Inference Latency vs. Network Bandwidth (TinyBERT-6l)