# 网页防篡改和自动恢复系统①

冶忠林，王相龙

(四川大学 软件学院，成都 610225)

**摘 要**：目前存在的网页防篡改和自动恢复技术主要有三种：时间轮询技术，事件触发技术+核心内嵌技术和文件过滤技术+事件触发技术。这三种方式都是对目标文件进行监控，当目标文件被篡改时就从备份文件中还原出原文件，但是却没有对备份文件做保护，如果备份文件被破坏，则无法正常的恢复原网页文件，所以在文件过滤+事件触发技术的基础上，研究了如何去使用 MD5 校验，DES 加密,文件重命名等三种方式去保护备份文件的安全性。

**关键词**：网页安全；网页防篡改；文件过滤；事件触发；网页自动恢复；DES 加密；MD5 校验

## Web Page Tampering Prevention and Automatic Recovery System

YE Zhong-Lin, WANG Xiang-Long

(Software Institute, Sichuan University, Chengdu 610225, China)

**Abstract**: The existing web prevent tampering and automatic recovery technology mainly have three at present: time polling technology, event trigger technology + core embedded technology, file filtering technology + event trigger technology. All the three ways monitor the target file when the target file is tampered, it will return to the original condition from backup files in the file, but the backup files weren't protected, if the backup files were destroyed, the web pages won't be returned to its original condition. So in file filtering + event trigger technology foundation, this paper studies how to use MD5 calibration, DES encryption, file renaming, three ways to protect the safety of backup files.

**Key words**: web security; Web prevent tampering; file filtering; event trigger; Web page automatic recovery; DES encryption; MD5 calibration

## 1 引言

网络数据传输安全已经变成计算机网络研究的一个重要方向，众所周知，网络数据一般借助网页在传输，所以如何保护网络数据安全已经成为亟待解决的问题。根据 2010 年 7 月底中国互联网络信息中心(CNNIC)发布的统计数据，截至 2010 年 6 月，中国网民人数已超过 4.2 亿，突破了 4 亿大关，网站数量超过 70 万个，互联网已深入到社会生活的方方面面,如此大规模的网民数量和网站分布给我国互联网产业发展带来了前所未有的安全挑战。又据国家计算机网络应急技术处理协调中心（CNCERT/CC）统计，2006 年 3 月，全球被篡改网站数量超过 3 万个，平均每 1.5 分钟，就有一个网站被篡改，在这样一个背景下，网页防篡改技术就应运而生。

## 2 国内外研究状况

目前国内外主要的网页防篡改和自动恢复技术主要有三种：

(1) 时间轮询技术是利用一个网页的轮询检测程序，每过一个时间段，就开始以轮询的方式对监控文件内的每一个网页文件与备份的网页文件进行对比，如果发现网页文件被篡改，则从备份文件中还原出原

① 收稿时间:2011-06-10;收到修改稿时间:2011-09-05

文件。

时间轮询技术是每隔一段时间检验一次网页文件是否被篡改，所以如果网页文件被篡改则不能在检查间隙时间段内被及时的发现和恢复。

(2) 事件触发技术+核心内嵌技术是采用了非对称的加密算法，把所有的网页使用非对称加密算法加密，当客户机发送一个访问请求时，服务器首先要进行密码验证，如果密码验证不通过，则调用备份文件中进行解密，然后再对外发布页面。这种方式的事件触发机制只要是依靠检验文件的一些属性，例如文件大小，创建时间等。但是这种技术也有一个弊端就是要对访问的每一个页面发布之前都要进行完整性检查，所以需要占用巨大的系统资源，给服务器带来很大的负载。

(3) 文件过滤技术+事件触发技术使用了文件底层驱动技术，通过事件触发的方式对目标文件进行实时的监控，文件过滤技术只要是实时的检测文件的属性，如果文件属性发生了变化，则从备份文件拷贝原网页到目标文件的相应位置，通过文件驱动技术，可以很容易的获取一个文件属性的变化，而且整个文件的复制过程是非常快捷的，通常用毫秒来计，所以在这么段得事件内，用户根本无法看到已经被窜改的页面，所以这种网页防篡改技术在实时性上达到了最高水准。

