第四章 入侵检测

概述

- 入侵检测（Intrusion Detection）是对企图入侵、正在进行的入侵或者已经发生的入侵进行识别的过程，旨在识别针对计算机网络或系统的网络攻击（Cyber Attack）。
- 传统的入侵检测方法在防护网络攻击方面逐渐暴露出诸多弊端，比如严重的误报（False Positives）和漏报（False Negatives）问题。数据挖掘的特性使得它在解决传统入侵检测的误报和漏报问题方面具有优势，目前，已经有大量数据挖掘技术中的数据分析方法用于改善入侵检测的性能。

目录

入侵检测技术

1. 入侵检测技术的发展
2. 入侵检测的分析方法
3. 入侵检测系统

- **入侵检测技术**是防火墙技术的合理补充，它扩展了网络空间安全管理人员或系统的安全管理能力。

入侵检测技术的发展

- **概念诞生阶段**

- 1980年，Anderson及其公司为一个保密客户写了一份题为《计算机安全威胁监控与监视》（Computer Security Threat Monitoring and Surveillance）的技术报告，第一次详细阐述了入侵检测的概念。

入侵检测技术的发展

- **模型发展阶段**

- 1984年到1986年期间，乔治敦大学的Denning和SRI公司计算机科学实验室的Neumann研究出一个实时入侵检测系统模型，命名为“入侵检测专家系统”（IDES）。该系统模型为构建入侵检测系统提供了一个通用的框架。1987年，Denning在前期工作的基础之上，发表论文“一种入侵检测模型”（An Intrusion Detection Model），文中提出了入侵检测的基本模型。

入侵检测技术的发展

- **百家争鸣阶段**

- 1990年，加州大学戴维斯分校的Heberlein发表论文“一种网络安全监控器”（A Network Security Monitor），该论文设计的NSM第一次将网络数据包作为入侵检测的信息源，它利用捕获的TCP/IP分组数据，监控异构网络环境下的异常活动。该成果第一次构建了基于网络的入侵检测技术，为入侵检测技术的发展史翻开了新的一页。

入侵检测技术的发展

- **继续演进阶段**

- 1990年以后，逐步出现集成了实用入侵检测技术的商用入侵检测产品，并形成了更多的入侵检测技术及其分类。早期的入侵检测技术仅仅执行网络攻击监测或者提供有限的数据分析功能，随着技术的发展，许多新型的入侵检测技术或者增加了应用层数据分析的能力，或者能够配合防火墙进行联动，形成功能互补，可更有效的阻断攻击事件，形成了入侵防御的功能。

入侵检测的分析方法

- **误用检测**

- 误用检测又称为特征检测(Signature-based Detection)，它将已知的入侵活动用一种模式来表示，形成网络攻击特征库，或称为网络攻击规则库。该方法对输入的待分析数据源进行适当处理，提取其特征，并将这些特征与网络攻击特征库中的特征进行比较，如果发现匹配的特征，则指示发生了一次入侵行为。

入侵检测的分析方法

- **误用检测**
  - **优点**
    - 误报率低，能够准确地识别已知的攻击，可详细地报告出网络攻击类型
  - **缺点**
    - 漏报率高，对新的入侵方法无能为力

入侵检测的分析方法

- **异常检测**

- 异常检测搜集正常活动的规律，将待检测的活动与这些正常活动的规律进行比较，对于违反正常活动统计规律的活动，认为该活动可能是入侵行为。

入侵检测的分析方法

- **异常检测**
  - **优点**
    - 不需要对每种入侵行为进行定义，可以检测到未知的入侵行为，不需要维护庞大的网络攻击特征库
  - **缺点**
    - 无法明确入侵行的类型，误报率高，难点在于如何不把正常的活动误认为“入侵”或忽略真正的“入侵”行为

入侵检测系统

- **入侵检测系统**

- **入侵检测系统**（IDS）是指执行入侵检测任务和功能的系统，它将入侵检测技术植入到可部署的系统中，对计算机网络或系统中违反安全策略的行为进行识别和相应处理。
- 根据IDS输入数据源（待检测的数据）的不同，IDS通常被分为：
  - 基于主机的入侵检测系统（HIDS）
  - 基于网络的入侵检测系统（NIDS）
  - 混合入侵检测系统（Hybrid IDS）

