x-cloud challenges

Jakub Krzywda
Dept. of Computing Science
Umeå University
SE-901 87 Umeå, Sweden
Email: jakub@cs.umu.se

William Tärneberg
Dept. of Electrical and Information Technology
Lund University
Ole Römers väg 3, 223 63 Lund, Sweden
Email: william.tarneberg@eit.lth.se

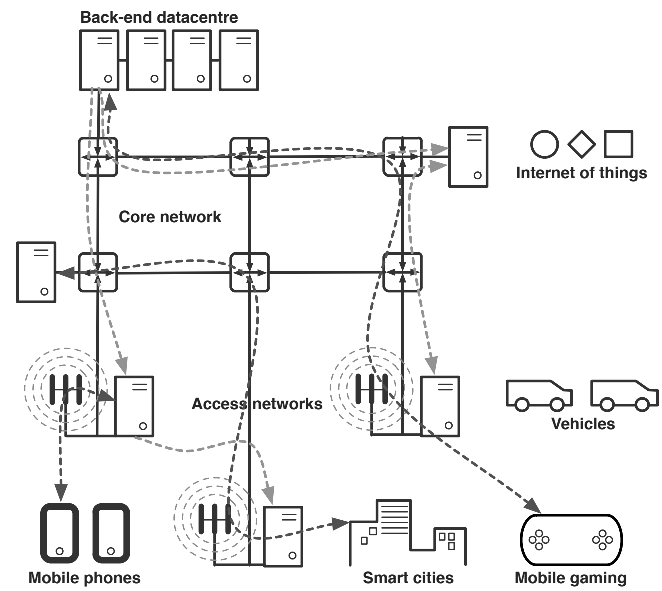

Fig. 1: x-cloud

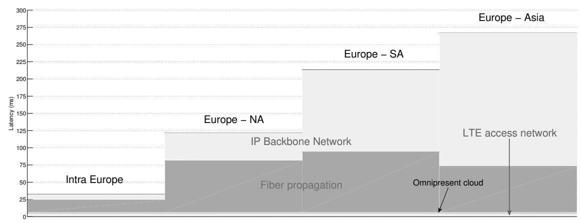

Fig. 2: IP Internet latency in western Europe [10] over LTE [13]

***Abstract*—The abstract goes here.**

I. Introduction

II. The case for the x-cloud

A. The bandwidth case for x-cloud

- Internet structure, latency, and bandwidth: [28]

B. The latency case for x-cloud

The intermediate latency between a client and a data centre is a product of propagation, modulation, and network routing and traffic shaping. Propagation is a clear physical obstacle to reducing latency, and there is very little evidence to suggest that information will propagate faster than $\frac{2}{3}$ of the speed of light, at scale, in the near future. Furthermore, the delay in the backbone network is incurred to the most part by routing. A full point to point network where the propagation speed is the only limit, is not economically viable and would dissolve the fabric of the Internet. As such, we can always expect a certain amount of network contributed latency and jitter. At best, an LTE mobile access network adds about 5 ms of latency [13]. Radio access network latency can be expected to diminish over the next few generations of mobile networks.

Moving the cloud data centres closer to the IP backbone networks eliminates some of the additive latency on one side of the connection. Doing so, not only eliminate the propagation delay, but will over time, add more complexity to peripheries of the backbone as more severs nodes make their home there.

The x-cloud remedies this latency challenge in a more sustainable way. By moving compute resources to the mobile networks, IP backbone network propagation and routing delays are eliminate without disrupting the Internet topology. The resulting distributed infrastructure is capable of delivering content and services at latencies less than 10 ms.

The x-cloud will thus enable latency-sensitive services to be migrated to the cloud, such as, gaming, financial trading, process control, and most real-time human-machine interaction process.

C. The infrastructure case for x-cloud

Distributed virtualized mobile networks will rely on centralized compute nodes for higher level link management. One node will proposedly host multiple base stations, to which they connect over a network link, much like the Ericsson Radio Dot System [**?**], but at a larger scale. The size of these virtualization resource nodes is proportional to the maximum distance they can reside from the radio nodes, given the induced propagation delay. Supposedly these virtualization resource nodes will the placed in the vicinity of the core IP network. The virtualization resource nodes can be seen as to define geographic areas whos boundaries are defined by the reach of the mobile network which it serves. Depending on the level of desired provision and load balancing flexibility, these geographic domains will overlap to varying degrees.

The virtualization resource nodes are conceivably constructed of generic x86 or ARM servers, hosting VMs or containers within which the virtualized mobile network infrastructure is executed. Given the placement of the virtualization resource node, any free or designually excess capacity can be used hist other services.

The topology is designed to optimize the use of radio resources, the geographic domains which the virtualization resource nodes constitute do not necessarily overlap or map the demographic area which x-cloud services operate.

III. Proposed research topics

A. Placement

IV. Simulation

A. Constituent models

1) Data Centre: Cloud model : [20] Web serve model : [15]

2) Radio access network: Because the mobile access network is a service access qualifier, the mechanisms of the network is relatively irrelevant to the primary research topics. The network can appropriately be modelled with a series of delays.

