

5. 入侵检测

Intrusion Detection System

D. E. Denning. An intrusion detection model. IEEE Transactions on Software Engineering, Special issue on computer security and privacy, 13(2):222–232, 1987.



入侵检测系统分类

入侵检测系统

分 类 依 据	入侵检测系统类别	备　　注
检测方法	基于行为的入侵检测系统	也称异常性检测
	基于入侵知识的检测系统	也称误（滥）用检测
被保护的目标系统	基于主机的入侵检测系统	主机环境适用
	基于网络的入侵检测系统	适用于网络环境
分析数据源	主机系统的审计、系统日志等	根据分析数据源可分为针对不同分析数据源的入侵检测系统。
	网络数据报、网管信息	
	应用程序的日志	
	其它入侵检测系统的报警信息	
响应方式	主动的入侵检测系统	对检测到的入侵进行主动响应处理。
	被动的入侵检测系统	对检测到的入侵仅进行报警



入侵检测的分类（1）

- 误用检测模型（**Misuse Detection)**：收集非正常操作的行为特征，建立相关的特征库，当监测的用户或系统行为与库中的记录相匹配时，系统就认为这种行为是入侵。

- 有时也被称为特征分析**(Signature Analysis)**或基于知识的检测**(Knowledge-based Detection)**

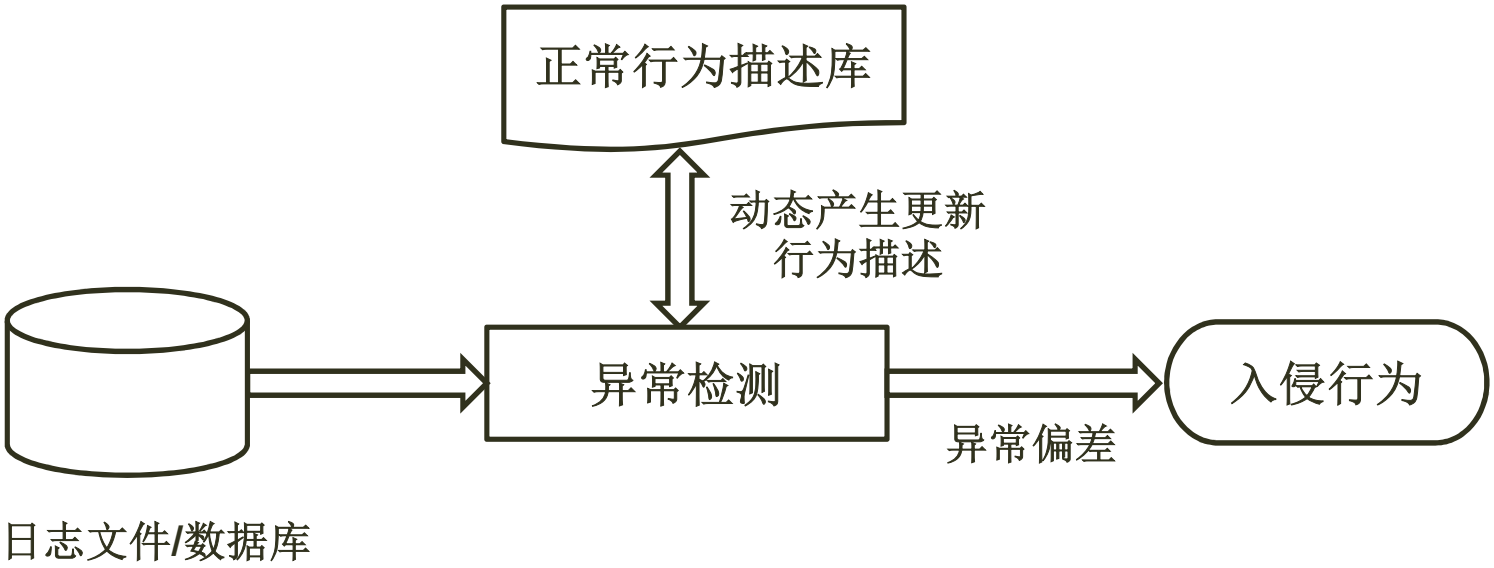
- 检测准确率高，但检测范围受已知知识的局限，将具体入侵手段抽象成知识也很困难

- 对目标系统依赖性高，移植性差



入侵检测的分类

- 异常检测模型（**Anomaly Detection):**首先总结正常操作应该具有的特征（**Normal Usage Profile)**），当用户活动与正常行为有重大偏离时即被认为是入侵（度量及门限）
- 通用性较强，甚至有可能检测出以前未出现过的攻击方法。
- 性能：异常与入侵并不一定总是相关；如何全面正确地对用户行为进行描述；可能被恶意训练欺骗。用户的行为是经常改变的，误检率高





