Paper1: Probius-Automated approach for VNF and service chain analysis, SOSR'18

The paper argues that there is lot of performance uncertainty in NFV environments and there is no systematic framework to identify the root cause of performance problems in VNFs. Although, there is some prior work on improving the performance of VNFs, they are focused on individual elements of NFV infrastructure. The performance problems are mainly at: VNF services, virtual switch, hypervisor, hardware. Moreover, they show that the throughput changes significantly with different network chains or adding more NFs, which makes it hard for network operators to distinguish which VNF is causing the problem.

So, they take an approach that systematically understand the NFV platforms and behaviours of NFV service chains. Then, they create various service chains based on some configurations and resource constraints, which provides insights about performance issues before the VNFs are deployed. Next, they build a comprehensive monitoring and tracing tool for all the NFV entities. Finally, they develop a performance analyzer to understand the root cause of performance issues using the collected traces.

The service chain builder generates all possible combinations of service chains while VNF manager creates these service chains based on all possible resource constraints. The workload manager emulates network workloads based on the given configurations from network operator. The configurations from operators are range of resource constraints, workload configurations, VNF configurations such as type and CPU memory constraints, and lastly service chaining policies such as which VNFs are together or followed by which.

To understand the NFV environment, they examine the VNFs workflow/relation with the underlying hypervisor and service chains. This investigation extracts important tracing points at all entities of an NFV platform, that affect the performance of VNFs. They use three monitors to collect these features and build graph based representation to maintain the workflow synchronization. After collecting the data, they first filter the outliers and then identify the suspicious service chains with anomaly detection (Cook's distance, regression analysis). Finally, they analyse the performance individual VNFs as well as difference combinations of service chains and reason about common causes of performance issues.

Paper2: CoCo-Compact and optimized consolidation of modularized NFs, ICC'18

The paper argues that modularizing NFs creates performance problems because of frequent inter VM packet transfers (latencies). They also say that modularized NFs requires more hardware resources to accommodate all processing elements (not sure why) compared to SFC with monolithic vNFs. So, they propose to consolidate some common processing elements on the same VM with one CPU core to improve the hardware resource consumption. Furthermore, they carefully consider what elements to consolidate and reduce the inter-VM packet transfers. The packet transfer cost is analyzed in terms of delayed bytes.

They also solve the related scaling problems when the NFs are overloaded. The propose a new push-aside scale-up procedure for performance and greedy scale-out for resource efficiency. The push-aside scale-up adjusts the processing elements between the upstream and/or downstream VMs to instead of bringing up a new VM to avoid inter-VM packet transfers and state synchronization. When there are no such elements that can be migrated or the overload cannot be alleviated, they bring a greedy scale-out mechanism which tries to place the replica on already working VM instead of bringing a new VM. If it can't find the required amount of resources on any working VMs, it brings a new VM just as the traditional.

They also propose a scheduler that ensures fair resource allocation. They try to match the processing speeds with respective packet arrival rates. This way, the element with a lower processing speed and smaller flows can get an appropriate proportion of CPU. This, they study in terms of buffer variations which is also useful in informing the scaling methods under overload situation.

Comments:
1. They could have shown actual performance overhead due to modularized NFs over monolithic vNFs.

Paper3: PRAM-Priority-aware Flow Migration Scheme in NFV Networks, SOSR'17

In this poster, the authors argue that there is no work in selecting flows for migration in NFV networks. Selecting flows that are not latency sensitive (e.g p2p download flows) for migration lets latency sensitive flows have better SLAs. Therefore, they propose a priority-aware flow selection for migration. They migrate flows with lower priority and larger size (PRAM) to minimize total migration time and not to affect the latency sensitive flows. They evaluate the system with different heuristics such as random flow selection, flow size based, priority only and PRAM method.