", IEEE 12th Int. Conf. on Data Engineering, pp.152-159, 1996.

[14] J. Srivastava, R. Cooley, M. Deshpande, P. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, Vol.1, Issue 2, 2000.