入侵检测系统

- **基于主机的入侵检测系统**

- 该系统分析的数据源来自于主机操作系统的事件日志、应用程序的事件日志、系统调用、端口调用和安全审计记录等，它一般保护所在的主机系统。

入侵检测系统

- **基于主机的入侵检测系统**
  - **优点**
    - 不需要额外的硬件，对网络流量不敏感，入侵检测的效率高，能准确定位入侵并及时进行反应
  - **缺点**
    - 直接依赖于主机的可靠性，可移植性差，所能检测的攻击类型受限，占用主机资源，影响主机运行性能

入侵检测系统

- **基于网络的入侵检测系统**

- 该系统分析的数据源来自于网络上传输的原始流量，它保护其所处网段的计算机网络。该类型系统被动地监听网络上传输的原始流量，对获取的网络数据进行处理，从中提取有用的信息，再通过协议分析、特征匹配、统计分析等手段发现当前发生的网络活动是否为网络入侵行为。

入侵检测系统

- **基于网络的入侵检测系统**
  - **优点**
    - 不依赖操作系统，可以检测协议攻击、特定环境的攻击等多种攻击，不影响主机的运行性能
  - **缺点**
    - 无法检测发生在应用级别的入侵行为，无法得到主机系统的实时状态，精确度较差

入侵检测系统

- **混合入侵检测系统**

- 该系统综合了基于主机的入侵检测系统和基于网络的入侵检测系统的数据源，它分析的数据来自于网络流量和主机操作系统的日志等本地主机信息。

入侵检测系统

- **混合入侵检测系统**
  - **优点**
    - 入侵检测的手段更加科学，能够更加准确地发现入侵行为
  - **缺点**
    - 需要大量的设备和具有良好性能的硬件资源，系统实现难度大，成本高

数据挖掘在入侵检测中的应用

- **基于分类方法的入侵检测**
- **基于关联分析方法的入侵检测**
- **基于聚类分析方法的入侵检测**
- **数据挖掘在入侵检测规避与反规避中的应用**

- 数据挖掘技术一直是入侵检测技术中重要的数据分析方法。数据挖掘是一个交叉的研究领域，它采用统计学模型（Statistical Model）、数学方法（Mathematical Algorithm）、机器学习（Machine Learning）等方法，可以发现大型数据集中信息的未知模式和关系。近年来，数据挖掘技术在大数据分析中的作用愈加明显，这也恰好可以为降低传统入侵检测技术的误报和漏报问题提供解决途径。

基于分类方法的入侵检测

- **目标**
  - 将待检测的源数据分类为正常活动和入侵行为

基于分类方法的入侵检测

- **过程**

- 首先，使用包含正常和各种入侵的历史数据作为训练数据；
- 其次，应用分类算法在数据上进行学习，建立分类模型，分类模型可以转化为识别正常活动和各种入侵行为的规则；
- 最后，使用这些规则对新的待检测数据进行分类判断，判断它是正常活动还是入侵行为。

