

一种网络不良图像过滤软件框架

陈智飞,王映辉

(西安理工大学 计算机科学与工程学院,陕西 西安 710048)

摘 要:网络信息监控技术针对网络中不良信息传播的问题,运用网络数据捕获、流量处理和内容过滤等技术实现对网络信息传播的有效监管。而不良图像过滤软件的架构是互联网信息监控中的一个关键问题。文中针对以 BHO 技术实现的图像过滤器插件存在的对非 IE 内核的浏览器无效等缺陷,设计并实现了一种基于 Winsock SPI 技术的图像过滤软件开发框架。实验表明,该图像过滤软件框架具有良好的普适性和兼容性,并对其它网络信息过滤软件的开发具有一定的借鉴作用。

关键词:软件框架;COM 构件;图像过滤

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2010)07-0149-03

A Software Framework for Filtering Sensitive Images on Internet

CHEN Zhi-fei, WANG Ying-hui

(School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Network information monitoring technology for the transmission of harmful information issues, use network data capture, traffic management and content filtering technologies to achieve the effective dissemination of information on network monitoring. The software architecture for filtering sensitive images on Internet is a key in the field of Internet information supervision. To compensate the drawback with BHO-based image filter plug-in which is invalid for non-IE kernel browsers, design and implement a framework for image filtering software based on Winsock SPI. The experiments show that the image filtering software framework has good universality and compatibility, and also can serve as a reference for developing other network information filtering software.

Key words: software framework; component object model; images filtering

0 引言

互联网在给人们带来资源共享的同时,也带来诸如暴力、反动,尤其是色情等不良信息的传播与泛滥,严重影响了青少年的身心健康。如何实现对互联网不良信息的监控,为人类营造一个健康的网络环境已经成为世界各国共同面对的难题与挑战,也成了人们所关注的全球性热点问题。为了防止网络不良信息的侵害,人们利用技术手段检测和过滤网络中的不良信息,主要包括基于内容的审计技术和基于内容的过滤技术。基于内容的审计技术一般采用旁路监听的方式,虽然不影响所监控网络的网络服务,但是只能提供事后报警,不能实时过滤不良信息。基于内容的过滤技

术通过收集网络中的关键数据包,对其传送的内容进行实时分析,将不安全的信息隔离在防火墙以外。这种技术能最大限度地保障网络信息内容的安全,因此成为内容过滤的关键技术^[1]。

在利用基于内容的过滤技术设计图像过滤软件时,现有的软件主要是基于 IE 浏览器助手对象 BHO (Browser Helper Objects) 技术将图像过滤器插件挂接到 IE 浏览器上来实现的^[2,3],其工作方式是当 IE 浏览器开始工作时,过滤器同时启动,监控 IE 事件,并在相应事件发生时进行网页图片信息的检测,根据检测结果决定对图片的取舍处理。BHO 插件虽然可以与 IE 进行无缝结合,但是对于非 IE 内核的浏览器如 Firefox, Chrome 等无效,并被当作“流氓软件”而禁用;此外还存在软件升级不便等问题。

为解决上述问题,文中给出了以采用 Winsock 服务提供者接口 SPI (Service Provider Interface) 为基础的不良图像过滤软件框架,其采用 SPI 技术编写客户服务处理函数,以 COM 构件的形式封装图像过滤模块,进而对该框架以防火墙的形式进行了实现。

收稿日期:2009-11-10;修回日期:2010-02-12

基金资助:陕西省科技项目(2007F51, 2008K4-11);西安市创新支持计划重点项目(CXY08003(3)),博士点基金(20070700002)

作者简介:陈智飞(1985-),男,硕士研究生,研究方向为软件框架和模式识别;王映辉,博士,教授,博导,研究方向为软件工程和模式识别。

1 图像过滤框架的设计

文中设计的图像过滤软件框架基于 Winsock SPI 技术,包括客户 SPI(Service Provider Interface)程序、数据捕获模块和图像过滤模块三大部分。其中客户 SPI 程序和浏览器等应用程序通过进程间共享变量方式保护客户 SPI 程序的地址空间,客户 SPI 程序和数据捕获模块之间通过命名管道实现进程间的相互通信,并以 COM 构件的形式实现图像过滤模块,通过 COM 接口和数据捕获模块进行交互,从而避免了在图像特征库更新和过滤算法改进时引起软件结构的改变。软件开发框架的工作流程如图 1 所示。

