文章编号: 1673 5439(2006) 01-0001-07

基于综合统计特征的 Skype流量分析与识别

王振华, 王 攀, 张顺颐
(南京邮电大学 网络技术研究中心, 江苏 南京 210003)

摘 要: 首先对 Skype通信机制进行了深入研究, 并在此基础上提出了 基于流统计特征和静荷统计特征的 Skype流量识别策略, 然后综合分析上面两种方法的优缺点, 提出了基于综合统计特征的 Skype流量识别方案, 并设计出基于综合统计特征的 Skype流量识别系统模型, 最后设计并实现了相应的实验系统对相关结论进行验证。

关键词: Skype; 流量识别; 流统计特征; 静荷统计特征
中图分类号: TP393. 06 文献标识码: A

An Analysis and Identification of Skype Network Traffic Based on Integrated Statistical Characteristics

WANG Zhen-hua, WANG Pan, ZHANG Shun-yi
(Research Center of Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In this paper, we first provide a study on Skype communication mechanisms and propose two policies to identify Skype traffic based on statistical flow characteristics and statistical payload characteristics. Then, after synthetically comparing the advantages and disadvantages of the two policies, we describe a solution and design a system structure to identify Skype traffic based on integrated statistical characteristics. Finally, we design and implement an experimental system to test the system and give some conclusions.

Key words: Skype; Traffic identification; Statistical flow characteristics; Statistical payload characteristics

1 引 言

计算机对等网络 P2P, 作为目前改变现有 Internet应用模式的主要技术之一, 成为新一代互连网技术研究的热点问题。基于第三代 P2P技术, Skype[1]正以优质的语音质量和低廉的通话费用吸引着越来越多的用户。它的出现给传统的 VoIP业务带来了巨大的冲击。为此, 研究 Skype的通信机制, 并能准确识别并控制 Skype流量有着很大的意义, 一方面可以借鉴其中的关键技术用于改进目前网络电话技术, 另外可以对 Skype业务进行有效控制。

目前国内外对 P2P网络流量的识别与管理有了一定的研究[2-4], 提出了比如基于传输层的 IP Pair[2]以及第七层静荷特征[3]等识别方法, 这些方法可用于目前大多数比较流行的 P2P软件的流量识别, 如 BT、Emule、eDonkey等。但是对于 Skype, 由于其协议不公开, 并且使用了比较安全的加密算法, 因此, 目前尚没有对 Skype协议的详细研究, 也没有识别 Skype流量的有效方案。

本文在文献[5]的基础上, 进一步对 Skype的通信机制进行了详细研究, 并分别从流统计和静荷特征统计的角度出发, 研究了 Skype的流统计特征和静荷统计特征, 提出了基于综合统计特征的 Skype应用识别方案及其系统模型, 并进行了实验验证。

收稿日期: 2005-10-31
基金项目: 国家高技术研究发展计划(863计划) (2003AA121560和 2005AA121620)资助项目

2 Skype的机制研究

下文将从 Skype的通信实体、连接机制等方面来描述 Skype的通信机制。

2.1 Skype的通信实体

Skype网络中有 4种主要节点:用户节点(User Node, UN)、超级节点(SuperNode, SN)、登录服务器(Login Server, LS)和事件服务器(EventServer, ES)。

(1) UN:用户节点是普通用户客户端,实现一般的语音呼叫、即时消息和文件传送等功能,用户节点对安装客户端的主机性能没有严格要求;

(2) SN:超级节点是 Skype网络中具有特殊功能的节点,是普通 UN的连接终点,接受并受理普通 UN行为请求以及和其他超级节点进行信息交互,同时也具有普通用户客户端的功能;

(3) LS:登录服务器是 Skype网络中的集中服务器,用来存储用户信息,保证用户唯一性,以及用户登录认证;

(4) ES:事件服务器也是集中服务器,Skype用户在每次新用户登录和客户端程序退出时和 ES进行数据传输,可能用来存储某些日志信息。

整个 Skype网络结构模型如图 1所示,其中实线连接为网络持续连接,虚线只是在网络实体加入网络瞬间传输认证信息、版本更新信息等建立的瞬间连接。

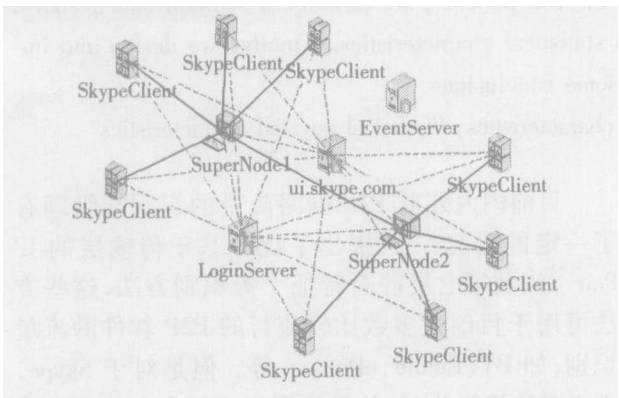


图 1 Skype网络模型

2.2 Skype的连接机制

(1) 端口特性:Skype和目前大多数 P2P软件一样,使用随机动态端口,其独有特征为:首先,客户端在第一次连接时随机产生一个本地端口,该端口用于对外进行 UDP连接。在连接过程中,始终侦听该端口的 TCP连接,并开放本地 80和 443作为备用 TCP端口;其次,登录服务器使用端口为 33033,事件服务器使用端口为 12350,另外,软件代码固化的 7个自举超级节点所使用端口和登录服务器一样,为 33033。

