计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2014.20121202
51(8):1695-1703,2014

StegoP2P:一种基于 P2P 网络的隐蔽通信方法

谭庆丰[1,2,4]　方滨兴[1,2,3]　时金桥[1,2]　徐钒文[3]　陈小军[1,2,4]
[1](中国科学院信息工程研究所　北京　100093)
[2](信息内容安全技术国家工程实验室　北京　100093)
[3](北京邮电大学计算机学院　北京　100876)
[4](中国科学院大学　北京　100049)
(tanqingfeng@iie.ac.cn)

StegoP2P: A Hidden Communication Approach in P2P Networks

Tan Qingfeng[1,2,4], Fang Binxing[1,2,3], Shi Jinqiao[1,2], Xu Fanwen[3], and Chen Xiaojun[1,2,4]
[1](*Institute of Information Engineering*, *Chinese Academy of Sciences*, *Beijing* 100093)
[2](*National Engineering Laboratory for Information Security Technologies*, *Beijing* 100093)
[3](*Department of Computer Science*, *Beijing University of Posts and Telecommunications*, *Beijing* 100876)
[4](*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract　With the development of Internet, privacy-preserving has become an increasingly prominent problem. Existing anonymous communication systems, such as Tor and Freenet, can conceal who communicate with whom. However they can't hide the fact that the users are using the anonymous communication technologies. File share software, such as BitTorrent and emule, has become the most popular application in Internet with users all over the world. In this paper, we present StegoP2P, a peer-to-peer based hidden communication method, which doesn't rely on a single system or a set of entry points. It is based on embedding the steganographic marker in the peer-to-peer meta-data exchange protocol, unlike other existing covert communication methods that rely on timing channel, and requires time synchronization. An efficient covert handshake protocol with steganographic marker techniques over peer-to-peer networks is proposed for unobservable communications, which allows users in peer-to-peer networks to exchange information secretly for circumventing Internet censorship. The steganography makes it easy for users to find the targeted content and difficult for a censor to identify them. Experimental results and security analysis show that our system has high performance and can defense against certain traffic censorships.

Key words　privacy-preserving; secret handshakes; covert channel; peer-to-peer; DHT

摘　要　随着互联网的发展,隐私保护问题越来越突出,现有的匿名通信系统如 Tor,Freenet 等能够隐藏用户的身份,但是不能够隐藏通信的事实.以 BitTorrent,emule 为代表的 P2P 文件分享软件已经成互联网的主要应用,而且其用户规模庞大,分布在全球各地.在此基础之上提出一种新的隐蔽通信方法——StegoP2P,该方法不依赖于任何单一的系统或者接入点,而是利用 P2P 协议的隐蔽通道;并设计一个有效的隐蔽握手协议,让 P2P 网络中互为好友(合谋)但素未谋面的节点隐蔽握手,秘密交换信息,从而绕过网络审查.实验结果和安全性分析表明我们的系统具有较高的性能和健壮性,并能够抵御常见的流量审查.

收稿日期:2012-12-21;**修回日期**:2013-05-15
基金项目:国家自然科学基金项目(61100174);国家"八六三"高技术研究发展计划基金项目(2011AA010701,2012AA013101);国家科技支撑计划项目(2012BAH37B04)

关键词 隐私保护;隐蔽握手;隐蔽通道;对等网络;分布式散列表

中图法分类号 TP393.08

如今,无处不在的网络监视、流量审查已经对用户隐私和网络安全构成严重威胁,传统的隐私保护技术都是基于加密隧道和重路由技术(如 Tor,VPN 等).首先,加密隧道具有明显的协议特征,即流指纹,审查者可以通过通信行为、流量指纹特征识别这些加密隧道的连接;其次,代理节点容易被发现.因此,审查者可以监视或者阻断所有到这些代理节点的连接.为了解决这些问题,本文提出一种基于 peer-to-peer(P2P)的隐蔽通信技术.众所周知,P2P 系统及其应用是一种分布式的系统,由于其可伸缩性和对等的通信特性,已经被互联网用户广泛使用.然而,今天的 P2P 网络(如 emule,BitTorrent)其本身并不提供给用户较好的隐蔽匿名的文件分享.一方面,emule,BitTorrent 网络具有很好的伸缩性和较高的性能,但是对等节点很容易被任意其他对等的节点监视[1-3];另一方面,匿名 P2P 网络(如 Freenet[4],oneswarm[5])提供匿名文件分享的特性,能够保护用户隐私,但是不能够隐藏其通信的事实.