基于分类方法的入侵检测

- **应用实例**

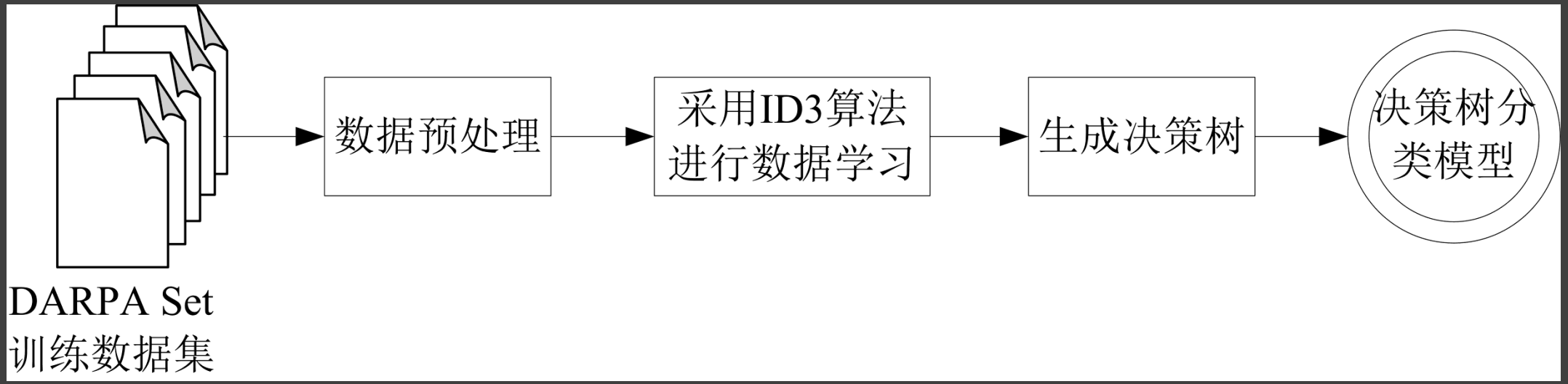
- 决策树分类方法在入侵检测中的应用
- 贝叶斯分类方法在入侵检测中的应用

决策树分类方法在入侵检测中的应用

- 采用DARPA 98林肯实验室评测数据集（DARPA Set）作为训练数据集和测试数据集，介绍了从DARPA Set中采用ID3决策树分类算法构建决策树的全过程，其过程如图所示，其构建的决策树可以用于网络攻击分类。

决策树分类方法在入侵检测中的应用

- 采用DARPA 98林肯实验室评测数据集（DARPA Set）作为训练数据集和测试数据集，介绍了从DARPA Set中采用ID3决策树分类算法构建决策树的全过程，其过程如图所示，其构建的决策树可以用于网络攻击分类。



贝叶斯分类方法在入侵检测中的应用

- 基于朴素贝叶斯分类器的异常检测方法是通过在任意给定的时刻，测量A_1、A_2、…、A_n变量值，推理判断系统是否有入侵事件发生。其中每个A_i变量表示系统或网络活动不同方面的特征。假设A_i变量有两个值，1表示异常，0表示正常。I表示系统当前遭受入侵，每个异常变量A_i的异常可靠性和敏感性表示为$P(A_i=1/I)$和$P(A_i=1/\neg I)$，则在给定每个A_i值的条件下，由贝叶斯定理得出I的可信度如式(4-1)所示。

$$P\left(I\left|A_1,A_2,\ldots,A_n\right.\right)=P\left(A_1,A_2,\ldots,A_n\left|I\right.\right)\frac{P\left(I\right)}{P\left(A_1,A_2,\ldots,A_n\right)} \quad (4\text{-}1)$$

贝叶斯分类方法在入侵检测中的应用

- 其中，要求给出I和$\neg I$的联合概率分布，又假定每个测量A_i仅与I相关，且与其它的测量条件A_j无关，$i \neq j$，则有式(4-2)和式(4-3)。

$$P\left(A_1, A_2, \ldots, A_n \middle| I\right) = \prod_{i=1}^{n} P\left(A_i \middle| I\right) \quad (4\text{-}2)$$

$$P\left(A_1, A_2, \ldots, A_n \middle| \neg I\right) = \prod_{i=1}^{n} P\left(A_i \middle| \neg I\right) \quad (4\text{-}3)$$

- 由以上各式得到式(4-4)

$$\frac{P\left(I \middle| A_1, A_2, \ldots, A_n\right)}{P\left(\neg I \middle| A_1, A_2, \ldots, A_n\right)} = \frac{P(I)}{P(\neg I)} \frac{\prod_{i=1}^{n} P\left(A_i \middle| I\right)}{\prod_{i=1}^{n} P\left(A_i \middle| \neg I\right)} \quad (4\text{-}4)$$

- 因此，根据各种异常测量的值、入侵的先验概率及入侵发生时每种测量到的异常概率，能够检测判断入侵的概率.

基于关联分析方法的入侵检测

- **关联分析方法**

- 关联分析（Association Analysis）方法用于发现隐藏在大型数据集中的有意义的联系，其发现的联系可以用关联规则（Association Rule）或频繁模式（Frequent Pattern）的形式表示。

基于关联分析方法的入侵检测

- **应用实例**

- 网络入侵的关联规则分析
- 网络入侵事件关联分析
- 网络入侵报警关联分析

网络入侵的关联规则分析

- 从网络连接记录中提取网络数据特征，并分析这些数据特征间的频繁模式，以此建立网络入侵的关联规则。
- 采用tcpdump捕获网络连接记录，然后从这些记录中提取其原始特征信息。部分网络连接记录信息的例子如表所示。

