基于同步点的 IDS 评估评分方法

杨望[1,2]，龚俭[1,2]，吴雄[1,2]

(1. 东南大学 计算机科学与工程学院，江苏 南京 210096;
2. 江苏省计算机网络技术重点实验室，江苏 南京 210096)

摘 要：评分方法是判定 IDS 的检测率和误报率的重要环节，评分方法的准确性直接影响评估结果的有效性。现有的考虑误报和不考虑误报的评分方法均存在不同程度的准确性误差，并且不能适应日益增加的带宽下的流量需求。本文分析了评分方法的判定窗口所需要的性质，并基于 IDS 对报文处理的 FIFO 队列特性，提出了基于同步点的评分方法。经过理论证明和实验验证，这个新方法相对于现有的评分方法有更高的准确性和更好的可扩展性。

关键词：IDS 评估；评分方法；同步点；误报率；可扩展性

中图分类号：TP393 **文献标识码**：A **文章编号**：1000-436X(208)09-

SyncPoint based scoring method in IDS evaluation

WANG Yang[1,2], GONG Jian[1,2], WU xiong[1,2]

(1. School of Computer Science and Technology, Southeast University, Nanjing 210096, China;
2. Key Laboratory of Computer Network Technology of Jiangsu Provincial, Nanjing 210096, China)

Abstract: Scoring the true positive rate and the false positive rate is a key component in IDS evaluation. The accuracy of the scoring method affects the effectiveness of the evaluation results. There are two kinds of scoring methods existed, one considering the false positive and the other not. But both of them aren't accurate enough and don't scale to the traffic volume increase. The paper analyzes the characteristics required by the evaluating window, and proposes a SyncPoint based scoring method utilizing the features that the IDS processes the packet in a FIFO queue way. The theoretical analysis and the experiment show that the SyncPoint based scoring method is better than the current methods in accuracy and the scalability.

Keyword: IDS Evaluation, scoring method, SyncPoint, false positive rate, scalability

1 引言

随着入侵检测技术的不断发展，IDS(intrusion detection system)已经成为安全防御体系中必备的一环，而对 IDS 评估的需求也不断增长。在 IDS 评估过程中，研究者通过对 IDS 检测率、误报率等测度的测量，发现了影响 IDS 检测精度的原因，从而提高了 IDS 系统的检测率，降低了误报率。目前主要的 IDS 评估方法是混合背景流量和攻击流量构造测试数据，测量 IDS 对测试数据的检测率和误报率，在此基础上通过 ROC 曲线等方法评价 IDS 检测能力。

评分方法是判定 IDS 的检测率和误报率的量化手段。评分方法通过比对标准答案和 IDS 给出的报警结果，区分 IDS 生成的真实警报和误报，再根据标准答案给出的攻击数量算出检测率，根据非攻击

收稿日期：2008-03-25；**修回日期**：2008-07-23
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2003CB314804）
Foundation Items: The National Key Basic Research Program of China(973 Program) (2003CB314804)

数量算出误报率。所以评分方法的准确性直接影响评估结果的有效性。

完全正确地判断警报和攻击的对应关系需要 IDS 保留完整的攻击报文信息，这种情况下对 IDS 提供的数据内容有严格的要求。由于目前 IDS 输出的报警格式并没有统一的标准，强制要求所有 IDS 提供相同的输出内容是不现实的，而如果要求 IDS 记录所有的报文信息，则由于磁盘 I/O 的速度远低于网络速度，将对 IDS 的测试性能产生严重影响，所以其测试代价不一定是可接受的。一般的解决方法是使用减弱的判断条件进行评分。目前的评分方法有两类。一类以 NSS 为代表，不考虑误报，通过简单统计 IDS 的警报数量获得结果[1]。另一类以麻省理工学院的林肯实验室(MIT/LL, massachusetts institute of technology/lincoln lab)为代表，考虑误报，通过时间窗口和警报信息区分正确的警报和误报，获得结果[2,3]。本文首先讨论了 2 种方法在评分准确性方面的缺陷，并分析了基于时间窗口的方法在准确性和扩展性方面的不足，归纳出评分方法所需要的 5 条性质。基于 IDS 对报文的处理是 FIFO 队列的特性，定义了同步点概念，并提出了基于同步点的评分方法，以满足评分准确性和可扩展性的需求。

本文的结构如下，第 2 节介绍目前评分方法的工作，指出目前各类方法中的不足；第 3 节讨论了评分方法的需求，归纳了准确的评分方法需要的五条性质；并提出了基于同步点的评分方法；第 4 节介绍了基于同步点的评分方法的实现；第 5 节通过试验对比了传统方法和基于同步点的评分方法的效果，第 6 节是结论和未来的工作。