本文关注 P2P 网络的隐蔽通信方法,考虑互为好友(合谋)但素未谋面的对等节点匿名认证的场景,P2P 网络中合谋的对等节点需要秘密地找到对方,相互认证,然后隐蔽地交换信息,因此,我们提出一种基于 P2P 网络的隐蔽握手协议,该协议的目标是提供足够的隐蔽性,同时具有一定的可用性和较高的性能.设计一个基于 P2P 网络的隐蔽通信系统存在以下挑战:1)由于 P2P 网络固有的分布式特性,其节点具有高动态性,且用户数量非常庞大,分布在全球范围内,如何在 P2P 网络中找到合谋的对等节点,即隐蔽握手的问题;2)握手成功以后,如何通过 P2P 网络的隐蔽通道将消息发送到合谋的对等节点,本文设计和实现一个基于 Distributed hash table(DHT)通道的隐蔽匿名数据分享原型系统,该系统将编码后的数据秘密地插入到 P2P 网络的 DHT 中,通过 P2P 网络将它们分散到整个 DHT 网络中,信息的接收者可以通过基于共享秘密的资源发现算法找到这些文件分片,并重构文件.

1 相关工作

1.1 隐蔽信道

隐蔽信道为一种非授权的信道,在匿名认证和隐蔽通信中扮演非常重要的角色.通常,隐蔽信道分为两种类型:基于计时的隐蔽信道和基于存储的隐蔽信道.传统的基于计时的隐蔽通道一般是通过数据包的发送时间间隔、频率等特性来编码目标信息[6-7].如 2004 年 Cabuk 提出了一种基于 IP 计时的隐蔽信道[8],在该信道中,发送者和接收者通过事先约定的时间间隔周期性地产生数据包,0 表示在该时间间隔内不发包,1 表示在该时间间隔内发送包.其检测方法主要针对数据包的时间序列特性,统计其包的时间分布,并根据阈值判断包的时间间隔是否被修改[9-10].为了解决这些问题,现在比较先进的隐蔽计时信道可以根据模型产生不同的包的时间分布,以此提高检测隐蔽计时信道的难度.

基于存储的隐蔽信道一般是利用协议中没有使用的字段,或者某些随机字段,即利用隐写术做标注的方法,例如 IP 头的 TTL 字段[11]、TCP 时间戳[12]、甚至 TCP 的初始化序列(ISN)[13]等,Murdoch 和 Lewis 在他们的文献[13]中详细分析了基于 TCP/IP 协议的隐蔽存储信道.

1.2 隐蔽通信

Feamster 等人基于非对称通信理论设计出一个抗流量审查的隐蔽通信系统 Infranet[14],该系统利用 Web 通信行为,将一个隐蔽的 HTTP 请求编码为一系列正常的掩体 HTTP 请求,并利用隐写术将目标内容隐藏在掩体资源文件的图片中.然而 Infranet 需要部署自己隐蔽服务端,且具有迭代次数过多、延时过大等缺点.Wustrow 等人提出 Telex[15],其基本思想是构建一个基于 TCP/IP 协议的隐蔽通道(利用隐写术在 HTTPS 协议的随机域做秘密标注的方式),然后在骨干网的路由器上检测该通道,并利用路由重定向技术来实现不可观测的通信,然而,Telex 方法需要友好的 ISP 在系统的客户端和掩体服务器之间的骨干网上部署一种特别的路由器(如 Telex Station),而 Eidenbenz 等人提出一种基于 P2P 网络的隐蔽通信(hidden communication)[16]:该方法利用 P2P 文件分享,将目标信息编码到一系列文件块(block)的请求序列中.比如一个文件分割成 m 块 $B_1,\cdots,B_m$,将合谋节点之间共享的秘密编码成一个特别的请求顺序,并按照这个顺序请求该文件,信息的接收方则解码该顺序,从而判断对方是否为一个合谋节点.

2 系统架构

本节将介绍隐蔽通信系统的架构和设计考虑,并给出系统的威胁模型和设计目标.其系统架构如图 1 所示:

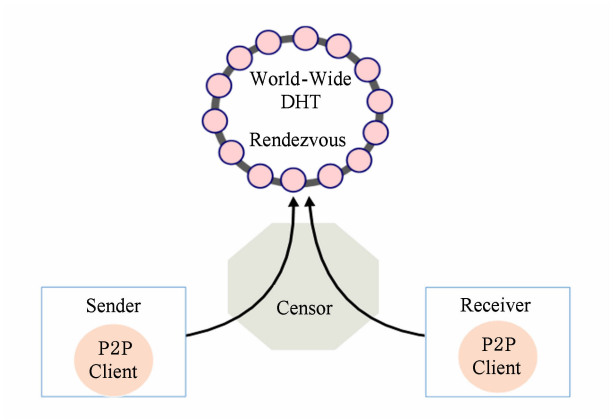


Fig. 1 System architecture.

图 1 系统架构

2.1 概 述

信息的发布者 Alice 和接收者 Bob 为 P2P 网络中互为好友(合谋)但素未谋面的节点,其中 Alice 要秘密地找到 Bob,并实现相互秘密认证,协商约会地点,最后,Alice 和 Bob 通过协商好的第三方约会地点实现隐蔽的通信.整个隐蔽通信过程分成两个阶段:首先,Alice 和 Bob 需要在全球范围内的大规模 P2P 网络环境中找到对方,实现相互认证,这个过程我们称之为隐蔽握手阶段;第 2 个阶段为数据传输通道,即通过第 1 个阶段,我们知道对方为好友节点,这个时候我们需要基于 P2P 网络建立一个隐蔽匿名的通道以秘密交换信息.

