



For now we do this because we do not have much real-world data. The problem is that the test data (and any anomalies) eventually joins the training data, polluting it



Eventually we will separate the traffic into two databases

