Abstract:

Title: METHOD AND APPARATUS FOR PROVIDING AUTOMATIC GENERATION OF WEBPAGES

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 May 2008 (15.05.2008)

(10) International Publication Number
WO 2008/057755 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2007/082223

(22) International Filing Date: 23 October 2007 (23.10.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/555,137 31 October 2006 (3.1.2006) US


(72) Inventors; and


(74) Agent: TONG, Kin-Wah; Patterson & Sheridan, LLP, 595 Shrewsbury Avenue, Shrewsbury, New Jersey 07702 (US).


(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

(54) Title: METHOD AND APPARATUS FOR PROVIDING AUTOMATIC GENERATION OF WEBPAGES

(57) Abstract: A method and apparatus to quickly provide a plurality of customized deceptive web pages that is designed to feed an intruder with a number of intrusion opportunities are disclosed. For example, the present method receives a plurality of parameter of interest and "look and feel" information. The method then applies the plurality of parameter of interest and the look and feel information to automatically generate a plurality of web pages for use in a security system, e.g., a honeypot system.
METHOD AND APPARATUS FOR PROVIDING AUTOMATIC GENERATION OF WEB PAGES

[0001] The present invention relates generally to communication networks and, more particularly, to a method and apparatus for providing automatic generation of web pages for use in a network security system.

BACKGROUND OF THE INVENTION

[0002] In order to improve existing network security and to provide early detection of network security intrusions, companies often need to understand the behaviors of and the techniques used by hackers. Very often, companies use honeypot systems to meet this need. A honeypot is a closely monitored decoy system (e.g., one or more web pages) placed on a network or the Internet that is set up to attract and trap attackers such as hackers who attempt to penetrate systems or networks. However, the setup of a honeypot system and its associated decoy environments, such as web pages, files, and data, can be time consuming and manually intensive.


SUMMARY OF THE INVENTION

[0004] In one embodiment, the present invention enables a method to quickly provide a plurality of customized deceptive web pages that is designed to feed an intruder with a number of intrusion opportunities. For example, the present method receives a plurality of parameter of interest and "look and feel" information. The method then applies the plurality of parameter of interest and the look and feel information to automatically generate a plurality of web pages for use in a security system, e.g., a honeypot system.
The teaching of the present invention can be readily understood by considering the following detailed description in conjunction with the accompanying drawings, in which:

FIG. 1 illustrates an illustrative network of the present invention;
FIG. 2 illustrates a flowchart of a method for a honeypot system generation flow of the present invention;
FIG. 3 illustrates a flowchart of a method for handling intrusion flow by a honeypot system of the present invention; and
FIG. 4 illustrates a high level block diagram of a general purpose computer suitable for use in performing the functions described herein.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

Fig. 1 illustrates an illustrative network 100 of the present invention. For example, a plurality of users 112a-n may access a network 110, e.g., the Internet, to access an application server 120. The application server may be operated by an enterprise customer of a service provider for providing information its customers or employees. As such, the application server may have real databases that contain valuable information. For example, these real databases may contain product/service information that will support a website that will allow a customer to purchase products or to subscribe to services. Furthermore, these real databases may contain data that are intended to be used only by authorized users, such as employees of an enterprise customer. These private databases may contain corporate information, accounting information, financial information, billing information, transaction information, medical information, customer information, and the like.

Although most of the users 112a-n are legitimate users who will access application server 120 for legitimate reasons, there are those who are hackers who will attempt to access protected information located on the
application server without the proper authorization. In one embodiment, the application server may deploy one or more honey pots 124 for thwarting the hackers’ attempt to breach security measures and also to possibly gain insights into the intent of the hackers.

[0013] In order to improve existing network security and provide early detection of network security intrusions, companies often need to understand the behaviors of and the techniques used by hackers. Very often, companies use honeypot systems to meet this need. A honeypot is a closely monitored decoy system (e.g., one or more web pages) put on a network or the Internet that is set up to attract and trap attackers such as hackers who attempt to penetrate systems or networks. The key idea is to deceive the attacker by making the honeypot seems like a legitimate system. A honeypot system can also be a decoy-based Intrusion Detection System (IDS) in corporate network environments that provide effective early detection of intrusion threats. Honeypot is typically a virtual system residing in a network that emulates a real system by simulating running services and open ports; therefore, honeypot can help to gather information about new attacks, exploitation trends, and hacking techniques used by intruders. Generally, a honeypot consists of a computer, data or a network site that appears to be part of a network which is actually isolated and protected and contains information that would be of value to hackers. The primary value of a honeypot is to provide information which can be used for early intrusion detection, warning, prediction, and awareness.

