# 第五周

1. **分类方法**
2. 基于端口分析，没用
3. 基于包分析，使用数据包内容，提取子序列、正则表达式等，用数据挖掘算法（不太热门）
4. 基于流分析，使用统计特征，如包平均大小、时间间隔等，用机器学习算法（这个比较流行）
5. 其他一些新提出的方法，主要根据网络工作特点进行设计。
6. **机器学习算法**
7. 无监督，主要是K均值和各种改进算法
8. 有监督，SVM、决策树、贝叶斯比较常用，深度学习（神经网络）比较新，还有一些没人用过的算法比如XGboost
9. 半监督，结合上面两种方法进行改进
10. **数据挖掘**
11. 关联分析，Apriori、FP-Growth等
12. 序列挖掘，Prefixspan等
13. **协议/应用**
14. P2P、媒体流量的识别比较有价值
15. 云应用正在变多，这类流量可能也有研究价值
16. 加密流量识别很重要