(2) 登录过程:刚刚安装的 Skype客户端首次运行时在本地 shared xml文件中生成 3个事件服务器(EventServer)、3个登录服务器(LoginServer)、200个超级节点的列表(hostcache)。安装后首次运行登录过程如下:(1) 开放侦听端口:生成一随机端口作为本地客户端端口,并尝试开放该端口连接,随机端口开放成功后,会继续开放本地 80和 443作为 TCP备用侦听端口,至此开放本地端口完成;(2) 连接信息 UDP协商:开放侦听端口之后,客户端利用本地开放 UDP端口,通过 UDP数据与超级节点列表中的某些主机进行连接信息协商,如果本地 UDP端口不能实现与超级节点进行连接信息协商,则利用备用 80和 443端口通过 TCP进行连接协商,并最终与某个超级节点建立 TCP连接;(3) 认证、版本更新与事件信息交互:成功连入 Skype网络后,需要与登录服务器进行连接并进行用户认证,与 ui.skype.com连接获取版本更新信息,与事件服务器连接进行日志及事件信息交互,与上面 3类服务器的连接是瞬态连接,完成信息交互之后即断开连接。安装后首次登录过程如图 2所示。

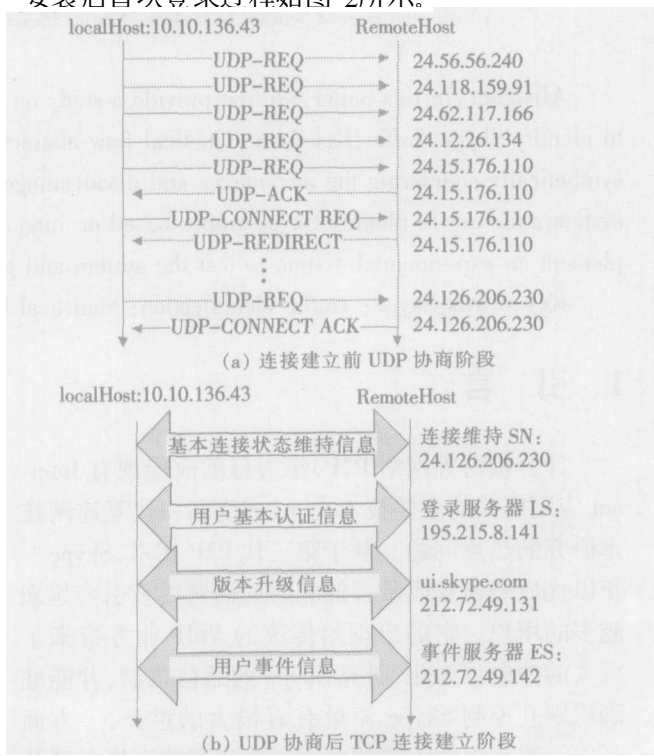


图 2 Skype首次登录协议分析

在图 2中,给出了首次登录过程连接建立的详细流程,其中连接建立前 UDP协商阶段的信令名称均为自定义,解释如下:(1) UDP-REQ:静荷长度为 18,第一个协商请求信息,每个超级节点对该请求有 3种回应:UDP-ACK、UDP-REDIRECT和 UDP-CON-

NECT ACK;(2) UDP-ACK:静荷长度为 11,是对第一个协商请求的回应信息,类似 SIP协议的临时响应,请求方收到该回应需要进一步进行协商请求;(3) UDP-CONNECT REQ:静荷长度为 23,是请求方在收到 UDP-ACK后的进一步协商请求;(4) UDP-REDIRECT:静荷长度为 58,如果被请求超级节点不能满足请求方需求,发送该消息进行超级节点重定向,使请求方重新与其他超级节点进行协商;(5) UDP-CONNECT ACK:静荷长度为 18,超级节点回应该消息表示可以接受并建立 TCP连接,请求方收到该消息后便可与该超级节点建立 TCP连接,如果连接建立成功,该请求客户端便成功接入 Skype网络。