2 目前评分方法的问题

在讨论目前的评分方法前，先定义以下概念：

攻击指一个入侵事件，在 IDS 评估中，由于真实攻击工具收集的困难[4]，攻击可以是符合 IDS 攻击描述的事件[5]，而不是真实入侵事件。

警报指一个事件被检测到的消息[6]。如果该事件为攻击，则警报为真，反之为假。

误报指一个假警报。

检测率指真警报和攻击数的比率。

误报率是误报和非攻击数的比率。非攻击数的定义和 IDS 的分析单元有关[3]。如果 IDS 分析单元是报文，则非攻击数是所有非攻击的报文数；如果 IDS 的分析单元是流，则非攻击数是所有非攻击的流数。不同的分析单元对于后文讨论的方法只有数量级的差别（报文数和流数的数量级差异随平均流长变化而变化），因为报文仍是 IDS 分析攻击的基本数据单元，Snort、Bro 等主流的开源 IDS 仍然支持基于报文的检测，所以后文的讨论均以报文数为 IDS 的分析单元，非攻击数即非攻击的报文数，误报率为误报与非攻击的报文数的比率。

2.1 不考虑误报的评分方法

不考虑误报的评分方法统计警报个数计算检测率，如独立商业测试组织 NSS Group 和 Wisconsin–Madison 大学的 Joel Sommer[7]进行的 IDS 评估都使用这种类型的方法。这种方法的优点是简单，对被测试者提供的警报格式没有要求。如果测试中只包含攻击流量，不包含背景流量，则测试中没有误报的干扰，采用该方法是适用的。但在模拟真实环境的测试中存在背景流量对 IDS 造成误报，则需要考虑误报对统计结果的影响。

根据警报个数统计的检测率等于所有的警报数（正确的警报数加误报数）除以总攻击数，而正确警报数等于攻击数乘以检测率，误报数等于非攻击数乘以误报率，所以估计的检测率 p_{tptest} 是一个和检测率、误报率以及攻击比率相关的条件概率[8]。设 p_{tp} 为检测率，p_{fp} 为误报率，攻击比率为 R_{attack}，则通过估计个数所得到的估计检测率 p_{tptest} 为

$$p_{\text{tptest}} = \frac{p_{\text{tp}} \times R_{\text{attack}} + p_{\text{fp}} \times (1 - R_{\text{attack}})}{R_{\text{attack}}} = p_{\text{tp}} + p_{\text{fp}} \times \frac{1 - R_{\text{attack}}}{R_{\text{attack}}} \quad (1)$$

当 R_{attack} 为 100%，相当于没有背景流量的情况，p_{tptest} 和 p_{tp} 相等，采用统计警报个数的方法是适用的。当 p_{fp} 和 p_{tp} 不变，随着 R_{attack} 变小，$\frac{1 - R_{\text{attack}}}{R_{\text{attack}}}$ 变大，p_{tptest} 单调递增，和检测率 p_{fp} 产生偏差，R_{attack} 越小，偏差就越大。在现实环境中，网络背景流量是必须要考虑的，而且它比攻击流量要大得多。

当 R_{attack} 和 p_{tp} 不变时，p_{tptest} 将随着误报率 p_{fp} 的变化而变化，p_{fp} 越大，p_{tptest} 和 p_{tp} 的偏差就越大。p_{tp} 是 IDS 和流量属性之间存在函数关系，IDS 的检测方法和阈值，以及流量的属性不同，p_{fp} 都会发生变化。

异常入侵检测通过建立系统的正常模式轮廓，如果实时获得的系统轮廓值和正常值的差异超过指定的阈值，就进行入侵报警。轮廓的建立方法是从一组测度中选择能够检测出入侵的测度，构成子集[9]。设用于检测的测度为$\{m_1, m_2, \mathbf{L}, m_n\}$，每种测度地阈值为$\{t_1, t_2, \mathbf{L}, t_n\}$，背景流量中每种测度的分布为$\{p_1, p_2, \mathbf{L}, p_n\}$，则 IDS 的误报率是测度阈值和背景分布的函数 $f(p_1, p_2, \mathbf{L}, p_n, t_1, t_2, \mathbf{L}, t_n)$。定性地来说，如果 IDS 选择的测度子集不适合入侵类型，使 IDS 在测度子集上很难定义有效的阈值，则误报率较高，反之则较低。Matthew V. Mahoney 对 3 种不同的基于异常 IDS 的测度选择和背景流量属性进行过比较，其误报率分布最高可达$1e-3$[10]。

利用入侵检测定义攻击的描述规则，如果报文匹配描述规则，则检测出攻击[9]。这样每条规则都可能产生误报，IDS 的误报率为所有规则的误报率之和。每条规则的误报率和该规则在检测的各分类域上的值以及各分类域的值域分布有关，设 IDS 共有 m 维分类域$(f_1, f_2, \mathbf{L}, f_m)$，每个分类域上报文取值的概率分布分别为$(p_1, p_2, \mathbf{L}, p_m)$，规则在每个分类域上的取值分别为$(v_1, v_2, \mathbf{L}, v_m)$，则第 i 条规则的误报率为该规则在分类域上的取值和背景流量在分类域上分布的函数，即 $fp_i = F(v_1, v_2, \mathbf{L}, v_m, p_1, p_2, \mathbf{L}, p_m)$，IDS 的误报率 $fp_{ids} = \sum_{i=1}^{N} fp_i$。滥用 IDS 的误报率分析较为复杂，目前没有专门的文献进行讨论。根据作者在 CERNET 华东北地区网上的实际观察，误报率在$1e-6$水平。

