AUTO-SCALING OF NETWORK RESOURCES

Sabidur Rahman

Nov. 4, 2016

Friday Group Meeting, Netlab

Agenda

Define auto-scaling

Network Function Virtualization

Motivation

Literature review

Problem statement and high level design

Future directions

Auto-scaling (1)

“Autoscaling, … 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

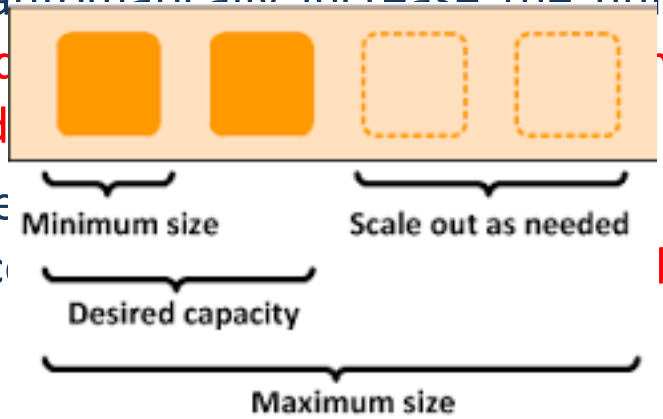
1. https://en.wikipedia.org/wiki/Autoscaling

Auto-scaling (2)

"Auto Scaling helps you maintain application availability and allows you to scale your Amazon EC2 capacity up or down automatically according to conditions you define.

….Auto Scaling can also automatically increase the number of Amazon EC2 instances during demand to maintain performance and decrease capacity during lulls to reduce costs.

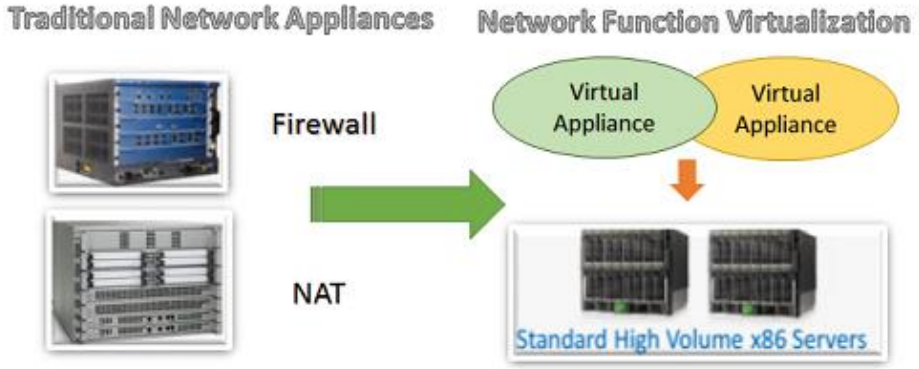
Auto Scaling is well suited to applications that have stable demand patterns or that experience hourly, daily, or weekly variability in usage."



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

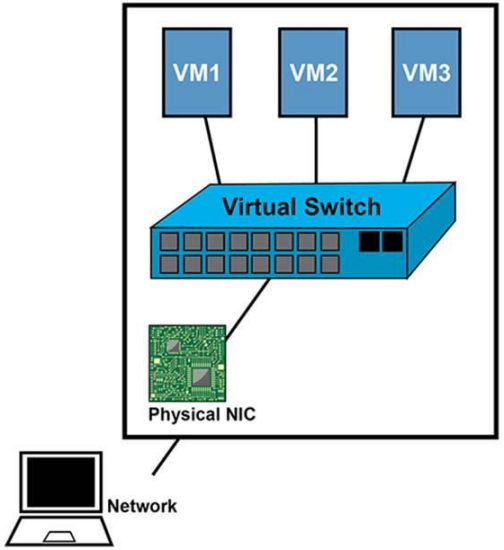
Scalable network resources

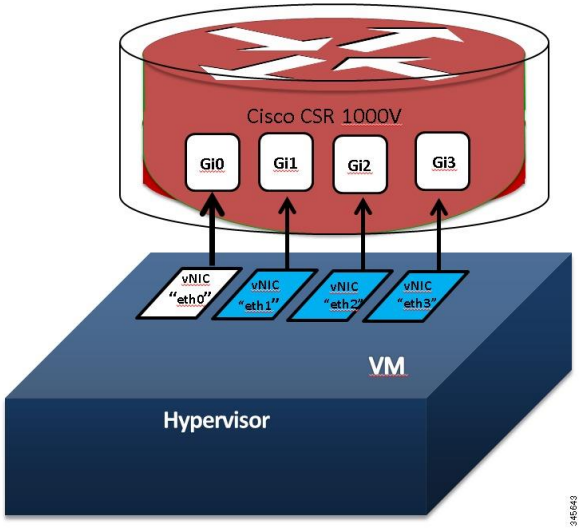
- Broadband Network Gateway (BNG)
- Evolved Packet Core (EPC)
- Firewalls
- Deep Packet Inspection (DPI)
- Data exfiltration system
- NAT
- DNS
- Web Proxies
- Load balancers
- Content caching
- Parental control



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.
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.

