信息融合技术在入侵检测系统中的应用

Application of Information Fusion Technology on Intrusion Detection System

摘　要：本文将入侵检测技术和信息融合技术相结合，针对大型异构网络提出了基于Internet的IDS即Cyber IDS的概念，给出了Cyber IDS的体系结构和相关融合问题，并提出了用于实现Cyber IDS的相关融合算法，揭示了新一代IDS的发展趋势和特点。

关键词：入侵检测　信息融合　代理　多代理系统

姜建国（四川大学数学学院　中国工程物理研究院）
周仲义（四川大学数学学院）

1 概述

1.1 基本概念

信息融合是一种多层次的、多方面的处理过程，这个过程是对多源数据进行检测、结合、相关、估计和组合以达到精确的状态估计和身份估计，以及完整、及时的态势评估和威胁估计。

信息融合的另一种普遍说法是数据融合，但就信息和数据的内涵而论，用信息融合一词更广泛、更确切、更合理、更具概括性。

将多传感器信息融合技术应用于入侵检测和网络监控，就是需要对来自大型异构网环境中的分布传感器的数据进行融合处理。下一代的网络管理和IDS将基于一个统一的协作模型相互通信和合作，将数据融合成信息和知识，使得能够对Cyberspace的网络状况和安全形势作出明确的判断。

1.2 Cyber

IDS：新一代的入侵检测系统

由于传统的IDS一般局限于单一的主机或网络架构，对异构系统及大规模网络的监测明显不足，同时不同的IDS系统之间不能协同工作，为了实现Internet环境下的入侵检测和安全响应，就需要设计和开发一个基于因特网环境的IDS即Cyber IDS，这种新一代IDS必须能够自动鉴别和追踪网络空间中动态的网络活动，从而可以监控网络空间中的各种攻击。显然，如何建立一个有效的检测体系，以适应网络规模化、高速化的发展；如何有效地融合来自各地异构网中的采集器、系统管理器及数据库的数据、威胁和环境信息，以得到关于网络活动的一致的、更好的安全估计。这些都是技术关键中最具挑战性的问题。

现行入侵检测系统大多数都采用单一体系结构，即所有的工作包括数据采集、分析都是由单一主机上的单一程序完成的，而一些分布式的入侵检测系统的体系结构采用分布集中处理，只是在数据采集上实现了分布式，数据的分析、入侵的发现还是由单个程序完成。在目前入侵检测技术的研究中，一方面在检测技术上针对越来越复杂的攻击方法，如何提高检测能力；另一方面，利用Agent技术在检测系统结构设计上，实现对大型网络、高速千兆网、分布异构平台环境的适应。此外，利用多传感器信息融合技术在分布式入侵检测系统中，实现多层次、多方面的信息处理，以达到对网络安全状况的监控和评估，这方面的研究相对较少。

2 Cyber IDS的结构与相关融合问题

2.1 层次分布式结构

图1、图2所示为Cyber IDS的一种结构，系统由分布于网络中的多个功能代理组成，代理之间既可以独立工作，又相互协作，整个系统形成一个层次体系。每台主机上可以有多个独立运行的具有不同功能的代理，代理之间相互通信、协作，底层的代理向上级代理提交报告并接受上级代理的控制。所有代理按功能可分为三种：入侵检测代理（IDA）、通信服务代理（TSA）、管理控制代理（MCA）。

注意这里的入侵检测代理（IDA）监控包括基于主机的和基于网络的所有活动，它既可以只是一个监视一个特定系统变量或事件的简单程序（如统计最后5秒里TELNET连接的次数），也可以是一个复杂的软件系统（如

基于主机或基于网络的IDS），只要它产生的输出能够被TSA接受。

通信服务代理（TSA/TA）：每个主机的外部通信接口，主要有控制和数据处理两大功能。IDS中的每个主机都必需有TSA，向下控制本机的活动监控代理，向上则可与多个管理控制代理通信。

管理控制代理（MCA/MA）：处于系统结构中的最高层，它们具有类似TSA的控制和处理功能，其主要区别在于MCA可以控制网络中的多个TSA，而TSA只能控制本机的入侵检测代理。

