ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org

Vol 12, Issuse.4 Oct 2021

Examining the Cloud/IX OS on ARM-based Data Center Servers

DR.V.V.SUNIL KUMAR[1], K.SUNIL KUMAR[2],

Abstract

A transition from costly hardware to a large number of inexpensive servers has become a dominant trend in data center design architecture, creating new challenges for data center architects and necessitating the adoption of fresh approaches. In this paper, we explore alternative approaches to designing distributed systems in the spirit of the Plan9 operating system. We begin with an overview of application and research initiatives such as the porting of Plan9 to the IBM Blue Gene/L supercomputer, the usage of Plan9 in data centers and clouds, and the development of distributed embedded systems. Then, we present Cloud/IX, an OS for ARM-based server systems that, like Plan9, is based on the Plan9 architecture and runs on top of a variant of Plan 9 called 9front. We also detail the infrastructure and findings of an experimental evaluation of Cloud/IX on a real-world, multi-server farm in a data center.

Keywords:

operating systems; distributed systems; Plan 9 operating system model; server platforms; data centers; Cloud/IX operating system.

Introduction

Modern data centers look very different than they did just 10 years ago and it's not just that 10 years is perhaps longer than any computer lifespan. The emergence of Big Data, Cloud Computing, 4G Mobile data, and other modern trends drastically changed the spectrum of industry's tasks and problems. Recent advances in science suchas genetic sequencing and nuclear research made sure we can safely assume that data production and massively parallel processing (MPP) continues its exponential growth. Integration with private and public clouds, server consolidation and full virtualization, more extensive skill set of data center personnel are all consequences of this explosive growth. This poses new tasks and demands the use of different strategies for data center (DC) architects. An increase in power consumption in DCs constitutes the major problem to the DC market development. Energyrelated costs account for about half of the total server maintenance costs of data center ownership, while most of these goes to the provision of power supply and cooling to the servers. In general, we can consider two approaches to solution of the energy efficiency problem, namely: an efficient use of existing facilities and resources (e.g., use of virtualization and cloud computing, allowing to increase utilization rate of available resources and to decrease the equipment needs), and new architectural solutions to the data center designs (e.g., Cisco Unified Computing System).

The energy efficiency problem of the server design is approached at different levels from a processor core, through the single server, and up to the server farm. At processor level, the ARM energy-efficient processor architecture [1] is nowadays considered a solution of choice to the development of server hardware. According to IDC forecasts, by 2015 the ARM architecture can win more than 15% market share of server hardware. In September 2010, the British company ARM Holdings first entered this market with Cortex-A15 processor based on ARMv7-architecture.Their version of the ARM Cortex A15 MPCore, designed for CPU clocks of up to 2.5 GHz, demonstrated the five times performance improvement with respect to the processors used in smart phones while maintaining the same energy consumption levels. Recently, ARM has released the first 64-bit processors - CortexA57 and A53, and a new 64-bit architecture - ARMv8, which is nearly ready for mass production

.

PROFESSOR[1], ASSOCIATE PROFESSOR[2],
DEPARTMENT OF CSE
PBR VISVODAYA INSTITUTE OF TECHNOLOGY AND SCIENCE::KAVALI

To leverage the advantages that new energy-efficient processor architecture offers, the server architects request new software solutions both at the operating systems level and at level of the server virtualization layers. An example of developments in this area is the x86 binary code compiler for the ARM architecture designed by the Elbrus Technology [2]. It allows for migration of software written for x86-based servers to the ARM processor architecture. Software emulator will allow unchanged run of applications compiled for the x86-architecture on the ARM server. Recently, ARM have published package additions to the Linux kernel, which provide support to the instruction set of the ARMv8-core. These additions are now implemented in a number of flavors of Linux, including Ubuntu. Thus, the design and development of an operating system for servers based on ARM processor architecture continues to be a task of great importance, a quality solution to which is expected to allow creation the efficient distributed server farms in terms of performance, power consumption, and scalability. In this paper, we describe our solution to the distributed systems design problems that we approach with the development of a new operating system called Cloud/IX. The design of our own operating system follows the Plan9 model and is implemented on top of one of Plan 9 derivatives called 9front - a free software distributed operating system [3]. We present the general characteristics of Plan 9-based approach to the design of distributed systems, and introduce the Cloud/IX operating system for data center servers based on ARM processors. Section 2 provides quick discussion of distributed systems application spectrum in its relation to the computer architecture problems. Section 3 presents key features of the Plan 9, accompanied with the examples of Plan 9 use in diverse distributed computing application areas. These include project of porting Plan 9 to the supercomputer platform (for MPP scientific computing applications), project of using Plan 9 in data center server platform (for distributed and cloud systems applications), and a research project of adding real-time scheduling into Plan 9 for distributed embedded systems (DES) applications. Section 4 describes our Cloud/IX operating system for the ARM-based server platforms. The experimental testbed and results of experimental tests of the Cloud/IX are discussed in Section 5. Finally, some conclusions and future work are shown in Section 6.