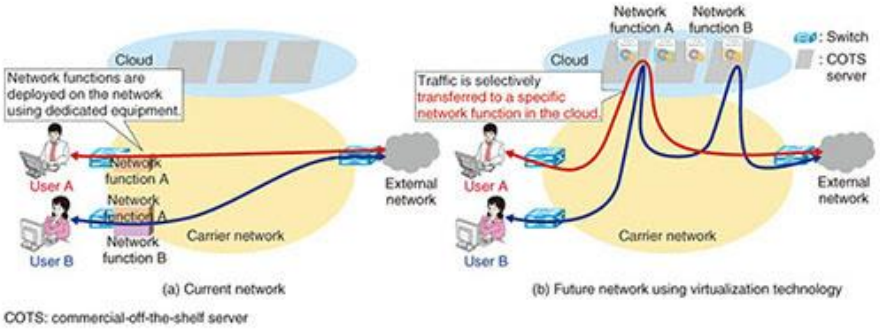
Virtual switch and routers

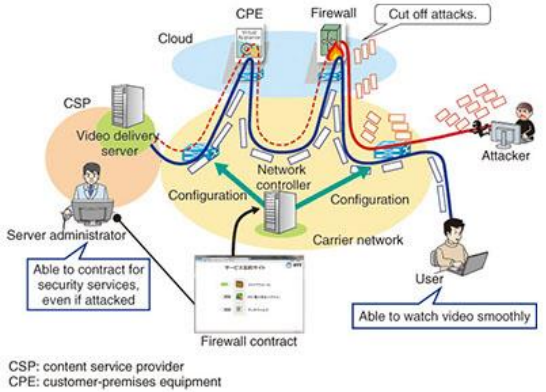




5. http://searchvmware.techtarget.com/tip/Virtual-network-switches-in-the-VMware-admins-world-Pros-and-cons

Service chaining

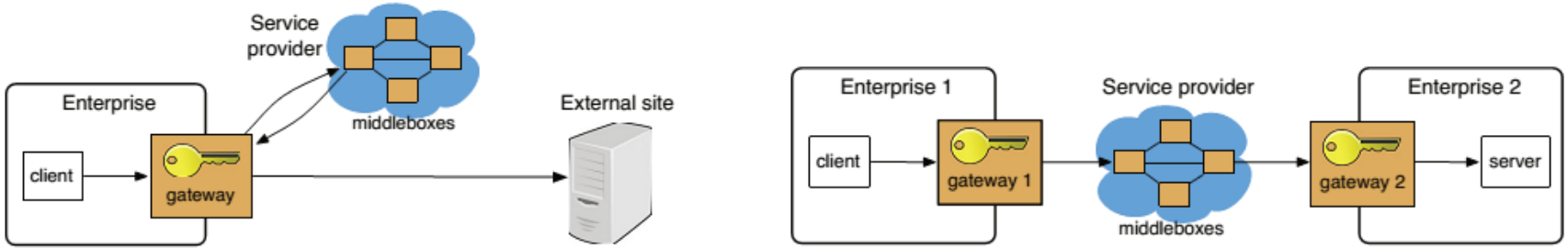




6. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201408fa2.html

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 (it's a desired feature!)
- Better management and control
- Cost savings: Operational and Capital (cheaper hardware?)
- Energy efficiency (Saving the world?)

Motivation (2)

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

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. In Proceedings 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. In NSDI 2010 Apr 28 (Vol. 10, pp. 249-264).
14. http://www.vmware.com/solutions/software-defined-datacenter.html?src=phd709

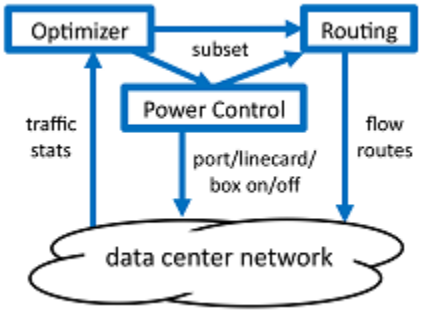
Literature review (1)

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
- More content replicas during peak load and less replicas during off-peak load
- Energy consumption model, analysis and content-placement techniques to reduce energy cost
- Storage power consumption and transmission power consumption
- Time-varying traffic irregularities

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).

