文章编号:1001 - 9081(2007)07 - 1606 - 03

一种基于分组填充 Mix 策略的匿名通信机制

汤 念,王 雷,姚焯善,张大方,徐红云
(湖南大学 软件学院,长沙 410082)
(wanglei@hnu.cn)

摘 要:Mix 机制为实现匿名通信技术提供了可行的解决方案,但攻击者仍可以在一定程度上通过流量分析来获取通信关系。为了进一步提高 Mix 机制对抗信息量分析攻击的能力,提出了分组填充的思想,并基于分组填充 Mix 策略给出了一种新型匿名通信机制。算法分析和仿真实验表明,与传统 Mix 机制相比,新的基于分组填充的 Mix 机制,能在有效抵御信息量分析攻击的同时,进一步降低网络的额外带宽开销。

关键词:匿名通信;Mix 机制;分组填充;网络安全
中图分类号:TN915.04;TP393.08 **文献标志码**:A

Anonymous communication mechanism based on group padding MIX approach

TANG Nian, WANG Lei, YAO Zhuo-shan, ZHANG Da-fang, XU Hong-yun
(*College of Software, Hunan University, Changsha Hunan* 410082, *China*)

Abstract: The Mix mechanism has provided a feasible solution to the realization of anonymous communication. But to a certain degree, the attackers can still obtain the correspondence relations through traffic analysis. In order to further improve the ability of Mix mechanism to resist the traffic analysis of message volume, a kind of scheme named group padding was proposed, and based on which, a new anonymous correspondence mechanism was designed. Algorithm analysis and the simulation results show that, compared with the traditional Mix mechanism, the newly proposed Mix mechanism based on group padding scheme can resist traffic analysis of message volume effectively, and needs lower extra bandwidth at the same time.

Key words: anonymous communication; Mix approach; group padding; network security

0 引言

随着互联网应用的深入,对于通信私密性和匿名性的要求越来越引起人们的重视[1,2]。私密性研究的是通信内容的保密和安全问题,匿名性研究的是对通信关系的保密,也就是隐藏谁在和谁通信的问题,具体可以分为发送者匿名、接收者匿名、通信关系匿名三种。针对匿名性而提出的匿名通信技术在现实生活中起到了积极的作用[3,4]。

Chaum[5]于 1981 年提出了一种基于 Mix 策略的匿名通信机制,其核心思想是通过定义一个经过多个中间节点转发数据的多级目标路径(Mix),以实现通信的匿名。Mix 机制最初是用于电子邮件系统,以隐藏邮件用户与邮件服务器之间的通信。在 Chaum 之后,很多人对 Mix 做了改进,提出了基于 Mix 机制的新协议和系统,提高了匿名系统的安全性与应用范围。Goldschlag[6]和 Philippe[7]针对多个 Mix 节点合作方式提出了不同的匿名通信系统,它们都属于 Core Mix Net[6,7],即其中 Mix 网络中的节点为整个系统的核心,用户可通过选取网络中的某些 Mix 节点构成一条重路由路径以获得相应的匿名性需求。Rob 和 Michael 基于 Mix 机制分别研究了 P2P 系统中的匿名通信技术,提出了 P2P 系统中基于 Mix 机制的匿名通信系统[8,9]。在基于 P2P 的 Mix 匿名通信系统中,每个 Mix 节点既是 Mix 又是用户,它既可以充当重路由路径上的节点,为其他匿名用户转发信息,又可匿名地发送和接收信息。基于 Mix 的应用系统也有很多,比如 Lance Cottrell 等提出的 Mixmaster[10],以及在其之上 George Danezis 等提出的 Mixminion[11]。基于连接的应用系统包括 Morphmix[12,13],Real-time Mixes,Onion Rerouting,Tor[14]和 Crowds[15]等。

上述基于传统 Mix 机制的匿名通信技术,虽然为实现匿名通信提供了可行的解决方案,但攻击者仍可以在一定程度上通过流量分析来获取通信关系。当然,传统 Mix 机制可通过随机位串的填充将所有信息包填充成一样大,来抵御匿名系统中的信息量分析攻击,可是这样将导致带宽的浪费。为了进一步提高传统 Mix 机制对抗信息量分析攻击的能力,本文提出了分组填充 Mix 的思想,并在此基础上提出了一种基于分组填充 Mix 策略的匿名通信机制。新机制通过在 Mix 节点转发信息包之前进行分组填充,从而使得新机制能有效抵御信息量分析攻击,并相比原有填充节约了带宽。算法分析和仿真实验表明,与传统 Mix 机制相比,新的基于分组填充的 Mix 机制,具有更低的网络额外带宽开销,且能有效抵御信息量分析攻击。

