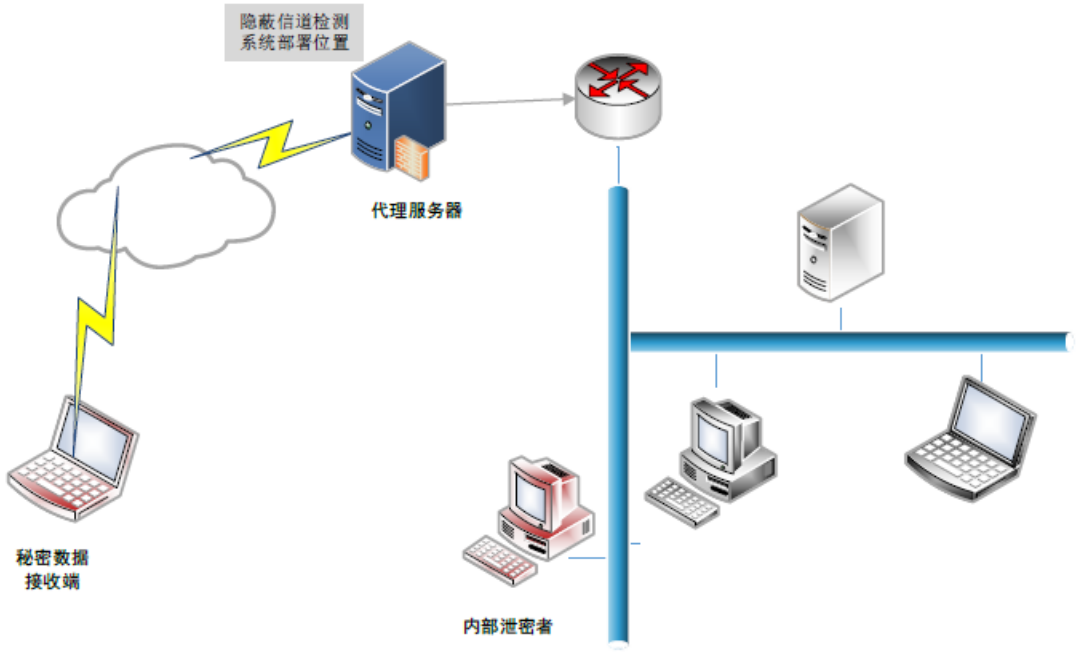
隐蔽信道检测系统设计方案

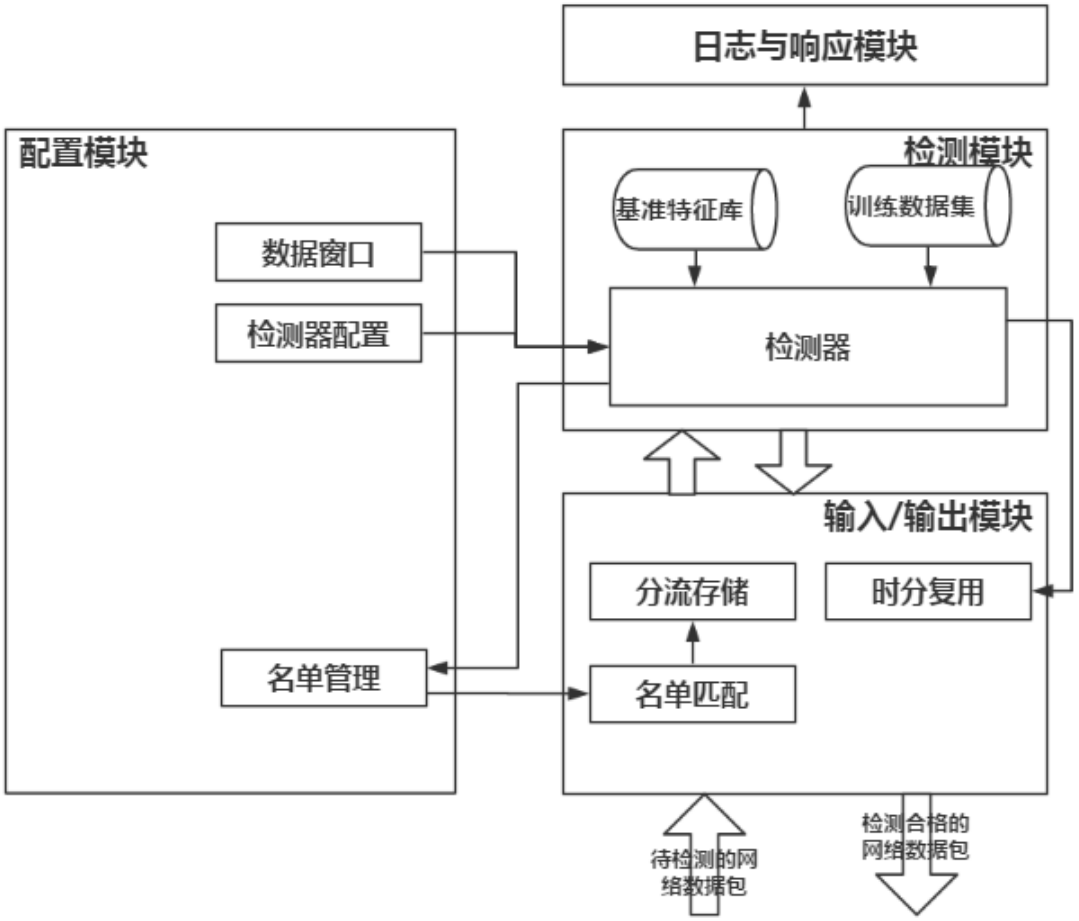
1. 系统部署

隐蔽信道检测系统部署在局域网的边界上，以对所有试图流出内部网络的数据包进行分析。如图所示，系统部署在一台代理服务器上，也可以为对外的网关服务器。



2. 系统架构

系统的整体架构图如下所示：



系统主要由四个模块组成：配置模块，输入/输出模块，检测模块以及日志与响应模块。下面分别阐述各模块的功能以及模块之间的联系。

1. 配置模块：配置模块允许对系统的运行参数进行配置，包括黑白名单的设置，检测器所用分类方法的选择以及数据窗口的设置。
2. 输入/输出模块：输入/输出模块实现对网络数据包的抓取、过滤以及转发，将存在于黑名单中的数据包阻拦、销毁，存在于白名单中的数据包进行转发，而其他数据包则在分流存储、拼接之后交由检测模块进行检测。另外，该模块接收来自于检测模块的数据包，并将其转发。
3. 检测模块：该模块实现对不存在与名单中的数据包的检测，主要包括基准特征库以及检测器。基准特征库储存的是当前网络状况下正常数据包的特征，其数据来源于输入/输出模块中存在于白名单中的数据包。
4. 日志与响应模块：该模块记录检测模块检测到的异常数据包的相关信息，并进行报警等一些响应。

