Title: INTERNET PROFILING SYSTEM AND METHOD

Abstract: An Internet profiling system and method allows identifying and associating a user's previous Internet sessions with the user's current session. A session data stream is used to create a user's Internet session fingerprint pattern including unique strings. The fingerprint pattern for a current session is compared against previous fingerprint patterns maintained in a database to associate Internet inter-session data streams conducted by the same user at different times.
INTERNET PROFILING SYSTEM AND METHOD

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority from United States Provisional Application Serial No. 60/150,217, filed August 23, 2000, by the inventor herein.

FIELD OF THE INVENTION

An Internet profiling method and system for client/user identification by statistical examination of streaming data.

BACKGROUND OF THE INVENTION

This invention relates, in general, to a method and system for identifying an Internet or network user who is engaged in repetitive Internet or network inter-sessions. More specifically it relates to an expandable method for a network device, which identifies and associates Internet or network inter-sessions conducted by a user using information sent/received to/from the user's computer to/from any remote host on the Internet or the network. The method is adaptable to many Internet and network users.

Identifying an Internet user has become a major issue since the Internet became a media for advertisement and targeted content. Remote content suppliers (e.g., web sites) who wish to identify users can use two common methods. First, the web sites may use resources on the client computer, e.g., cookies or client software, and second, the web sites may ask the client to identify himself/herself when the client asks for the content. These two methods are commonly used in client server interactions.

As is well known, clients connect to a public network, e.g., the Internet, through an ISP (Internet Service Provider). Currently, the ISP identifies its clients by their login information, e.g., username and password. Another method for the ISP to identify a client is to track their unique IP address, which is assigned to a client after login. The first method allows the ISP to associate a current user Internet session to a specific user and can...
create a user log file. The second method provides the ISP with a way to track the client Internet session in real time until the client disconnects. After disconnecting, the IP address is taken from the client and generally given to another client, so this method is effective only for the current Internet session.

The system and method presented herein allows for identifying Internet users by using information extracted from the users' network sessions and from the users' past activity pattern. The system and method allows associating various Internet inter-sessions that are conducted by the same user. The system and method is not dependent on login information or any other registry information.

SUMMARY OF THE INVENTION

A method and system for determining the identity of an unknown individual participating in a network session is provided. A database is maintained which includes records of each previously identified individual. The records include unique strings associated with each individual which were extracted from prior network sessions of that individual. When an individual to be identified participates in a network session, the data stream of information transmitted to and from that individual is read, and known data elements are identified in the information. A subset of information, which includes at least one unique string associated with a known data element, is extracted from the data stream. The subset of information is analyzed to determine if the individual participating in the current network session is a previously identified individual from the database. The analysis includes comparing the unique strings extracted for that individual with unique strings associated with previously identified individuals. If the individual is determined to be a previously identified individual, his or her identity is set as the identity of the previously identified individual, and the record in the database of previously identified individuals is updated with the new subset of information extracted from the current network session of the individual. Otherwise, if the individual is not identified, his or her identity is set as a new individual, and a new record is created in the database of the new individual with the subset of information extracted from the current network session.
One object of the present invention is to provide a generic, intelligent point which associates and identifies Internet sessions conducted by the same user using information and patterns, or a fingerprint, extracted from prior user sessions.

A further object of the present invention is to provide to an ISP a statistical method for monitoring and associating an anonymous user participating in a current Internet session using only the streaming data passing to/from the user to/from a remote server and a database of past Internet session information to compare to this streaming data.

Another object of the present invention is to provide a way for an ISP to share information with clients and remote content suppliers, e.g., web sites.

Still other objects and advantages of the invention will in part be obvious and will in part be apparent from the specification.

The invention accordingly comprises the several steps and relation of one or more of such steps with respect to each of the others, and the system embodying features of construction, combinations of elements and arrangements of parts which are adapted to effect such steps, all as exemplified in the following detailed disclosure, and the scope of the invention will be indicated in the claims.

**BRIEF DESCRIPTION OF THE DRAWINGS**

For a fuller understanding of the invention, reference is had to the following description taken in connection with the accompanying drawings, in which:

**FIG. 1** is a flowchart representation of a typical global computer network in accordance with the prior art;

**FIG. 2** is a flowchart representation of a global computer network in accordance with a preferred embodiment of the present invention;

**FIG. 3** is a detailed flowchart representation of the profiling system of **FIG. 2** constructed in accordance with the present invention; and
FIG. 4 is a flowchart representation depicting the steps performed during the profiling and identifying process according to a preferred embodiment of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Reference is first made to FIG. 1, which depicts a typical ISP junction in accordance with the prior art. In such a typical ISP junction, the main ISP site, generally indicated at 10 includes an ISP access device 18 which allows, for example, a dial-in access through a modem or the like as indicated at 14, direct access through a router or any other communication means, generally indicated at 16, thereby enabling a client 12, or a network 13 of clients 12a, 12b, 12c to connect to ISP junction 10. The site also includes a hub 22, a domain name server (DNS) 20, client access control such as a Radius 24, an e-mail server 25, hosted servers 26, and a router 30 which connects the ISP junction to a global computer network such as Internet 32. Generally, the identified named ISP devices are connected together via a network such as a local area network (LAN). It is noted that the particular configuration is shown as an example only and other ISP network configurations can be used with the present invention. The arrangement and set up of such configurations are well known to those skilled in the art. The present invention as described below in detail can be used in conjunction with any of these possible configurations.

