文章编号:1001 - 9081(2008)07 - 1812 - 04

P2P 文件共享网中病毒传播建模

冯朝胜[1,3],秦志光[1],劳伦斯·库珀特[2],罗瑞莎·托卡库克[2]
(1. 电子科技大学 计算机科学与工程学院,成都 610054; 2. 伦敦大学玛丽皇后学院 电子工程系,伦敦 E1 4NS, 英国;
3. 四川师范大学 计算机科学学院,成都 610066)
(csfenggy@126.com)

摘 要:鉴于病毒在 P2P 文件共享网中的巨大危害性,对 P2P 文件共享网和 P2P 病毒的传播特点进行了深入的分析,并在此基础上提出了 P2P 病毒的传播模型。为了考查模型中各参数对病毒传播的影响,使用了专门的数字分析软件进行了大规模仿真实验。不同参数对病毒传播有着不同影响这一仿真实验结果表明,可以通过控制那些影响大的参数来抑制病毒传播。

关键词:P2P 网络;文件共享;病毒;建模;仿真
中图分类号: TP393.08 **文献标志码**:A

Modeling of virus propagation in peer-to-peer file-sharing networks

FENG Chao-sheng[1,3], QIN Zhi-guang[1], Laurence Cuthbet[2], Laurissa Tokarchuk[2]
(1. *School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan* 610054, *China*;
2. *Department of the Electronic Engineering, Queen Mary, University of London, London E*1 4*NS, UK*;
3. *School of Computer Science, Sichuan Normal University, Chengdu Sichuan* 610066, *China*)

Abstract: To counter the attacks of virus in P2P file-sharing networks, the model of virus propagation in P2P networks was proposed based on deep analysis on the features of file sharing and virus propagation. In order to examine the effect of different parameters in this model, large scale simulating experiments were carried out. The fact that different parameters have varied effect on virus propagation shows that virus propagation can be throttled by controlling those key parameters.

Key words: P2P networks; file sharing; viruses; propagating model; simulation

如今,数以百万计的因特网用户通常都使用大规模文件共享 P2P 网络(如 eDoney、BitTorrent 等)来上传和下载文件。据统计,仅 eDonkey2000 网络在任意时间都有 200 万用户同时在线[1],而使用过 BitTorrent 网络用户早已超过千万,成为最受欢迎的文件传输网络[2]。

伴随着 P2P 文件共享网络的迅猛发展,该网络的安全问题日益突出,P2P 病毒就是一个很严重的安全威胁。造成 P2P 文件共享网络很适合病毒攻击和传播的主要因素有两个:一是病毒可以被置入到可执行文件中。当置入有病毒的文件被其他用户下载后,用户一旦执行了该文件就会被感染,并且这个感染主机会成为新的感染源。二是许多病毒文件通常都冠以最受欢迎文件的文件名,这无疑大大增加了感染的概率[3]。

本论文对 P2P 文件共享网上病毒的传播行为进行了深入研究,目的是构造出能较准确反映和预测 P2P 病毒传播行为和模式的数学模型,为将来提出的病毒遏制和免疫方法的有效性提供验证方法。

1 相关研究工作

1.1 相关研究工作及进展

1926 年 McKendrick 在他的学术论文中率先将数学引入生物学中帮助进行疾病传播建模,开创了数学流行病学[4]。鉴于计算机病毒和生物病毒的相似性,研究人员将数学流行病学引入到计算机病毒的传播建模中,结果也取得了很好效果。在 20 个世纪 90 年代初 Kephart 和 White 率先将经典流行病模型引入计算机病毒传播建模中[5]。最近,一些研究人员利用数学流行病学为因特网上的"红色代码"病毒[6]和电子邮件病毒[7]建立了传播模型,较准确地预测了这些病毒的传播趋势和行为。2004 年 Qiu 和 Srikant 建立了 BT 网络的性能模型[8]。2005 年 Dumitriu 等人[9]对感染文件在 P2P 网络上的传播进行了建模。R. W. Thommes 和 M. J. Coates[10]对 P2P 文件共享网上的病毒传播和感染文件传播分别进行了建模。D. Stutzbach 等人[11]通过实验的方法分析了 P2P 文件共享网的拓扑结构,指出 P2P 文件共享网呈现出"小世界"的特征。2006 年,Guanling Chen 等人[3]对非扫描型 P2P 蠕虫进行了仿真分析,然而,他们并没有给出 P2P 病毒传播的数学模型;Jie Ma 等人[12]利用数字模拟的方法分析了 P2P 系统参数对被动式 P2P 蠕虫传播的影响,他们考查的重点是蠕虫而不是 P2P 病毒(这里指寄宿型病毒)。