根据式(1)，对滥用和异常分别固定 R_{attack} 和 P_{fp} 对不同 p_{tptest} 的计算可以得到以下结果

图 1 和图 2 是攻击比率分别为 1e-3 和 1e-7 的条件下使用警报个数计算检测率随误报率变化的趋势。当攻击比率等于 1e-3，误报率对估计检测率的影响相对较小，当误报率小于等于 1e-5，估计的检测率和真实检测率基本吻合，当误报率大于等于 1e-4 时，估计检测率开始明显大于真实检测率，当误报率等于 0.01 时，估计的检测率接近于 2，属于无意义的结果。当攻击比率为 1e-7 时，误报率对估计检测率的影响更加明显，当误报率大于等于 1e-8 时，估计检测率开始明显大于真实检测率，当误报率为 0.001 时，估计检测率达到了夸张的 10 000。

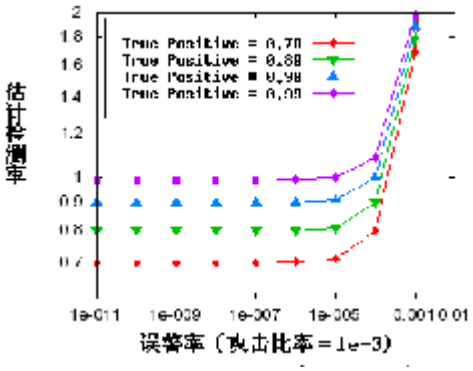


图 1　误警率对估计检测率影响（*ar*=1e-3）

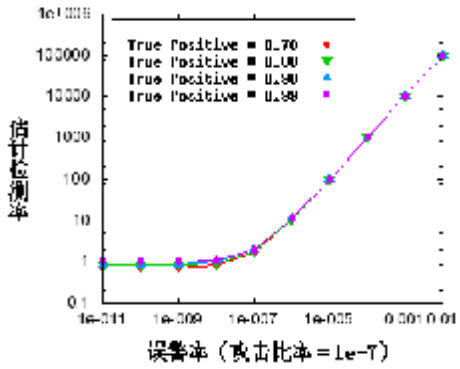


图 2　误警率对估计检测率影响（*ar*=1e-7）

图 3 和图 4 是误报率分别在 1e-6 和 1e-3 的条件下使用警报个数统计计算检测率随攻击比率变化的趋势。当误报率为 1e-6 时，攻击比率对估计检测率影响较小，但当攻击比率下降到 1e-6 时，估计的检测率接近于 2，也超过合理的范围，只有当攻击比率大于等于 1e-4 时，估计的检测率才接近于真实的检测率。当误报率为 1e-3 时，攻击比率对估计的检测率影响极大，当攻击比率为 1e-6 时，估计的误报率达到了 1 000，只有攻击比率大于等于 0.01 时，估计误报率才近似于真实检测率。

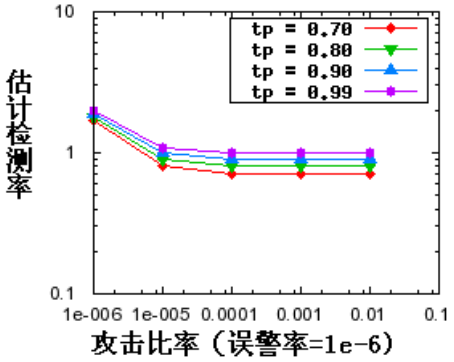


图 3　攻击比率对估计检测率影响(*fp*=1e-6)

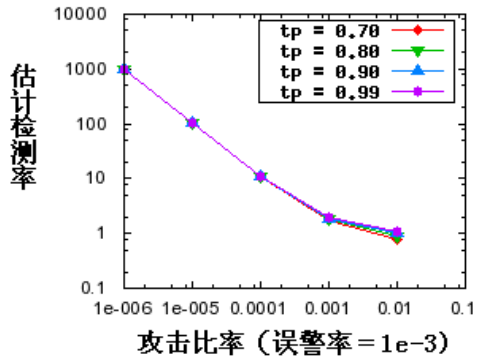


图 4　攻击比率对估计检测率影响(*fp*=1e-3)

根据上述分析，无论滥用还是异常入侵检测方法，使用警报个数统计只有在攻击比率极大（大于 1e-3）或误报率极低（小于 1e-6）时，其估计结果才接近于真实结果，其他情况将获得大于真实检测率的结果，甚至会出现检测率大于 1 的荒谬结果。由于误报率在测试前一般不能确定，所以在评估中只有在攻击比率极大或为 1 的情况下，才可以使用统计个数的办法，在模拟真实场景的评估中，攻击比率一般都极低，不考虑误报的评分方法给出的评估结果是不合理的。

2.2　考虑误报的评分方法

一般测试者使用 IP 和攻击类型对警报进行判断，由于背景流量报文和攻击流量报文可能使用相同的 IP 地址和协议，所以在不对报文信息进行完全比对的情况下，仅使用攻击类型和 IP 不足以判定警报的真实性。如图 3 所示，MIT/LL 在 IP 判定的基础上通过时间窗口对警报的范围进行限定，如果警报的时间位于攻击的时间窗口内（前后一分钟），则属于真实警报。使用时间窗口进一步增加警报对应真实攻击的可能性。下面分别从讨论基于时间窗口的方法存在的问题。

