

基于源-目的端 ISP 包标记方案研究

叶小涛¹, 吕爱丽¹, 赵林²

(1. 河南理工大学 现代教育技术中心, 河南 焦作 454010;

2. 东北大学 东软信息学院, 辽宁 大连 116023)

摘要:提出了一个基于源-目的端 ISP 包标记方案。此方案不再用来重构攻击路径,而主要用于刻画 DDoS 攻击流特征。这些特征对于受害者过滤攻击非常有效。在过滤方面,提出了一个比率控制方案,通过限制攻击流并保持合法数据流不受影响来有效保护受害者。在经济方面 ISP 能提供更好的安全措施作为对客户的增值服务,因此也就更有积极性来部署。

关键词:分布式拒绝服务攻击;包标记;比率控制

中图分类号:TP309

文献标识码:A

文章编号:1673-629X(2009)04-0098-03

Research of Packet Marking Scheme Based on Source and Destination - End ISP

YE Xiao-tao¹, LÜ Ai-li¹, ZHAO Lin²

(1. Modern Education Technology Center, Henan Polytechnic University, Jiaozuo 454010, China;

2. Neusoft Institute of Information, Northeastern University, Dalian 116023, China)

Abstract: In this paper, propose new packet marking models. It's not used for reconstructing the attack path, but characterizing DDoS attack streams. Such common characterization can be used to make filtering by the victim more effective. In terms of filtering, propose a rate control scheme that protects destination domains by limiting the amount of traffic during an attack, while leaving a large percentage of legitimate traffic unaffected. On economic front, it enable providers to offer enhanced security protection against such attacks as a value-added service to their customers, and hence offer positive incentives for them to deploy the proposed models.

Key words: DDoS; packet marking; rate control

0 引言

目前 Internet 安全的威胁主要来自于黑客入侵攻击、计算机病毒和拒绝服务(DoS)攻击三个方面。拒绝服务攻击以及分布式拒绝服务攻击(DDoS)成为目前最严重的安全问题之一^[1],其主要原因是还未归纳出用于抵御和过滤攻击数据流的普遍特性。攻击者经常使用伪造的 IP 地址,使得在如今的网络基础结构上识别和屏蔽这些攻击非常困难。

由于 DDoS 攻击的危害性和紧迫性,许多学者都在研究防御措施,通用的措施是追踪机制。这种机制有个共同的前提,就是必须重构完整的攻击路径,而且大多数机制设定受害者只能发起追踪或接收追踪信息,而不能有效地参与对数据包的过滤。然而重构确

切的攻击路径往往是没有必要的,因为需要得到的只是攻击数据包包含的特定路径信息。另外,管理者需要手工增加过滤规则和访问控制列表,这样导致反应时间滞后而且缺乏对攻击模式的适应性。文中提出一种基于源-目的端 ISP 包标记模型,用于给受害者 ISP 提供稳定而安全的关于接受数据包的信息,受害者可以根据攻击数据包的路径信息自己进行过滤。

1 源-目的端 ISP 包标记策略

此模型根据数据包通过的路由器路径嵌入一个标识,受害者只需要分离出首个恶意数据包,便能够过滤出后续所有有相同标记的数据包。这些标记可以用于对攻击或可疑数据流的检测而不受攻击者伪造数据流影响,并可以提供一个通用的特性用于过滤。数据包经过同样的路径时会有同样的标记值,而且标记是确定性的^[2],每个数据包都有一个路径标识,不需要像概率包标记时需要受害者收集大量的数据包,即使收到

收稿日期:2008-07-14

基金项目:河南省自然科学基金(2003520257)

作者简介:叶小涛(1980-),男,河南焦作人,助教,研究方向为计算机网络安全。

一个攻击数据包也可以开始过滤。不管是在路由器作标记还是受害者检测或过滤过程中,此方案负载较低。

另外提出一个比率控制方案,该方案可以在攻击期间限制攻击流而保持合法流不受影响,这样 ISP 就可以作为对客户的增值服务而更好地保护客户,从而在攻击过程中提高合法用户的可用带宽。