1.2 P2P 文件共享网络

在像 BitTorrent 和 eDonkey2000 这样的网络中,每个用户都有一个共享文件夹,用户将所有可共享的文件都放到共享文件夹以便其他用户共享,网络中的任何用户都可以从其他任意一个用户的共享文件夹中下载文件。当用户想要下载某

收稿日期:2008 - 01 - 28;**修回日期**:2008 - 03 - 24。 **基金项目**:国家自然科学基金资助项目(60473090),国家自然科学基金与英国皇家科学会合作项目(60711130232);四川师范大学重点项目(07ZD018)。

作者简介:冯朝胜(1971 -),男,四川广元人,讲师,博士研究生,主要研究方向:网络与信息安全; 秦志光(1956 -),男,四川荣昌人,教授,博士生导师,CCF 会员,主要研究方向:网络与信息安全; 劳伦斯·库珀特,男(英国人),教授,博士,主要研究方向:通信与智能系统;罗瑞莎·托卡库克,女(英国人),讲师,博士,主要研究方向:人工智能、模式识别。

个文件时,他会发出搜索文件请求。在 BitTorrent 和 eDonkey 中通过 Tracker 服务器来处理这个请求,而在 Gnutella 中,通过"洪泛法"搜索文件。无论哪种 P2P 文件共享网络,请求文件用户最终都会收到与请求相匹配的文件列表。获取了文件列表后,用户可以从中列表中选择一个或多个主机来下载该文件。文件下载后被放在共享文件,可被网络中其他主机下载[10]。

P2P 文件共享网络中文件一经下载马上就可共享这一特点给用户共享文件带来了极大方便,同时也给病毒的传播留下了可乘之机,一些利用这一特点进行传播的病毒已经出现,像 Achar[13] 和 Gotorm[14] 这样的病毒,在其依附的文件被执行时将在共享文件夹中生成固定个数的感染文件,这些感染文件在所有被感染的主机上有着相同的文件名称。攻击力更强的病毒像 Bare[15] 和 Krepper[16],则从一个很大的命名域中选取名称。

2　P2P 病毒传播建模

2.1　建模参数和假设

建模的目的是为了预测 P2P 病毒在 P2P 网络中的传播趋势和行为。在模型中,主机的状态分成三种:易感的、暴露的和感染的。为了简化建模,作了如下假设:

1)网络中在线的用户数量没有发生变化。

2)文件被下载后被放到共享文件夹。

3)主机状态转移都在一个时间单元(Time Unit)内完成。

4)一台主机一旦被感染,将在共享文件夹中生成 c 个文件。所有的感染主机共享同样 c 个文件名称。

5)建模时考虑的文件都是可执行文件,包括被压缩的可执行文件,不能包含病毒的文件如媒体文件不被考虑。

为了便于下面的病毒建模分析,将建模时要用到的参数列举如下。

$N(t)$:t 个时间单元后网络中主机台数。在本文中这个值不随 t 变化。$N(0) = 50\,100$。

$S(t)$:t 个时间单元后易感主机数。$S(0) = 50\,000$。

$I(t)$:t 个时间单元后感染主机数。$I(0) = 50$。

$E(t)$:t 个时间单元后暴露主机数。$R(0) = 50$。

$K(t)$:t 个时间单元后感染文件数。$K(0) = 500$。

$M(t)$:t 个时间单元后未感染文件数。$M(0) = 100\,200$。

$h(t)$:t 个时间单元后下载感染文件的概率为 $h(t) = \dfrac{K(t)}{M(t) + K(t)}$。

λ_s:每个时间单元内每个主机下载文件的平均个数。$\lambda_s = 0.01$。

λ_e:每个时间单元内执行感染文件的暴露主机数。$\lambda_e = 0.01$。

λ_r:每个时间单元内恢复为易感状态的感染主机比例。$\lambda_r = 0.001$。

p_{ei}:每个时间单元内暴露主机成功执行感染文件的概率。$p_{ei} = 0.5$。

c:执行了下载的感染文件后在共享文件中增加的感染文件数。$c = 10$。

2.2　P2P 病毒传播状态机分析

根据 P2P 中计算机病毒传播的特点,为其建立的病毒传播状态机如图 1 所示。

2.2.1　主机所处状态说明

易感染的(S):当 P2P 中的主机有下载染毒的风险时,该主机就处于容易感染病毒状态;