Distributed Applications and its Impact on Computer Architecture Design

There is a huge class of tasks that can be well parallelized, which makes it much easier to run them on multi-node computer architectures – this includes mostly computational tasks, such as fluid dynamics and optical modeling as well as generic data processing, i.e. genetic sequencing and text processing, including web search and indexing. Web search and indexing is, in fact, one of the most, if not the most widespread use of computing resources today.

Some estimates put the percentage of worldwide CPU cycles spent on it as high as 35%, which sounds believable when you count for a fact that a search engine consumes not just its own hardware's resources, but also each and every server's it indexes. This caused a prominent trend in data center architecture design - a shift from powerful and expensive hardware (like mainframes 25 years ago and HP Superdome about a decade later) towards a multitude of simple servers. These massively parallel architectures can use either Google-like server farms built from off-the-shelf hardware or proprietary blocks forming what today is called a supercomputer (IBM Blue Gene).

Trends in operating systems research and development

While computer hardware evolved at the blazing speed, the software counterpart obviously could not remain the same. Early operating systems creators didn't care much for architecture – nobody at the time had the experience of writing a program this big. As one of the consequences, those early OS's lacked the modular structure, each and every subroutine could be called globally, and the entire thing was a huge monolithic “blob”. This made scaling and expansion extremely difficult. First OS/360 release took 5 years and 5000 people to write and amassed just over 1 million lines of code. Its successor Mastics, released in 1975, already grew to 20 million lines. It was obvious that without radical review of design principles further advances were impossible. Thus, modular paradigm was born, and most of the development in modern software engineering is still based on it or its variants. Modular design naturally led to modules with similar functionality grouping together and stratification of OS into hierarchical

model. Practically all modern OS's can be subdivided into following levels:

Hardware support

Machine-dependent code

Common kernel mechanisms

Resource manager

System calls API

Utilities.

Sometimes levels are split or combined, like in nanokernel and microkernel architectures, sometimes even swapped (exokernel), but the basic structure remains more or less the same. Another huge step in OS design was made when IBM introduced its Virtual Machine that abstracted the underlying hardware from the lowest level of OS. This made possible system partitioning and running several instances of OS on the single physical machine. For a couple of decades, virtualization was exclusively mainframe feature, but it was, of course, bound to propagate into server and workstation world. Nowadays, all major processor manufacturers include hardware virtualization support (VT for x86, TrustZone for ARM) [4]. It is estimated that today over 50 percent of all server workloads are virtualized and this figure is projected to reach 86 percent by 2016 [5]. Virtualization also spread to workstations where it is widely used as a low-cost software alternative to acquiring dedicated hardware for test and debugging purposes. Lately, it got even to the mobile and embedded segments of the market, where its benefits are security, interoperability and, once again saving on hardware – a virtual multimedia processor can be as good as a dedicated physical one [6]. Virtualization allows computational resource sharing and partitioning, but there is also need for exactly the opposite – not slicing the existing system into a number of virtual machines, but uniting the resources of multiple systems into a bigger and more powerful "supersystem". While virtualization techniques are nowadays ubiquitous, including hardware manufacturers' support (VT for x86, TrustZone for ARM) and well known software solutions (VMWare servers and stations, Oracle VBox, Xen), aggregation is a much more complex problem. Simply speaking, if a virtual node, from software viewpoint, is indistinguishable from a physical system, the topology of an aggregated system is quite different from a single node. And, of course, effective use of these aggregated resources requires some sophisticated techniques.