(3)文件传输过程:客户端 A准备向客户端 B发送文件时,对于 A和 B均为 NAT用户的情况,其传输过程为:(1) A发送 TCP连接请求,并同时利用 TCP和 UDP进行某些传输协商信息交互,最终确定传输效率最高的一条公网传输路由;(2) A端利用 UDP向中转节点传送文件数据;(3)中转节点利用 UDP将文件数据转发给用户 B;(4) B收到经中转节点转发的数据之后,利用 UDP向 A发送接收确认信息。即一旦文件传输开始,A向中转节点单向发送文件数据,而中转节点不向 A发送确认信息;B在正确接收数据后向 A发送确认信息,而不直接接收 A的任何信息。其简单传输流程如图 3所示。

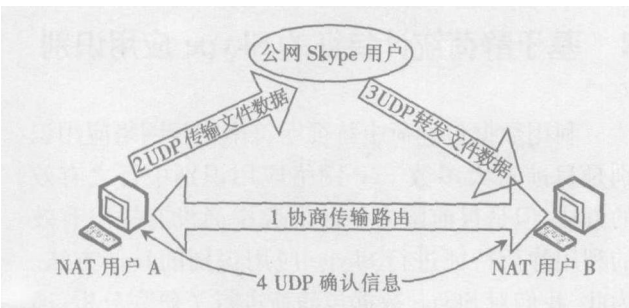


图 3　文件传输流程图

3　基于流统计特征的 Skype应用识别

流信息主要是指数据包的五元组信息,即源地址、源端口、目的地址、目的端口和协议类型。通过对一定时间维度内这些参数的综合统计分析,结合需要研究应用的流统计特征,可对多种业务进行有效识别。为此,首先分析 Skype的流统计特性,进而确定用于有效识别 Skype应用的流统计策略。

3.1　Skype流统计特征

虽然 Skype协议为非公开协议,并且使用随机动态端口机制,但是通过对数据流进行统计分析,发现 Skype流量在端口、使用协议、远端地址和连接方面仍然明显的存在某些自己的特征。

(1)端口特性:Skype客户端连接成功后,主要有以下几类端口:(1)本地侦听端口主要是 80、443和本地 Skype端口,该 Skype端口在安装之后只要用户不去修改并且能够开放侦听成功就不会改变;(2)本地 UDP探测与数据传输端口,该端口使用本地 Skype端口,进行超级用户探测和数据传输;(3)本地连接端口,这些端口是本地按顺序开放的 TCP端口,一般 1 000－4 000左右,所连接的远端端口为超级节点或普通用户的 Skype端口,和超级节点的 TCP连接用于保证客户端是否连接在 Skype网络中,并且进行一些控制信息的传输。

(2)协议特性:同一对 IP地址对同时使用 TCP和 UDP是 P2P网络的普遍特性,Skype也不例外,但是也有自己的特点:(1)单协议端口多连接特性:进行超级节点探测以及搜索利用本地 Skype端口的 UDP协议,或者利用 80和 443端口的 TCP协议,这些数据传输一般使用单一协议,本地使用单一端口,但快速连接多个远端 IP;(2)先 UDP后 TCP特性:在 Skype客户端连入 Skype网络过程中,首先利用 UDP对超级节点列表地址进行连接信息协商,之后才和某个超级节点建立 TCP连接以维持客户端和 Skype网络的连接;(3)双协议特性:数据传输过程中,使用本地某个 TCP端口通过 TCP协议实现连接协商信息的传输,使用本地 Skype端口通过 UDP协议实现数据的传输以及确认信息的接收。

(3)连接特性:(1)快速探测性:本地 Skype端口快速连接多个远端地址,因为 Skype客户端不断更新超级节点列表,并向多个超级节点提供自己的状态信息,因此不断利用本地 Skype端口或 http端口对超级节点进行探测;(2)不同协议同端口性:同一 IP地址对源端或目的端两者中总有一个对应的 TCP和 UDP使用同一端口,这几乎可以当作是 Skype所独有的特性;(3)独立端口特性:虽然有多个远端 IP与本地 Skype客户端进行 UDP或 TCP数据传输,但是每个 IP都有自己的端口,没有重复,这主要是 Skype是随机分配端口的缘故。

(4)特殊地址:在 Skype网络中,有几个特殊地址:(1)软件更新地址 ui skype com:客户端每次运行时,都会根据用户的需要,登录这个地址进行新版本更新检查;(2)登录服务器地址:客户端在用户注册或新用户登录时都要在连接成功后到登录服务器

进行用户信息认证,登录服务器共有 3个,这些地址使用固定的 33033端口;(3)事件服务器地址:登录服务器也有 3个固定的地址,使用固定的 12350端口;(4)自举超级节点地址:这些地址如果不是在超级节点列表丢失或全部连接失败的情况下是不会连接的,共有 7个自举超级节点,使用固定的 33033端口。

3.2 基于流统计特征的 Skype应用识别策略

通过上述 Skype流统计特征的分析与研究,可在一定时间维度内有针对性地进行流信息统计,制定出专门针对 Skype的有效的流统计应用识别策略。在此,提出了二维条件性综合流统计识别策略。

