IM即时通信统一监控管理技术研究与实现

郭思薇 马兆丰 蒋 铭 钮心忻 杨义先
(北京邮电大学信息安全中心 北京 100876)

摘 要 通过对主流即时通信软件的通信机制进行分析,解析整理出多种即时通信文本传输协议,设计并实现了一个即时通信统一监控管理系统-IMSMMS。IMSMMS通过会话关联,解决了先前的消息发送方和接收方的有效截取问题。IMSMMS不仅能对主流即时通信(MSN、Fetion、雅虎通等)的文本信息进行提取,还能够对MSN和雅虎通这两个不同即时通信软件之间互相发送的文本信息进行有效提取,并且通过设置敏感词,能够过滤出涉及敏感词的消息。实验数据表明,IMSMMS在通过网关携带即时通信文本信息的数据包每秒小于1000个时,仅有小于0.21%的漏检率。所以,IMSMMS对于中小型规模企业网,能够进行有效的监控。

关键词 即时通信,协议分析,会话关联,监控管理
中图法分类号 TP393 **文献标识码** A

Research and Implementation of Instant Messenger Standard Monitor Management Technology

GUO Si-wei MA Zhao-feng JIANG Ming NIU Xin-xin YANG Yi-xian
(Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract Based on the analysis of communication mechanism of mainstream instant messenger software, summary of the various instant messengers' transmission protocol of texts, a standard monitor system for instant messenger-IMSMMS was designed and realized. Through session association, IMSMMS can intercept the sender and the receiver of the message effectively. IMSMMS not only extracts the text message of the mainstream instant messengers (such as MSN, Fetion, Yahoo! Messenger etc.) for multiple versions, but also extracts the text message between two different kinds of software: MSN and Yahoo! Messenger. Through setting the sensitive words, the system can filter out messages that contain sensitive words. The experiment results show that when the IM text message packets though the gateway is less than 1000 per second, IMSMMS only has a rate of missing packet less than 0.21%. The result indicates that IMSMMS can meet the network security requirements of small scale Intranet effectively.

Keywords Instant messenger, Session association, Protocol analysis, Monitor management

1 引言

作为一种便捷的通信手段,即时通信(Instant Messenger,IM)软件已经越来越深入人心,应用范围从单纯的网络聊天工具变成了工作所不可或缺的信息交流平台。即时通信软件的功能从最初的文字聊天发展到如今的能够支持文件传输、语音聊天、视频聊天、远程协助、发送邮件、浏览网络等功能。国内的飞信甚至能够发送短信到手机或者对应的飞信客户端,实现手机与PC间免费的文本通信;MSN与雅虎通从2006年7月份开始实现文本之间传输的互通,形成全球最大的IM社区,共同占领超过44%的全球市场份额,全球将近一半的IM用户第一次实现互通。

即时通信技术在给个人以及企业带来高效便捷的信息交流的同时也产生了一系列的安全问题。首先,即时通信软件由于本身的安全性不高,为黑客攻击和病毒入侵提供了方便。其次,对于企业而言,即时通信软件不仅使得员工工作效率降低,占用大量网络带宽,而且容易导致泄漏机密、窃取资料等情况发生,从而给企业造成无法估量的损失。因此,对IM进行统一监控管理是网络管理面临的迫切问题[1-5]。

目前已有多篇文献研究了即时通信统一监控管理的相关问题。文献[6]对即时通信协议进行分析,并提出了5种阻断IM的方法,验证了每一种阻断算法的效率。文献[7]提出了一个通用协议解析处理模型,并采用LEX工具自动生成协议解析器,实现了对MSN文本消息的监控和解析。文献[8]利用网络驱动接口规范技术,设计与实现了一种监控系统来对飞信的文本与文件传输进行有效监控。

但是目前的IM通信统一监控管理系统还是存在许多的不足,主要方面有:

1)现有的安全管理软件一般只能针对特定的IM软件,不能进行多个主流IM软件的统一监控管理;

到稿日期:2012-06-12 返修日期:2012-09-21 本文受国家自然科学基金(60803157,90812001),国家242项目(2009A105)资助。
郭思薇(1990-),女,硕士生,主要研究方向为信息安全等,E-mail:guosiwei627@163.com;马兆丰(1974-),男,博士,讲师,主要研究方向为数字版权保护、数字内容安全、计算机网络安全;蒋 铭(1984-),男,博士,主要研究方向为信息安全、数字水印、多媒体信号处理;钮心忻(1963-),女,博士,教授,主要研究方向为数字水印、信息隐藏、隐写分析;杨义先(1961-),男,博士,教授,主要研究方向为密码学、计算机网络与信息安全。