入侵检测的分类（2）

按照数据来源：

- 基于主机：系统获取数据的依据是系统运行所在的主机(系统日志、应用程序日志)，保护的目标也是系统运行所在的主机
- 基于网络：系统获取的数据是网络传输的数据包，保护的是网络的运行,往往将一台机器的网卡设于混杂模式,监听所有本网段内的数据包并进行判断。
- 混合型: 基于主机又基于网络，一般是分布式的

更多的数据来源： 文件内容分析、信誉等

如**NGIDS（Fireeye)**， 加上了文件的**anomaly detection**

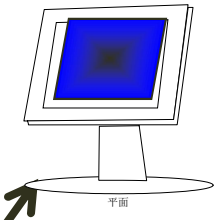


入侵检测的分类（3）

- 网络IDS
  - 侦测速度快
  - 隐蔽性好
  - 视野更宽
  - 较少的监测器
  - 占资源少
- 主机IDS
  - 视野集中
  - 易于用户自定义
  - 保护更加周密
  - 对网络流量不敏感
  - 大规模应用的成本高
  - 占用主机资源

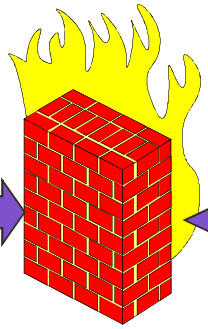
基于网络入侵检测系统

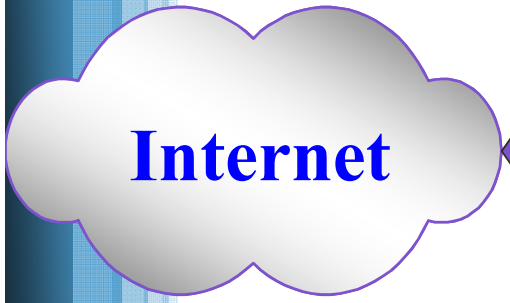
- ✈通过网络适配器捕获数据包
- ✈分析数据包

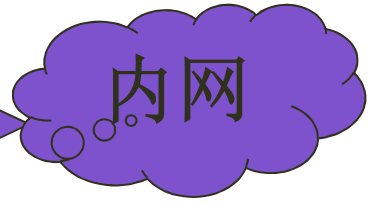


- 能够检测超过授权的非法访问
- IDS发生故障不会影响正常业务的运行
- 配置简单

- 只能检查单一网段的通信
- 通常采用特征检测的方法
- 无法检测系统级的入侵



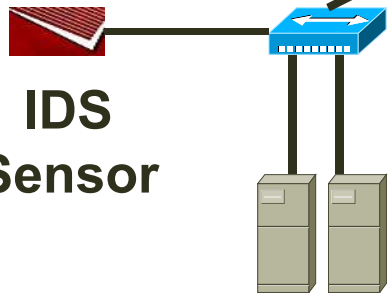






共享媒介

Consol
e



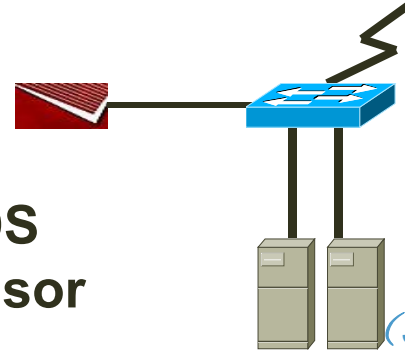
HUB

IDS
Sensor

Monitored
Servers

交换环境

Console



Switch

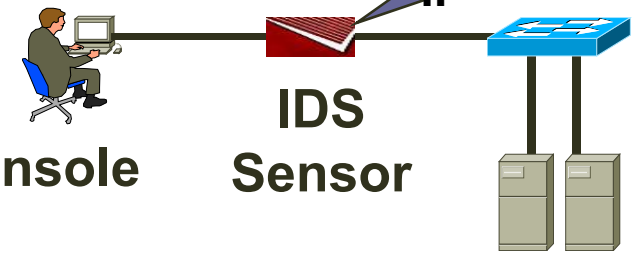
IDS
Sensor

