University of Glasgow

NETLAB
NETWORKED SYSTEMS RESEARCH LABORATORY

In-Network Placement of Security VNFs in Multi-Tenant Data Centers

Abeer Ali, Christos Anagnostopoulos, and Dimitrios P. Pezaros
School of Computing Science, University of Glasgow , UK

Agenda

- Introduction
- Background
- Literature Review
- Proposed System
- Evaluation

Introduction

- Problem
  - The placement of Security Functions in Multi-tenant Data Centers.
- Objective and Research questions
  - Identify unique characteristics of security functions as VNF.
  - Design of a placement framework that
    - Considers the unique characteristics of security functions
    - Provides customised services
    - in multi-tenant environments

Background

University of Glasgow

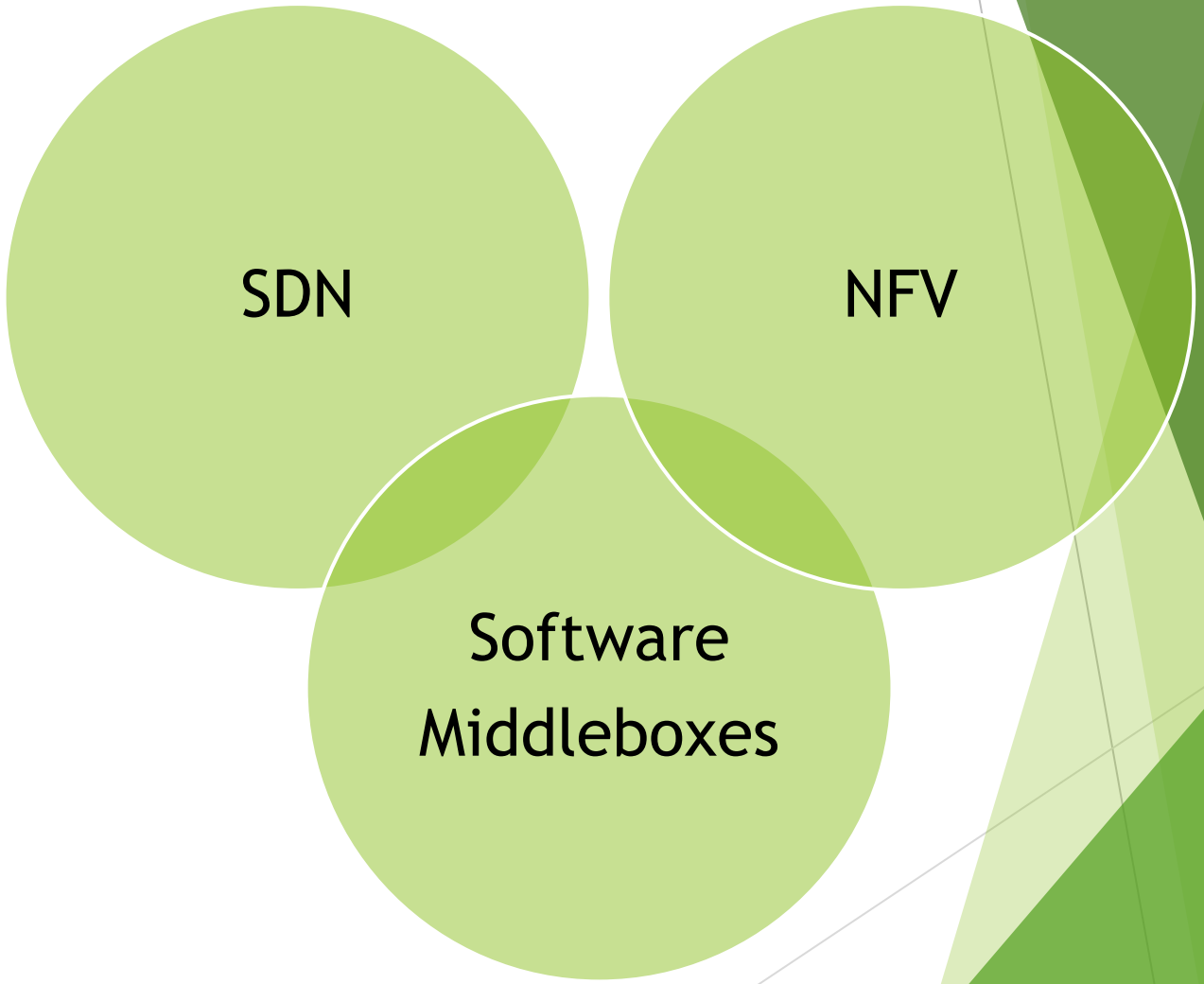
Hardware-based Middleboxes	Software-based Middleboxes
Fixed allocation	Rapid and Flexible deployment
Centralized & Monolithic systems	Scalable resources
Limited extent of functionality	Allow extension of functionality
Vendor lock-in	No Vendor lock-in
Expensive	Inexpensive compared to HW

