第一部分：隐蔽信道类型及现有检测方式

网络存储隐蔽信道

核心原理：将编码后的隐蔽信息隐藏在各层协议的特定字段中，利用协议报文达到传输隐蔽信息的目的。

举例：

ACK Command：构造特殊的 ACK 报文，将隐秘信息直接放在 ACK 数据包的载荷中进行传输。

ICMP SHELL：利用防火墙等防护机制一般不会对 ICMP 报文的不限制性，将隐蔽信息加载在 ICMP 上，隐蔽信道可以获得较强的生存能力。

TCP/IP：将隐蔽信息直接填充在 TCP/IP 协议头的特定字段（如 TCP 序列号，IP 的 ID 号等），在接收端按照隐蔽信道的协议解析相应的字段，达到隐蔽信息的处理。

UDP：根据端口的多样性，如采用 8 个不同的端口号，每个端口号采用二进制编码，于是每个端口号可对应 3bit 信息。

应用层协议（DNS,HTTP 等）

现有检测方式：

网络存储隐蔽信道的检测方法目前有：

1. 基于特征库的模式匹配，即解析出可用于构建隐蔽信道的字段的值，与已有的特征库进行匹配，若匹配上，则存在隐蔽信道。其中，特征库来自于已知的公开隐蔽信道。该方法**简单，容易实现，**且**精确度高，**但是**耗时长**。

2. 基于数据挖掘、机器学习的隐蔽信道检测，即利用隐蔽信道对正常数据包所修改信息的规律性，收集大量数据包（包括隐蔽信道的数据包和正常数据包），分析是否存在明显的规律性，若存在，则说明存在隐蔽信道。在众多方法中，聚类、SVM 与 PCA（主成分分析）在一些研究中得到使用。该方法**效率高**，**适用性强**，且对网络环境的**依赖性较小**，但不足是实现过程较为**复杂**，且**精度不够高**。

3. 其他方法如马尔可夫链、CSP 等对网络行为进行建模，再进行检测。

网络时间隐蔽信道

核心原理：基于正常网络数据包的时间特性来构建的隐蔽信道，在时间上有较强的规律性。

举例：

基于时间间隔：发送者和接收者提前商定相邻分组到达时间间隔的调制机制，不同的时间间隔代表着不同的隐蔽信息。

一段时间内的数据包行为：

1. 发送者和接收者约定一定的时间间隔，在每个时间间隔内，发送者选择发送数据包或保持静默，以代表不同的信息。（如）。
2. 在每个时间间隔内，通过改变不同的数据流速率，来表示不同的信息。

数据包到达顺序：对于固定的若干个正常的数据包，通过对其到达接收方的不同顺序进行编码，可达到传输不同信息的目的。

现有检测方式：

1. 基于**统计**的检测方式。对于稳定的网络，网络流量一般会基本满足特定的数学分布与一些约束条件，而网络时间隐蔽信道的存在势必会改变这些特性，因此，通过检测和分析数据包流量的统计特性，隐蔽信道就能够被检测出来。

2. 利用数学知识，计算相应的**检测指标**来确定是否存在隐蔽信道。如数据包间隔方差指标、数据包间隔概率熵等。

基于用户行为的隐蔽信道

核心原理：基于发送端用户的行为，对用户行为进行编码，用正常的数据包构建隐蔽信道。

举例：

用户行为：规定一系列用户操作序列，对不同的操作序列进行编码，以表示不同的信息。

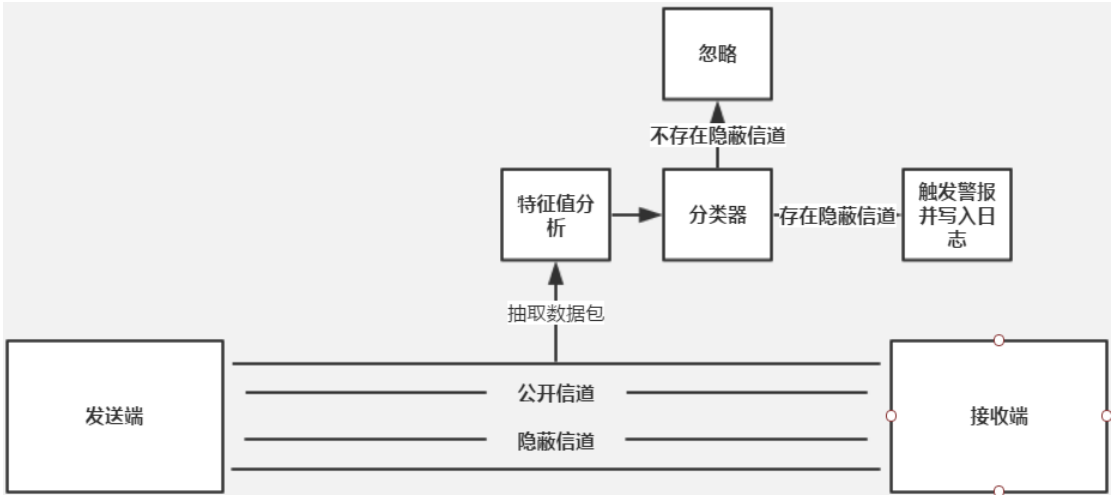
数据包长度：通过控制正常数据包的长度，并对其进行编码，达到传输隐蔽信息的目的。

现有检测方式：

基于用户行为的隐蔽信道与网络时间隐蔽信道类似，数据包、操作之间有着一定的规律性，因此，通过衡量数据包、数据包长度等之间的一些相关性（如熵的衡量），即可检测出是否存在隐蔽信道。