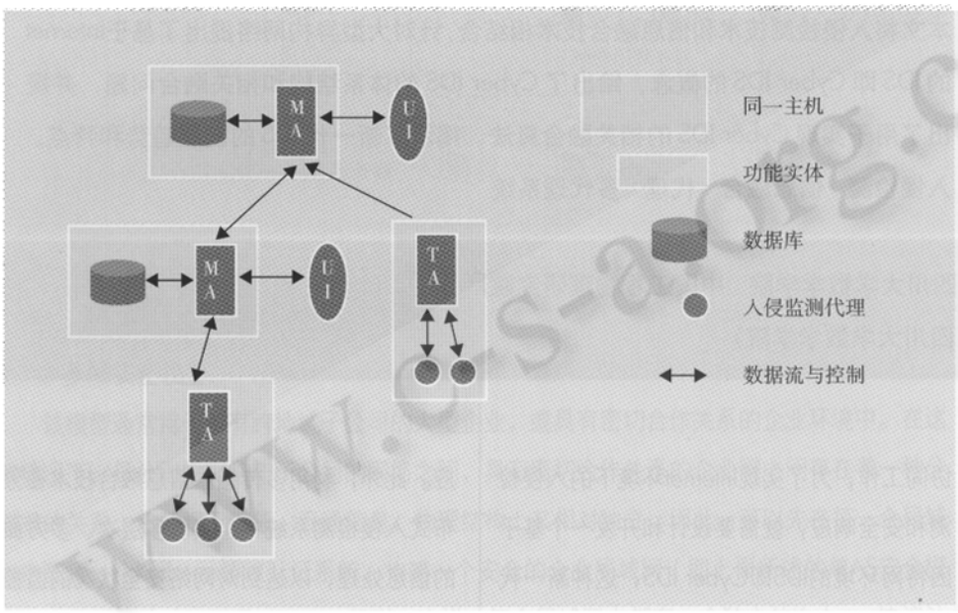

图 1　系统物理结构

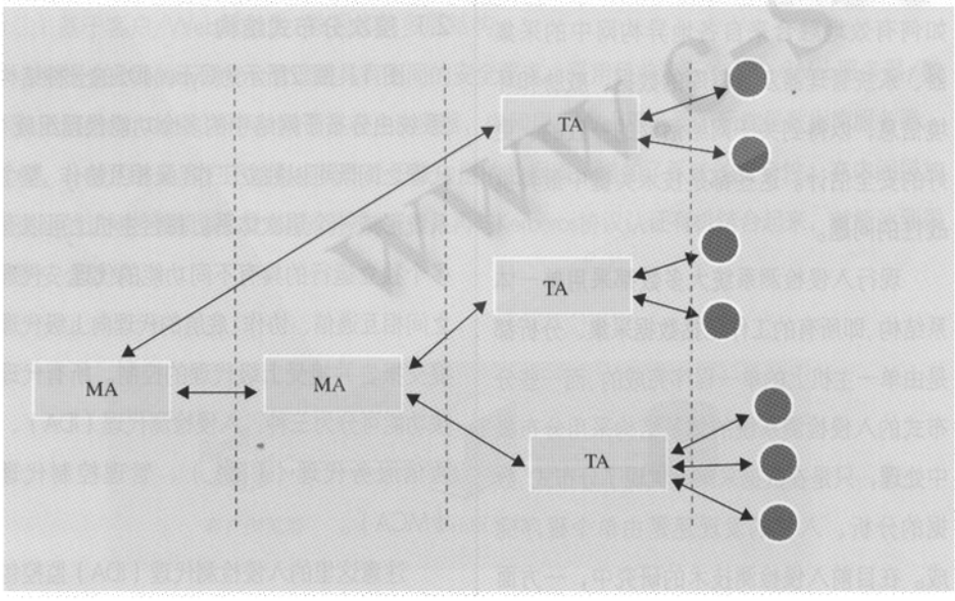

图 2　系统逻辑分层结构

注意设计Cyber IDS体系结构时需要考虑其应具备的主要特点：

（1）可伸缩性：可不断扩大规模而不显著增加网络流量。在一般分布式系统中，随着单个IDS代理数量的增加，各代理之间的相互联系也大大增加，控制程序因此越来越复杂。Cyber IDS的层次式体系结构充分考虑了这一点，通过隔离每一层必需的相关处理，减少了复杂性；将大量的低层实时处理与较少的复杂高层处理分离开，使每一层处理时间近似相等。同时每一层又可以实行并行处理，使得层数的增加不会引起复杂性的指数增长；而且通过使用网络环境的先验知识，每一层的处理量还可以进一步地减少。

（2）鲁棒性：不存在单一失败点，抗攻击性强。由于IDS代理可以与多个上层代理通信，某个IDS代理的失效只影响其所在的网络段的监控，并不会影响Cyber IDS的整体功能；

（3）智能性：IDS代理具有一定的自适应、自学习功能；

（4）互操作性：IDS代理间可以按照一定的控制关系相互通信、协作。

2.2 相关融合问题

下面以Cyber IDS 的树形分层结构来说明相关融合问题。

（1）入侵检测代理（IDA）内部之间的数据融合。IDA的主要功能就是收集活动在网络和主机的事件或对象的数据，并根据入侵检测规则进行判断。基本的入侵检测方法主要有两种，即异常发现技术和模式发现技术。

