

SYSTEM AND METHOD FOR DETECTING FALSE CODE

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of Korean Patent Application No. 2008-0047551
5 which was filed on May 22, 2008, which are hereby incorporated by reference as if fully set
forth herein.

BACKGROUND

1. FIELD

10 The present invention relates to a system and method for detecting a false code, and
more particularly, to a system and method for detecting a false code in a web site linked to a
plurality of web pages.

2. DESCRIPTION OF THE RELATED ART

15 With development of the Internet, the number of web sites is sharply increasing. In
such a web site environment, contents provided by Web 2.0 and user created content (UCC)
service users are also increasing. Since a general user can organize contents of web sites,
actions for malicious purposes against the management policies of web sites are also
continuously increasing.

20 For example, methods of embedding malicious codes in vulnerable web sites and
spreading the malicious codes to a large number of general users visiting the web sites are
becoming widespread. In particular, attacks on national public institutions are rapidly
increasing and causing serious problems. Here, a malicious code is a program developed to
disturb a user in smooth use of a computer, including computer viruses, computer worms



(self-replicable malicious codes), Trojan horses (malicious codes not infecting other targets but causing a system to malfunction), and so on.

The administrator of a web site must allow use of hypertext markup language (HTML) tags, such as IFRAME and SCRIPT, in order to use content provided by users and a
5 malicious code may be inserted into such HTML tags. In other words, the problem is in that the administrator of a web site must allow use of HTML tags because the web site cannot perform its proper functions unless use of HTML tags is allowed.

In particular, a normal HTML tag and malicious HTML tag have basically the same format, and thus it is difficult to distinguish them from each other. Also, tags having
10 different purposes are input according to non-formal characteristics of a web site, for example, portals, moving pictures, price comparisons, and shopping sites, and thus it is impossible to limit a user input to a predetermined pattern. As such, an administrator must examine all web pages and determine whether or not there is a problem. However, in a huge web site in which tens of thousands of postings are registered, it is substantially impossible to monitor all
15 the web pages.

Consequently, a system that can rapidly check an entire web site in a short time period is necessary.

SUMMARY

20 The present invention is directed to a system and method for detecting a false code in a web site, involving collecting a part or whole of web pages linked from a web site, determining whether tags included in the collected web pages are false codes or normal codes, providing the determination result to an administrator terminal, and thereby enabling an administrator to efficiently detect a false code.

The present invention is also directed to a system and method for detecting a false code in a web site, involving collecting a generated or updated web page among web pages linked from a web site, determining whether tags included in the collected web pages are false codes or normal codes, and thereby reducing check time.

5 The present invention is also directed to a system and method for detecting a false code in a web site, involving collecting a part or whole of web pages linked from a web site, informing an administrator terminal of false codes or normal codes included in the collected web pages to enable rapid countermeasure against false codes, and thus allowing a system to be stably managed.

10 The present invention is also directed to a system and method for detecting a false code in a web site, involving rapidly detecting and coping with a false code included in a part or whole of web pages linked from a web site, and thus enabling a user to safely use the web site.

 According to an aspect of the present invention, there is provided a system for
15 detecting a false code, including: a web server for administering a web site linked to a plurality of web pages; and a surveillance server for collecting a generated or updated web page among the linked web pages, determining whether tags included in the generated or updated web page are false codes or normal codes, and providing the determination result to an administrator terminal such that an administrator can check the determination result.

20 The surveillance server may collect location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server, and may determine whether a previously set tag among the tags included in the generated or updated web page is a false code or normal code.

 The tag may be a hypertext markup language (HTML) tag for inserting information
25 into the web page, and may be one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and

SPAN. The false codes may violate a policy of the web site and may affect an administrator or user of the web site.

The system for detecting a false code may further include a database for separately storing the false code and the normal code according to the determination result.

5 According to another aspect of the present invention, there is provided a system for detecting a false code, including: a collector for collecting location information of generated or updated web pages among web pages linked from a web site administered by a web server; a loader for loading the web pages according to the collected location information of the web pages; and a determiner for determining whether tags included in the loaded web pages are
10 false codes or normal codes.