3. 模块设计

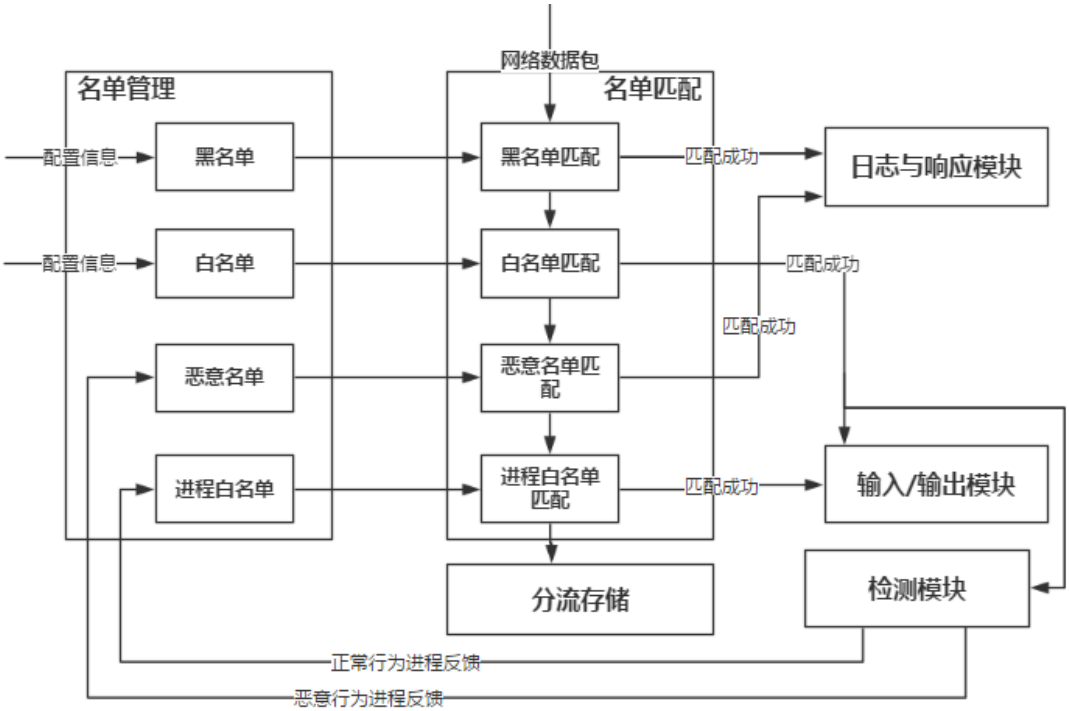
3.1 配置模块

1. 数据窗口：允许对检测模块单个线程所处理数据包的数量进行设置。
2. 检测器配置：包括选择检测器所用的分类方法，以及对该分类方法的一些参数的设置。
3. 名单管理：存有名单所带数据，并允许对黑白名单进行增、删、改、查的处理。

3.2 输入/输出模块：

1. 名单匹配

名单匹配功能如下图所示。



名单存储格式：在系统中，我们用数据包上的 IP 地址和端口组合来表示一个数据源。各个名单的存储格式如下图所示。

IP地址	端口号
192.168.0.2	33666
192.168.0.3	56548
192.168.0.3	48515
⋮	⋮

黑、白名单、进程白名单存储格式

IP地址
192.168.0.6
192.168.0.8
192.168.0.10
⋮

恶意名单存储格式

黑名单：手动设置的禁止向外界发送数据包的数据源，在黑名单中的数据源所发的数据包将被系统直接丢弃。

白名单：手动设置的安全数据源，在白名单中的数据源所发的数据包将被直接转发。

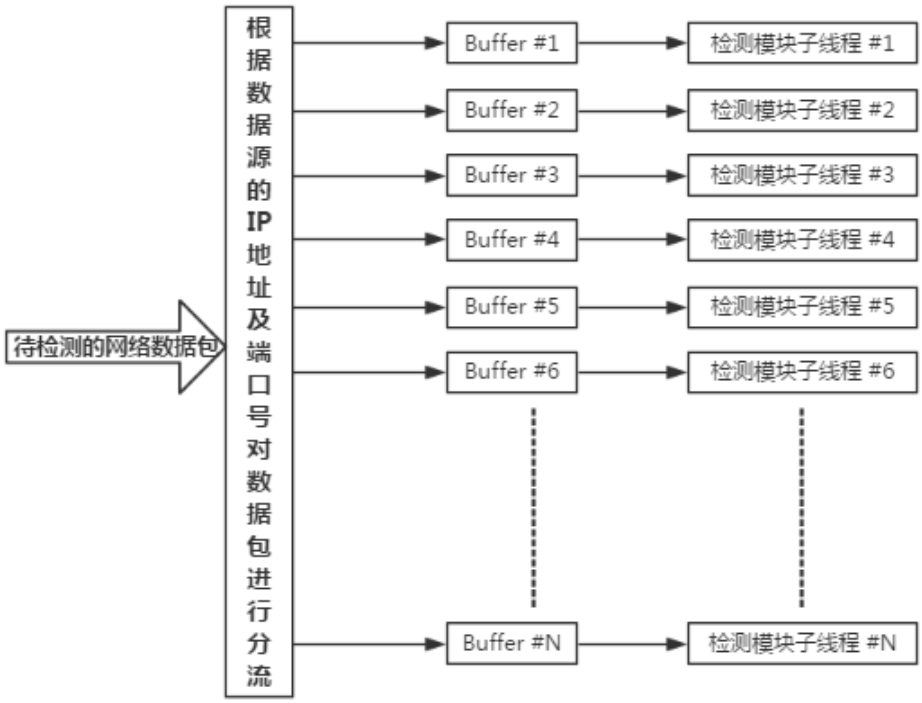
恶意名单：若数据包经过检测模块后表现的状况为异常，则我们认为数据包的数据源所属的主机为恶意主机，任何来自该主机的数据包都将直接被丢弃，直到管理员将其从恶意名单中删除。

