

micro data server using raspberry pi

Research Proposal

P.j.e. Velthuis
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
p.j.e.velthuis@student.utwente.nl

ABSTRACT

This paper describes the performance of raspberry pi load balancing service using video streaming.

Keywords

Cloud computing, Raspberry Pi, micro data server, video streaming, load balancing

1. INTRODUCTION

Today cloud computing becomes more important. The amount of cloud computing services is increasing fast [3]. Despite the attention from the research community, research and development of Cloud Computing services is still in its childhood [34]. Cloud computing is a trend IT that people move computing and data away from desktop and portable PCs into large data centers [10]. People will in the future access internet services over lightweight portable devices. In this research the Raspberry Pi cloud services will be investigated. A reason for this is that the density of servers has increased a lot in the past [20]. There are new technologies such as for example the ARM processor. Many companies want to explore the possibilities of for example the Raspberry Pi and his ARM Processor. Some data centers even offer some cloud computing using the Raspberry Pi. A reason for this is the small location that the Raspberry Pi needs, and the low power usage of a Raspberry Pi [31, 27]. A Raspberry Pi has a power usage between the 3-5 Watt. A normal server has a power usage between the 75 and 250 watt [28, 4]. This could mean that it is better to use a Raspberry Pi for specific small tasks that do not demand a whole server. During this research there will be an investigation on the performance of the Raspberry pi as a server. A Raspberry Pi can be the small data center for the future [34]. The Raspberry is a rather cheap device for 35 euros. This makes it cheaper to do research in compared to a normal server. This is also the main reason why the Raspberry pi was created. The current versions of the Raspberry Pi cannot perform the task of a large scale x86 server [34]. Building a cloud like this can be a cost effective scale model [34]. It's an ideal testbed for testing distributed software. During this research we will investigate the performance of a raspberry pi as a video streamer to other computers. The reason I

want to investigate it with video streaming is because this demands a lot of bandwidth from the network and could be one of the first things that should be improved.

Today a lot of people watch their movies online using video streaming. The Netherlands have for example one million subscribers for Netflix [9]. Netflix is a video streaming service that makes HD movies watching possible. For this Netflix makes use of a content distribution network (CDN). On the internet this is also known as a on demand service [2]. Netflix makes use of MPEG-DASH a protocol that makes streaming over http possible [22]. The problem is that Netflix is responsible for 29.7% of the peak downstream traffic in US [2]. In the Netherlands more and more users are making use of video streaming as Youtube and Netflix. In this research there will be an investigation on the possibility to make small micro data centers with raspberry pi's. This research will try to find out how this cloud works. There will be taken a look in for example the load balancing of video streaming. In this way it can be possible to see improvements that can be made in the load balancing so that the servers can be used more efficient. This is also build to learn more from load balancing and a computer cluster. This can be really useful, because nowadays it's still done on relatively expensive servers.

In this paper we will try to measure the performance of different load balancing techniques using the Raspberry Pi. For this the main research question will be:

What is the performance of a raspberry pi load balancing techniques using video streaming?

In order to come to a good answer we need to research several things. Therefore several sub-questions will be researched. These sub-questions will altogether provide an answer to the main research question.

1. What is small scale cloud computing?
2. What is cloud video streaming?
3. What kind of different load balancing techniques are there?
4. Is it reachable to build small personal Raspberry Pi clouds for services?
5. Is it possible to improve load balancing using small scale Raspberry Pi's?

This research will be part literature study and a part of it will be building a small Raspberry Pi cloud to analyze balancing techniques on the Raspberry Pi. For this research cloud video streaming is one of the specific aspects that will be investigated. The load balancing techniques will be

optimized for the video streaming. The first three research questions will describe the state of the art. The last two sub-questions will go into the technical research that will investigate the performance and possibility of Raspberry Pi's for cloud video streaming. The last one will investigate if there are improvements in load balancing possible. The five sub-questions are being looked at in the next five sections. Followed by a final section for conclusions and identified areas for future work.