2)由于 MSN 文本消息格式的特殊性以及没有完善的会话关联体系,因此难以提取出 MSN 的发送方和接收方;

3)现有的 IM 安全监控软件缺少对 MSN 和雅虎通两个不同的即时通信软件实现互通之后的安全监控;

4)IM 安全管理还面临着 IM 软件升级之后的监控问题。

本文在实验基础之上,对主流 IM 通信机制进行了研究,深入分析并概况了其交互过程中的协议格式,总结出其应用层的特征字符串。设计并实现了一个 IM 统一监控管理系统 IMSMMS(IM Standard Monitor Management System)。该系统不仅能对 MSN、飞信、雅虎通等即时通信软件进行单独管理,还能对 MSN 与雅虎通之间的文本通信进行监控,实现统一监控管理。并且捕获到的消息若含设定的敏感词,会向前台提出报警提醒监控者,同时保存记录在日志文件中,保证监控域内的安全性。

2 即时通信软件的通信架构

主流即时通信软件的基本通信架构同样采用 C/S 模式,即客户端与客户端之间的文本消息通过服务器转发。MSN 的应用广泛而且其通信协议 MSNP 被广泛地应用到各种聊天工具的开发中[9]。而飞信以及雅虎通的通信协议比 MSN 的通信协议简单,所以本文主要分析 MSN 的文本消息传输。

MSN 使用 TCP 传输协议,除了文件传输和语音聊天是直接的点对点通信之外,其他所有的情形全部通过服务器中转。主要有 3 种类型的服务器[10]:

1)派遣服务器(Dispatch Sever,DS),是客户端最先连接的服务器。主要功能是协商协议版本和向客户发送可用的通知服务器 IP 及端口。派遣服务器与客户端连接后,根据从客户端处得出的信息,给客户端分配合适的通知服务器,然后断开与客户端的连接。派遣服务器的端口号是 1863。

2)通知服务器(Notification Server,NS),是客户端需要一直连接的服务器。通知服务器负责大部分非中转的交互,包括登录,改变状态(隐身、在线等),获取用户列表,修改用户列表,发起聊天等。通知服务器的服务端口号由派遣服务器指定,一般为 1863。

3)接线服务器(Switchboard Server,SB),是客户端间通信所使用的中转服务器。每开一个聊天窗口,客户端就与服务器建立一个 TCP 会话。文件传输和语言聊天等,为发起点对点的会话,也需要从接线服务器上得到对方的 IP 地址等信息。接线服务器的服务端口由通知服务器指定,一般为 1863。

MSN 的 C/S 模式通讯架构如图 1 所示。

图 1 中,过程依次为:

1)MSN 软件用户登录客户端,客户端通过 DS 找到 NS;

2)MSN 客户端与 NS 建立 TCP 连接;

3)在 TCP 连接成功之后,客户端将用户输入的 ID 和密码进行加密,之后发送到 NS 中;

4)服务器在验证通过后,NS 与客户端进行各种非中转的交互,如登录、获取用户列表和更新个人信息等。客户端与 NS 一直连接,除非客户退出 MSN。当然,当客户端打开一个聊天窗口发出聊天请求时,NS 会给客户端分配一个 SB,然后 5)和 4)就并行运行了;

5)客户端通过连接服务器 SB 与好友聊天。聊天过程中,若 A 为会话发起端,客户端与 SB 间的交互具体过程如图 2 所示。图中的通信过程一直持续到某客户端关闭聊天;

6)断开连接,响应 NS。

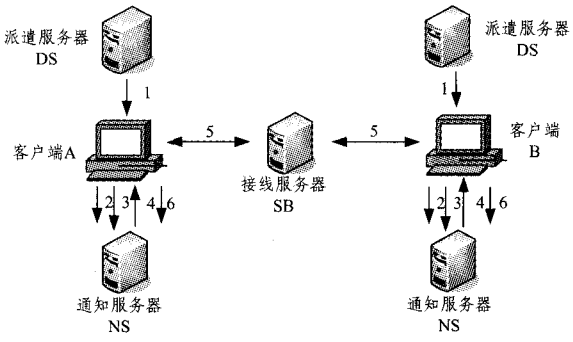

图 1 即时通信系统架构

图 2 典型文本消息通信过程

3 MSN 文本协议分析

为了实现对 MSN 的监测,需要对其协议进行分析。由于 MSN 文本数据在应用层是以明文传输,因此在分析测试中,通过有目的的发送,利用 Wireshark 的过滤规则,获取发送的文本数据包,然后分析其数据包的组成规律[11],从而可以总结出 MSN 协议的特征信息,如表 1 所列。