The collector may collect location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server, and the loader may load content of a previously set tag included in the web pages or a web page including the previously set tag among the web pages according to the
15 collected location information of the web pages.

Here, the tag may be an HTML tag for inserting information into the web pages, and may be one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN. The false codes may violate a policy of the web site and may affect an administrator or user of the web site.

According to still another aspect of the present invention, there is provided a method
20 of detecting a false code, including: administering, at a web server, a web site linked to a plurality of web pages; and collecting, at a surveillance server, a generated or updated web page among the linked web pages, determining whether tags included in the generated or updated web page are false codes or normal codes, and providing the determination result to an administrator terminal such that an administrator can check the determination result.

The providing of the determination result to an administrator terminal may include: collecting a web page generated or updated within a predetermined time among the linked web pages; searching tags included in the collected web page for a tag previously set to be checked; determining whether the tags included in the web page are false codes or normal
5 codes according to the search result; and providing the determination result to the administrator terminal such that the administrator administering the web server can check the determination result.

The tag may be an HTML tag for inserting information into the web page, and may be one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN. The false codes may
10 violate a policy of the web site and may affect the administrator or user of the web site.

The method may further include separately storing the false codes and the normal codes according to the determination result.

According to yet another aspect of the present invention, there is provided a method of detecting a false code, including: collecting, at a surveillance server, location information
15 of generated or updated web pages among web pages linked from a web site administered by a web server; loading, at the surveillance server, the web pages according to the collected location information of the web pages; and determining, at the surveillance server, whether tags included in the loaded web pages are false codes or normal codes.

The collecting of the location information of generated or updated web pages may
20 include collecting location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server. The loading of the web pages may include loading content of a previously set tag included in the web pages or loading a web page including the previously set tag among the web pages according to the collected location information of the web pages.

The tag may be an HTML tag for inserting information into the web pages, and may be one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN. The false code may violate a policy of the web site and may affect the administrator or user of the web site.

5

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other features and advantages of the present invention will become more apparent to those of ordinary skill in the art by describing in detail preferred exemplary embodiments thereof with reference to the attached drawings in which:

FIG. 1 schematically illustrates a constitution of a system according to an exemplary
10 embodiment of the present invention;

FIG. 2 is a block diagram of a surveillance server shown in FIG. 1;

FIG. 3 is a flowchart illustrating a method of detecting a false code according to an exemplary embodiment of the present invention;

FIG. 4 illustrates an example of details for detecting a false code according to an
15 exemplary embodiment of the present invention;

FIG. 5 illustrates an example of a screen in which a false code is detected according to an exemplary embodiment of the present invention; and

FIG. 6 illustrates examples of a screen for managing a false code according to an exemplary embodiment of the present invention.

20

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

The invention is described more fully hereinafter with reference to the accompanying drawings, in which embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the

embodiments set forth herein. Rather, these embodiments are provided to fully enable those of ordinary skill in the art to embody and practice the invention.

Hereinafter, a system and method for detecting a false code in a web site according to exemplary embodiments of the present invention will be described in detail with reference to
5 FIGS. 1 to 6.

The present invention provides a technique for efficiently detecting a false code in a web site. More specifically, the exemplary embodiments of the present invention involve collecting a part or whole of web pages, for example, generated or updated web pages linked from a web site, determining whether codes or tags included in the web pages are false codes
10 or false tags, and providing the determination result to an administrator terminal such that an administrator can efficiently detect a false code.

FIG. 1 schematically illustrates a constitution of a system according to an exemplary embodiment of the present invention.

As illustrated in FIG. 1, the system for detecting a false code in a web site according
15 to an exemplary embodiment of the present invention may include a user terminal 110, a web server 120, a surveillance server 130, a database (DB) 131, and an administrator terminal 140.

The web server 120 may administer a web site linked to a plurality of web pages and provide a variety of information. For example, web pages administered by vendors are linked to a web site corresponding to an online open market, and the vendors can upload
20 information relating to products onto the web pages. Then, a user can access a web page in the web site using a user terminal, check the information relating to a desired product, and purchase the product.