网络入侵的关联规则分析

timestamp	duration	service	src_host	dst_host	src_bytes	dst_bytes	flag	...
1.1	0	http	spoofed_1	victim	0	0	S0	...
1.1	0	http	spoofed_2	victim	0	0	S0	...
1.1	0	http	spoofed_3	victim	0	0	S0	...
1.1	0	http	spoofed_4	victim	0	0	S0	...
1.1	0	http	spoofed_5	victim	0	0	S0	...
1.1	0	http	spoofed_6	victim	0	0	S0	...
1.1	0	http	spoofed_7	victim	0	0	S0	...
...	...	...	...	...	...	...	...	...
10.1	2	ftp	A	B	200	300	SF	...
12.3	1	smtp	B	D	250	300	SF	...
13.4	60	telnet	A	D	200	12100	SF	...
13.7	1	smtp	B	C	200	300	SF	...
15.2	1	http	D	A	200	0	REJ	...
...	...	...	...	...	...	...	...	...

网络入侵的关联规则分析

- 入侵检测可以被认为是一种分类问题，它将网络连接记录分类为正常和入侵网络活动。采用RIPPER分类算法对上述表格中的网络连接记录进行分类，分类结果如下表所示。

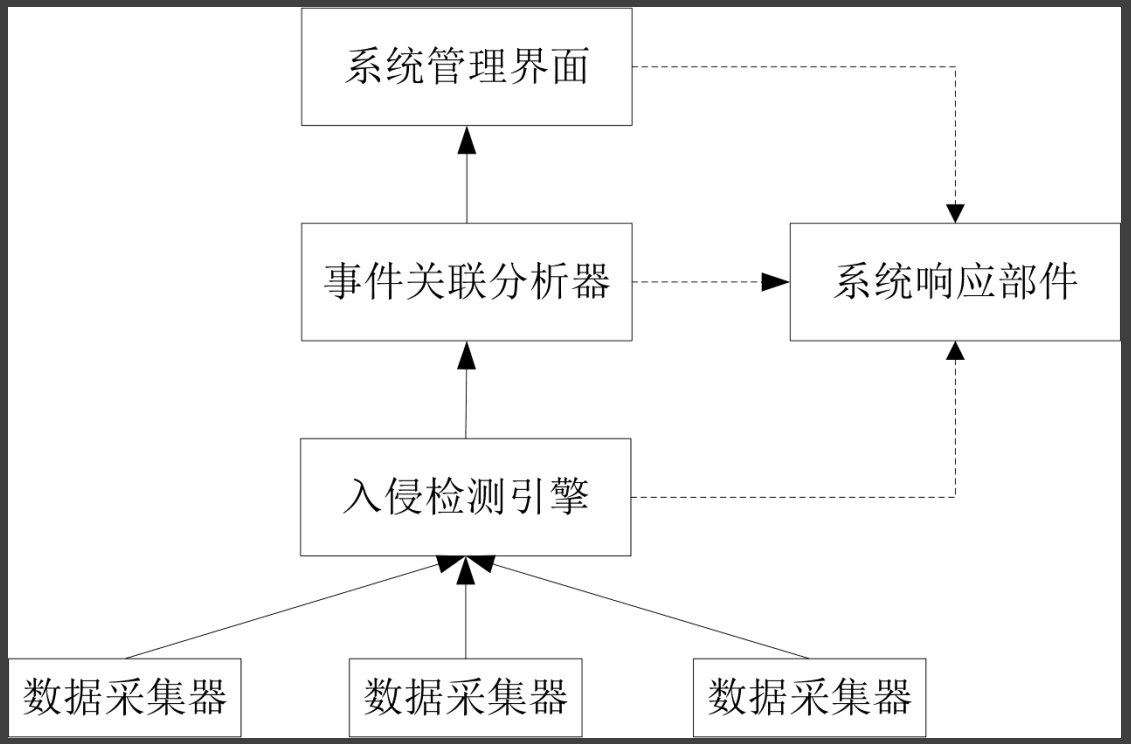
网络入侵的关联规则分析

- 入侵检测可以被认为是一种分类问题，它将网络连接记录分类为正常和入侵网络活动。采用RIPPER分类算法对上述表格中的网络连接记录进行分类，分类结果如下表所示。

label	service	flag	host_count	srv_count	host_REJ_%	host_diff_srv_%	duration	...
normal	ecr_i	SF	1	1	0	1	0	...
smurf	ecr_i	SF	350	350	0	0	0	...
satan	user-level	REJ	231	1	85%	89%	0	...
normal	http	SF	1	0	0	1	3	...
...	...	...	...	...	...	...	...	...