Related works: Plan9 operating system revisited

Nowadays, there is a renewed interest in another OS – Plan9 and its derivatives. Plan9 is an OS developed in the late 1980s and early 1990s at AT&T Bell Laboratories by the group of researchers and engineers that included some of the original UNIX creators [8]. In Plan9 design they attempted to straighten out what they thought went wrong with UNIX and its ancestors. When introduced to the USENIX community in 1992, it was received very well, with reviews ranging from carefully optimistic to outright ecstatic branding it "a UNIX killer". Killing UNIX did not happen – we can only guess for specific reasons, but the general consensus seems to be that while Plan 9 was in many ways superior to UNIX, it just failed to gain critical mass on the improvements [9]. Simply speaking, UNIX and later Linux as one of UNIX flavors were not as elegant but still good enough. This, combined with its massive code base, put it in an industry leader position. Plan9, meanwhile, found a niche as hobbyist and research system. It has, as any great but underachieving project would, a small but dedicated army of followers. Its impeccable pedigree and elegant design also make it very attractive as a subject in Operating Systems courses in academia. Plan9 is based on three major principles: x All resources are named and represented by files in a filesystem x There is a standard protocol, 9P, for accessing files across node boundaries x Separate filesystems can be joined into a single private name space It was aggressive application of these principles that kept Plan9 consistently compact and robust through the years and a major rework in 2000-2004. Some of Plan9 features turned out to be so attractive they were adopted by mainstream UNIXs. Most prominent of those is, perhaps, a filesystem interface to system per-process statistics - /proc filesystem. Linux's /sys filesystem representing system-wide resources is another nod in that direction. Plan9 also introduced UTF-8, a full and honest n2 set of native and cross-compilers and linkers for all supported architectures, and some other nifty innovations.

One of the most attractive Plan9 qualities is its compact size. Historically, it was introduced "when things were small" and even Linux was not the monster we know today. And it managed to stay that through the years. For example, cat utility resident footprint on Ubuntu 12.04 is 384K while its Plan9 counterpart is just 11K. Similarly, most standard utilities common for both systems show a factor of 10 to 30 in memory footprint. Cache usage is, of course, much more conservative in Plan9 as well, which it even more important for performance. This 'tight and robust' paradigm made Plan9 an attractive candidate for embedded systems design. There it was always, although marginally, present, particularly in network equipment and

storage systems. The distributedprocessing model of Plan 9 is very effective and flexible, and it is attractive for embedded systems. The 9P protocol is useful for inter-system communication. The private name space of Plan 9 also enables flexible and safe distributed processing in embedded systems. Plan 9 can run on various hardware platforms and is highly suited to building large distributed systems. A typical Plan 9 installation would comprise one or more file servers, some CPU servers and a large number of terminals (user workstations). The small size and straightforward structure of (most of) its source code, and low system management overhead, makes it particularly suitable for distributed embedded systems (DES) applications. The server world, though, lusts for Plan 9's other features – and first of all a relatively thin 9P protocol and the ease of inter-node communication by manipulating name spaces. This leads to a higher level of abstraction – applications are agnostic of their execution details. They can run anywhere on any node in the system, on any architecture. Client can run, within one session, several programs on geographically separated machines. This improves modularity of any project by representing any information or data as a set of plain files [10]. 9P protocol was implemented for several foreign systems, including Linux. Actually, for Linux there are is a 9P server allowing accessing files on a Linux server from Plan9 station and 9P client to access Plan9 files from Linux computer. This is very handy for cross-platform development. As proof of the renewed interest in Plan9 OS we'll take a look on three projects.

Cloud/IX operating system – a Plan9-based solution to ARM-based server platforms

When selecting the basic operating environment for the development of Cloud/IX OS we used multiple criteria, including among others the support for distributed operations, scalability, license purity, easy porting of device drivers and applications, easy deployment and support, standard interfaces, minimal system services overhead. The main advantage of the 9front system for distributed server application is its ixP protocol, which allows for managing local and distributed resources by simple mappings onto the namespace. Perhaps the most notable disadvantage of 9front is the difference between its set of system interfaces and the POSIX, which is a traditional standard solution for the similar products.

