网络入侵检测系统的设计与实现

第 1 章　概　　述

入侵检测系统（IDS, Intrusion Detection System）用来识别针对计算机系统和网络系统，或者更广泛意义上的信息系统的非法攻击，包括检测外界非法入侵者的恶意攻击或试探，以及内部合法用户的超越使用权限的非法行动。使用 IDS 的目的各有不同。有的人是对法律方面的事务感兴趣，包括对入侵者的跟踪、定位和起诉，而有些人使用 IDS 是为了保护自己重要的计算资源。还有一些使用者则对发现和纠正系统安全漏洞更加感兴趣。

本书中所指的“网络入侵检测系统”是一个相对广泛意义上的概念，包括针对网络基础设施以及其中主机系统的非法攻击行为的检测。

1.1　入侵检测系统的组成部分

从功能逻辑上讲，入侵检测系统由探测器（Sensor）、分析器（Analyzer）和用户接口（User Interface）组成。下面分别对这 3 大部分进行简要介绍。

1）探测器（Sensor）

探测器主要负责收集数据。探测器的输入数据流包括任何可能包含入侵行为线索的系统数据，比如说网络数据包、日志文件和系统调用记录等。探测器将这些数据收集起来，然后发送到分析器进行处理。

2）分析器（Analyzer）

分析器又可称为检测引擎（Detection Engine），它负责从一个或多个探测器处接受信息，并通过分析来确定是否发生了非法入侵活动。分析器组件的输出为标识入侵行为是否发生的指示信号，例如一个警告信号。该指示信号中还可能包括相关的证据信息。另外，分析器组件还能够提供关于可能的反应措施的相关信息。

3）用户接口（User Interface）

IDS 的用户接口使得用户易于观察系统的输出信号，并对系统行为进行控制。在某些系统中，用户接口又可称为“管理器”、“控制器”或者“控制台”等。

除了以上 3 个必要组件之外，某些 IDS 可能还包括一个所谓的“蜜罐”（Honeypot）诱饵机。该诱饵机被设计和配置成为具有明显的系统安全漏洞，并对攻击者明显可见。诱饵机能够作为 IDS 中一个专门提供给攻击者进行入侵的探测器来使用，从而提供关于某次攻击行为发生过程的相关信息。

根据检测引擎的实现技术，入侵检测系统可分为“滥用检测（Misuse detection）”和“非规则检测（Anomaly detection）”两种系统。下面将对这两种系统做一个分析和比较。

1.2 滥用入侵检测系统

滥用入侵检测系统的应用是建立在对过去各种已知网络入侵方法和系统缺陷知识的积累之上，它需要首先建立一个包含上述已知信息的数据库，然后在收集到的网络活动信息中寻找与数据库项目匹配相关的蛛丝马迹。当发现符合条件的活动线索后，它就会触发一个警告，这就是说，任何不符合特定匹配条件的活动都将会被认为是合法和可以接受的，哪怕其中包含着隐蔽的入侵行为。因此，滥用入侵检测系统具备较高的检测准确性，但是，它的完整性（即检测全部入侵行为的能力）则取决于其数据库的及时更新程度。

可以看出，滥用入侵检测系统的优点在于具有非常低的虚警率，同时检测的匹配条件可以进行清楚地描述，从而有利于安全管理人员采取清晰明确的预防保护措施。然而，滥用入侵检测系统的一个明显缺陷在于，收集所有已知或已发现攻击行为和系统脆弱性信息的困难性以及及时更新庞大数据库需要耗费大量精力和时间，这是一项艰苦工作。另一个存在的问题是可移植性，因为关于网络攻击的信息绝大多数是与主机的操作系统、软件平台和应用类型密切相关的，因此带来的后果是这样的入侵检测系统只能在某个特定的环境下生效。最后，检测内部用户的滥用权限的活动将变得相当困难，因为通常该种行为并未利用任何系统缺陷。

在滥用入侵检测系统中，研究者们提出基于各种技术类型的检测器，如专家系统（Expert system）技术、特征分析（Signature Analysis）技术、Petri 网络分析、状态转移分析（State-transition analysis）等等。

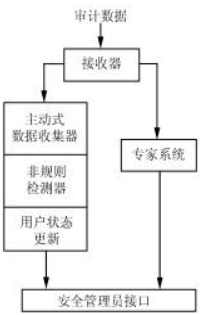


图 1.1 IDES 原型系统的结构

专家系统技术在各种开发模型（Prototypes）中得到广泛应用。通常，专家系统中包含一系列描述攻击行为的规则（Rules），当审计数据事件被转换成为能够被专家系统理解的包含特定警告程度信息的事实（Facts）后，专家系统应用一个推理机（Inference engine）在事实和规则的基础上推理出最后结论。这里，原始的审计数据被抽象成系统能够理解的事实，有利于进一步应用更高层次的各种分析技术。

采用专家系统技术的典型例子有 SRI 公司开发的入侵检测专家系统（IDES，Intrusion Detection Expert System），其系统结构如图 1.1 所示。

由于处理速度的原因，专家系统技术目前只是在各种研究原型中得到应用，而商业化的软件产品采用了其他效率更高的技术，其中目前应用最广泛的就是特征分析技术。与专家系统技术比较，相同之处是同样要收集关于网络入侵行为的各种知识，不同点是特征分析技术更直接地运用收集到的各种知识，例如入侵行为可以被转化成它们在实施过程中所产生的一个事件序列或某种系统审计文件以及网络数据包中的数据样板模型。