2. SMALL SCALE COMPUTING

Today there are many kinds of cloud computing. The amount of cloud computing services is increasing fast [3]. It's now even possible to build your own small cloud using Raspberry Pi's [6, 1]. In this research there will be small scale cloud computing with the Raspberry Pi. The main motivation behind this is that it is ways to expensive to simulate large scale cloud computing. Another reason is the small location that is needed used with small scale cloud computing for Raspberry Pi's [27]. The Raspberry Pi is really small compared to a normal server [7]. Because of this it could be better in executing small tasks. Small scale cloud computing is cloud computing with smaller computation amounts than normal [7]. More motivation is that it does not use a lot cooling and power [34]. Most clients of cloud computing buy the services instead of the hardware and having to set up the servers themselves. The Amazon Elastic Compute Cloud (Amazon EC2) is one of the most widely used infrastructure platforms that these clients use [15]. This elasticity of computing resources is very important [23].

2.1 What is it?

To understand what small scale cloud computing really is we will look more at the specific cloud service called Infrastructure as a Service. To this specific cloud service belongs the online storage. Today a lot of companies have their own virtual machines running in large data centers [5]. These instances can most often be scaled dynamically to the customer it's need. Using these instances makes it easier management and varying workloads better possible. The Raspberry Pi is a lot cheaper and has less power consumption, because of this it can be interesting to host virtual machines on the Raspberry Pi. The reason we call it small scale is because it is using small Raspberry's instead of large servers. Some providers offer a special hybrid cloud. You can then connect a private cloud to a public cloud [30].

2.2 Storage

In the cloud there are several storage systems. The cloud storage has become an integral part of our modern and mobile lives. Nowadays people own more often multiple devices. Cloud storage made it possible that the files we need are available where and when we want them. There are several ways to do this storage. Dropbox for example stores your information on a local folder on your device, after it has done that it will sync with an online version [11, 12]. In this way your other devices or the friends you shared the information with will have that information. In this way offline editing is also possible and this makes dropbox really powerful.

Today there are a lot of problems with security for cloud storage and that's why some people like to do it by themselves. For this there is a cloud service called ownCloud [26]. This service gives you the possibility to host the data where you want. You can even build your own cloud storage system by using ownCloud open source. The security

aspect is important, because this can lead to loss or leak of data. Sometimes this can even have legal consequences [15]. For all important software used for cloud storage this means that they should be stored server side [15].

Another problem that is very important is the reachability of the data centers that are used by the cloud. Netflix responsible for 29.7% of the peak downstream traffic in US [2].

2.3 Power usage

Power usage efficiency is nowadays becoming a more important aspect in cloud computing. This is one of the main problems in cloud computing. Data centers require a lot of electrical power. Therefore we perhaps need to more Green Cloud Computing solutions that reduce the electrical power consumption and reduce the environmental impact of data centers [4]. Besides the power the data center consumes it can be important to look at the traffic pattern. It can make a huge different in sending data in small bursts or in one large burst [23]. With the increase of lightweight portable devices the energy efficiency becomes more important. This means that a trade-off has to be made between local processing and computation offloading [23]. This is computation offloading makes cloud computing more and more important.

3. VIDEO STREAMING

There are a lot of cloud services nowadays. These include a webserver, a DNS server, a chat server, a mail server, a vpn server or a monitoring server or even a video streaming service.

Today a lot of people watch their movies online using video streaming and this amount of people is increasing. The netherlands have for example one million subscribers for Netflix [9]. One thing that these users want is the on demand video streaming. They don't want to store the data itself and they want to have a wide choice of different videos. Netflix is a video streaming service that makes HD movies watching possible. For this Netflix makes use of a content distribution network (CDN). On the internet this is also known as a on demand service. They make use of amazon its AWS, Amazon simpleDB, S3 and cassandra for file storage [2]. Netflix makes use of MPEG-DASH a protocol that makes streaming over http possible. Currently Netflix has 30% of the downstream in the United States [18]. To make this video streaming possible Netflix does some things in the cloud. These are content ingestion, content processing and uploading different versions to the CDN (Content Distribution Networks) [2]. Content ingestion means that Netflix receives the studio master version of the movies and uploads these to the cloud. In the cloud there are then created many different formats for each movie. Once all these versions have been created they are distributed over the multiple CDNS.

