I-CaN-MaMa: Integrated Campus Network Monitoring and Management

Shuai Zhao[1], Kelsey Leftwich[1], Matthew Owens[1], Frank Magrone[2],
James Schonemann II[2], Brian Anderson[2,3], Deep Medhi[1]
[1]Computer Science & Electrical Engineering Department, University of Missouri–Kansas City, USA
[2]Campus Information Services Division, University of Missouri–Kansas City, USA
([3]now with EAGLE Software Inc., Kansas City, Missouri, USA)

***Abstract*—A campus network provides a number of services to its stakeholders. Due to the complexity of having many services, it is often difficult to pinpoint an issue quickly. The currently available tools for monitoring a campus network are often designed for a specific piece of the network. On the other hand, systems administrators running a campus network with data centers and supporting a number of internal and external applications face several challenges with the current set of tools for monitoring and management. To fill this void, we propose the I-CaN-MaMa (Integrated Campus Network Monitoring and Management) framework. Our approach takes an integrated view that was developed in consultations with issues faced by systems administrators. We present the key components in this framework and through a brief study, we show the utility of our approach.**

I. Introduction

A campus network at an academic institution provides services to a number of customers: students, faculty members, academic departments, and administrative offices. For instance, students and faculty members may simply maintain their own web pages, while departments may require website maintenance, as well as servers, for course work and research projects. Faculty research groups may use dedicated servers (virtually) for compute-intensive research projects and may require services for long-term data maintenance. Administrative offices may have a number of services, from student record management to file systems management for different administrative user groups. Furthermore, there are campus-wide needs such as course management tools for courses such as Blackboard. While it may be tempting to say that all services could run using external cloud services, this is not possible for all services due to regulatory requirements. At the University of Missouri–Kansas City (UMKC), the students email service is currently provided through an external cloud provider, but most other services are internally handled. The campus has also moved away from each department running their own physical machines and networks to these being handled through the Campus Information Services Division (*CampusIS*, for short). Thus, the campus information services has a number of diverse requirements to meet so that the best quality of services can be provided to the end customers.

In this paper, we focus on two important components that the CampusIS needs to address. The first is the campus physical network, and the second are the campus data centers that connect users to computing resources through the network. The goal of this project has been to take an integrated approach that gives the CampusIS system administrators (sysadmin, in short) the ability to understand the impact on the network. The impact may be due to different services using the data centers over the network, or due to users accessing external sites. The aim is to allow the CampusIS sysadmin to have a comprehensive view so that they can see how the network is being used by the data centers as well as external applications through a common framework. Because of virtual machines (VMs) provided at the data centers, it is also necessary to see how different traffic flows are incoming or outgoing from a specific VM. While the CampusIS sysadmin uses a number of commercial and public-domain tools, most of them are geared for a specific situation such as `mrtg` [1] that gives views on a link use.

The rest of the paper is organized as follows. Related work is described in Section II. Section III presents the background on the UMKC campus networks and issues of interest. In Section IV, we present our I-CaN-MaMa framework. In Section V, study results are presented using the I-CaN-MaMa framework. Finally, we conclude with a summary and future work in Section VI.

II. Related Work

NetFlow [2] data has become the dominant source to conduct network analysis [3]–[12]. NetFlow data has been used for intrusion detection system development [7], [10], [13], [14] and bandwidth evaluation and provisioning in both Campus [4] and public networks [9], [15].

The existing software for parsing and analyzing NetFlow data includes public-domain tools such as `nfdump` [16], `nfsen` [17], Most of these use `mysql` database schema depending on which NetFlow version is used. `nfdump` [16] is perhaps the most commonly used tool for researchers since it directly deals with NetFlow raw data. Scrutinizer [18] provides many optimizations in terms of both data query speed and the user interface that includes a number of additional features.

There are several works [13], [19], [20] that address network monitoring applications. But most of them focus on one or two

978-1-4799-0913-1/14/$31.00

types of network data sources. The methodology to implement network monitoring varies based on both requirements and expectation. In general, network monitoring can be done in a passive or active mode or in a combined manner. [15] presents a toolkit comparison between passive and active methods.