通过端口镜像实现

（SPAN / Port Monitor）

Monitored Servers

NIDS的部署



不设
IP

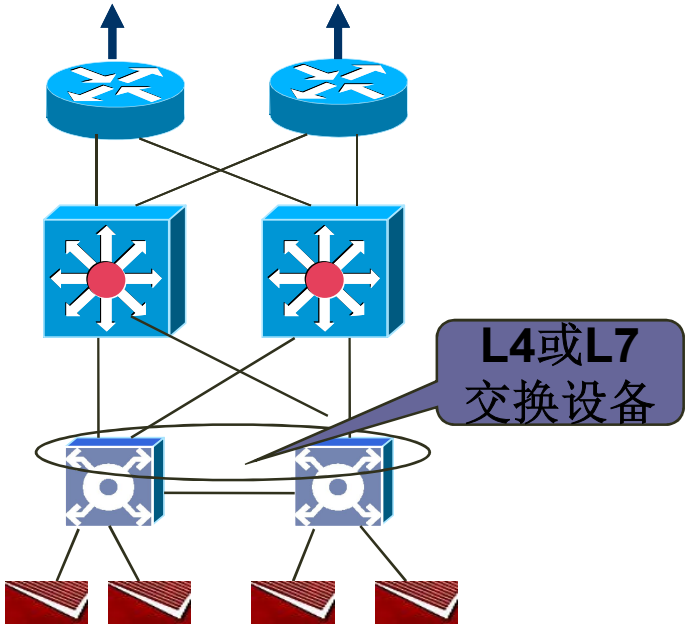
Console

IDS
Sensor

Switch

Monitored
Servers

隐蔽模式

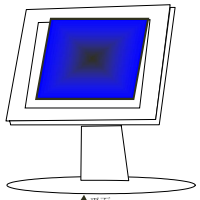


L4或L7
交换设备

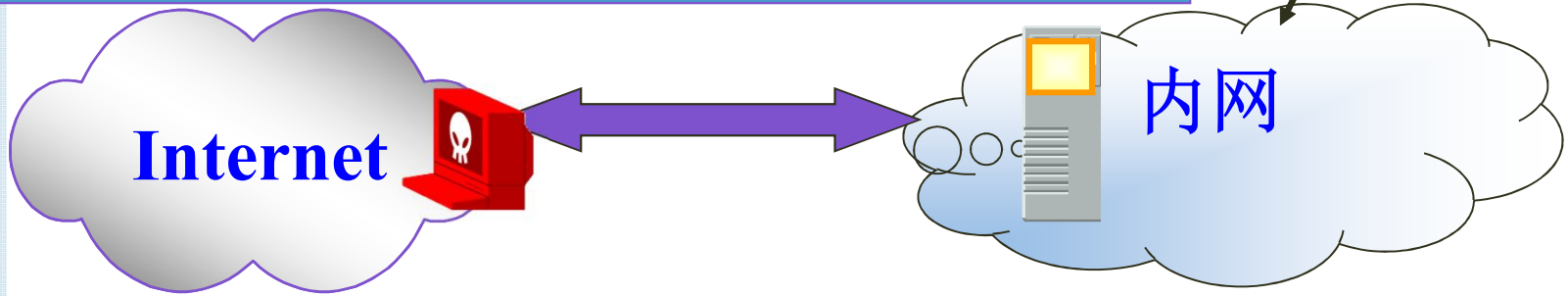
IDS Sensors

基于主机的入侵检测系统的功能

- 基于主机的IDS与基于网络的IDS相比通常能够提供更详尽的相关信息。
- 基于主机的IDS通常情况下比基于网络的IDS误报率要低，因为检测在主机上运行的命令序列比检测网络流更简单，系统的复杂性也少得多。



- IDS安装在需要保护的设备上,增加了被入侵的风险
- 依赖于服务器固有的日志与监视能力
- 全面布署基于主机的IDS代价较大.
- 日志文件容易被删除或修改





入侵检测的分类（3）

- 按系统各模块的运行方式
  - 集中式：系统的各个模块包括数据的收集分析集中在一台主机上运行
  - 分布式：系统的各个模块分布在不同的计算机和设备上



二、IDS产品介绍

1. Snort.是一个免费，开放源代码的基于网络的入侵检测系统，它具有很好的配置性和可移植性。除此之外，它还可以用来截获网络中的数据包并记录数据包日志。(主要采用的是模式匹配方法)
   http://www.Snort.org

2 Bro：http://www.bro-ids.org/ The Bro Network Security Monitor， *Lawrence Berkeley National Laboratory*.

Bro is a powerful network analysis framework that is much different from the typical IDS you may know.