表 1 MSN 协议的特征信息

通信方式	C/S P2P 混合
网络传输协议	TCP/IP
握手协议	HTTP
服务器端口号	1863
应用层协议	MSNP
应用层协议组织方式	0x0D 0x0A 分割应用层协议数据段
文本/文件消息格式	Text/plain
文本/文件编码方式	UTF-8
文本/文件传输形式	明文

下面对于关键数据包的协议格式进行分析:

1)客户端登录 DS 服务器成功数据包

这是客户端在登录 MSN 成功(包括验证登录帐号和密码正确、成功获取用户列表)之后,DS 服务器发送给客户端的数据包,该数据包格式如图 3 所示(空白表示分割符"0x20",阴影表示分割符"0x0d 0x0a",下同)。

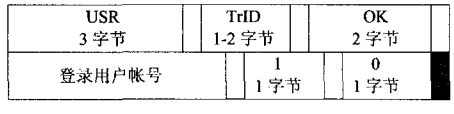

图 3 客户端登录 DS 成功数据包的协议格式

2)SB 通知客户端会话建立成功的数据包

这是本地用户作为会话的邀请者向 SB 服务器发起会话要求,SB 服务器在确认被邀请者加入会话之后发送给本地用户(邀请者)的会话建立的成功的数据包,如图 4 所示。

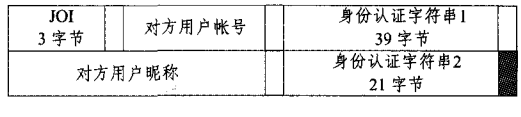

图 4 SB 通知客户端会话建立成功数据包的协议格式

3)客户端发送文本消息的数据包

这是文本信息传输时客户端发送给服务器的包含文本信息的数据包,如图 5 所示。

MSG 3 字节	TrID 1-2 字节	N 1 字节	消息总长度 2-4 字节
MIME-Version: 版本号 17 字节	其它数据		消息内容(明文)

图 5 客户端发送文本消息数据包的协议格式

4)客户端接收文本消息的数据包

这是文本信息传输时服务器端发送给客户端的包含文本信息内容的数据包,协议格式如图 6 所示。

MSG 3 字节	发送方账号	其它数据
MIME-Version: 版本号 17 字节	其它数据	消息内容(明文)

图 6 客户端接收文本消息数据包的协议格式

5)客户端发送消息给脱机用户的数据包

这是文本信息传输时,客户端发送给 NS 服务器的包含文本信息的数据包,如图 7 所示。

UUM 3字节	TrID 1-2字节	对方用户帐号	未知数据	1 1字节
消息总长度 2-4字节	MIME-Version:版本号 17字节	其他数据		消息内容(明文)

图 7 客户端发送给脱机用户数据包的协议格式

6)客户端发送消息给雅虎通用户的数据包

MSN 客户端发送消息给雅虎通用户的数据包与 5)格式一致。

7)客户端接收雅虎通用户发送的文本消息的数据包

MSN 客户端接收到雅虎通用户发送的文本消息,是由 NS 发送到 MSN 客户端的,其数据包格式如图 8 所示。

UBM 3字节	对方用户帐号	未知数据	接收用户帐号	1 1字节	1 1字节
消息总长度 2-4字节	MIME-Version:版本号 17字节		其他数据		消息内容(明文)

图 8 客户端接收雅虎通用户发送的文本消息数据包的协议格式

通过以上关键数据包的协议格式分析,发现各个数据包的格式都是唯一的。所以,可以通过捕获不同的包获取所需信息。例如,能够根据 1)获取登录的 MSN 帐号及其对应的 IP;可以根据 2)获取对方的 MSN 帐号;利用 3)—7)的协议格式分析能够提取出消息中的文本信息。对于飞信和雅虎通的协议格式分析方法也是类似。

4 IMSMMS 的设计与实现

4.1 IMSMMS 网络拓扑

IM 即时通信统一监控管理软件可以工作在以太网络中的任何位置。通过集线器可以与客户端在同一网段中,捕获所有通往客户端的数据包。也可以与网关在同一网段中,捕获所有通过网关的数据包。如图 9 所示,当监听位置位于 1 时,可以监听到所有达到用户机 3 的数据包。当监听位置位于 2 时,可以监听到所有从内网到达网关的数据包和网关转发给内网的数据包。