已暴露的(E):当主机上的 P2P 共享文件夹中至少拥有一个染毒文件(下载而来,还没有执行)时,该主机就处于已暴露状态。

已感染的(I):当暴露主机执行了下载的染毒文件后,染毒文件的数量就会变成 c 个(假设),此时主机处于已感染状态。

2.2.2　主机状态转移说明

$S \to E$:当用户下载了染毒文件,主机就由易感染状态转化到已暴露状态。

$E \to S$:当暴露主机执行下载的染毒文件时,如果下载的染毒文件都被杀毒软件清除,那么主机就由已暴露状态恢复到易感染状态。

$E \to I$:当暴露主机执行下载的染毒文件时,如果下载的染毒文件中只要有一个成功执行,那么主机就由已暴露状态进入到已感染状态。

$I \to S$:当用户发现自己的主机已被感染并采取措施将共享文件夹中所有的染毒文件删除后,主机就由已感染状态回到易感染状态。

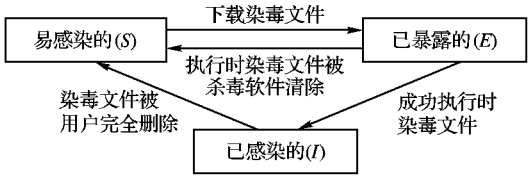

图 1　P2P 网络中病毒传播状态机

2.3　病毒传播模型分析

2.3.1　易感主机的变化率

t 个时间单元后易感主机的数量为 $S(t)$,每台易感主机下载文件成为暴露主机的概率为 $\lambda_s h(t)$,因而单位时间单元内有 $\lambda_s S(t)h(t)$ 台主机成为暴露主机。由状态机知,有两种状态可能转变成易感状态。一种是感染状态,另一种是暴露状态。显然,在单位时间单元内有 $\lambda_r I(t)$ 台感染主机转变成易感主机。由于单位时间单元内共执行 $\lambda_e E(t)$ 个感染文件,而单个感染文件执行失败率为 $(1 - p_{ei})$,故单位时间单元内转变为易感主机的暴露主机数为 $\lambda_e E(t)(1 - p_{ei})$。根据以上分析,得到反映易感主机变化率的微分方程为:

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = -\lambda_s S(t)h(t) + \lambda_r I(t) + \lambda_e E(t)(1 - p_{ei}) \quad (1)$$

2.3.2　暴露主机的变化率

根据易感主机变化率的分析知道,t 时刻单位时间单元内增加的暴露主机数为 $\lambda_s S(t)h(t)$,而单位时间单元内有 $\lambda_e E(t)$ 台主机执行染毒文件,无论成功与否,它们的状态都要发生转变,所以暴露主机的变化率为:

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = \lambda_s S(t)h(t) - \lambda_e E(t) \quad (2)$$

2.3.3　感染主机的变化率

$\lambda_e E(t)$ 台暴露主机中,成功执行染毒文件的概率为 p_{ei},所以有 $\lambda_e E(t)p_{ei}$ 台主机转变成了已感染主机,与此同时,有 $\lambda_r I(t)$ 台易感主机转变成已感染主机,故易感主机的变化率为:

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \lambda_e E(t)p_{ei} - \lambda_r I(t) \quad (3)$$

2.3.4 染毒共享文件的变化率

由于染毒文件的比率对病毒传播有重要影响，所以有必要研究下染毒文件的变化率。在 t 时刻，一台主机下载感染文件的数量为 $\lambda_s h(t)$ 个，$S(t)$ 台易感主机下载的染毒文件数为 $\lambda_s S(t)h(t)$ 个，成功执行染毒文件的暴露主机使得增加的感染文件数量为 $\lambda_e E(t)p_{ei}(c-1)$，而执行时被杀毒软件清除的染毒文件的数量是 $\lambda_e E(t)(1-p_{ei})$，显然被用户清除的文件数为 $\lambda_r I(t)c$，所以感染文件的变化率为：

$$\frac{dK(t)}{dt} = \lambda_s S(t)h(t) + \lambda_e E(t)p_{ei}(c-1) - \lambda_e E(t)(1-p_{ei}) - \lambda_r I(t)c \quad (4)$$