The video streaming happens with TCP. TCP sends frames to the client and in these frames is the video stored. Of course if the available TCP send rate is large enough then the video will play without any delays. But if the TCP send rate is less then needed then the video play will alternate between periods of continuous playout and periods of freezing [18]. Its very important that video freezing or audio playing faster then the video does not happen. This is why multimedia applications have high performance requirements. For this there are models to predict the performance and techniques to find a minimal cost bandwidth allocation that will result in all user requirements being met.

3.1 HTTP streaming

HTTP streaming is one way to stream a video. In HTTP streaming the HTTP GET request specifies the specific range of bytes the client wants to retrieve from the desired video [18]. In this way the client can correctly receive the video. If the client receives too much then its buffer for video playing can get full. To prevent throwing data away the server will then stop sending. To prevent throwing data away the client will have a not too large buffer.

There is also a new kind of HTTP streaming called DASH (Dynamic Adaptive Streaming over HTTP) [18]. In DASH the video is encoded into several different versions. If the bandwidth is high, then the client selects chunks from a high-rate version and the way around [18]. The possibility to switch between bit rate versions is in video streaming very important. To make a smooth transition between these versions possible there are intermediate versions so that the client will see less suddenly the change in video quality. In HTTP streaming the server stores the video with a different URL.

To investigate video streaming there will be an experimentation cloud with Raspberry Pi's to see if these small mini computers will work as a cloud. Resources I could use for cloud video streaming: [17, 19, 33, 29]

4. LOAD BALANCING

In distributed servers it is very important that good load balancing takes place. There are several choices that have to be made by choosing a load balancer for a service. These choices can for example be performance, reliability and features. There are several load balancing techniques out there, for example priority queuing [18]. This means that packets in the output queue are prioritized. This is typically done in a first in first out way. Another type of queuing is the round robin. Round robin puts packets into different classes. It will first look for a packet in a given class and if there is none in that class it will look for the next packet [18]. A more abstract form of round robin is weighted fair queuing. The difference between these is that each class may receive a differential amount of service in any interval of time [18].

4.1 Video streaming and load balancing

For video streaming several load balancing techniques can be used. There can be a non-uniform demand for different movies. One solution to overcome this problem would be to replicate the most popular movies. One obvious problem is then that it is expensive in terms of storage space required. To solve this you can use dynamic replication to solve the load demand [8]. By using this you move portions of a movie to less used storage devices. The allocation of these movies only happen when there is a too high increase on one specific storage device. In this way load imbalances can be prevented [8].

Load balancers are generally grouped in two layers. Layer 4 (FTP, IP, UDP, TCP) and layer 7 the application layer. Layer 7 load balancers distribute request based upon data in the application layer, this can for example be HTTP. A lot of video streaming services use http streaming [2].

[35, 24]

5. RESEARCH METHODS

This research will investigate a cloud computer consisting of Raspberry Pi's. The research method for this would be the Design Science research method. This is a method to solve field problems. This research will make use of

the design Science method proposed by Hevner [14]. This design method has three variants.

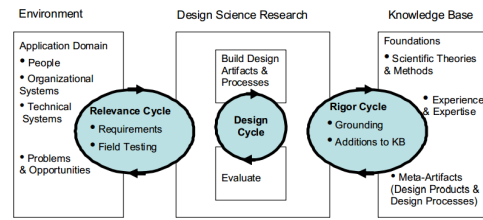


Figure 1: Design Science

The **relevance cycle** is important, because you first have to identify the problems and opportunities. The **design cycle** is where you actually build the design. The **rigor cycle** matters because it can give you two types of additional knowledge. It provides past knowledge to ensure its invention and it's there to guarantee that the new product is a research contribution.