2.2 威胁模型

考虑 n 个对等节点组成的 P2P 网络,其节点集合为 P,$|P|=n$,通常,一个对等节点 $v\in P$ 要加入 P2P 网络,它需要通过种子节点查找与它在逻辑上相邻的节点,并将这些节点加入到自己的路由表中,然后跟其他节点交换信息.在这里,我们将对等节点分成普通节点和合谋节点.普通节点只是正常的 P2P 用户,不主动提供隐蔽通信服务.合谋节点为这些 P2P 节点之间可以主动提供隐蔽通信服务的节点,即合谋节点之间通过隐蔽握手协议,协商约会地点,建立隐蔽信道的节点集合,这个集合我们记为 C.集合 C 的大小:$|C|=m$,且 $m\ll n$.因此,我们考虑系统的威胁模型分为没有监视、部分监视和完全监视 3 种情况:

没有监视.对抗者没有能力监视 P2P 网络中的任何其他对等节点的连接以及信息交换,除非一个合谋节点被其他对等节点发现,并举报给攻击者.

部分监视.对抗者可以对部分对等节点的连接和交换内容进行监视和审查,但是不能够监测到所有节点的活动,即对抗者具有局部视野.

完全监视.对抗者有能力对整个 P2P 网络的活动节点进行监控,对其流进行审查.即对抗者具有全局视野.

在本文我们假定对抗者具有部分监视能力.

2.3 设计目标

信息发布者的可抵赖性:即从通信行为来说是一个计算上的迷题,审查者不能够确认或者区分该用户是否有意发布一些特别的信息.

私密性原则:通信双方的内容对于任何第三方是不可见的.

隐蔽性原则:由于对抗者可能会监视 P2P 网络的部分节点的活动规律,因此,在整个通信过程中审查者很难从通信行为和传输的内容上区分一个对等节点是一个普通节点还是一个特别的节点.

容错性:隐蔽信道应该是可靠的,即信息具有一定的冗余,不会因为某些分片丢失而导致整个内容不可重构.

3 通信协议设计

3.1 理解 DHT

DHT 是一种分布式的、为 P2P 网络提供元数据存储的散列表,将 P2P 用户的节点以其分享的原文件映射成一系列$\langle key,value\rangle$对,存在本地的路由表.每一个 Node ID 通常由 160 b 的散列值生成.目前 P2P 网络中的 DHT 的实现主要有 KAD 协议、Vuze 和 Mainline 等.一般如 KAD 网络为上层应用提供一系列 API,包括 *PING*,*FIND_NODE*,*FIND_VALUE*,*STORE* 等操作,其中 *FIND_NODE* 以 160 b 的 *Node-id* 为查询参数,返回与该 *Node-id* 在逻辑上最近的节点信息如$\langle$*IP address*,*UDP port*,*Node-id*$\rangle$,*FIND_VALUE* 跟 *FIND_NODE* 的行为相似,通过查询 *STORE* 存储的 *key*,并返回与该 *key* 相应的 *value*. *STORE* 向 DHT 发布$\langle key,value\rangle$,其过程包括两个部分,先通过迭代或者递归查找与 *key* 在逻辑上相邻的节点,然后向该节点插入$\langle key,value\rangle$.

3.2 隐蔽握手协议

隐蔽握手在匿名认证和隐蔽通信过程中扮演非常重要的角色,本文提出一种基于对称密码学的秘密标注方法,在相同的好友节点之间,隐蔽握手协议通过共享的秘密 K 为通信双方建立一个安全的匿名和隐蔽的通道,其中共享的秘密可以通过带外的方式发布(如 Email). 假设 Alice 和 Bob 为 P2P 网络的互为好友节点,他们之间要相互认证,其认证过程分为如下 3 个步骤:

第 1 步. Alice 向 Bob 认证,即 Alice 通过共享的秘密,生成带隐写标记的键值对,然后发布到 P2P 的 DHT 网络中,其算法如下所示:

算法 1. Alice 向 Bob 认证算法.

输入: Alice 和 Bob 共享的秘密(K);

输出: Alice 和 Bob 的约会地点,即键值对.

$init\_seed = SR(SEED\_LENGTH)$;

$init\_key = MAC(FIRST\_HALF(K) | init\_seed)$;

$encrptext = SE_{init\_key}(LAST\_HALF(K) | InitLocator)$;

$init\_keyword = MAC(FIRST\_HALF(K)$ XOR $LAST\_HALF(K))$;

$init\_value = init\_seed | encrptext$;

$STORE(init\_keyword, init\_value)$.

算法 1 中 SR 为随机数生成算法,MAC 是散列算法,SE_{init_key} 为 AES 对称加密算法,其中密钥为 $init\_key$,长度为 16 B,$a|b$ 为字符串 a 与 b 相连接,c XOR d 为 c 和 d 做异或运算,$FIRST\_HALF(e)$ 为字符串 e 的前半部分,$LAST\_HALF(e)$ 为字符串 e 的后半部分.