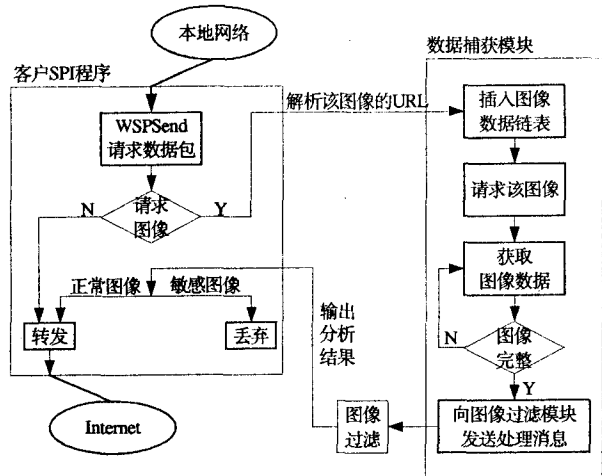


图 1 图像过滤软件框架流程图

与许强等人^[4]提出的基于图像内容的智能防火墙的设计思路不同的是,本图像过滤软件框架以本地网络作为工作流程的起点。首先对经过客户 SPI 程序中的 WSPSend 发送函数的 HTTP 请求数据包进行分析,如果为图像请求数据包,则解析出所请求的图像的 URL 地址,然后将该 URL 地址发送到数据捕获模块,数据捕获模块将生成相应的图像结点,插入图像数据链表,并根据该 URL 地址请求图像数据。当收到 HTTP 响应数据包时,开始对图像数据链表进行查找,当找到相应的图像结点时,将其插入到该链表的子链表中相应的位置处。如果该图像数据包已经全部进入链表,则向图像过滤模块发送处理消息,使其对该图像进行分析。图像过滤模块分析完成之后将分析结果发送到 WSPSend 函数,如果是正常图像,则转发该请求数据包到外部 Internet 网络,否则为不良图像,则丢弃该请求数据包,中断此次 Socket 连接。

下面将对该框架的客户 SPI 程序设计和图像过滤模块的设计等几个关键技术进行详细阐述。

2 客户 SPI 模块的设计

Winsock2 是一个可扩展的 API,通过引入 SPI 的

编程接口,可以在 Socket 中插入一层,以实现网络安全监控等功能,Winsock2 体系结构如图 2 所示^[5-8]。SPI 程序以动态链接库(DLL)的形式存在,它工作于应用层,为上层的 Winsock API 调用提供接口函数。当由用户编写的客户 SPI 程序安装后,所有的 Winsock 请求都会自动地发送到客户 SPI 程序,并由客户 SPI 程序完成对底层网络功能的调用,从而截获所有的应用层数据包。

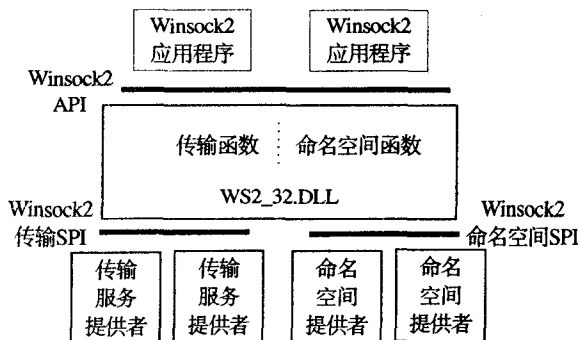


图 2 Winsock2 的体系结构图

客户 SPI 程序定义的初始化函数为 WSPStartup 函数,其中参数 lpProcTable 的类型为 WSPPROC-TABLE 结构,该结构保存了传输服务提供者的 30 个服务函数的指针。只要将客户 SPI 程序的函数地址赋给这些函数指针,当相应的函数请求发生时就会首先由客户 SPI 程序的函数来处理。由于只对发送函数 WSPSend 的请求数据包和接收函数 WSPRecv 的响应数据包进行分析和过滤,因此只需要重载这 2 个函数即可。图 3 给出了调用系统 SPI 程序和该软件的客户 SPI 程序的对比结构。