1.1 IP 地址的 Hash 函数标记

Burch 和 Cheswick 提出了在数据包的头部写入路由器地址信息的包标记方案^[3]。由于数据包在途中经分段(fragment)处理的情况是很少出现的(不超过 0.25%)^[4],因此 IP 报头中的标示域(Identification field)也很少使用,于是, Savage 等人建议将路径信息嵌入到 16 位的标示域中,如图 1 所示。

最简单的标记方案是 n -bit 方案,就是将路由器 IP 地址的最后 n 位在 IP 地址的 16 位标示域进行标注。确定标示域每位需要标记的内容,首先把标示域分成块 $[16/n]$,然后根据路径中的路由器顺序标注其中某一块。

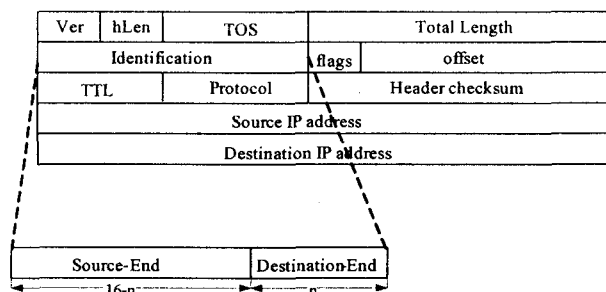


图 1 IP 报头标示域作为标记域

因为许多 ISP 会倾向于给路由器分配最后字段为 0 的 IP 地址,那么通过此类路由器的数据包其被标记内容也会为 0,所以直接将 IP 地址的最后 n 位作为标记内容进行标记会造成受害者收到的数据包经过的路径就不容易区分。为解决这个问题, A. Yaar 提出路由器使用 IP 地址 Hash 值的最后 n 位来作为标记内容^[5,6],本模型选定使用 IP 地址的 MD5^[7] Hash 值的最后两位作为标记内容。

1.2 基于源-目的端的标记策略

文中的源-目的端 ISP 标记模型中,标记内容分为两部分,如图 1 所示:前一部分是在源端与 ISP 相连的边界路由器上实施,这部分标记是确定的。为了实现统一的标记值,对与源域相连的路由器,取其 IP 地址的 MD5 Hash 值的最后两位。后一部分是在目的端 ISP 与外网相连的边界路由器上实施的,这些边界路由器利用 16 位 IP 标示域中的 n ($n < 16$) 位标记一

个有别于其它路由器的值,剩下的 $16 - n$ 位供源端 ISP 使用,如图 2 所示。

前 $16 - n$ 位的标记内容跟标记路由器的 IP 地址有关,而后 n 位仅仅是为了区分目的端 ISP 与外网相连的边界路由器,并不使用 Hash 函数。 n 的值取决于目的端 ISP 与外网相连的边界路由器数,路由器越多, n 越大。比如,如果目的端 ISP 连接到 Internet 需要 4 个路由器,那么 2 位就足够来区分这些路由器了;如果此类路由器只有 1 个,则 16 位全部用于源端的标记。性能分析见下文 3.2 节。

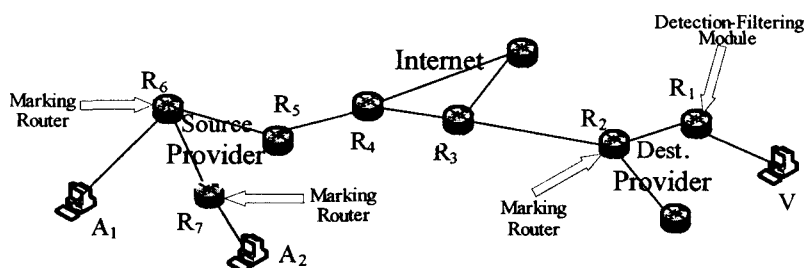


图 2 标记模型

1.3 部分部署时处理带虚假标记的攻击

一般情况下所有从源域发出的合法包都可以通过目的端 ISP 的一个或多个边界路由器到达目的域^[8],然而,在带虚假标记攻击时,目的端 ISP 会从大部分边界路由器收到相同的标记值。如果 16 位 IP 标示域全部都在源端 ISP 标记,即 n 为 0,那么在模型部分部署的环境里不能处理这些带虚假标记的攻击。因为模型只有部分部署,所以对于源端 ISP 没有实施标记的数据包,攻击者可以命令在和 ISP 相连域中的傀儡机使用其它源域的合法值来标记。如果检测到攻击后实施过滤,那么被引用的源端发出的合法通信会受影响。而如果在目的端 ISP 实施标记,即 $n \neq 0$,就可以通过不同的边界路由器进入到目的端,目的端 ISP 标记内容不同,因此标记信息被伪造的可能性大大降低。除非这些攻击流伪造了与受攻击源端相同的标记,而且和此源端合法数据流通过相同的边界路由器才能有相同的合法标记值。