网络入侵事件关联分析

- 在入侵检测系统中加入事件关联分析，就是在入侵检测引擎和系统管理界面之间增加一个起到事件再加工作用的事件关联层，即事件关联分析器，并用其关联到系统响应部件，如图所示即IDS及事件关联分析器。



网络入侵事件关联分析

- **网络事件之间的关系**
  - 冗余关系
    - 是由一个简单攻击或一个复杂攻击的某个步骤触发多个入侵检测引擎或多次触发一个入侵检测引擎所引起网络事件间的联系
  - 因果关系
    - 在一个完整的攻击过程中，由单个攻击步骤触发的多次网络事件之间的联系，在因果关系中，前因事件导致了后果事件

网络入侵事件关联分析

- **冗余关系关联分析方法**
- **因果关系关联分析方法**

- 网络入侵事件的关联分析中，一般先进行冗余关系分析，将重复的多个事件聚合为一个事件，再进行因果关系分析，以减少不必要的重复计算。

网络入侵事件关联分析

- **冗余关系关联分析方法**

- 对于冗余关系所采用的事件关系分析方法主要是依据事件攻击特征(*Attack_Specif*)中相关属性之间的相似度。

网络入侵事件关联分析

- **因果关系关联分析方法**

- 判断事件间因果关系所采用的事件关联分析方法主要基于攻击事件模型*E*的三个字段：*Attack_Precond*，*Attack_Postcond*，*Attack_Specif*，其基本思想是：寻找一个攻击事件的前因(*Attack_Precond*)和另一个攻击事件的后果(*Attack_Postcond*)之间是否存在逻辑联系，如果存在联系，就表明这两个攻击事件是关联的。

网络入侵报警关联分析

- 对一种频繁闭模式挖掘算法CLOSET[1]进行了改进，并采用改进的算法对网络入侵的报警进行关联分析。

网络入侵报警关联分析

- **具体算法**

- Step1 将报警消息存入事务数据库TDB;计算各项目数量: Count_items( ) ;
- Step2 将频繁项列表*FrequentItemList*倒排；
- Step3 根据*frequentItemList*, 获得事务中的频繁项列表, 并排序；
- Step4 根据事务长度对事务进行索引, 并返回事务列表, 生成构建FP-Tree的数据；
- Step5 初始化频繁闭项集*FCI* ；
- Step6 调用CLOSET子程序, 获得频繁闭项集*FCI* ；
- Step7 for each list in *FCI*

  if(list.length<min_depth) delete list;

基于聚类分析方法的入侵检测

- 该检测技术的基本流程为：
- 首先，对网络数据进行预处理，提取其关键特征信息，并对一些属性的值进行标准化；
- 其次，采用一定的聚类算法，利用测试数据集进行聚类操作，将测试数据分成若干的簇；
- 然后，对分出的簇进行标类操作，按照一定的原理，将不同的簇分为正常网络活动或入侵行为；
- 最后，将待检测的数据与分出的簇进行距离计算，判断其属于哪个簇，根据该簇是什么分类来判断带检测的网络数据是正常网络活动还是入侵行为。

基于聚类分析方法的入侵检测

- **应用实例**

- 基于无监督聚类（Unsupervised Clustering）的入侵检测算法，该算法是一种基于异常检测的入侵检测分析方法。
- 该算法包括三部分：
  - 数据预处理
  - 无监督聚类算法
  - 检测算法

数据挖掘在入侵检测规避与反规避中的应用

- **入侵检测规避**
  - 指对网络攻击的表现形式进行伪装转换，致使IDS无法识别，从而逃避IDS检测的活动。

数据挖掘在入侵检测规避与反规避中的应用

- **入侵检测规避与反规避技术**

- 由于IDS与被保护目的主机分析的数据的不一致性、网络协议分析的复杂性、入侵检测技术的缺陷等原因，目前已出现了大量入侵检测的规避技术。
- 入侵检测的规避技术可以在保留原有攻击行为的同时逃避IDS的检测，对IDS保护的目标主机或系统构成了严重的威胁，因此，入侵检测的反规避技术（Counter-evasion）也应运而生。