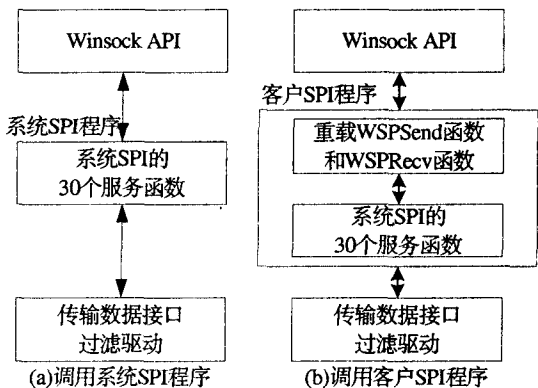


图 3 调用系统 SPI 程序和客户 SPI 程序的对比结构

3 基于 COM 的图像过滤模块设计

图像过滤模块采用基于 COM 的 VC 和 Matlab 混合编程技术^[9,10],通过 Matlab COM Builder 工具将 Matlab 编写的过滤程序封装成 COM 构件,通过 COM 接口和其他模块进行无缝结合。该方法实现简单,通

用性强,几乎可以调用 Matlab 的任何函数,并且在图像处理方面也具有很好的优势,因此文中采用这种方法实现图像过滤模块。

以下简单介绍如何在 Matlab 下编译成 COM 构件,以及如何在系统中调用该构件。

- (1)创建工程,并向工程添加 NudeImage.m 文件;
- (2)编译工程,将 NudeImage.m 文件编译成 COM 构件,即 NudeImage_1_0.dll;
- (3)调用该 COM 构件。将工程中的 mwcomtypes.h,nudeimage_idl.h,nudeimage_idl.i.c 文件拷贝到 VC 的工程目录下,将其添加在头文件中,并在程序中调用,核心代码如下:

```
Inudeimage * ptr=NULL;
HRESULT hr = CoCreateInstance (CLSID_ nudeimage, NULL,
CLSCTX_ ALL,IID_ Inudeimage,(void * *)&ptr);
VARIANT n,filename;
VariantInit(&n);
VariantInit(&filename);
n.vt=VT_BOOL;
filename.vt=VT_BSTR;
filename.bstrVal = _com_ util:: ConvertStringToBSTR ( ImagePath);
if(SUCCEEDED(hr))
{ ptr->nudeimage(1,&n,filename);
bool value = false;
if (n.vt == VT_BOOL)
{ value = n.boolVal ? true:false; }
if(value)
{ AfxMessageBox (“此图片为不良图片!”);
return true; }
else
{AfxMessageBox (“此图片为正常图片!”);
return false; }
ptr->Release();}
else
{ AfxMessageBox(“COM 构件初始化失败!”);
return false; }
```

- (4)打包和发布 COM 构件。需要注意的是,编译结束后产生的 COM 构件只能在当前计算机上使用,如果要在另一台机器上运行的话,必须打包该 COM 构件,并在目标计算机上安装和注册后才能使用。

当新的过滤算法构件安装时,旧版本的算法构件将被自动替换,在用户无障碍的情况下实现软件的版本升级。

4 实验结果分析

测试基于文中给出的框架开发的图像过滤防火墙

软件对于不同内核浏览器的兼容性。针对 Windows IE7、Mozilla Firefox3、Google Chrome2.0、Opera Opera10 和 Apple Safari4 五款独立内核的浏览器,以 BHO 图像过滤器插件和文中给出的软件框架两种方法进行兼容性测试,测试结果如表 1 所示。