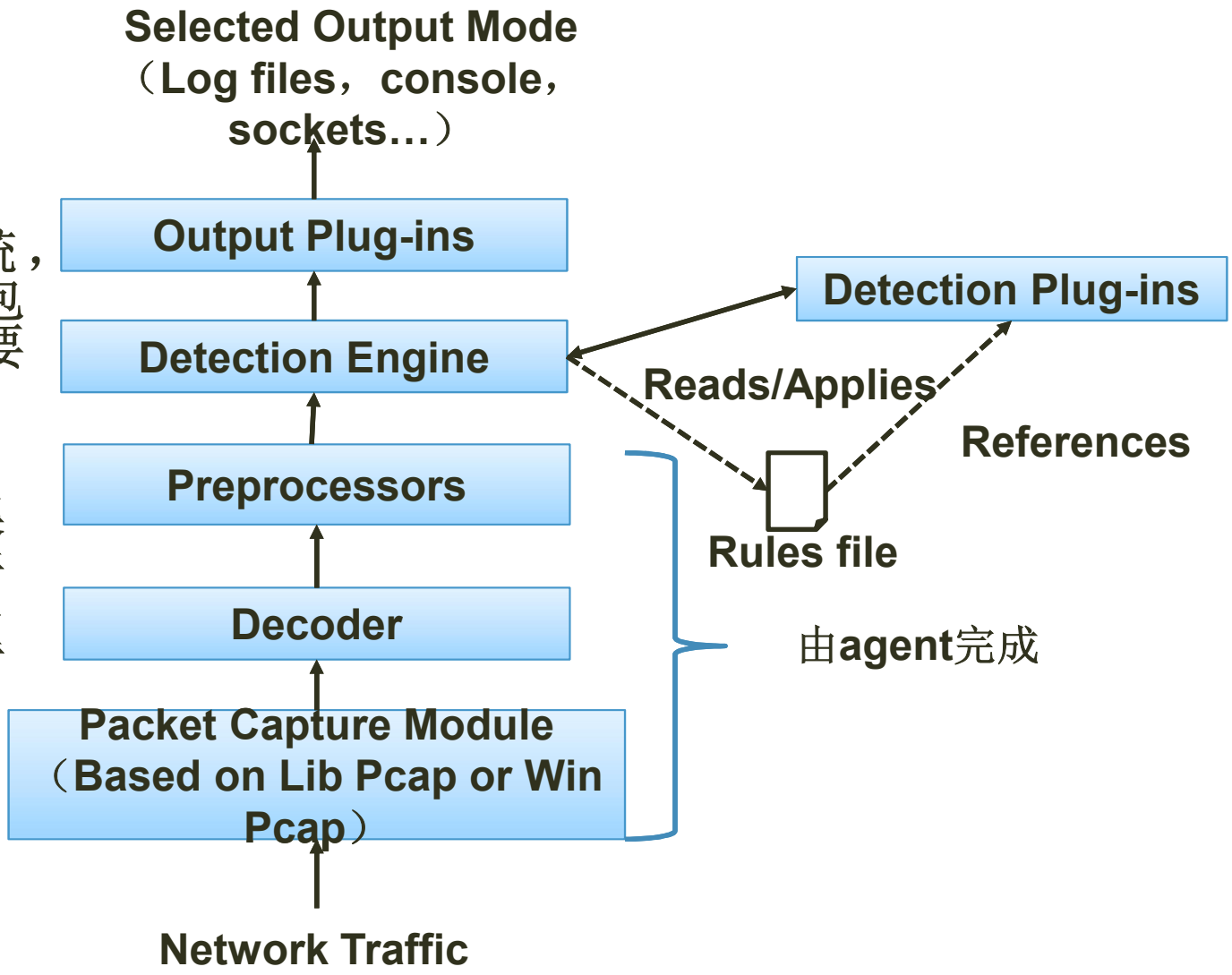
三、入侵检测关键技术



例：SNORT结构

入侵检测系统，通常来说包括三个主要功能部件

（1）信息收集
（2）信息分析
（3）结果处理





（1）信息收集的来源

- 系统或网络的日志文件
- 系统目录和文件的异常变化
- 程序执行中的异常行为
- 网络流量
  - 通过网络侦听，数据包捕获
  - **SNMP/RMON的流量采集**
  - 基于网络探针**(Probe)**的流量采集
  - 基于网络流**(NetFiow)**的流量采集
  - 基于Openflow的流量信息收集



NIDS （Network Driver Interface Specification）抓包

- **PF_PACKET**
  - linux中提供了PF_PACKET接口，可以操作链路层的数据从链路层抓包，PF_PACKET协议族是与系统TCP/IP协议栈并行的同级别模块，即从PF_PACKET协议族得到的数据包是没有经过系统TCP/IP协议栈处理的
- **libpcap**
  - unix/linux平台下的网络数据包捕获函数包http://www.tcpdump.org/，http://sourceforge.net/projects/libpcap/
    - tcpdump及ids snort都是基于libpcap编写的
- **winpcap(windows packet capture)**
  - Win32平台下用于抓包和分析的系统。包括：
    - 内核级别的packet filter
    - 底层的DLL(packet.dll)提供一个底层的API，通过这个API可直接访问网络设备驱动，而独立于Microsoft OS
    - 高级的独立于系统的DLL（Wpcap.dll）：高层的强大捕获程序库，与Unix下的libpcap兼容。它独立于下层的网络硬件和操作系统