Example of Security Services

- Amazon's AWS Multitenant virtualized infrastructures
  - (2015) Firewall web application(WAF) ,Dec 2016 AWS Shield (DDoS protection services) , Nov 2017 GuardDuty (Intelligent threat detection)
  - Alert Logic , Armor, Cisco and Barracuda

Management of Software middleboxes

- NFV
  - Software-based NFs
  - Efficient resource provisioning
  - Flexibility of Placement
- SDN
  - Centralised control
  - Programmability
  - Global view of the network



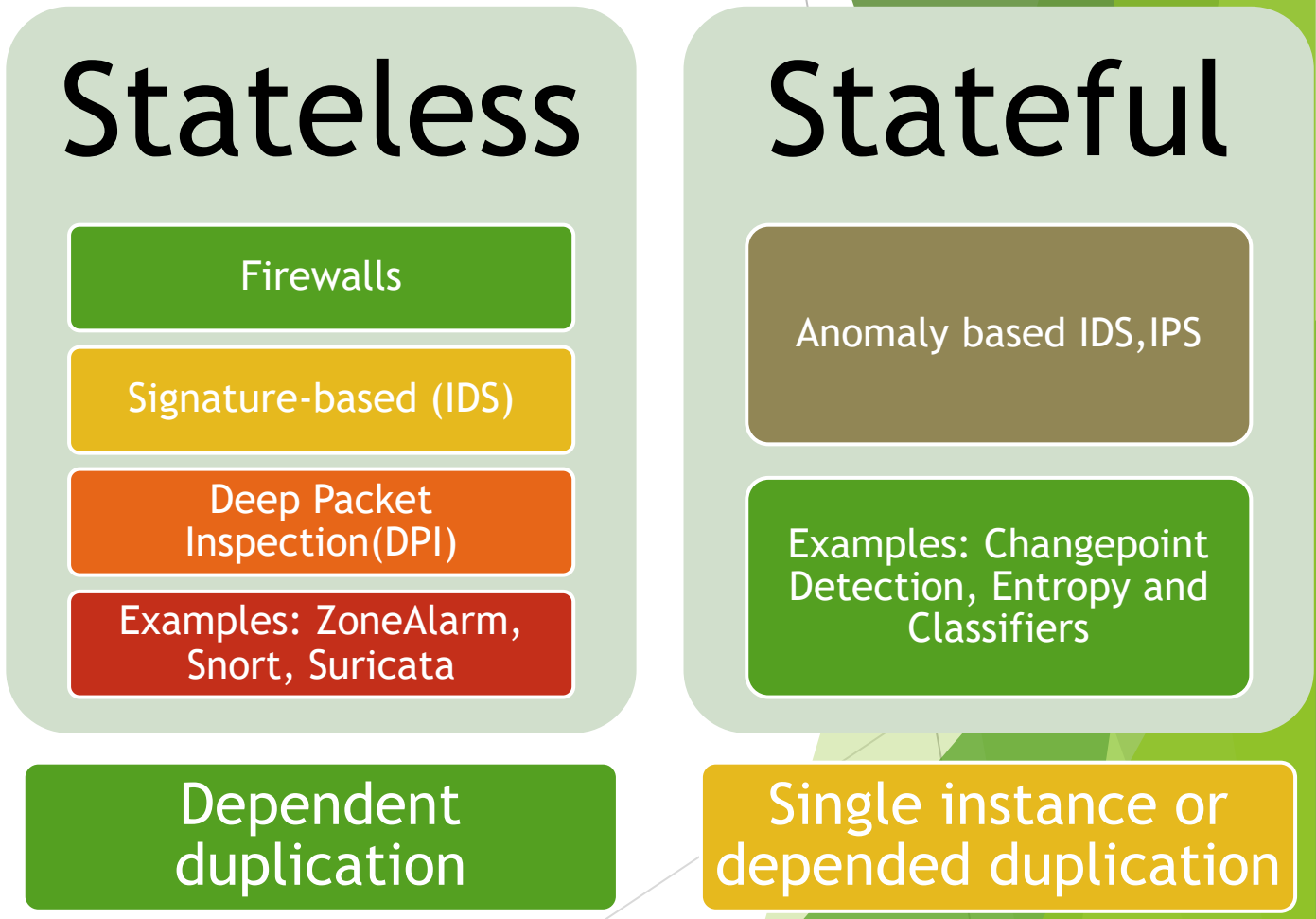
Literature Review

University of Glasgow

- Management of Softwarized middleboxes is complex and expensive
- Reviewing VNF placement and security function processing
- Issues and limitations for security function in Multitenant infrastructure
  - Traffic constraints (Stateless, Stateful)
  - Duplication of security functions (instances)
  - Shared security functions among tenants