表 1 不同内核浏览器的兼容性

采用的方法	IE7	Firefox3	Chrome2.0	Opera10	Safari4
BHO 插件	√	×	×	×	×
图像过滤防火墙	√	√	√	√	√

实验结果表明,基于该框架开发的图像过滤防火墙软件不仅能支持 IE 浏览器,还能支持多种非 IE 内核的浏览器,可以克服 BHO 插件存在的缺陷。

从实验结果的表中可知,BHO 图像过滤器插件仅仅对 IE 浏览器有效,而对于其他非 IE 浏览器无效。这是因为 BHO 插件主要是基于 IE 浏览器助手对象 BHO 技术实现的,只能监测到基于 IE 内核的浏览器的典型事件,而无法监测到非 IE 内核的浏览器如 Firefox,Chrome 等的典型事件,所以对其无效。另外,BHO 插件还被众多杀毒软件当作“流氓软件”而禁用,因而其具有很大的局限性。

而文中提出的图像过滤软件框架是基于 Winsock SPI 接口技术设计的,而该技术对于应用层软件透明。安装基于该框架开发的防火墙软件后,任何内核的浏览器的 HTTP 请求都会自动地发送到客户 SPI 程序,并由客户 SPI 程序完成对底层网络功能的调用,从而截获所有的 Winsock 调用,因此具有良好的兼容性。

5 结束语

文中的重点不在于不良图像的判定,而是给出了一个网络不良图像过滤软件开发框架。在该框架中,只要利用 COM 技术,将实现了的不良图像过滤算法进行封装并嵌入之,即可实现对不良图像的有效过滤。此外,基于 Winsock SPI 接口技术设计的网络图像过滤软件开发框架,具有良好的兼容性,克服了传统的以 BHO 实现的图像过滤器插件的缺陷,并以 COM 接口作为各个模块之间相互通信的桥梁,有利于软件的版本升级;同时,该框架对网络上其它信息过滤软件的开发也能提供很好的借鉴和支撑。

参考文献:

[1] 孙钦东,管晓宏,周亚东.网络信息内容审计研究的现状及趋势[J].计算机研究与发展,2009,46(8):1241-1250.
[2] 范 晓,申铨京.基于 IE 浏览器的色情图片过滤器[J].吉林大学学报:信息科学版,2004,22(6):631-637.

3200m 数据进行了读写测试。结果如图 5 所示,漂移之前原始逻辑卷的读写平均速度为 28~29MB/s,漂移后通过 iSCSI 读写 IO 的平均速度为 10MB/s。造成这样的主要原因在于本系统使用的是百兆网卡,理论上的最高下行速度为 12.5MB/s,所以 iSCSI 传输速度实际上已经达到了网卡传输速度的上限。虽然目前漂移后 iSCSI 的性能比起原始磁盘的读取速度还有一定差距,但是在使用了千兆网卡后性能还能够获得进一步的提升。

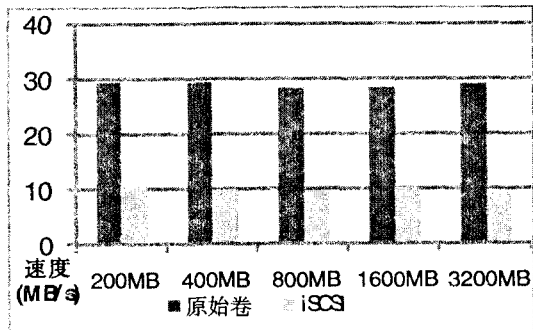


图 5 服务漂移的性能测试

在服务漂移切换时间的测试中,启动脚本主要执行两个操作,一是启动 Windows Server2003 上的 IIS 服务,二是启动 Microsoft SqlServer 服务。结果如图 6 所示,服务漂移的平均切换时间为 17.2s,这样的切换时间能够比较快速地完成服务漂移的切换,不会对服务中断造成太大的影响。