## 3 改进方案

通过以上三种主要防篡改和自动恢复系统的分析可以发现，不论是时间轮询技术，事件触发技术+核心内嵌技术还是文件过滤技术+事件触发技术都没有对备份文件保护，如果备份文件遭到破坏或者备份文件中的网页文件被替换，那么以上三种网页防篡改方式就无法从备份文件中还原出原网页。

所以本文研究如何在文件过滤技术+事件触发技术基础上使用 DES 加密，MD5 校验，文件重命名等三种技术对备份文件保护。

在本文中，要对备份文件中的每个网页文件进行 DES 加密，然后计算 MD5,最后重命名每个网页文件，所以在本文中需要建立一个表，该表中要保存每个网页文件的重命名后的网页文件名称，网页文件原名称，MD5 码。

该表维护了重命名后的网页文件名称，网页文件原名称和 MD5 码值之间的映射关系，当文件被篡改后，首先使用从表中读出 MD5 码，并计算出当前网页的 MD5 码，校验备份网页文件是否被篡改，如果没有被修改则使用 DES 解密，然后读出网页文件的原名称，并用原名称重命名该文件，最后复制该网页文件到目标文件。

表 1　系统维护重命名后的文件名，原名称，MD5 码的映射表（示例）

| 重命名后的文件名 | 文件原名称 | MD5 码 |
|---|---|---|
| Test.php | Index.php | 68f52509917615e8 06c530e1ffe48d0c |

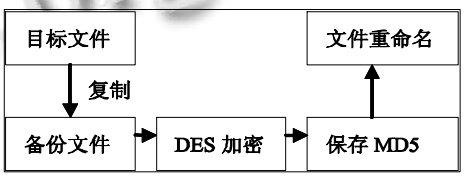


图 1　文件备份过程

从图 1 可以看出，当开始备份时，首先把所有的网页文件复制到备份文件，然后对备份文件中的每个网页文件进行 DES 加密和计算 MD5，并随机产生一个文件名，并用随机产生的文件名重命名该文件，然后把该随机产生的文件名，文件原来的名称，MD5 码都存入到数据表中。

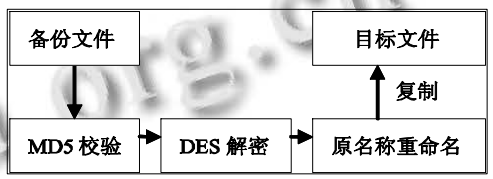


图 2　文件被篡改后的操作过程

从图 2 可以看出，当网页文件被篡改后，首先从读数据库中读出该网页文件的 MD5 码，并计算出当前网页文件的 MD5 码，校验该网页文件是否被篡改，如果没有被篡改，则使用 DES 解密出原文件，最后从数据表中读出该网页文件的原名称，并用该原名称重命名该文件，然后把该文件复制到目标监控文件。这样网页的恢复过程就完成了。但是备份文件中此时的文件是解密后的网页文件，所以还需要使用 DES 加密一次，并重新获取一个随机文件名，最后更新数据表，在这个过程中，不需要再重新获取 MD5 码，因为文件

没有被修改。

## 4 系统设计

系统软件总共有五个模块：开始监控模块，停止监控模块，查看日志模块，导出文件模块和告警模块。系统运行后维护着一个线程，该线程负责对目标文件进行监控。当目标文件中有创建文件，删除文件，修改文件和重命名文件时，系统中会产生相应的消息 FILE_ACTION_ADDE，FILE_ACTION_REMOVED，FILE_ACTION_MODIFIE 和 FILE_ACTION_RENAMED_OLD_NAME，该线程负责获取这些系统的消息，然后判别消息的类型，做相应的处理。这次处理如图 2 所示，有 MD5 校验，DES 解密，获取网页文件的原真实名称，然后用该真实名称重命名该网页文件，然后把该网页文件复制到目标监控目录下。此时网页文件是已被解密的文件，所以还需要 DES 加密一次，然后再重新获取一个随机的名称并存入数据表，最后删除已被解密的原网页文件。

