Toward A GPU-VDI Computer Room: Concepts and Problems

Ken'ichi Fukamachi*

Jun 16, 2017

Abstract

Today students need to use computers anytime, anywhere via multi-devices. To help it, we propose a concept system which is a virtual infrastructure provisioning VDI powered by GPU. It should be a multi-purpose infrastructure. We want it to have a feature that we build a trial and do stepwise replacement. The network design depends on the profile of software usage. The most important question is latency but the solution will be a provider/carrier issue.

1 Introduction

Today computer rooms in universities need to change drastically. The computer room has been used for traditional programming class. However it is indirectly needed for almost all classes since students use computers for preparation, review and homework.

The traditional design of university facility does not suppose such situation. There are not enough computers and rooms for them. Hence, BYOD[1] is expected to be a solution for it. It implies first-year undergraduates inevitably buy their own PC or tablet[2].

It is a well-known fact that the university education rebuilding is required. Ministry of Education Culture, Sports, Science and Technology (MEXT) has been promoting "Acceleration Program for University Education Rebuilding" (AP) [1]. One theme of AP contains "how to ensure the study time" which is a part of quality assurance in university education. MEXT defines that the granting of a credit requires 45 hours [2] [3] where 45 hours includes a set of preparation, lecture/exercise and review. To accomplish this criterion, students need to use computers anytime, anywhere via multi-devices. This problem has something in common with work-style reforms of companies.

2 Computer Room Case Studies

2.1 Traditional Case

Chitose Institute of Science and Technology (CIST), the author belongs to, has traditional computer rooms. There are FAT clients running Windows/CentOS (dual-boot).

Computer rooms exist within the campus network. The Remote access to the computer rooms is not allowed.

2.2 BYOD Cases

One famous example of large scale BYOD (over 10,000 students[4]) is Kyushu University[5] though we can find smaller advanced cases [6]. First-year undergraduates buy their own PCs. Computer rooms are not required for them. Kyushu University announced it would phase out the computer rooms.

We can find another hybrid solution such as BYOD with a few computer rooms. This example is

*Chitose Institute of Science and Technology. 758-90, Bibi, Chitose, Hokkaido, 066-8655, Japan. mailto:k-fukama@photon.chitose.ac.jp http://www.nsrg.fml.org/

[1]BYOD in university education implies that "students bring their own note PCs or tablets" which is not needed to be under management such as MDM (Mobile Device Management) or MAM (Mobile Application Management) mandatary in company network.

[2]In some cases, the intuition contains a BYOD device.

Tokyo University of Agriculture and Technology[7]. The students have their own devices and also can use HTML5 based VDI hosted on the private cloud. Computer rooms on the private cloud provides virtual but traditional computer rooms where all students can use the same computer environments. This solution is a little expensive but must be helpful to avoid BYOD troubles described below.

2.3 BYOD Problems

It is possible that BYOD increases operation, support and license costs though computer room hardware decrease. These costs are considered to be needed to solve the shortage of computers and rooms.

Firstly, consider operation and support costs. In all likelihood, a traditional computer room has a central management system for client systems. It provides the same OS (+ security patches) and 3rd party software for all users. This OS master image is updated once per month/term et.al. When the user log-off/out, the user specific data on the OS is initialized.

In the case of BYOD, each user has each computer environment. It is very difficult to fix a troubled BYOD device since there are slightly different thousands of computers. Therefore BYOD must need help desk support. This support cost may be larger than a traditional computer room.

Secondly, consider license cost. Traditional computer rooms are designed for just classes not the simultaneous use of all students. In the case of BYOD, we need to assume the simultaneous use of all students.

If we can use the site license for all software, we can estimate the upper limit of license cost. However licenses of department specific scientific software are various. In the worst case, we need to buy (device) licenses for all students even if they use them or not.

It is difficult to evaluate the license cost in the case of BYOD. It highly depends on software we use. It varies a lot from universities to universities.

