A Detection Method for Anomaly Flow in Software Defined Network

IEEE/2018

一、整体结构

采集流量

控制器

特征提取

预处理

异常检测机制

异常检测

检测流程：

（1）OpenFlow交换机向中央控制器提交信息，通知中央控制器当前的流状态，同时向控制器发送请求最新的流表转发规则。

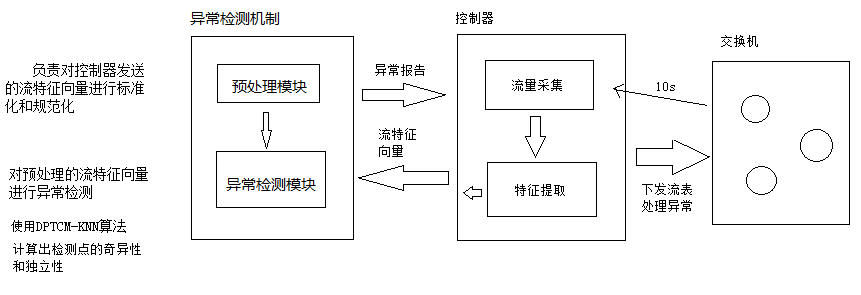
（2）中央控制器从交换机发送的流量表信息中提取流量特征，由特征提取模块进行汇总，形成多维的流量特征向量，并将其发送到异常检测机制。

（3）异常检测机制对接收到的流量特征向量进行预处理（标准化、规范化），采用DPTCM-KNN算法计算出检测点的奇异性和独立性，完成异常检测。

（4）异常检测机制在接收到检测结果后生成异常检测报告，并通过安全通道定期向中央控制器发送报告。中央控制器可以根据网络情况控制与异常检测机制相关的参数。

（5）中央控制器根据接收到的报告修改流表转发规则，并将其发送给OpenFlow交换机进行异常处理。

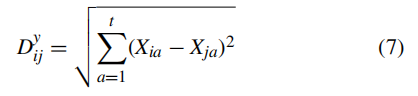
详细图解：



二、算法分析：

1.两点之间的欧几里得距离：

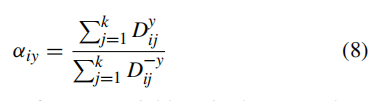
欧几里得距离通常用于测量两点之间的接近度，并表示为（y类中点i和点j之间的欧几里得距离）



其中，Xi和Xj分别是训练点i和j的提取特征的值集；是第i个向量X的第a个属性的值；t是向量第i个向量X的长度。

2.奇异性：

y和-y是正常元素的集合和异常元素的集合

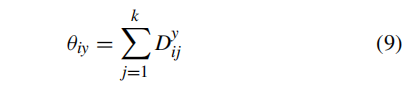


这个比值就叫做奇异性

其中k是knn算法中最近邻的个数，是i相对于y的奇异性。奇异性描述了一个点对一个类别的归属程度。奇异性越大，点属于类别的可能性就越小。

3.独立性：

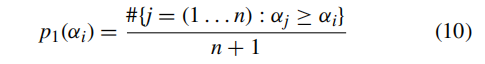
独立性是检测点与其k个最近邻点之间欧几里得距离之和，即两者之间的绝对偏差。



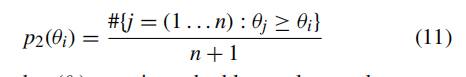
式中，为i相对于y的独立性；k为最近邻数；为i与第j最近点之间的距离。同时，表示检测点对某一类别的隶属度。独立性越小，点越可能属于该类别。

4.双P值：

一个检测点对一个类别的隶属度。p值越大，点越可能属于y（法向点）。



式中，αi为y中i的奇异性，αj为第j个训练点相对于y的奇异性，#为有限集的单元数，即奇异性大于y中i的奇异性的训练点个数，p的计算公式如下：



p1（αi）和p2（θi）构成双p值，作为dptcm-knn算法的异常检测标准，只有满足这两个要求，检测点才被认为是正常的。

算法步骤

步骤1（计算训练集样本之间的欧几里得距离）

步骤2（计算训练集样本的奇异性和独立性）

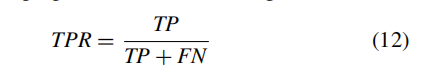
步骤3（计算检测点的奇异性和独立性）

步骤4（获取检测点的双p值）

步骤5（识别异常）

三、性能对比：

1. 检测率TPR

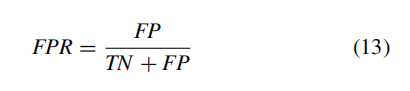


其中TP表示被识别为异常流的异常流的数量，FN表示被识别为正常流的异常流的数量。 换句话说，TPR是正确分类的异常流量与异常流量总数的百分比。

* 1. 本文选择（K，τ1，τ2）=（20,0.01,0.05）作为DPTCM-KNN算法的最优参数组合。
  2. 当R取0.75时，DPTCM算法达到最高TPR，换句话说，当正常和异常子集 的数量达到4到3时，该算法具有最高的检测率。

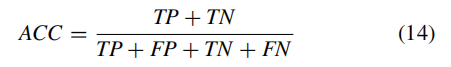
在本文中，选择（K，τ1，τ2，R）=（20，0.01，0.05，0.75）作为DPTCM-KNN算法的最佳参数组合。

1. 误报率FPR



其中FP表示被识别为异常流的正常流的数量，FN表示被识别为正常流的正常流的数量。 换句话说，FPR是正确分类的正常流量与正常流量总数的百分比。

1. 精度ACC



ACC是正确分类的样本数与样本总数的百分比，反映了分类器的辨别能力。

1. ROC曲线，AUC（曲线下面积）

四、实验结果：

与其他算法相比，DPTCM-KNN算法显著提高了异常流的分类性能，可以看出DPTCM-KNN算法具有较高的检测率、准确性和较低的假阳性率，可以在sdn环境下实现高精度的异常检测。