Since a honeypot system is not a production system, it should not receive any traffic or activity at all. If a honeypot does capture any activity, the activity is most likely unauthorized and originated by a hacker with malicious intents. However, the setup of a honey pot system and its associated decoy environments, such as web pages, files, and data, can be time consuming and manually intensive.

[0014] To address this criticality, the present invention enables a method to quickly and accurately implement or provide specialized deceptive web pages needed by a honeypot associated website that is designed to feed an intruder with a number of intrusion opportunities. One aspect of the present invention is that it reduces the time needed to create usable deceptive web pages to a point
where the time needed to construct these deceptive web pages is negligible. For example, an owner of an application server or a website may provide a plurality of parameters that are likely to be of interest to unauthorized users. Examples of parameters may include, but are not limited to, customer information (e.g., name, address, social security number, phone number, birthday, and the like), security information (e.g., passwords, access codes, private keys), financial information (e.g., account number, billing information, transaction information, credit information and the like), product information (e.g., product designs, product launch date, product cost, product blueprints, product part lists, and the like), contact information (e.g., client listings, client phone numbers, client email addresses) and so on. The owner of an application server or a website may also provide look and feel information, e.g., company trademarks, company slogans, company logos, company colors, company images, and the like. Using the provided parameters and look and feel information, the present invention is able to quickly generate a plurality of specialized deceptive web pages.

[0015] In one embodiment, the present invention can identify the motives and intentions of an intruder. For instance, a honeypot service provider may provide the prospective intruder with a number of choices including, but not limited to, those that address the following topics:

- financial malfeasance;
- corporate espionage;
- national security espionage;
- unauthorized access to classified information;
- unbecoming conduct.

Note that the aforementioned items are only a very small subset of the possible types of potential information paths that can be presented to the un-authorized honeypot intruder who may have already ignored the warnings to stay away and has, therefore, identified oneself as a person of interest.

[0016] In one embodiment, the present invention enables a provider of honeypot services to change, in a close to real-time basis, the web pages that will be presented to the intruder. One feature of the present invention enables the ability to design a customer facing page for a honeypot website that is
capable of creating appropriately structured subsequent links. If the customer facing honeypot web page is properly constructed, a honeypot service provider can craft a set of predefined links on the web page to determine the intent of the intruder. If the intruder's intent is known, then the honeypot service provider can focus on the proper forms on subsequent web pages to respond to the intrusion.

[0017] Thus, the present invention enables a honeypot service provider to include and/or change honeypot web page contents at will. This gives the service provider the ability to change the apparent and the actual contents of the website with a minimal number of alterations. This implies that the visual contents of the website can be customized differently for each individual customer.

[0018] In one embodiment, the present invention enables a table driven script, or program. The program is constructed in such a way that it can generate a file that comprises a set of web pages that is immediately usable in a honeypot system. For example, the program receives two sets of inputs: 1) parameters of interest and 2) look and feel information for generating a honeypot website. These two sets of information are derived from inputs provided by owners of application servers or web sites. The speedy and nimble execution of the program will allow a honeypot service provider to provision and support a large number of potential honeypot customers in a relatively short time. More importantly, the program will greatly reduce the amount of time a honeypot service provider will need to spend on provisioning a new instance of a honeypot for an individual customer.


[0020] In step 210, the method receives a plurality of parameters of interest from an owner of an application server or a web site. For example, the plurality of parameters of interest may include customer information (e.g., name, address, social security number, phone number, birthday, and the like), security information (e.g., passwords, access codes, private keys), financial information (e.g., account number, billing information, transaction information, credit...
information and the like), product information (e.g., product designs, product launch date, product cost, product blueprints, product part lists, and the like), contact information (e.g., client listings, client phone numbers, client email addresses) and so on.

