Networking for the Cloud: Challenges and Trends

I. Drago, R. de O. Schmidt, R. Hofstede, A. Sperotto, M. Karimzadeh, B. R. Haverkort and A. Pras
Design and Analysis of Communication Systems
University of Twente, The Netherlands
{i.drago,r.schmidt,r.j.hofstede,a.sperotto,
m.karimzadeh,b.r.h.m.haverkort,a.pras}@utwente.nl

***Abstract*—Cloud services have changed the way computing power is delivered to customers, by offering computing and storage capacity in remote data centers on demand over the Internet. The success of the cloud model, however, has not come without challenges. Cloud providers have repeatedly been related to reports of major failures, including outages and performance degradation. The internal network of cloud data centers has frequently been identified as a root-cause of these problems, showing that network provisioning and monitoring is still a major challenge for the deployment of cloud services. This paper argues that today's technologies for measuring and monitoring Internet traffic could be applied in the context of the internal network of cloud data centers as well. To support that, we first show the suitability of flow-based traffic measurements for monitoring cloud services. Then, we present a case on bandwidth capacity provisioning to exemplify how flow-based measurements can be used to guarantee the performance of cloud services. Finally, we discuss future directions we believe will guide the development of new cloud services. We advocate that next generation cloud services will not only rely on the Internet as a means to reach users, but also influence how the Internet itself is organized. We illustrate this trend by describing our ongoing research on *mobile clouds*.**

I. INTRODUCTION

Cloud services have changed the way computing power is delivered to customers. Cloud services abstract away the complexity of system management, by offering computing and storage capacity in remote data centers on demand. In retrospective, this advent can be seen as a natural step in the evolution of the Internet [1]. The extreme growth of Web services popularity in the early 2000's led cloud providers, such as Amazon, Google and Microsoft, to invest both in data center provisioning for their own services and in the development of scalable software solutions [1]. Even though the later conversion of this infrastructure into a utility may have involved major technical challenges, the way for a new computing model was certainly starting to be paved.

It is not surprising that many companies are considering to migrate services to the cloud [14]. Outsourcing to the cloud is deemed advantageous given the gains obtained from the reduced costs, flexible provisioning and high scalability. However, this migration also has drawbacks. Cloud providers have been repeatedly related to reports of major failures [6]. Among the most common causes, network failures have been pointed as a recurring problem, *e.g.,* because of internal reconfigurations in cloud data center networks [6]. Addressing the problems in the cloud internal networks is, therefore, essential for the deployment of dependable cloud services.

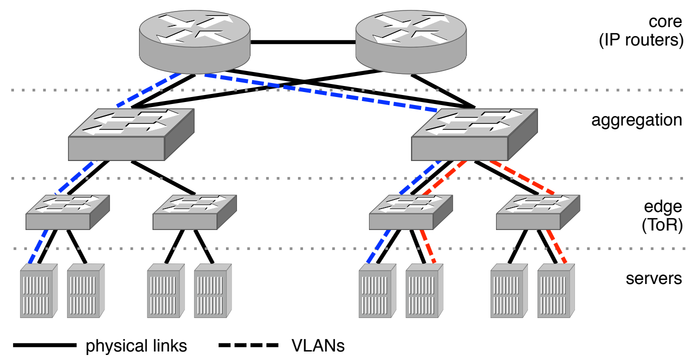


Fig. 1. An example of a generic data center topology.

It is known that some cloud providers (*e.g.,* Google) engineer their own networks starting from hardware components, making the data center application-specific. However, the tendency is to generalize data centers to well-known topologies, thus enabling the evolution of such infrastructure as the application mix changes [3]. Therefore, such a generic internal cloud network would have a topology similar to the one presented in Fig. 1. In this simplified 3-tiered topology, the edge switches – also known as *Top-of-Rack* (ToR) switches – interconnect servers that host the services. The ToR switches are interconnected by switches at the aggregation level, and these are interconnected by devices (*e.g.,* IP routers) in the core tier. The core tier also connects the data center to external networks (*e.g.,* the Internet). In the topology of Fig. 1, traffic between services running in the same rack goes via a two-hop path: from the server to the ToR switch and back. Communication between services running in different racks, however, may require up to six hops.