2.3.5 正常共享文件的变化率

一台易感主机下载感染文件的概率是 $h(t)$，那么下载正常文件的概率就是 $1-h(t)$。P2P 网络中每台主机下载的文件数为 $\lambda_s h(t)$，$N(t)$ 台主机共下载 $\lambda_s N(t)(1-h(t))$ 个正常文件。于是，正常文件的变化率为：

$$\frac{dM(t)}{dt} = \lambda_s N(t)(1-h(t)) \quad (5)$$

2.4 传播模型

根据以上分析，可以到如下病毒传播模型。

$$\frac{dS(t)}{dt} = -\lambda_s S(t)h(t) + \lambda_r I(t) + \lambda_e E(t)(1-p_{ei}) \quad (6)$$

$$\frac{dE(t)}{dt} = \lambda_s S(t)h(t) - \lambda_e E(t) \quad (7)$$

$$\frac{dI(t)}{dt} = \lambda_e E(t)p_{ei} - \lambda_r I(t) \quad (8)$$

$$\frac{dK(t)}{dt} = \lambda_s S(t)h(t) + \lambda_e E(t)p_{ei}(c-1) - \lambda_e E(t)(1-p_{ei}) - \lambda_r I(t)c \quad (9)$$

$$\frac{dM(t)}{dt} = \lambda_s N(t)(1-h(t)) \quad (10)$$

$$N(t) = S(t) + E(t) + I(t) \quad (11)$$

其中 $h(t) = \frac{K(t)}{M(t)+K(t)}$。

3 仿真验证及分析

3.1 实验说明

流行的 P2P 文件共享网络通常由上百万个节点构成，在这样的网络上直接验证提出的病毒传播模型是否有效以及模型中各参数对病毒传播的影响是不现实的，因此只有通过仿真实验来做到。通过 Matlab 中的 Simulink 来仿真验证了上面模型，并且基于该模型得到了病毒传播曲线。为了考查模型中各参数对病毒传播的影响，将同一参数不同取值对应的传播曲线在同一个图中做对比。因篇幅所限，下面仅展示部分实验结果，在不作特别说明的情况下，实验中变量的初值和参数的值就是 2.1 节中的取值。

3.2 实验结果和分析

图 2 考查了感染初值对 P2P 病毒传播的影响。图 2 清楚地表明在病毒传播起始阶段，被感染的节点越多，感染的高峰到来的越快，最终被感染的节点越多。图 2 还表明，无论初始感染节点数为多少，完全消除网络中病毒的影响都要花费相同的时间。

图 3 考查了下载率对病毒传播的影响。图 3 表明下载率越高，感染高峰的到来的越快，感染高峰的值越大，消除网络中所有病毒花费的时间也越多。感染率由 0.01 到 0.02 时传播曲线的变化远比感染率由 0.02 到 0.03 时传播曲线变化大的事实表明，下载率达到一定的值后，下载率对病毒传播的影响达到极限，即无论如何增加这个值，其影响不会变化很大。

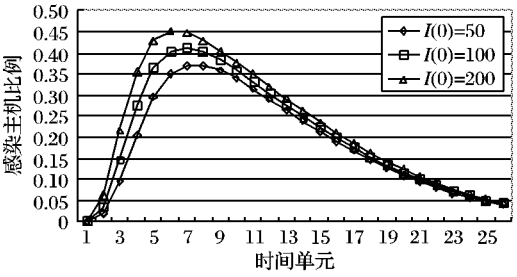

图 2 不同感染主机初值对病毒传播的影响

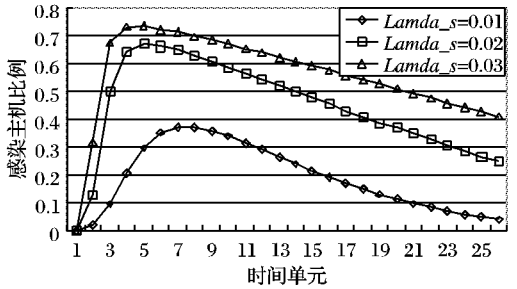

图 3 不同下载率对病毒传播的影响

图 4 考查了执行感染文件的暴露主机比例对病毒传播的影响。图中的曲线表明，执行感染文件的暴露主机比例越大，感染高峰到来的越早，感染高峰的值越大，但消除网络中所有的病毒所需要的时间是相同的。