第二部分：检测思路

本报告致力于设计一个隐蔽信道的检测系统，该检测系统是基于数据挖掘分类算法，且结合了前人的研究，力求得到高精确度的同时，不影响正常信道的通信。
该系统的运行模块图如下：



图中，隐蔽信道与公开信道同时存在于同一信道中。在网络数据包传输的任意一个节点中，数据包以复制的方式被抽取出来，这样一来，该隐蔽信道检测系统对网络的影响就会降低，设抽取的数据包数量为 N，称这 N 个数据包为一个检测单元。将该检测单元传送到特征值分析模块，提取出若干特征值，以便分类器进行分类。这边的分类器主要使用的是二分类器，对到来的检测单元进行分类。

特征值分析模块：

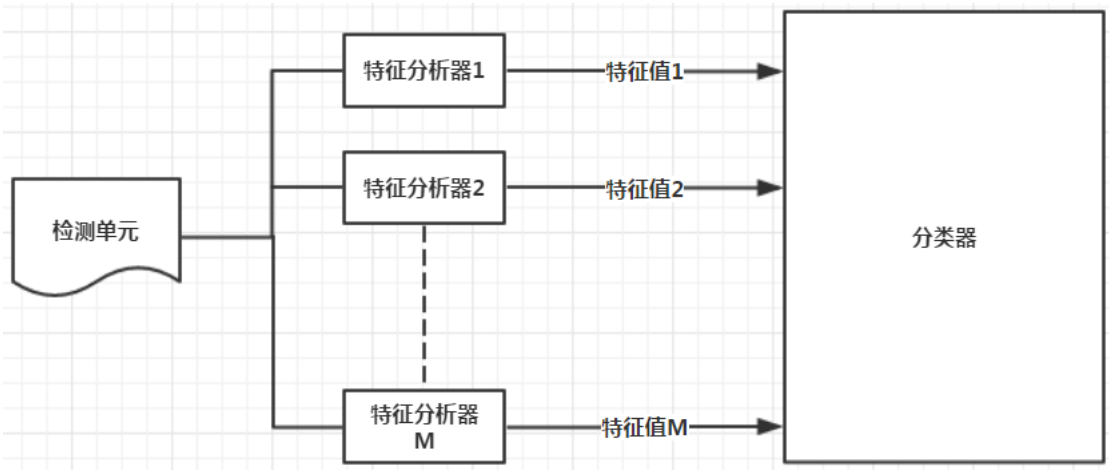
功能：对一个检测单元中的 N 个数据包进行特征值分析，以便分类器进行分类。

特征值包括（但不限于）：

1. 数据包传输时间间隔间的相关性指标
2. 数据包长度间的相关性指标
3. 数据包各层协议特定字段的分布特性
4. 相同数据包的数量

其中，相关性指标的衡量之前的很多研究都有过讨论，这边可以加以引用、改进。

特征值分析模块的运行图如下：

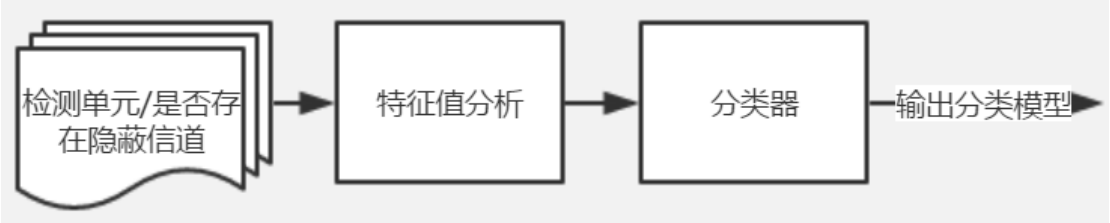


分类器：

隐蔽信道的检测可以考虑成一个二分类问题，因此数据挖掘分类算法可以运用于隐蔽信

道检测。

可用的分类器有：二项逻辑回归、决策树、朴素贝叶斯等。
分类器的学习流程如下：



图中，训练集来自于不同的网络传输信道，其是否存在隐蔽信道是已知的，其中检测单元中的数据包数量与从待检测信道中抽取的数据包数量 N 相等。通过对训练集的学习，各个学习器可生成各自的分类模型。

分类器的选择：
由于待测信道网络环境的差异，以及不同分类器的性能不同，需要对分类器进行选择，而分类器选择的标准即是分类器学习时准确率的高低。

N 值的确定：
N 值的确定与分类器的选择一样，选择使分类准确率最高的 N 值进行学习。