Thesis Abstract

In the current era of pervasive Internet access and rapidly expanding network infrastructures, real-time communications (RTC) have ascended as a linchpin of modern digital interaction, underpinning a wide spectrum of use cases ranging from remote work and telemedicine to online gaming and live media streaming. These applications entail ever-escalating demands for reliability, ultra-low latency, adaptive performance, and technologically advanced, robust, and scalable frameworks. In this context, machine learning (ML) techniques, which have flourished in recent years, have emerged as promising candidates for intelligently capturing complex network dynamics and enabling adaptive, data-driven, and proactive optimization of RTC systems.

This thesis centers on Real-time Transport Protocol (RTP)-based RTC, exploring various facets of ML algorithms---from coarse to fine features, from single to multiple objectives, and from theoretical examination to practical implementations---to progressively augment network performance and eventually elevate the Quality of Experience (QoE) in RTC applications.

We commence the research with an in-depth investigation of packet loss phenomena, aiming to preemptively identify the onset of future losses and thereby potentially avoid subsequent ramifications. We discover the predominance of continuous losses and implement as well as compare multiple ML approaches based on aggregated traffic features, successfully classifying the majority of lossy events without significantly penalizing normal conditions.

Secondly, we turn our attention to the well-established problem of traffic prediction, striving to estimate network throughput with a dedicated emphasis on traffic extremes, including peaks, valleys, and abrupt changes. Leveraging fine-grained packet-level information, we develop an innovative deep learning (DL) algorithm capable of adeptly recognizing/accommodating extremities.

Moving forward to multiple objectives, we next propose a multi-task learning framework that integrates packet- and flow-level correlations with a knowledge distillation paradigm to efficiently predict a range of Quality of Service (QoS) metrics for each RTP flow in a single shot.

Lastly, we transition from theoretical feasibility analyses to the evaluation of practical performance gains. We scrutinize the congestion control (CC) mechanism in RTC, refuting the necessity of reinforcement learning (RL) and ascertaining the deterministic role of bandwidth estimation (BWE), which motivates the design of a simple yet effective regressor.

In summary, we conduct a comprehensive study of RTC at both the application and network levels, uncovering latent deficiencies and formulating various problems. In response, we provide ML-based remedies with different, specific targets, while progressively addressing pertinent constraints. Our work is anticipated to not only advance the state of the art in RTC optimization but also pave the way for more resilient, adaptive, and efficient RTC systems in increasingly complicated network environments.