1.3 非规则入侵检测系统

非规则入侵检测系统的工作是建立在如下假设基础上的，即任何一种入侵行为都能由于其偏离正常或者所期望的系统和用户的活动规律而被检测出来。描述正常或者合法活动的模型是从对过去通过各种渠道收集到的大量历史活动资料的分析中得出来的。入侵检测系统将它与当前的活动情况进行对比，如果发现了当前状态偏离了正常的模型状态，则系统发出警告信号，这就是说，任何不符合以往活动规律的行为都将被视为入侵行为。因此，非规则入侵检测系统的检测完整性很高，但要保证它具备很高的正确性却很困难。

此类检测技术的优点在于它能够发现任何企图发掘、试探系统最新和未知漏洞的行为，同时在某种程度上，它较少依赖于特定的操作系统环境。另外，对于合法用户超越其权限的违法行为的检测能力大大加强。

较高的虚警概率是此种方法的主要缺陷，因为信息系统所有的正常活动并不一定在学习建模阶段就被全部了解。另外，系统的活动行为是不断变化的，这就需要不断地在线学习。该过程将带来两个可能后果，其一是在此学习阶段，入侵检测系统无法正常工作，否则生成额外的虚假警告信号。还有一种可能性是，在学习阶段，信息系统正遭受着非法的入侵攻击，带来的后果是，入侵检测系统的学习结果中包含了相关入侵行为的信息，这样，系统将无法检测到该种入侵行为。

在非规则入侵检测中，最广泛使用的技术是统计分析（Statistics Analysis）。系统或者用户的当前行为通过按一定时间间隔采样并计算出的一系列参数变量来描述，如每个会话进程的登录和退出时间，占用资源的时间长短及其在每个进程中占用的 CPU—内存—硬盘等资源的多少等。采样的时间间隔从几分钟到一个月，时间长短不等。在最初的模型中，系统计算出所有变量的平均值，然后根据平均偏差检测当前行为是否超过了某一阈值，当然，这样的模型是很简单和粗糙的，无法准确检测异常活动。进一步的算法将单个用户的参数变量数值与积累起来的群体参数变量值进行比较，但是检测能力的提高还是不大。目前在几种非规则检测系统中使用了一种更加复杂的模型，检测系统同时计算并比较每个用户的长期和短期活动状态，而状态信息随着用户行为的变化而不断更新。

另一种主要的非规则检测技术是神经网络技术。神经网络技术通过学习已有的输入—输出矢量对集合，进而抽象出其内在的联系，然后得到新的输入—输出的关系；这种技术在理论上能够用来在审计数据流中检测入侵行为的痕迹。然而，目前尚无可靠的理论能够说明神经网络是如何理解学习范例中的内在关系的，所以同样也无法清楚地解释它是如何发现并理解入侵行为的。神经网络技术和统计分析技术的某些相似之处已经被理论证明，而使用神经网络技术的优势在于它能够以一种更加简洁快速的方式来表示各种状态变量之间的非线性关系，同时，能够自动进行学习/重新训练的过程。

1.4 两种分析技术的比较

本节将从以下 3 个方面来描述上述两种分析技术的不同之点。

1）所需的知识

对入侵行为的检测需要关于可能攻击行为的知识，或者系统已知和期望行为的知识。对于基于滥用检测技术的 IDS，如果要检测到所有攻击行为的话，那么它就需要知道所有可能攻击行为的先验知识。该 IDS 为此必须识别任何攻击行为的细节过程，或者标识该类攻击行为的特征模式。而对于非规则检测系统而言，它必须拥有系统已知和期望行为的所有信息，才能够检测到所有的入侵行为。而在实践中，这两种情况都不可能存在。它们只是代表了理想的状况。

2）易于配置性

一般来说，滥用检测系统比起非规则检测系统来说，需要少得多的配置工作，因为后者需要更多的数据收集、分析和更新工作。有些滥用检测系统允许用户创建自己的特征模式文件，这样将会增加建立系统配置的难度。

一般来说，非规则入侵检测系统都是比较难以配置的。因为它需要对系统的已知和期望行为模式做全面综合的定义。这就要求用户去发现、理解并表示和维护目标系统的所有正常行为状态。

3）报告的数据

滥用检测系统一般在模式匹配的基础上生成最后结论。其具体输出形式可以是一条指示特定攻击行为发生的警告信号，其中还可包含相关的提示数据。而非规则检测系统的输出结论通常是建立在实际活动行为与系统期望行为的统计相关处理的基础上。非规则检测系统常常生成更多的数据量，因为任何超出期望行为范围的事件都将被报告给系统管理员。

1.5 入侵检测系统的层次体系

尽管每一个 IDS 从概念上都由 3 部分构成：探测器、分析器和用户接口，但是每个特定 IDS 所检查的数据类型和生成数据的类型都大不相同。根据所检查数据类型的不同，IDS 可以划分为以下几类层次。