VMware's `vSphere` [21] is another tool that is available with VMware's products for data centers, which can be used for VM monitoring and management. Note that such a tool does not provide any network level information nor is it designed to do so.

It is clear from the current tools and approaches that most of them are developed for specific situations or for a specific type of analysis. Seldom do we find an integrated approach that considers both NetFlow data in core locations in a campus network and data centers while correlating with a variety of applications to understand their impact on the network.

III. Background on the UMKC Campus Network and Issues

The University of Missouri–Kansas City consists of two physical campus sites, the Volker campus and the Hospital Hill campus, approximately 10 km (6 miles) apart. The network core routing site for the campus is located on the Volker campus with a redundant core routing site located at a co-location facility in downtown Kansas City, which is outside the two campus sites. There is a redundant fiber optic connectivity between these three locations. To make the best use of this connectivity, the campus uses optical wavelength-division multiplexing (DWDM). The UMKC network topology is shown in Fig. 1.

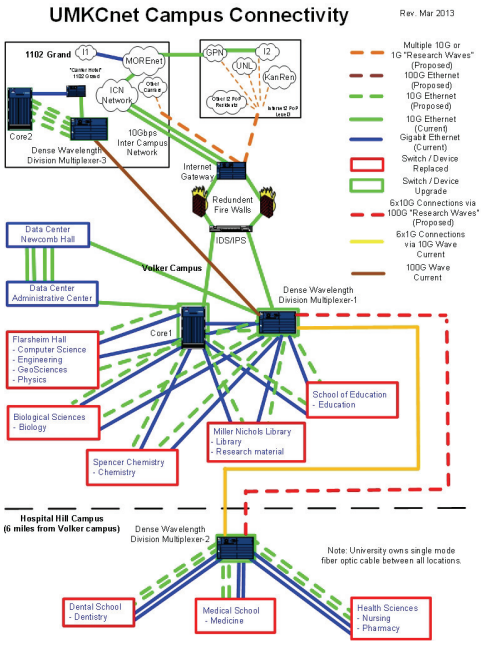


Fig. 1. UMKC Network Topology

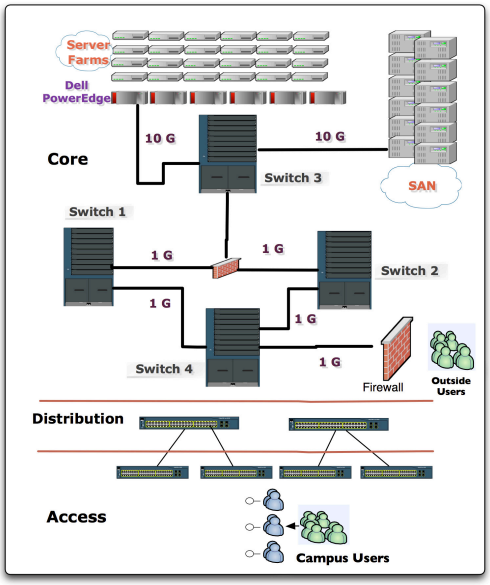


Fig. 2. UMKC data center Three-tier Design

The UMKC campus network maintains three data centers, two at the Volker Hill campus and a smaller one at the Hospital Hill campus. The data centers use a three-tiered architecture (see Fig. 2) where the VMs are deployed using VMware's products. VMware virtualization tools are used for virtual machine configuration and management for end users' needs. Some of these VMs are exposed externally through public IP addresses such as web-sites, while others are strictly for internal research or administrative use on the private IP address space. Certainly, authorized users can use the remote-access VPN service to access the campus first and then access resources in the private address space. Some users' works are compute-intensive, while others' are data-intensive (in terms of the amount of data generated), still others have a combination of both compute- and data-intensive needs.

There are also external and internal applications accessed by users on campus, including students who live in the campus dormitories whose Internet access goes through the campus network. Here, we use the term *application* broadly. For example, YouTube and NetFlix are examples of external applications, while BlackBoard is an internal application for students and faculty members to use for course management. The campus phone service now goes through the campus data network using VoIP, which is another application. In addition, there are service applications such as iSCSI traffic due to the storage area network (SAN). While the primary SAN is located at one of the data centers, the backup for redundancy is done at another building in another data center. Any copy function for redundancy between two different SANs uses the campus network again for transfers that show up as iSCSI traffic.