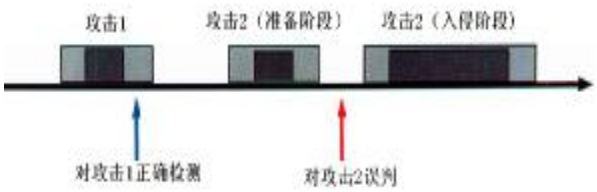


图 5　MIT 时间窗口判定方法

1) 准确性

使用时间窗口提高准确性的前提是时间窗口能将误警和真实警报区分在不同的区间内，如果误警落在时间窗口内的概率较大，则时间窗口不能有效区分真假警报。假设攻击和误报随机均匀分布在测试流量中，设 N_{attack} 为攻击数量，T_{window} 为判定窗口长度，T_{test} 为测试时间长度，则一个误报落在攻击的时间窗口内的概率为

$$p_{\text{hit}}=\frac{\left(N_{\text{attack}}\times T_{\text{window}}\right)}{T_{\text{test}}} \tag{2}$$

设 IDS 的误报率为 p_{fp}，测试流量的非攻击数为 N_{test}，则误报数 N_{fp} 为 $N_{\text{test}}\times p_{\text{fp}}$，根据概率式（2）可知至少有一个误报落在时间窗口的概率为

$$P_{\text{aohit}}=1-(1-P_{\text{hit}})^{N_{\text{fp}}} \tag{3}$$

以 MIT/LL 的 99 年测试的最后两周数据为例（这两周为测试数据，其他周为训练数据），N_{test} 为 500k，T_{window} 为 120 秒，T_{test} 为 864000 秒，N_{attack} 为 201，MIT/LL 没有公布参与评估系统的具体结果，但由于没有系统达到 MIT/LL 期望的 0.01 误报率，P_{fp} 取 0.01 作为理想情况。则没有误报落入攻击的时间窗口的概率为 3e-62，即至少有一个误报落入真实攻击的时间窗口内概率近似于 100%。通过仿真计算平均有 140 个误报落在攻击的时间窗口内，相对于攻击的数量，尽管有时间窗口，误报对结果判定仍会产生相当大的干扰。

2) 扩展性

MIT/LL 采用绝对时间来描述攻击发生的时间点，并用绝对时间作为报警的标准。实际测试的时候，在调整了原始数据的播放速率或被测系统记录攻击的绝对时间情况下，被测系统检出攻击的相对位置在时间轴上发生了变化，无法直接对比攻击检出的时间和原始数据中的攻击时间，必须对两个时间坐标轴进行调整，将两个时间坐标映射到同一个时间坐标轴上进行比较。调整的依据是在相同的流量速率下，标准答案中攻击出现的相对时刻和被测系统检测出攻击的相对时刻相等。由于被测系统从检测出攻击到记录攻击的延时，当时间轴调整后依靠时间来评分会对准确性造成影响。

另一方面，MIT/LL 为了避免时间窗口的冲突，攻击间的间隔必须大于 2 分钟。但不同的网络环境下攻击的密度是不同的。MIT/LL 的工作开展在 10 年前，面向低带宽环境，这种假设是合理的。但在当前的高带宽环境中，攻击密度已显著增加。例如据观察，CERNET 某省网边界平均每天的警报数是

35 万，某个重点大学校园网边界每天平均是 4.5 万，使用低于 2 个/分钟的攻击速率对于测试资源是浪费的，也是不真实的。如果缩小判定的时间窗口，则在高速网络环境下，由于系统时钟精度的有限性，不同攻击的时间会出现重合，时间窗口由于无法避免的冲突而导致失效。

3　基于同步点的评分方法改进

根据第二节的分析，不考虑误报的评分方法只有在攻击比率高的情况下适用，而在攻击比率低的情况下，目前使用时间窗口的评分方法的准确性和扩展性都不能满足评分需求。为了满足评分过程的准确性和扩展性要求，需要采用其他判定窗口替代时间。本节首先讨论了判定窗口必须具有的性质，然后定义同步点作为一种新的判定窗口，并证明了同步点满足本节所提出的所有性质。

3.1　性质

根据式(3)，p_{aohit} 越大，误报对评分结果的影响越大。合成式(2)和式(3)为式(4)，可以发现在攻击数和误报数固定的情况下，p_{aohit} 和 T_{window} 以及 T_{test} 相关。

$$p_{\text{aohit}} = 1 - \left(1 - \frac{N_{\text{attack}} \times T_{\text{window}}}{T_{\text{test}}}\right)^{n_{\text{fp}}} \tag{4}$$

根据式(4)分别固定 T_{test} 和 T_{window}，对攻击和误报在测试中的分布采用独立的均匀分布进行模拟计算，可以得出图 6 和图 7 的结果。

图 6 是在判定窗口恒为 60s 时落在判定窗口内的误警个数随测试区间长度 T_{test} 变化的趋势，当测试区间长度为 1e6s 时，落在判定窗口内的误警个数为 10，对评分结果有显著影响。随着测试区间长度的增加，落在判定窗口内的误警个数下降，当测试区间长度大于 1e7s 时，落在判定窗口内的误警个数小于 1，属于可接受的范围。图 7 是测试区间长度恒为 1e6s 时，落在判定窗口内的误警个数随判定窗口大小 T_{window} 变化的趋势。当判定窗口长度为 2s 时，错落在判定窗口内的误警个数小于 1，对评分结果没有显著影响；随着判定窗口的增大，落在判定窗口内的误警个数随之增大；当判定窗口为 256s 时，落在判定窗口内的误警个数达到 56，评分结果不可接受。以上的模拟基于对攻击和误报在测试区间中的分布采用相互独立的均匀分布，而实际的测试流量中误报和攻击的出现往往相关，实际的判定结果的误报的影响要大于上述模拟结果。