1）应用程序

基于应用程序的 IDS 负责监测特定应用程序的运行，它通常检查的是各种系统日志文件（包括系统调用记录）。

2）主机

基于主机的 IDS 通常检查诸如日志文件、进程记账信息、用户行为信息以及主机上运行的基于应用程序的 IDS 的输出数据等。

3）网络

网络 IDS 主要检查网络范围内的数据流量。它一般可以访问所监控网络环境中的基于主机和应用程序的 IDS 所发来的输出数据，同时也可以直接检查网络数据包信息。

4）多个网络/基础设施

多网络环境下的 IDS 通常表现为紧急事件反应小组（IRT，Incident Response Team）的形式。它的输入来自下属网络范围内的各个站点。每个站点是一个管理域内的一个安全监控实体。该种类型的 IDS 接受来自上述其他各种类型 IDS 的输出数据。

1.6 进一步发展的若干方向

1.6.1 宽带高速网络的实时入侵检测技术

大量高速网络技术如 ATM、千兆以太网、G 比特光纤网等在近年内不断出现，在此背景下的各种宽带接入手段层出不穷，其中很多已经得到了广泛的应用。如何实现高速网络下的实时入侵检测已成为一个现实的问题。目前，国外市场已经推出了几款基于千兆以太网环境的入侵检测系统产品，但是其性能指标还远未成熟。

这需要考虑两个方面的问题。首先，入侵检测系统的软件结构和算法需要重新设计，以适应高速网络的新环境，重点是提高运行速度和效率。开发与设计相适应的专用硬件结构，加上配合设计的专用软件是解决这方面问题的一个途径。另一个问题是，随着高速网络技术的不断进步和成熟，新的高速网络协议的设计也成为未来的一个发展趋势，如对 TCP/IP 协议的重新设计等，所以，现有的入侵检测系统如何适应和利用未来新的网络协议结构是一个全新的问题。

1.6.2 大规模分布式入侵检测技术

传统的集中式入侵检测技术的基本模型是在网络的不同网段中放置多个传感器或探测器用来收集当前网络状态信息，然后这些信息被传送到中央控制台进行处理和分析。或者更进一步的情况是，这些传感器具有某种主动性，能够接收中央控制台的某些命令和下载某些识别模板等。

这种集中式模型具有几个明显的缺陷。首先，面对在大规模、异质网络基础上发起的复杂攻击行为，中央控制台的业务负荷将会达到不可承受的地步，以致于无法具有足够能力处理来自四面八方的消息事件。这种情况会造成对许多重大消息事件的遗漏，大大增加漏警概率。其次，由于网络传输的时延问题（这在大规模异质网络中尤其如此），到达中央控制台的数据包中的事件消息只是反映了它刚被生成时的环境状态情况，已经不能反映可能随着时间已经改变的当前状态。这将使基于过时信息做出的判断的可信度大大降低，同时也使得反回去确认相关信息来源变得非常困难。异质网络环境所带来的平台差异性也将给集中式模型带来诸多困难。因为每一种攻击行为在不同的平台操作环境中都表现出不同类型的模式特征，而已知的攻击方法数目非常之多。这样，在集中式模型的系统中，想要进行较为完全的攻击模式的匹配就已经非常困难，更何况还要面对不断出现的新型攻击手段。

面对诸多难题，很多新的思路已经出现，其中一种就是攻击策略分析（Attack Strategy Analysis）方法。它采用了分布式智能代理的结构方式，由几个中央智能代理和大量分布的本地代理组成，其中本地代理负责处理本地事件，而中央代理负责整体的分析工作。与集中式模型不同的是，它强调的是通过全体智能代理协同工作来分析入侵者的攻击策略，中央代理扮演的是协调者和全局分析员的角色，但绝不是惟一的事件处理者，其地位有点类似于战场上的元帅，根据对全局形势的判断，指挥部下开展行动。这种方法有其明显的优点，但同时又带来了其他的一些问题，如大量代理的组织和协作问题、相互之间的通信、处理能力和

分析任务的分配等等。

1.6.3 入侵检测的数据融合技术

目前的入侵监测系统还存在诸多缺陷。首先，现有的实时检测系统在技术上还不具备是以检测到由受到良好训练的黑客发起的复杂隐蔽攻击行为的能力。其次，检测的虚假警告问题也是一个令许多网络管理员头疼的事情。同时，来自各种渠道的大量泛滥数据、系统消息等常常没有得到很好和及时的处理，这样非但无助于解决问题，反而浪费和降低了 IDS 系统的处理能力和检测性能。

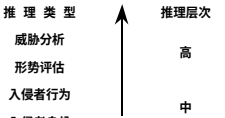


图 1.2 IDS 数据融合推理层次

为解决上述问题，多传感器数据融合技术提供了一条重要的技术途径。它能够把从多个异质分布式传感器处得到的各种数据和信息综合成为一个统一的处理进程，来评估整个网络环境的安全性能。数据融合入侵检测系统的输入可以是从网络嗅探器处得到的各种网络数据包，也可以是系统日志文件、SNMP 信息、用户资料信息、系统消息和操作命令，系统输出是入侵者的身份估计和确定位置、入侵者的活动信息、危险性信息、攻击等级和对整个入侵行为危险程度的评估等。

入侵检测数据融合推理的层次结构如图 1.2 所示。

入侵检测数据融合技术同样面临着若干挑战，例如开发一种通用的结构化“元语言”，用于来描述入侵检测和网络管理的对象以及对动态网络攻击行为的检测技术，还有将具有强烈数学背景的多传感器数据融合理论应用到实际的 IDS 系统所面临的若干复杂问题等。