It is a constituent objective of our research to determine relevant X-Cloud/NGN [1] symbiotic topologies. These topolgies will be feed into the simulation model and will conceivably encompass, resource placement and dimension, cell sizes, and radio resource provisioning [21], [27], [29].

Our basic research topics will require a homogeneous, equidistant, and equirange cell topology. Although it is reasonable to assume that future networks hosting an X-Cloud will be distributed.

3) Base station: The base station can conceivably be modelled with a queue and a delay proportional to its propagation distance to its associated "C-RAN" node.

4) Core network: The essential property of the core network is bandwidth and delay. Both of which can be modelled with queues.

- Latency
  - "Point-to-Point" core network delay model [17]
  - "One-hop" core network router queue delay model [26]

5) Mobility: A smooth random walk, unobstructed, bounded, edge-aware mobility model will provide a uniformly distributed dispersion of users across the simulation domain [12]. The model is two-dimensional and provides pedestrian, bicycle, and auto mobile mobility modes. The model is uniform and does thus not take into account any socio-demographic variations, and local clusters. Nevertheless, exploring specific demographic and urban settings is beyond the scope of our basic research topics. Furthermore, in the absence of a socio-demographic and urban scenarios, an aggregate mobility mode will be deployed.

[1] Next Generation Network

6) Service: There is a multitude of appropriate service models.

- Light weight 1-tiere web service model from 1998 [11]
- Modern light weight 1-tiere web service model [22]
- YouTube workload generator [19]
- 3-tiered open-loop web service model [25]
- Web browsing behaviour : [24]
- Cloud service usage patterns : [31]

B. Simulation framework

Below are the candidate simulation tools and frameworks proposed during the third Cloud Control Workshop. [32]

1) SimJava:

2) SimPy: Python and SimPy has the ability to run powerful statistical analyses with R [8], interact with a MATLAB workspace [7], and bind NS-3 modules [6]. Nevertheless, not able to confirm weather or not you can call uncompiled MATLAB SimEvent modules.

3) CloudSim: [1]

CloudSim adaptations:

- NetworkCloudSim [18]
- CloudAnalyst [30]

4) GreenCloud: [3]

5) iCanCloud: [4]

6) MDCSim: [23]

7) SimGrid: [9]

- [16]
- [14]

8) CoolSim: [2]

9) ns-3: [5]

10)Matlab+SimEvent: (and TrueTime)

V. Papers

A. Comparison of existing simulators from the perspective of x-cloud

A survey of existing simulators with comparison of their capabilities (and limitations) to simulate x-cloud .

Simulators of:

- Data centers,
- BTS,
- Network (BTS — DC),
- Mobile network,
- Mobile devices,
- Users (mobility).

What is different in operation/simulation of x-cloud ?

B. *Limitations of current infrastructure & the setup/structure of x-cloud*

1) Limitations of current infrastructure: Simulate the current infrastructure (mobile network + remote/big data centers) and show the limits of it.

- What will happen when the number of mobile devices increases by order(s) of magnitude? The influence on a network connection between a base station and a big/remote data center.
- How many mobile devices can be handeled by the current infrastructure (depending on a latency limits)?

2) The setup/structure of x-cloud : The x-cloud consists of antennas, small (edge) data centers and big (remote) data centers. Small data centers are located close to the antennas and can host both virtualized base station software and VMs with applications. Big data centrs are located far away from users. Small data centers have smaller amount of resources than big ones (maybe also performance is lower) and running applications there is more expensive. However, latency is is much lower than in a case of big (remote) data centers.

Questions about the setup/structure of x-cloud :

- how many antennas should be associated with one small (local) data center? (probably this will be limited by the latency between an antenna and a small data center)
- how big should small (local) data centers be (#CPUs etc)?

C. *x-cloud model*

D. *Throughput and bandwidth limitations in the x-cloud*

E. *Virtual Machine placement and migration in x-cloud*

Regarding placement of Virtual Machines (VMs) in the edge data centers:

- Should a VM that serves all users (even these outside of the range of the directly connected antennas) be placed in an edge data cener or should it be rather an additional instance that serves users that are in the close proximity (duplicating a VM in a big data center)?
- When a VM should be placed/duplicated in a small data center?
- While users are moving from one antenna to another when VM should be migrated from one edge data center to another one?

F. *Other thoughts*

- Different workload patterns?
- How to perform monitoring? (System is very distributed)
- Maybe new metrics to monitor (eg. distance from antenna, velocity, direction, etc.)
- Changes in the architecture of mobile applications

REFERENCES

[1] Cloudsim. Available online at http://www.cloudbus.org/cloudsim/.

[2] Coolsim. Available online at http://www.coolsimsoftware.com/.

[3] Greencloud. Available online at https://greencloud.gforge.uni.lu/.

[4] icancloud. Available online at http://www.arcos.inf.uc3m.es/~icancloud/Home.html.

[5] Ns-3. Available online at http://www.nsnam.org/.

[6] Ns-3 python interaction. Available online at http://www.nsnam.org/docs/manual/html/python.html.

[7] pymatlab. Available online at https://pypi.python.org/pypi/pymatlab.