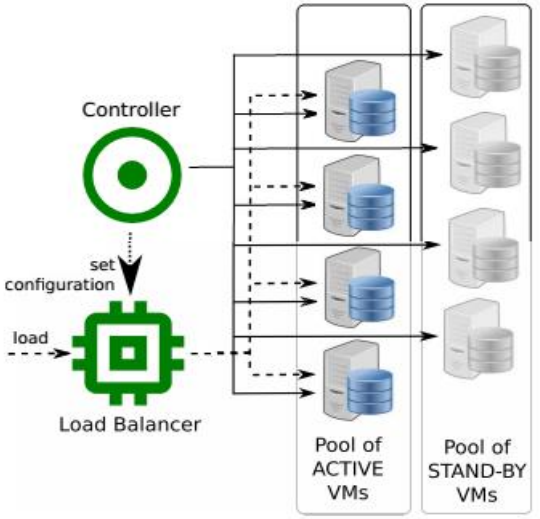


- **Focus**: Data center networks
- Scale up and down the network to save energy
- Dynamically adjust link and switches to satisfy changing traffic load
- Optimizer monitors traffic to choose set of elements needed to meet performance and fault tolerance goals.
- Formal model, topology-aware heuristic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

- Scale up / scale down technique for VMs
- 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

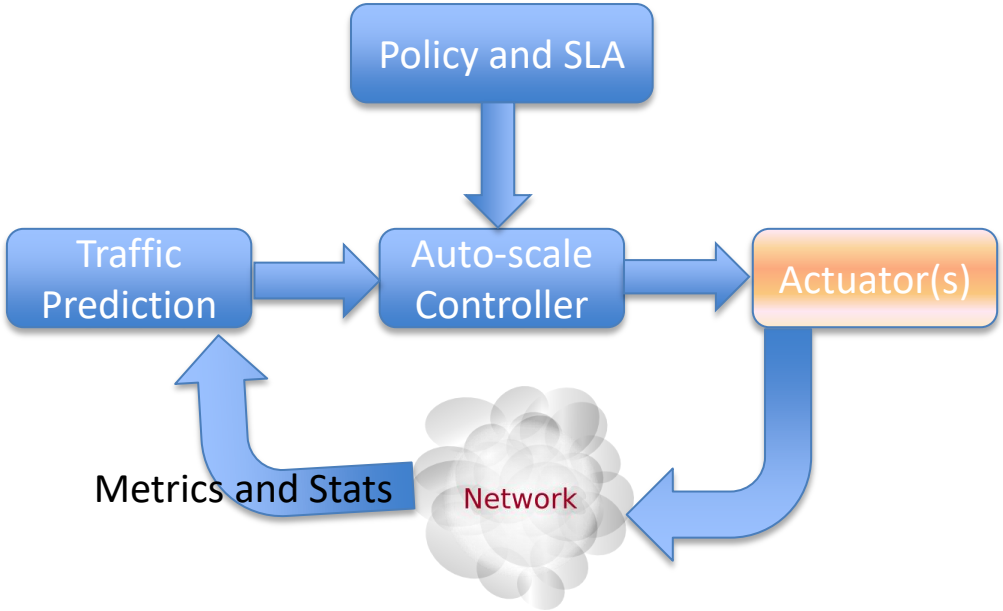


Problem statement

ScaleNet: Traffic prediction based auto-scaling of virtual network resources for flexible management at reduced cost.

- Generic: to be applied to many the usecases
- Simple: easy to understand the motivation (Ease of use? Cost? Energy?)
- Unique: different from what others have done so far

High level design



Traffic prediction

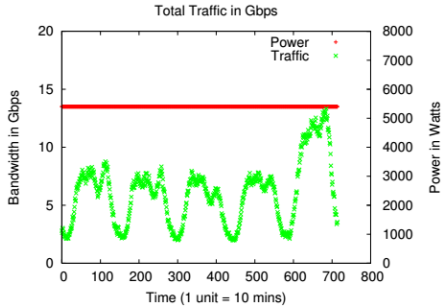


Figure 2: E-commerce website: 292 production web servers over 5 days. Traffic varies by day/weekend, power doesn't.

Auto regression [13] and Exponential Moving Average (EMA) [19]

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).
19. Jokhio F, Ashraf A, Lafond S, Porres I, Lilius J. Prediction-based dynamic resource allocation for video transcoding in cloud computing. In2013 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing 2013 Feb 27 (pp. 254-261). IEEE.

What is new?

- Network Function Virtualization and virtual switching/routing
- Traffic prediction (instead of assuming that we know traffic load)
- Power model for such networks
- Methods to scale up/down the resources
- Periodic traffic and ever increasing traffic
- Constraints from Policy (redundancy? Net neutrality?) and SLA.

More to study…

16. 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.

17. 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.

18. Mastorakis G, Mavromoustakis CX, Markakis E, Pallis E, Skourletopoulos G. Virtual Network Functions exploitation through a prototype resource management framework. In Telecommunications and Multimedia (TEMU), 2014 International Conference on 2014 Jul 28 (pp. 24-28). IEEE.


Thank You