基于随机包标记方案的IP追踪性能分析*)

周 曜[1] 徐长江[2] 徐 佳[1] 刘凤玉[1]

(南京理工大学计算机科学与技术学院 南京 210094)[1] (海军兵种指挥学院 广州 510430)[2]

摘 要 在匿名DDoS攻击源追踪的研究领域中,基于随机包标记(probabilistic packet marking)的攻击源追踪方案以其高效和灵活成为关注的焦点,业界已经提出了多种方案,但存在着性能上的差异。本文对目前最具代表性的方案相关性能指标进行了深入探讨,指出了导致差异的关键因素是标记与重构算法以及标记概率的取值,并且伪造包会对性能造成较大的干扰。

关键词 匿名DDoS攻击,攻击源追踪,包标记,性能指标

On the Performance of IP Traceback Based on Probabilistic Packet Marking

ZHOU Yao[1] XU Chang-Jiang[2] XU Jia[1] LIU Feng-Yu[1]

(Department of Computer, Nanjing University of Science and Technology, Nanjing 210094)[1]

(Naval Arms Command Academy, Guangzhou 510430)[2]

Abstract To defend agaist anonymous DDoS attacks on the Internet, it is necessary to locate the attack source quickly and correctly. In such field, IP traceback based on packet-marking is being most concerned for its efficiency and flexibility. Many mechanism with different performance have been proposed and to reveal the reasons caused the diversities is important to future research. In this paper, we analyze and evaluate the performance of some most representive shemes. We find that the performance depend not only on the marking and reconstruct algorithm but also on the marking probability and may be greatly decreased by the spoofing packets.

Keywords Anonymous DDoS attack, IP traceback, Packet marking, Performance

1 引言

因特网中的攻击手段繁多。根据规模的不同,我们可以把网络中的攻击分为两种类型:洪泛攻击和少数包攻击。前者以分布式拒绝服务(Distributed Denial of Service, DDoS)攻击为代表,特点是发送数量庞大的数据包,以耗尽目标节点的资源,发起拒绝服务攻击或造成瘫痪,其实施简单,危害巨大。为了避免被定位,有经验的黑客常常会以匿名的方式发起DDoS攻击,即使用伪造的IP源地址。为了有效地与之对抗,被攻击方必须快速准确地找出攻击者真实的位置,与之相关的定位技术称为攻击源追踪或IP追踪(IP Traceback)技术。

2002年,Savage等人在文[1]中率先提出了针对DDoS的基于包标记进行攻击源追踪的方案FMS(Fragment Marking Sheme),其中心思想是通过中间路由器随机地在转发的IP报文头部标记地址信息来重构攻击路径。该方案一经提出,就以其高效性和灵活性引起了业界的广泛注意。在此基础上,研究者又相继提出了高级帧标记方案AMS(Advance Marking Scheme)[2]、可变概率包标记方案APPMS(Adjusted Probabilistic Packet Marking Scheme)[3,4]、权重包标记策略[5]、验证包标记策略等,形成了攻击源追踪研究领域的一个体系和热点。

这些方法在性能上存在差异,产生原因是多种多样的,而找出其中的关键点对攻击源追踪方案在性能上的优化有着现实意义。本文对包标记技术中三种最具代表性的方案:FMS、AMS和APPMS的性能做了深入探讨,指出影响性能指标的关键因素,为后续的相关研究提供了理论依据。

2 性能研究

衡量一个攻击源追踪方案的性能高低主要以收敛时间、误报率和抗干扰能力这三个指标来考量,它们分别体现了系统的效率、精度和健壮性。

2.1 收敛时间

影响收敛时间的主要因素是标记包的收集时间和重构攻击路径的计算时间。

2.1.1 FMS的收敛时间

FMS使用概率 p 决定中间路由器是否对转发的数据包加以标记,标记值是(*start*, *end*, *distance*)这一三元组,其中*start*和*end*分别是本跳和下一跳的地址,*distance*为本机距离终点的跳数。当有足够多的包到达终点后,将能够得到路径上每一个路由器所标记的包。目的节点将使用这些标记包里所包含的边的信息重构出整条路径。具体的标记算法和重构算法可见文[1]。

FMS的标记包收集时间取决于何时收到来自所有中间路由器的标记包。文[1]中指出收敛所需包的期望值 $E[X]$ 是可以控制的,即

定理1 对于长度为 d 的路径,若其中每个中间路由器都以概率 p 随机地对转发包进行标记,则为了得到来自每个路由器的至少一个标记包,所需收集的包数目期望值:

*)国家自然科学基金资助项目(60273035)。周 曜 博士生,主研方向:信息安全与移动自组织网络;徐 佳 博士生,主研方向:拥塞控制与网络管理;刘凤玉 教授,博士生导师,主研领域:网络安全、软件性能保持和多媒体。

$$E[X]<\frac{\ln(d)}{p(1-p)^{d-1}}$$

文[1]中并没有给出定理1的证明，此处证明之。

证明：设数据包所经过的路径为 $S\rightarrow R_1\rightarrow R_2\rightarrow\cdots\rightarrow R_{d-1}\rightarrow R_d\rightarrow D$，其中 S、$R_i(1\leqslant i\leqslant d)$、$D$ 分别为源点、中间路由器和终点，易见某个包被 R_i 采样（被标记且在到达终点前未被重新标记，下同）的概率 $P_i=p(1-p)^{d-i}$。

设 $p_s=\sum P_i/d$ 为整个路径上的平均采样概率，F_i 为被 R_i 采样的包，若 F_i 未被之前的 $i-1$ 个路由器采样，则定义 F_i 采样成功。记 X_i 为从 F_i 采样成功到 F_{i+1} 采样成功后所需要的包数，易见

$$X=\sum_{i=0}^{d-1}X_i$$

在 F_i 采样成功之后，每个包采样成功的概率 $p_i=\frac{d-i}{d}\sum P_i=(d-i)p_s$，并且 X_i 符合 p_i 上几何分布，故 $E[X_i]=\frac{1}{p_i}=\frac{1}{(d-i)p_s}$，且 $E[X]=\sum_{i=0}^{d-1}E[X_i]=\frac{1}{p_s}\sum_{i=0}^{d-1}\frac{1}{d-i}=\frac{1}{p_s}\sum_{i=1}^{d}\frac{1}{i}<\frac{H_d}{p(1-p)^{d-1}}$ 其中 H_d 为 d 阶调和级数，因为 $H_s=\ln(d)+O(1)$，所以

$$E[X]<\frac{\ln(d)}{p(1-p)^{d-1}}$$

证毕。

实际上，由于路径上第 i 个路由器 R_i 处的标记包出现的概率为 $p(1-p)^{d-i}$，随 i 递增，在 R_1 处具有最小的概率 $p(1-p)^{d-1}$，那么在期望情况下，至少要收到来自 R_1 的1个标记包才可以保证收到整个路径上的标记包，即有

$$\frac{1}{p(1-p)^{d-1}}\leqslant E[X]<\frac{1n(d)}{p(1-p)^{d-1}} \quad (1)$$

设 $f(p)=p(1-p)^{d-1}$，因为

$$f'(p)=(1-p)^{d-2}(1-pd)$$

当 $f'(p)=0$ 时，即 $p=1$ 或 $p=1/d$ 时，$f(p)$ 有极大值。但 p 不能为1，故当 $p=1/d$ 时，$f(p)$ 极大。

从上面的分析可以看出，FMS的标记包收集时间取决于标记概率 p 和攻击路径的长度 d，$p=1/d$ 时具有最佳的收敛性能。

攻击路径的重构时间取决于重构算法的计算复杂度。FMS为了节省存储空间，采取了把地址信息分片分别存储的策略，为了验证地址的有效性，把原地址和其32位的HASH值分别放在奇位和偶位，共64位分成8片，组合后验证奇位的HASH值是否与偶位相同。由于受害者重构时无从知晓某8个分片是否来自同一路由器，它必须尝试所有的组合来验证是否有效。

FMS的重构算法从终点开始递归地得到攻击路径上路由器的地址，设 S_k 为 k 处已得到的路由器集合，C_{k+1} 为 $k+1$ 处所有分片的组合的集合，$\psi_{k+1,f}$ 为 $k+1$ 处第 f 个分片，显然

$$|C_{k+1}|=\prod_{0\leqslant f\leqslant 7}|\psi_{k+1,f}|$$

在距离 k 处，为得到 $k+1$ 处的值，需进行 $|S_k|\cdot|C_{k+1}|$ 次XOR计算。在长度为 d 的路径上，需检验的组合数目

$$|\Gamma|=\sum_{1\leqslant k\leqslant d}(|S_{k-1}|\cdot\prod_{0\leqslant f\leqslant 7}|\psi_{k,f}|)$$