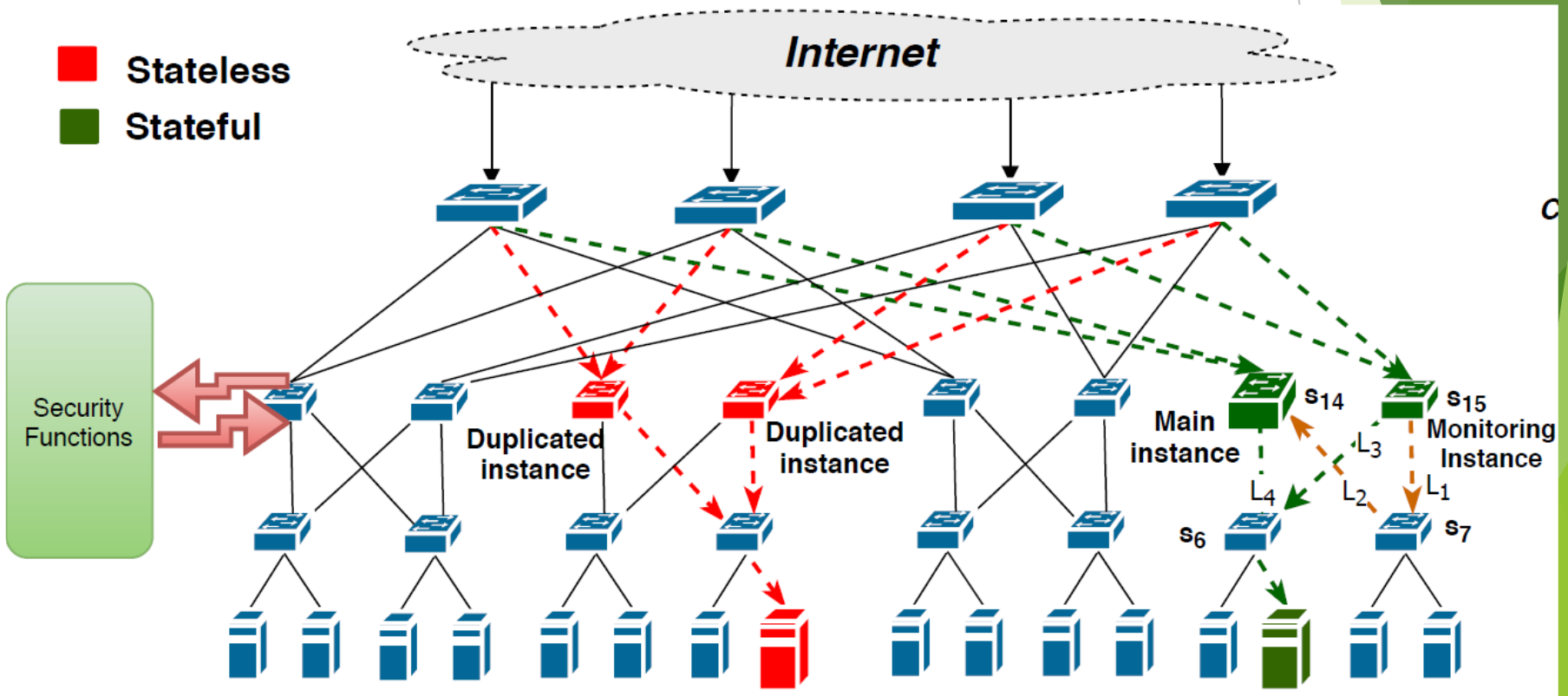
Security Functions Equivalence Classes

- Stateless modules
  - process traffic at the individual flow or packet level. Therefore, it can operate independently at different links in case of per-flow routing.
  - VNF of this class can be duplicated across network locations where tenant's traffic is being split, as long as the routing is per-flow (independent duplication)
- Stateful modules
  - process traffic to detect anomalies based on a coarser traffic granularity such as flow aggregation to build a behaviour model.
  - It will require all the flows of the monitored traffic to be rerouted to one instance of the network function of this type.
  - Or monitoring nodes could be deployed wherever traffic is distributed. they will extract the needed features from the traffic and share this information with the main node that takes the detection decision. (Dependent duplication)



