# 基于Spark Streaming的实时数据分析系统及其应用

韩德志，毕坤，戴永涛,刘罕

上海海事大学

在并发网络访问中，快速有效的从海量数据流中分离出异常访问，识别网络攻击并及时做出反馈具有重要意义。为了实现对实时网络数据的快速分析，本文设计一种分布式的实时数据流分析系统，能有效解决并发访问数据的收集、存储和实时分析问题，为大数据环境的网络安全检测提供了一种有效的数据分析方法；根据Spark运行的原理设计并实现了一种动态采样的K-Means并行算法，能快速有效地检测大数据环境下的各种DDoS攻击。分析和实验结果显示，基于Spark Streaming的实时数据分析系统有好可扩展性、容错性和实时处理能力，与DDoS检测系统融合后能大大缩短攻击的检测时间。