此时备份文件中，所有的网页文件都是经过 DES 加密的，同时每个网页名称都是经过了重命名的，有一张数据表，专门维护网页文件的原真实名称，临时名称和 MD5 码值的映射关系。所以在解密的时候需要通过这张数据表找回原网页文件的真实名称和 MD5 码值。

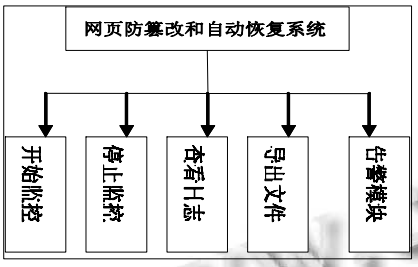


图 3 系统模块图

### 4.1 开始监控模块

首先把目标文件中的所有网页保存到备份文件中，然后对备份文件中的每一个网页文件获取其名称，把该名称保存在一个数据表中，之后随即产生一个文件名，并用该文件名重命名该网页文件，同时，把该随即产生的临时文件名称也存入到数据表中。数据表中的原真实名称和临时名称是一个一一对应的关系。然后计算每个文件的 MD5 码，然后把 MD5 码保存在数据表中，最后使用 DES 加密算法对每个文件进行加密，经过重命名，获取 MD5，加密等三种操作，完成了开始监控模块的具体操作。

在开始监控模块中，要开启一个线程，该线程就是用来对目标文件进行监控和从备份文件中进行文件恢复用的。

监控中的监控线程要实时的获取系统的消息，然后对获取的消息做相应的判断，然后进行相应的操作，如果文件被修改和删除等非法的操作，那么就开始进行文件的恢复，其恢复的具体流程是：首先从数据表中获取文件的真实名称，然后计算其 MD5 码，再从文件中读取 MD5 码，对两个 MD5 码进行相应的比较，如果 MD5 校验后是一致的，则文件没有被修改，然后使用 DES 进行解密，然后把该解密后的网页文件复制到目标文件中，然后再使用加密的过程对该文件再一次进行加密，并删除其被解密后的原网页文件。

### 4.2 停止模块

终止对目标文件进行监控的线程。对备份文件和监控文件不做任何修改。

### 4.3 察看日志

对监控文件的每一个操作都要进行相应的记录，然后保存到日志文件中，可以察看日志文件，但不能对日志文件做修改。

### 4.4 导出文件

因为有的时候由于物理的原因，原文件和备份文件都被遭到破坏，或者备份文件中的文件被删除，那么当目标文件中的文件被修改和删除时就无法进行恢复，这个时候就必须需要一个非本主机的备份文件，所以该模块就是把目标文件和备份文件备份到一个文件中，用来非本地的存储。然后当备份文件和原文件遭到破环时用来恢复文件。

### 4.5 告警模块

当文件被修改，删除和添加，以及重名命时，给用户一个右下角弹出提示和声音告警提示。

## 5 实验结果

选择目标文件和备份文件，然后开始监控目标文件，其过程如图 1 所示，包括复制，DES 加密，计算 MD5，重命名四个过程。当目标文件中的网页文件发生变化时，其过程如图 2 所示，包括 MD5 校验，DES 解密，用原名称重命名，复制到目标文件等四个过程，并且在界面上显示出目标文件中的操作时间，操作类

型，变化的文件和最后的恢复结果等信息。

| 序号 | 时间 | 操作类型 | 变化的文件 | 是否已恢复 |
|---|---|---|---|---|
| 0 | 2011:7:26:23:53:46 | 删除了文件 | testDelete.php | 已恢复 |
| 1 | 2011:7:26:23:54:1 | 添加了新文件 | testAdd.php | 已恢复 |
| 2 | 2011:7:26:23:54:11 | 重命名了文件 | testRename.ph... | 已恢复 |
| 3 | 2011:7:26:23:54:20 | 修改了文件 | testModify.php | 已恢复 |