6. RESEARCH APPROACH

For the experiment we need a Raspberry Pi with a static IP address and we need a SD card for the OS. For the experiment we need a Raspberry Pi B that can make use of the Ethernet. This is because a Raspberry Pi can be used as a webserver and this is needed to have this in order to make video streaming possible. This webserver will make video streaming possible. It will have a VPN to make it possible to put video's on the Raspberry pi [25]. So I will first make this video streaming webserver. After this we will make a cluster with one Raspberry Pi as a load balancer and the other two as a video stream pi. If I have made such a cluster I can start doing some tests on it. For example some test in streaming in different quality with load balancing. The setup of the Raspberry Pi can be seen in the figure below 2.

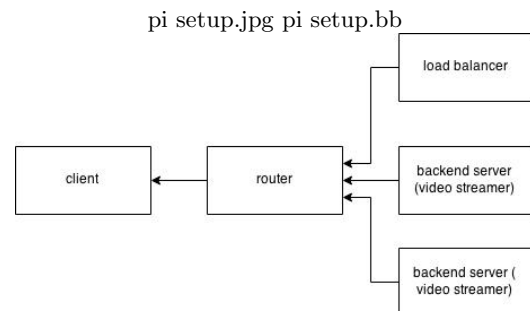


Figure 2: Raspberry Setup

For the load balancing we will make use of Nginx for this we will use the following sources: [32, 16]

7. RASPBERRY PI

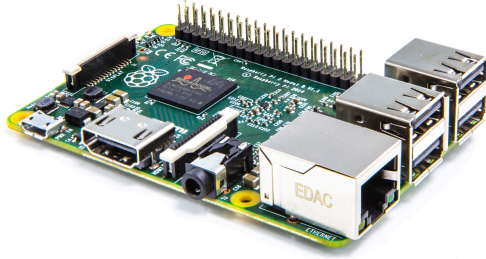


Figure 3: Raspberry Pi 2 model B

Ethernet	100 MBps
USB	4 x USB 2.0
Video out	HDMI 1.4
Audio	2 x analog
CPU	900MHz quad-core ARM Cortex-A7
card slot	Micro SD

Table 1: Specificaties

The Raspberry Pi only consumes 3 watts, because of this it doesn't need active cooling. With this subquestion I want to find out if it is possible to let users have small Raspberry Pi servers I also want to know what the performance of this small Raspberry Pi cloud is. Resources: [27, 32, 16]

For this research I want to investigate if it is possible to do more research in the load balancing techniques that can be used in a raspberry pi. A example of this would be the round robin queuing [32, 16].

7.1 Raspberry Pi cloud projects

there have been several cloud projects with the raspberry Pi.

One of these has been the supercomputer build by Southampton [7]. Here they build it with 64 raspberry Pi's . They used a Message Passing interface to communicate between the raspberry pi's. The research was done in order to see what the performance of a low-power high performance cluster is.

The project of the university of Glasgow [34]. This data center has 56 Raspberry Pi's. It is build for research and education purposes for a cloud data center. They have a hadoop running on their servers.

There was another project called the Beowulf cluster [21]. This project was created for a phd assignment. This cluster is build for collaboratively processing sensor data. A Beowulf cluster is simply a collection of identical, (typically) commodity computer hardware based systems [21]. The big advantage of the Raspberry Pi in this project was that it is cheap and it does not need a cluster administrator to watch over everything you do.

8. STATE OF THE ART

People want to have everything has to be accessible through the network around the clock [36]. One thing that people want on demand are there film series. Currently Netflix has 30% of the downstream in the United States. This downstream is huge and that means that there is optimization in this branch possible. To make the video on demand service of Netflix possible they use servers from Amazon.

In this paper there will be research done on making a load balancing service for the Raspberry Pi. This service will

consist of several raspberry Pi's. These Raspberry Pi's make together a small cloud computer. The Raspberry Pi is made for research and education purposes [13]. The Raspberry Pi is a cheap device costing around 35 euro. It has a power consumption of only three watt and has quite good processor. For this reason it might be very useful for small scale cloud computing. The raspberry Pi is a small device and is excellent for a lot of small devices in a relatively small place. The Raspberry Pi doesn't use cooling and it can be used for very rapid elasticity and on-demand self-service. Because of it's cooling features and low power consumption it might be a good alternative for the nowadays high power consuming data centers. The raspberry pi is really small so it is a lot easier to place extra pi in a data center compared to a normal server. The Raspberry Pi has one drawback and that is that the processing power is relatively low. By using load balancing there will be taken a look at what load balancing techniques are successful for the Raspberry pi.

