基于 Winpcap 的邮件还原系统的实现*

吴志强,马春波,敖发良
(桂林电子科技大学 信息与通信学院,广西 桂林 541004)

摘 要:随着互联网的普及,SMTP/POP3 协议传输方式下的邮件已成为最简便、最经济的通信方式,但许多有害的邮件信息也随之而来。针对这些问题,在 VC 开发环境下,研究并实现了基于 Winpcap 的邮件监控及还原系统。该系统设计了多线程模块,可以同时处理 50 个邮件及附件信息,并对网络数据包的捕获、过滤与重组、信头解码、附件处理、信体内容提取等模块进行了算法优化。功能测试和验证结果表明,该系统稳定、可扩展,达到实时性要求。

关键词:Winpcap;数据包捕获;邮件重组;多线程;邮件解码
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1674-7720(2011)02-0058-04

Implementation of an email recovering system based on Winpcap

Wu Zhiqiang, Ma Chunbo, Ao Faliang
(College of Information & Communication, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: With the prevalence of the Internet, email of SMTP/POP3 protocol has become the most convenient and economical communications means, yet along goes a great deal of harmful information. According to these problem, this article proposes an email security monitoring and recovering system which is based on Winpcap in VC software environment on Windows system. This system which designs the multi-thread modul can deal with 50 email and information of accessories at one time the system presents some improvement in algorithm, includeing capturing, filtering and reordering of the network packages, email-header decoding, attachment-processing and bodies-content and so on. The experiment indicates that this system can be extended greatly, good stability and meet the real-time request.

Key words: Winpcap; packets capture; email reassemble; multi-thread; mail decoding

当前流行的电子邮件收发主要有两种方式:第一种是使用 Fire-Fox、IE 等浏览器软件登录 ISP 的 Web 站点;第二种是使用 Foxmail、Outlook 等邮件客户端软件连接邮件服务器,通过 SMTP/POP3 协议收发邮件。这里主要讨论基于 SMTP/POP3 协议邮件信息的获取与还原技术。实验结果表明,本系统有效实现了邮件信息捕获和信息还原等功能,具有较高的效率和较好的时实性。

1 网络监控相关理论基础

1.1 Winpcap 原理

Winpcap[1-3]为数据包捕获提供了一套标准接口,它是由伯克利分组捕获库派生而来的分组捕获库,在 Win32 平台上实现对底层包的捕获,其体系结构包括 3 个模块:(1)NPF(内核级的数据报过滤器)核心的包过滤驱动程序;(2)底层的动态连接库 Packet.dll(数据包低级驱动程序库)为 Win32 平台提供了一个公共的接口;(3)高层的独立于操作系统的库 Wpcap.dll(数据包高级驱动程序库)[4]。

1.2 网络监控及邮件信息获取原理

监控主体可以分为邮件服务器端监控、邮件客户端监控和第三方网络监控。本文阐述的是基于第三方网络监控的邮件安全监控系统,即将网络上的关于邮件协议的数据包进行截获,并结合内容检测技术和协议分析技术对邮件进行有效监听。这种监控方式允许在各个组织

* 基金项目:国家自然科学基金(60862001);广西教育厅基金(200808MS004);
广西信息与通讯技术重点实验室基金(10908);广西研究生科研创新项目(20101059508m15)

内部灵活配置监控规则，并且响应迅速。

网卡具有 4 种工作模式：直接模式、多播传送模式、广播模式和混杂模式[5]。网卡的缺省工作模式包含广播模式和直接模式，即它只接收广播帧与发给对应地址的帧。如果采用混杂模式，网卡将接收同一网络内所有主机所发送的数据包。

1.3 SMTP/POP3 协议的邮件信息获取

由于系统主要分析 SMTP/POP3 协议传输方式下的邮件信息获取，因此主要针对 SMTP 协议、POP3 协议网络数据进行分析。根据协议特点只需要对感兴趣的网络连接进行数据重组，其他协议的网络数据可以直接丢弃。所以只要获取端口地址为 25 和 110 的数据包即可，如图 1 所示。