1 基本概念

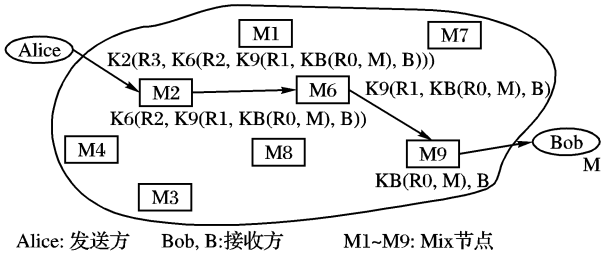
1.1 基于 Mix 策略的匿名通信机制

基于 Mix 策略的匿名通信机制是 Chaum 于 1981 年首次提出来的。该机制通过使用称为 Mix 的中继服务器来进行重路由,并使用 Mix 公钥加密邮件信息。其原理如图 1 所示。当 Alice 要与 Bob 进行匿名通信时,Alice 的访问程序将在 Mix 网络中选择某些 Mix 节点,如图 1 中的 M2、M6、M9,建立一条从发送方 Alice 到接收方 Bob 的通道。通道建立也就意味着,

收稿日期:2007 - 01 - 03;**修回日期**:2007 - 03 - 14。 **基金项目**:国家自然科学基金资助项目(60473031)。

作者简介:汤念(1984 -),女,湖南长沙人,硕士研究生,主要研究方向:计算机网络; 王雷(1973 -),男,湖南长沙人,副教授,博士,主要研究方向:计算机网络、机器学习、生物计算; 姚焯善(1983 -),男,广西桂平人,硕士研究生,主要研究方向:计算机网络; 张大方(1959 -),男,上海人,教授,博士生导师,主要研究方向:计算机可信网络; 徐红云(1959 -),女,湖南长沙人,教授,主要研究方向:网络信息安全、匿名通信。

Alice 要发送的数据将首先转发给 M2,接着由 M2 转发给 M6,M6 再转发给 M9,最后 M9 将数据传送给 Bob。发送者到接收者所经过的所有 Mix 节点构成了发送者的重路由路径,路径上的第一个节点 M2 知道信息的发送者(入口),路径上的最后一个节点 M9 知道信息的接收者(出口),所以,在这条路径上,只要有一定数量的 Mix 没被攻击者控制,则能保证发送者和接收者的不连接性,从而实现了匿名。



Alice: 发送方　Bob, B:接收方　M1~M9: Mix节点
R0, R1, R2, R3: 随机位串　M: Alice要发送给Bob的信息
K2, K6, K9, KB: 分别是M2, M6, M9, Bob的公钥

图 1　Mix 结构

为了使得 Alice 发送给 Bob 的数据内容以及他们之间的通信关系得到隐藏,在 Mix 网络中采用了公钥嵌套加解密数据的方法来实现。如图 1 中,Alice 要将数据 M 发送给 Bob,那么它将按照通道中 Mix 节点的逆序一层一层用它们的公钥嵌套加密构造信息包。最内层,将要发送的数据 M 前添加一个随机位串 R0,然后用接收方 Bob 的公钥加密——KB(R0,M),在其后添加接收方 Bob 的地址。再在其基础上添加随机位串 R1,然后用通道中最后一个 Mix 节点 M9 的公钥 K9 加密——K9(R1,KB(R0,M),B),依序使用通道中其他 Mix 节点的公钥加密添加了一个随机位串的内层信息包,故 Alice 最后要发送的信息包为 K2(R3,K6(R2,K9(R1,KB(R0,M),B)))。这样发送出去的信息包沿着通道传递,当 M2 接收到了这个包以后,用它的私钥解密所收到的信息包,得到 R3,K6(R2,K9(R1,KB(R0,M),B)),然后丢弃随机位串 R3,将剩下的数据包发给 M6;M6 也将用自己的私钥解密收到的信息包,并丢弃随机位串 R2,然后将剩下的信息包传给 M9,同理 M9 将 KB(R0,M)传给 Bob;最后 Bob 用自己的私钥解密,得到了 Alice 要发送给他的数据 M。