The campus has the ability to collect and store a large amount of information for monitoring and management. However, there are also a number of issues as noted below:

- All NetFlow data are collected from core switches. However, the NetFlow data by themselves do not identify the semantic information on the source or the destination of each flow since NetFlow records store IP addresses of sources and destinations, and not their domain names. Thus, we need additional information associated with the IP addresses to be able to filter and quantify as (and when) needed for different applications.
- Virtual machines at the campus data centers have certain information available through the `vSphere` software [21] available from VMware. However, this only provides information such as CPU, memory, and so on, but does not have the associated information on where and how a flow starting (or ending) at a VM goes through different segments or switches in the campus network to either internal or external destinations.
- The CampusIS uses VLAN tags extensively to identify different types of internal applications (such as VoIP or iSCSI traffic). There are over 108 VLANs that have been configured for different applications and services. However, they are not natively known to the VMs nor to the NetFlow data collected at the switches.
- The campus maintains information about switches ("SwitchConfig DB") and VLANs ("VLAN DB"), which is available through the tool `switchmap` [22]. On the other hand, information such as VLAN, available through `switchmap`, contains simply VLAN names as VLAN 1 and VLAN 2, but it does not identify the specific address associated with each VLAN or its semantic meaning. This then requires another separate table to relate the VLAN names to the VLAN addresses.

A. What Do We Want to Know?

As we can see from the above, different information is available at different repositories. Furthermore, certain information (such as the address associated with a VLAN) may change over time, which is not captured anywhere. This makes it difficult to analyze network conditions when processing historic NetFlow data. Similarly, the IP addresses associated with external applications (such as NetFlix, YouTube) are not limited to a single address for such applications due to multiple content distribution servers used by these services. In addition, the destination IP address list also changes with time due to changing content server locations. Thus, it is difficult to answer questions such as the following:

1) Which application flows are dominating a particular switch, and how much bandwidth are they consuming at a particular point in time?
2) How much bandwidth is used by campus users with a particular external application (such as watching NetFlix or NCAA Basketball final four tournaments) at a certain point in time?
3) How much bandwidth is used by an internal service such as iSCSI traffic due to redundancy of storage area networks located in separate physical locations?
4) How much traffic is incoming (or outgoing) to a VM at a data center and which part of the campus network are they coming from (or going to)?
5) What is the impact of configuration changes or VM migration on traffic flow through the network?

While the above queries can be currently answered with considerable effort, there has been a lack of an integrated approach to correlate data so that the campus network can be managed efficiently and such queries can be answered *quickly*. The goal of the I-CaN-MaMa (Integrated Campus Network Monitoring and Management) framework is to have a comprehensive view and a tool set that can help CampusIS sysadmin in monitoring and managing network resources from a single point of view.

IV. I-CaN-MaMa

We now present the Integrated Campus Network Monitoring and Management framework that we have been internally developed. The framework is shown in Fig. 3. Our framework is divided into three major components: 1) the campus network, 2) the data integration environment, and 3) the front end environment and tool set. Since the campus network has already been discussed, we focus next on the the second and the third components.

There were a number of design decisions made while conceiving this integrated environment. First, additional tools needed for I-CaN-MaMa should let the network operate independently and efficiently without being a hindrance. For instance, we want to minimize on data processing for our environment between the front end and data sources, which are again recorded as new flows at switches (by NetFlow) due to our distributed network environment. Secondly, the CampusIS currently uses certain useful tools such as `vSphere` or `mrtg`, which are not meant to be replaced by I-CaN-MaMa. They do have their place for the CampusIS sysadmin for specific situations. Furthermore, where possible, we keep the data at their source locations and access them on an as needed basis. Consider NetFlow data that is available from the core switches; these data run into multiple terabytes for historic collection stored at a SAN that is accessible by a dedicated VM for this purpose. Thus, we access them on an as (and when) needed basis.

