Chapter 29

Nonparametric Bayesian Prediction of Primary Users' Air Traffics in Cognitive Radio Networks

Ju Bin Song
Kyung Hee University, South Korea

Zhu Han
University of Houston, USA

ABSTRACT

In cognitive radio networks a secondary user needs to estimate the primary users' air traffic patterns so as to optimize its transmission strategy. In this chapter, the authors describe a nonparametric Bayesian method for identifying traffic applications, since the traffic applications have their own distinctive air traffic patterns. In the proposed algorithm, the collapsed Gibbs sampler is applied to cluster the air traffic applications using the infinite Gaussian mixture model over the feature space of the packet length, the packet inter-arrival time, and the variance of packet lengths. The authors analyze the effectiveness of their proposed technique by extensive simulation using the measured data obtained from the WiMax networks.

INTRODUCTION

Due to non-cooperative nature and limitation of signaling in cognitive radio networks, it is difficult for secondary users (unlicensed users) to be aware of the activities of primary users. On the other hand, it is very important for the secondary users to be aware of primary user's activity (such as air traffic patterns) in order to efficiently utilize primary user channel and prevent primary user from any harmful interference caused by secondary users. Channel occupancy statistics of primary users have high impacts on the performance of cognitive radio networks. For example, a secondary user can cluster the air traffic patterns of primary users, consequently, the secondary users can utilize the air traffic patterns to optimize their transmission strategy and save the unnecessary sensing overhead. Therefore, it is of significant importance for secondary users to detect the primary users' air traffic patterns.

DOI: 10.4018/978-1-4666-6571-2.ch029
The rapid proliferation of Peer-to-Peer (P2P) and Skype applications have enriched the traffic over the Internet. The P2P traffic is using more than 50% of the total traffic on the Internet (Mellia, Pescape, & Salgarelli, 2009). Different types of applications show different behaviors in terms of features that have distinguishing characteristics (Bonfiglio, Mellia, Meo, Rossi, & Tofanelli, 2007; Choi & Hossain, 2011; Gu, Zhang, & Huang, 2011; Nguyen & Armitage, 2008). Features for different payloads can be packet (or frame) length (observed from packets’ TCP header), connection time (from SYN and FIN flags from the packet header), packet inter-arrival time (time between two consecutive packets of same application), the variance in packet lengths, and many others that are specific to different applications. Existing solutions for the identification of traffic payload falls into three categories. 1) Port based, this traffic payload identification, (Erman, Mahanti, Arlitt, Cohen, & Williamson, 2007), is not valid now anymore due to the fact that many applications use dynamic ports (like P2P) and try to disguise themselves by using known port numbers. 2) Payload based techniques that rely on deep packet inspection, (Haffner, Sen, Spatscheck, & Wang, 2005; Moore & Papagiannaki, 2005; Zander, Nguyen, & Armitage), that matches pre-defined signatures of applications. This technique performs well if the payload is not encrypted. Due to following reasons we cannot rely on payload based techniques, a) Encryptions of payloads prevent us to inspect the packet, b) Signatures of applications may change with time, different versions have different signatures (MSN2009 and MSN2011), c) Obfuscated data can lead to serious problems, Skype application uses data obfuscation (Bongfiglio, Mellia, Meo, Rossi, & Tofanelli, 2007). 3) Heuristic approaches are not accurate for classifying traffic applications (Bongfiglio, Mellia, Meo, Rossi, & Tofanelli, 2007; Mellia, Pescape, & Salgarelli, 2009).

In this chapter, we adopt a nonparametric Bayesian approach for clustering traffic applications, where the nonparametric means the number of hidden traffic applications are unknown. In the proposed scheme, secondary users observe primary channels by taking information from the packet header. Our approach differs from other clustering algorithms in the following ways 1) It is based on observations without any exchange of control messages and increased overhead. 2) Our approach is based on the unsupervised clustering algorithms (Ghosh & Ramamoorthi, 2003). 3) It is also worth mentioning that the proposed scheme is flexible in such a way that more features can be added to increase the accuracy of clustering algorithm. It does not assume any prior knowledge about the number of different traffic applications. We model the feature space of a single traffic cluster as a multivariate Gaussian distribution with unknown parameters. For multiple traffic clusters, we model it as infinite Gaussian mixture model. We formulate our problem as a nonparametric Baysian approach to cluster the traffic with unbounded number of applications in an unsupervised manner. The Dirichlet distribution is used to define the prior, the collapsed Gibbs sampling algorithm is used in conjunction with Dirichlet prior, and the collapsed Gibbs sampling samples from the posterior and estimates the number of clusters based on the Chinese restaurant process (Ghosh & Ramamoorthi, 2003). The effectiveness of the proposed scheme is validated via simulated data and real measurement data traces. For real measurement data, we used the WiMax wireless traces for different applications like Game, VoIP and UDP (Available from http://crawdad.cs.dartmouth.edu/~crawdad/snud　wow_via_wimax/; http://crawdad.cs.dartmouth.edu/~crawdad/kaist/wibro/). The data is measured at different places (campus, subway, residential) with traffic variations. The results validate the effectiveness of the proposed scheme in traffic clustering.