1.2　信息量分析攻击

信息量分析攻击是匿名系统常见的攻击方式之一。信息量分析攻击通过分析传送信息的长度,对手可以关联不同的客户端 - 服务器对。信息量分析攻击是对通信方式的研究,对象不是电文内容本身,而是它们的特点:谁和谁联系过、什么时候、电文的长度是多少、在多短的时间答复、答复有多长,这些就是信息量分析的问题,其答案可以揭示出大量的信息。

在 Mix 网络中传输的信息包中添加 Ri,其目的是用以消除攻击者验证两个加密信息包是否相同带来的威胁。即防止攻击者将从某个 Mix 节点发送出去的信息包,用该 Mix 的公钥重新加密之后,再与进入 Mix 的信息包进行比对,达到攻击目的。可是在一定程度上不能够抵御基于信息量分析的攻击。因为随机位串 Ri 相对于要发送的 M 来说是非常小的。

Mix 节点在传送信息包时可采取重新排序、延迟和填充手段使攻击者获取通信关系的概率更低,从而加大攻击者进行流量分析的难度。在抵御信息量分析攻击方面,显然,传统 Mix 机制可通过随机位串的填充将所有信息包填充成一样大,来抵御匿名系统中的信息量分析攻击,但是给系统带来了更多的开销。

2　基于分组填充 Mix 策略的匿名通信机制

Mix 节点输出报文的规则关系到 Mix 网络的匿名性能,目前关于 Mix 节点转发报文的规则主要有阈值(Threshold)Mix[16],缓冲池(Pool)Mix[10,17],停止转发(Stop-and-Go)Mix[7]等。在阈值 Mix 系统中,Mix 节点转发报文的规则是:事先设定一个阈值 n,只有当 Mix 收到的信息包数目达到阈值 n 时,才将 n 个信息包同时发送出去。

为了在有效保证 Mix 机制抵御信息量分析攻击能力的前提下,尽量降低不必要的通信开销。本节在阈值 Mix 系统中 Mix 节点转发报文的规则基础上,提出了一种基于分组填充 Mix 策略的新型匿名通信机制,其核心思想如下:事先设定一个阈值 n,当 Mix 收到的信息包数目达到阈值 n 时,Mix 先对 n 个信息包进行分组填充,然后打乱顺序将 n 个信息包同时发送出去。

基于上述分组填充 Mix 思想,综合考虑通信开销与匿名度之间的平衡关系,本节分别给出了两种分组填充算法如下:

分组填充算法一

Step 1: Mix 接收 n 个信息包;

Step 2: 对接收到的每一个信息包,用自己的公钥解密,丢弃随机位串;

Step 3: 把 n 个信息包随机打乱顺序;