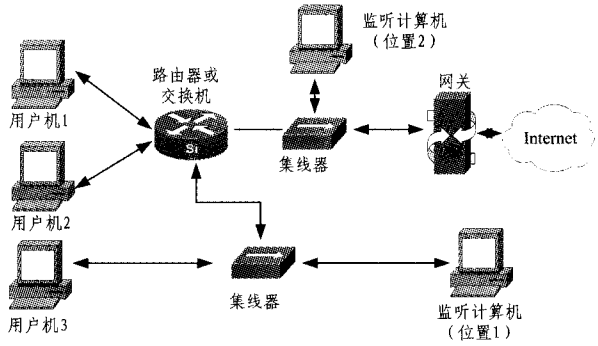

图 9 设计方案的网络拓扑图

4.2 系统设计与实现

总体设计框架如图 10 所示。从图 10 可以看出,系统分析的 4 个层面依次为数据采集层、协议分析层、业务感知层和业务应用层。

4.2.1 数据采集层

数据采集层面采用旁路监听的方式采集经过监听节点的数据,网络监听系统首先依靠于一套捕捉网络的数据包函数库。由于所设计的系统在 Windows 平台下不仅需要捕获数据包,还需要完成数据包过滤、流量统计、数据包存储,因此选取 WinPcap 作为网络模块的开发包[12]。

在实现过程中,着重考虑如何针对即时通信协议特点设置相对应的过滤规则,进行初级过滤。以 MSN 通信协议为例,根据之前分析的 MSN 协议的特征,其通信是基于客户端/服务器模式,服务器端口是 1863,并且消息传输是采用 TCP 协议的。因此,在捕获数据包时挑选出符合要求的数据包,即符合传输层使用 TCP 协议、源端口或者目的端口为 1863 的 IP 数据包,将符合规则的数据包转发给协议分析层。

4.2.2 协议分析层

协议分析层先是对上层传输来的数据帧进行处理,主要完成数据链路层的数据验证、拆包。并将数据交给 IP 层进行处理。IP 层数据包的处理主要完成 IP 层的数据验证、拆包,并将数据提交给传输层进行处理。由于 MSN 文本信息发送具有大小限制,每次发送只能在一个包中传输,因此不需进行 TCP 会话重建,只需要完成 TCP 层数据的验证、拆分以及 TCP 重复处理。将处理之后的信息交与上一层(业务感知层),以满足其对业务的识别和感知。

4.2.3 业务感知层

业务感知层是 IMSMMS 的核心层,主要作用是特征匹配和会话关联。特征匹配具体是将协议分析层提交的数据部分与特征字符串进行匹配,识别是否是 IM 文本消息业务。通过匹配的数据进行会话关联,否则将其丢弃。特征字符串是由协议格式分析总结得出的。匹配过程需要搜索符合特定模式(特征字符串)的字符串,使用正则表达式可以高效快速地完成对特征字符串的匹配。由于即时通信软件升级时特征字符串改动较小,只需通过修改正则表达式就能够监控到新版本的即时通信软件。

以 MSN 会话关联为例,MSN 会话关联是指根据 MSN 协议的特点,采用 MSN 帐号与 IP 地址的绑定技术,达到识别 MSN 数据包发送方与接收方的目的,其分为 3 部分:本地 MSN 帐号绑定、远程 MSN 帐号绑定和会话关联识别。