Lastly, consider hardware cost. In the case of BYOD, hardware (client and server PCs in computer rooms) is not required, so that hardware cost decrease.

For BYOD WiFi network all over the campus is required. If we do not have it, we should construct it.

3 Our Concept

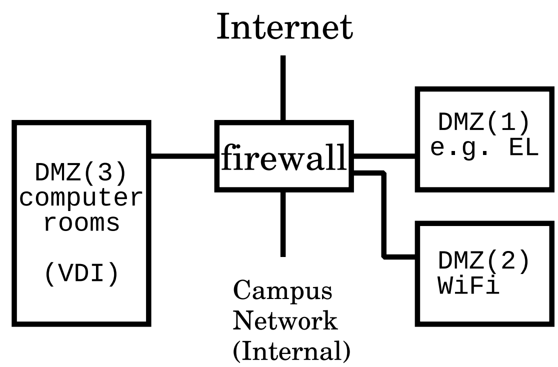

Figure 1: Basic network structure: DMZ(1) has several learning assistant and management systems such as a portal and E-Learning system. DMZ(2) is WiFi network covering the whole campus. DMZ(3) is our virtual computer room. DMZ(3) has virtual infrastructure with GPU of NVIDIA GRID. It provisions VDI virtual machines on demand. It can also provide surplus resources for any purposes. The internal network contains laboratories and the back office.

CIST promotes AP to ensure quality assurance in university education[8]. We suppose students want to use computer rooms anytime, anywhere.

It seems we need to accept a BYOD solution but we would like to avoid the increase of the operation and help desk support cost as could as possible.

Hence we, "our laboratory" [3], propose the following concept and have been evaluating the system as a next-next generation computer room model. We assume the following requirements:

- virtual infrastructure

[3] CAUTION: This is just our laboratory project and vision. It is not authorized by the institute.

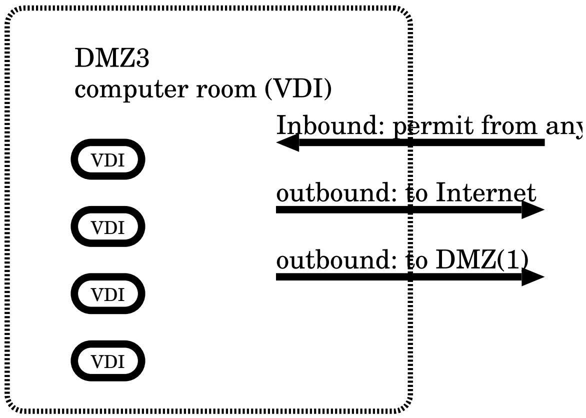

Figure 2: DMZ(3) (computer room) traffic flow: Students are able to use DMZ(3) anytime, anyware via multi-devices. Fundamental access control rules are as follows: (a) The inbound traffic to DMZ(3) should be allowed from anywhere. (b) The outbound traffic from DMZ(3) is limited to DMZ(1) and Internet.

- GPU (NVIDIA GRID[9]) powered computer resources
- VDI usable anytime, anyware via multi-devices
- multi-purpose use
- trial and stepwise replacement
- minimum operation cost
- (on-premises)

3.1 Basic Configuration and ACL

See Fig. 1.

DMZ(1) has several learning assistant and management systems such as a portal and E-Learning system. DMZ(2) is WiFi network covering the whole campus. DMZ(3) is our virtual computer room.

Students are able to use DMZ(3) anytime, anyware via multi-devices.

Fig. 2 shows the fundamental access control rules: (a) The inbound traffic to DMZ(3) should be allowed from anywhere, (b) The outbound traffic from DMZ(3) is limited to DMZ(1) and Internet.

DMZ(3) has virtual infrastructure with GPU of NVIDIA GRID [4]. It provisions VDI virtual machines on demand. The surplus resources can be used for any purposes.