Each client 12 is generally a computer such as a PC or laptop with video and audio capabilities, having a processor and programs or applications associated therewith. Internet 32 is a networked collection of clients and servers which are adapted through software and communication links to communicate with one another. The clients, typically through a browser program, can send a request message to a server and await a response. The response is displayed or presented by the browser. For a more detailed description of the Internet, browsers, Internet communication and protocols, reference is made to Ruvolo U.S. Patent No. 5,928,363, the description therein being incorporated by reference herein as though fully set forth.

FIG. 2 depicts the network configuration of FIG. 1 in which a profiling system, in the form of a session identifier, generally indicated at 40, and arranged and constructed in accordance with the present invention, has been installed. Like elements in
FIG. 2 as shown in FIG. 1 have the same reference numbers. It is noted that session identifier 40 is provided in ISP junction 10 in this example. However, it will be readily understood that the profiling system and method described herein may be used at other points on the network, such as at the hub of a web site.

FIG. 3 depicts a flowchart representation of the profiling system 40 of FIG. 2 arranged and constructed in accordance with the present invention. The profiling system preferably includes the following modules: system administrator 104, which enables a system operator to set the profiler policy, storage medium 100 including a database, which stores the profiling records, and sniffer 102, which collect the streaming data passing through hub 22 of the ISP.

The profiling process includes the following steps. First, the system extracts information from an Internet session and creates a session fingerprint pattern based on the user browsing activity and data stream. Second, the system tries to match previously created profile records in a database to the current browsing session. At this stage, the current Internet session may be identified by its unique IP address.

FIG. 4 depicts a flowchart of the profiling and identifying process. Task 110 reads the entire data stream passing through hub 22. The sniffer 102 distinguishes between different current user Internet sessions by extracting the unique IP address which is assigned to each client while connecting to the Internet. Task 112 extracts information from the Internet session data stream. The Internet session fingerprint pattern is a collection of unique strings which the user sends/receives to/from Internet servers during the Internet session. The unique strings are embedded in the application protocols stream, e.g., HTTP, SMTP, etc. As mentioned above, the profiling system uses a profiling policy. The profiling policy includes known data elements to look for and additionally determines the sources from which the unique string will be extracted. The source includes, but is not limited to: HTTP URL and domain, e.g., private home page, which includes special URL strings, user SMTP connection streams, cookies sent by an Internet host to a client, and information about the client computer and browser (e.g., OS type and version, browser type and version).

Task 114 analyzes and classifies the extracted unique strings. The task stores the unique strings in patterns according to the profiler policy. Task 118 creates and updates a
profile record in the database for an Internet session conducted by a user. The task assigns a unique serial number to the record in order to identify it later on. The record contains a group of unique strings which was extracted from the user Internet session and later will serve as a reference. Task 122 saves the profile record in the database of storage medium 100. Task 116 checks if the unique strings extracted from the Internet session match one of the profile records stored in the database contained in the storage medium 100. If a match is found, the Internet session is assigned, in task 120, to the matched profile. The match will be dropped when another better match is found and contradicts the first match which was found in the current Internet session.

The profiling system uses two tables for the profiling process, Current Session Table and Reference Session table. The Current Session Table stores information about an Internet session in real time. Each session is identified by its unique IP address. An example of the fields in this table is set forth below. The table can store any information passing to/from the Internet user during an Internet session. The table fields are determined by the profiling policy.

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Popular Web Sites</th>
<th>Unique Information Related to Popular Web Sites, e.g., Cookies, Customized Page, etc.</th>
<th>Browser Information, e.g., Type, Version, Screen Size, etc.</th>
<th>OS Information, e.g., Type Version, etc.</th>
<th>Fields of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1.1</td>
<td><a href="http://www.adwise.net">www.adwise.net</a></td>
<td>674cx99034</td>
<td>Ie5 v3</td>
<td>Win95</td>
<td>Sports</td>
</tr>
</tbody>
</table>

The Reference Session table stores information about previous Internet sessions. The information stored is used to associate current sessions with previous sessions. This is done by comparing unique keywords which are stored in the table fields. The table includes all the fields that are used in the Current Session Table except the IP address which is replaced by the unique serial number for the user. The table is updated every time a new unique subset of information is found and associates to a previous session.

The profiling process occurs at the ISP site where the profiling system can read the information passing from/to the Internet user.
An Example of the profiling process is set forth below:

1. Client connects to the ISP and gets a unique IP address.

2. Client starts an Internet browsing session.

3. The profiling system extracts information from the browsing stream and stores it in the Current Session Table using the IP address as an Internet session ID.