Alice 要向 Bob 秘密认证,首先 Alice 生成长度为 $SEED\_LENGTH$ 的种子 $init\_seed$,然后通过种子和共享的秘密 K 的前半部分生成长度为 16 B 的密钥;然后 Alice 通过生成的 $init\_key$ 加密 K 的后半部分和 Bob 向 Alice 认证的地点(即 $InitLocator$). 最后,Alice 生成〈$init\_keyword, init\_value$〉对,并向 P2P 网络的 DHT 发布这个键值对.

第 2 步. Bob 为了能够与 Alice 接上头,则需要周期性地到协商好的约会地点搜索 Alice 的接头暗号,即 Alice 发布的键值对.

算法 2. Bob 搜索并验证 Alice 发布的键值对.

输入: Alice 和 Bob 共享的秘密(K);

输出: 如果 Alice 是 Bob 的好友,则返回真,否则返回假.

$init\_keyword = MAC(FIRST\_HALF(K)$ XOR $LAST\_HALF(K))$;

$init\_value = FIND\_VALUE(init\_keyword)$;

$init\_key = MAC(FIRST\_HALF(K) | init\_value[:SEED\_LENGTH])$;

$plaintext = SD_{init\_key}(init\_value[SEED\_LENGT:])$;

验证 $plaintext$ 的前半部分是否等于 $LAST\_HALF(K)$.

算法 2 中 $FIND\_VALUE$ 为 P2P 递归查询算法,即向 DHT 网络递归查询健为 $init\_keyword$ 的值,并取 $init\_value$ 前 $SEED\_LENGTH$ 个字节,SD_{init_key} 为 AES 解密算法,其密钥是 $init\_key$,并验证其共享密钥的后半部分.

第 3 步. 如果 Alice 为 Bob 互为好友节点,则响应之,并向 Alice 认证,同时发布其下一次约会的位置索引信息,即 $resp\_keyword$.

算法 3. Bob 向 Alice 认证算法.

输入: Alice 和 Bob 共享的秘密(K);

输出: Alice 和 Bob 的数据通道约会地点,即键值对.

$resp\_seed = SR(SEED\_LENGTH)$;

$resp\_key = MAC(FIRST\_HALF(K) | RESP\_SEED)$;

$encrptext = SE_{resp\_key}(LAST\_HALF(K) | RespLocator)$;

$resp\_keyword = MAC(FIRST\_HALF(K)$ XOR $LAST\_HALF(K) | InitLocator)$;

$resp\_value = resp\_seed | encrptext$;

$STORE(resp\_keyword, resp\_value)$.

Alice 和 Bob 相互认证,并向 Alice 发布随机索引,即 Alice 和 Bob 隐蔽通信的约会地点. 至此,隐蔽握手过程完成,在 3.3 节我们将讨论如何构建隐蔽通道.

3.3 隐蔽信道构建

3.2 节我们讨论了如何在 P2P 网络的好友节点之间相互认证,本节我们将详细讨论如何在 DHT 中利用阈值秘密分享算法和 Bob 响应的位置索引 $RespLocator$ 构建一个隐蔽信道. 这个方案可能存在以下几个方面的挑战:首先,为了实现客户端可抵赖性,所有好友节点之间的通信行为应该是一个正常的通信方式,好友节点之间是通过看似随机的第三方无辜节点进行通信的,而实际是通过秘密协商的,在攻击者看来,该节点是一次正常的查询、存储

操作,而不具有行为上的异常,任何其他 P2P 用户也都可以到该节点上搜索、存储元信息.其次,信息发布方能够保证信道的可靠性,即不会因为某些节点下线或者分片丢失而导致整个信息的不可重构.最后,信息的接收方能够正确地识别该信息,即该信息没有被伪造、篡改.

DHT 网络为 P2P 用户的元信息提供了存储空间,每一个 P2P 用户将自己的 IP 地址和端口映射成一个 *Node-id*,然后该用户将需要分享的文件名以及对该文件的描述信息映射成位置索引信息(一般为文件名或者内容的散列值).因此,我们可以将目标内容加密、签名、冗余编码后利用 DHT 的〈*key*,*value*〉对的 *value* 域,并通过 3.2 节秘密握手方式协商好的约会地点在好友节点之间秘密地交换信息,而不需要双方直接交互.

由于 P2P 节点状态的不确定性,即某个对等节点可以随时加入和退出 P2P 网络,且 DHT 网络的信道容量非常有限,如 KAD 网络设定其每一条记录的键值对大小为不超过 2 KB,为了提高隐蔽信道的容量和可靠性,我们首先利用阈值秘密分享算法[17]将需要交换的目标内容 D 分成 N 个分片,$D_1,\cdots,D_N$,其中任意 M 个分片($M<N$)就可以重构原内容信息.例如,我们将一个文件分散成 100 片,即 $N=100$,如果选择阈值比率大小为 0.5 即只需要其中的 50 片就可以重构原文件内容,参数 M 为 50.其次,为了在合谋节点之间秘密交换信息,我们需要利用 3.2 节协商好的位置索引 ID(*RespLocator*),生成若干个 sub-id:$I_1,\cdots,I_N$,sub-id 的生成方法有很多种,比如通过按位与的方法、计数方法等,只需要接收者和发布者生成算法一致即可.并将每一个编码后的分片映射到这些 sub-id.最后,向 DHT 网络发布〈I_1,D_1〉,…,〈I_N,D_N〉,即对于每一个 $i\in\{1,\cdots,N\}$,信息的发布者将在索引位置 I_i 存储与之相应的 D_i.

