SCALING VIRTUAL NETWORK FUNCTIONS TO MINIMIZE OPERATIONAL COST

Sabidur Rahman

Friday Group Meeting, Netlab

UC Davis

krahman@ucdavis.edu

http://www.linkedin.com/in/kmsabidurrahman/

Happy Bengali New Year 1424!

Agenda

Define auto-scaling

Motivation

Literature review

Problem statement and dimensions

Methods and progress

Auto-scaling (1)

“Autoscaling, also spelled **auto scaling** or **auto-scaling**, is a method used in cloud computing, whereby the amount of computational resources in a server farm, typically measured in terms of the number of active servers, scales automatically based on the load on the farm.”

Amazon Web Services (AWS)
Netflix
Microsoft's Windows Azure
Google Cloud Platform
Facebook

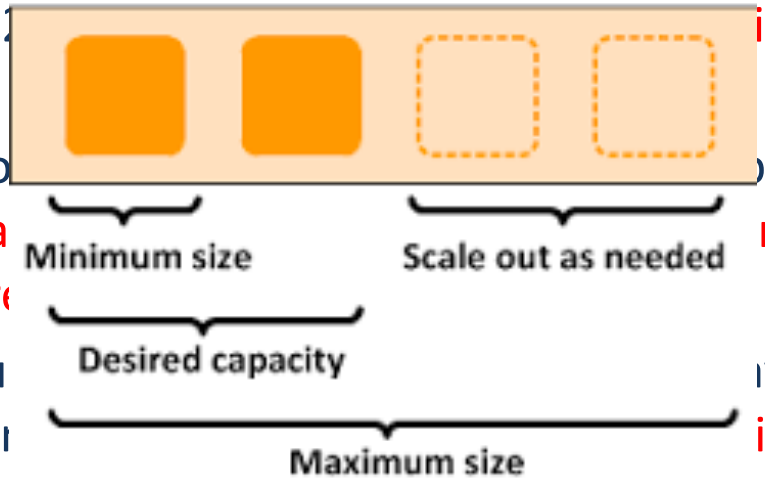
1. https://en.wikipedia.org/wiki/Autoscaling

Auto-scaling (2)

"Auto Scaling helps you maintain application availability and allows you to scale your Amazon EC2 ically according to conditions you define.

….Auto Scaling can also ber of Amazon EC2 instances during dema nce and decrease capacity during lulls to r

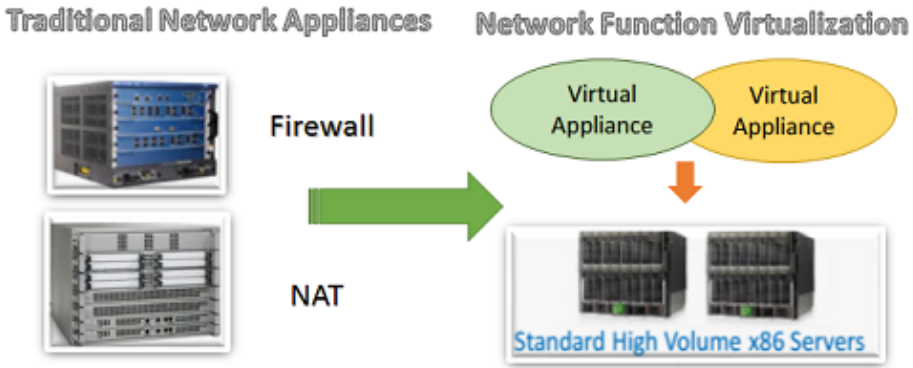
Auto Scaling is well su ve stable demand patterns or that experie ity in usage."



