

**Online Scaling of NFV Service Chains Across Geo-Distributed Datacenters**

**Abstract:**

Network Function Virtualization (NFV) is an emerging paradigm that turns hardware-dependent implementation of network functions (i.e., middleboxes) into software modules running on virtualized platforms, for significant cost reduction and ease of management. Such virtual network functions (VNFs) commonly constitute service chains, to provide network services that traffic flows need to go through. Efficient deployment of VNFs for network service provisioning is a key to realize the NFV goals. Existing efforts on VNF placement mostly deal with offline or one-time placement, ignoring the fundamental, dynamic deployment and scaling need of VNFs to handle practical time-varying traffic volumes. This work investigates dynamic placement of VNF service chains across geodistributed datacenters to serve flows between dispersed source and destination pairs, for operational cost minimization of the service chain provider over the entire system span. An efficient online algorithm is proposed, which consists of two main components: 1) A regularization-based approach from online learning literature to convert the offline optimal deployment problem into a sequence of one-shot regularized problems, each to be efficiently solved in one time slot and 2) An online dependent rounding scheme to derive feasible integer solutions from the optimal fractional solutions of the one-shot problems, and to guarantee a good competitive ratio of the online algorithm over the entire time span. We verify our online algorithm with solid theoretical analysis and trace-driven simulations under realistic settings.

**Existing System:**

Existing efforts on VNF placement mostly deal with the offline or one-time placement, ignoring the fundamental, dynamic deployment and scaling demand of VNF service chains to handle practical time-varying traffic flows. In contrast, this work investigates dynamic placement of VNF service chains in geodistributed cloud datacenters to serve dynamically-generated flows between various source/destination pairs across the globe, for service cost and delay minimization. We show that even in the offline setting, the problem we consider

renders an NP-hard combinatorial nature, leading to significant difficulty in efficient online algorithm design. Looking deep into the structure of the problem, we design an efficient online algorithm based on the state-of-the-art online learning techniques and a well-designed dependent rounding scheme.

**Proposed System:**

We formulate a practical online cost minimization problem enabling dynamic deployment and removal of VNF instances in different datacenters, as well as dynamic traffic flow routing among VNF instances in the respective service chains. Various deployment and running costs of VNF instances in different datacenters are considered, in addition to time-varying bandwidth costs to transmit flows into and out of a datacenter. As a key QoS performance indicator for network service provisioning, the average end-to-end delays of the flows are also formulated and minimized as part of our objective.

We leverage a regularization based technique from the online learning literature to transform the relaxation of the integer offline optimization problem into a sequence of regularized sub-problems. In particular, the regularization eliminates temporal correlation among decisions across time slots by lifting the precedence constraints coupling successive time slots into the objective function, such that each of the subproblems can be efficiently and optimally solved in each time slot, using only information at the current time. By solving each sub-problem, VNF instance deployment/removal and flow routing decisions at the time are obtained in polynomial time, constituting part of the feasible (fractional) solution to the relaxed offline problem. Based on the KKT optimality conditions, we are able to show that an upper-bounded overall cost, as compared to the optimal offline solution, can be guaranteed by this (fractional) feasible solution. Moreover, we adapt the regularization framework into a general network flow model to handle the end-to-end delay.