A. Data Integration Environment and Knowledge Base

Our data integration environment focuses on creating a few key tables in the `PostgreSql` database to store a few key pieces of information on a periodic basis. We have created a number of tables. Collectively, we have built a KnowledgeBase through these tables.

The KnowledgeBase keeps track of relevant IP addresses including private and public IPs. The main knowledge table (see Table I) keeps track of common external applications such as YouTube, NetFlix, and their associated IP addresses. Note that while we show only one per entry in this example, we store all the addresses that we have been able to identify for an application through a number of means (such as a DNS query

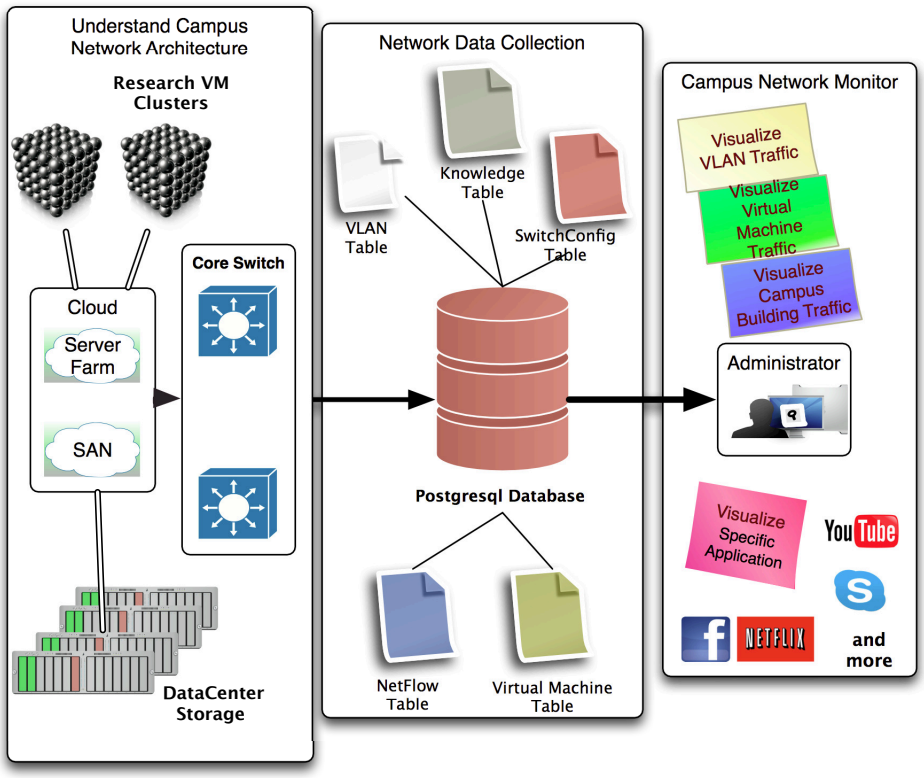


Fig. 3. I-CaN-MaMa Framework

TABLE I
APPLICATION DOMAIN TABLE

Domain	IPv4 Address
YouTube.com	206.111.x.y
www.Google.com	206.111.x.y
NetFlix.com	4.214.x.y

TABLE II
VMCONFIG TABLE

Server Name	IPv4 Address	Main Memory allocated
Server 1	10.10.x.y	2 GB
Server 2	192.168.x.y	4 GB
Server 3	134.193.x.y	8 GB

and the work done in [23]). Another knowledge table called the SwitchConfig table has the configuration information for the core switches, obtained from `switchmap` periodically.

The VM table in the KnowldegeBase stores the VM configuration in a temporal context by accessing the `vSphere` database on a periodic basis using the Powershell language. The VLAN table stores the mapping between VLAN names, their addresses, and to which applications and services they are assigned to—this is also associated with the timestamp information so that we can identify any changes over time (see Table III).

TABLE III
VLAN TABLE EXAMPLE

VLAN Name	IP Gateway	IP Subnet Mask	Subnet Bits
VLAN 1	10.10.x.1	255.255.255.0	24
VLAN 2	10.10.y.1	255.255.255.240	28
VLAN 3	10.10.z.1	255.255.240.0	20
...	...	...	...