2. https://aws.amazon.com/autoscaling/

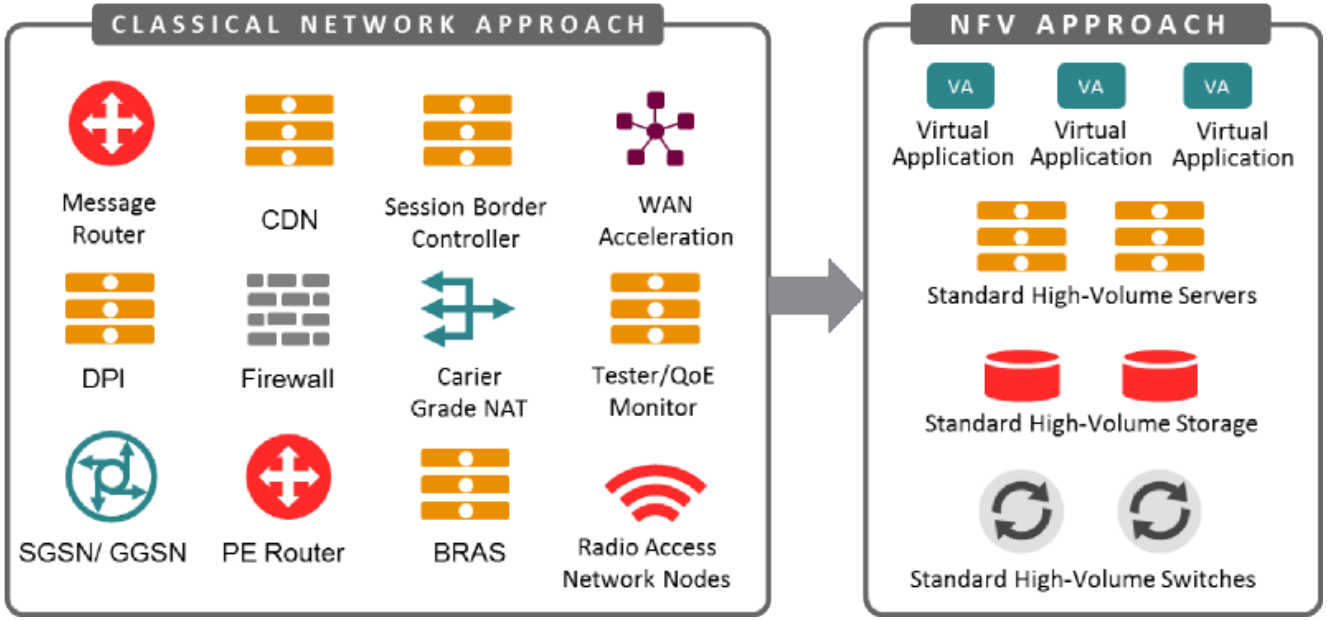
Auto-scaling of network resources

- Broadband Network Gateways (BNGs)
- Evolved Packet Core (EPC)
- Firewalls
- Deep Packet Inspection (DPI)
- Data exfiltration systems
- NATs
- Web Proxies
- Load balancers
- Content caching
- Parental control



3. Palkar S, Lan C, Han S, Jang K, Panda A, Ratnasamy S, Rizzo L, Shenker S. E2: a framework for NFV applications. InProceedings of the 25th Symposium on Operating Systems Principles 2015 Oct 4 (pp. 121-136). ACM.

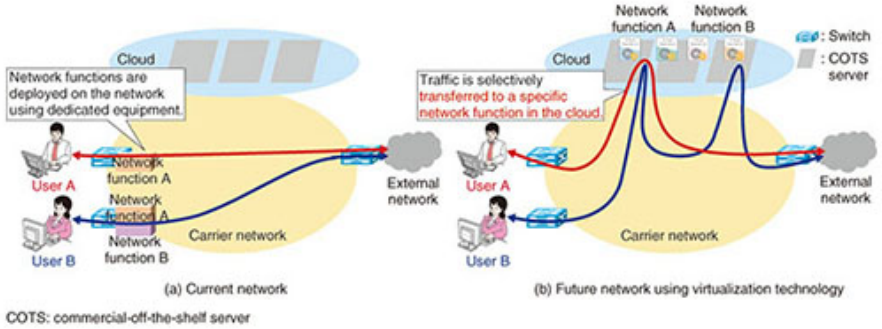
Network Function virtualization

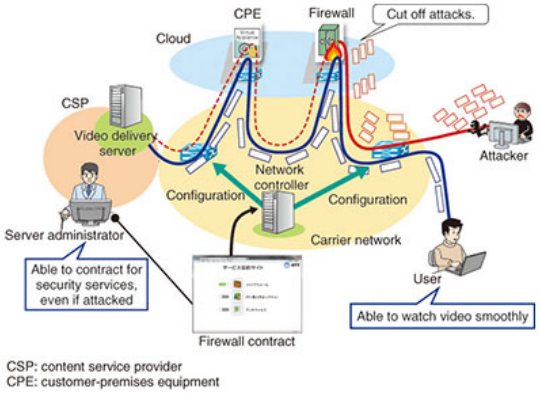


4. http://www.alepo.com/white-papers/alepo-in-the-virtualized-core-network/

5. Gupta A, Habib MF, Chowdhury P, Tornatore M, Mukherjee B. Joint virtual network function placement and routing of traffic in operator networks. UC Davis, Davis, CA, USA, Tech. Rep. 2015 Apr 20.