Here we can consider two different approaches to solution of this problem. First, an APE (ANSI / POSIX Environment) package was developed for 9front - the best approximate of the system interfaces to the POSIX. Second, our team in association with AltLinux company undertake efforts of porting Linux on the ARM and microTCA-based server platform (ARM, microTCA). This will permit easy adaptation to the target platform of many applications developed for the Linux, including traditional cluster applications. It should be noted that many useful features of 9front design were adapted to Linux. This applies, in particular, to the ixP protocol, the use of which in Linux is now possible at the level of mounted file systems that allows for exchange of files between Linux and 9front. Since our Cloud/IX is based on 9front, it inherits all the features of its prototype. The system is based on three principles: x Resources are named and are available as files in a hierarchical file system x ixP standard protocol for access to local and remote resources x unbounded hierarchies provided by diverse services are linked together into an own hierarchical file namespace. ixP protocol implements multiple transactions, each of which sends a request from the client process to the local or remote server and returns the result. ixP controls the file system, not just files. Access to the files occurs at the byte level, not blocks, which distinguishes ixP from protocols such as NFS and RFS [19]. At present, a β-version of the Cloud/IX operating system is developed, and a work is performed on porting the most popular and commonly used software applications (e.g., nginx – a web-server and a mail proxy-server running on Unix-like operating systems).

5Experiments with Cloud/IX

We have carried out tests of the software prototype of the ARM-based server platform in order to study the stability of the ported nginx http-server on heterogeneous Cloud/IX system and its scalability. The tests were performed in the Data Center at the Systems and Solutions Ltd. on a distributed computer system that comprises 24 x86-based computers (blade servers), organized into 3 system racks each with 8 blade servers. All blade servers are equipped with at least one Ethernet 1000Mbps controller.

Experimental

Setup During the experimental testing, we have monitored the load level in cluster nodes with OS services (separately for each subsystem), which allows to conclude about potential ways of performance improvement. To display the results of the monitoring, a special purpose software was developed to collect statistics from multiple nodes in a cluster, to aggregate it on the single node, and to transform monitoring results into format suitable for the analysis and display in real-time. The software receives data about the node's and network interface's workloads and displays it in a visual form on a web page. The solution is

implemented using the following technology stack: x Server part of application is written in Clojure – a lisp-dialect implementation for the JVM and libraries: Ring, Composure, Web bit, Clj-json x Client part is implemented in Clojure Script – a dialect of Clojure that translates into a regular JavaScript, executed in the browser x Implementation of message passing mechanism from the server to the client is based on WebSocket's technology x Dynamic rendering of the graphical elements is realized by working with Canvas element (HTML5 specification). To generate requests to the load balancer,nix the httpperf utility is used. Httpperf measures the performance of web server and provides a flexible environment for generating workloads for the HTTP-server and for measuring its performance.

At the end of test run, it generates a report, which contains three sections:

general results, a groupdedicated to compounds and group. In the load generating mode, it generates requests with the substitution of the growing numbers, which digits are used as components of the path to the resource. Scenario of testing the effectiveness of load balancing require the creation of a nginx file hierarchy one each node, so that their paths with respect to the root directory of nginx match the query, formed by httpperf, and thus the total amount of data would exceed the size of RAM in each node. Under these conditions, the disk subsystem becomes the bottleneck at each node, so that it becomes possible to evaluate the effect of workload parallelization. The same sequence of non-recurring requests is submitted to the balancer and to the separate node, and the results are compared. Performance of individual components and of the entire cluster is also tested by repeated (identical) queries. Httpperf allows you to adjust the number of requests per unit of time, which is reflected in the number of requests processed in parallel. The test is carried out separately for downloading large files and for downloading small files, allowing you to identify the various potential bottlenecks in the ported nginx. In this test scenario, the entire contents of the file in the cache of the operating system, and a disk subsystem is no longer a bottleneck. We wanted the test results to reflect the performance of the server solutions (nginx), as opposed to the entire client-server complex. For this, it requires either a presence of multiple client computers, generating queries simultaneously, or the use of a system for running the client, which outperforms significantly a set of all nodes in the cluster (without disk subsystem that is not used by the client actively). For this study, we have chosen the second solution.

Conclusion

In this article, we presented the Plan 9 operating system model-based distributed systems design strategy. Using examples from the realms of the supercomputer, server platforms, and distributed embedded systems, we demonstrated that the fundamental concepts of the Plan 9 OS are ideally adapted to capture the distributed processing mechanisms originating from parallel and distributed computational/programming models. Applications may be written independently of the specifics of the hardware on which they run thanks to the standardization of the filesystem interface and the simplicity of inter-node communication made possible by the manipulation of file name spaces. They are completely portable and can function on any system node, regardless of hardware.