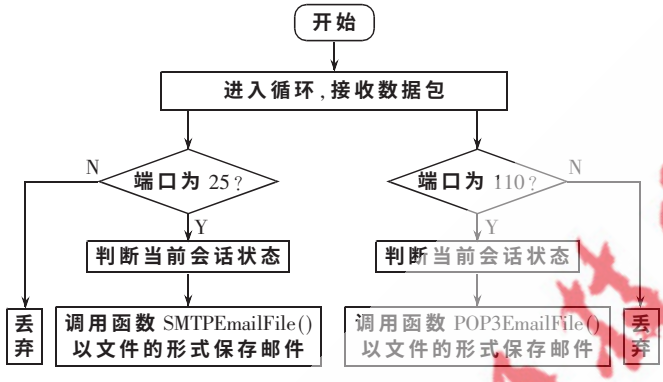

图 1 SMTP/POP3 协议的邮件信息获取流程图

2 监控邮件及邮件信息还原的分析与设计

整个系统的设计可分为两个部分：一个邮件监控部分，计算机运行在 Windows 系统下，目的是截获经过网卡的邮件数据包；另一个是邮件解析和还原部分，主要目的是对 SMTP/POP3 文件进行解析，还原成原来的可读邮件信息，并对还原后的文件进行阅读查看、删除等管理操作，系统的总体框架如图 2 所示。系统由包捕获功能模块、IP 协议数据解析模块、TCP 协议数据解析过滤模块、邮件协议解析模块、邮件解码模块、邮件保存模块组成。

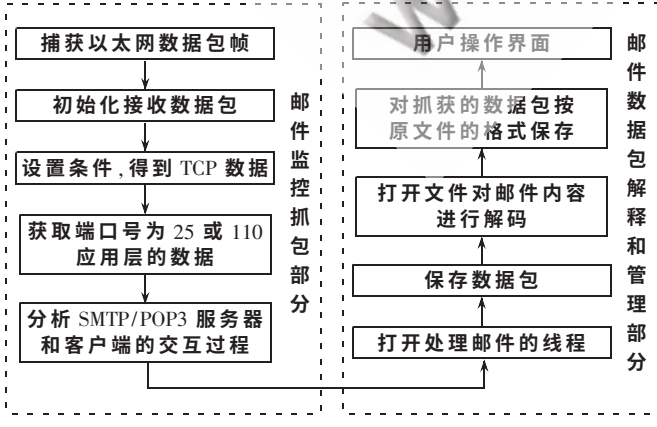

图 2 系统的总体框架

下面重点分析系统中邮件重组和解码等关键问题和算法的设计。

2.1 SMTP/POP3 协议通信方式

SMTP/POP3 协议采用会话方式工作，通信双方通过命令请求和命令响应进行交互，完成邮件的发送或接收。提取基于 SMTP/POP3 协议传输的邮件信息关键在于邮件会话的识别。下面以 POP3 为例，说明交互过程[6]。