无论哪种检测方法，都需要收集大量的事件数据，这些原始数据需要汇集在一起进行校准和提炼，以便作出最优判断，如图3所示。

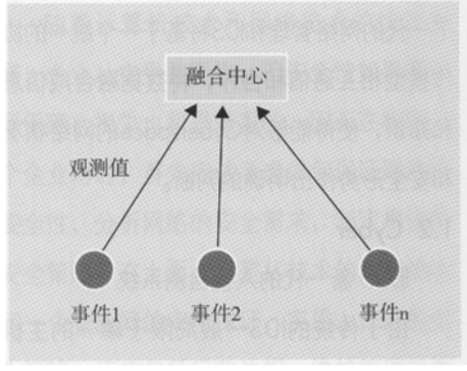

图 3　最优判断

对于同时采用了多种方法进行检测的IDA，同样存在融合问题：如何融合不同检测方法得到的结果，以便作出更好的判断。

（2）各个IDA在通信服务代理（TSA）的信息融合来自多个IDA的报警数据量可能相当大，其中同一个攻击的许多相关部分可能会分布在不同的IDA报警数据中，如何融合由多个IDA产生的报警，进行相关、关联处理，以便作出融合判断。如图4所示。

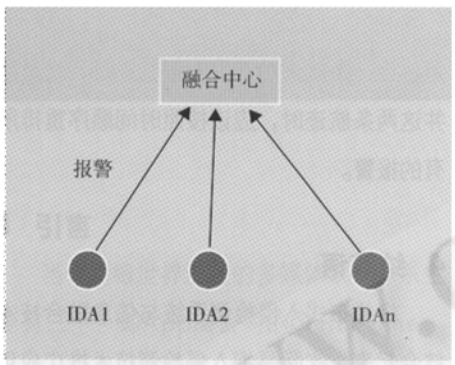

图4 融合判断

需要结合由多个IDA产生的报警并形成航迹，每个航迹用于体现某个网络中的实体或组织所采取的一系列动作，简单说就是具有一个共同属性的报警集。对每一个新到报警，需要与以前的报警形成的多个航迹进行比较，决定新报警是否归属某个航迹，否则开始一个新的航迹。

（3）各个TSA在管理控制代理（MCA）的信息融合类似于IDA在TSA的融合，多个TSA产生的航迹需要上传到MCA作进一步的融合处理，多条航迹之间也可能存在大量的相关部分，需要进行航迹与航迹的关联处理，通过结合和合并，形成更大范围的能反映网络状况的新航迹。

显然，航迹将所有的报警分组并提供不同的报警集组构成的反映网络状况的多层视图，从而使网管员能够集中注意力，更容易发现破坏或DOS行为，不至于被大量的报警数据所淹没。同时还可以将错误报警概率分配到整个航迹而非个别报警，使得在增加相关IDA的感应能力（或灵敏度）的同时减少误报率。

3 设计与实现

在上述介绍的通信控制代理中（TSA和MA），除了建立树型层次结构所必须的通信控制功能外，还有一个共同的处理功能，即数据分析处理模块。其主要功能就是报警的聚类和相关处理（Aggregation and Correlation Components，简称ACC），涉及到报警-报警、报警-航迹、航迹-航迹等方面的相关处理。

每一个ACC从IDA或其他ACC（下级TSA/MA）得到报警，每当收到一个报警，执行两个任务，一是基于ACC的航迹数据库分析接受到的报警（航迹数据库是由早先收到的报警和相关配置信息构成的），再就是提供本地输出，这样每一个ACC无论处于哪一层，都可由用户查询。

3.1 报警相似度

由于在信息融合的过程中，报警-报警、报警-航迹、航迹-航迹等的关联分析最终都可以归结到报警与报警的比较，如何定义两条报警是相同的或相似的，由此导致报警相似度的概念，即两条报警相似的程度或相似的概率。

3.1.1 属性相似度、入侵检测系统中的应用

要考虑两条报警的相似度，首先要考虑的是这两条报警的共有属性，这些属性包括攻击源、目标（主机和端口）、攻击类型、和时间信息。对每一种属性，需要定义一个在0和1之间的相似度函数，1表示属性的完全匹配。

考虑攻击类型的相似度，系统需要维护一个关于各类攻击类型的相似度矩阵，其对角线上的值全为1，而非对角线位置上的值则表示对应攻击类型的相似度。

考虑IP地址的相似度，主要因为攻击者可能从不同但又合法的IP地址上发起攻击，这些IP地址大都位于同一子网。定义因子r表示IPV4子网中两个IP地址异或取反后较高位中1的个数，因此，