最后将通过特征匹配和会话关联的数据，提取出业务应用层所需的信息(包括发送方、接收方和文本消息内容等)。

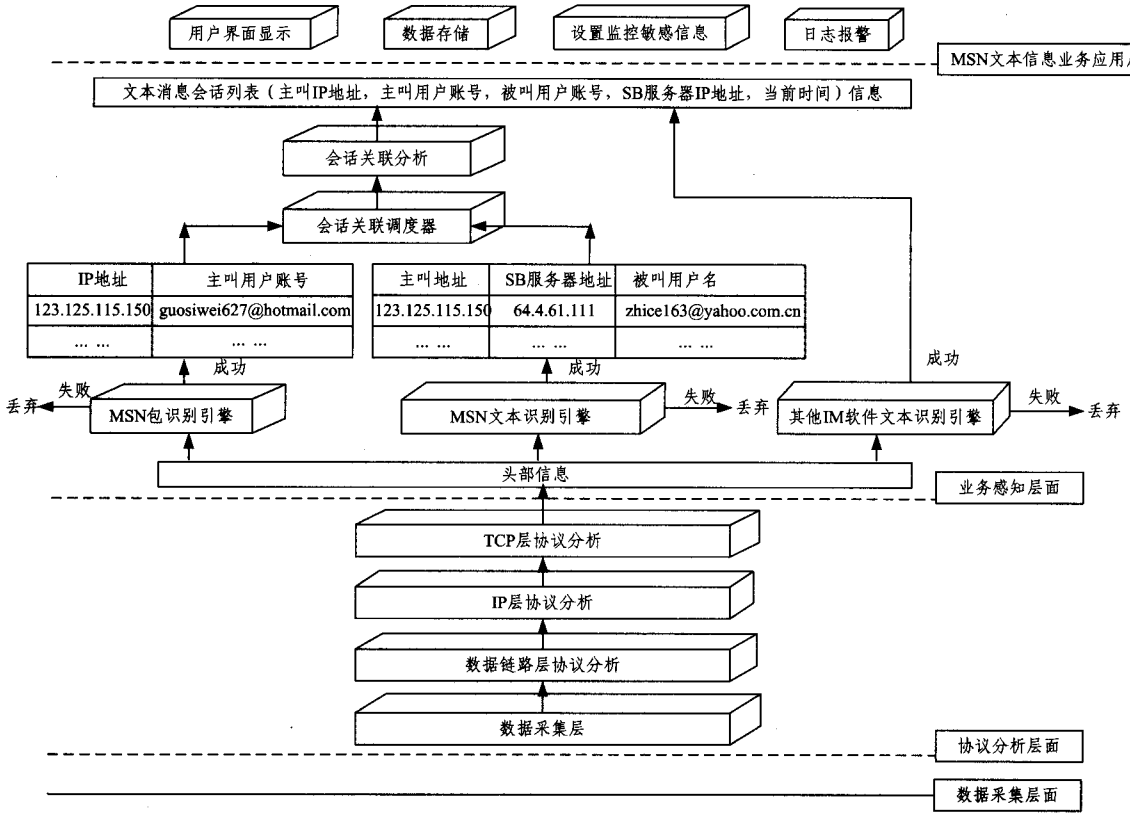

图 10　系统总体框架

4.2.3.1　本地 MSN 帐号绑定

本地 MSN 帐号绑定处理过程如图 11 所示，其具体处理流程如下。

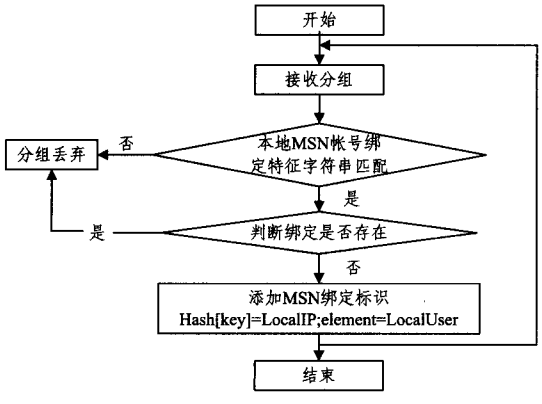

图 11　本地 MSN 帐号绑定

1)初始化本地用户哈希表。由于哈希表最优可实现时间复杂度为 O(1)的查找操作，有极高的查找效率，本系统选择将 MSN 绑定标识存储于本地用户哈希表中(LocalUserList)，MSN 绑定标识由 MSN 帐号和本地 IP 地址两元组来表示，通过绑定标识可在通信任何阶段识别出本地 MSN 账号。初始化阶段需建立 LocalUserList 并将 LocalUserList 中所有的元素初始化为 0。

2)接收分组。

3)根据本地 MSN 帐号绑定特征字符串进行匹配，根据 MSN 协议特点，在一台电脑上只能登录一个 MSN 帐号，这就意味着 MSN 帐号与 IP 地址有唯一映射关系，这里的本地 MSN 帐号绑定特征串匹配是指 MSN 成功登录消息的特征，以正则表达式{USR \d+ OK (. *) 1. \r\n}对数据包进行匹配，如果匹配成功则可以成功提取登录帐号，并与登录电脑的 IP 地址进行绑定，记录于哈希表 LocalUserList 中，然后转 4)，如匹配失败，丢弃分组，转 2)。