对于每一种组合，至少需要进行一次HASH计算来检验。

对于DDoS，由于每一 k 处结点会很多，分片后所有组合数会相当多。例如对于仅25个攻击者，$\prod_{0\leqslant f\leqslant 7}|\psi_{k,f}|$ 为 $25^8\approx 1.5\times 10^{11}$，所以FMS在试图重构DDoS的攻击路径时，计算量会非常可观。

2.1.2 AMS的收敛时间

AMS使用地址的哈希输出值来取代原值，从而减少了在IP报头的存储空间。由于没有地址原值，被攻击者必须知道整个网络的拓扑，才能重构出攻击的路径。AMS的标记算法与FMS相比，不同之处只在于标记的内容，因此在路径上不同路由器的标记包出现概率与FMS呈现同样的分布，故标记包的收集时间与FMS相同。

AMS用地址的哈希输出取代原值，可以把32位的地址压缩到8位。为了降低冲突率，采用HASH函数族，同时在标记时加入了3位的HASH函数索引。由于不存在地址的分片，重构时的计算量大大降低。与FMS比较，AMS不需要检查所有分片的组合，而只需检查地址的单个HASH值，其重构算法计算复杂度仅为 $O(\sum_d|S_d|\cdot|\psi_{d+1}|)$，因此路径重构时间大为降低。

2.1.3 APPMS的收敛时间

APPMS使用动态概率进行标记，其目的是要使长度 d 的攻击路径上所有的中间路由器的标记包都以同样的概率 $1/d$ 出现。即对于攻击路径 $S\rightarrow R_1\rightarrow R_2\rightarrow\cdots\rightarrow R_{d-1}\rightarrow R_d\rightarrow D$，如果使用固定概率 p 标记，则 R_i 处的采样概率 $P_i=p(1-p)^{d-i}$ 各不相同。只有采用变概率，即对不同的 R_i 采用不同的标记概率 p_i，才能使得最终每个节点处的采样概率相等[3]。

文[4]中提出了根据数据报头里的TTL值来决定标记者距离发送者的跳数 i（TTL值通常在发送时被设为32的整数倍），并取标记概率为 $p_i=1/i$，从而在 R_i 处有 $P_i=\frac{1}{i}\cdot(1-\frac{1}{i+1})\cdot(1-\frac{1}{i+2})\cdot\cdots\cdot(1-\frac{1}{d-1}\cdot(1-\frac{1}{d})=\frac{1}{d}$ 可见采样概率为固定值，与标记者所处的位置 i 无关。

由于所有中间路由器都具有相同的采样概率 $1/d$，故收敛所需包的数目期望值

$$E[X]=d(\ln(d)+O(1))$$

（此式只需在定理1的证明过程中把 p_s 替换成 $1/d$ 即可得到）。例如在 $d=20$ 时，APPMS需约60个包即可得到来自所有中间结点的标记包。此时FMS或AMS（取 $p=1/20$ 的最优值）所需包数在53和159（见式(1)）之间。一般来说，APPMS具有比FMS或AMS更短的标记包收集时间。

与FMS或AMS相比，APPMS的区别在于标记算法中概率的取值，其重构算法与上述两者相比并无不同，因此路径重构时间相同。

2.2 误报率

2.2.1 FMS的误报率

由于FMS的奇偶检验策略使用32位的HASH函数来验证，冲突率为 $1/2^{32}$，故在所有的 $|\Gamma|$ 个组合中，冲突数目的期望值为 $|\Gamma|/2^{32}$，设为 α；由于IP地址为32位，对于某一个不在攻击路径上的路由器B来说，这 α 个组合里至少有一个与B的IP地址相同的概率是 $1-(1-2^{-32})^{\alpha}$，B最多有 2^{32} 种可能，故误报数（一个误报即 α 中的某个数字串与某一个不在攻击路径上的路由器地址相同）的期望值为

$$E[false]=(1-(1-2^{-32})^{\alpha})\cdot 2^{32}$$

当 $\alpha<<2^{32}$ 时，$E[false]\approx\alpha=\frac{|\Gamma|}{2^{32}}$。按上一节对计算量的讨论，可知 $|\Gamma|$ 在DDoS攻击时会非常大，因此哪怕在面对仅几

十个 DDoS 攻击者时，也会产生数千个误报。

2.2.2 AMS 的误报率

AMS 采取了两方面的措施来降低误报率：一是引进了异常阈值 m，在收到超过 m 的标记包时才认为处于攻击路径上，这样可以过滤转发正常数据包的路由器；二是采用了一组相互独立的 HASH 函数来取代单个函数，可以降低冲突率，若采用 2^w 个相互独立的 HASH 函数，则可以将冲突率降低到原来的 $1/2^w$。