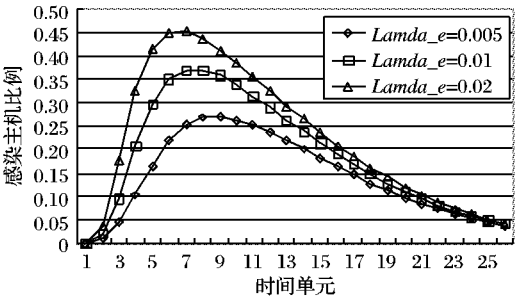

图 4 执行感染文件的暴露主机比例对病毒传播的影响

图 5 考查了恢复率对病毒传播的影响。图中曲线表明，恢复率越大，感染高峰到来的越早，感染高峰的值越小，消除网络中病毒的影响所需时间越少。这表明，提高恢复率可以有效控制病毒传播，然而要通过让用户删除病毒文件的方法来提高恢复率是不现实的。

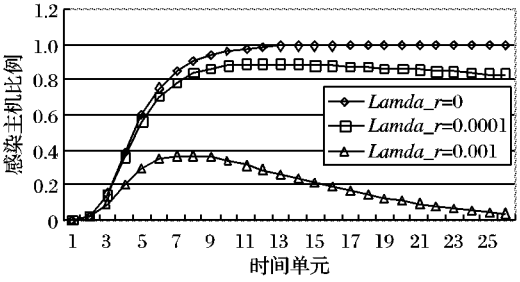

图 5 不同恢复率对病毒传播的影响

图 6 考查了参数 c 对病毒传播的影响。总的来看，c 的值越大，感染高峰来的越早，高峰的值越大，消除病毒在网络中的影响所需时间越多。在图 5 中，c 由 10 变成 50 引起的传播曲线的变化比 c 由 50 变成 100 引起的传播曲线的变化大得多，

这表明 c 的值达到一定值后,对病毒传播的影响趋于平稳。

图 7 考查了暴露主机成功执行感染文件的概率对病毒传播的影响。图中传播曲线表明,执行的成功率越高,感染高峰值越大,但感染高峰到来的时间和消除网络中所有的病毒所需时间都基本相同。

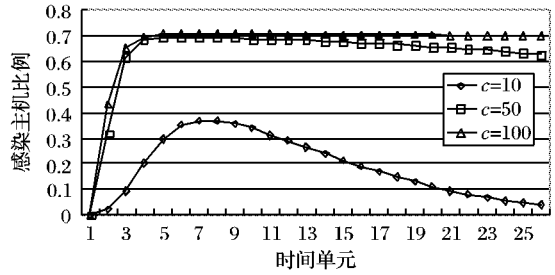

图 6　不同的 c 值对病毒传播的影响

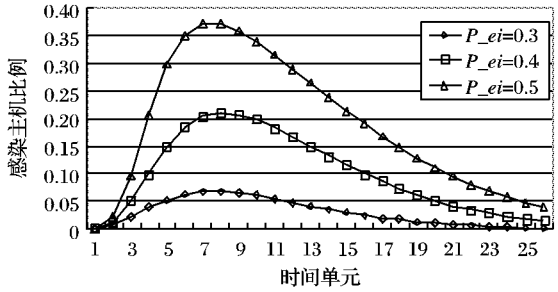

图 7　病毒文件成功执行概率对病毒传播的影响

4　结语

过去,由于 P2P 网络的规模不是很大加之病毒在 P2P 网络上传播速度较慢,因此,P2P 网络上的病毒并未怎么引起研究人员的关注。然而,在今天,随着 P2P 网络的发展,特别是 P2P 文件共享网规模和受欢迎程度的巨大提高,P2P 病毒已成为 P2P 网络的重大安全隐患。本文主要对 P2P 病毒及其传播进行了深入研究。首先,对 P2P 文件共享网络的特点和病毒传播的研究情况进行了介绍;接着在对病毒进行深入分析的基础上,提出了病毒在 P2P 网上传播的模型;最后,为了考查哪些参数对病毒传播有着重大影响,进行了大规模仿真实验。实验结果表明,不同参数对病毒传播的影响有很大不同。今后,在制定抑制病毒传播的策略时,可以考虑通过控制模型中的关键参数(如 c 和 p_{ei})来制定抑制策略。

参考文献:

[1] eDonkey2000 server list[DB/OL]. [2007 - 12 - 11]. http://ocbmaurice. no-ip. org/slist/serverlist. html.

[2] Bittorrent Protocol Specification v1.0 [DB/OL]. [2007 - 12 - 11]. http://www. bitconjurer. org/BitTorrent/protocol. html.

[3] CHEN G L, GRAY R S. Simulating non-scanning worms on peer-to-peer networks[C]// Proceedings of the 1st international conference on Scalable information systems. Hong Kong: ACM Press, 2006.