4)判断该绑定是否已存在哈希表中，如果是，则丢弃分组，转 2)；如果不是，转 5)。

5)保存 MSN 绑定标识。Key 为 MSN 登录用户的 IP，element为该登录 IP 地址所对应的 MSN 帐号。

6)转 2)。

4.2.3.2　远程 MSN 帐号绑定

远程 MSN 帐号绑定处理过程如图 12 所示，由图可以看出远程 MSN 帐号绑定与本地 MSN 帐号绑定十分相似，其区别在于特征字符串匹配与绑定标识。

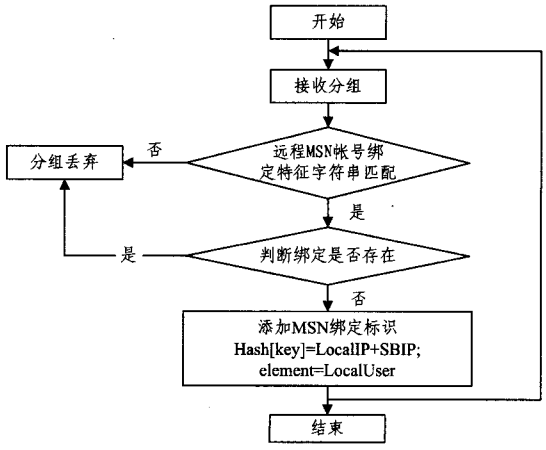

图 12　远程 MSN 帐号绑定

通过实际抓包分析，远程 MSN 帐号绑定特征字符串匹配过程为：若为本地 MSN 帐号主动连接远程 MSN 帐号，匹

配正则表达式为{JOI(. *);\{. *\}. * \d+:\d+\r\n},若为远程 MSN 帐号主动连接本地 MSN 帐号,匹配正则表达式为{MSG(. *). *\d +\r\n MIME-Version:. *\r\n Content-Type:. *\r\nX -MMS-IM-Format:. *\r\n\r\n([\s\S] *)},如果匹配成功则可以成功提取本地 MSN 帐号登录电脑 IP、SB 服务器 IP 和远程帐号 3 个信息一并进行绑定,记录于哈希表 RemoateUserList 中,在以后的消息识别中可以用该绑定信息实现对远程帐号的识别。

绑定标识区别在于本地绑定为本地 MSN 帐号登录电脑 IP 与本地 MSN 帐号绑定,而远程 MSN 帐号绑定为本地 MSN 帐号登录电脑 IP、SB 服务器 IP 和远程 MSN 帐号 3 个信息的绑定。这里 KEY 为本地 MSN 帐号登录电脑 IP 和 SB 服务器 IP,element 为远程 MSN 帐号。

4.2.3.3 会话关联识别

虽然在每条 MSN 消息中无法获取到发送方与接收方,但通过会话关联识别可有效解决这个问题。

MSN 文本消息业务会话关联识别处理流程如图 13 所示。

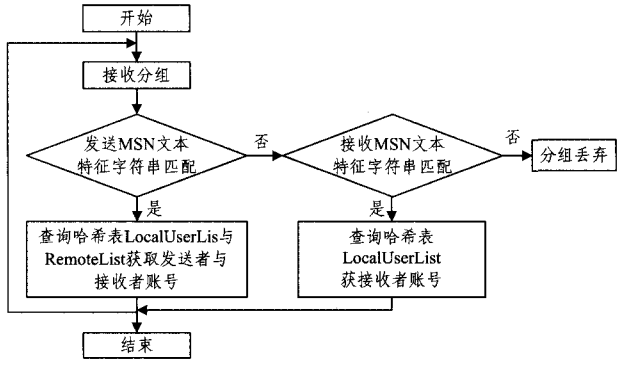

图 13 MSN 文本消息业务会话关联识别

1)接收分组。

2)根据 MSN 文本特征字符串进行匹配,通过实际的抓包分析,MSN 发送与接受消息都具有明显的特征,因此,对获取发送的 MSN 消息构造了正则表达式{MSG \d+ [AUN] \d+\\r\n MIME- Version:. *\r\nContent-Type:. *\r\n X-M -MS-IM-Format:. *\r\n\r\n([\s\S] *)}进行匹配,如果匹配成功,则可成功提取出文本信息,则转 3);若不成功,则转 4)。

3)这种情况下,数据包被本地用户发送给远程用户,抓取的文本数据包中目的地址为 SB 服务器 IP 地址,源地址为本地帐号登录电脑的 IP 地址,因此根据源 IP 地址对哈希表 LocalUserList 进行查询可得到本地登录 MSN 帐号(即消息发送者帐号),根据目的 IP 地址和源 IP 地址查询哈希表 RemoateUserList 可得到远程 MSN 帐号(即消息接收者帐号),转到 5)。

4)这种情况下,数据包被远程用户发送给本地用户,通过实际的抓包分析,这里可以直接得到远程用户的帐号和文本信息,因此构造了正则表达式{MSG (. *). * \d+\r\nMIME-Version:. *\r\n Content -Type:. *\r\nX-MMS-IM -Format:. *\r\n\r\n([\s\S] *)},如果匹配成功,其第一个匹配项为远程 MSN 帐号,第二个匹配值为文本消息,再通过目的地址查询哈希表 LocalUserList 可得到本地帐号。为了提高效率,这里不再查询哈希表 RemoateUserList,转到 5)。若匹配失败,则丢弃分组,转 1)。