1.6.4 先进检测算法的应用

下面将简要介绍 3 种机器学习算法在入侵检测中的应用情况。它们分别是计算机免疫技术、神经网络技术和遗传算法。

计算机免疫技术是直接受到生物免疫机制的启发而提出的。根据该种理论，由于计算机网络受到安全策略、计算机程序以及系统配置等多种因素中所可能包含的错误的影响，所以总是处于易受入侵的状态。生物系统中存在的种种脆弱性因素都是由免疫系统来妥善处理的，而这种免疫系统机制在处理外来异体时呈现了分布的、多样性的、自治的以及自修复的特征。生物免疫系统通过识别出异常或者以前从未出现的特征，来确定入侵异体。

受到上述思想的激励，计算机免疫技术为入侵检测提供了以下思路，即通过正常行为样本的学习来识别出不符合常态概念的行为序列。在此方面已经做了若干研究工作。早期的一些实验建立在训练系统识别代表正常执行程序活动的系统调用序列的基础上。这些实验发现若干系统调用形成的序列能够形成稳定的检测特征，可用于检测在使用诸如 Sendmail 和 lpr 程序时的异常情况。

神经网络技术在入侵检测中的应用历史比较长，并且一直在不断发展。早期的一些研究人员通过训练后向传播神经网络（BP 神经网络）来识别已知的网络攻击行为。更进一步的实验通过训练后向传播神经网络来识别未知的网络入侵行为。基本的思想是首先使用若干正

常行为的样本来训练神经网络，然后检测任何偏离这些行为样本的行为模式。研究人员还在“正常”训练样本集中加入随机模式（这些通常被认为是异常状态），期望减小检测的漏警率。

时至今日，神经网络技术已经体现出了强大的攻击模式分析能力，它的优势包括能够较好地处理带噪声的数据，不需要费力的模型构建过程并且分析计算速度很快而能用于实时分析等。早期的实验结果已经证明，该种技术能够检测到未知的入侵模式，因此，具有更进一步发展的潜力。另外，现在提出了各种其他的神经网络架构诸如自组织特征映射网络等，以试图克服后向传播网络的若干限制性弱点。

遗传算法在入侵检测中的应用历程还比较短，所取得的成果也有限。在一些研究试验中，使用了若干字符串序列来定义用于分析检测的指令组。用于检测识别正常或者异常行为的这些指令在初始训练阶段中不断进化，以提高分析能力。在训练之前，代表指令的这些字符串所具有的分析检测能力很弱，但是通过任意重组这些字符串的片段，就可以生成新的字符串，然后再选出检测能力最强的那些指令字符串。这样重组、测试和选择的循环过程一直进行，直到检测能力不再提高。此时，这些指令字符串的检测能力已经得到很大提高，可以用于实际检测之中了。迄今为止，只在若干小特征样本集的情况下进行了该算法的训练过程，所取得的结果表明这种方法还需要提高其区分异常与正常行为的能力。

1.7 面临的挑战

目前，入侵检测系统也面临着若干重要的挑战。这些挑战有些来自技术方面，有些则来自非技术方面。技术方面的主要挑战包括：

1）网络规模和复杂程度的不断增长。在一个大型的异构网络环境中，入侵检测系统所遇到的主要问题有：如何集成并处理来自分布在网络各处实体的具有不同格式的各种相关信息，如何在相互合作但是并不完全相互信任的组织之间来共享敏感的相关入侵行为信息，如何进行管理域间的合作进程以及如何保证在局部入侵检测系统失效的情况下仍能维护系统全局的安全等。

2）如何在造成损失前及早发现入侵活动，即预警技术。

3）网络繁忙情况下的系统性能问题。为了保证发挥效能，网络入侵检测系统必须能够分析所有的内向数据包。如果一个入侵检测系统无法应付网络吞吐量的话，它就可能漏掉不少反映入侵活动的特征数据，从而造成安全漏洞。

4）入侵模式特征的准确性。用来描述异常入侵行为的模式特征是滥用检测系统最重要的基石。如何保证所采用的特征集能够准确而又足以描述已知的各种攻击模式（包括复杂的分阶段攻击行为）及其变种，是一个重要而敏感的问题。

5）入侵检测系统的评估。时至今日，在这方面所做的工作非常少。对入侵检测系统的评估测试是一项复杂的工作，因为 IDS 不能在独立环境中检测，首先必须建立一个实际网络平台环境。同时，还需要大量的包含各种测试入侵模式的复杂数据，这些数据还要根据不同的操作系统平台和版本加以调整。

非技术因素包括如下 3 个方面。

1）攻击者不断研究新的攻击模式，同时随着安全技术的普及，越来越多的人进行了越

来越多的入侵攻击尝试。

2）自动攻击的软件工具不断得到改进，使普通用户也能够利用它来进行网络攻击。

3）各种机构（包括政府、公司等）对包括 IDS 在内的安全技术的认识不足或者缺乏足够熟练的安全管理员。

我国计算机系统及网络以国外产品为主，软硬件系统中难免也存在各种潜在威胁和安全“陷阱”（诸如操作系统后门、路由器漏洞等）。因此，利用这些设备建立的网络系统在其安全性方面得不到根本性的保障。