[8] R project. Available online at http://www.r-project.org/.

[9] Simgird. Available online at http://simgrid.gforge.inria.fr/.

[10] Monthly network summary, 05 2014.

[11] Paul Barford and Mark Crovella. Generating representative web workloads for network and server performance evaluation. *ACM SIGMETRICS Performance Evaluation Review*, 26(1):151–160, 1998.

[12] Christian Bettstetter. Smooth is better than sharp: A random mobility model for simulation of wireless networks. In *Proceedings of the 4th ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWIM '01, pages 19–27, New York, NY, USA, 2001. ACM.

[13] T Blajić, D Nogulić, and M Družijanić. Latency improvements in 3g long term evolution. *Mipro CTI, svibanj*, 2006.

[14] Laurent Bobelin, Arnaud Legrand, David A González Márquez, Pierre Navarro, Martin Quinson, Frédéric Suter, and Christophe Thiéry. Scalable multi-purpose network representation for large scale distributed system simulation. In *Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)*, pages 220–227. IEEE Computer Society, 2012.

[15] Jianhua Cao, M. Andersson, C. Nyberg, and M. Kihl. Web server performance modeling using an m/g/1/k*ps queue. In *Telecommunications, 2003. ICT 2003. 10th International Conference on*, volume 2, pages 1501–1506 vol.2, Feb 2003.

[16] H. Casanova, A. Legrand, and M. Quinson. Simgrid: A generic framework for large-scale distributed experiments. In *Computer Modeling and Simulation, 2008. UKSIM 2008. Tenth International Conference on*, pages 126–131, April 2008.

[17] Baek-Young Choi, Sue Moon, Zhi-Li Zhang, Konstantina Papagiannaki, and Christophe Diot. Analysis of point-to-point packet delay in an operational network. *Computer networks*, 51(13):3812–3827, 2007.

[18] S.K. Garg and R. Buyya. Networkcloudsim: Modelling parallel applications in cloud simulations. In *Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on*, pages 105–113, Dec 2011.

[19] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube traffic characterization: A view from the edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, pages 15–28, New York, NY, USA, 2007. ACM.

[20] H. Khazaei, J. Misic, and V.B. Misic. Performance analysis of cloud computing centers using m/g/m/m+r queuing systems. *Parallel and Distributed Systems, IEEE Transactions on*, 23(5):936–943, May 2012.

[21] Raymond Kwan, Rob Arnott, Robert Paterson, Riccardo Trivisonno, and Mitsuhiro Kubota. On mobility load balancing for lte systems. In *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pages 1–5. IEEE, 2010.

[22] Jeongeun Julie Lee and Maruti Gupta. A new traffic model for current user web browsing behavior. *Intel Corporation*, 2007.

[23] Seung-Hwan Lim, B. Sharma, Gunwoo Nam, Eun Kyoung Kim, and C.R. Das. Mdcsim: A multi-tier data center simulation, platform. In *Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on*, pages 1–9, Aug 2009.

[24] Chao Liu, Ryen W White, and Susan Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 379–386. ACM, 2010.

[25] Xue Liu, J. Heo, and Lui Sha. Modeling 3-tiered web applications. In *Modeling, Analysis, and Simulation of Computer and Telecommu-*

nication Systems, 2005. 13th IEEE International Symposium on, pages 307–310, Sept 2005.

[26] Konstantina Papagiannaki, Sue Moon, Chuck Fraleigh, Patrick Thiran, and Christophe Diot. Measurement and analysis of single-hop delay on an ip backbone network. *Selected Areas in Communications, IEEE Journal on*, 21(6):908–921, 2003.

[27] Andras Racz, Andras Temesvary, and Norbert Reider. Handover performance in 3gpp long term evolution (lte) systems. In *Mobile and Wireless Communications Summit, 2007. 16th IST*, pages 1–5. IEEE, 2007.

[28] Venugopalan Ramasubramanian, Dahlia Malkhi, Fabian Kuhn, Mahesh Balakrishnan, Archit Gupta, and Aditya Akella. On the treeness of internet latency and bandwidth. *SIGMETRICS Perform. Eval. Rev.*, 37(1):61–72, June 2009.

[29] J Salo, M Nur-Alam, and K Chang. Practical introduction to lte radio planning. *A white paper on basics of radio planning for 3GPP LTE in interference limited and coverage limited scenarios, European Communications Engineering (ECE) Ltd, Espoo, Finland*, 2010.

[30] Bhathiya Wickremasinghe, Rodrigo N Calheiros, and Rajkumar Buyya. Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications. In *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*, pages 446–452. IEEE, 2010.

[31] Gansen Zhao, Jiale Liu, Yong Tang, Wei Sun, Feng Zhang, Xiaoping Ye, and Na Tang. Cloud computing: A statistics aspect of users. In *Cloud Computing*, pages 347–358. Springer, 2009.

[32] Wei Zhao, Yong Peng, Feng Xie, and Zhonghua Dai. Modeling and simulation of cloud computing: A review. In *Cloud Computing Congress (APCloudCC), 2012 IEEE Asia Pacific*, pages 20–24. IEEE, 2012.