- **SNMP/RMON的流量采集**
  - 所有的网络设备都提供标准的SNMP功能。通过一些系统可提供固定时长的流量曲线和重要数据，能够满足一般的端口链路流量监视的要求，但其相对其它技术而言功能单一，信息量少
- **基于NetFlow的流量采集技术**
  - NetFl ow是Cisco公司开发的专用流交换技术，同时又用于记录流量统计信息。NetFl ow 中定义流为从特定的源主机到目的主机的单向数据报集合。数据流由以下七个参数标识:源IP地址、目的IP地址、协议类型、源传输层端口、目的传输层端口、业务类型TOS以及设备的输入接口。上报数据中含有流的起止时间、报文数和字节数等信息，还包括AS域或路由信息
- **基于Openflow的流量信息收集**



数据 采 集 探针

- 流量探 针可以实时对流量数据进行采集记录，经过汇聚和预处理将流量信息发送到后端数据库。通过分析软件可进行实时监视，图表显示分析统计结果或导出报表文件;经条件设置还能够利用流量探针的数据捕获功能对网络流量进行实时采集或流量镜像，进行报文的协议分析
- 使用时通过分光分路设备、交换机流量镜像端口或直接将其串接在待观测的链路上，对链路上所有的数据报文进行处理，提取流量监测所需的协议字段甚至全部报文内容。

NetScout Probe Sniffer硬件探针



（2）数据预处理模块：从各种数据源采集上来的数据，需要经过预处理才能够加以分析。

- 去除一些明显无用的信息
- 进行数据的分类，将同种类型的数据分在一起
- 将相关的数据进行合并，合并的过程中也可以再去除一些冗余、无用的信息
- 将数据进行格式转换，使得这些数据可以被分析模块识别和处理。



〈3〉. 分析模块:分析模块是入侵检测系统的核心模块，它完成对事件的分析和处理。分析模块可以采用现有的各种方法对事件进行分析，在对事件进行分析后，确定该事件是否是攻击，如果是就产生报警，如果不能够确定，也要给出一个怀疑值。分析模块根据分析的结果，决定自己怀疑的数据是否要送给关联模块进行数据融合。

〈4〉. 关联模块:关联模块进行数据融合的主要目的就是综合不同分析模块送报上来的已给出怀疑值的事件，判断是否存在分布式攻击。

〈5〉. 管理模块:管理模块接到报警等信息后，决定是否采取响应，采取何种响应。



指标

- The most popular performance metrics is detection rate (DR) together with false alarm rate (FAR). An IDS should have a high DR and a low FAR. Other commonly used combinations include Precision and Recall, or Sensitivity and Specificity

	Negative Class (Normal)	Positive Class (Attack)
Actual Normal →	True Negative (TN)	False Positive (FP)
Actual attack →	False Negative (FN)	True Positive (TP)

True Negative Rate (TNR): $\frac{TN}{TN+FP}$, also known as Specificity.

True Positive Rate (TPR): $\frac{TP}{TP+FN}$, also known as Detection Rate (DR) or Sensitivity. In information retrieval, this is called Recall.

False Positive Rate (FPR): $\frac{FP}{TN+FP} = 1 - specificity$, also known as False Alarm Rate (FAR).

False Negative Rate (FNR): $\frac{FN}{TP+FN} = 1 - sensitivity$.

Accuracy: $\frac{TN+TP}{TN+TP+FN+FP}$

Precision: $\frac{TP}{TP+FP}$, which is another information retrieval term, and often is paired with "Recall".



检测算法

根据检测方法的不同，入侵检测技术可分为两类：

- 误用检测(Misuse Detection)：首先定义违背安全策略的事件的特征，检测主要判别这类特征是否在所收集到的数据中出现。
- 异常检测 (Anomaly Detection)：建立系统“正常”情况的模型，然后将系统运行时的数值与所定义的“正常”情况比较，得出是否有被攻击的迹象。

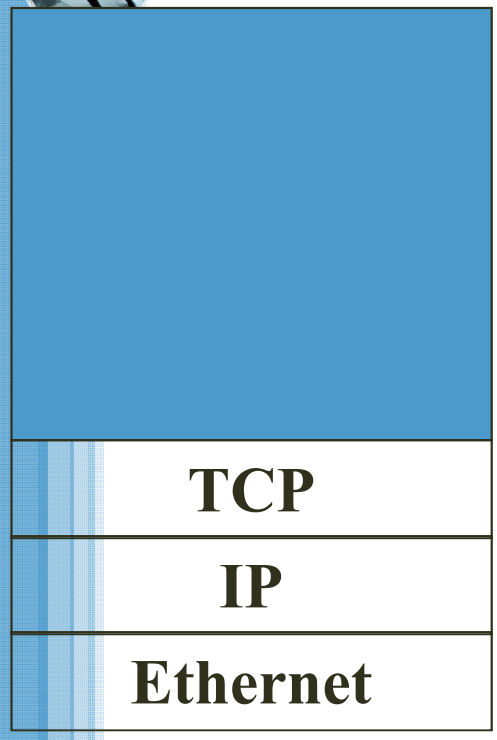
还有混合系统

- Varun Chandola, Arindam Banerjee and Vipin Kumar, Anomaly Detection:Survey, ACM Computing Surveys, Vol.41,No.3,Article 15,July 2009