在我国计算机网络的建设现状下，基于防火墙和加密技术的安全防护固然重要，但是，发展网络入侵检测以及预警技术也同样重要。入侵检测技术已经成为当前网络安全技术领域内的一个研究热点。它的快速发展和极具潜力的应用前景需要有更多的研究和工程技术人员投身其中，在基础技术原理的研究和工程项目开发等多个层面上同时开展工作，才有可能开发出领先的产品系统。

第 2 章　网络编程基础知识

本章将简要地介绍网络协议方面的基本知识，同时还将介绍国际标准化组织（ISO）指定的开放系统互联参考模型（OSI）和伴随着因特网而发展起来的 TCP/IP 模型，并简单地介绍 UNIX 系统环境下网络编程技术的基础知识。最后，对事实上的工业标准 TCP/IP 协议族中各主要协议的报文格式做简要介绍。

2.1　分层协议模型

为了减少网络协议在设计上的复杂性，大多数的协议族采用了层次模型。每一较低层次的协议都向较高一层的协议提供一定的服务，这些服务的具体实现方法对上一层协议而言是透明不可见的。相邻层的协议之间采用原语进行交互，这样的设计提供了各层协议在实现上的独立性。同样层次的协议在不同的机器和操作系统上可能有不同的实现方式，但是只要它正确实现了与上下协议层的交互界面原语，提供了一致的服务，就可以确保网络通信的正常进行。

网络通信是在两个通信实体之间进行的。这两个实体的对应层采用协议来交换信息。协议是对应层之间进行通信的一组约定规则。举例来说，两个不同国家的以色列学者和中国学者通过电话进行学术交流。他们一个说希伯莱语，一个说汉语，其各自的翻译除了会本国语言外都会说英语和法语。那么，怎么进行交流呢？显然可以使用下面的过程：说希伯莱语的以色列学者可以先用希伯莱语告诉他的翻译要说的内容，然后他的翻译将其译为英语，再通过电话告诉中国学者的翻译。中国学者的翻译把接收到的英语内容翻译成汉语，然后告诉说汉语的中国学者。这个过程可以看成是两个不同国家的学者通过一个“翻译网络”进行通信的过程。图 2.1 所示为此过程的示意图。

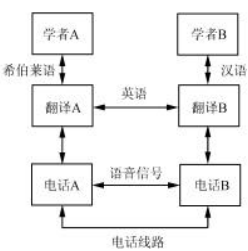


图 2.1　“翻译网络”的通信过程

翻译们之间具体使用什么语言进行交流，就是翻译层的协议规则。在这个例子中，翻译之间使用的协议语言是英语，但是如果翻译们约定使用法语（相当于另一个协议），并不会影响两个学者之间的交流。这也就是前面所指的各层协议在具体实现上的独立性。同样，也可以把电话看做是通信网络中的一层，它们之间的通信协议就是语音信号。电话之间用何种调制方式的电波脉冲并不会影响它们传递的语音信号。层与协议一起构成了网络协议结构。

2.2 开放系统互联参考模型 OSI/ISO

在早期的计算机发展中，只有同一厂商生产的系列计算机之间才能进行互相通信。为了促进异构机器互联网络的发展，国际标准化组织（ISO）于 20 世纪 70 年代后期指定了开放系统互联参考模型 OSI（Open System Interconnection），该参考模型共分为 7 层：

- 应用层

提供用户网络分布信息服务的接口，如文件传送、电子邮件服务等。

- 表示层

提供两个应用层协议实体之间数据表示的语法，如加、解密算法等。

- 会话层

提供应用层实体会话通道的建立和清除以及会话过程的维护。

- 传输层

提供上述面向应用的高 3 层和以下所述面向网络的低 3 层之间的接口，为会话层提供与具体网络无关的可靠的端对端通信机制。主要有面向连接的服务（字节流）和面向无连接的服务（数据报）两种服务类型。

- 网络层

建立传输层实体之间的网络（WAN 或者 LAN）连接，包括路由选择等服务。

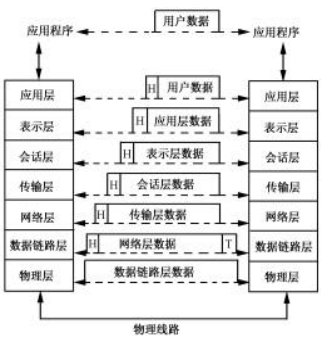


图 2.2 OSI 参考模型

- 数据链路层

建立在特定网络（LAN）的物理连接上，为网络层提供可靠传送通道，提供传输错误检测与数据重发。

- 物理层

提供网络端设备接口的物理和电气接口，与物理传输介质直接相连。

图 2.2 为 OSI 参考模型的示意图，显示了用户数据是如何穿过各层，被加上各层协议头后发送到相邻下一层，直至通过物理介质传送，然后在到达目的地后再逐层上传，在此过程中去掉相应的协议信息头，最终被应用程序接收的全过程。虚线表示同层之间按照一定的协议规则所建立起来的虚拟通道。

2.3 TCP/IP 参考模型

TCP/IP 参考模型是伴随着 Internet 的出现而产生的，也是目前应用最为广泛的网络协议结构。TCP/IP 不仅是 TCP 协议和 IP 协议，而是一个协议族，包括了大部分在互联网上广为使用的协议。该参考模型共分 4 层，如图 2.3 所示。

