Abstract: Anomaly detection refers to the process of identifying unexpected objects or patterns, which do not conform to the usual behavior. The detection of “not-normal” observations has attracted a lot of research interest from the machine learning community since it has a wide variety of practical applications.

In this talk, I will briefly present an overview of the challenges of unsupervised anomaly detection. I will also present our novel model-based approach that relies on the multivariate probability distribution associated with the observations. Since the rare events are present in the tails of the probability distributions, we use copula functions, which are able to model the fat-tailed distributions well. The presented procedure scales well; it can cope with a large number of high-dimensional samples and also with missing values.

I will also demonstrate the usability of the method through a case study, where we analyze a large dataset consisting of the performance counters of a real mobile telecommunication network.