此处的二维条件性综合流统计识别策略主要是从流信息五元组中根据统计需求选出其中的二元组合按给定条件过滤并进行综合统计,如 X-Y(C)二元组合要求统计出 List Y Group By X Where Packets Satisfy C,再综合二元组合统计的结果进行分析识别,以达到有效实现应用识别的目的。此方法从普遍的连接特性统计出发,高于一般的应用层静荷特征字符串识别以及其他识别方法,具有普遍适用性,可实现多种网络应用的识别,尤其对于目前新的网络体系结构,如 P2P网络应用,可以从总体上进行识别。

针对 Skype应用识别,可用以下二元组合策略:

(1) LocalPort-RemoteIP(ALL):利用本地 Skype客户端端口在较短时间内与多台远程主机进行数据传输的特性,通过本地端口所连接主机数目,可识别出非常规网络体系业务的本地端口。此非常规业务主要是针对目前越来越多的 P2P业务提出,因为对大多数 P2P业务,客户端都会通过某个特定端口(一般是随机分配)不断连接多台远程主机,实现 P2P网络的对等特性,以便于资源搜索,因此本地 P2P端口一般会在比较短的时间内与多台远程主机进行数据传输,从这个角度可以识别出大部分 P2P流量。

(2) RemoteIP-Protocol(ALL):P2P网络应用或者某些媒体流业务等在进行抢占资源的数据传输时一般都是利用可靠的 TCP保持基本连接实现信令交互,利用高传输效率的 UDP实现数据的传输,如 Skype用户在进行通话或文件传输时,就是同时使用 TCP和 UDP两种协议。此外,还有几种熟知的网络应用,如 DNS服务等,也是同时使用 TCP和 UDP两种协议,但是这些网络应用都可以通过常规端口或其他特性很容易识别。因此,通过对同时使用 TCP和 UDP的源 -目的 IP对,并且去掉那些已知的正常业务,可以对这些非常规业务进行准确识别。

(3) RemotePort-LocalPort(UDP):通过对和远程某个端口利用 UDP连接的本地端口以及本地端口数的统计,可在上一步的基础上进一步从远端主机端口的角度进行应用识别。对于某个远端主机端口,如果连接的本地端口中包含上一步中统计所得的非常规端口,则可判断出该远端端口上的应用与本地非常规端口对应的业务相同。例如,如果识别出本地 2558端口为 Skype本地端口,某个远端主机的 3380端口通过 UDP与本地主机进行数据传输,则也可断定该远端端口上所有数据传输均为 Skype业务。

(4) RemotePort-Protocol(同一 IP对):通过对 Skype流量统计特性的研究发现,虽然 Skype客户端与某些超级节点间以及某些普通客户端之间进行数据传输时同时使用 TCP和 UDP,但是通常远端主机的 TCP和 UDP使用同一端口,即 Skype随机开放的端口。因此,结合 Local-RemoteIP(ALL)策略,对于与本地非正常业务端口连接的主机,如果远端端口同时使用 TCP和 UDP,则可以初步判定为 Skype业务。当然这里只是初步判定,如果发现有其他业务同样具有该统计特征,需要对该策略进行修正。

4 基于静荷统计特征的 Skype应用识别

利用数据包静荷中特征字符串进行网络应用识别是目前对大多数 P2P网络应用识别中行之有效的方法,但是目前国内外的研究中尚没有提出有效的利用静荷特征进行 Skype应用识别的具体方法。为此,我们对 Skype数据包静荷进行了研究分析,通过分析,达到准确识别 Skype应用的目的。

对 Skype数据包统计分析发现,TCP数据包没有明显的静荷特征,但是 UDP数据包中,自始至终都存在一定的特征,这对于进行 Skype应用识别及其协议研究非常有效。

4.1 登录过程 TCP连接建立前的静荷特征分析

Skype进行网络连接时,首先利用 UDP与超级节点进行连接信息协商,满足连接条件后,才能与超级节点建立 TCP连接,通过统计分析发现,在建立 TCP连接之前的不同的 UDP协商数据包具有如表 1所示的不同静荷特征:

(1) UDP-REQ: 由客户端发出的第 1个 UDP数据包请求中的第 3字节为 02, 静荷长度为 18;

(2) UDP-ACK: 确认包中前两个字节和 UDP-REQ相同, 第 4到 7的 4个字节用来表示本地客户端的公网 IP, 静荷长度为 11;

(3) UDP-CONNECT REQ: 进一步连接请求数据包中前两字节和 UDP-REQ相同, 第 3字节后 4位为 3, 第 4字节为 01, 第 9到 12的 4个字节表示远端超级节点的 IP, 静荷长度为 23;

(4) UDP-CONNECT ACK: 连接确认信息数据包中, 静荷前两个字节不同于 UDP-REQ, 并且第 3字节为 02, 静荷长度为 18;

(5) UDP-REDIRECT: 重定向信息数据包中, 静荷前两个字节也不同于 UDP-REQ, 并且第 3字节为 02, 静荷长度为 58.