Step 4: 这 n 个信息包组成集合 U,令 direction = UP(UP 表示后续包大小单调递增),first = 0

/ * direction 表示后续包大小的单调增减,first 表示每个分组的第一个包的序号 * /

Step 5: 依次取 U 中的包 i,令 G_i 表示包 i 的大小,$d = G_i - G_{i+1}$,MAX = 0

/ * MAX 表示一个组内包的大小的极大值 * /

Step 6: IF i 是最后一个包

IF 包 i 的大小 > MAX

MAX = 包 i 的大小

END IF

把 first 到该分组之间的分组全部填充为 MAX 大小

转 Step 7

END IF

Else

IF direction = UP 并且 d > 0　　//到达极大值,拐点

direction = DOWN, MAX = G_i

/ * DOWN 表示包的大小单调递减 * /

END IF

IF direction = DOWN 并且 d < 0　　//到达极小值,拐点

把 first 到该分组之间的分组全部填充为 MAX 大小

direction = UP, first = 0

END IF

转 Step 5

Step 7: 打乱 n 个包的顺序,发送 n 个包,算法结束

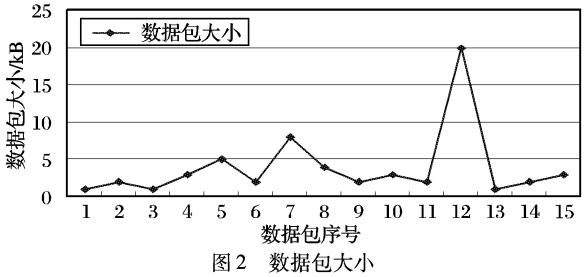


图 2　数据包大小

Mix 接收到 n 个数据包的一个典型例子如图 2 所示。图中包序列(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)的大小分别为(1,2,1,3,5,2,8,4,2,3,2,20,1,2,3)(单位 kB)。上面的算法是要求出每个拐点(极值点),把两个极小值之间的数据包分成一个组,这个组的包全部填充为组内的极值点的大小。在这个例子中,第 1 ~ 2 个包都填充为 2 kB,第 3 ~ 5 个都填充为 5 kB,6 ~ 8 填充为 8 kB,9 ~ 10 填充为 3 kB,11 ~ 12 填充为 20 kB,13 ~ 15 填充为 3 kB。假定在 1 s 内发完这些数据包,这种填充方法需要带宽为:2 × 2 + 5 × 3 + 8 × 3 + 3 × 2 + 20 × 2 + 3 ×

3 = 98 kBps。如果全部填充为最大值 20 kB，则需要带宽为 20 × 15 = 300 kBps 可见这种方法可以较好地减小带宽开销。

这种方法也有可以改进的地方，比如说第 3 个包，如果分在第一组，只需要填充成大小为 2 kB 的包，但是分在第二组，却要填充成大小为 5 kB 的包，所以可以把算法改进，当到达极小值时，计算该包的大小与前一个包的大小的差值以及该包的大小与后一个包的大小的差值，然后比较两个差值的大小，把这个包归入差值较小的分组。

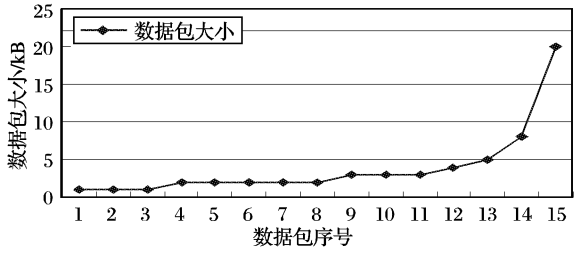


图 3 排序后再分组

基于上述分析，下面给出一种更能节约带宽，不过要增加额外的处理时间的算法二。在算法二中，首先把 n 个包按从小到大的顺序排列，如图 3 所示，然后把相邻的一些包分为 1 组，同时根据曲线的斜率决定分组情况。当斜率比较大时，比如图 3 中第 13 个包到第 14 个包之间出现了较陡的折线，这时候就可以把它们分别归入不同的分组。还要采取措施保证每个组里面包的数目 > 1，否则就相当于没填充，降低了匿名性。

分组填充算法二

Step 1: Mix 接收 n 个信息包

Step 2: 对接收到的每一个信息包，用自己的公钥解密，丢弃随机位串；

Step 3: 把 n 个信息包按包的大小从小到大进行排序；

Step 4: 每相邻的 k 个包分为一个组，填充为组内的包的最大值；如果最后只剩下一个包，把这个包加入到前一个组，重新把这个组的包填充为组内最大的包的大小

在这个算法中，k 取值越小，带宽越节约；k 的取值越大，匿名性越好。如果取 $k = 1$，就是每个包都不填充，都按原来的大小发送，这样是最节约带宽的，但是达不到理想的匿名效果。如果取 $k = n$，就相当于全部填充成该 n 个数据包中最大的值，这样就有比较高的匿名性。所以，k 的取值要在带宽和匿名度之间做一个权衡。

3 仿真结果

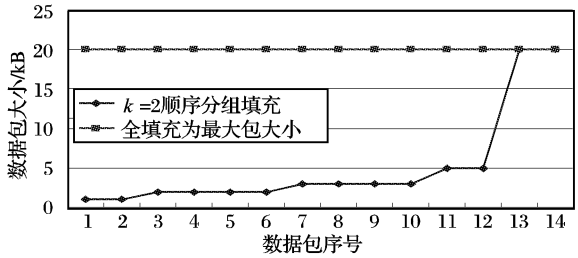


图 4 分组填充算法一和全部填充带宽使用情况比较

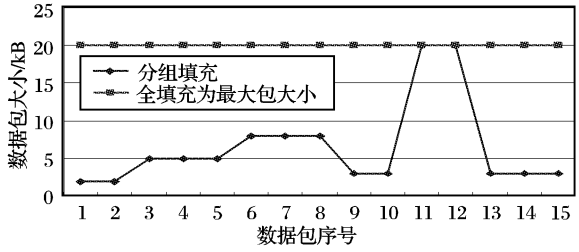


图 5 分组填充算法二和全部填充带宽使用情况比较

为了更好地描述本文提出的分组填充算法的优越性，下面分别对算法一和 $k = 2$ 时算法二与把所有发送给 Mix 的信息包填充成一样大小的算法所使用带宽使用情况进行比较。