Resource-Aware Security Placement Framework

- On-path deployment
  - Collocated with the network switches
- Target efficient management of resources, minimum overhead and consider ECMP
  - independent duplication (stateless class)
  - Dependent duplication (stateful class)
- Constraints
  - Traffic
  - Resources
  - Security



Fat-tree k=4 data center

Mathematical Models

$\max.\left[\sum_{\forall s\in S} s.c - \sum_{\forall r\in Q}\sum_{\forall p\in P}\sum_{\forall s\in S} \mathbf{x}_{r,p}\cdot u_{p,r,s}\right]$	Max Residual Resources RS	Objectives
$\min.\left[\sum_{\forall r\in Q}\sum_{\forall p\in P}\sum_{\forall l\in L} \mathbf{x}_{r,p}\cdot v_{p,r,l}\cdot l.w\right]$	Min Communication Overhead CO	
s.t. $\sum_{\forall r\in Q}\sum_{\forall p\in P} \mathbf{x}_{r,p}\cdot u_{p,r,s} \leq s.c \quad \forall s\in S$	Switches Capacity	Constraints
$\sum_{\forall r\in Q}\sum_{\forall p\in P} \mathbf{x}_{r,p}\cdot v_{p,r,l} \leq l.b \quad \forall l\in L$	Links Capacity	
$\mathbf{x}_{r,p}=0, \quad \forall r\in Q,\ \forall p\in P \qquad if\ w_{p,r}=0$	Location Validity	
$\sum_{\forall p\in P} \mathbf{x}_{r,p}=1, \quad \forall r\in Q$	One allocation	

Solutions

- Constraint Programming
  - Model solved using Cplex optimiser
  - CP RS+CO adding the two objective with equal weight
  - CP_two_Pass optimising the stateful requests CO then the stateless for RS
- Heuristic
  - RANDOM
  - First Fit Decreasing (FFD)
  - Best Fit Decreasing (BFD)
- Legacy one-instance strategy
  - The one instance simulates the legacy allocation of hardware middleboxes where in case of traffic distributed over multiple links, all traffic must be steered to one instance of the module.

BFD

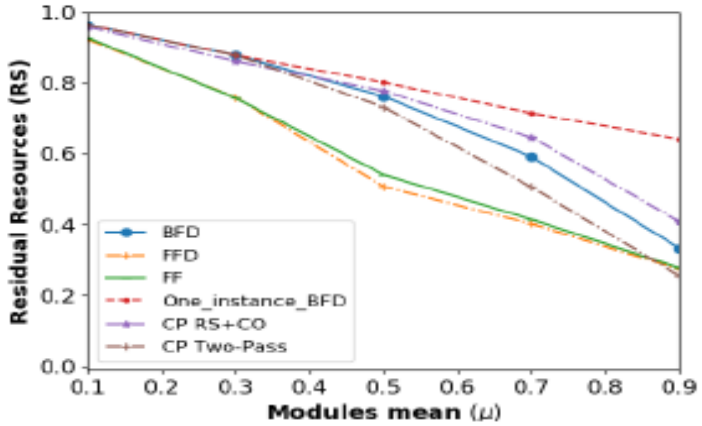
- In FFD the requests are ordered in decreasing order based on resource consumption and are allocated to first fit (module types with high resource consumption allocated first).
- In BFD, the requests are ordered the same way as in FFD, and then they are allocated to the best-fit location where the total cost is minimised.

Input: Set of requests Q, set of locations L
Output: Set of requests allocated to locations A

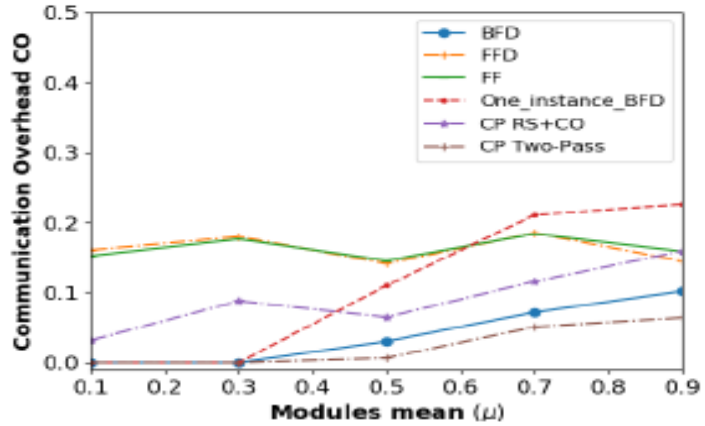
1: $A \leftarrow \emptyset$ // initialisation
2: $Q^* \leftarrow sort(Q)$ // sort request w.r.t. resources
3: **for all** $r \in Q^*$ **do**
4: **for all** $l \in L$ **do**
5: **if** $(capacity(A, r, l)$=TRUE$) \bigwedge (validation(r, l)$=TRUE$)$ **then**
6: $l^* = \arg\min_{l' \in L} total\_cost(r, l')$
7: **end if**
8: **if** $(l^* \neq 0)$ **then**
9: $A \leftarrow A \cup \{(r, l^*)\}$ // allocate request r to location l^*
10: **end if**
11: **end for**
12: **end for**
13: **return** Set of allocated requests A