表 1　登录过程 TCP连接建立前静荷特征统计表

连接协商 UDP包	静荷长	字节 1, 2	字节 3	字节 4
REQ	18	*m*, *n*	02	xx
ACK	11	*m*, *n*	xx	xx
CONNECT REQ	23	*m*, *n*	x3	01
CONNECT ACK	18	不同于 *m*, *n*	02	xx
REDIRECT	58	不同于 *m*, *n*	02	xx

(注: x表示不确定)

4.2　Skype登录成功后其他操作的 UDP静荷特征分析

Skype登录成功后的其他操作过程中, UDP数据包仍然具有一定的静荷特征:

(1) UDP协商: 几乎每次客户端与超级节点进行数据传输前, 都要通过 UDP进行信息协商, 其中第 1个 UDP数据包的第 3字节均为 02;

(2) 文件传输: 文件传输过程中, 有很大比例的 UDP数据包的第 3字节的低 4位为 d;

(3) 搜索用户: 搜索用户过程中, 超级节点返回的查询结果的 UDP数据包的第 3字节低 4位为 f 且第 7字节为 02.

4.3　基于静荷统计特征的 Skype应用识别策略

由于对所有 Skype的 UDP数据包中不含有固定字节的特征字符串, 因此不能根据某一个固定字节的特征来判断是否 Skype应用, 可以根据以上对 Skype数据包中的静荷的统计分析, 不断收集 Skype数据包中的静荷特征, 生成 Skype静荷特征集合, 制定基于静荷统计特征的 Skype应用识别策略, 对每一个数据包, 只要其中包含静荷特征集合的特征字符串, 便可判定为 Skype应用。

这种方法理论上简单, 但是如果仅仅根据大多数 UDP数据包所具有的第 3字节 02来判断, 会将其他应用错误地判断为 Skype应用, 如 Emule某些 UDP数据包中第 3字节同样为 02, 而要实现多种情况特征字符串的匹配, 会严重影响识别效率, 并且由于 Skype既有 TCP, 又有 UDP, 因此在判断准确性上仍然存在一定的问题。

5　基于综合统计特征的 Skype应用识别系统模型

前面两节中, 分别提出了基于流统计特性和基于静荷统计特性的 Skype应用识别, 这两种方法都有各自的优点和不足之处, 基于流统计特性的识别方法具有通用性, 适用于多种应用识别, 但缺乏具体的识别依据; 基于静荷统计特征的识别方法理论上实现简单, 并且有明确的识别标准, 但是对于 Skype应用, 很难保证识别的准确性和识别效率。为此, 综合上面两种方法, 结合以上对 Skype的研究分析, 首先提出了基于综合统计特征的 Skype应用识别策略, 进而给出相应的 Skype应用识别系统模型。

5.1　基于综合统计特征的 Skype应用识别策略

为了能够准确有效地进行 Skype应用识别, 我们结合了上面讨论的流统计和静荷统计的方法, 制定了如图 4所示的识别策略。

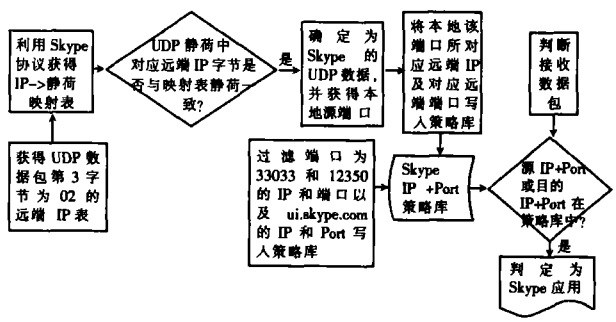


图 4　基于综合统计特征的 Skype应用识别策略

识别策略具体描述如下: (1) 根据 UDP静荷中远端 IP字节所对应的 IP与该数据包的远端 IP一致, 确定该 UDP数据包为 Skype数据, 从而获得本地 Skype端口; (2) 将所有与本地 Skype端口对应的 UDP数据包的远端 IP和端口写入 Skype IP+Port策略库; (3) 由于 Skype除了与超级节点或普通客户端通信外, 还与某些特殊节点通信, 如 ES、LS和 ui skype com, 因此, 也要将这 3类地址和端口写入

Skype IP+Port策略库;(4)判断每一个数据包,无论是TCP数据包还是UDP数据包,只要源地址+源端口或者目的地址+目的端口中有一个在Skype IP+Port策略库中,则可以判定该数据包为Skype应用。

上述识别策略中,首先用Skype协议特有的静荷指定字节表示远端IP的特征来识别本地Skype端口,进而根据本地地址和端口识别远端IP和端口,此外根据Skype数据中TCP和UDP对于通信双方总有一方使用同一Skype端口这一特点,数据包中只要源IP+Port和目的IP+Port两者中有一个在Skype IP+Port中,就可肯定地判定为Skype应用。