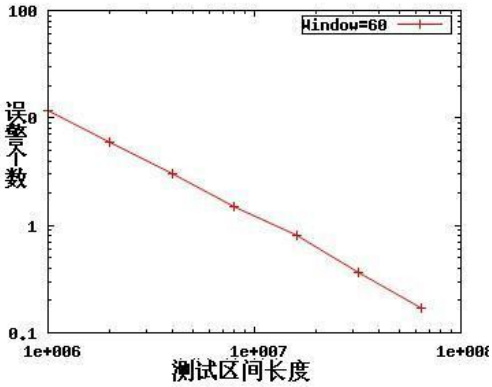


图 6　测试区间长度对误判警报数影响

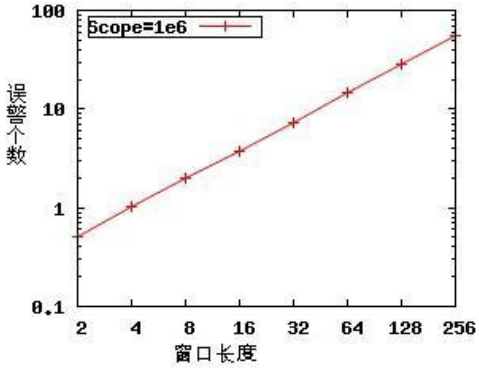


图 7　判定窗口长度对误判警报数影响

从图 6、图 7 的分析可知，如果要获得准确的结果，要么扩大测试区间，要么缩小判定窗口。扩大测试区间意味着测试成本的增加，扩大一倍的窗口就必须增加一倍的测试流量，和测试时间，以 MIT/LL 的测试为例，MIT/LL 在 99 年测试中提供了 7 周的数据，如果窗口扩大一倍，就必须提供 14 周的测试数据，其代价对于评估过程可能是不可接受的，所以，评分方法选择的判定窗口必须能在缩小窗口长度的同时而不影响判定的准确性。

根据 2.2 节中扩展性的分析，判定窗口还必须能适应高速流量测试和流量速率变形测试的需求，综上所述，判定窗口必须具有以下属性：

1) 缩小判定窗口不影响判定的准确性；

2) 在高速流量下保持判定的准确性；

3) 流量的速率可以实现变形而不影响判定方法的实现；

4) 判定窗口对 IDS 的性能影响很小，可以忽略不计；

5) 判定窗口对 IDS 的检测结果不产生影响。

基于时间的方法虽然有性质 4 和性质 5，但无法保证性质 1~3 的实现，所以必须定义一种新的满

足性质 1~5 的判定窗口的方法。

3.2 同步点

IDS 对报文的处理基于 FIFO 队列，这意味着不论报文以低速还是高速或者以不同的速率播放，其相对于 IDS 处理的顺序是不变的。如果可以在攻击的前后加入标识信号，其在 IDS 的警报中出现的顺序和在测试中播放的顺序是一致的。如果对每个信号进行唯一标识，并在 IDS 的警报中能反映出该标识，那么这种信号机制可以作为一个判定窗口来进行作为评分过程的一个标准，且这种判定窗口可以满足 3.1 所提出的性质 1-5。本文中这种信号被定义为同步点。

同步点 S 是一个有限的信号序列 $\{s_1, s_2, \mathbf{L}, s_n\}$，相邻序号的同步点代表了一个判定窗口，每个判定窗口的长度是相等的。则攻击 a_i 必在某个判定窗口 $(s_i, s_{i+1})\ \mathbf{L}\ 0 \leqslant i \leqslant n-1$ 中，判定窗口的长度取决于同步点的个数，如果缩小判定窗口长度，则可以相应增加同步点个数,例如当将判定窗口缩小为原来的一半,则同步点序列 $\{s_1, s_2, \mathbf{L}, s_n\}$ 扩展为 $\{s_1, s_2, \mathbf{L}, s_{2n}\}$，则攻击 a_i 必可以用新的判定窗口 (s_{2i}, s_{2i+1}) 和 (s_{2i+1}, s_{2i+2}) 中的一个来判定，从而保证缩小判定窗口不影响判定的准确性，所以同步点满足性质 1 的要求。

由于同步点的序号代表的是警报的相对位置关系，若攻击 a_i 必在某个判定窗口 $(s_i, s_{i+1})\ \mathbf{L}\ 0 \leqslant i \leqslant n-1$ 中，则在 IDS 的警报中同步点和警报将总以 s_i, a_i, s_{i+1} 的顺序出现，这种顺序关系是不随测试的速率而改变的，所以同步点也同时具有性质 2 和性质 3。