Service chaining

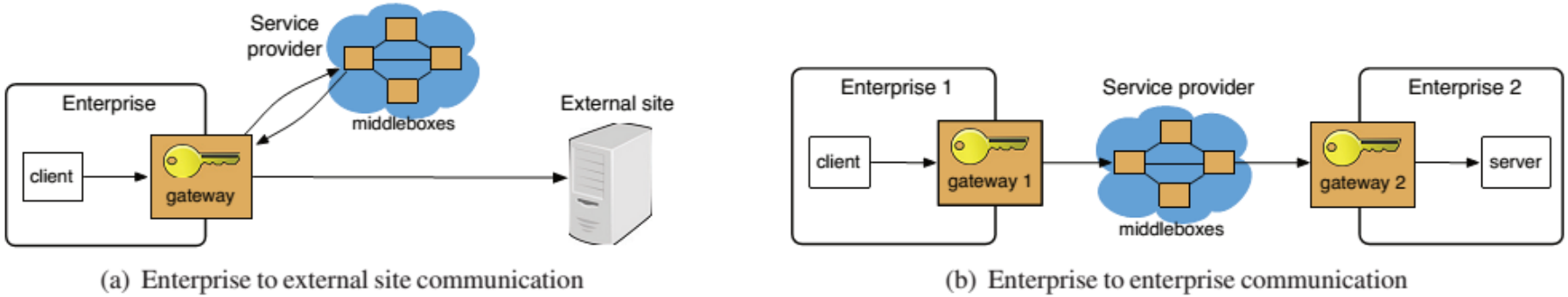




6. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201408fa2.html

7. Gupta A, Habib MF, Chowdhury P, Tornatore M, Mukherjee B. On service chaining using Virtual Network Functions in Network-enabled Cloud systems. In2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS) 2015 Dec 15 (pp. 1-3). IEEE.

Network function outsourcing



(a) Enterprise to external site communication

(b) Enterprise to enterprise communication

Figure 1: System architecture. APLOMB and NFV system setup with Embark encryption at the gateway. The arrows indicate traffic from the client to the server; the response traffic follows the reverse direction.

8. Lan C, Sherry J, Popa RA, Ratnasamy S, Liu Z. Embark: securely outsourcing middleboxes to the cloud. In13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16) 2016 Mar 16 (pp. 255-273).

9. Fayazbakhsh SK, Reiter MK, Sekar V. Verifiable network function outsourcing: requirements, challenges, and roadmap. InProceedings of the 2013 workshop on Hot topics in middleboxes and network function virtualization 2013 Dec 9 (pp. 25-30). ACM.

Motivation (1)

- Autonomous
- Better management and control
- Cost savings
- Energy efficiency (Saving the world?)

Motivation (2)

- Content Distribution Networks (CDNs) [10]: Netflix, Akamai.
- Telecom networks [11]: AT&T, Verizon.
- Data Center Networks [13]: Google, Amazon, Facebook.
- Mobile Virtual Network Operators [12]: Boost Mobile (Sprint), Cricket Wireless (AT&T), MetroPCS (T-Mobile US)
- Software-defined Data Center [14]
- Network function outsourcing

10. Mandal U, Chowdhury P, Lange C, Gladisch A, Mukherjee B. Energy-efficient networking for content distribution over telecom network infrastructure. Optical Switching and Networking. 2013 Nov 30;10(4):393-405.
11. Zhang Y, Chowdhury P, Tornatore M, Mukherjee B. Energy efficiency in telecom optical networks. IEEE Communications Surveys & Tutorials. 2010 Oct ;12(4):441-58.
12. Zarinni F, Chakraborty A, Sekar V, Das SR, Gill P. A first look at performance in mobile virtual network operators. InProceedings of the 2014 Conference on Internet Measurement Conference 2014 Nov 5 (pp. 165-172). ACM.
13. Heller B, Seetharaman S, Mahadevan P, Yiakoumis Y, Sharma P, Banerjee S, McKeown N. ElasticTree: Saving Energy in Data Center Networks. InNSDI 2010 Apr 28 (Vol. 10, pp. 249-264).
14. http://www.vmware.com/solutions/software-defined-datacenter.html?src=phd709