在这一策略中,由于静荷是利用与远端IP对应的连续4个字节精确匹配来判断,因此识别准确率不容质疑,并且对于正常网络,只要允许UDP数据传输,一般也可以保证识别的完整性。但是对于某些特殊网络,如果不允许进行UDP数据传输,Skype仍然能够利用443和80端口最终连接成功,这种情况便不能准确识别,除非之前该网络允许UDP传输,并在Skype IP+Port策略库中有一定的策略积累。

5.2 Skype应用识别系统模型

基于综合统计特征的Skype应用识别一方面要对采集到的数据进行统计,生成识别策略;另一方面要利用生成的识别策略对采集到的数据进行网络流量识别,从而对各种不同的应用进行不同的统计分析与处理。两者是并行处理的,即网络流量识别以动态识别策略为识别依据,从而保证识别的准确性和可扩展性。系统模型如图5所示。

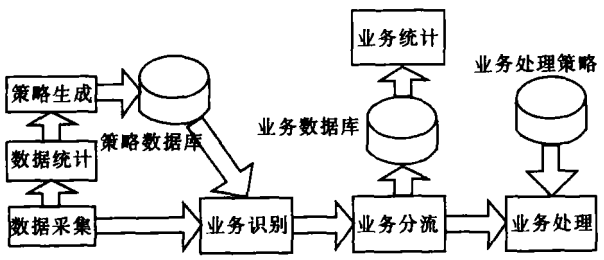


图5 基于综合统计特征的Skype应用识别系统模型

6 实验验证

为了验证本文提出的结论,设计一个验证系统,利用RAW_SOCKET截获流经本机的所有IP数据包,将所有IP数据包报头基本信息及部分静荷信息写入数据库,按本文提出的基于综合统计特征的Skype应用识别策略,生成Skype IP+Port策略库,并最终实现Skype流量识别,从而验证了本文的结论。

(1)测试环境:本地测试环境,Windows2000操作系统,同时使用验证系统Skype流量分析与识别系统和CommView 5.0(Build 455)[6]进行流量分析对比,运行Skype、Emule、BT等多种P2P软件,并且保证网络中有一定的正常网络应用,如http、FTP等,从而保证对结论的验证不失一般性。

(2)流统计特征验证:LocalPort-RemoteIP(ALL)的统计结果如图7所示。从图6中可以看出,本地端口中9640、1334、4672、1340几个端口都在很短的时间内与多个远程主机有数据传输,因此,这几个端口均为非常规业务,但是目前仍然不能确定哪个是Skype本地端口。

本地端口号	协议类型	远程主机数	数据包数
9640	UDP	884	1511
1334	UDP	85	226
4672	UDP	60	117
1340	UDP	28	64
-	ICMP	9	60
1347	UDP	3	3
1348	TCP	1	3
1349	TCP	1	9
1350	TCP	1	3
1351	TCP	1	3
1353	TCP	1	11
1354	TCP	1	3

图6 LocalPort-RemoteIP(ALL)统计结果

为了分析出本地Skype端口,进一步对RemoteIP-Protocol(ALL)进行统计分析,如图7所示。虽然有多个远程主机同时使用了TCP和UDP两种协议,但是只有195.62.22.211使用同一端口4174进行TCP和UDP数据传输,因此可以判定,该主机为本地客户端连入Skype网络的超级节点,与之对应的本地UDP端口1334为本地Skype端口。与图6结论进行对比可以发现,1334确实为非正常业务端口,从而进一步验证了结论的正确性。

本地主机地址	远程主机地址	协议类型数	数据包数
10.10.136.43	147.32.49.69	2	5
10.10.136.43	195.62.22.211	2	55
10.10.136.43	140.123.108.139	2	5
10.10.136.43	194.146.227.8	2	5
10.10.136.43	207.150.166.150	2	5
10.10.136.43	207.44.140.73	2	8
10.10.136.43	211.233.41.235	2	5
10.10.136.43	203.218.81.13	2	4
10.10.136.43	193.225.80.117	2	5
10.10.136.43	195.245.244.243	2	13
10.10.136.43	213.158.119.104	2	5
10.10.136.43	217.227.114.102	2	5

协议类型	本地主机地址	本地端口	方向	远程主机地址	远程端口
UDP	10.10.136.43	1334	->	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
UDP	10.10.136.43	1334	<-	195.62.22.211	4174
TCP	10.10.136.43	1336	<-	195.62.22.211	4174
TCP	10.10.136.43	1336	<-	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
TCP	10.10.136.43	1336	->	195.62.22.211	4174
TCP	10.10.136.43	1336	<-	195.62.22.211	4174
TCP	10.10.136.43	1336	<-	195.62.22.211	4174