虽然同步点的数量在定义上与测试区间的长度成比例关系，但是由于攻击比率低，有大量的判定窗口 (s_i, s_{i+1}) 不包含任何攻击，在实现中可以省略不包含任何攻击的判定窗口，仅在评估中加入包含攻击的判定窗口，通过同步点的序号 i 同步判定窗口的位置。所以同步点的数量 N_s 实际与攻击的数量成 N_a 成比例关系，由于同步点的应用环境是攻击比率较低的环境，所以同步点的数量远少于背景流量的非攻击报文，对 IDS 的测试性能影响也很小，满足性质 4 的要求。在第 5 节中将通过试验来比较增加同步点与否对 IDS 的性能影响。

同步点虽然作为一种警报在 IDS 中出现，但是同步点作为一种判定信号，其作用是判定警报的位置是否与攻击发生的位置相符合，可以通过定制同步点的内容防止 IDS 对同步点产生误报。所以同步点也满足性质 5 的要求。

4 基于同步点的评分方法实现

4.1 同步点的设计

根据第 3 节的讨论，同步点是一种可被 IDS 准确识别的有限序列信号，通过信号的定位实现评估可接受的判定窗口。同步点可以通过包含特征的报文实现。由于 IDS 对报文的处理是顺序的，所以信号报文和攻击报文出现在 IDS 检测结果中的顺序和它们在流量中的顺序是一致的。而报文的顺序和测试速率无关，所以测试速率的改变不会改变判定窗口和攻击的相对关系。按报文顺序定义的窗口精度也远高于时间窗口。以千兆测试流量为例，其每秒报文数应该在 200k 的数量级，即使时间窗口精确到毫秒级，每个时间窗口内也平均有 200 个报文，而这样的精度由于测试环境的误差是不可接受的。以报文信号来定义窗口，则窗口精度最小可以为一个报文。综上所述，报文可以作为同步点实现的有效方式。

作为同步点实现的报文包含 2 种特征：信号特征和空间顺序特征。前者保证同步点报文可以被准确的识别并不会影响原有测试流量的检测结果，后者实现判定窗口。

信号特征是 IDS 检测域 $\{f_1, f_2, \mathbf{L}, f_n\}$ 上的一组约束 $\{F_1, F_2, \mathbf{L}, F_n\}$。理论上只要满足性质 5，信号特征可以使用任何检测域的组合，但是不同检测域对检测过程的代价不同，由于性质 4 的要求，信号特征应避免选择检测代价大的检测域，如定义在负载部分的特征使用字符串匹配算法进行检测，定义在协议头部的特征可以使用分类算法进行检测，信号特征应尽量选择协议头部的检测域。

空间顺序特征是 IDS 警报给出信息域的一个子集 $\{a_1, a_2, \mathbf{L}, a_n\}$，每个信息域都可以被顺序编码。顺序特征的域选择在满足判定窗口的要求同时，需要注意避免对性质 4 的影响，例如选择报文的 IP 或端口字段作为信息域进行顺序编码，容易产生类似扫描的行为，导致 IDS 对同步点报文的序列做出误报，造成检测结果的偏差。同时信息域必须保证足够的空间，以满足高速流量测试的需求，例如只选择报文的 TOS 字段作为信息域，由于 TOS 字段

的空间限制，只能实现 256 个同步点，最大形成 255 个判定窗口，当攻击所处的判定窗口大于 255 时，同步点会失效。

所以同步点可以定义如下：

$$S=\{Sig=\{F_1,F_2,\mathbf{L}\,F_n\},Seq=\{a_1,a_2,\mathbf{L}\,a_n\}\}$$

代入具体的检测域和信息域的约束后，就可以得到同步点的实现，例如：

$$S=\{Sig=\{ip,ipproto=111,iplen=66\},Seq=\{ipid\}\}$$

该同步点定义同步点使用 IP 协议，信号特征选择了 IP 协议和长度两个头部字段，保证了低检测代价，其中 IP 协议字段值为 111，IP 报文长度为 66 字节，空间顺序编码使用 IPID 字段，该字段一般不作为检测域，同时该字段的空间为 2^32，可以充分满足判定的窗口数量需求。

