Identification of encrypted VOIP application using deep learning approach

**KeyWords** deep learning; bi-flow; VOIP; traffic identification

**Abstract** We proposed a new way to identify the VOIP application in this paper. We construct the matrix by the traffic, which we got from a call . There is a rule that we extracting the voip traffic. In this paper, we applied DNN to analyze the voip traffic in a complete connection. We

文章针对加密的voip应用提出了一种使用deep learning提取通话开始到结束过程完整流量特征的方法，并使用所获得的特征进行voip应用识别。我们使用dnn网络对一次完整的voip通话所产生的流量进行分析，将获取的特征应用到svm进行应用分类。（文章中提到了不同的应用有其特有的连接开始的标志，我们分析了几种广泛使用的voip应用的建立连接的模式，并建立了确定的规则提取voip完整通话产生的流量。）我们捕获了大量的voip流量数据，并按照特定规则进行处理后输入dnn网络。最后对svm分类的结果进行了客观的评估。我们的结果显示，文章所提出分类方法具有较高的准确率。

**1. Introduction**

**Recnetly,** voip应用随着其服务质量的提高凭借其优越的价格优势，在广大人群中的得到了越来越多的使用。Voip应用的广泛使用使网络监管的难度大大提高。对voip应用流量进行有效的识别可以帮助解决网络监管问题。

流量识别方面最传统的使用ip地址和应用端口号进行识别的方案在voip应用使用nat技术和使用特殊端口号的前提下已变得不可行。深度包检测的技术也因为Voip应用普遍的加密特性变得不可行。因此，有研究人员开始使用机器学习的方案对voip应用进行识别，通过对voip流建立特征集进行训练（Identification of VoIP encrypted traffic using a machine learning approach），他们得出了很好的结果。在此基础上，本文将深度学习的思想应用到voip应用识别，通过深度学习生成特征集，从而进一步提高识别精度。

最后，第二节为相关工作介绍；第三节介绍使用深度学习提取应用特征；第四节介绍使用第三节中所提取的特征应用svm进行分类；第五节展示试验结果；第六节总结工作并对下一步工作进行探讨。

**2. Related work**

我们收集了大量的voip流量，本文采用监督学习方法，首先对数据贴上标签。之后按照特定规则进行数据预处理。将其处理成我们所需的形式输入到dnn网络。 得到特征后，使用svm进行分类。对试验结果进行评估。

**3. Methodology**

本文主要对加密的voip应用进行分类操作，学习的过程采用有监督的学习方法。传统的机器学习方法需要人为的去创建特征集，此类方法不但效率低下，并且其准确率没有保障，如采用特征平均包长度来作为特征进行分类，当碰到两个应用的udp报文长度相等时（如本文uucall和altcall）准确率会大大降低。前面提到，本文采用深度学习的思想去获取流量特征，减少人为工作的同时也提升了分类准确率。