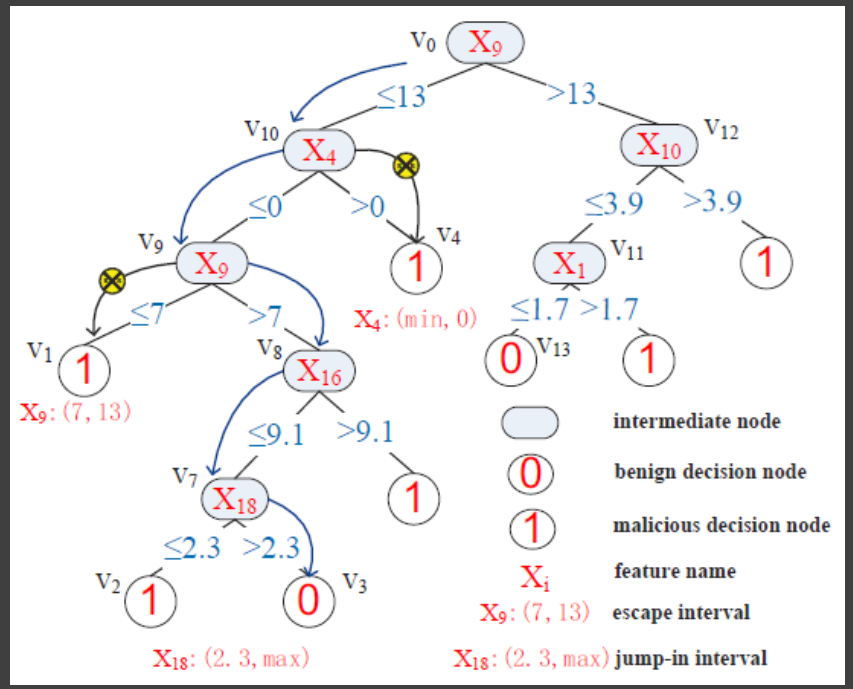
数据挖掘在入侵检测规避中的应用

- **应用实例**

- 以J48决策树分类器作为例子，介绍恶意网站检测的规避算法。
- J48分类算法构建的决策树如图所示。

数据挖掘在入侵检测规避中的应用

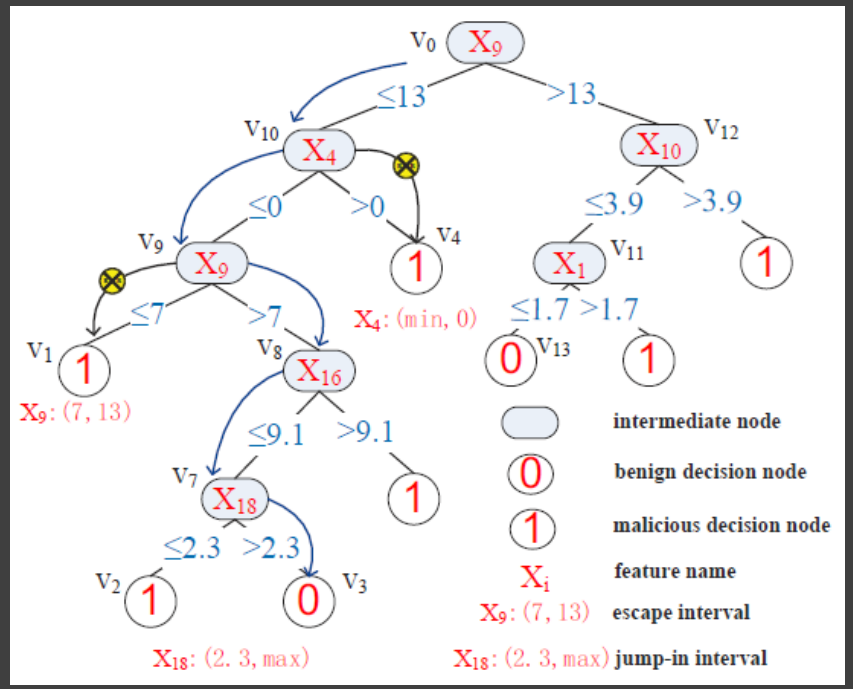
- **应用实例**



- 以J48决策树分类器作为例子，介绍恶意网站检测的规避算法。
- J48分类算法构建的决策树如图所示。

数据挖掘在入侵检测规避中的应用

- **应用实例**

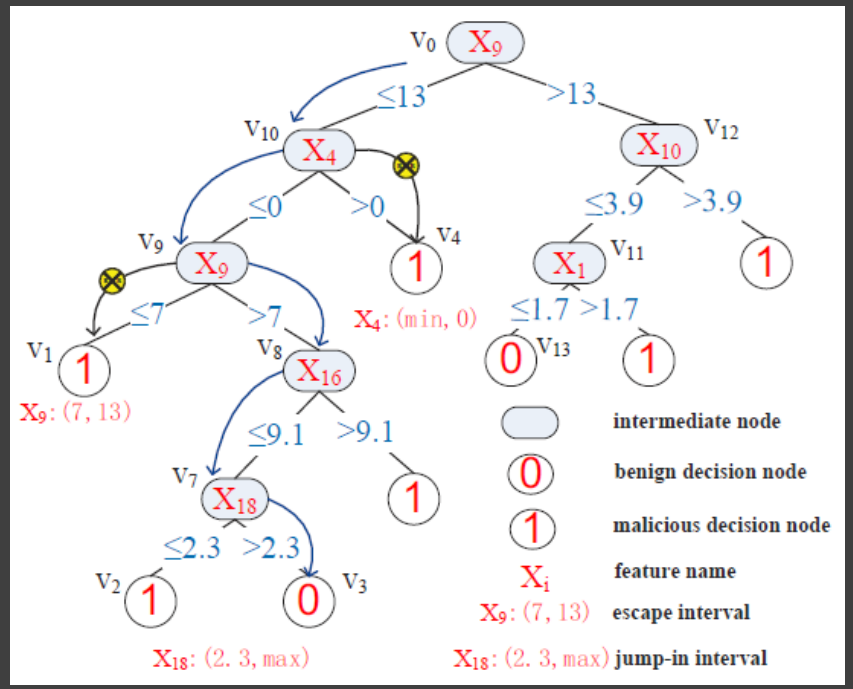


- 以J48决策树分类器作为例子，介绍恶意网站检测的规避算法。
- J48分类算法构建的决策树如图所示。

其中，叶子结点0表示正常网站分类，叶子结点1表示恶意网站分类，X_i（1≤i≤18）表示网站数据记录的属性名称。

数据挖掘在入侵检测规避中的应用

- **应用实例**



- 针对图示的决策树分类器，提出了两种规避算法：
  - 自顶向下操纵属性值
  - 自下向上操纵属性值

数据挖掘在入侵检测反规避中的应用