信息的接收者根据位置索引 *RespLocator*,并通过与信息发布者同样的算法生成一系列的 sub-id,其中每一个 sub-id 对应一个分片,因此,信息的接收者只需要通过 DHT 的 *FIND_VALUE* 操作即可在 DHT 网络找到相应的分片,然后重构这些分片信息.

4 实现和性能评估

我们基于 amule(一个开放源码的 P2P 软件系统)实现了一个原型系统.实验的网络环境为 ADSL 拨号网络.实验步骤分为 2 部分:1)信息秘密分享,即将一个文件秘密分享给信息的接受者;2)通过隐蔽通道在 DHT 网络中搜索秘密分享的文件信息.系统性能测试主要关注隐蔽信道的容量即在成功搜索到秘密分享的文件的前提下,其信道带宽(即信息传输的速度)与冗余比率,搜索的并行数量之间的关系.

在分析系统性能之前,有必要先了解系统中跟性能有关的参数及其含义,如表 1 所示:

Table 1 System Parameters

表 1 系统参数

Parameters	Value	Description
SEARCHSTOREKEYWORD_LIFETIME/s	140	Maximum time limit for storing a key word
SEARCHSTORENOTES_TOTAL	10	The number of nodes for notes
SEARCHKEYWORD_LIFETIME/s	45	Maximum time limit for searching a key word
KADEMLIATOTALSTOREKEY	10	Total hashes to store
ENCRIPTION	0	Setting 0 if symmetric encryption; setting 1 if not
SEARCH_NUM	16	Total parallel number to search key words
PIECE_SIZE / B	1 024	Pieces size
R	0.5	Threshold ratio

本文的实验环境为 2 Mbps 带宽的 ADSL 拨号网络环境,系统采用如表 1 所示的参数设置,匿名信息分享者发布一个大小为 308 KB 的文件.实验结果如图 2、图 3 所示.

由图 2 所示,由于网络环境不错,分片在上传过程中基本没有超时(即在 140 s 之内就已经完成 10 个节点的存储过程),因此,其实际测试得到的速度比理论值的下界要快,而隐蔽信道的上行速度的理论

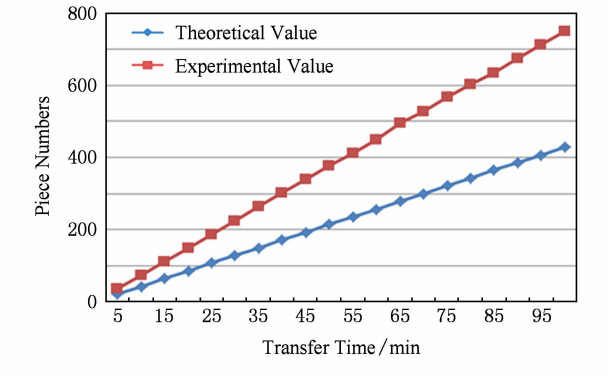


Fig. 2 Piece numbers w. r. t transfer time.

图 2 文件片数与发布时间的关系图

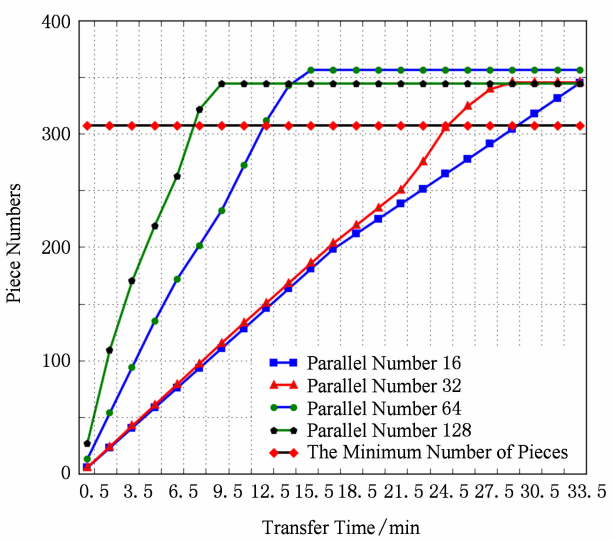


Fig. 3 Search numbers and pieces w. r. t transfer time.

