

With the sheer scale of devices that are being built and used day by day, there has come with it the simultaneous growth of new types of applications that are highly latency-sensitive. For this class of applications the traditional cloud model, whereby an application communicates with a data centre(s) is not always optimal for its latency sensitive requirements.

The fog, initially proposed by Cisco, is a middle layer between the cloud and the edge of the network, allowing applications to be run on hardware and resources throughout the network, such as routers and switches. This allows the application to meet its latency sensitive requirements by using already provisioned network hardware close to the edge of the network.

With the rise of the fog, a new set of research challenges have emerged. Some of these include how to program for the fog and the edge, how to manage and place applications throughout the different hardware resources on the network, and how to scale those applications in an optimal manner. There is currently no way of allowing developers to be able to program their applications across the “full stack” from the edge to the fog to the cloud (and make full use of the advantages that come with them).

My PhD research focuses on building a resource-aware custom scheduler on Kubernetes that accommodates for all of these different parameters – allowing applications to be optimally placed across the “full stack” of nodes on the edge, the fog and cloud. This custom scheduler, taking in different parameters such as the CPU power needed by the application, memory, energy requirements, etc. will choose and place the application on the most suitable node. The end goal is for the developer to be able to simply submit their applications with their requirements, and for the scheduler to decide where to place the applications, 2) how to scale the applications 3) when to move applications across the stack depending on the user’s requirements 4) how to manage the applications effectively in the face of heterogeneity of the different cloud providers, network hardware, etc.