In turn, this leads to an application that takes into account both the model's specified functionality and the communication requirements of a distributed computing environment. Plan 9's implementation of the 9P protocol offers a framework for designing scalable distributed system architectures and a hint for allocating workloads among the available nodes in the system. The ability to represent any kind of data or information as a collection of simple files greatly enhances the project's modularity. In addition, Plan 9's system servers are user mode processes, making it a breeze to write new software for. In conclusion, the rising popularity of Plan 9 and its offshoots is a definite trend. Extensions to Plan 9's real-time and MPP support are the subject of many active initiatives.

References

[1] S. Orlev. Revolution ARM. Journal of network solutions. LAN №11, 2012. Available at: http://www.osp.ru/lan/2012/11/13032394/.

[2] Startup Elbrus Technologies' emulator will allow ARM processors to work with x86-applications. Available at: http://servernews.ru/596643.

[3] "Plan 9 from the People's Front of cat-v.org (9front)", NineTimes, June 17, 2011, retrieved September 13, 2012.

[4] T. Laplante, Virtualization has surpassed 50 percent of all server workloads, DataCenterPost.com, March 20, 2014. Available at: http://datacenterpost.com/2014/03/virtualization-surpassed-50-percent-server-workloads.html.

[5] O. Kharif, Virtualization goes mobile, Bloomberg Businessweek. Technology, April 22, 2008. Available at: http://www.businessweek.com/stories/2008-04-22/virtualization-goes-mobilebusinessweek-business-news-stock-market-and-financial-advice.

[6] J. Dean, S. Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, 6th Symposium on Operating Systems Design & Implementation (OSDI'04), December 6-8, 2004, USENIX 2004, pp.137-149. [PDF].

[7] M. Isard, M. Budiu, Y. Yu, A. Birrell, D. Fetterly, Dryad: Distributed Data-Parallel Programs from Sequential Building

Blocks, European Conference on Computer Systems (EuroSys'07), Lisboa, Portugal, March 21-13, 2007. [PDF].

[8] R. Pike, D. Presotto, S. Dorward, B. Flandrena, K. Thompson, H. Trickey, and P. Winterbottom, Plan 9 from Bell Labs, Computing Systems, vol. 8, no. 3, 1995, pp. 221–225. Available at: http://plan9.bell-labs.com/sys/doc/9.html.

[9] E.S. Raymond, The Art of Unix Programming, Thyrsus Enterprises, 2003.

[10] D. Presotto, P. Winterbottom, The Organization of Networks in Plan 9. Available at: http://plan9.bell-labs.com/sys/doc/net/net.html.

[11] E. Van Hensbergen, C. Forsyth, J. McKie, and R. Minnich, Petascale Plan 9 on Blue Gene, USENIX 2007 Annual Technical Conference (USENIX ATC'07), June 17-22, 2007, Poster Session. [Abstract].

[12] R.G. Minnich, J. Floren, and A. Nyrhinen, Measuring kernel throughput on Blue Gene/P with the Plan 9 research operating system, in: Proceedings of the 6th International Workshop on Plan 9 (IWP9), Athens, GA, USA, October 12, 2009. [PDF].

[13] J. McKie, J. Floren, Edging Towards Exascale with NIX. [PDF]

[14] NIX is a new multicore OS based on Plan9. Available at: http://code.google.com/p/nix-os/.

[15] F.J. Ballesteros, CSP-style Network, File, and System Services in Clive. Lsub Systems Lab, Universidad Rey Juan Carlos, Madrid, TR Draft, May 23, 2014. [PDF].

[16] S.J. Mullender, P.G. Jansen, Real Time in a Real Operating System, in: Herbert, Andrew James (et al.) (Eds.), Computer Systems. Theory, Technology, and Applications, Springer, 2004, pp. 213-221. ISBN 978-0-387-21821-2. [PDF].

[17] S.J. Mullender, J. McKie, Real Time in Plan 9, in: Proceedings of the 1st International Workshop on Plan 9 (IWP9), December 4-5, 2006, Madrid, Spain. [PDF].

[18] Y. Sato, K. Maruyama, LP49: Embedded system OS based on L4 and Plan 9, in: Proceedings of the 4th International Workshop on Plan 9 and Inferno (IWP9), Athens, GA, USA, February 21-23, 2009. [PDF].

[19] H. Trikey, APE - The ANSI/POSIX Environment, Plan 9 Programmer's Manual, Volume 2, AT&T Bell Laboratories, Murray Hill, NJ, 1991