2.2 POP3/SMTP 还原总流程图及还原算法

基于 POP3/SMTP 协议的邮件信息还原涉及四方面的内容：数据包保存、重要域值提取、邮件内容的解析以及附件的解析还原。

下面主要介绍面向传输层的数据分析还原，总流程如图 3 所示。

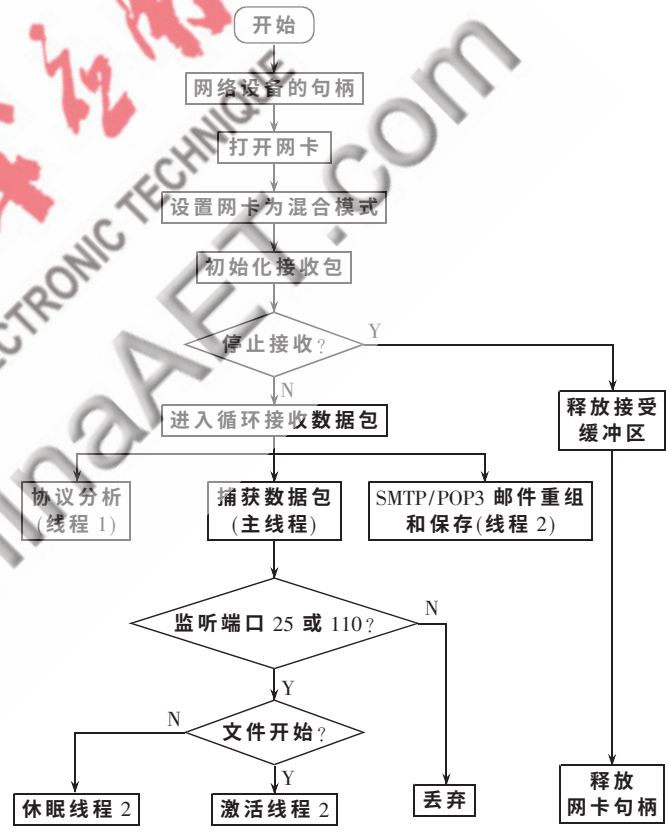

图 3 网络数据包分析流程图

在对邮件协议交互过程有很深刻的认识后可对电子邮件数据包进行截获。为了提高运行速度，防止丢包现象，提高还原文件效率，优化程序，作者创建 3 个线程分别执行任务：

(1)主线程是 Winpcap 抓包，整个系统在贯彻主线程的基础上对网络层、传输层和应用层进行了分析和研究；

(2)线程 1 是协议分析，网络通信有很多协议，因此协议分析是关键，针对不同的协议有不同的分析方法，因此系统具有良好的可扩展性，可方便地添加对新的网络协议的支持；

(3)线程 2 是对电子邮件数据包的解析并还原,因此当捕获邮件文件数据包时,打开线程组 2,可以同时循环接收和保存 50 个邮件,不过此时这些线程是休眠状态,可能客服端只是对 SMTP/POP3 服务器的一般交互过程,没有邮件传输,因此要仔细观察和分析整个邮件服务器交互过程,如果分析客户端邮件发送动作时,触发线程 2 进行邮件信息的重组、保存,直到完整地保存了整个邮件及附件后关闭线程。

此时作者设计了 SMTP/POP3 邮件信息重组子流程图,如图 4 所示。

邮件数据包重组的策略也是文件还原系统的核心,下面详细说明捕获和重组算法:

(1)根据 SMTP/POP3 命令的参数,监听端口号 25(或 110)来分析服务器与客服端之间的交互过程;

(2)对经过网卡数据包的命令进行字符串匹配,当匹配的字符串 DATA(Data)成功时;客服端要向 SMTP 服务器发送邮件,创建文件名为 EmailFile[file_num/50],这时打开可以同时处理 50 个邮件的线程 2 进行监听,跳至(4);

(3)当匹配的字符串 RETR 成功时,客服端要向 POP3 客户端接收邮件,创建文件名为 EmailFile1[file1_num/50],这时打开可以同时处理 50 个邮件的线程 2 进行监听,跳至(5);

(4)把 SMTP 数据包的当前序列码为 sequence 和数据偏移量 data_len 保存,设置标记位 STRAT=1 下一个包的序列码为 next_seq_num[file_num/50]。其大小为:next_seq_num[file_num/50]=sequence+data_len;保存发送邮件;跳至(6);

(5)把当前 POP3 数据包的序列码为 sequence 和数据偏移量 data_len 保存,设置标记位 STRAT1=1 下一个包的序列码为 next_seq_num[file1_num/50],其大小为:next_seq_num[file1_num/50]=sequence+data_len;保存接收邮件,跳至(7);

(6)开始重组 SMTP 邮件数据包并把当前序列码记为 sequence 和数据偏移量 data_len,上次一个包的序列码为 next_seq_num[i]。以大小为 50 的循环序列进行一一判断;当满足 next_seq_num[i]==sequence,则为该文件 EmailFile[file_num/50]的数据,不等则丢弃,这样的目的是为了在监听多邮件传输时确保不会导致捕获的数据混乱,跳至(8);