图 3 文件片数与搜索时间以及搜索并行数的关系

下界为式(1)所示,主要取决于每一个分片上传速度 V(V 跟本地网络环境如带宽,以及搜索的并行数有很大的相关性),以及冗余编码 R,每个分片的大小 $PIECE\_SIZE$(理论上限 2 KB)和加密方式 $ENCRIPTION$(如果对称加密则为 0,非对称加密为 1):

$$f=V\times PIECE\_SIZE\times\frac{1}{1+R}\times\frac{1}{1+ENCRIPTION\times 0.33}. \tag{1}$$

因此,系统搜索的并行数越高,系统消耗的网络带宽越高,若阈值比例系数 R 越小,则编码后的分片数量就会越小,系统的上传速度就越大,反之 R 过大,上传速度则会变小.

在下行阶段,即搜索并下载分片阶段,从图 3 的测试结果可以看出,搜索成功率和搜索并行数关系不大,而总共所需要的时间与搜索并行数有一定的关系,当并行数量为 128 时(amule 中能修改的理论搜索并行数为 256),约 8 min 就可以恢复 308 KB 的文件,效率相当高.

下面将分析下行阶段文件重构的性能以及不同的参数选择对系统性能的影响. 成功获取一个文件的时间主要取决于以下参数:并行搜索数 $SEARCH\_NUM$、阈值比率 R、分片数量 N、每轮搜索成功率 p 以及速度 L(单位为每分钟所下载的分片数量).

$$t=\frac{N\times(1-(1-p)^n)}{(1.5\times SEARCH\_NUM)\times L}; \tag{2}$$

$$\frac{R}{1+R}>(1-p)^n. \tag{3}$$

由此可见,总共花费的搜索时间 t 可以通过式(2)计算出,其中搜索轮次 n 可以通过式(3)求得,而从实际测试的情况看,网络的性能对于搜索的成功率影响比较大.

因此,在网络环境(带宽,负载)一定的情况下,成功搜索到 M 个文件(重构文件所需的分片数)所需的时间主要取决于并行搜索数量、阈值比率、以及文件大小(分片数量),而网络环境对于搜索的成功率以及上行和下载速度的影响非常大.

5 安全性分析

由于隐蔽通信的对抗性非常强,在这个部分我们将从审查者的角度来讨论系统的安全性. 为了评估系统地安全性,我们将重点分析系统用户可能面临的一些安全性问题:如系统的可用性、客户端的可抵赖性、隐蔽性、以及传输内容的私密性.

5.1 攻击者

本文定义的攻击者可能为一个恶意对等节点,也可能为不友好的 ISP,或者是其他组织,他们能够通过主动或者被动的方式监听、篡改、拦截 P2P 通信流,监视其通信行为. 本文假设攻击者具有局部视野,即能够监视部分对等节点的活动情况,能够对其通信行为和流量进行拦截和分析.

P2P 网络节点分布在全球各地,而这种分布式的架构和性质决定了攻击者很难监视整个 P2P 网络的行为;其次,DHT 超时机制使得存储在散列表的分片信息会被自动刷新(即超过一定的时间,其键值对会被较新的键值对替换),这样增加了攻击者在规定时间内收集、检索这些分片的难度. 最后,P2P 节点动态变化的特性:假设在某一时间,某一对等节点被秘密存储一个分片,在过去某个时间间隔内,该节点从 P2P 网络下线后,然后重新上线,其节点身份($Node\text{-}id$)信息可能已经改变了,这样增加了攻

击者持续追踪和监视某一些节点的难度.

5.2 容错和可用性分析

系统的可用性包括两个方面含义:一方面,由于攻击者怀疑某个信息为有害信息,对该信息进行审查和过滤,甚至屏蔽该节点;另一方面,由于系统设计时没有考虑到容错,比如没有考虑到 P2P 网络的动态变化的特性,致使某些分片的丢失而导致系统的不可用(不能够重构原文件).在设计目标部分,我们提到系统的设计目标包括信道的可靠性,即系统不会因为某些分片的丢失而导致整个信息不可用,也不会因为攻击者对 P2P 网络的审查或者是某个不友好的 ISP 屏蔽某些恶意节点而使得系统不可用.下面我们将从攻击者的角度来分析系统的可用性.

攻击者通常是不愿意屏蔽整个 P2P 网络,或者说屏蔽 P2P 网络的合法用户.但是他能够监视 P2P 网络的节点行为,由于目标信息被编码到 DHT 的元信息存储空间,该信道的字段 *key* 为一个散列值,字段 *value* 为该节点将要向 P2P 网络分享的文件描述信息.从攻击者的角度来看,很难从海量的对等节点中区别出某个字段 *value* 的信息是一个特别的信息,因此,如果攻击者在不确定的情况下屏蔽某个节点的元信息,将导致其他合法用户同样不可能使用该信息.

其次,系统的可用性很大程度取决于阈值秘密分享算法,其参数选择如阈值比率、分片数量等都可以影响系统的可靠性.通常,增加分片数量或者是提高阈值比率都可以提高系统的可用性,但是增加分片数量又会增加系统负载,降低信道容量.

最后,攻击者的规模对于系统的可用性也会有很大影响.在 DHT 网络规模大小一定情况下,可以接受的攻击者规模以及成功收集到某个分片的概率,将在另一论文详细讨论,在这里我们只做定性的分析.通常,攻击者规模越大(也可能同一攻击者变换多个身份如 sybil 攻击),系统的可用性也就越低.