BFD Algorithm

Evaluation
The effect of the modules' sizes workload



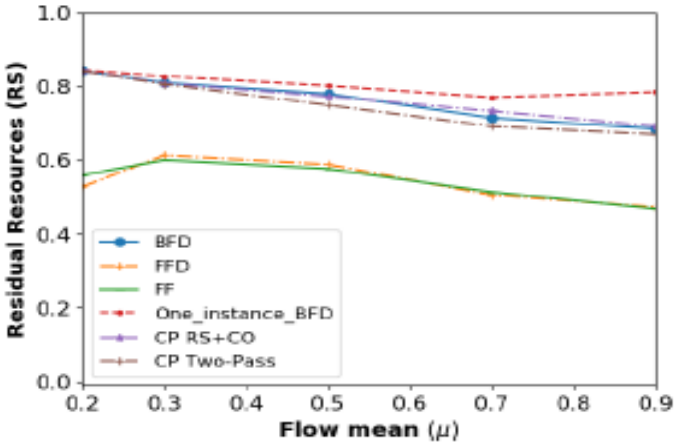
(a) Residual Resources



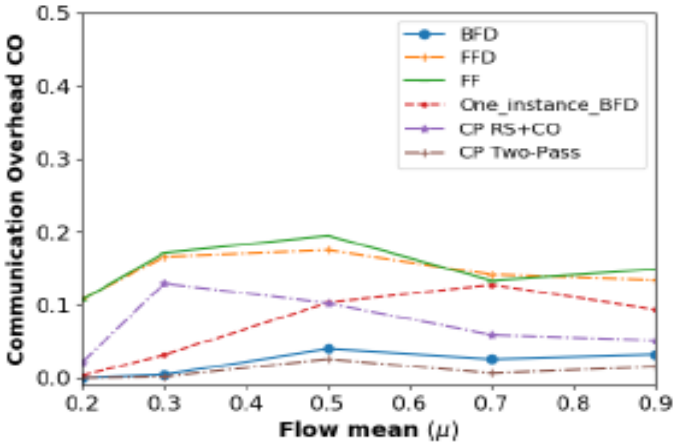
(b) Communication Overhead

- BFD offers more RS that can reach 50% of other algorithms.
- BFD offers reduced communication overhead up to 80%.
- Single-instance BFD has more spare resources than other algorithms but has more communication overhead.
- CP Two-pass exhibits the least communication overhead compared to other algorithms where is priorities CO over RS functions.
- CP RS+CO exhibits more RS than CP two-pass because it balances between the two objectives while CP two-pass gives priority to CO over RS, however, CP RS+CO only show 10% more RS than BFD.

The effect of traffic demand of tenant as workload



(a) Residual Resources



(b) Communication Overhead

- RS slightly decreases as the traffic part of functions' required resources is depending on traffic rate and results in a reduction in spare resources.
- BFD still shows less reduction than FFD and FF that reach up to more than 20% in spare resources.
- CP RS+CO, CP Two-pass and BFD still show a significantly higher RS compared to FF and FFD.
- CO shows steady results for the BFD algorithm due to being normalised to the total consumed bandwidth which also increases with the workload.
- CP two-pass has the least communication overhead compared to other types including the CP RS+CO due to prioritising CO over RS.

Conclusion

- Security functions impose unique constraints to the placement problem such as limited duplication that can cause increase in resource consumption.
- Sharing security functions in multitenant environments increases management complexity.
- Security placement framework exploits on-path allocation to minimise resource overhead while satisfying the security placement imposed constraints.
- The BFD algorithm has shown higher performance compared to other heuristic algorithms utilising computing and communication resources.
- BFD showed less RS than legacy single-instance strategy but less Communication overhead.
- BFD showed near optimal CP performance.