Here, the vendors must use hypertext markup language (HTML) tags, such as IFRAME and SCRIPT, in order to display information relating to products, and a false code
25 or false tag may be inserted into the HTML tags. There are various types of false codes

including 1) a type that hacks a personal computer (PC) by inserting a malicious code distributing a computer virus, or uses a PC as an agent for a distribute denial of service (DDoS) attack, 2) another type that inserts a code installing an anti-virus vaccine and tricks a user into settling accounts using false diagnosis, and 3) still another type that unfairly
5 manipulates the benefit of accumulating a part of a settlement fund, that is, a reward process in a shopping mall, or leads to price comparison or direct transaction.

Thus, the surveillance server 130 may check HTML tags at a predetermined interval. More specifically, it is possible to determine whether HTML tags included in a web page in which information relating to a product is uploaded are false codes or normal codes. Then,
10 the surveillance server 130 may provide the determination result to the administrator terminal 140.

Also, the surveillance server 130 may inform the administrator that a false code is detected using an alarm, message, and so on.

Here, an HTML tag is used to insert information into the web page, and may be one
15 of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN.

The surveillance server 130 may store location information of web pages including detected false codes, that is, a false code list, in the DB 131 and manage it. Also, the surveillance server 130 may store an exception list including reliable web sites, e.g., Gmarket, Samsung Electronics, and Microsoft, not to be searched for a false code in the DB 131 and
20 manage it. In other words, the DB 131 may be classified into a first DB storing a false code list and a second DB storing a normal code list or an exception list.

In addition, an HTML tag to be checked may be set by the administrator in the DB 131.

In this way, an exemplary embodiment of the present invention determines whether
25 or not a false code exists only in generated or updated web pages among web pages linked

from a web site, and provides the determination result to an administrator terminal, thereby narrowing a detection range. Consequently, the time taken to detect a false code in a web site can be reduced.

FIG. 2 is a block diagram of the surveillance server 130 shown in FIG. 1.

5 As illustrated in FIG. 2, the surveillance server 130 may include a first interface 210, a collector 220, a loader 230, a notifier 240, a determiner 250, a second interface 260, and a third interface 270.

The surveillance server 130 may interoperate with the web server through the first interface 210, with the DB through the second interface 260, and with the administrator
10 terminal through the third interface 270.

Through the first interface 210, the collector 220 may collect location information, e.g., uniform resource locators (URLs), of generated or updated web pages, i.e., web pages to be collected, among a plurality of web pages linked from a web site administered by the web server.

15 The loader 230 may be provided with the collected location information of the web pages, load the web pages based on the location information, and provide the loaded web pages to the determiner 250. Upon receipt of the loaded web pages, the determiner 250 may check HTML tags set to be checked, e.g., some or all of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN, among all the HTML tags in the web pages from the DB through
20 the second interface 260.

The determiner 250 may determine whether HTML tags in the loaded web pages are false codes or normal codes based on the HTML tags to be checked, and provide the determination result to the administrator terminal through the third interface 270.

According to the determination result, the notifier 240 may output an alarm to inform
25 an administrator that there is a web page including a false code, and also generate and send a

message, e.g., a short message service (SMS) message, a long message service (LMS) message, a multimedia message service (MMS) message, e-mail, or a voice message, including text or voice, e.g., "One hundred false codes have been detected," input in advance to the administrator.

5 When no false code is detected, needless to say, it is unnecessary to send a message.

 In this way, an exemplary embodiment of the present invention enables a part or whole of a web page linked from a web site to be collected and an administrator terminal to be informed of a false code or normal code included in the web page. Thus, an administrator administering/managing the web site can rapidly cope with a false code, for
10 example, treat or block the corresponding web page.

 FIG. 3 is a flowchart illustrating a method of detecting a false code according to an exemplary embodiment of the present invention.

 As illustrated in FIG. 3, a surveillance server may periodically monitor whether or not all web pages linked from a specific web site are updated or generated, and when there
15 are generated or updated web pages, may collect location information of the generated, modified or updated web pages, that is, the web pages to be collected (S310).

 This is because a huge web site, such as a portal site or a shopping mall site, is linked to tens of thousands or hundreds of thousands of web pages and it is substantially impossible to check all the web pages. Thus, the surveillance server may pick only web pages
20 generated or updated by users within a predetermined time period and collect location information of the web pages.