B. Front End Environment

The front-end environment and tool set are for the CampusIS sysadmin to be able to query and display information about situations such as the ones listed in Section III-A. For this, we have developed a number of functionalities. The front end environment is designed to fetch data from databases storing the configuration and NetFlow data to create tables and visualizations based on user queries. Users can view real-time information and visualizations as well as those based on historical data.

The front end tool is built using the Ruby on Rails web application framework, a framework that is especially useful for applications that query databases. Additionally, a web application format is ideal for flexibility (such as cross-browser, cross-platform) and accessibility (accessible from any computer connected to the Internet). Underneath this environment, we are currently building several compute engines. Currently, the main engine implemented is the bandwidth

Algorithm 1: Bandwidth Calculator

Input: NetFlow Dataset $\mathcal{F}$, start and end time pair for each flow f record $T^f_{start,end}$, source and destination IP address set $IP^f_{src,dst}$ for flow f, VLAN IP addresses IP_{vlan}, Bandwidth rate R^f_{ip} of flows f, Application's IP addresses IP_{app}, Bandwidth check time window $\mathcal{T}$

Output: Total Bandwidth R_{total}, VLANs' Bandwidth R_{vlan}, application's Bandwidth R_{app}

```
for f ∈ F, (t_s, t_e) ∈ T^f_{start,end} , ip ∈ IP^f_{src,dst} do
    for t_c ∈ T do
        if t_s ≤ t_c ∧ t_c ≤ t_e then
            R_total += R^f_ip;
            if ip ∈ IP_vlan then
                R_vlan += R^f_ip ;
            if ip ∈ IP_app then
                R_app += R^f_ip ;
```

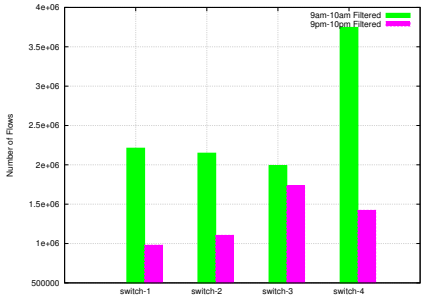


Fig. 4. Number of Flows

compute engine, which is briefly described here.

The bandwidth compute method is shown in Algorithm 1. In this method, we set up checkpoints for a selected time window for a query where the CampusIS sysadmin can choose an application, a set of applications, or a VLAN. We assign a time length for the checkpoint delimiter; currently, this is set to 1 sec but is configurable to any suitable value of interest. Each NetFlow record has the flow start and end time. If a flow goes across a check point, the average bandwidth for this flow, along with important network information, is stored and summed up in the appropriate buckets such as for a VLAN or an application, which is then associated with the KnowledgeBase.

V. Study Results

We present here three brief studies using the I-CaN-MaMa framework. For this analysis, we used two hours of NetFlow data (9AM-10AM and 9PM-10PM on 1 July 2013) collected in five minute intervals from each core switches in order to understand different impacts.

A. NetFlow Summary at Switches

The first study is on flow assessment at four core switches, labeled Switch 1 to Switch 4, in the campus network. Even though the first study does not need to rely on the information from the KnowledgeBase, it gives us important information regarding how many flows go through each core switch. When we calculate the flow count for each switch, we group flows into a filtered table if the flows' bps (bits per second) are not equal to 0. The reason for doing this is because flows with their bps equal to 0 has no meaning in bandwidth calculation. The total number of flows through each switch at two different time intervals is shown in Fig. 4. This information, along with bandwidth (discussed later), gives a good sense of cumulative flow through the network. In particular, it helps us identify most heavily used switches (in this case, switch 3 has the most amount of flows) and determine if some changes are necessary in the network configuration.