进程白名单：若数据包经过检测模块后表现的状况为正常，则我们认为该信道不存在隐蔽信道，且将该数据包的数据源加入进程白名单，直到该信道关闭为止。

名单匹配：对到来的数据包根据四种名单依次进行匹配，并根据匹配结果作出相应的处理。需要特别提到的是，若白名单匹配成功，在转发的同时将数据包的副本作为基准数据交付给检测模块，以便保证检测模块的时效性。

2. 分流存储

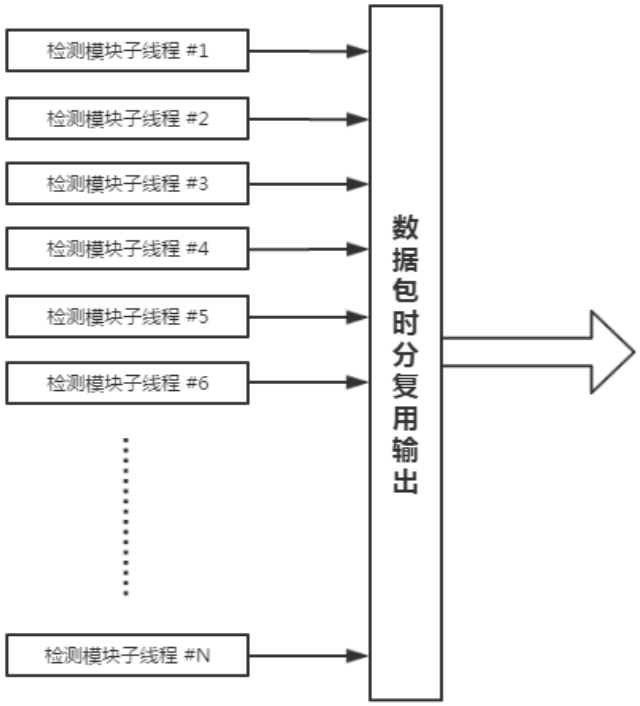
分流存储的功能如下图所示。



无法与任何一个名单相匹配的数据包将被分流存储，分流的依据是数据包上 IP 地址以及端口号组合。数据包被分流后，将被存放进存储器中，整流后供检测模块检测。

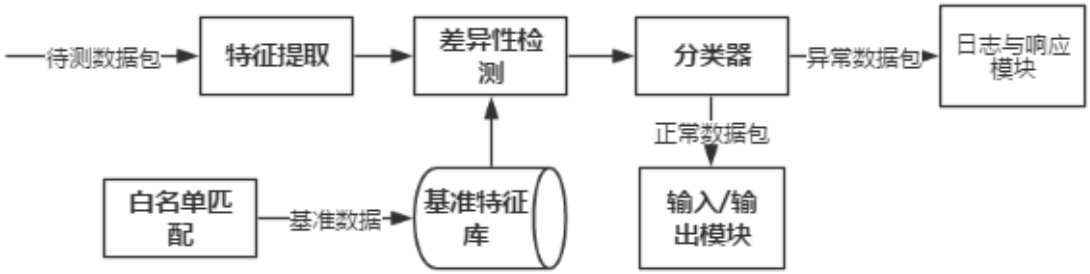
3. 时分复用

参考了通信系统中的时分复用技术，当要把检测正常的包转发出去时，由于分流的原因，需要将其再次整成一个数据流。如图所示。



3.3 检测模块

检测模块是检测系统中的核心模块，它的性能直接决定了系统的有效性和可用性。其构架如图所示。



1. 基准特征库：基准特征库来自于名单匹配中白名单中数据源所发的数据包，它们作为基准数据，为检测其提供当前网络状态下的基准特征。（关于特征的选择将在后文说明）
2. 特征提取：对待分析的数据包进行特征提取，以便检测器的检测。

   特征选取：由于输入/输出模块中存储器的使用，使得时间隐蔽信道不复存在，因此，我们在这里提取协议中可能存在隐蔽信道的字段特征，以便对数据

包进行分类。

a) 数据包长度的分布：隐蔽信道可能通过不同的数据包长度来进行编码，因此判断待测数据包长度的分布是否与基准特征库中的一致，可以检测出是否有隐蔽信道的存在。

b) TCP 各个可疑字段的分布：对于 TCP 协议来说，很多字段比如保留字段、序列号等都能被用来构建隐蔽信道。然而，无论隐蔽信道通过什么编码方式，一旦隐蔽信道被建立，则所用字段的值的分布必然和原始分布有所区别，因此通过衡量隐蔽信道构建前后字段值分布的差异，隐蔽信道就能被检测出来。如 ISN（初始序列号）的随机性可以被用来构建隐蔽信道。根据之前的研究，隐蔽信道的数据大多以 ASCII 码的形式存储在 TCP 序列号的高 8 位，通过比较待测数据包与基准特征库中 ISN 高 8 位的分布，可以检测出隐蔽信道的存在。