 The surveillance server may load the web pages according to the collected location information of web pages (S320). When the surveillance server loads the web pages, it may search for HTML tags set to be checked and check HTML tags in the web pages (S330).

At this time, the surveillance server may load the entire contents of the web pages or may load contents of a specific HTML tag in the web pages or a web page including the specific HTML tag. This is because not all the HTML tags, but only some of them are typically used for a false code.

5 The surveillance server may determine whether the HTML tags to be checked are false codes or not (S340). When a false code exists in the web pages, the surveillance server may make a false code list including location information of the corresponding web page or the corresponding HTML tag (S350).

 When no false code exists in the web page, the surveillance server may determine
10 whether the HTML tags to be checked are normal codes or not (S360). When the HTML tags to be checked are normal codes, the surveillance server may make an exception list (S370). On the other hand, when the HTML tags to be checked are not normal codes, the surveillance server may make a pattern application list including the HTML tags (S380). This is because the HTML tags included in the web pages are not included in either a false
15 code DB or normal code DB, and an administrator determines whether the HTML tags are to be stored in the false code DB or normal code DB.

 At this time, location information of HTML tags included in the pattern application list may be newly applied to the false code list or normal code list according to the administrator's determination. Then, when the same pattern is detected later, it can be
20 processed as a false code or a normal code.

 When the surveillance server finishes checking HTML tags of all the generated or updated web pages, it may provide the checked details to an administrator terminal such that the administrator can check the checked details (S390). The checked details displayed on the administrator terminal will be described with reference to FIG. 4.

FIG. 4 illustrates an example of details for detecting a false code according to an exemplary embodiment of the present invention.

As illustrated in FIG. 4, a false code detection window according to an exemplary embodiment of the present invention may display 1) HTML tags set to be checked, 2) a list of web pages or HTML tags set to be collected, 3) a pattern application list, 4) a false code list, and so on.

These will be described in detail now. First, 1) HTML tags to be checked, for example, IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN may be set by an administrator.

2) A part or whole of web pages set to be collected, that is, web pages generated or updated within a predetermined time period, may be displayed using, for example, types of HTML tags and content of the HTML tags.

It is determined whether 1) the set HTML tags among HTML tags in the collected web pages are false codes or normal codes, and 3) a pattern application list may be displayed using, for example, types of HTML tags and contents of the HTML tags. This is because a false code has a uniform pattern and it is necessary to detect the pattern. Here, 3) the pattern application list includes HTML tags that cannot be automatically determined as false codes or normal codes and are determined by an administrator later, and patterns of the HTML tags will be applied to a false code or normal code.

According to the determination result, 4) a false code list may be displayed using, for example, types of HTML tags, URLs of web pages including the HTML tags, and content of the HTML tags.

In this way, an exemplary embodiment of the present invention collects only generated or updated web pages among web pages linked from a web site, and informs an

administrator terminal of a false code or false tag included in the web pages. Thus, it is possible to rapidly cope with the false code or false tag, and a system can be stably managed.

FIG. 5 illustrates an example of a screen in which a false code is detected according to an exemplary embodiment of the present invention.

5 As illustrated in FIG. 5, content of a web page including a false code (a rectangular section) is shown. Content of a false code may be examined by clicking any one of false codes in 4) the false code list displayed in the detection window of FIG. 4.

FIG. 6 illustrates examples of a screen for managing a false code according to an exemplary embodiment of the present invention.

10 As illustrated in FIG. 6, a surveillance server according to an exemplary embodiment of the present invention can separately store and manage an upper false code list and a lower normal code list. Based on such lists, it is possible to determine whether a collected web page includes a false code or not. Thus, 3) the pattern application list of FIG. 4 may be determined later by an administrator as a false code or normal code, and the pattern may be
15 applied to the false code list or normal code list.

The above-described method can be implemented as computer-readable code in a computer-readable recording medium. The computer-readable recording medium is any recording medium for storing data that can be read by a computer system. Examples of the computer-readable recording medium include a read-only memory (ROM), a random access
20 memory (RAM), a compact disk-read only memory (CD-ROM), a magnetic tape, a floppy disk, and optical data storage. Alternatively, the medium may be implemented in the form of carrier waves (e.g., Internet transmission).