[4] MCKENDRICK A G. Applications of mathematics to medical problems[C]// Proceedings of the Edinburgh Mathematical Society. [S. l.]: Cambridge University, 1926, 44: 98 - 130.

[5] KEPHART J O, WHITE S R. Directed-graph epidemiological models of computer viruses[C]// Proceeding of the IEEE Symposium on Security and Privacy. Oakland, California: IEEE Press, 1991: 343 - 359.

[6] ZOU C C, GONG W, TOWSLEY D. Code red worm propagation modeling and analysis[C]// Proceedings of the 9th ACM Conference on Computer and Communication Security. Washington, DC: ACM Press, 2002.

[7] ZOU C, TOWSLEY D, GONG W. Email worm modeling and defense[C]// Proceedings of the 13th International Conference on Computer Communication and Networks. [S. l.] : IEEE Press, 2004: 409 - 414.

[8] QIU D, SRIKANT R. Modeling and performance analysis of BitTorrent-like peer-to-peer networks[C]// Proceedings of ACM SIGCOMM. Portland, USA: ACM Press, 2004.

[9] DUMITRIU D, KNIGHTLY E, KUZMANOVIC A. Denial-of-service resilience in peer-to-peer file-sharing systems[C]// Proceeding ACM Sigmetrics. Banff, Canada : ACM Press, 2005.

[10] THOMMES R W, COATES M J. Modeling virus propagation in peer-to-peer networks[R]. Department of Electrical and Computer Engineering, McGill University, 2005.

[11] STUTZBACH D, REJAIE R, SEN S. Characterizing unstructured overlay topologies in modern P2P file-sharing systems[C]// Proceedings of the 15th ACM Internet Measurement Conference. Berkeley, California: ACM Press, 2005: 49 - 62.

[12] MA J, CHEN X M, XIANG G L. Modeling passive worm propagation in peer-to-peer system[C]// Proceedings of the IEEE 2006 International Conference on Computational Intelligence and Security. [S. l.] : IEEE Press, 2006: 1129 - 1132.

[13] P2p-worm. win32. achar. a[DB/OL]. [2008 - 01 - 02]. http://www. viruslist. com/en/viruses/encyclopedia?virusid = 23893.

[14] W32. hllw. gotorm[DB/OL]. [2008 - 01 - 03]. http://securityresponse. symantec. com/avcenter/venc/data/w32. hllw. gotorm. html.

[15] W32/bare. worm[DB/OL]. [2008 - 01 - 05]. http://www. virus-scan-software. com/latest-virus-software/latest-viruses/w32bare-worm. shtml.

[16] Sophos virus analysis: Troj/krepper-g[DB/OL]. [2008 - 01 - 05]. http://www. sophos. com/virusinfo/analyses/trojkrepperg. html.

(上接第 1809 页)

4　结语

本文首先对 IKE 协议第一阶段基于数字签名认证的原理进行了描述。然后分析了协议容易遭受的两种中间人攻击,针对攻击导致的身份保护缺陷提出改进方案。从安全性及性能分析的结果看出,提出的两种改进方案都能有效防止中间人攻击,尤其是方案二的综合性能更优于方案一。由于 IKE 协议的灵活性及复杂性,不仅使得对其分析的难度增大,还导致其存在种种其他已知或未知的安全缺陷,为了更好地保证数据的通信安全,IKE 协议还有待于进一步完善。

参考文献:

[1] RFC2409, The Internet Key Exchange(IKE)[S]. 1998.

[2] 宋育芳, 张宏科. Internet 密钥交换协议的安全性分析[J]. 计算机工程与应用, 2004, 40(8): 136 - 139.

[3] PERLMAN R, KAUFMAN C. Analysis of the IPSec key exchange standard[C]// Proceedings of the 10th IEEE International Workshops on WEIICE: [S. l.]: IEEE Press, 2001: 150 - 156.

[4] 卫剑钒, 唐礼勇, 陈钟. IKE 协议两种身份保护缺陷的改进[J]. 计算机工程与应用, 2004, 40(26): 33 - 35.

[5] 陈艳红, 韩秀玲, 刘文超. IKE 协议主模式认证机制的分析与改进[J]. 计算机工程与应用, 2006, 42(9): 120 - 121.

[6] 张琳, 王汝传. IKE 协议中基于数字签名验证的主模式研究[J]. 南京邮电大学学报, 2007, 27(1): 70 - 73.