In one embodiment, the received parameters of interest can be organized hierarchically. For example, a parameter of interest can be "accounting information" which can be perceived as a parent parameter, whereas "tax information" can be a child parameter, whereas "tax return for 2005" can be a grandchild parameter, so on. This hierarchical organization of the parameters of interest will allow the program to not only automatically generates a set of links, but to generate a hierarchical set of links as further described below.

In step 220, the method receives look and feel information from an owner of an application server or a web site. For example, the owner of an application server or a web site may also provide look and feel information, e.g., company trademarks, company slogans, company logos, company colors, company images, and the like.

In step 330, the method uses the parameters of interest and the look and feel information as inputs to a program to generate a plurality of deceptive (e.g., honeypot) web pages. Namely, the program is designed to simply apply the look and feel information in one or more predefined ways to generate the plurality of deceptive web pages. For example, if the look and feel information contains a company color then the background of the web page can be set to that color. If a border pattern is provided, then the border pattern is deployed on the web page. If a company logo is provided, the company logo is deployed on one of the corners of the web page. If an image or an image sequence is provided, then the image or an image sequence is deployed on the web page.

Furthermore, the program will generate a plurality of links in accordance with the parameters of interest on the web page, e.g., a front end customer facing web page. For example, a link called "financial information" can be generated if a "financial information" parameter of interest is received. Furthermore, if the parameters of interest are hierarchically organized, then a plurality of hierarchical links can be generated. For example, if "tax information"
is a child parameter to the parent "financial information" parameter, then a link for tax information is also created, but it will only be displayed if the "financial information" link is accessed first, e.g., clicked. Thus, a customized web page can be easily and automatically generated by the program.

[0025] In step 240, the method deploys the generated honeypot web pages in a honeypot system. The method ends in step 250.

[0026] FIG. 3 illustrates a flowchart of a method 300 for handling intrusion flow by a honeypot system of the present invention. Method 300 starts in step 305 and proceeds to step 310.

[0027] In step 310, the method receives an access to a front end customer facing web page of a honeypot system. For example, an intruder has uncovered the front end customer facing web page which was intentionally deployed to be found by hackers.

[0028] In step 320, the method identifies the intent of an intruder using pre-designed links on the front end web page. As discussed above, the program in method 200 is able to automatically generate a plurality of links that are correlated to a plurality of parameters of interest. Depending on which link that the intruder is trying to access, method 300 is then able to determine the intent of the intruder. For example, if the intruder is presented with three (3) links on the front end web page, "financial information", "customer information" and "transaction information", and the intruder attempts to gain access to the "financial information" link, then method 300 is able to identify that the intruder's intent to gain access to the financial information that is believed by the intruder to be stored on the web site. Thus, method 300 may deduce the intent of the intruder by assessing a pattern of access of the plurality of web pages.

[0029] In step 330, the method provides subsequent web pages that can be traversed by the intruder to fit the intruder's intent. For example, when the intruder clicks on the "financial information" link, method 300 may present the intruder with a different webpage. However, in one embodiment, method 300 simply presents a response that will stop the intruder from progressing any further, but at the same time, allows the intruder an opportunity to return. For example, the response may be "The financial information database is currently undergoing an update process, but will be available tomorrow". This response
will stop the intruder, but it also presents an opportunity for the intruder to return, thereby allowing method 300 to assess the degree of interest of the intruder. However, if the intruder does in fact return the next day, then method 300 may present the next hierarchical link or links to the intruder, e.g., a "tax information" link and a "corporate profit report" link. If the intruder clicks on the "tax information" link, then method 300 may again present another response that will stop intruder but at the same time, allows the intruder an opportunity to return. For example, the response may be "Tax documents have not been validated, where validation will be completed in a week". As such, this process can be continued again and again by method 300 to gauge the degree of interest of the intruder. One novel aspect of the present invention is the automated nature of the present invention, where a realistic set of web pages can be quickly generated for a client by using a program that simply uses two simple sets of inputs.

In step 340, the method records all activities performed by the intruder. The method ends in step 350.