The system and method for detecting a false code in a web site are not limited to the configurations and methods of the exemplary embodiments described above, and all or some
25 of the exemplary embodiments may be selectively combined to yield variants.

While the invention has been shown and described with reference to certain exemplary embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A system for detecting a false code, comprising:
 - a web server for administering a web site linked to a plurality of web pages; and
 - 5 a surveillance server for collecting a generated or updated web page among the linked web pages, determining whether tags included in the generated or updated web page are false codes or normal codes, and providing the determination result to an administrator terminal such that an administrator can check the determination result.
- 10 2. The system of claim 1, wherein the surveillance server collects location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server, and determines whether a previously set tag among the tags included in the generated or updated web page is a false code or normal code.
- 15 3. The system of claim 2, wherein the tag is a hypertext markup language (HTML) tag for inserting information into the web page, and is one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN.
- 20 4. The system of claim 1, wherein the false codes violate a policy of the web site and may affect an administrator or user of the web site.
- 25 5. The system of claim 1, further comprising a database for separately storing the false codes and the normal codes according to the determination result.

6. A system for detecting a false code, comprising:
- a collector for collecting location information of generated or updated web pages among web pages linked from a web site administered by a web server;
- a loader for loading the web pages according to the collected location information of the web pages; and
- a determiner for determining whether tags included in the loaded web pages are false codes or normal codes.
7. The system of claim 6, wherein the collector collects location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server.
8. The system of claim 6, wherein the loader loads content of a previously set tag included in the web pages or a web page including the previously set tag among the web pages according to the collected location information of the web pages.
9. The system of claim 8, wherein the tag is a hypertext markup language (HTML) tag for inserting information into the web pages, and is one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN.
10. The system of claim 6, wherein the false codes violate a policy of the web site and may affect an administrator or user of the web site.
11. A method of detecting a false code, comprising:
- administering, at a web server, a web site linked to a plurality of web pages; and

collecting, at a surveillance server, a generated or updated web page among the linked web pages, determining whether tags included in the generated or updated web page are false codes or normal codes, and providing the determination result to an administrator terminal such that an administrator can check the determination result.

5

12. The method of claim 11, wherein the providing of the determination result to an administrator terminal includes:

collecting a web page generated or updated within a predetermined time among the linked web pages;

10 searching tags included in the collected web page for a tag previously set to be checked;

determining whether the tags included in the web page are false codes or normal codes according to the search result; and

15 providing the determination result to the administrator terminal such that the administrator administering the web server can check the determination result.

13. The method of claim 12, wherein the tag is a hypertext markup language (HTML) tag for inserting information into the web page, and is one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN.

20

14. The method of claim 11, wherein the false codes violate a policy of the web site and may affect an administrator or user of the web site.

15. The method of claim 11, further comprising separately storing the false codes
25 and the normal codes according to the determination result.

16. A method of detecting a false code, comprising:
- collecting, at a surveillance server, location information of generated or updated web pages among web pages linked from a web site administered by a web server;
- 5 loading, at the surveillance server, the web pages according to the collected location information of the web pages; and
- determining, at the surveillance server, whether tags included in the loaded web pages are false codes or normal codes.
- 10 17. The method of claim 16, wherein the collecting of the location information of generated or updated web pages includes collecting location information of a web page generated or updated within a predetermined time among the web pages linked from the web site administered by the web server.
- 15 18. The method of claim 16, wherein the loading of the web pages includes loading content of a previously set tag included in the web pages or loading a web page including the previously set tag among the web pages according to the collected location information of the web pages.
- 20 19. The method of claim 18, wherein the tag is a hypertext markup language (HTML) tag for inserting information into the web pages, and is one of IFRAME, SCRIPT, OBJECT, EMBED, DIV, and SPAN.
- 25 20. The method of claim 16, wherein the false codes violate a policy of the web site and may affect an administrator or user of the web site.

21. A computer-readable recording medium storing a program for executing the method of any one of claims 11 to 20.