2 比率控制

ISP 可以在与目的域相连的边界路由器上根据数据包标记内容进行检测和过滤。由于检测系统会出现误报的情况,因此直接丢弃包含攻击流特征的数据包会对合法数据流造成影响。文中提出一种比率控制的方法,该方法保证合法数据流量在一定的比率,而不是把所有认定为攻击流的所有数据包完全丢弃,从而保证合法的数据流不受影响。

假设 l_i 为发生攻击之前标记为 i 的概率, I 为标记值的集合。现在假定发生 DDos 攻击, 取 A 为确定为攻击流的标记值集合, L 为非攻击流标记值集合, 因此 $I = A \cup L$ 。如果 ISP 边界路由器到受害者的带宽为 C , 那么 ISP 可以分配 C_{legit} ($C_{\text{legit}} = (\sum_{j \in L} l_j / \sum_{i \in I} l_i) C$) 为包含集合 L 中标记值的合法数据流带宽, 并且保证它的平均值在攻击前后保持一致。而分配 C_{attack} ($C_{\text{attack}} = (\sum_{j \in A} l_j / \sum_{i \in I} l_i) C$) 为被确定为攻击流的带宽。

对于标记值为 i ($i \in A$) 的攻击流带宽 a_i , 如果满足 $\sum_{i \in A} (l_i + a_i) > C_{\text{attack}}$, 则攻击流带宽被限制为 C_{attack} , 超出的部分 $\sum_{i \in A} (l_i + a_i) - C_{\text{attack}}$ 被丢弃, 被丢弃

比率为 $1 - \frac{\sum_{j \in A} l_j}{\sum_{i \in I} (l_i + a_i)} \frac{C}{\sum_{i \in A} l_i}$ 。对于每个 j ($j \in A$) 可分配带宽为 $C_j = (l_j / \sum_{i \in I} l_i) C$ 。这样对于属于集合 A 的

每个标记 j , 弃包比率为 $1 - \frac{l_j}{l_j + a_j} \frac{C}{\sum_{i \in I} l_i}$ 。因此对于标

记值为 j 的数据包, 其弃包比率是一个和攻击流有关的函数。与包含攻击标记的直接过滤机制相比, 这种动态的过滤机制允许部分误标记为攻击流的合法数据包进入目的域。

3 实验结果

文中利用网络拓扑结构评估在模拟 DDos 攻击时对合法数据流和攻击数据流的区别效果。另外, 研究了攻击者数目、在目的端标记需要的字节数以及 ISP 实施此策略的程度对区别效果的影响。文中对攻击检测过程不做讨论, 假定模型具有理想的检测效果, 即 100% 的检测率和 0% 的误报率。

3.1 实验介绍

实验利用的拓扑结构为 Burch/Cheswick 因特网地图, 假设受到 DDos 攻击的受害者为根节点, 攻击者和合法用户为叶子节点。在实验中, 选择 5000 个节点来作为合法用户, 它们相互发送 10 个数据包, 还有一部分节点作为攻击者, 它们相互发送 100 个数据包, 两组节点集合不交叉。实验关心的既不是绝对的流量, 也不是攻击流和合法流的相对比率, 而是接受的流量比率。实验中除非特殊说明, 第一次标记都发生在 3 跳以外的路由器上, 而每个实验结果是在同一个参数下 5 次实验的平均值。

3.2 性能分析

实验的性能用接收率差额 (Acceptance Ratio Gap)

来度量, 它是用户包接收率减去攻击包接收率之值。本实验假设模型 100% 被部署, 从而攻击包接收率为 0, 接收率差额就和用户包接收率相同。从图 3 可以看出在 2000 攻击者时用户包接收率和接收率差额为 70%, 就是说有 70% 的合法用户数据包不被过滤系统影响。随着攻击者数目增加, 合法数据包和攻击数据包的冲突增加, 从而导致用户包接收率减少。