设在网络上距离被攻击者 d 处存在 $|M_d|$ 个攻击者，对于 $d-1$ 处某个有 t_y 个子孙的节点 y 来说，其子孙中的误报数为

$$t_y \times \prod_{1\leqslant i\leqslant 2^w} \frac{\Phi_{d,i}}{2^{11-w}}$$

其中 $\Phi_{d,i}$ 是来自距离为 d 的路由器使用第 i 个函数的标记包数目，假设 HASH 函数是理想的，则

$$E(\Phi_{d,i})=(1-(1-\frac{1}{2^{11-w}})^{|M_d|})\times 2^{11-w}$$

例如取 $w=3, t_y=32, |M_d|=128$，则 y 的子孙中预计的误报数将不大于 1。文[2]中的研究表明，采用此种策略可以对抗来自约 1500 个节点的 DDoS 攻击。

2.2.3 APPMS 的误报率

由于 APPMS 和 AMS、FMS 唯一的区别在于标记时所取概率，而从上面的分析看出，引起误报率的主因是 HASH 函数的冲突率，与标记概率无关。因此，APPMS 的误报率与 FMS 或 AMS 相比并无不同。

2.3 抗干扰能力

2.3.1 FMS 和 AMS

FMS 和 AMS 的强制距离递增只能防止攻击者假冒自己与受害者之间的路由器，却不能防止它冒充之外的路由器，如图 1 所示。

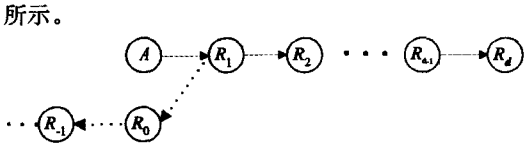

图 1 伪造路径的形成

若攻击者 A 在包的 *start* 域里写入 R_0 的地址，*distance* 域里写入 0，当 R_1 决定不标注时，它会在 *end* 域里写入自己的地址，同时把 *distance* 加 1。这样，带有边信息 (R_0, R_1)，*distance* 为 $d+1$ 的标记包将以 $(1-p)^d$ 的概率到达受害者 V。由于 FMS 是迭代地得到路径上的路由器地址，V 将会得到 $(R_0, R_1, \cdots, R_d)$ 这一错误的攻击路径。同样，A 也可以制造 (*start*, *end*, *distance*) 为 $(R_{-1}, R_0, 1)$ 的包，和上面的包一起将会使 V 得到 $(R_{-1}, R_0, R_1, \cdots, R_d)$ 的路径。依次类推，A 可以通过伪造标记包，把攻击路径引向任意的起点。因为 (*start*, *end*) 域分别为 (R_0, R_1) 和 (R_{-1}, R_0) 的包到达 V 的概率都为 $(1-p)^d$，看起来似乎可以通过比较收到包的数目来鉴别伪造包（当收到足够多的包时，距离越远的路由器标注的包将越少。当来自最远的两跳的标记包数目相同时，则认为该两跳是伪造的），但这只是期望的情况，并不一定符合实际情况。

按照 FMS(AMS) 固定标记概率的策略，由于伪造包到达 V 的概率为 $(1-p)^d$，因此可以考虑提高 p 值来降低这种可能性。但是这会增加路由器的标记成本，也不能完全屏蔽伪造包（p 不能为 1，否则只能收到来自 R_d 的包）。在不能完全杜绝伪造包的前提下，在受害者和标记路由器之间必须存在有效的鉴别机制。

2.3.1 APPMS

APPMS 的可变概率机制可以杜绝来自攻击者的伪造包，这是因为在长度为 d 的路径，每个中间路由器的标记包出现的概率为 $1/d$，总的概率为 1，也就是说攻击者发送的伪造包在到达终点前，一定会被某个中间路由器重新标记。

但是，由于 APPMS 的标记概率完全由 TTL 值决定，攻击者可以伪造 TTL 值来使总的标记概率不为 1。在这种情况下，APPMS 不能完全地避免干扰。文[4]中提出了一种解决方案来降低这种影响，即设定一个 TTL 域值 T_{ini}（可以设为 32 或 64，因为绝大多数通信都不会超过此范围），当发现数据包的 TTL 值大于 T_{ini} 时，即改为 T_{ini}，并按此时距攻击者为 1 跳（认为该包是伪造的 TTL）来确定标记概率。在这种情况下，P_i 同样为 $1/d$，路径上总的标记概率为 1。攻击者为了不被发现，只能以小于 T_{ini} 的 TTL 值来发送包。设 $TTL=T_{ini}-x$，此时 $P_i=1/i+x$，而某个包在整个路径上不被标记的概率为 $x/d+x$。