4.2　基于同步点的评分方法

同步点的算法包括两个部分：生成算法和匹配算法。

基于同步点的测试流量生成算法如下：

1) 定义同步点的属性，并对 IDS 进行测试，检查是否能正确报出同步点攻击。

2) 扫描测试流量，确定包含攻击的同步点判定区间序列。

3) 扫描测试流量，在相应位置插入同步点。

4) 生成测试的标准答案。

基于同步点的测试结果比较算法：

1) 遍历标准答案，对每一对定位同步点间的攻击进行第 2 步和第 3 步的检查，标准答案遍历完毕，跳至第 4 步。

2) 首先查找警报文件中是否存在对应的同步点，如果警报中同步点缺失，则为漏报。

3) 如果对应的定位同步点存在，检查警报中对应的定位同步点间是否存在和标准答案文件信息一致的攻击，如果不存在，则为漏报；如果存在，则为正确的警报。

4) 遍历警报文件，对每一对定位同步点间的攻击，进行第 5 步的检查，对所有不位于定位同步点间的攻击，都定义为误报。遍历完成后，跳至第 6 步。

5) 查找标准答案中对应的定位同步点间是否有和警报信息一致的攻击答案，如果有，为正确警报，记录该警报前最近的一个计时同步点的顺序编号，如果没有，则为误报。

6) 根据统计的正确的警报数和误报数，计算 IDS 的检测率和误报率。根据正确警报附带的计时同步点信息，比较对不同攻击的检测及时性。

5　试验

本节对 MIT/LL（基于时间的）评估方法和基于同步点的评估方法进行了对比，并测试了同步点对被测系统的性能影响。测试的背景流量选用 MIT Lincoln 1999 年测试数据中第三周第一天的 inside.tcpdump，根据 MIT/LL 的说明，该流量不含任何攻击。攻击流量和同步点的生成，以及和背景流量的混合通过 IDS 测试平台 AOLES 完成。被测 IDS 选用 Snort 2.6.14[11]，采用 Snort 的缺省配置和规则，规则库版本为 2007 年 3 月 22 日版本，同时增加同步点的规则到 Snort 检测的规则集中测试平台采用两台使用 1G 以太网直连的测试机器 A 和 B，测试流量从测试机器 A 向测试机器 B 播放，被测 IDS 配置在测试机器 B 中。

5.1　基于时间 vs.基于同步点

为了对比基于时间和基于同步点的评分方法，基于 inside. tcpdump 的背景流量数据，不同数量的攻击数据和背景流量中的误报在判定窗口内的冲突数分别进行了比较。根据 MIT/LL 的说明，inside. tcpdump 不包含任何攻击，所以背景流量被 Snort 检测出的任何攻击都可以看成是误报。缺省配置下 Snort 共会产生 252 个误报。背景流量的时间为 22h，共 1 492 231 个报文。攻击的数量分别采用 100，500，1 000 三组值，均以均匀随机概率分布在背景流量中。基于时间的方法采用了 10s 和 60s 两种判定窗口（tl10 数据和 tl60 数据），基于同步点的判定窗口采用 10 报文和 60 报文两种判定窗口（sp10 数据和 sp60 数据），在不同攻击数量下根据不同方法和不同判定窗口大小内得到的冲突数量如图 8 所示。

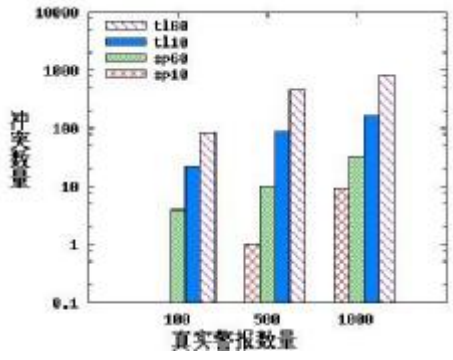


图 8　基于时间 vs 基于同步点的冲突率

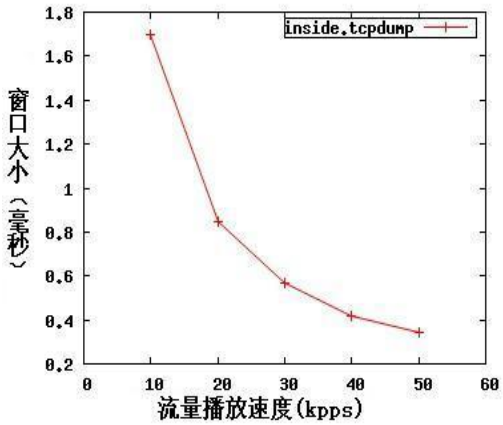


图 9　判定窗口大小 vs 测试流量速率

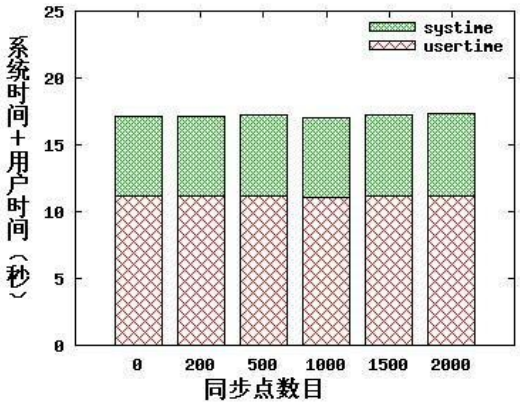


图 10　同步点对 IDS 检测性能影响

在攻击数量为 100 的情况下，采用判定窗口为 10 的同步点方法的冲突数量为 0，所以图 8 中第一组数据没有 sp10 的值。从图 8 可以看出，采用同步点方法的冲突数量远小于采用时间窗口的方法，图 9 是 inside. tcpdump 在不同的测试速率下，避免误报落在判定窗口内的最大时间窗口变化趋势。随着测试速率不断提高，时间窗口从 1.7ms 下降到 0.3ms，说明在高速测试流量下，时间判定窗口的精度不足以准确评分。