FIG. 4 depicts a high level block diagram of a general purpose computer suitable for use in performing the functions described herein. As depicted in FIG. 4, the system 400 comprises a processor element 402 (e.g., a CPU), a memory 404, e.g., random access memory (RAM) and/or read only memory (ROM), an automatic generation of web pages module 405, and various input/output devices 406 (e.g., storage devices, including but not limited to, a tape drive, a floppy drive, a hard disk drive or a compact disk drive, a receiver, a transmitter, a speaker, a display, a speech synthesizer, an output port, and a user input device (such as a keyboard, a keypad, a mouse, and the like)).

It should be noted that the present invention can be implemented in software and/or in a combination of software and hardware, e.g., using application specific integrated circuits (ASIC), a general purpose computer or any other hardware equivalents. In one embodiment, the present automatic generation of web pages module or process 405 can be loaded into memory 404 and executed by processor 402 to implement the functions as discussed above. As such, the present automatic generation of web pages process 405
(including associated data structures) of the present invention can be stored on a computer readable medium or carrier, e.g., RAM memory, magnetic or optical drive or diskette and the like.

[0033] While various embodiments have been described above, it should be understood that they have been presented by way of example only, and not limitation. Thus, the breadth and scope of a preferred embodiment should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.
What is claimed is:

1. A method for generating a plurality of web pages, comprising:
   receiving a plurality of parameter of interest;
   receiving look and feel information; and
   applying said plurality of parameter of interest and said look and feel
   information to automatically generate a plurality of web pages for use in a
   security system.

2. The method of claim 1, wherein said security system is a honeypot
   system.

3. The method of claim 1, wherein said plurality of parameter of interest is
   organized into a hierarchical order.

4. The method of claim 1, wherein at least one of said plurality of web
   pages contains a plurality of links that are correlated to said plurality of
   parameter of interest.

5. The method of claim 4, wherein said plurality of links is organized in a
   hierarchical order.

6. The method of claim 5, further comprising:
   deploying said plurality of web pages in said security system.

7. The method of claim 6, further comprising:
   receiving an access request to one of said plurality of web pages in said
   security system.

8. The method of claim 7, further comprising:
   presenting said plurality of links in a sequence of web pages in
   accordance with said hierarchical order.
9. The method of claim 8, further comprising:
determining an intent of an intruder based upon a pattern of access of said plurality of web pages.

10. A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of a method for generating a plurality of web pages, comprising:
    receiving a plurality of parameter of interest;
    receiving look and feel information; and
    applying said plurality of parameter of interest and said look and feel information to automatically generate a plurality of web pages for use in a security system.

11. The computer-readable medium of claim 10, wherein said security system is a honeypot system.

12. The computer-readable medium of claim 10, wherein said plurality of parameter of interest is organized into a hierarchical order.

13. The computer-readable medium of claim 10, wherein at least one of said plurality of web pages contains a plurality of links that are correlated to said plurality of parameter of interest.

14. The computer-readable medium of claim 13, wherein said plurality of links is organized in a hierarchical order.

15. The computer-readable medium of claim 14, further comprising:
deploying said plurality of web pages in said security system.

16. The computer-readable medium of claim 15, further comprising:
receiving an access request to one of said plurality of web pages in said security system.
17. The computer-readable medium of claim 16, further comprising:
   presenting said plurality of links in a sequence of web pages in
   accordance with said hierarchical order.

18. The computer-readable medium of claim 17, further comprising:
   determining an intent of an intruder based upon a pattern of access of
   said plurality of web pages.

19. An apparatus for generating a plurality of web pages, comprising:
   means for receiving a plurality of parameter of interest;
   means for receiving look and feel information; and
   means for applying said plurality of parameter of interest and said look
   and feel information to automatically generate a plurality of web pages for use
   in a security system.

20. The apparatus of claim 19, wherein said security system is a honeypot
    system.
FIG. 2

START 205

RECEIVE PARAMETERS OF INTEREST

210

RECEIVE LOOK AND FEEL INFORMATION

220

PROVIDE SAID PARAMETERS OF INTEREST AND SAID LOOK AND FEEL INFORMATION TO A PROGRAM FOR AUTOMATED GENERATION OF A PLURALITY OF DECEPTIVE WEB PAGES

230

DEPLOY SAID DECEPTIVE WEB PAGES IN A HONEYPOT SYSTEM

240

END 250