In this research we will make a small raspberry cloud to do research on video streaming using HTTP. This will be done using the design Science method proposed by Hevner[14]. After we have done this for the Raspberry Pi there will be looked at streaming using Amazon. In this way we can see how the results are compared to streaming via Amazon. This research will try to find out if the raspberry has enough processing power to stream to multiple web clients. The cloud performance can now be better researched, because everything is at one location and the information about what processes are running on the pi are well defined. In this way we can take a better look at algorithms used in video streaming. This is mostly because there is not a lot of overhead.

9. CONCLUSION AND FUTURE WORK

10. ACKNOWLEDGMENTS

This paper would not have been possible without the help of Björn Postema and Nick Shot.

11. REFERENCES

- [1] P. Abrahamsson, S. Helmer, N. Phaphoom, L. Nicolodi, N. Preda, L. Miori, M. Angriman, J. Rikkila, X. Wang, K. Hamily, et al. Affordable and energy-efficient cloud computing clusters: The bolzano raspberry pi cloud cluster experiment. In *Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on*, volume 2, pages 170–175. IEEE, 2013.
- [2] V. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang. Unreeling netflix: Understanding and improving multi-cdn movie delivery. In *INFOCOM, 2012 Proceedings IEEE*, pages 1620–1628, March 2012.
- [3] M. Armbrust, O. Fox, R. Griffith, A. D. Joseph, Y. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, et al. M.: Above the clouds: a berkeley view of cloud computing. *Communications of the ACM*, 2009.
- [4] A. Beloglazov, J. Abawajy, and R. Buyya. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future generation computer systems*, 28(5):755–768, 2012.
- [5] A. Beloglazov and R. Buyya. Energy efficient allocation of virtual machines in cloud data centers.