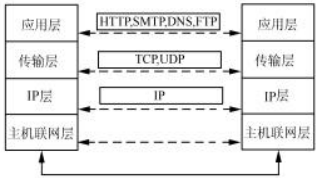


图 2.3　TCP/IP 参考模型及其协议族

- **应用层**

应用层提供各种常用的高层协议：FTP（**文件传输**）、TELNET（**远程登录**）、SMTP（**简单邮件传送**）、HTTP（**超文本传输**）、DNS（**域名服务**）等。**该层与** OSI **模型的应用层对应。**

- **传输层**

与 OSI **的传输层功能类似，提供了两种端对端的会话协议**（TCP **和** UDP）。**端对端是指传输层协议从自己的一个端口发送数据到对方的一个端口。端口是为了方便同一个主机上同时执行多个网络应用程序而设置的。因此在传输层的通信中，除了要指明对方的** IP **地址外，还要指定对方应用程序所用的端口。对同一个服务器的** IP **地址，指定不同的端口值将会获得不同的应用服务。大多数的因特网服务器不止提供一项服务，根据端口的不同能区分同一主机上的不同服务。表** 2.1 **给出了常见的主机服务端口。**

表 2.1　常见的因特网服务端口

端　口　号	所提供的服务
7	**回送**（Echo）
13	**日期时间**（Daytime）
15	**网络状态**（Netstat）
20	FTP **数据**（FTP-Date）
21	FTP **控制**（FTP）
23	**终端连接**（Telnet）
25	**简单邮件传输协议**（SMTP）
37	**时间**（Time）
42	**名字服务器**（Name Server）
53	**域名服务器**（Domain）
67	**引导协议服务器**（Bootps）
68	**引导协议客户**（Bootpc）
69	**简单文件传送协议**（TFTP）
101	NIS **主机名字服务器**（Hostname）
103	X400 **邮件服务**（X400）
110	**邮件协议第** 3 **版**（POP3）
119	**新闻传送协议**（NNTP）

TCP（**传输控制协议**）**是面向连接的协议，提供无差错字节流的可靠传递。要发送的字**

节流分成若干个块顺序传递到 IP 层，到达目的地后再由对方的 TCP 层按照顺序重新组装起来。如果数据包在这一过程中丢失，发送方的 TCP 层还负责重发丢失的数据。TCP 还进行流量控制，防止发送速度大于接收速度而出现数据包丢失的现象。

UDP（用户数据包协议）是不可靠的非连接的协议。它仅仅负责将数据打包发送出去，不保证数据包一定到达接受方，同时接受方 UDP 也不会调整所接受数据包的顺序。这一协议主要用于强调数据传送的速度而不是可靠性的场合，例如语音和视频信号的传送。

- 网际协议（IP）层

因特网上的任何一台主机都拥有一个独一无二的 IP 地址。目前的 IP 地址（IPv4）是由 4 个字节组成，如 210.27.2.64。IP 地址格式的规定包含在 IP 协议中。IP 层是整个 TCP/IP 模型中最关键的一环。该层上实现了 IP 分组数据包在网络中的点对点的传输，也就是说，发送 IP 分组包只需要指明它要到达的地址，而不必关心它经过什么样的路径到达。一般来说，如果进行通信的双方不是在同一个局域网内，就很难知道所发送的 IP 分组数据将如何抵达对方。

IP 分组在网络中传输所经过的路径叫做路由。连接因特网中的各个局域网的服务器负责把 IP 分组数据从一个子网发送到另一个子网，这样的服务器叫做路由器。路由的选择将在 IP 分组到达各路由器时进行，路由器根据它所存储的网络信息和 IP 分组所携带的目的 IP 地址信息，决定向何处发送 IP 分组（如果这个路由器与目的地址直接相连，则将直接转发到目的地址；否则，路由器将根据一定的算法选择发送到与本路由器相连的其他中继路由器）。这一系列的转发过程与发送方和接受方无关，这样就实现了异种网络之间的分组交换，使整个因特网看起来像一个大的单一网络。当前，使用最为广泛的 IP 协议是 IPv4 协议，定义的是 32 位 IP 地址。但是随着 Internet 的飞速发展，32 位地址已经无法满足需要，为此业界已经提出了使用 128 位地址的 IPv6 协议。

2.4 UNIX 网络编程技术概述

UNIX 的 Socket 技术提供了在 TCP/IP 模型中各个层次上的编程支持。Socket 的英文原意为“插座”，一般译为“套接字”。Socket 技术首先出现在 Berkeley 发布的 BSD 版本中，它是指在 UNIX 内核中处理接收到的各层协议数据包，然后应用程序使用读取文件的方式来接受内核传来的数据。通常使用一个套接字描述符来作为文件读取的句柄，文件读取的方式与普通文件类似。

UNIX 中通常有 4 种常用的套接字（Socket），分别是流套接字（Stream Socket）、数据报套接字（Datagram Socket）、原始套接字（Raw Socket）和本地套接字（Domain Socket）。