3.2 Multi-Devices

We should support multi devices.

Today a modern person has plural devices: smartphone(s), tablet(s) and note PC(s) et.al. Hence we assume a student probably have a smartphone even if he/she have no PC at home.

Our preliminary questionnaire suggests that a smartphone is sufficient for young people to use as a multipurpose tool.

A modern smartphone is a powerful computer for us to use for almost all work though it has no large display and keyboard for fast program typing. However if they need a keyboard/mouse, they can use Bluetooth keyboard/mouse. If they need a large display, they can use a modern TV with HDMI. If they have an Amazon Fire Stick (2017 new model)[10], it is useful to connect a smartphone with TV. It does not seem one more note PC is required.

Hence it is preferable to support students (except for programmer applicants) with unlimited access to VDI environments by their smartphones without extra charge in order to study on attending school and at home.

3.3 GPU Power and Multi-Purpose Use

Today it seems that science and engineering institutes need to have a high power computing resources shared with all users in the institute. In order to build an efficient system, the system should be virtualized. This virtual infrastructure needs to have GPU based high power computing facilities to assist the research

[4] Actually it is possible to reach a settlement on a hybrid of small VDI environments and traditional computer rooms smaller than now. Figures in this paper assumes full VDI environments.

and exercises of deep learning, VR using Unity[11], scientific calculation/simulation and others.

This infrastructure should be used for several purposes: typically (a) virtual computer rooms (VDI) daytime (b) high power computing not daytime and (c) surplus resources can be used for any purposes. The surplus resources can be assigned for laboratory VDI resources to decrease the help cost for teaching staff.

We also need GPU even for clients (VDI VM) since the modern operating systems premise GPU.

3.4 Importance of Trial System

It is important to build and run a trial system for at least several months since people cannot accept a new invisible unknown IT infrastructure without using it actually. We verified it through the network system replacement[12].

It is preferable to replace the system step by step. Our infrastructure needs to support small start and easy extension.

3.5 Operation

3.5.1 Virtual Infrastructure

Operation of traditional computer rooms seems already almost minimized. The operation should be inherited to one of the virtual infrastructure.

One of the simplest example of the operation is as follows: There is one OS master image. On the virtual infrastructure, a control system provisions a user VDI based on this master image on demand. The master image is updated once per month. When the user log-off/out, the VDI is initialized. Each user cannot customize the OS e.g. each user cannot install a new software. Each user property can be stored on the common network storage (e.g. Microsoft Moving Profile). Each user can have its home directory/folder (e.g. Z: drive) to save the data on the network storage.

3.5.2 User Devices

Each user device itself is out of support. We should help the part of client software e.g. Citrix Receiver installed in the user device.

3.6 On-premises vs Cloud

In many cases universities and institutes locate in country. The phrase “Cloud First” is true when we locate at the center of a big city. Hence, the on-premise has been one of choices.

3.7 License

License issue in our case is regarded as same as the BYOD one. The difficulty depends on the profile of computer software usage.

4 Discussion

VDI using NVIDIA GRID is enough mature. It is widely used in secure closed environment. One of the most famous example is Nissan[13] using VDI for 3D CAD of the car design. Next, we should apply it to open network. It seems that a trial to use VDI over the Internet has begun. One example is to share BIM (Building information modeling) among the main office and building sites[14]. Our concept is regarded as a variation in this direction.

The network design of our system may be changed according to the profile of usage whether a lot of students want to use computer room specific software or not in commuting to school and from home. It needs further investigation. If the bandwidth is required more than we imagine, we may need to develop a hybrid system of on-premise and cloud which are connected through VPN where a new VDI at cloud is launched for the user who request a VDI by time slot or by the distance.