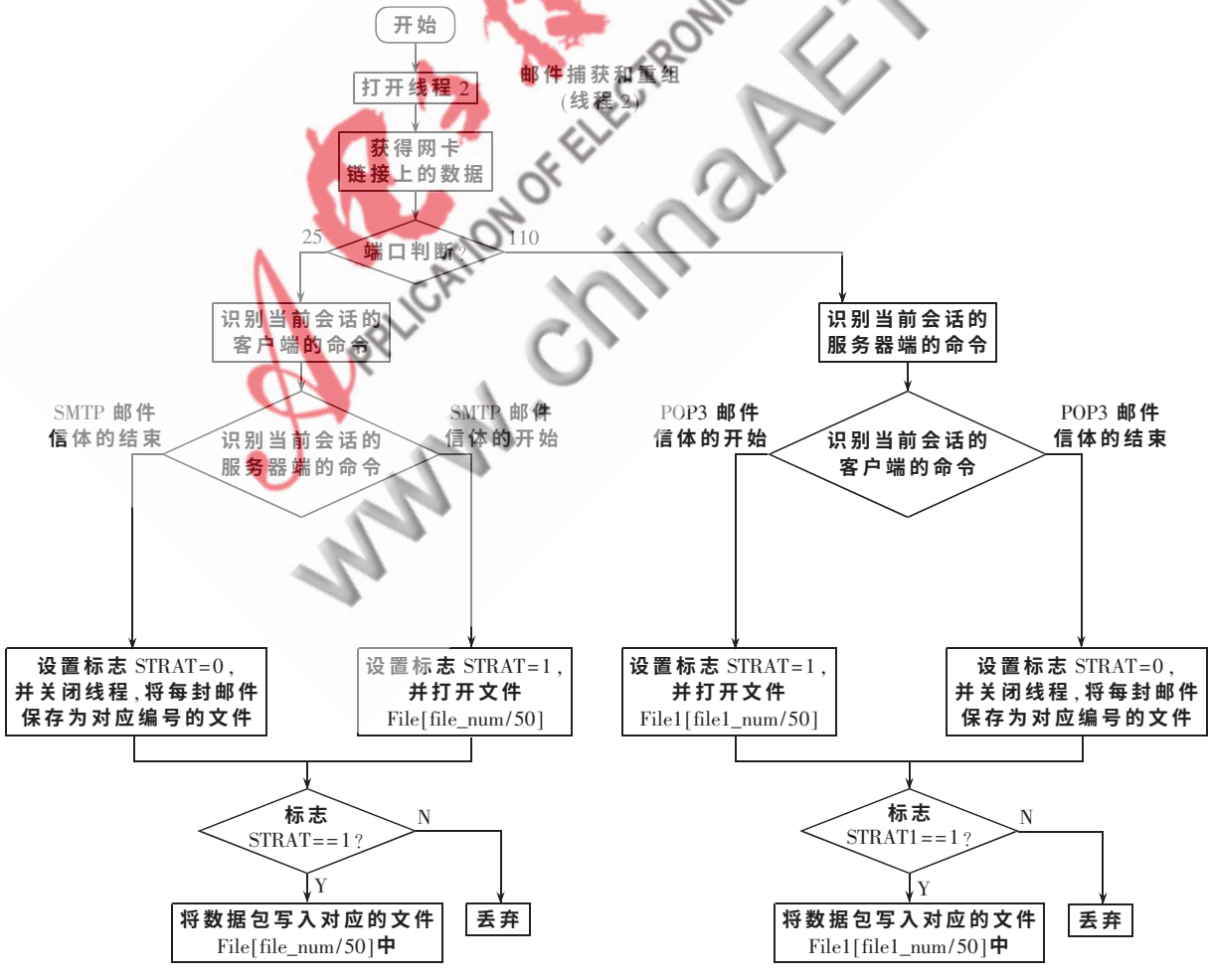

图 4 邮件捕获和重组流程图

(7)开始重组 POP3 邮件数据包并把当前序列码即为 sequence 和数据偏移量 data_len，上次一个包的序列码为 next_seq_num[i]。以周期为 50 的序列进行判断：next_seq_num[i]==sequence，则为该文件 EmailFile1[file1_num/50]的数据，不等则丢弃，这样的目的是为了在重组多邮件传输时确保不会导致捕获的数据混乱；

(8)当数据包以一个 QUIT 命令来结束数据连接，则邮件 SMTP/POP3 邮件传输结束，设置标记位 STRAT 为 0，设置标记位 STRAT1 为 0；

(9)关闭文件，关闭线程，停止邮件捕获和重组子程序。

从算法可以得出，首先识别邮件会话状态的方法来确定邮件信息、重组邮件数据包，并保存在临时文件夹，待邮件数据包解析和还原时调用。只需捕获端口是 80、25 或 110 的数据包。

2.3 电子邮件内容的提取与解码模块的设计与实现

互联网上使用的电子邮件基本都遵循 MIME 规范，MIME 邮件传递实际是一个经过特殊编码并以约定格式进行网络传输的，因此只需提取存储在邮件各种域中的位置、格式和编码信息，根据这些信息从字符序列中提取出对应的字符内容对其进行解码，就可以得到需要的有关内容。下面是带附件和不带附件的邮件信息提取和解析模块设计。流程图如图 5 所示。