- 流套接字：提供基于 TCP 协议的双向、可靠、有序且不重复的无记录边界的字节数据流服务。
- 数据报套接字：提供基于 UDP 协议的双向数据流服务，但是不一定可靠、有序和不重复。
- 原始套接字：提供支持 Socket 概念的基本通信协议的访问。通过这个套接字可以对主机的网际协议层进行编程。它可以用来开发新的或者自定义的网络层协议，如新的 ICMP 协议等。

- **本地套接字：通过此种 Socket 技术，可以实行同一主机上不同进程之间的通信，即单一主机上的进程间通信（IPC，Inter-Process Communication）也可以用 Socket 技术来实现。本地套接字主要用于移植远程客户/服务程序到单一主机上。**

除了 Socket 技术，UNIX 还提供 TLI（Transprot Layer Interface）技术来支持网络编程。TLI 技术是基于 OSI 参考模型的技术，提供传输层上面向连接和不面向连接的接口。因为 OSI 模型同样包含了 TCP/IP 模型，因此 TLI 同样也支持 TCP/IP 的结构。但是，本书中主要针对特定 TCP/IP 结构，所以，采用了 Socket 网络编程技术，也只对 Socket 技术的应用做了介绍。对 TLI 技术感兴趣的读者，可参考相关文献。

2.5　TCP/IP 协议

从体系结构看，TCP/IP 可分为应用层、传输层、网络层和网络接口或主机联网层，如 2.3 节中所述。其中，网络接口层相当于 OSI 定义的 7 层模型中的物理层和数据链路层。每层包含的主要协议类型如表 2.2 中所示。

表 2.2　TCP/IP 协议族及分层结构

应　用　层	简单邮件传输协议（SMTP）	超文本传输协议（HTTP）	文件传输协议（FTP）	域名服务协议（DNS）	简单网络管理协议（SNMP）
传输层	TCP			UDP	
网络层	IP，ICMP，ARP，RARP				
网络接口层	Ethernet	Token-Ring		100 BASE-T	其他

2.5.1　网络接口层协议

实际上，TCP/IP 协议并不包括物理层和数据链路层协议，只定义了各物理网络协议与 TCP/IP 之间的接口信息。这些物理网络包括了多种广域网（如 ARPANET，MILNET 和 X.25 公用数据网）以及各种局域网（如 IEEE 所定义的 Ethernet，Token-Ring 等标准局域网类型）。由于该层与各种具体的物理网络打交道，所以其协议帧的格式随着所采用的网络类型的不同而不同，如以太网（Ethernet）的帧格式和令牌环网（Token-Ring）的帧格式就不一样。这里只介绍最常见的以太网 IEEE802.3 的帧格式。图 2.4 给出以太网帧的帧头格式。

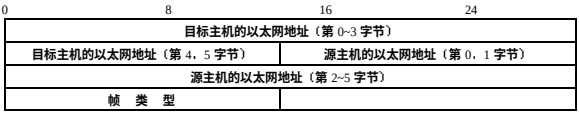


图 2.4　以太网帧头格式

IEEE 802.3 的帧头格式包括 6 个字节（48 位）的目标主机以太网地址、6 个字节的源主机以太网地址和 2 个字节的帧类型，其中帧类型指明所采用的协议。常见的协议类型如下。

1）**IP 协议，类型值** 0x0800；

2）ARP 协议，类型值 0x0806；

3）RARP 协议，类型值 0x8035。

2.5.2 ARP 协议和 RARP 协议

为了使 TCP/IP 协议与具体的物理网络无关，并隐藏物理地址而统一使用 IP 地址进行网际通信，就必须提供一种在 IP 地址和物理地址之间进行映射的机制。对于像以太网这样具备广播能力的网络，TCP/IP 使用地址解析协议 ARP（Address Resolution Protocol），来提供从 IP 地址到物理地址的映像服务。而提供从物理地址到 IP 地址映像服务的是逆向地址解析协议 RARP（Reverse Address Resolution Protocol）。

1）ARP 协议

当一个主机向另一个主机发送报文时，只有知道了对方 IP 地址所对应的物理地址后，才能在具体物理网络上进行传输。这种地址解析服务是由 ARP 协议来提供的。注意，ARP 只用于解析对方的物理地址，而不用于解析本机的物理地址。

ARP 采用一种称为“动态绑定”（Dynamic Binding）的技术来解析对方物理地址。具体过程是，发送方主机向网络中发送一个广播式 ARP 请求报文。报文中包含了目的主机的 IP 地址，用来请求与该 IP 地址匹配的目的主机的物理地址。网络中的所有主机都能接收到这个 ARP 请求报文，但是只有本机 IP 地址与报文中目的 IP 地址匹配的主机才会发送一个 ARP 响应报文，报告自己的物理地址。如此，发送方就得到了目的主机的物理地址。

ARP 协议的报文格式如图 2.5 所示。



图 2.5 ARP/RARP 报文格式

各个字段的内容说明如下。

- 硬件类型：指明硬件接口类型，若为以太网，该值为 1。
- 协议类型：指明发送方使用 ARP 获取物理地址的高层协议类型，其中 0x0800 代表的是 IP 协议。
- 硬件地址长度：对于以太网，长度为 6 个字节。
- 协议地址长度：对于 IP 协议，值为 4 字节。
- 操作类型：表示报文类型，定义如下所示。