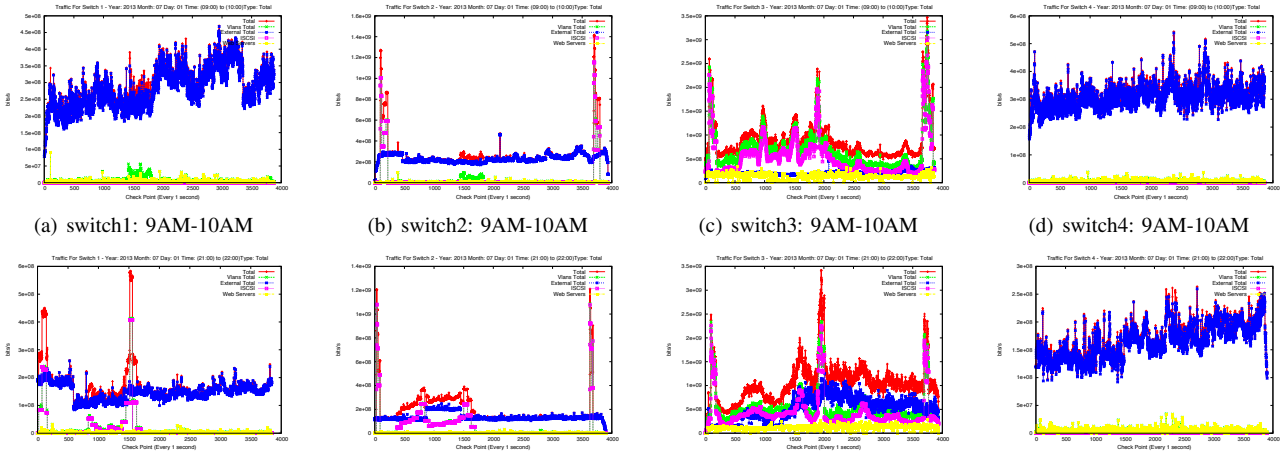
B. Bandwidth from the Switches' Perspective

Using the I-CaN-MaMa framework, we performed a study on bandwidth used by both internal and external applications. Fig. 5(a) to Fig. 5(h) show the bandwidth use at four core switches during two different time intervals. We observed the following:

- Switch 3 had the highest number of flows (Fig. 4). Not surprisingly, it also had the highest total bandwidth rate, which went up to 3.5 Gbps (at the peak during the study window), out of which 2.5 Gbps was for the VLAN traffic. The other three switches did not have such a high bandwidth rate pattern.
- Switch 4 had the hightest external bandwidth. There was no iSCSI traffic at Switch 4.
- Flow behavior at different time intervals for each switch was unpredictable. It reflected different application traffic at different time periods. For example, Fig. 5(a) shows that Switch 1 had very high iSCSI usage from 9:00AM to 10:00AM, but it is not much at 9:00PM-10:00PM (Fig. 5(e)). Switch 2 and Switch 4 showed a relatively steady traffic trend and Switch 4 had the most external traffic. By comparing Fig. 5(c) and Fig. 5(g), we see that, even though Switch 3 has the most dramatic network traffic movement, it had the most iSCSI traffic at all time.
- Switch 3 had the largest total of VLAN traffic usage.

By reviewing the campus network architecture, we found that most of the iSCSI traffic went through Switch 3. This is the reason for the largest traffic and the highest bandwidth at this node during the time windows. The campus data centers periodically did an iSCSI backup; one of the windows happened to be when a full backup was performed. Also, Switch 3 was like a root switch among all core switches inside of the campus network. Due to the specific design, many flows are routed through Switch 3 and then go to the destination that causes Switch 3 to have the highest number of flows at all times. Most of the campus servers, like web or the remove access VPN server, were located behind Switch 3, which also made Switch 3 the busiest switch. For example, whenever a student attempts to login to a campus account, it will cause at least one traffic flow through Switch 3.