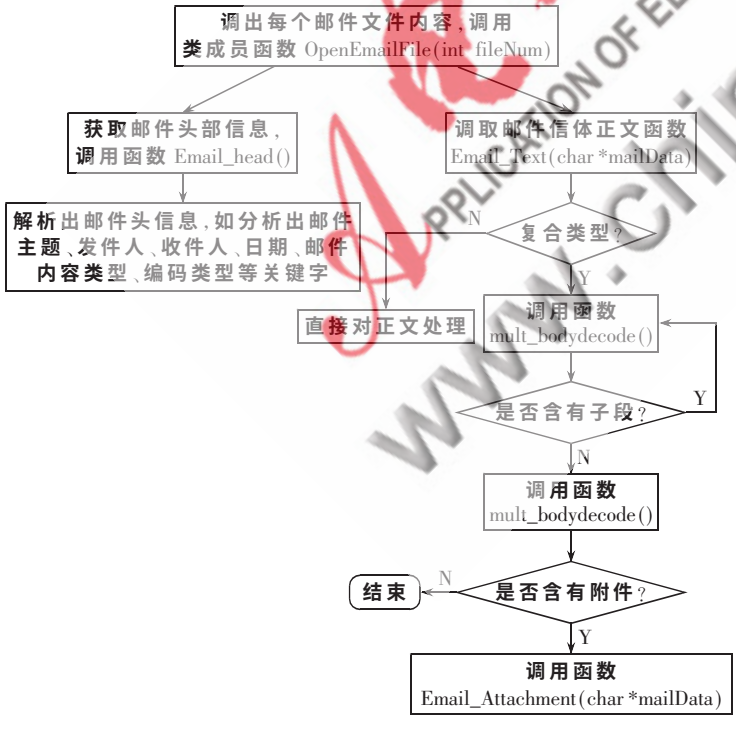

图 5 邮件内容解析流程图

这个部分主要是邮件内容的关键字匹配，主要采用精确关键字匹配，它将待检索的数据串和关键词组成的模式串进行逐字比较，只有在数据串中发现与模式串完全一致的部分之后，关键词匹配才算成功。把邮件各部分内容读取到字符串数组中，再根据 MIME 规范进行编写相应的解码函数，经过邮件解析，提取电子邮件各部分(如发信人、收信人、主题、正文等)，并对包含编码的部分进行解码(Base64、Quoted-Printable 等)，还原为可理解的电子邮件。邮件正文数据包可能分几个子段进行传输，此时要循环调用函数 mult_bodydecode()，当有附件时，先把正文内容解码之后，才进行附件的解码，这也是调用附件解码函数 Email_Attachment()对附件进行还原。直到整个邮件解析完成才关闭文件和线程，完成邮件的还原。

本系统主要是在编程上优化，提高数据包的分析和解码效率和速度，由本文可以看出邮件数据分析、过滤、重组、解析、还原算法建立在基于 Winpcap 基础之上，因此，该系统保留了 Winpcap 的抓包特点的同时，也克服了 Winpcap 部分不足。主要体现在以下两点：进一步提高了分析的速度，同时大大提高分析的准确性；节约了协议分析时间。

文中设计的网络安全监控系统，选择 Windows2000 作系统平台，利用 VC++6.0 编写程序，它是基于系统的底层进行设计，与操作系统紧密结合。通过在局域网中的使用和分析，它能实时地、动态地对局域网内的所有上网主机进行监视、控制与管理，系统稳定、效果好。

参考文献

[1] 张伟，王韬，潘艳辉，等.基于 Winpcap 的数据包捕获及应用[J].计算机工程与设计，2008，29(7)：1649-1651.

[2] 循序渐进学习使用 Winpcap[EB/OL].中国协议分析网.http://www.cnpaf.net/，2005.

[3] 李雪莹，刘宝旭，许榕生.基于 WinPcap 的网络监控系统性能优化[J].计算机工程，2004，30(1)：8-9.

[4] 赵英男，张秉权.MIME 邮件结构格式分析[J].软件技术，2001，20(2)：50-53.

[5] 秦根建，张秉权.网络数据包截获机制研究[J].兵工自动化，2003，22(6)：2-3.

[6] 唐燕.POP3 协议解析及简单实现[J].网络通讯与安全，2007，16(2)：951-952.

(收稿日期：2010-06-28)

作者简介：

吴志强，男，1983 年生，硕士研究生，主要研究方向：网络安全，信息还原。