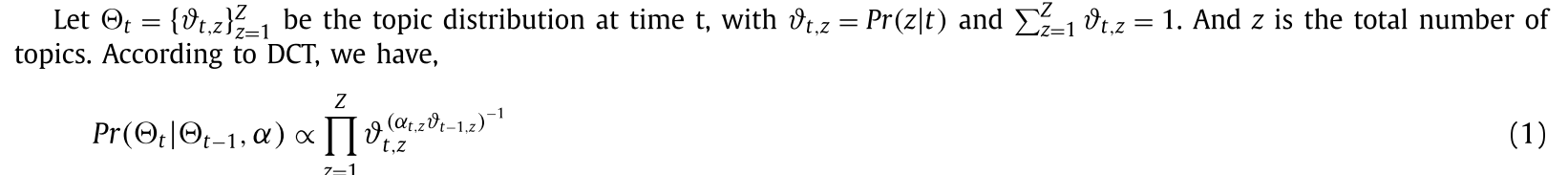
DMTCS考虑动态了时间上前后两个包的关系，他假设t时刻包的协议分布与t-1时刻包的协议分布有关。

把数据包当作document，用了一种新的聚类算法DCT，这是一种主题模型聚类算法。大致意思是主题模型中的两个分布在t时刻只与t-1时刻有关，相当于在LDA中加入了时间关系，下一篇文档的两个分布只和上一篇文档两个分布有关。



通过DCT得到每个包的主题分布，然后给他打上的标签，标签就是最可能的那个主题。然后相同的有相同标签的包聚为一类，这个聚类的类别就是该聚类中包出现最多的类别。

这样分完一个流量包之后，聚类就有可能更新。

Vocabulary的大小W是500-1500。在做n-gram generator的时候，生成的词只保留出现频率最大的W个。