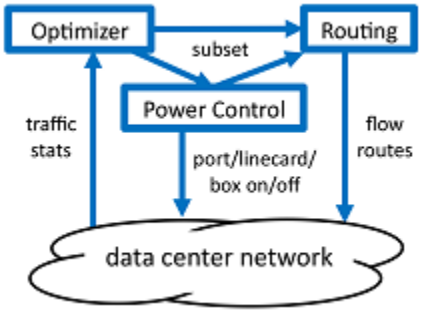
Literature review (1)

10. Mandal U, Chowdhury P, Lange C, Gladisch A, Mukherjee B. Energy-efficient networking for content distribution over telecom network infrastructure. Optical Switching and Networking. 2013 Nov 30;10(4):393-405.

- **Focus**: Content distribution over telecom network
- Energy consumption model, analysis and content-placement techniques to reduce energy cost
- Storage power consumption and transmission power consumption
- Time-varying traffic irregularities
- More content replicas during peak load and less replicas during off-peak load

Literature review (2)

13. Heller B, Seetharaman S, Mahadevan P, Yiakoumis Y, Sharma P, Banerjee S, McKeown N. ElasticTree: Saving Energy in Data Center Networks. In NSDI 2010 Apr 28 (Vol. 10, pp. 249-264).

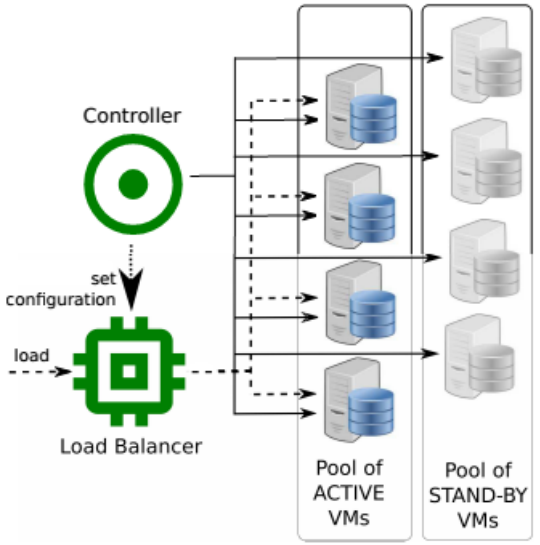


- **Focus**: Data center networks
- Scale up and down to save energy
- Dynamically adjust link and switches to satisfy changing traffic load
- Optimizer monitors traffic to choose set of elements needed to meet performance and fault tolerance goals.
- Formal model, Greedy bin-packer, topology-aware heuristic and demand prediction-based method

Literature (3)

15. Avresky DR, Di Sanzo P, Pellegrini A, Ciciani B, Forte L. Proactive Scalability and Management of Resources in Hybrid Clouds via Machine Learning. In Network Computing and Applications (NCA), 2015 IEEE 14th International Symposium on 2015 Sep 28 (pp. 114-119). IEEE.

- A proactive system scale up / scale down technique
- Machine learning models for predicting failures caused by accumulation of anomalies (Software/Hardware)
- When a VM joins (or leaves) a region, the region workload is automatically spread across local VMs



Literature review (4)

Phung-Duc T, Ren Y, Chen JC, Yu ZW. Design and Analysis of Deadline and Budget Constrained Autoscaling (DBCA) Algorithm for 5G Mobile Networks. arXiv preprint arXiv:1609.09368. 2016 Sep 29.

- VNFs can be dynamically scale-in/out to meet the performance desire
- Auto-scaling algorithm for desired characteristics with low operation cost and low latency
- Tradeoff between performance and operation cost
- NFV enabled Evolved Packet Core (EPC) is modeled as queueing model
- Legacy network equipment are considered as **reserved a block of servers**
- VNF instances are powered on and off according to the number of job requests present.

Literature review (5)

Tang P, Li F, Zhou W, Hu W, Yang L. Efficient Auto-Scaling Approach in the Telco Cloud Using Self-Learning Algorithm. In 2015 IEEE Global Communications Conference (GLOBECOM) 2015 Dec 6 (pp. 1-6). IEEE.