3 三种方案的比较

表 1 显示了三种方案在性能上的比较。表中四项性能指标从左到右分别为标记包收集时间、攻击路径计算时间、误报数和抗伪造包能力，其中 APPMS 的路径计算时间与误报数视其使用相对或绝对地址标记而与 AMS 或 FMS 相同。

表 1 三种方案性能对比

	Packets Collection Time	Pathes Compution Time	False Positives Number	Resistibility against Spoofing
FMS	Long	Long	High	Bad
AMS	Short	Short	Low	Bad
APPMS	Short			Good

AMS 比 FMS 具有更快的收敛速度和更低的误报率，但前提是被攻击者必须了解外部网络的拓扑结构，在许多研究者已开展的网络拓扑发现的技术基础上，这一前提是合理的。在没有相关网络资料的情况下，FMS 是较好的选择，虽然其在重构路径时计算量很大，但是可以采用具有高速计算能力的设备来减小这一影响。

与 FMS 和 AMS 采用固定概率不同，APPMS 使用动态的标记概率，这使其在收敛速度和抗干扰能力上具有较大的优势，但是 APPMS 在路由器处的标记成本比较高，这体现在两个方面，一是要根据包的 TTL 值来计算 p，二是标记概率比较高，尤其是在距源点最近处会达到 1，这在现实中是很难实现的。

评价一种攻击源追踪方案的优劣并没有绝对的标准，选择何种方案往往要根据追踪者占有的资源、被攻击的规模以及所追求的效果来最终确定。

目前基于包标记法的追踪技术都未能完全解决攻击者反追踪的问题，在这方面表现最好的 APPMS 也只能尽量减少伪造标记包的干扰而无法完全杜绝，因此如何解决标记者身份验证的问题将是以后研究的重点之一。

在攻击源追踪研究领域，包标记法的优势是明显的，但是其同样存在难以大规模部署、管理成本高等缺点，目前有关研究也只是停留在试验阶段，而少见实践中的成功报道，这同样需要研究者在架构的易实施性及协议的向下兼容性方面做进

一步的研究。

结束语 基于包标记的攻击源追踪技术以其灵活高效的特点，引起业界的广泛注意。研究者提出了很多方案，但存在性能上的差异，本文对目前有代表性的几种方案的性能指标做了深入的探讨，为后续工作的性能优化提供了依据。

参 考 文 献

1 Savage S, Wetherall D. Network support for IP traceback. IEEE/ACM Transactions on Networking, 2001, 9(3): 226～237

2 Song D, Perrig A. Advanced and authenticated marking schemes for IP traceback. In: Proceedings of the IEEE INFOCOM, Anchorage, Alaska USA, 2001, 2: 878～886

3 Peng T, Leckie C, Kotagiri R. Adjusted probabilistic packet marking for IP traceback [A]. In: Proceedings of the Second IFIP Networking Conference (Networking 2002) [C]. Pisa, Italy, May 2002. 697～708

4 Liu J, Lee Z J, Chuang Y C. Efficient dynamic probabilistic packet marking for IP traceback [A]. Networks, ICON2003 [C], 2003. 475～480