5)转到 1)。

以上 3 部分即完成了 MSN 的会话关联识别。在对 MSN 文本的监控中,会话关联是一个非常重要的过程,该过程采用的关联元素(关联会话)为 RemoteUserList 表中的登录用户 IP 匹配 LocalUserList 中的用户 IP。采用消息驱动的机制可保证 MSN 文本会话的识别准确率,也就是说,仅仅在 MSN 的文本消息到来的时候其才会被激活。在没有 MSN 的文本消息传递和 MSN 消息传输合法信息期间,会话关联处于安静的等待阶段。其特点是效率高,实时性好,实现较为复杂。

4.2.4 业务应用层

业务应用层主要由用户界面显示、数据存储、设置监控敏感信息、日志报警组成。

用户界面显示将业务感知层获取的文本消息发送到前台用户,可以查阅所有捕获的 IM 消息,也能够查阅特定 IM 软件消息。

数据存储主要是将监控到的涉及敏感信息的消息送往数据库存储,若截获的消息中无需要监控的敏感信息,则存储所有截获的消息,用于日志处理。

设置监控敏感信息的主要功能是设置所需监控 IM 消息中的敏感消息。

日志报警主要功能是通过设置查阅规则,例如 IP、时间、发送方、接收方等,将涉及敏感信息的消息显示到日志报警页面,实现安全功能。

5 IMSMMS 性能测试与分析

5.1 测试环境

为了对 IMSMMS 做进一步的性能分析,对 IMSMMS 的数据采集模块、协议分析模块、业务感知模块和业务应用模块进行了功能测试。

实验拓扑图如图 14 所示,主机 A 与主机 B 之间采用发包工具 EtherPeek NX 3.0.0,通过设置不同的发包率进行漏检测试,网关计算机上装有 IMSMMS 进行监控。

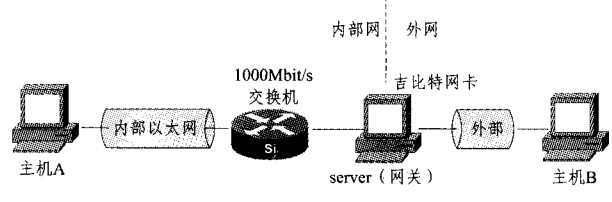

图 14 IMSMMS 性能测试拓扑图

5.2 测试结果分析

本系统开发环境为 Visual Studio 20008,加载了 WinPcap 库函数的 wpcap. lib 和 Packet. lib,最终结果可见界面显示,完成了对多种即时通信文本信息的解析和还原。

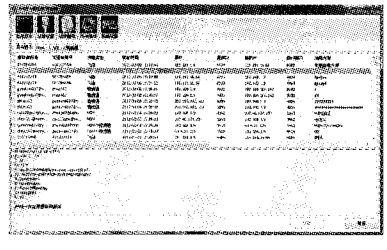

图 15 测试结果界面

实验表明,对于即时通信软件文本消息的解析和还原结

(下转第 82 页)

=35、B2=17、B3=24、B4=40；KEY3：S=341769205802、B1=35、B2=17、B3=24、B4=39，密码矩阵 T_1，T_2，…，T_n 由算法随机选取。图 4 中，图 4(a)为应用 KEY1 加密的结果；图 4(b)为应用 KEY2 加密的结果；图 4(c)为应用 KEY1 对图 4(a)进行解密的结果；图 4(d)为应用 KEY3 对图 4(a)进行解密的结果(密钥 KEY3 与 KEY1 的差，$SUB_{KEY3-KEY1}=1$)；图 4(e)为应用 KEY1 对图 4(b)进行解密的结果(密钥 KEY2 与 KEY1 的差，$SUB_{KEY2-KEY1}=1$)。

以上结果表明，密钥的细微差别导致完全不能解密，该算法具有高度敏感性。

结束语 本文提出了一种改进 Arnold 变换算法，其充分利用矩阵的特性，结合伪随机序列和安全哈希函数，并进行了实验验证和性能分析。该算法在抗穷举攻击和时间复杂度方面比经典 Arnold Cat 变换有明显的优势，而且密钥量大、置乱效果好，具有可行性和有效性。

参 考 文 献

[1] Kumar B K S, Patil C R. JPEG image encryption using fuzzy PN sequences[J]. SIViP, 2009, 0131(6): 2-3

[2] 陈铭. 基于 Arnold 变换的图像信息伪装算法[J]. 计算机应用研究, 2006(1): 235-237

[3] 樊昌信, 曹丽娜. 通信原理(第六版)[M]. 北京：国防工业出版社, 2006: 326-338

[4] 黄谆, 白国强, 陈弘毅. 快速实现 SHA-1 算法的硬件结构[J]. 清华大学学报, 2005, 45(1): 123-129

[5] 任洪娥, 尚振伟, 张健. 一种基于 Arnold 变换的数字图像加密算法[J]. 光学技术, 2009, 35(3): 384-390

[6] 马进, 卢雷, 朱宁. 基于划分思想的 Arnold 变换算法[A]// 中国电子学会第十六届信息论学术年会论文集[C]. 北京：电子工业出版社, 2009: 90-94