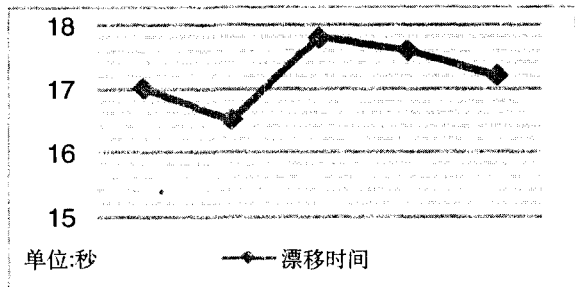


图 6 服务漂移的切换时间测试

3 结束语

提出并实现一种基于现有数据容灾系统的服务漂移方案,该服务漂移功能可以在客户服务器遭到灾难或故障时有效地实现服务连续性和数据一致性,并且具有较快的服务切换时间和,在 IO 性能上也具有较大的提升潜力。该方案是对现有数据容灾系统功能一个有力的扩充,适合作为中小型系统的服务容灾方案。

参考文献:

- [1] 李 涛,刘晓洁,曾金全,等. 信息系统容灾抗毁原理与应用[M]. 北京:人民邮电出版社,2007.
- [2] 李瑞浴,刘晓洁,李 涛,等. 一种基于 ICE 通信的数据库容灾系统[J]. 信息安全与通信保密,2007(3):121-123.
- [3] 卢正添,李 涛,赵 奎,等. 一种数据库服务多点容灾系统[J]. 计算机应用研究,2008,25(12):3707-3709.
- [4] 易固武,刘晓洁,李 涛,等. 一种网络备份系统的数据一致性检测方法[J]. 计算机应用研究,2008,25(12):3700-3701.
- [5] 张 丁,刘晓洁,李 涛,等. 数据库服务容灾系统的设计与实现[J]. 计算机工程与设计,2008,29(12):3247-3251.
- [6] Stevens W R. TCP/IP Illustrated Volume1: The Protocols [M]. [s.l.]:Addison Wesley, 1994.
- [7] 董欢庆,李战杯,林 伟,等. Linux 平台远程逻辑卷复制系统的设计[J]. 计算机工程与应用,2004(18):109-112.
- [8] 王彦龙,李战杯,董欢庆,等. Linux 平台远程逻辑卷复制系统的灾难恢复策略[J]. 微电子学与计算机,2005,22(3):68-72.
- [9] Symantec(VERITAS). VERITAS global cluster manager: technical over view[R]. [s.l.]:[s.n.],2002.
- [10] EMC. EMC SRDF family: powerful remote replication solutions improve organizational productivity, enhance online business continuity[R]. [s.l.]:[s.n.],2008.
- [11] Bryant R E,O'Hallaron D. Computer Systems: A Programmer's Perspective[M]. [s.l.]:Prentice Hall, 2003.
- [12] Stevens W R,Rago S A. Advanced Programming in the Unix Environment[M]. [s.l.]:Pearson Education, 2000.

(上接第 151 页)

- [3] Girgis M R, Mahmoud T M, Abd - El - Hafeez T. An Approach to Image Extraction and Accurate Skin Detection from Web Pages[J]. International Journal of Computer Science and Engineering,2007,12(2):88-96.
- [4] 许 强,江 早,赵 宏. 基于图像内容过滤的智能防火墙系统研究与实现[J]. 计算机研究与发展,2000,37(4):458-464.
- [5] 甘利杰,丁明勇,杨永斌. 基于 Winsock SPI 技术的包过滤研究[J]. 计算机科学,2007,34(8):122-124.
- [6] 吴 星,陈明锐. 恶意网页防护系统的设计与实现[J]. 计

算机技术与发展,2009,19(5):154-157.

- [7] 李石君,李 洲,余 军,等. 基于 URL 过滤与内容过滤的网络净化模型[J]. 计算机技术与发展,2009,19(1):5-8.
- [8] 周瀚宇. 基于 URL 的网页内容过滤器的设计与实现[J]. 计算机工程,2006,32(7):81-83.
- [9] 王映辉,冯德民. 大规模软件架构设计[M]. 北京:科学出版社,2003.
- [10] Drimbarean A F. Image processing techniques to detect and filter objectionable images based on skin tone and shape recognition[M]. [s.l.]:IEEE,2001.