Proactive Configuration of Data Centre Networks for Big Data Processing

Harpreet Singh, M.Sc.
University of Dublin, Trinity College, 2016

Supervisor: Stefan Weber

Various studies have determined that the network is a performance *bottleneck* in Big Data processing applications running in the cloud such as Hadoop. Numerous attempts have been made to alleviate this network bottleneck by traffic engineering during execution of the applications, using Software-Defined Networking. Such measures of traffic engineering are overwhelmingly *reactive* in nature and are bound to induce control traffic overhead in the network. In this project, we propose a *proactive* approach for configuring Data Centre Networks as the means to optimize application traffic, specifically Hadoop; thereby accelerating the execution of applications in the cloud.

We configure the network before execution of the application, to determine if there is a performance gain when there is no control overhead in the network. The network is configured *proactively*, by logging the flow decisions made by the *reactive* algorithms from previous studies. These flow rules are subsequently installed in the routing devices before the execution of the application, after which, the flows are routed *reactively*. We demonstrate an *average gain in network bandwidth utilization* between 11.9% to 59.9% in comparison to *reactive* approaches, while Hadoop job completion times are reduced by 10% to 33.5%.