图7 RemoteIP-Protocol(ALL)统计结果

(3)综合统计特征识别策略的验证:在该实验系模型系统中,验证了流统计特征的结论,重点根据得出的基于综合流统计特征的策略,即结合静荷特征字符串和流统计特征的综合策略,对 Skype流量进行了识别。基于综合统计特征策略的 Skype流量识别结果如图 8所示。从图 8可以看出 Skype运行期间所有数据传输的详细信息,包括利用 http协议,从 ui skype com获取版本更新信息、连接协商、连接建立、与 LS、ES交互对应信息等。另外可以从静荷前 10位对静荷特征进行观察研究,进一步验证静荷特征的结论。

数据时间	协议类型	本地主机地址	本地端口	方向	远程主机地址	远程端口	包长	静荷长	BYTE1	BYTE2	BYTE3	BYTE4	BYTE5	BYTE6
2005-8-16 9:02:34	TCP	10.10.136.43	1332	<-	212.72.49.131	80	52	0						
2005-8-16 9:02:34	TCP	10.10.136.43	1332	->	212.72.49.131	80	40	0						
2005-8-16 9:02:34	TCP	10.10.136.43	1332	->	212.72.49.131	80	152	112	47	45	54	20	2f	75
2005-8-16 9:02:34	TCP	10.10.136.43	1332	<-	212.72.49.131	80	40	0						
2005-8-16 9:02:34	TCP	10.10.136.43	1332	<-	212.72.49.131	80	491	451	48	54	54	50	2f	31
2005-8-16 9:02:34	TCP	10.10.136.43	1332	<-	212.72.49.131	80	40	0						
2005-8-16 9:02:34	TCP	10.10.136.43	1332	->	212.72.49.131	80	40	0						
2005-8-16 9:02:35	TCP	10.10.136.43	1332	->	212.72.49.131	80	40	0						
2005-8-16 9:02:37	TCP	10.10.136.43	1332	->	212.72.49.131	80	40	0						
2005-8-16 9:02:38	TCP	10.10.136.43	1332	<-	212.72.49.131	80	40	0						
2005-8-16 9:02:43	UDP	10.10.136.43	1334	->	24.15.176.110	59089	46	18	c6	3b	02	4f	2d	af
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.56.56.240	35932	46	18	c6	3d	02	68	b3	6a
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.90.217.72	61789	46	18	c6	3f	02	5e	5b	13
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.118.159.91	37289	46	18	c6	41	02	42	d7	df
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.60.59.109	32638	46	18	c6	43	02	5c	d8	67
2005-8-16 9:02:44	UDP	10.10.136.43	1334	<-	24.56.56.240	35932	39	11	c6	3d	07	da	02	d8
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.56.56.240	35932	51	23	c6	3d	53	01	ee	88
2005-8-16 9:02:44	UDP	10.10.136.43	1334	<-	24.118.159.91	37289	39	11	c6	41	57	da	02	d8
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.118.159.91	37289	51	23	c6	41	73	01	18	f5
2005-8-16 9:02:44	UDP	10.10.136.43	1334	<-	24.60.59.109	32638	39	11	c6	43	07	da	02	d8
2005-8-16 9:02:44	UDP	10.10.136.43	1334	<-	24.90.217.72	61789	39	11	c6	3f	47	da	02	d8
2005-8-16 9:02:44	UDP	10.10.136.43	1334	->	24.60.59.109	32638	51	23	c6	43	43	01	41	f9

图 **8**　基于综合统计特征策略的 **Skype**流量识别结果

为进一步验证综合统计特征策略流量识别的准确性和完整性,将基于综合统计特征识别策略的 Skype流量分析与识别系统所识别的 Skype数据,与 CommView 的 Latest IP Connections 中 process为 Skype exe的数据进行多次对比,发现两者所识别的 Skype远程连接和数据包数完全一致。

7　结束语

本文对 Skype通信机制及其流量识别进行了深入研究,并设计出验证系统对本文提出的结论进行了验证。但关于 Skype的研究仍然存在很多问题,比如对于纯 TCP的 Skype流量识别, Skype协议的详细分析研究,对于某个具体 Skype呼叫信令流程的追踪等,这些对于 P2P网络和 VoIP网络技术的研究是非常有意义的,需要进一步做更深入的研究。

本文所研究的 Skype通信机制以及提出的 Skype流量分析识别策略,能够对 Skype流量进行准确识别与分析,有助于 P2P网络技术以及 P2P网络流量识别与管理的研究。对于 P2P网络技术研究人员,可以更好的提高 P2P网络的可靠性以及资源利用的合理分配,提供用户满意的 P2P服务;对于网络管理人员可以有助于其对 P2P网络流量进行有效的管理与控制;对于传统电信运营商,可以从中获得改进现有 VoIP的关键技术,提高传统 VoIP业务的竞争力,并且可以对 Skype应用进行有效的控制。目前国内外还没有对 Skype通信机制及流量识别的详细研究,通过对 Skype通信机制及流量识别的研究,必将有效地推动 VoIP网络技术以及 P2P流量识别技术的进一步深入研究。

参考文献:

[1] Skype Limited Skype Explained[EB/OL]. http://www.skype.com/products/explained.html

[2] KARAGIANNIS T, BRODIO A, FALOUTSOS M, et al. Transport Layer Identification of P2P Traffic[C]// International Measurement Conference. Taormina Italy Oct 2004.