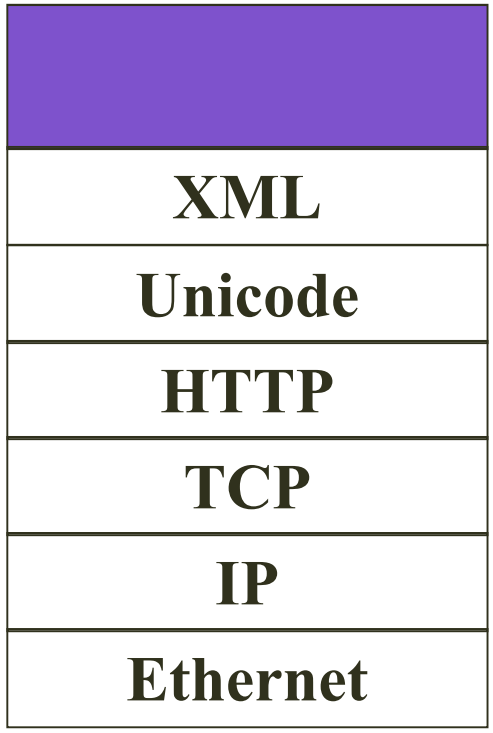
入侵检测的常用方法

特征检测/模式匹配，协议分析

模式匹配



协议分析



Challenges:

1.Realtime detection

2. False positive 误　报

False negative漏报

规则发现：

专家系统、统计分析
利用人工智能自动规则发现：神经网络，模糊系统，遗传算法，免疫系统，数据挖掘，深度学习



- 模式匹配的方法用于误用检测。它建立一个攻击特征库，然后检查发过来的数据是否包含这些攻击特征，判断它是不是攻击。
- 算法简单，准确率相对异常检测高
- 缺点
  - 只能检测已知攻击，对于无经验知识的入侵与攻击行为无能为力
  - 模式库需要不断更新，且模式更复杂（实时性，误报率）
  - 对于高速大规模网络，由于要处理分析大量的数据包，这种方法的速度成问题。
- 例：Port 25:{“WIZ”|“DEBUG”}
- 检查25号端口传送的数据中是否有“WIZ”或“DEBUG”关键字。



基于统计分析的检测技术

- 基于统计分析的检测技术根据系统中特征变量（如：事件的数量、间隔时间、资源消耗、流量等）的历史数据建立统计模型**historical statistical profile)**
  - 对正常数据的各个特征进行统计，根据统计结果对每一个特征设定一个正常范围的门限。这些特征和相应的门限组成检测的统计模型
  - 动态更新:模式向量随时间衰减，并将新的用户行为所产生的审计数据嵌入到知识库中，计算出新的模式向量存储在知识库中。

- 运用该模型对特征变量未来的取值进行预测和检验偏离，从而判断是否发生入侵。(异常检测）

PCA（ Principal Component Analysis ）： minor components of PCA (the subspace obtained after removing the components with largest eigenvalues) revealed anomalies

统计检测

1、操作模型:假设异常可通过测量结果与一些固定指标相比较得到，固定指标根据经验值或一段时间内的统计平均得到，例：短时间内的多次失败的登录可能是口令尝试攻击；

2、多元模型，操作模型的扩展，同时分析多个参数实现检测；

3、方差模型，计算参数的方差，设定其置信区间，当测量值超过置信区间的范围时表明有可能是异常；

4、马尔柯夫Markov过程模型，将每种类型的事件定义为系统状态，用状态转移矩阵来表示状态的变化，当一个事件发生时，或状态矩阵该转移的概率较小则可能是异常事件；

5、时间序列分析，将事件计数与资源耗用根据时间排成序列，如果一个新事件在该时间发生的概率较低，则该事件可能是入侵。



例子：

统计分析方法在入侵检测中的应用：

对于网络流量，可以使用统计分析的方法进行监控，这样可以防止拒绝服务攻击(DDos)等攻击的发生。如果设定某个端口处每秒钟允许的最大尝试连接次数是1000次，那么如果检测发现某个时间段内的连接此次超过此限，就视为异常，需要进行异常处理，以判断是否存在攻击。描述如下：

```
set max-connect-number = 1000/s;
set state =normal;
connect-number = count(connect);
if(connect-number> max-connect-number)
{
set state= abnormal;
进行异常处理;
}
```