c) 报文类别的分布：隐蔽信道可能通过对不同的报文种类进行编码从而达到传输隐蔽信息的作用，然而这就改变了原本网络中各个类别报文的数量，将其与基准特征库进行对比，即可检测出来。

d) ICMP 的特征：ICMP 能被用来构建隐蔽信道的主要是它的荷载，荷载的内容取决于数据源所在的操作系统的类型，通过直接检测荷载值可检测隐蔽信道。下图是不同系统下的 ICMP 荷载值。

	ICMP Payload
Null Packet	0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
Win Packet	0900 6162 6364 6566 6768 696a 6b6c 6d6e 6f70 7172 7374 7576 7761
Solaris Packet	50ec f53d 048f 0700 0809 0a0b 0c0d 0e0f 1011 1213 1415 1617 1819
Linux Packet	9077 063e 2dbd 0400 0809 0a0b 0c0d 0e0f 1011 1213 1415 1617 1819

3. 差异性检测：使用特定的算法计算待检测数据的特征与基准特征之间的差异性，以供分类器进行分类。

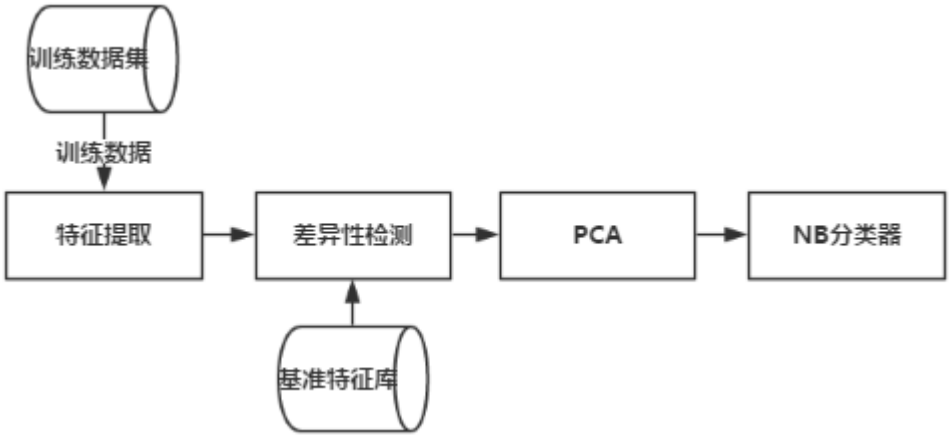
K-S 检测：由于所提取的特征含有分布特征，我们利用 K-S 检测来衡量待测数据包的分布与特征数据库中的差异性。在实际操作中，我们认为正常的数据包分布成正态分布，并使用 EM 算法估算其参数，然后再比较待测数据包分布与其的差别。

ICMP 特征匹配：对于 ICMP 来说，我们将正常的载荷记为“1”，异常的记为“0”以方便分类器的分析。

4. 分类器：分类器先根据数据窗口的大小从输入/输出模块的存储器中抓取响应的数据包，再根据待测数据包特征与基准特征库的差异性，判断待测数据包是否存在隐蔽信道，然后交由其他模块处理。

   分类器的选择：分类器采用数据挖掘的分类方法对数据包进行分类，在之前的研究中，SVM，聚类等分类方法在隐蔽信道检测中得到了运用。在这个系统中，我们计划初步实现基于主成分分析（PCA）和朴素贝叶斯（NB）的方法来实现，在后期实现自适应的分类方法选择，以适应不同的网络环境。

   分类器的学习：在进行有效的检测之前，分类器需获得当前网络状态的相关知识，因此需要进行学习。基于 PCA 和 NB 的分类器学习过程如下。



   训练数据来自于系统所部署的主机，其中包括了带有隐蔽信道的数据以及正常数据，分类器通过对训练数据的学习，建立起一套评判标准，以准确地对待检测数据进行检测。

   训练数据集中带有隐蔽信道的数据将使用各类基于 TCP 和 ICMP 的存储型隐蔽信道生成。