- 1、针对规避技术采用的数据属性改变方法，对朴素贝叶斯分类算法进行了扩展，对被成功规避的入侵实例进行重新分类，达到入侵检测反规避的目的。
- 2、采用一类SVM分类算法构建了一种基于异常检测方法的高速IDS，它从有效载荷（Payload）里提取待检测数据源的数据特征，并采用特征聚类算法对特征空间进行降维处理，达到检测基于规避技术的网络攻击的目的。
- 3、提出了针对Web攻击的入侵检测反规避方法，采用主动训练和检测策略，学习出检测适应性攻击的分类规则，达到恶意网站攻击检测的反规避目的。

小结

- **入侵检测作为一种主动的网络空间安全保障技术，它是防火墙技术的合理补充。**

- 首先，概述了入侵检测的基本概念及其面临的严峻挑战；
- 其次，介绍了入侵检测技术的发展状况、入侵检测的分析方法和入侵检测系统；
- 然后，重点介绍了几种数据挖掘技术在入侵检测中的应用，其中：
  - 分类方法部分介绍了决策树分类方法在入侵检测中的应用实例；
  - 关联分析方法部分介绍了网络入侵的关联规则分析、事件关联分析和报警关联分析的应用实例；
  - 聚类分析方法部分介绍了一种基于无监督聚类的入侵检测算法的应用实例；
  - 入侵检测规避与反规避部分概述了入侵检测规避与反规避技术，并介绍了数据挖掘技术在入侵检测规避与反规避中的应用情况。

Thanks!