图 4 程序运行结果截图

## 6 结语

该系统在文件监控的时候是开启了一个线程，该线程对监控文件进行实时的消息获取，如果文件被进行了任何操作，则会产生一个相应的消息，该线程就是专门获取该消息，然后根据消息的类别进行相应的处理，比如对修改文件，添加文件，删除文件，修改名称等消息做相应的处理，但是这里存在一个问题，就是如何保证该线程的安全？

众所周知，线程是可以终止的，假如该线程被终止了，那么就无法对目标文件进行实时监控，所以在本系统中需要对该监控线程做一个保护，这样就制止了非法的终止该线程的执行。

本系统在一定程度上对备份文件做了保护，可以防止备份被非法的篡改以及通过 MD5 码制止了被篡改的网页文件被恢复到监控目录，但是并没有彻底的实现对备份文件的保护，如果备份文件被彻底的删除则无法恢复网页文件，在下一步的研究中，可以通过非授权时禁止访问备份文件的方式对备份文件加以保护，或者只有最高权限的用户账号才能访问备份文件的方式保护备份文件，通过这两种方式，可以解决备份文件被删除后无法还原网页文件的问题，然后再结合使用本文中的 MD5 校验，DES 加密，文件重命名对备份文件中的网页文件加以保护，最终达到备份文件无法被篡改和删除的效果，从而在根本上解决备份文件的安全性问题。

## 参考文献


1 牛少彰,江为强.网络的攻击与防范—理论与实践.北京:北京邮电大学出版社,2006.

2 李枫,刘志永,马丽.网页反篡改技术研究及应用.电力信息化,2008,(7):124–126.

3 杨飞.网页防篡改技术.计算机安全,2008,(9):76–77.

4 张磊,王丽娜,王德军.一种网页防篡改的系统模型.武汉大学学报,2009,(1):121–124.

5 吴新华,周建.网页防篡改软件的设计与实现.南通纺织职业技术学院学报,2009,(4):18–20.

6 杨敏.网页防篡改安全研究.中国高新技术企业,2010,(25):72–74.

7 杨雷.网页防篡改安全技术的研究与实现.西安,西安电子科技大学,2008.

8 王海涛,杜宏伟.网站内容安全防护技术浅析.信息化研究,2010,(12):1–3.


（上接第 265 页）


2 Brown AW, Wallnau KC. The current state of CBSE. IEEE Software,1998,9:37–46.

3 Cai X, Lyu MR, Wong KF, et al Component-Based Software Engineering: Technologies, Development Frameworks and Quality Assurance Schemes. Proc. of the 7th Asia-Pacific Software Engineering Conference. 2000: 372–379.

4 徐征,陈雪飞,刘晓铭,等.一种构件化的动态软件系统模型.小型微型计算机系统,1999,(2):98–101.

5 张世琨,张文娟,常欣,等.基于软件系统结构的可复用构件制作和组装.软件学报,2001,12(9):1351–1359.

6 杨芙清,梅宏,李克勤.软件复用与软件构件技术.电子学报,1999,(2):68–75

7 杨芙清,王千祥,梅宏,陈兆良.基于复用的软件生产技术.中国科学,2001,31(4):363–371.

8 Draft IEC61970: Energy Management System Application Program Interface. Part 301: Common Information Model (CIM).Draft 6.

9 Draft IEC61970: Energy Management System Application Program Interface. Part 302: Common Information Model. Financial, Energy Scheduling and Reservation. Draft 2.

10 Draft IEC61970: Energy Management System Application Program Interface (EMS-API)-Part 302: Common Information Model (CIM) SCADA. Draft 2.

11 张慎明,刘国定.IEC61970 标准系列简介.电力系统自动化,2002,26(14):1–6.

12 郑志琴,钟叔玉.柔性 MIS 及其支撑技术.昆明理工大学学报,2001,26(2):8–11.