[3] SEN S SPATSCHECK O, WANG D. Accurate scalable in network identification of p2p traffic using application signatures[C]// Proceedings of World Wide Web Conference NY, USA, May 2004.

[4] 李江涛, 姜永玲. P2P流量识别与管理技术[J]. 电信科学, 2005, 21(3): 57-60.

[5] BASET S A, SCHRULZRINNE H. An analysis of the Skype peer to peer Internet telephony protocol[R]. New York: Columbia University Technical Report 2004.

[6] TamoSoft CommView[EB/OL]. http://www.ethernet analyzer com

(下转第 12页)

参考文献:

[1] 蒋景瞳, 刘若梅, 贾云鹏. 国际元数据标准的发展和研究现状[EB/OL]. http://www.sdinfo.net.cn/ngcc/sdinfo/prodtecteddoc/mt1.htm

[2] MANOLA F, MILLER E. RDF Primer[EB/OL]. http://www.w3.org/TR/rdf-primer

[3] CHALMERS D, DULAY N, SLOMAN M. Meta Data to Support Context Aware Mobile Applications[C] // Proceedings of the 2004 IEEE International on Mobile Data Management (MDM'04). 2004.

[4] ITU-T. F.750. Metadata Framework Recommendation[S]. February 2005.

[5] McCARTHY D. Jena简介[EB/OL]. http://www-128.ibm.com/developerworks/cn/java/j-jena/?ca=dwcn-newsletter-java

[6] KLYNE G, REYNOLDS F, WOODROW C, et al. W3C Recommendation Composite Capability/Preference Profile (CC/PP): Structure and Vocabularies 1.0[EB/OL]. http://www.w3.org/TR/2004/REC-CCPP-struct-vocab-20040115/

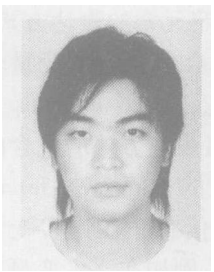
[7] 郭志红. 元数据的多角度透视[EB/OL]. http://www.lib.sjtu.edu.cn/chinese/teaching&research/4041.htm

[8] 赵永平, 承继成, 李琦. 基于我国NSII关键技术研究的元数据标准内容体系[EB/OL]. http://www.sdinfo.net.cn/ngcc/sdinfo/ProdtectedDoc/mt6.htm

[9] 肖珑, 陈凌, 冯向云. 中文元数据标准框架及其应用[EB/OL]. http://www.idl.pku.edu.cn/pdf/metadata_framework.pdf

[10] 高登凤, 杨冬青, 唐世渭. 元数据管理与空间信息共享[EB/OL]. http://www.sdinfo.net.cn/ngcc/sdinfo/prodtecteddoe/mt4.htm

作者简介:

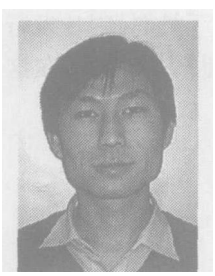


杨 赟(1982-),男,浙江宁波人。南京邮电大学通信与信息工程学院硕士研究生。2005年毕业于南京邮电大学通信工程系。目前主要的研究方向为IP与宽带网络技术。

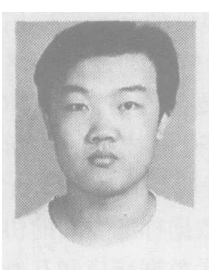
糜正琨(1946-),男,浙江上虞人。南京邮电大学通信与信息工程学院教授、博士生导师,ITU-T中国专家组成员。1967年毕业于复旦大学物理系,1981年在南京邮电学院获工学硕士学位。目前主要研究方向为宽带交换和通信网技术。

(上接第7页)

作者简介:



王振华(1982-),男,河北衡水人。南京邮电大学计算机学院硕士研究生。2004年毕业于南京邮电学院光信息技术系。目前主要研究方向为计算机通信网络。



王 攀(1977-),男,新疆阿克苏人。南京邮电大学网络技术研究中心助教。2004年在南京邮电学院计算机科学与技术系获硕士学位。目前主要研究方向是IP网络服务质量检测与管理。

张顺颐(1944-),男,江苏南京人。南京邮电大学副校长、教授、博士生导师。1968年毕业于天津大学,1983年至1985年在日本国立电气通信大学电子工学科进修。目前主要研究方向是计算机通信网及IP技术。