4. The Profiling system tries to find a match between a Current Session Table Entry and a Reference Session Table Entry.

5. If a match is found, the entry at the Reference Session Table is marked for future use and the user is associated with the serial number of this record.

6. If a match is not found, the profiling system continues to extract information until a match is found or creates a new record in the database.

Thus, a method and system for determining the identity of an unknown individual participating in a network session is provided. A database of records of each previously identified individual is maintained. The records include unique strings associated with individuals which were extracted from prior network sessions. When an individual to be identified participates in a network session, the data stream of information transmitted to and from that individual is read by the sniffer, and known data elements are identified in the information. A subset of information, which includes at least one unique string associated with a known data element, is extracted from the data stream. The subset of information is then analyzed by comparing the information extracted from the current network session to information stored in the database to determine if the individual participating in the network session is a previously identified individual. The analysis generally includes comparing the unique strings extracted for that individual with unique strings associated with previously identified individuals in the database. If the individual is determined to be a previously identified individual, his or her identity is matched with the identity of the previously identified individual, and the record in the database of the previously identified individuals is updated with the new subset of information extracted from the current network session. Otherwise, if the individual is not identified, his or her identity is set as a new individual, and
a new record is created in the database for the new individual which includes the subset of information extracted from the current network session.

The system and method according to the present invention allows identifying and associating a user’s previous Internet sessions with the user’s current one. The current session data stream is used to create the user’s Internet session fingerprint pattern, and includes unique strings extracted from the data stream. The fingerprint pattern for a current session is compared with fingerprint patterns from previous sessions maintained in a database to associate Internet inter-session data streams conducted by the same user at different times.

It will thus be seen that the objects set forth above, among those made apparent from the preceding description, are efficiently attained and, since certain changes may be made in carrying out the above method and in the construction set forth without departing from the spirit and scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings will be interpreted as illustrative and not in limiting sense.

It is also to be understood that the following claims are intended to cover all of the generic and specific features of invention herein described and all statements of the scope of the invention which, as a matter of language, might be set default therebetween.
CLAIMS

What is claimed is:

1. A method of determining the identity of an unknown individual participating in a current session on a computer network, comprising the steps of:

   providing a database of records of previously identified individuals including associated unique strings extracted from previous sessions on a computer network of said previously identified individuals;

   reading a data stream of information transmitted to and from an unknown individual participating in a current session on said computer network, and identifying known data elements in said information;

   extracting from said data stream a subset of information including at least one unique string associated with said known data elements;

   analyzing said subset of information, including comparing said at least one unique string with unique strings in said database, to determine if said unknown individual is a previously identified individual;

   if said unknown individual is determined to be a previously identified individual, assigning the identity of said unknown individual to the identity of said previously identified individual, and updating the record of said previously identified individual in said database with said subset of information; and

   if said unknown individual is determined to not be a previously identified individual, assigning the identity of said unknown individual to a new known individual, and creating a new record of said new known individual in said database with said subset of information.

2. The method of claim 1, wherein said computer network is a global computer network.

3. The method of claim 2, wherein said global computer network is the Internet.
4. The method of claim 3, wherein said unique strings include information associated with popular Internet web sites.

5. The method of claim 3, wherein said unique strings include information associated with Internet web browsers.

6. The method of claim 1, wherein said unique strings include information associated with a computer used by said unknown individual.

7. The method of claim 1, wherein said unique strings include information associated with fields of interest of said unknown individual.

8. A system for determining the identity of an unknown individual participating in a current session on a computer network, comprising:

   a database of records of previously identified individuals including associated unique strings extracted from previous sessions on a computer network of said previously identified individuals;

   means for reading a data stream of information transmitted to and from an unknown individual participating in a current session on said computer network, and identifying known data elements in said information;

   means for extracting from said data stream a subset of information including at least one unique string associated with said known data elements;

   means for analyzing said subset of information, including comparing said at least one unique string with unique strings in said database, to determine if said unknown individual is a previously identified individual;

   if said unknown individual is determined to be a previously identified individual, means for assigning the identity of said unknown individual to the identity of said previously identified individual, and updating the record of said previously identified individual in said database with said subset of information; and

   if said unknown individual is determined to not be a previously identified individual, means for assigning the identity of said unknown individual to a new known
individual, and creating a new record of said new known individual in said database with said subset of information.

9. The system of claim 8, wherein said computer network is a global computer network.

10. The system of claim 9, wherein said global computer network is the Internet.

11. The system of claim 10, wherein said unique strings include information associated with popular Internet web sites.

12. The system of claim 10, wherein said unique strings include information associated with Internet web browsers.

13. The system of claim 8, wherein said unique strings include information associated with a computer used by said unknown individual.

14. The system of claim 8, wherein said unique strings include information associated with fields of interest of said unknown individual.