As one can see, internal networks of cloud data centers closely resemble those supporting the Internet itself. Therefore, we believe that widely-deployed technologies to measure and monitor Internet traffic could be brought to the context of internal cloud networks as well. This paper illustrates this point of view by discussing the use of *flow-based measurements* to monitor and provision networks for cloud services. Flow-based measurements are typically exported by network devices, such as routers and dedicated probes, using protocols such as Cisco NetFlow [4] or IPFIX [5]. We first review the suitability of such measurements for monitoring cloud services [17], describing common pitfalls that need to be avoided for successfully employing the flow data. Then, we show how flow measurements can be used to provision the capacity of network links [8] and discuss how these provisioning approaches could be applied to internal cloud networks, in order to guarantee the performance of cloud services.

Finally, we will discuss what we believe to be a future direction in the development of cloud services. Cloud services have succeeded partly because they rely on the Internet to reach a larger user base. The resulting gains of scale allow cloud providers to offer competitive services. In the current scenario, network operators act solely as enablers, providing the infrastructure to connect users to services. In the future, we foresee a paradigm shift, in which network operators will also be able to allocate computing and network capacity on demand relying on cloud services, according to the workload faced by the network.

The remainder of the paper is organized as follows. Section II introduce the fundamentals of cloud services, including a definition for the term as well as key characteristics and current usage of such services. Section III illustrates how customers are exposed to dependability challenges when migrating to the cloud, showing examples of both Service Level Agreements (SLAs) offered by major providers and recent cases of dependability problems. Section IV and V show that *flow measurements* can be applied in the context of cloud networks, by reviewing the suitability of flow data to monitor cloud services and by using flow data for the capacity provision of cloud networks, respectively. After that, we will introduce our vision of the future of clouds and discuss a motivating example, namely *mobile clouds*, in Section VI. Finally, Section VII concludes the paper.

II. Clouds nowadays

A. Definition and Key Characteristics

Cloud services have been interpreted in several manners. Terminology is many times fuzzy and, therefore, a consistent definition is needed. In this paper, we follow the conservative point of view of [1] and consider a cloud service to be any application that relies on utility computing to be delivered on demand over the Internet. In its turn, utility computing is the model of offering computing resources and charging customers based on the utilization [23].

Recent surveys, for example [1], have identified a multiplicity of aspects that characterize cloud services. Given the above definition, we consider the following five aspects to be key characteristics of cloud services [9]:

- **Shared resources or multi-tenancy**: resources are shared among several customers in a cloud environment. In contrast, customers of conventional data centers normally do not share the same pool of resources.
- **Scalability, elasticity or dynamic provisioning**: users can allocate resources on-the-fly, without providers' assistance. For example, in a cloud storage service customers can increase their storage space by requesting it from the pool of resources. Some authors refer to this property as "the appearance of infinity computing resources" [1].
- **Abstract infrastructure or virtualization**: cloud customers do not know the details of the infrastructure and systems providing the services, but instead, control them using well-defined interfaces. Note that abstraction and virtualization do not necessarily mean that virtual machines are in place (*e.g.,* as it is the case when a platform is offered as a service).
- **Pay-per-use or utility-based pricing**: although the units used to charge customers vary greatly, cloud services adopt the pricing model of a utility, with customers paying based on usage.
- **Connectivity, ubiquitous accesses or Internet centric**: by definition, cloud services are delivered via the Internet. As a consequence, private enterprise systems are not considered cloud services in this paper.

