Latency overhead of network and network function virtualization in 5G

5G networks are envisioned to support diverse use cases and their underlying network requirements. These use cases include, but are not limited to emergency services, massive machine type communications,intelligent transport services and autonomous driving, smart cities, human machine communication and enhanced mobile broadband services [1]. The use cases have diverse service requirements. For example while critical communications like emergency services and autonomous driving may require ultra low latency but not so high bandwidth, enhanced mobile broadband will place higher emphasis on high bandwidth while relaxing constraints on latency. To meet such diverse service requirements, 5G networks are envisioned to implement logical network slices on top of shared physical resources wherein each use case of service level is assigned a dedicated slice which is tuned to the service requirements of that slice[2]. Network slices may be end to end and include the RAN or they may be limited to the core network. A network slice comprises all the network functions needed to implement a telco network, but the network functions shall be logical rather than the traditionally physical equipment. Logical as used here refers to software running on generic hardware.
The two key enabling technologies of network slicing are software defined networking (SDN) and network function virtualization (NVF), both leveraging cloud computing and virtualization technologies. Both SDN and NVF introduce virtualization overhead as one might guess. These overheads might become network bottlenecks in some scenarios, for example while video streaming. The objective of this thesis would be to quantify the latency overhead in a 5G network slice. As a starting point, the latency overhead caused by a single virtual network function (VNF) may be studied. Typically a VNF would run in a host machine which runs a hypervisor and the VNF would deal with network traffic on a l2/l3 tunnel. A more comprehensive investigation would simulate a full fledged 5G (core) network tunnel and measure end to end latency overhead as well. This may be done from a use case scenario perspective, for example video streaming or other multimedia application.

Related material:

1. Future mobile communication forum, White Paper 3: End to End Network Slicing http://www.wwrf.ch/files/wwrf/content/files/publications/outlook/White%20Paper%203-End%20to%20End%20Network%20Slicing.pdf
2. Franco Callegati, Walter Cerroni, and Chiara Contoli, “Virtual Networking Performance in OpenStack Platform for Network Function Virtualization,” Journal of Electrical and Computer Engineering, vol. 2016, Article ID 5249421, 15 pages, 2016. doi:10.1155/2016/5249421 https://www.hindawi.com/journals/jece/2016/5249421/
3. Trivisonno, Riccardo, Xueli An, and Qing Wei. "Network slicing for 5G systems: A review from an architecture and standardization perspective." Standards for Communications and Networking (CSCN), 2017 IEEE Conference on. IEEE, 2017.
4. 5G Americas White Paper – Network Slicing for 5G and Beyond

[1] https://www.ericsson.com/en/5g/use-cases
[2] https://5g.co.uk/guides/what-is-network-slicing/

5. Chauhan, Maneesh. "Measurement and analysis of networking performance in virtualised environments." (2014).