基于统计分析方法

- 优点：
  - 不需要很多先验知识，有较为成熟的统计方法可以应用
  - 动态性好，用户行为改变时，相对应的度量能产生一致性的变化，保证行为模式的更新
- 问题：
  - 难以提供实时检测和自动响应功能：大多 数统计分析系统是以批处理的方式对审计记录进行分析的，因此检测系统总是滞后于审计记录的产生
  - 对入侵发生的顺序不敏感：许多预示着入侵行为的系统异常都依赖于事件的发生顺序，但是统计分析的特性导致了它不能反映事件在时间顺序上的前后相关性，因此事件发生的顺序通常不作为分析引擎所考察的系统属性;
  - 阈值难以确定：门限值如选择得不当，就会导致系统出现大量的错误报警。

❖专家系统

入侵的特征抽取与表达，是入侵检测专家系统的关键。

基于规则的入侵检测技术：在系统实现中，将有关入侵的知识转化为**if-then**结构，条件部分为入侵特征，**then**部分是系统防范措施。

基于状态转移图的入侵检测技术：状态转移图用来描述复杂和动态入侵过程的时序模式特征，可以表示入侵事件发生的时序关系和相关性，使入侵的行为、状态、上下文环境背景和发生的过程与步骤得到直观的描述。



基于专家系统入侵检测技术

- 基于专家系统的检测技术的特点：
  - 误报少准确性高
  - 只能发现已知攻击，难以准确识别同一种攻击的变种，对未知的攻击不具备检测的能力。同时规则库的建立及维护代价高，且容易出现冗余、矛盾、蕴含等问题。
- 运用专家系统防范有特征入侵行为的有效性完全取决于专家系统知识库的完备性，知识库的完备性又取决于审计记录的完备性与实时性。



基于生物系统模拟的检测技术

- 基于神经网络：由神经元通过突触连接。如**BP**网络是一种多层前馈神经网络，包括输入层、隐层和输出层。当学习样本提供给网络后，在输出层得到对输入的响应，按照减少目标输出与实际输出误差的方向，从输出层经过各隐层逐层修正各连接权值，以达到神经网络的实际输出与期望输出的最大拟和，从而实现分类。
  - 例：异常 检 测 工具：**NNID(Neural Network Intrusion Detector)**
  - 浅层网络被**SVM**（**supporting vector machine)**超越（ 其在解决小样本、非线性及高维模式识别方面的优势）
  - 深度学习兴起： **DBN(deep belief network), CNN(convolutional neural netwok)**
- 特点：
  - 需要学习训练，系统有可能趋向于形成某种不稳定的网络结构，不能从训练数据中学习到特定的知识
  - 不使用固定的系统属性集来定义用户行为，具备了非参量化统计分析的优点
  - 通常无法对判断为异常的事件提供任何解释或说明信息，不利于对入侵进行分析并采取相应对策



- 人工免疫的检测技术：生物免疫系统具有健壮性、记忆能力、容错能力、动态稳定性以及异常检测等良好特性”这些特性与一个合格的网络入侵检测系统有很高的相似性

- 遗传算法：基于选择、交叉和变异等基因操作。以适应度函数**fittest function**为启发式搜索函数，通常以分类正确率为度量，确定能表达某一类攻击的各参数特征。



基于数据挖掘

- 数据 挖 掘 (data mining)也称为知识发现技术，其目的是要从海量数据中提取出我们所感兴趣的数据信息（知识）：统计学的数学理论+机器学习的计算机实践
  - 预测：根据数据其他属性的值来预测特定属性的值
    - 分类的任务是对数据集进行学习，从而构造拥有预测功能的分类函数或分类模型（分类器），把未知样本标注为某个预先定义的类别。
    - 离群点分析（**outlier mining**）：发现离群点并对其进行处理的过程。离群点是与数据集中大部分数据的观测值明显不同的数据。
  - 描述：发现概括数据中潜在的联系模式
    - 聚类分析特别适合用来讨论样本间的相互关联，在事先对数据集的分布没有任何了解的情况下，按照数据之间的相似性度量标准将数据集自动划分为多个簇。
    - 关联分析用于寻找数据集中不同项之间的潜在的联系。例如，通过关联规则挖掘发现数据间的关系，或通过序列分析发现有序事物间的先后关系。



- 入侵检测系统中的数据挖掘算法，目前主要包括3种::
  - 数据分类(data classification)：连接(会话)记录的误用检测
  - 关联分析(association analysis)：用户行为模式的异常检测
  - 序列挖掘(sequence mining)：用户行为模式的异常检测
- **MADAMID(Mining Audit Data for Automated Model for Intrusion Detection)**
  - 误用检测，离线检测，利用规则分类算法RIPPER对审计数据进行归纳学习来得到描述类的模型
- **ADAM(Audit Data Analysis and Mining)**项目
  - 异常检测，关联规则与分类