$r = 32$ 表示两个IP地址完全相同，

$r = 0$ 表示两个IP地址完全不同，

$0 < r < 32$ 表示两个IP地址中高位顺序r个位是相同的。

R_{ij} = 表示i、j两个IP地址的相似度。

3.1.2 期望相似度

当考虑两个报警的相关性时，期望相似度可以用来表示应该优先考虑匹配的属性，不同的攻击类别有不同的属性期望相似度。例如，来自同一主机的探测攻击可能扫描同一个子网上的不同机器的相同端口集，实际可能是同一类攻击，所以此时对目标IP地址匹配的期望值可以较低；同样，类似SYN FLOOD之类的攻击假冒源地址，所以只要其他属性匹配，两条报警就是相关的，而不需要考虑源地址是否匹配，此时对源地址匹配的期望值可以很低。

显然，期望相似度可以在进行报警属性相似度匹配时，有效地作为加权系数来调节。

3.1.3 最小相似度

对于每一类攻击，需要定义各个属性的最小相似度，以作为报警相关的必要但非充分的条件。当进行报警间的融合处理时，有些属性参数需要严格匹配，而有些参数只需要近似匹配，前者的最小相似度必须为1，后者的最小相似度则可以小于1但必须大于0。

如果存在相对应的属性参数匹配的相似度小于该属性的最小相似度，则可以确定进行相关处理的两条报警的总相似度为0，即这两条报警没有关系；否则，总相似度是对应属性的相似度的加权平均，各个属性的期望相似度分别用作加权系数。

3.1.4 属性的融合-报警相似度

当系统进行报警的融合处理时，需要计算两条报警的相似度，如上所述这可以通过计算各个属性相似度的加权平均得到。当然，其必要条件必须是所有对应属性参数的

相似度都超过对应的最小相似度。可以计算两条报警的总相似度如下：

$$SIM(X,Y) = \sum_j E_j SIM(X_j,Y_j) / \sum_j E_j$$

X = 备选航迹中的报警

Y = 新报警

j = 报警属性项索引

E_j = 属性j的期望相似度

X_j，Y_j = 分别对应报警X和Y中属性j的值（可能为列表值）

3.2 融合算法

3.2.1 报警预处理

预处理主要用于统一规范一条报警中的信息，并包括相关错误检查功能，以确认报警没有包含明显错误的信息，如一个无效的时间戳。

由于网络类IDA和主机类IDA对一个攻击的源和目标的表示方法不一样，如网络类IDA很可能用IP地址表示，而主机类IDA则用主机名表示，为确保唯一性表示，预处理将报警中的信息转换成用一个唯一确定的标识符来分别表示IDA的名称、源和目标。这种转换是通过预先定义的配置文件来完成的。

3.2.2 形成航迹并存入数据库

该部分对报警进行相关处理，将具有一定共同属性的逻辑相关的报警组合成航迹，并形成航迹数据库，具体算法如下。对每一个新到报警，执行如下过程：

（1）首先判断航迹数据库是否为空？若是，则用新报警开始一条新航迹，并存入数据库；否则进行下一步；

（2）对数据库中的每一条航迹，取其最近加入的报警，计算新旧两条报警的相似度s，保留s始终为较大值直至处理完所有的航迹。进行下一步；

（3）判断s值是否超过相关阀值，若是，则将新报警加入对应的航迹；否则，用新报警开始一条新航迹。

（4）再读下一条新报警，重复上述步骤。

3.2.3 航迹相关和聚类分析

该部分需要对航迹进行分类表示，根据以上对航迹类的定义，分别对每一条航迹进行聚类分析，并赋以相应的标识符。根据这些聚类信息，可以更好地评估和判断当前的网络状况。

航迹相关是指针对来源不同的航迹进行航迹-航迹相关分析，希望合并相关航迹，形成更高一级的航迹，体现更大范围的网络状况。实际上，航迹-航迹关联可以转化为报警-航迹关联，即取一条航迹的代表-最新加入该航迹的报警，则该报警与另一条航迹的关联度就是这两条航迹的关联度。只是需要注意，当一条航迹加入另一条航迹时，就是合并这两条航迹时，应该按照时间顺序重排所有的报警。

4 结束语

将分布式入侵检测系统与信息融合技术结合起来，实际是把入侵检测技术推广应用到整个Internet环境中，由于网络规模和结构发生了质的变化，因而Cyber IDS与现时IDS相比，体系结构和数据融合将显得更加重要，新一代入侵检测系统的更高目标将是构造整个网络的一个清晰的感知空间，并能及时发现、反映和消除网络中的异常活动。

参 考 文 献

1 Bass, T. "Intrusion Detection Systems and Multisensor Data Fusion: Creating Cyberspace Situational Awareness." Communications of the ACM. Forthcoming, 1999.

2 Waltz E, Llinas J. 多传感器数据融合，宗贵等译，电子部 28 所。

3 Bass, T. "Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems." 1999 IRIS National Symposium on Sensor and Data Fusion, May 1999.供本地输出，这样每一个ACC无论处于哪一层，都可由用户查询。