仿真结果表明，本文提出的分组算法大大减小了带宽开销。如果一个由多个 Mix 节点组成的匿名通信网络中，每一个 Mix 节点都对接收到的信息包采取分组填充，那么，即使 $k = 2$，新机制还是可以有较高的匿名性，可以抵御比较网络中包大小的信息量分析攻击。

参考文献：

[1] 眭鸿飞，陈松乔，陈建二，等. 基于重路由匿名通信系统的负载分析[J]. 软件学报，2004，15(2)：278 - 285.

[2] 王伟平，陈建二，陈松乔，等. 匿名通信中短距离优先分组重路由方法的研究[J]. 软件学报，2004，15(4)：561 - 570.

[3] 高虎明，陈晓峰，王育民. 一个新的(t，N - 2)弹性的 Mix Net[J]. 计算机学报，2003，26(10)：1361 - 1365.

[4] 王继林，伍前红，陈德人，等. 匿名技术的研究进展[J]. 通信学报，2005，26(2)：112 - 118.

[5] CHAUM D. Untraceable electronic mail, return addresses, and digital pseudonyms[J]. Communications of the ACM, 1981, 24(2), 84 - 90.

[6] GOLDSCHLAG D, REED M, SYVERSON P. Onion routing for anonymous and private internet connections[J]. Communications of the ACM, 1999, 42(2): 39 - 41.

[7] KESDOGAN D, EGNER J, BüSCHKES R. Stop-and-go MIXes: providing probabilistic anonymity in an open system[C]// Proceedings of the Second International Workshop on Information Hiding, LNCS 1525. Heidelberg: Springer - Verlag, 1998: 83 - 98.

[8] SHERWOOD R, BHATTACHARJEE B, SRINIVASAN A. P5: a protocol for scalable anonymous communication[C]// Proceedings of the 2002 IEEE Symposium on Security and Privacy. Washington, DC, USA: IEEE Computer Society, 2002: 58 - 70.

[9] FREEDMAN M J, MORRIS R. Tarzan: a peer-to-peer anonymizing network layer[C]// Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002). Washington, DC: ACM Press, 2002: 193 - 206.

[10] MÖLLER U, COTTRELL L, PALFRADER P, *et al*. Mixmaster protocol - version 2[S]. 2003.

[11] DANEZIS G, DINGLEDINE R, MATHEWSON N. Mixminion: design of a type III anonymous remailer protocol[C]// Proceedings of the 2003 IEEE Symposium on Security and Privacy. Berkeley, CA, USA: IEEE Computer Society, 2003: 2 - 15.

[12] RENNHARD M, PLATTNER B. Introducing MorphMix: peer-to-peer based anonymous internet usage with collusion detection[C]// Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2002). New York, USA: ACM Press, 2002: 91 - 102.

[13] RENNHARD M, PLATTNER B. Practical anonymity for the masses with MorphMix[C]// Proceedings of Financial Cryptography (FC'04), LNCS 3110. Berlin: Springer - Verlag, 2004: 233 - 250.

[14] DINGLEDINE R, MATHEWSON N, SYVERSON P. Tor: the second-generation onion router[C]// Proceedings of the 13th USENIX Security Symposium. Mateo: USENIX Association, 2004: 303 - 320.

[15] REITER M, RUBIN A. Crowds: anonymity for web transactions [J]. ACM Transactions on Information and System Security, 1998, 1(1): 62 - 92.

[16] SERJANTOV A, DINGLEDINE R, SYVERSON P. From a trickle to a flood: active attacks on several mix types[C]// Proceedings of Information Hiding Workshop (IH 2002), LNCS 2578. Noordwijkerhout, The Netherlands: Springer - Verlag, 2002: 36 - 52.

[17] SERJANTOV A, NEWMAN R E. On the anonymity of timed pool mixes[C]// Proceedings of the Workshop on Privacy and Anonymity Issues in Networked and Distributed Systems. Athens, Greece: Kluwer, 2003: 427 - 434.

[18] DíAZ C. Anonymity and privacy in electronic services [D]. Katholieke Universiteit Leuven, 2005.