5.3 客户端的可抵赖性

系统的主要目标是抵御网络审查和监视,同时,为使用该系统的用户提供一定的可抵赖性,不像匿名性、可抵赖性很难量化分析,如果审查者不能够从正常的通信行为或者通信流中区分出某个特别节点的通信行为,我们就认为该客户端的通信行为是可抵赖的,下面我们将讨论系统可能面临的各种针对客户端可抵赖性的攻击以及对抗方法.

攻击者可能加入 P2P 网络,伪装成一个 P2P 用户,监视其他对等节点的活动,以及连接关系、更强的攻击者,如不友好的 ISP 可能对所有经过的 P2P 节点都能够监视.为了提高用户的可抵赖性,我们可以选择分享一些可抵赖性的内容;其次,节点的通信行为应该尽可能地符合正常的通信行为,如通信行为应该不具有周期性,上行通信的流量和下行通信的流量应该服从 P2P 网络的流量分布特征.最后,位置索引信息应尽可能的随机,且分散在在整个节点的度量空间.

5.4 可检测性

本节我们将讨论系统在局部监视情况下的可检测性,由于对抗者可能监视所有可疑节点的通信行为,追踪并屏蔽所有可疑的节点.因此,从对抗者角度来看,攻击者可能采用流分析等方法识别出合谋节点共享的秘密 KEY.如果对于任意概率多项式时间的攻击者 A,其识别出该 KEY 值的概率是一个可忽略的值 ε,即概率多项式时间的攻击者 A 攻击成功的优势是可忽略的,我们认为该方案就是安全的,即带标记的秘密 KEY 具有不可区分性.形式化定义如下:

对于任意概率多项式时间的攻击者 A,定义 KEY 的度量空间为 M,带隐写标记 KEY 的度量空间为 D,且 $D\subseteq M$,如果存在一个可以忽略的值 ε,使得:

$$Pr(d\rightarrow M:indistinguishable_{A,d}=1)\leqslant\varepsilon. \quad (4)$$

对于任意 $d\in D$,记 $indistinguishable_{A,d}=1$ 为攻击者 A 的一次可区分性实验,这样 d 可以被区分的概率为 ε,理想情况下 ε 为 0.

在数据通信阶段,攻击者 A 可能通过女巫攻击(sybil attacks)方式收集分片信息,即一个攻击者,同时生成大量的 sybil 节点加入 P2P 网络,通过主动抓取或者被动等待方式收集分片信息.其概率模型可以等价于彩票收集模型[18].

若成功收集到一个分片的概率定义为 p,收集一个分片的概率可以看成二项随机变量,因此我们可以定义其重构一个原文件的概率等价于 sybil 攻击者成功收集到 k 个分片信息的概率:

$$Pr(\text{recover original file})=\sum_{i=n-k}^{n}\binom{n}{i}p^i(1-p)^{n-i}. \quad (5)$$

假设每个 sybil 节点平均可以看到 c 个分片,总共有 m 个 sybil 节点,则 sybil 攻击者 A 总共可以收集 cm 个分片.一旦 cm 达到一定的比例,则收集到新分片的概率将越来越小.因此,sybil 攻击监测到分片比例的期望值为

$$E(\text{pieces detected})=1-e^{-ms}. \quad (6)$$

其中 $s=\frac{c}{n}$.

根据上面的分析,系统的可检测性取决于两个方面的因素:首先,在隐蔽握手阶段,隐写标记的键值对的可区分性,如果 KEY 的可区分性比较高,则攻击者 A 很容易监测到秘密约会地点;其次,系统的安全性还取决于阈值秘密分享算法的阈值比率 $\frac{k}{n}$ 和攻击者 A 的攻击能力和付出的代价. 由于大部分 ISP 和组织内部广泛应用 NAT 技术,因此,现有的 Kad 网络和 Vuze DHT,OpenDHT[19] 等重叠网络都允许同一个 IP 拥有多个身份,但在 sybil 节点数量上都有一定程度的限制,比如 Vuze DHT 限制同一个 IP 最多可以拥有 64 个合法的身份,如果攻击者 A 拥有大量的 IP 地址,则成功获取到 k 个分片的可能性则越高;因此,为了保证其安全性可以在加密后再做冗余编码,编码后的分片可以分散到多种 DHT 网络,以提高系统的安全性.

6 结 论

本论文提出一种基于 P2P 网络的隐蔽通过方法,该方法利用 DHT 网络建立一个隐蔽握手协议,使互为好友而素未谋面的 P2P 用户能够秘密地交换信息,从而逃避网络审查. 信息的发布者和接收者共享一个秘密,并通过该秘密计算出位置索引信息(即约会的地点),只有知道这个共享秘密的好友节点才能够找到该约会地点,这个过程为隐蔽握手阶段,其目的是相互秘密认证,并协商数据通道的约会地点,握手成功以后,信息的发布者将目标文件编码后隐藏在 DHT 网络中,并将该消息插入到协商好的约会地点,信息的接收者通过同样的算法去协商好的秘密约会地点检索相应的掩体信息,并重构原始信息. 实验结果和安全性分析表明:该系统不仅具有较好的性能,而且具有很高的安全性,并能够抵御常规的流量审查.