- In *Cluster, Cloud and Grid Computing (CCGrid)*, 2010 10th IEEE/ACM International Conference on, pages 577–578. IEEE, 2010.
- [6] P. S. Cox. Steps to make raspberry pi supercomputer. http://www.southampton.ac.uk/~sjc/raspberrypi/pi_supercomputer_southampton_web.pdf, 2012. Last accessed February 20, 2015.
 - [7] S. J. Cox, J. T. Cox, R. P. Boardman, S. J. Johnston, M. Scott, and N. S. O'Brien. Iridis-pi: a low-cost, compact demonstration cluster. *Cluster Computing*, 17(2):349–358, 2014.
 - [8] A. Dan and D. Sitaram. Load balancing in video-on-demand servers by allocating buffer to streams with successively larger buffer requirements until the buffer requirements of a stream can not be satisfied, Aug. 6 1996. US Patent 5,544,327.
 - [9] de Volkskrant. Nederland op vijfde plek met bijna miljoen netflix-abonnees. <http://www.volkskrant.nl/media/nederland-op-vijfde-plek-met-bijna-miljoen-netflix-abonnees-a351034/>, 2015. Last accessed February 27, 2015.
 - [10] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali. Cloud computing: Distributed internet computing for it and scientific research. *Internet Computing, IEEE*, 13(5):10–13, 2009.
 - [11] I. Drago, M. Mellia, M. M. Munafo, A. Sperotto, R. Sadre, and A. Pras. Inside dropbox: understanding personal cloud storage services. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 481–494. ACM, 2012.
 - [12] Dropbox. Share the power of great sync with external collaborators. <https://www.dropbox.com/business/uses/file-sharing>. Last accessed March 2, 2015.
 - [13] R. P. FOUNDATION. What is a raspberry pi? <http://www.raspberrypi.org/help/what-is-a-raspberry-pi/>, 2015. Last accessed February 27, 2015.
 - [14] A. R. Hevner. A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2):4, 2007.
 - [15] C. Höfer and G. Karagiannis. Cloud computing services: taxonomy and comparison. *Journal of Internet Services and Applications*, 2(2):81–94, 2011.
 - [16] N. HOWTOS. Load balancing with nginx. <http://www.networkinghowtos.com/howto/load-balancing-with-nginx/>, 2013. Last accessed February 27, 2015.
 - [17] IBEX. Streaming video using gstreamer. <http://www.raspberry-projects.com/pi/pi-hardware/raspberry-pi-camera/streaming-video-using-gstreamer>, 2014. Last accessed February 27, 2015.
 - [18] K. W. R. James F. Kurose. *Computer Networking A Top-Down Approach*. Pearson, sixth edition edition, 2012.
 - [19] jelleke60. How-to: Raspberry pi media server (easy). <http://www.raspberrypi.org/forums/viewtopic.php?t=56149>, 2013. Last accessed February 27, 2015.
 - [20] T. D. Journal. The datacenter journal. raising data center power density. <http://www.datacenterjournal.com/it/raising-data-center-power-density/>, October 2013. Last accessed February 27, 2015.
 - [21] J. Kiepert. Creating a raspberry pi-based beowulf cluster. http://coen.boisestate.edu/ece/files/2013/05/Creating.a.Raspberry.Pi-Based.Beowulf.Cluster_v2.pdf, May 2013. Last accessed February 27, 2015.
 - [22] J. Martin, Y. Fu, N. Wourms, and T. Shaw. Characterizing netflix bandwidth consumption. In *Consumer Communications and Networking Conference (CCNC), 2013 IEEE*, pages 230–235. IEEE, 2013.
 - [23] A. P. Miettinen and J. K. Nurminen. Energy efficiency of mobile clients in cloud computing. In *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing*, HotCloud’10, pages 4–4, Berkeley, CA, USA, 2010. USENIX Association.
 - [24] T. U. of Tennessee. What load balancing methods are available? <https://help.utk.edu/kb/index2.php?func=show&e=1699>, 2015. Last accessed February 19, 2015.
 - [25] L. Orsini. Building a raspberry pi vpn part one: How and why to build a server. <http://readwrite.com/2014/04/10/creating-a-raspberry-pi-vpn-tutorial-server-secure-web-browsing>, 2014. Last accessed February 27, 2015.
 - [26] ownCloud. owncloud. <https://owncloud.org/>. Last accessed March 2, 2015.
 - [27] PCextreme. Raspberry pi colocatie. <http://raspberrycolocatie.nl/>, October 2013. Last accessed February 27, 2015.
 - [28] PCextreme. Waarom doen we dit. <http://raspberrycolocatie.nl/why/>, 2013. Last accessed February 27, 2015.
 - [29] L. Plissonneau and E. Biersack. A longitudinal view of http video streaming performance. In *Proceedings of the 3rd Multimedia Systems Conference*, pages 203–214. ACM, 2012.
 - [30] L. Qian, Z. Luo, Y. Du, and L. Guo. Cloud computing: An overview. In *Cloud Computing*, pages 626–631. Springer, 2009.
 - [31] raspberry hosting. Raspberry pi hosting. raspberry pi in the heart of europe. <https://raspberry-hosting.com/en>, 2015. Last accessed February 27, 2015.
 - [32] E. Sverdlöv. How to set up nginx load balancing. <https://www.digitalocean.com/community/tutorials/how-to-set-up-nginx-load-balancing>, 2012. Last accessed February 27, 2015.
 - [33] tpy’s. Live 1080p video streaming from the raspberry pi to browsers using nginx and rtmp. <http://tpy.tw/?p=109>, 2014. Last accessed February 27, 2015.
 - [34] F. P. Tso, D. R. White, S. Jouet, J. Singer, and D. P. Pazaros. The glasgow raspberry pi cloud: A scale model for cloud computing infrastructures. In *Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on*, pages 108–112. IEEE, 2013.
 - [35] J. L. Wolf, S. Y. Philip, and H. Shachnai. Disk load balancing for video-on-demand systems. *Multimedia Systems*, 5(6):358–370, 1997.
 - [36] L. Youseff, M. Butrico, and D. Da Silva. Toward a unified ontology of cloud computing. In *Grid Computing Environments Workshop, 2008. GCE’08*, pages 1–10. IEEE, 2008.