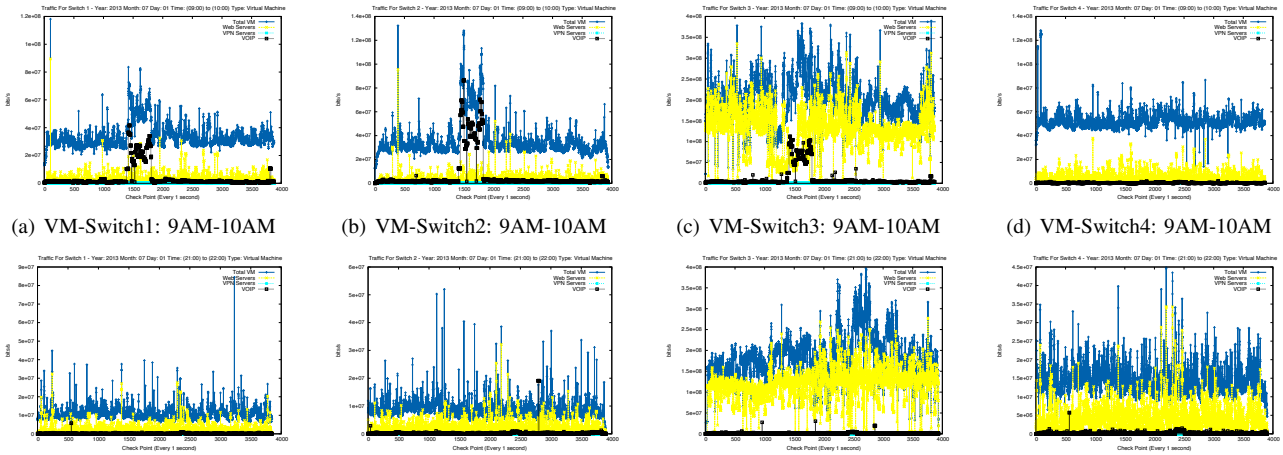
UMKC attempts to maintain campus bandwidth utilization under around 60 percent. Fig. 2 shows that iSCSI and Server Farm connectivity between Switch 3 had been configured at 20 Gbps bandwidth. The rest of the connectivities had been



(a) switch1: 9AM-10AM (b) switch2: 9AM-10AM (c) switch3: 9AM-10AM (d) switch4: 9AM-10AM

(e) switch1: 9PM-10PM (f) switch2: 9PM-10PM (g) switch3: 9PM-10PM (h) switch4: 9PM-10PM

Fig. 5. Bandwidth For Core Switches



(a) VM-Switch1: 9AM-10AM (b) VM-Switch2: 9AM-10AM (c) VM-Switch3: 9AM-10AM (d) VM-Switch4: 9AM-10AM

(e) VM-Switch1: 9PM-10PM (f) VM-Switch2: 9PM-10PM (g) VM-Switch3: 9PM-10PM (h) VM-Switch4: 9PM-10PM

Fig. 6. Bandwidth For Selected Virtual Machines

set to 1 Gbps. By reviewing our results, we can tell that the bandwidth during the time windows in our study was still under the desired level. The advantage with I-CaN-MaMa is that it allows us to *quickly* quantify how much was going through Switch 3 in near real time and whether we should do any route changes based on applications, or if the bandwidth rate goes beyond an acceptable threshold, or if we should change the time of the SAN backup (that generates iSCSI traffic) to another time window.

Switch 4 is located at the edge of the campus network. It connects the campus network to the public Internet. Thus, most of its traffic is due to external traffic.

C. Bandwidth from Virtual Machines' perspective

The third study we conducted was to monitor VM activities through the four core switches. The campus' main data center currently hosts nearly 400 VMs. We chose four VMs from the data center and analyzed the bandwidth rate for each of these (Fig. 6). We observed the following:

- Total VMs' usage at each switch varied at different time periods (see Fig. 6(a) to Fig. 6(h)). Selected VMs (such as the UMKC web server) had different traffic trend in each switch at selected time frame.
- All switches showed frequent traffic flows that connected to VMs hosting UMKC web servers. By taking a closer look at the VM traffic, we observed that Switch 3 had the highest web access traffic (see Fig. 6(c) and Fig. 6(g)), which went up to 400 Mbps at the peak time.

The VoIP and VPN servers did not show any significant traffic. It might be because there were limited activities (compared to the rest) during the time windows of our study.

VI. Summary and Future Work

In this paper, we have presented the I-CaN-MaMa framework. This framework allows us to have an integrated view of activities in a campus network that includes data centers. It also allows us to monitor and manage the network. Our approach allows the CampusIS sysadmin to react quickly to significant traffic changes in the network, which till now required a considerable effort to pinpoint. We are in the process of enhancing the environment by adding additional compute engines such as a latency calculator.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation Grant # CNS-0916505 and CNS-1217736.