5.2　性能试验

为了说明同步点在评估过程中对 IDS 性能造成的影响，不同数量等级的同步点被加入背景流量中，通过比较 IDS 离线处理测试流量文件的时间，根据第 2 节的分析，当攻击比例较大时，可以忽略误报的影响，此时可以不需要同步点进行判定，所以实验的同步点数量相对于背景流量的报文数量控制在 1:1e4 以下。实验所用流量报文的数量级在 1e7，所以同步点的数量级取在 1e2 到 1e3 之间。实验中通过 Linux 系统的 Time 程序获得 Snort 处理流量所消耗的系统时间 (systime) 和用户时间 (usertime)，首先记录 Snort 处理纯背景流量文件所消耗的时间，然后依次记录 Snort 处理加入不同同步点的流量文件所消耗的时间，每组实验进行 10 次取均值以降低程序运行过程中的随机因素对试验结果的影响。图 10 为 2 种时间在不同的同步点数量级下的对比。

从图 10 可以看出，不同的同步点数量对 IDS 分析流量文件的时间几乎没有影响，说明当同步点和背景流量的比率较低（即攻击比率较低）情况下，同步点对 IDS 的性能没有明显影响，对评估过程的影响可忽略不计。

6　结束语

不考虑误报的评分方法只有在攻击比率极大（大于 1e-3）或误报率极低（小于 1e-6）的情况下才接近真实结果，而真实环境的攻击比率都远低于 1e-3 的理想值，所以不考虑误报的评分方法不能满足模拟真实环境下的 IDS 评估需求。基于时间窗口的考虑误报的评分方法虽然能提高评分的准确性，但由于时间测量的精度限制，误报和真实警报在一个时间窗口中仍然存在着较大的冲突概率，并且只适用于百兆以下的慢速网络环境中的 IDS 评估，无法适应目前普遍是千兆并向万兆速度发展的网络环境。本文通过对上述评分方法的分析，提出了评分方法的判定窗口满足准确性和扩展性需求必须的性质，并基于 IDS 对报文的处理是 FIFO 队列的特性，定义了基于同步点的评分方法。由于同步点方法利用报文的相对顺序为判定依据，判定结果不受报文速率的影响，无论千兆或万兆网络环境下都可以适用，满足了评分方法的扩展性需求。同时同步点的最小测量单位为报文，可以实现高精度的判定窗口，使误报和真实警报在判定窗口内冲突的概率几乎为 0。基于同步点的评分方法相对于目前的考虑误报和不考虑误报的评分方法可以提供更准确的评分结果，使 IDS 评估的结论有更高的可信度。

基于同步点的评分方法依赖于 IDS 对报文的处理是 FIFO 队列的特性，而随着流量规模的不断扩大，分布式 IDS 开始不断出现，尽管在分布式 IDS 每一个节点内部对报文处理仍然遵循 FIFO 队列的特性，但是不同节点之间不存在 FIFO 的特性，作者的下一步工作，是研究如何将基于同步点的评分方法扩展到分布式 IDS 的评估中去。

参考文献：

[1] NSS Group. Intrusion Detection Systems Group Test (Edition 4)[R]. NSS Group, 2004.

[2] HAINES J, LIPPMANN R, FRIED D. Design and Procedures of the 1999 DARPA Intrusion Detection Evaluation: Design and Procedures[R]. MIT Lincoln Laboratory, 2001.

[3] MCHUGH J. Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory[J]. ACM Transactions on Information and System Security, 2000, 3 (4):262-294.

[4] MELL P, HU V, LIPPMANN R. An Overview of Issues in Testing Intrusion Detection Systems[R]. National Institute of Standards and Technology ITL, 2003.

[5] MUTZ D, VIGNA G, KEMMERER R. An experience developing an ids stimulator for the black-box testing of network intrusion detection systems[A]. Proceedings of the 19th Annual Computer Security Applications Conference[C]. Las Vegas, Nevada, USA, 2003. 374- 383.

[6] 陆晟: 基于规则的高速网络入侵检测[D]. 东南大学, 2003.
LU S. Rule-based Intrusion Detection for High-Speed Network[D]. Southeast University. 2003.

[7] SOMMERS J, YEGNESWARAN V, BARFORD P. Toward Comprehensive Traffic Generation for Online IDS Evaluation[R]. UW Technical Report, August, 2005.

[8] AXELSSON S. The Base-Rate Fallacy and its Implications for the Difficulty of Intrusion Detection[A]. Proceedings of the 6th ACM conference on Computer and communications security[C]. Singapore, 1999. 1-7.

[9] 卿斯汉, 蒋建春, 马恒太. 入侵检测技术研究综述[J]. 通信学报, 2004, 25 (7):19-29.
QING S H, JIANG J C, MA H T. Research on intrusion detection techniques: a survey[J]. Journal of Communications, 2004, 25(7): 19-29

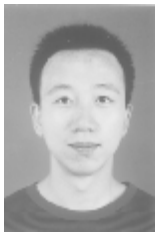
[10] MAHONEY M V, CHAN P K. An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection[A]. Proceedings of the Sixth International Symposium on Recent Advances in Intrusion Detection[C]. Pittsburgh, PA, USA, 2003. 220-236.

[11] Snort 2.6.14[EB/OL] http://www.snort.org/. 2007.

作者简介：

杨望（1979-），男，安徽宣城人，东南大学博士生，主要研究方向为入侵检测系统的测试与评估。

龚俭（1957-），男，上海人，博士，东南大学教授、博士生导师，主要研究方向为网络体系结构、网络安全和网络行为学。



吴雄（1980-），男，江苏常州人，东南大学硕士生，主要研究方向为入侵检测系统的测试和评估。