图 3 表示不同目的端字节数的源-目的端 ISP 标记模型的情况。在目的端分配更多的字节会带来如下问题: 第一, 减少了在源端标记的字节, 从而减少了在源端差异性; 第二, 增加在目的端 ISP 的标注字节也增加了差异性。两种影响是显著的, 因此, 单纯增加或减少差异性取决于字节数。尤其是在目的端使用了 2-5 位来标记, 留下 11-14 位在源端 ISP 标记, 结果要优于 16 位全部用于源端 ISP 来标注。相反的, 仅 4-7 位用于源端 ISP 标记效果也同样不佳。这个平衡点跟源端到目的端的距离和拓扑结构有关。

如果环境里有部分 ISP 没有部署本模型, 那么它们的边界路由器就不会标记通过的路由器, 这时攻击包接收率不再为 0%。随着部署本模型的比率的下降, 性能也相应下降。

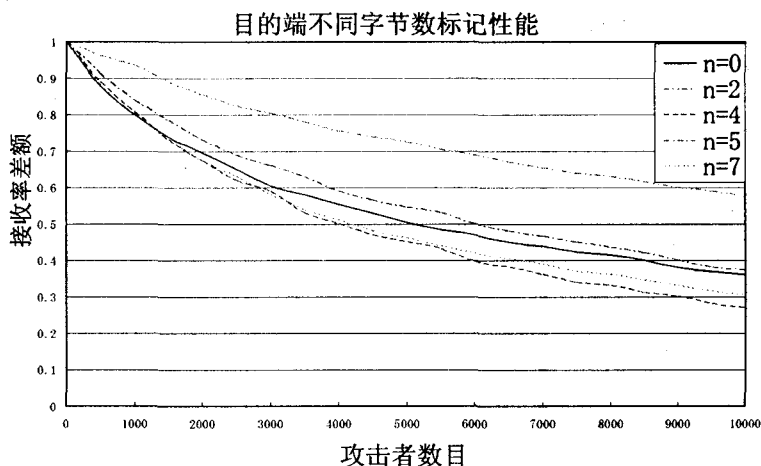


图 3 源-目的端标记模型性能

4 结束语

本模型采用在源端和目的端分别来做标记方法, 源端标记内容包括 MD5 加密后的 IP 地址 Hash 值的最后 n 位, 这样减少了相同路径标识的可能, 目的端标记记录了数据流到达时经过的边界路由器, 这样在有部分未部署的情况下数据流被伪造的可能性大大降低。另外, 采用频率控制方案, 尽可能减少了对于检测系统误报情况下对于合法数据流的影响。

本模型最重要的特性是在经济方面提高 ISP 对部

(下转第 104 页)

- Center, 1998.
- [3] 周明中, 龚 俭. 数据流管理系统综述[J]. 计算机工程, 2006, 32(2): 10-12.
 - [4] Post G V. Database Management Systems: Designing and Building Business Applications[M]. 3rd Edition. 冯建华, 刘旭辉, 周维续译. 北京, 机械工业出版社, 2006.
 - [5] 金澈清, 钱卫宁, 周傲英. 流数据分析与管理综述[J]. 软件学报, 2004, 15(8): 1172-1181.
 - [6] Terry D, Goldberg D, Nichols D, et al. Continuous queries over append-only databases[C]//In Proc. of the 1992 ACM SIGMOD Intl. Conf. on Management of Data. San Diego, California: [s. n.], 1992: 321-330.
 - [7] Arasu A, Babu S, Widom J. An Abstract Semantics and Concrete Language for Continuous Queries over Streams and Relations[R]. US: Stanford University, 2002.
 - [8] Sullivan M, Heybey A. Tribeca: A System for Managing Large Databases of Network Traffic[C]//In Proc. USENIX Annual Technical Conf. New Orleans, Louisiana: [s. n.], 1998.
 - [9] Carney D, Cetintemel U, Cherniack M, et al. Monitoring streams - A New Class of Data Management Applications [C]//In Proc. Int. Conf. on Very Large Data Bases. Hong Kong: [s. n.], 2002: 215-226.
 - [10] 萨师焯, 王 珊. 数据库系统概论[M]. 北京, 高等教育出版社, 2000.
 - [11] Golab L, Bijay K. On Concurrency Control in Sliding Window Queries over Data Streams[C]//In Proceeding 10th International Conference on Extending Database Technology. Munich, Germany: [s. n.], 2006: 608-626.
 - [12] Babcock B, Data M, Motwan I R. Sampling from a moving window over streaming data[C]//In ACM-SIAM Symposium on Discrete Algorithms. San Francisco, CA, USA: [s. n.], 2002.
 - [13] 葛君伟, 公丕强, 刘兆宏. 一种存储和索引历史数据流数据的方法[J]. 计算机应用研究, 2007, 24(6): 104-106.
 - [14] Giannella C, Han J, Pei J, et al. Mining Frequent Patterns in Data Streams at Multiple Time Granularities[C]//In Kargupta et al. Data Mining: Next Generation Challenges and Future Directions. [s. l.]: MIT/AAAI Press, 2004.

