There are proven benefits in terms of cost and convenience in delivering thin-client based virtual desktops, versus the use of traditional physical computers for end-user computing purposes. New cloud offerings such as "desktop-as-service" are rapidly being adopted among various communities, ranging from government agencies to research institutes and business entities. However, several challenges remain in evaluating hardware resources and thin-client protocol configurations for delivering virtual desktops with adequate user Quality of Experience (QoE). Suitable performance benchmarking methods and tools are necessary for service providers to optimize resources to provision large-scale virtual desktop requests, and achieve cost efficiency as well as user satisfaction. Moreover, performance benchmarking can serve as a troubleshooting capability to identify bottlenecks or to qualify whether a particular configuration of virtual desktop infrastructure is valid for the supported user application profiles.

In this thesis, we develop a novel benchmarking methodology and toolkit for virtual desktop environments by enhancing earlier works on slow-motion benchmarking and the VDBench toolkit. We focus on automation aspects of benchmarking, and extend the end-to-end performance traceability for common desktop applications such as Internet Explorer, Media Player and Excel Spreadsheets. Our approach prevents any invasive modification of thin-client systems, and allows emulation of user behavior with realistic workloads. We also address the user interface design issues of managing workflows between the client and server, and to easily instrument and generate comprehensive performance reports for complex environment setups. In a validation study, we deploy the enhanced VDBench toolkit in a real-world virtual desktop testbed that hosts applications that render 3D visualizations of disaster scenarios for scene understanding and situational awareness. Through the benchmarking results, we show how the toolkit provides user QoE assessments involving reliable video events display under different network health conditions and computation resource configurations.