5 徐永红，杨云，刘凤玉，等. 基于权重包标记策略的 IP 追踪技术研究[J]. 计算机学报，2003, 26(11): 1598～1603

（上接第 67 页）

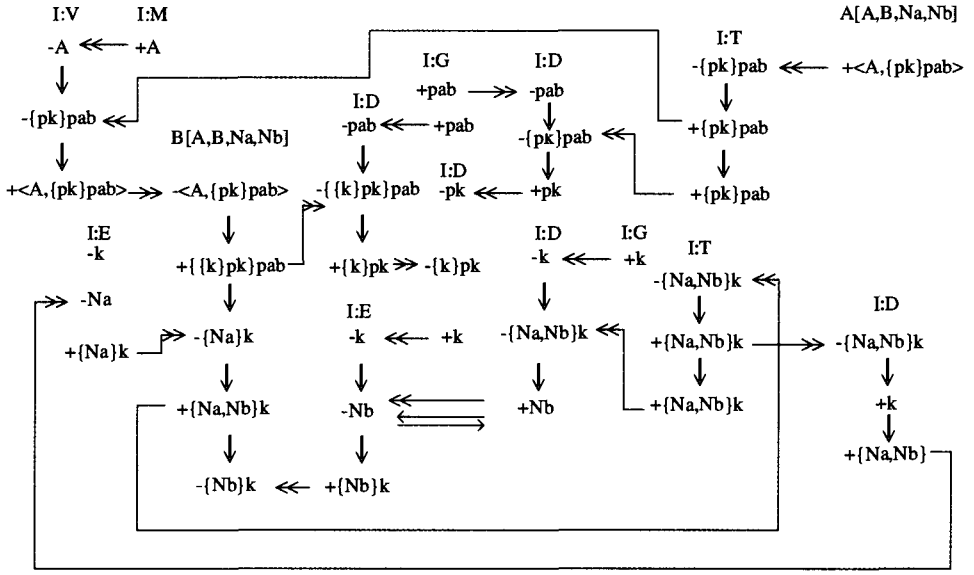

图 1 Athena 分析方法最终状态图

图 1 为运用扩展后的 Athena 方法生成的最后的状态图。因为在这里 $\Delta \cap l = \emptyset$，Final 规则不能运用，且 $G = \emptyset$ 说明认证性不满足。又由 $GG \neq \emptyset$ 可知，存在猜测攻击，所以该状态图为猜测攻击的反例。

$\leftrightarrows$ 两端对应的为两种不同方式产生的验证项 Nb，其生成过程便是猜测攻击的攻击过程：

(1)攻击者截获 $\{pk\}_{pab}$；

(2)猜测弱口令 pab，得到 pk；

(3)截获 $\{\{k\}_{pk}\}_{pab}$；

(4)利用 pab 和(2)中得到的 pk，得到 k；

(5)截获 $\{Na, Nb\}_k$，$\{Nb\}_k$；

(6)利用 k 解密 $\{Na, Nb\}_k$，得到 Nb；解密 $\{Nb\}_k$ 得到 Nb，存在两种不同的方式产生验证元 Nb。

结束语 本文在课题组前期工作的基础上针对猜测攻击对 Athena 分析方法进行了扩展，并运用扩展后的 Athena 方法对协议实例进行了分析，指出其攻击序列。本文进行扩展时对 Lowe 在文[6]中形式化定义的猜测攻击的条件进行了适当简化，假定私钥不可能丢失，因此没有将其条件(4)中第三部分($v \in$ ASYMMETRIC_KEYS $\wedge v^{-1} \in IK$)考虑其中。本文分析实例时猜测攻击之所以发生是由于协议中存在两种不同方法来产生验证元(将验证元直接在协议中出现也视为其中一种方法)。将一个单独运行的不存在猜测攻击的协议置于多协议环境中，由于消息项交互的可能性增加，由两种不同方法产生验证元的可能性也随之增加，即发生猜测攻击可能性会增加。因此将各自独立运行安全的多个协议混合运行，发生猜测攻击也是可能的。

参 考 文 献

1 卿斯汉. 安全协议 20 年研究进展[J]. 软件学报，2003, 14(10): 1740～1752

2 Gong L, Mark T, Lomas A, et al. Protecting poorly chosen secrets from guessing attacks[J]. IEEE Journal on Selected Areas in Communications, 1993, 11(5): 648～656

3 Steiner M, Tsudik G, Waidner M. Refinement and extension of encrypted key exchange[J]. ACM Operating Systems Review, 1995, 29(3): 24～29

4 Ding Y, Horster P. Undetectable on-line password guessing attacks[J]. ACM Operating Systems Review, 2000, 34(4): 12～20

5 Halevi S, Krawczyk H. Public-key cryptography and password protocols[A]. In: ACM Transaction on Information and System Security[C]. New York, USA, ACMPress, 1999, 23: 230～268

6 Lowe G. Analysing protocols subject to guessing attacks[A]. In: Workshop on Issues in the Theory of Security WITS'02[C], January 2002

7 Thayer F J, Herzog J C, Guttman J D. Strand spaces: Why is a security protocol correct? In: Proceedings of 1998 IEEE Symposium on Security and Privacy[C], 1998. 160～171

8 Song D, Perrig A, Berezin S. Athena: a novel approach to efficient automatic security protocol analysis[J]. Journal of Computer Security, 2001, 9(1-2): 47～74

9 吴光伟，董荣胜. 基于串空间的 Athena 分析技术研究[J]. 计算机科学，2006, 33(8): 9～13