The most important question is latency. It is presumed that about 200 ms is the threshold[15] to use VDI comfortably. See Table 1 which shows several Round Trip Time (RTT) snapshots. We can find the RTT varies widely up to discomfort when we use LTE in Japan. The variation looks to depend on the time slot such as morning, noon, evening and others. It is intuitively proper since a lot of people use Internet

RTT (ms)	type	from	gate to Internet	information
100	LTE	SAPPORO	OSAKA	Tue,16:00
150	LTE	SAPPORO	TOKYO	Mon,12:00
60	WiFi	NAHA	-	Fri,07:00 hotel
80	LTE	NAHA	TOKYO	Fri,07:00 hotel
100	LTE	NAHA	TOKYO	Fri,16:30 airport

Table 1: Round Trip Time (RTT) examples: These are just snapshots (not statistically valid). The ICMP Target machine locates at CIST which is connected to Internet via IIJ leased line. LTE and 3G is MVNO(IIJ), an NTT DOCOMO OEM. We can find the LTE gate to the Internet changes randomly. For your reference, the RTT from chitose to a San Jose router in IIJ backbone is 120-130ms.

in commuting time. We then need to review the relation between RTT and usability more closely and systematically.

It is well known that Internet is designed for continuity not latency. The path of packet routing on the Internet is not optimized to be geographically shortest since this path is determined mainly by capitalism not technical reason.

To resolve this difficulty, the day when local IX (Internet eXchange) is necessary at last may come. The idea of local IX has been proposed these 20 years. However local IX is of no use since there is no local content to exchange a lot. For the first time we may find local meaningful traffic. However the local IX problem is "which comes first a chicken or egg dilemma".

Instead, we expect 5G since one of 5G features is "low latency". From the point of the Internet infrastructure, 5G is not only for Tokyo Olympic but also for VDI and sensor network (IoT).

5 Conclusion

For students to use computers anytime, anywhere via multi-devices, we have proposed the concept system described above. Evaluating the system is underway. The result will be reported in the near future. We are interested in the statistical profile of 3rd party software usage and the verification of the hypothesis of our design that a lot of students want to use computer room specific software in commuting to school and from home. These directly lead to the network design and license cost. The further investigation is required.

References

[1] `http://www.mext.go.jp/a_menu/koutou/kaikaku/ap/` [Accessed Jun 3 2017]

[2] `http://www.mext.go.jp/a_menu/koutou/daigaku/04052801/003.htm` [Accessed Jun 3 2017]

[3] `http://www.mext.go.jp/b_menu/shingi/chousa/koutou/053/gijiroku/__icsFiles/afieldfile/2012/10/30/1325943_02_3_1.pdf` [Accessed Jun 3 2017]

[4] `https://www.kyushu-u.ac.jp/ja/university/publication/number` [Accessed Jun 3 2017]

[5] `https://axies.jp/conf2013cd/paper/axies_w2e-1.pdf` [Accessed Jun 3 2017]

[6] `http://www.ctc-g.co.jp/~caua/viewpoint/vol10/04.pdf` [Accessed Jun 3 2017]

[7] `http://www.uniadex.co.jp/casestudy/education/tuat.html` [Accessed Jun 3 2017]

[8] `https://www.chitose.ac.jp/info/pj_kaikaku.html` [Accessed Jun 3 2017]

[9] `http://www.nvidia.co.jp/object/grid-technology-jp.html` [Accessed Jun 3 2017]

[10] `https://www.amazon.co.jp/Staging-Product-Not-Retail-Sale/dp/B01ETRGGYI/` [Accessed Jun 3 2017]

[11] `http://unity3d.com/` [Accessed Jun 3 2017]

[12] https://reg.axies.jp/pdf2016/FF22.pdf [Accessed Jun 3 2017]

[13] https://www.citrix.co.jp/customers/nissan-motor-jp.html [Accessed Jun 3 2017]

[14] http://ja.community.dell.com/techcenter/b/weblog/archive/2014/05/27/japan-bim [Accessed Jun 3 2017]

[15] private communication.