参 考 文 献

[1] Piatek M, Kohno T, Krishnamurthy A. Challenges and directions for monitoring P2P file sharing networks-or: Why my printer received a DMCA takedown notice [C] //Proc of USENIX Association HOTSEC'08. Berkeley, CA: USENIX Association, 2008: 1-7

[2] Siganos G, Pujol J M, Rodriguez P. Monitoring the bittorrentmonitors: A bird's eye view [C] //Proc of the 10th Int Conf on Passive and Active Network Measurement. Berlin: Springer, 2009: 175-184

[3] Piatek M, Isdal T, Krishnamurthy A, et al. One hop reputations for peer to peer file sharing workloads [C] //Proc of the 5th USENIX Symp on Networked Systems Design and Implementation. Berkeley, CA: USENIX Association, 2008: 1-14

[4] Clarke I, Sandberg O, Wiley B, et al. Freenet: A distributed anonymous information storage and retrieval system [C] //Proc of Int Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability. Berlin: Springer, 2001: 46-66

[5] Isdal T, Piatek M, Krishnamurthy A, et al. Privacy-preserving P2P data sharing with OneSwarm [C] //Proc of ACM SIGCOMM'10. New York: ACM, 2010: 111-122

[6] Cabuk S. Network covert channels: Design, analysis, detection, and elimination [D]. West Lafayette, IN: Purdue University, 2006

[7] Shah G, Molina A, Blaze M. Keyboards and covert channels [C] //Proc of the 15th Conf on USENIX Security Symp. Berkeley, CA: USENIX Association, 2006: 59-75

[8] Cabuk S, Brodley C E, Shields C. IP covert timing channels: Design and detection [C] //Proc of ACM CCS'04. New York: ACM, 2004: 178-187

[9] Gianvecchio S, Wang H. Detecting covert timing channels: An entropy-based approach [C] //Proc of ACM CCS'07. New York: ACM, 2007: 307-316

[10] Berk V, Giani A, Cybenko G. Detection of covert channel encoding in network packet delays, TR536 [R]. Hanover, NH: Department of Computer Science, Dartmouth College, 2005

[11] Zander S, Armitage G, Branch P. An empirical evaluation of IP time to live covert channels [C] //Proc of IEEE ICON'07. Los Alamitos, CA: IEEE Computer Society, 2007: 42-47

[12] Giffin J, Greenstadt R, Litwack P. Covert messaging through TCP timestamps [C] //Proc of the 2nd Int Conf on Privacy Enhancing Technologies. Berlin: Springer, 2003: 194-208

[13] Murdoch S J, Lewis S. Embedding covert channels into TCP/IP [C] //Proc of the 7th Int Conf on Information Hiding. Berlin: Springer, 2005: 247-261

[14] Feamster N, Balazinska M, Harfst G, et al. Infranet: Circumventing Web censorship and surveillance [C] //Proc of the 11th USENIX Security Symp. Berkeley, CA: USENIX Association, 2002: 247-262

[15] Wustrow E, Wolchok S, Goldberg I, et al. Telex: Anticensorship in the network infrastructure [C] //Proc of the 20th USENIX Conf on Security. Berkeley, CA: USENIX Association, 2011: 30-30

[16] Eidenbenz R, Locher T, Wattenhofer R. Hidden communication in P2P networks steganographic handshake and broadcast [C] //Proc of IEEE INFOCOM'11. Piscataway, NJ: IEEE, 2011: 954-962

[17] Shamir A. How to share a secret [J]. Communications of the ACM, 1979, 22(11): 612-613

[18] Mitzenmacher M, Upfal E. Probability and Computing: Randomized Algorithms and Probabilistic Analysis [M]. New York: Cambridge University Press, 2005

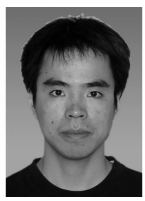
[19] Karp B, Ratnasamy S, Rhea S. Spurring adoption of DHTs with openhash, a public DHT service [C] //Proc of the 3rd Int Conf on Peer-to-Peer Systems. Berlin: Springer, 2004: 195-205

Tan Qingfeng, born in 1981. PhD candidate, assistant professor. Member of China Computer Federation. His main research interests include anonymous communication, covert communication and information warfare techniques.

Fang Binxing, born in 1960. Professor and PhD supervisor. Academician of Chinese Academy of Engineering. His current research interests include computer architecture, computer network and information security.

Shi Jinqiao, born in 1978. PhD, associate professor. Member of China Computer Federation. His research interests include network and information security, especially privacy enhancing techniques, data leakage detection and protection.

Xu Fanwen, born in 1989. Master. His main research interests include covert communication and anonymous communication.



Chen Xiaojun, born in 1979. PhD candidate, engineer. Member of China Computer Federation. His research interests include network security behavior analysis, network anomaly detection and data loss prevention.