[7] 陈亦欢, 严伟超. 应用 SIMD 并行技术的 SHA-1 加密算法的批量实现[J]. 重庆理工大学学报：自然科学版, 2012, 26(7): 74-80

(上接第 46 页)

果见图 15，其中，使用网卡选择，选择具体的网卡类型进行数据包捕获；可以通过混杂模式查看所有的消息捕获结果，包括消息的接收者帐号、发送者帐号、消息类型、获取时间、源 IP、源端口、目的 IP、目的端口、消息内容；也可以选择特定的即时通信软件进行查阅截获结果。通过记录查询可以查询到特定时间段截获的结果，并且能够另存为文本文件。同时通过设置过滤规则，可以设置需要匹配的有害信息关键字，以实时拦截有害信息。利用发包工具进行漏检测试，结果如表 2 所列。

表 2 漏检测试结果

每秒发包个数	共发送包数	共检测包数	漏检率
100	1456	1456	0.00%
300	4823	4823	0.00%
1000	11729	11704	0.21%
1500	20304	19396	4.47%

系统漏检率指在网络正常情况下，系统应该检测出的数据占全部通过数据的百分比。随着发包率的增加，在一定时间内底层提交给文本识别引擎的数据量相应增加，当超出文本识别引擎的最大吞吐量时，部分数据会被文本识别引擎丢弃，造成漏检率的出现，系统应工作在可接收的漏检率下。通过表 2 发现，当每秒发包率小于 300 时，系统的漏检率为 0；当每秒发包率大于 300 小于 1000 时，出现微量漏检；当每秒发包率大于 1500 时，丢包率急剧增加。所以，本系统可以完全满足中小型企业网对于监控的需求。

结束语 本文针对目前对即时通信监控的需要，通过分析 MSN 和其他几个主流的通信机制，得出了 IM 软件的文本消息传输协议。同时，针对 MSN 文本消息传输协议的特殊性，提出了一种会话关联识别来对 MSN 文本信息进行有效完整的提取。并且分析了 MSN 和雅虎通这两个即时通信软件之间相互传输文本消息的格式。在提取消息时采用正则表达式，使得匹配提取过程更为精确，并且只需通过修改相应正则表达式即可实现对不同版本即时通信软件的监控。通过实验数据证明，IMSMMS 可对中小企业进行有效监控。下一步将研究对即时通信中文件传输、视频、音频等的统一监控进行分析实现。

参 考 文 献

[1] Lu Rui, Mi Jia, Huang Bo. Design and Implementation of Instant Messenger Security Monitoring System Based on Protocol Analysis[C]// 2010 Chinese Control and Decision Conference. 2010: 4290-4293

[2] Osullivan S. Instant messaging vs. instant compromise [J]. Network Security, 2006(7): 4-6

[3] Hindocha N. Threats to Instant Messaging[EB/OL]. Symantec Security- Response WHITE PAPER, http://securityresponse.symante-c. com/avcenter/reference/threats. to. instant. messaging. pdf

[4] 付安民, 张玉清. 即时通实时监控系统的设计与实现[J]. 通信学报, 2008, 29(10): 165-172

[5] Xu Guo-tian. Design and realization of the MSN monitoring system[C]// Information Theory and Information Security (ICITIS). 2010: 178-181

[6] 胡振宇, 刘在强, 苏璞睿, 等. 基于协议分析的 IM 阻断策略及算法分析[J]. 电子学报, 2005, 33(10): 1830-1834

[7] 刘彬, 赵彩荣, 丛建刚. 即时通信协议分析与监控技术研究[J]. 计算机应用研究, 2007, 24(9): 260-265

[8] 付安民, 张玉清. 飞信即时通监控系统的设计与实现[J]. 计算机工程, 2008, 34(13): 229-231

[9] 章秩. 基于以太网的即时通信监控系统研究与实现[D]. 上海：复旦大学研究生院, 2010

[10] MSN Messenger Service 1. 0 Protocol[EB/OL]. http://www.hypothetic. org/docs/msn/sitev1/index. php, 2003. 9

[11] 郑有才, 郑光华, 张玉清. 即时通信息监听技术的研究与实现[J]. 计算机应用研究, 2005, 22(8): 113-116

[12] 严华, 蔡瑞英. 即时通信系统的设计与实现[J]. 计算机技术与发展, 2009, 7(19): 242-248