  1——ARP 请求报文　　3——RARP 请求报文

  2——ARP 响应报文　　4——RARP 响应报文
- 源主机硬件地址：用于存放发送方的物理地址，长度 6 个字节。
- 源主机 IP 地址：用于存放发送方的 IP 地址，长度 4 个字节。

- 目的主机硬件地址：用于存放目的方的物理地址，长度 6 个字节。对于 ARP 请求报文而言，该字段为空。
- 目的主机 IP 地址：用于存放目的方的 IP 地址。

发送方在发送 ARP 请求报文时，要填写除目的主机硬件地址字段之外的其他所有字段。接受方通过发送 ARP 响应报文来报告自己的物理地址，这时报文中的发送方和目的方字段要做相应的对换。

在同一网段中，发送方的 ARP 请求报文可以抵达网络中的任何一台主机。在互联网环境中，发送给另一网络中主机的数据包必须经过路由器的转发。此时，发送方必须首先获取路由器节点的物理地址，然后发送 ARP 请求报文给该路由器节点。

2）RARP 协议

在站点初始化后，如果只有自己的物理地址而没有 IP 地址，就可以通过 RARP 协议发送广播式请求报文来获取自己的 IP 地址，而 RARP 服务器专门负责对此请求做出应答。RARP 协议主要供无盘工作站用来获取自己的 IP 地址。

RARP 协议的报文格式与 ARP 相同。当发送方以广播方式发送 RARP 请求报文时，在源主机硬件地址和目的主机硬件地址字段中都填入本机物理地址。RARP 服务器接收到该请求报文后，就回送一个 RARP 响应报文，在其目的主机 IP 地址字段中返回发送方的 IP 地址。

2.5.3　IP 协议

IP 协议（Internet Protocol）是 TCP/IP 协议族的核心协议之一，它提供了无连接数据包传输和网际路由服务。

IP 通过互联网传输数据报，它是无连接的，各个 IP 数据报之间是相互独立的。它是不可靠的，不保证投递抵达目的地，对分组的丢失、重复、延迟和顺序错位等情况都不予检测。IP 协议只是做到尽力发送每个数据分组，但是在网络阻塞的情况下，可能会发生丢弃某些数据报的现象。

在传送数据分组时，IP 协议将高层协议数据封装成 IP 数据报，然后通过网络接口发送出去。如果目的主机在同一个网络中，则 IP 直接将数据报发送到该目的主机。如果目的主机位于远程网络中，它就将数据报发送给本地网络的路由器，再由本地路由器转发给下一个路由器，或者目的主机。如此，IP 数据报在网络间不断传递，直至到达目的主机时为止。

1）IP 数据报格式

IP 数据报由报头和报文数据两部分组成，如图 2.6 所示。



图 2.6　IP 数据报格式

各个字段的内容说明如下。

① 版本号（Version）长度为 4 位，指明 IP 协议的版本号，目前取值为 4。在处理 IP 数据报之前，先检查版本号，以保证与所期望的格式一样。

② 报头长度（Internet Header Length）长度为 4 位，以 4 个字节为单位计算 IP 报头的长度。取值范围 5~15，默认为 5，即 20 个字节。该字段值常用来计算报文数据在数据报中的起始位置，具体计算方法是用报文总长度值减去报头长度值。

③ 服务类型（Type of Service）长度为 8 位，指明本数据报所需要得到的服务质量。该字段允许指定传输的优先级别、传送速度、可靠性和吞吐量等参数，这些参数可在双方交互过程中加以协商。

④ 报文总长度（Total Length）为 16 位，以字节为单位计算数据报的总长度，包括报头长度，最大长度为 65 535 个字节。考虑到实际传输网络的质量，实际传输的数据报长度要比最大值小。

⑤ 报文标识（Identification）长度为 16 位。该字段为一个整数，并且惟一标识一个数据报，即不同数据报的标识不一样。如果数据报需要分段，则每一个分段的标识符都一样。当目的主机收到具有相同标识符的数据报分段时，认为它们来自同一个完整的数据报，并据此对其进行重组工作。

⑥ 标志（Flags）长度为 3 位，其中最高位未定义，设定为 0，其余两位分别是“不可分段位”（DF）和“更多分段位”（MF），用来控制 IP 数据报的分段情况，如图 2.7 所示。

（未定义）	DF 位	MF 位
设定为 0	0:可分段 1:不可分段	0:最后分段 1:更多分段

图 2.7 报文标志

⑦ 段偏移量（Fragment Offset）长度为 13 位。当数据报进行分段时，以 8 个字节为单位，指出该分段的第一个数据字在原始数据报中的偏移位置。

⑧ 生存期（Time to Live）长度 8 位，取值范围 0~255，以秒为单位，指出 IP 数据报能在网络中存活的最长时间，其值由发送方设定，并在每经过一个路由节点时减 1。当该值为 0 时，该数据报被丢弃。

⑨ 协议（Protocol）长度为 8 位，指明接收该数据报的高层协议。常见的协议值如表 2.3 中所示。

表 2.3 常见网络协议值

网络协议值	关 键 字	协 议 名 称
0		保留
1	ICMP	Internet 网间控制报文
2	IGMP	Internet 网组管理
3	GGP	网关-网关协议
4	IP	IP 里的 IP（IP in IP）
5	ST	数据流
6	TCP	传输控制协议
8	EGP	外部网关协议