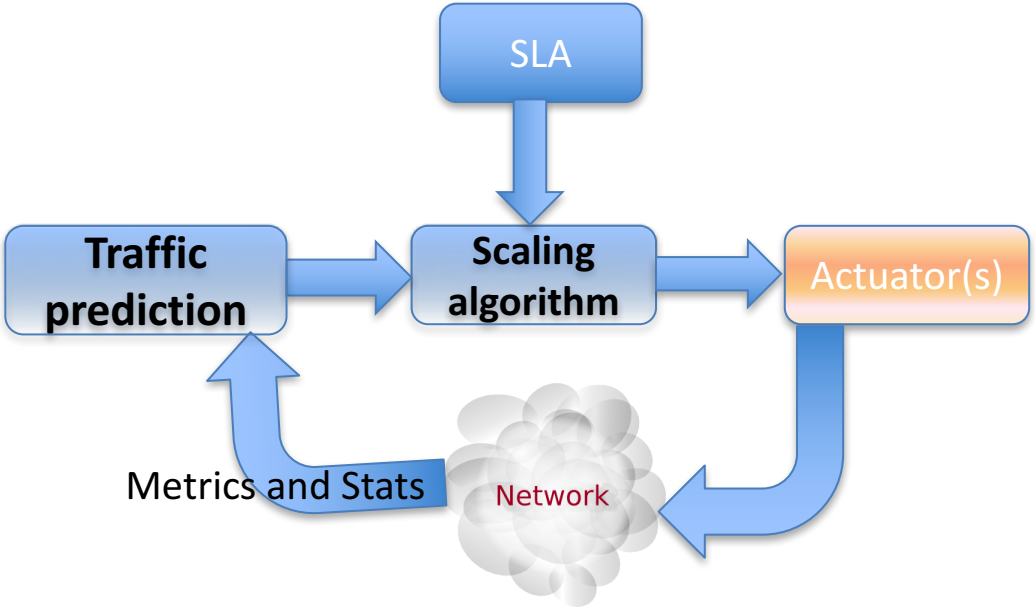
- Provision and orchestration of **physical and virtual resource** is crucial for both Quality of Service (QoS) guarantee and cost management in cloud computing environment.
- SLA-aware and Resource-efficient Self-learning Approach (SRSA) for auto-scaling policy decision
- Busy-and-idle scenario and burst-traffic scenario

Problem statement

Given: Network topology, network traffic data, SLA

Objective: Predict the future traffic and scale VNFs to minimize network operation cost (or network leasing cost).

High level design



Usecases

1) Network (or VNF) leaser (CDN, MVNO): Lower usage = lower rents
2) Network (or VNF) owner (AT&T, Time Warner): Lower usage = lower OPEx.

Topologies

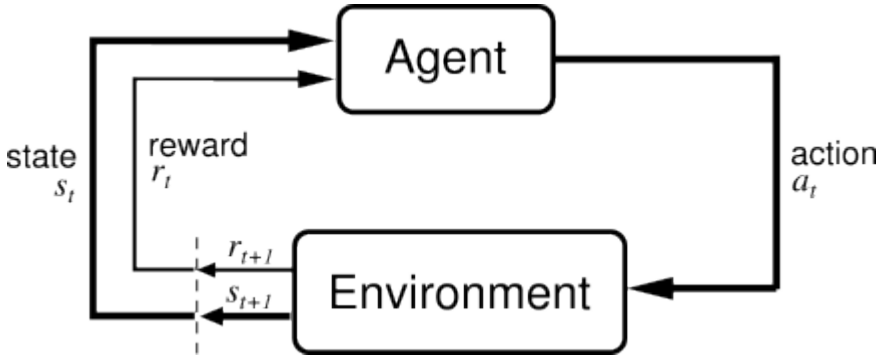
1) Data Center hosting VNFs (Fat Tree)
2) Access Network with COs (Compute node hosting VNFs) and WAN

Scaling techniques

1) Heuristic: threshold based; allocate enough VNFs to serve traffic
2) Reinforcement learning algorithm

Traffic prediction

1) Machine learning
2) Deep learning



Summary

- Machine learning can help with the prediction
- Reinforcement learning can help with the scaling decision
- Cost analysis from both network owner and leaser perspective would be interesting.

Questions?

krahman@ucdavis.edu
http://www.linkedin.com/in/kmsabidurrahman/