结果处理模块

- 告警处理
- 响应

❑当事件出现时显示攻击的特征信息

❑重新配置防火墙

❑阻塞特定的TCP连接

❑邮件，传真，电话提示管理员

❑启动其它程序来阻止攻击

❑SNMP陷阱

❑生成报告

动态响应

❑模式匹配

❑统计分析

❑文件对象完整性分析

❑系统的全面扫描

❑证据搜寻

静态响应



告警融合

- 网络入侵检测系统分析的数据源是网络数据包，在一些情况下很容易突然产生大量相似的警报，称之为警报洪流。
- 例如攻击者可以通过发送大量经过精心设计的数据包使得入侵检测系统出现警报洪流，或是所检测的网络中某些服务器提供的一些固有服务产生的数据可能被误检测为入侵数据从而出现警报洪流。
- 在出现警报洪流时，入侵检测系统检测到的真正入侵行为所产生的警报就会被淹没，很难被管理员发现 。因此有必要实现告警融合。

- 研究：关联分析方法对IDS产生的告警进行关联



告警聚集

- 由于警报洪流中的警报一般是相似的，相似的报警在一个较短时间内多次出现是没有必要的，因此可以通过将多条相似的警报合并成为一条警报从而避免出现警报洪流或降低警报洪流的规模。这就叫做告警聚类，将特征相似的警报合并在一起，聚类（cluster)算法所依据的规则是警报的相似规则。
- 通过事先定义好的攻击过程进行事件关联：通过机器学习或人类专家来得到各种攻击过程，将这些攻击过程作为模板输入到系统中去，然后系统就可以将新的报警同这些攻击过程模板相比较，进行实时关联。
- 通过事件的前因和后果进行事件关联：任何一个攻击都具有前因和后果。所谓前因就是攻击要实施所必须具有的前提条件，后果就是攻击成功实施后所造成的结果。在一个有多个攻击动作组成的入侵过程中，一个攻击的后果就是下一个攻击前因。基于这一思想，首先定义每一个单独攻击的前因、后果，然后就可以将具有因果关系的攻击关联在一起，重现整个攻击过程。



入侵检测技术发展方向

- 入侵或攻击的综合化与复杂化
- 入侵主体对象的间接化
- 入侵或攻击的规模扩大
- 入侵或攻击技术的分布化
- 攻击对象的转移

- 分布式入侵检测
- 智能化入侵检测
- 全面的安全防御方案



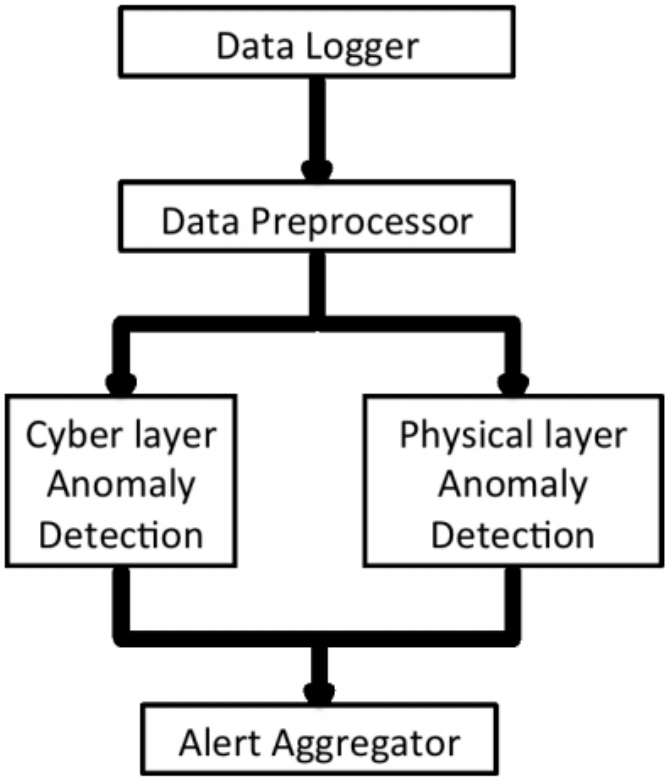
新一代的IDS

- 更为高速的协议分析
- 更为智能的模式语言
- 更低的误报率和漏报率
- 全局管理、知识共享
- 安全事件融合和相关、数据挖掘

与扫描器的相关
与漏洞管理系统的相关

- 专用硬件
- **IDP/IPS，从单纯检测到实时阻断**
- 各种不同网络的**IDS**

- …



从IDS到IPS

- IPS（Intrusion Prevention System，入侵防御系统）
- IDS（Intrusion Detection System，入侵检测系统）