REFERENCES

[1] "MRTG." [Online]. Available: http://oss.oetiker.ch/mrtg/

[2] [Online]. Available: www.cisco.com/go/netflow

[3] V. Sekar, M. K. Reiter, W. Willinger, H. Zhang, R. R. Kompella, and D. G. Andersen, "csamp: A system for network-wide flow monitoring." in *NSDI*, 2008, pp. 233–246.

[4] D. López, J. E. López-de Vergara, L. Bellido, and D. Fernández, "Monitoring an academic network with netflow," in *Proceedings of EUNICE*, 2004.

[5] R. d. O. Schmidt, A. Sperotto, R. Sadre, and A. Pras, "Towards bandwidth estimation using flow-level measurements," in *Dependable Networks and Services*. Springer, Towards2012, pp. 127–138.

[6] V. Mann, A. Vishnoi, and S. Bidkar, "Living on the edge: Monitoring network flows at the edge in cloud data centers," in *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*. IEEE, 2013, pp. 1–9.

[7] Y. Liu, J. Sun, R. Sun, and Y. Wen, "Next generation internet traffic monitoring system based on netflow," in *Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on*, vol. 1. IEEE, 2010, pp. 1006–1009.

[8] V. Krmicek, J. Vykopal, and R. Krejci, "Netflow based system for nat detection," in *Proceedings of the 5th international student workshop on Emerging networking experiments and technologies*. ACM, 2009, pp. 23–24.

[9] F. Dressler and G. Carle, "History-high speed network monitoring and analysis," in *Proceedings of 24th IEEE Conference on Computer Communications (IEEE INFOCOM 2005), Miami, FL, USA*, 2005.

[10] A. Bobyshev, P. DeMar, V. Grigaliunas, M. Grigoriev, and A. Fermilab, "Use of flow data for traffic analysis and network performance characterization." *CHEP07, Victoria BC, Canada*, pp. 2–7, 2007.

[11] D. Rossi and S. Valenti, "Fine-grained traffic classification with netflow data," in *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*. ACM, 2010, pp. 479–483.

[12] R. Hofstede, I. Drago, A. Sperotto, R. Sadre, and A. Pras, "Measurement artifacts in netflow data," in *Passive and Active Measurement*. Springer, 2013, pp. 1–10.

[13] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi, "Preserving the big picture: Visual network traffic analysis with tnv," in *Visualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on*. IEEE, 2005, pp. 47–54.

[14] T.-L. Pao and P.-W. Wang, "Netflow based intrusion detection system," in *Networking, Sensing and Control, 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 731–736.

[15] P. Calyam, D. Krymskiy, M. Sridharan, and P. Schopis, "Active and passive measurements on campus, regional and national network backbone paths," in *Computer Communications and Networks, 2005. ICCCN 2005. Proceedings. 14th International Conference on*. IEEE, 2005, pp. 537–542.

[16] [Online]. Available: http://nfdump.sourceforge.net

[17] [Online]. Available: http://nfsen.sourceforge.net

[18] [Online]. Available: http://www.plixer.com

[19] M. Fisk, S. A. Smith, P. Weber, S. Kothapally, and T. Caudell, "Immersive network monitoring," in *Proceedings of the 2003 Passive and Active Measurement Workshop*, 2003.

[20] J.-W. Choi and K.-H. Lee, "A web-based management system for network monitoring," in *IP Operations and Management, 2002 IEEE Workshop on*. IEEE, 2002, pp. 98–102.

[21] [Online]. Available: http://www.vmware.com/support.html

[22] [Online]. Available: http://switchmap.sourceforge.net

[23] P. Juluri, L. Plissonneau, Y. Zeng, and D. Medhi, "Viewing youtube from a metropolitan area: What do users accessing from residential isps experience?" in *Proc. of 13th IFIP/IEEE IM'2013)*. IFIP/IEEE, 2013, pp. 589–595.