(上接第 97 页)

参考文献:

- [1] 梁路洪, 艾海舟, 徐光祜, 等. 人脸检测研究综述[J]. 计算机学报, 2002, 25(5): 449-458.
- [2] Hafed Z M, Levine M D. Face recognition using the discrete cosine transform[J]. International Journal of Computer Vision, 2001, 43(3): 167-188.
- [3] 于威威, 滕晓龙, 刘重庆. 复杂背景下人眼定位及人脸检测[J]. 计算机仿真, 2004, 21(12): 185-188.
- [4] 陶 亮, 庄镇泉. 复杂背景下人眼自动定位[J]. 计算机辅助设计与图形学学报, 2003, 15(1): 38-42.
- [5] 冯建强, 刘文波, 于盛林. 基于灰度积分投影的人眼定位[J]. 计算机仿真, 2005, 22(4): 75-77.
- [6] 吕东辉, 王 滨. YCbCr 空间中一种基于贝叶斯判决的皮肤检测方法[J]. 中国图象图形学报, 2006, 11(1): 47-52.
- [7] Tao Liang, Zhuang Zhen-quan. An effective approach for frontal face verification[J]. Journal of Image and Graphics, 2003, 8(8): 860-865.

(上接第 100 页)

署的兴趣, 可以使用防御模型作为增值服务来为用户提供更好的服务, 从而增加收入项目, 因此 ISP 也就有相当大的兴趣来部署。另外, 所有必要的防御措施都由最后的 ISP 来管理, 它是防御模型的受益者。

在以后的工作中, 比率控制方法仍然是一个重点, 在网络拥塞的节点可以考虑对流量进行一定的疏导, 保证正常流量尽可能不受到攻击的干扰。另外在 IPv6 的情况下, 20 位标记位性能的提高也有待验证。

参考文献:

- [1] Schneier B. Secrets and Lies: Digital Security in a Networked World[M]. New York: John Wiley & Sons, 2000.
- [2] Belenky A, Ansari N. On deterministic packet marking[J]. Computer Networks, 2007, 51: 2677-2700.
- [3] Burch H, Cheswick H. Tracing anonymous packets to their approximate source[C]//Proc. USENIX LISA Conf. New Orleans, LA: [s. n.], 2000: 319-327.
- [4] Stoica I, Zhang H. Providing Guaranteed Services Without Per Flow Management [C]//Proc. the 1999 ACM SIGCOMM Conf. Boston, MA: [s. n.], 1999: 81-94.
- [5] Yaar A, Perrig A, Song D. Pi: a path identification mechanism to defend against DDoS attacks[C]//Proceedings of the IEEE Symposium on Security and Privacy. Berkeley: IEEE Press, 2003: 93-107.
- [6] Yaar A, Perrig A, Song Dawn. StackPi: new packet marking and filtering mechanism for DDoS and IP spoofing defense [R]. US: Carnegie Mellon University, 2003.
- [7] Rivest R L. The MD5 message digest algorithm[S]. RFC 1321, Internet Activities Board, Internet Privacy Task Force, 1992.
- [8] 孙知信, 李清东. 基于源目的 IP 地址对数据库的防范 DDoS 攻击策略[J]. Journal of Software, 2007, 18(10): 2613-2623.