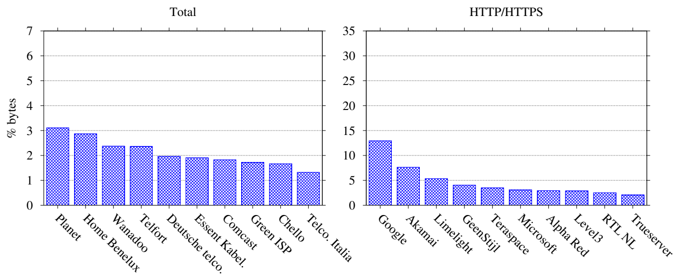
The importance of provisioning and monitoring networks in cloud data centers is clear from these characteristics. For example, multi-tenancy and the Internet centric nature of cloud services imply that the performance of a customer's application can be negatively impacted by the workload of other customers. Similarly, virtualization, elasticity and utility-based pricing imply that mechanisms *must* be in place to provision resources to services, in order to minimize customers' costs while delivering well-performing services.

B. Current Usage

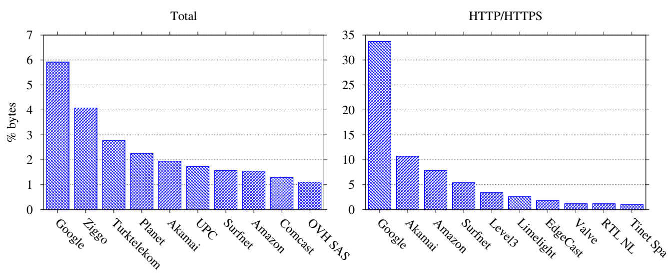
The success of the cloud service model is reflected by the increasing traffic to the biggest cloud providers. In a measurement study covering around 25% of the Internet interdomain traffic between 2007 and 2009, [18] showed that a very small number of networks is involved in most Internet traffic. Among more than 30,000 Autonomous Systems (AS), only 30 are responsible for around 30% of all inter-domain exchanges. The top 150 AS are already involved in more than 50% of the transfers. Major cloud providers are topping the list, with Google being responsible for around 5% of the traffic and others, like Microsoft and Akamai, among the ones with fastest growth.

Other works [12], [13] reported a similar strong concentration (in 2012) when measuring from edge networks, with up to 65% of the HTTP and HTTPS traffic going to the top 10 providers. Fig. 2 illustrates this trend using traffic flows that crossed the border routers of the University of Twente. The remote IP addresses of these flows were translated into IP owners using the MaxMind GeoIP Organization dataset[1] and the top organizations exchanging traffic with the university were calculated. Two datasets are plotted: Fig 2(a) shows the distribution of traffic among several Internet Service Providers (ISP) in Sept. 2008; Fig. 2(b) shows that four years latter (Oct.–Dec. 2012) the traffic at the university has become much more concentrated around a few remote organizations, including the ones offering cloud services, such as Google, Akamai and Amazon.

[1] http://www.maxmind.com



(a) 2 weeks in Sept 2008



(b) Oct–Dec 2012

Fig. 2. Top organizations exchanging traffic with the University of Twente.

III. Dependability of Cloud Services

In [2], dependability is defined as *the ability to deliver service that can justifiably be trusted* or, in another words, *the ability to avoid service failures that are more frequent and more severe than is acceptable*. Therefore, in this paper, we consider that a system is dependable when reliance can justifiably be placed on the services it delivers.

Cloud providers have been involved in numerous performance incidents. A recent survey of media articles, in [6], revealed evidence of 49 outages in 20 providers worldwide during the 6-year period ending in 2011. The causes are various, ranging from power outages to software updates. Given that this study has only considered events that received media attention, the frequency of such problems is likely to be much higher. Moreover, due to shared resources and multi-tenancy, these problems impact much more people than similar outages in private data centers.

Considering that the loss of turnover in case of failure can be remarkable, dependability has become a key issue for cloud providers. Despite modern networks are usually transparent to the end user, a poorly managed network can severely impact the performance of the service. For example, in data centers deadlines may be assigned to flows. If congested links are found at any tier of the cloud data center network, they may cause performance degradation and, consequently, flows deadlines may be missed. These problems directly affect data centers credibility and, ultimately, the performance perceived by the end users. A recent study in [16] showed that delays of 100 ms on flows completion time can cost Amazon 1% of its sales, and that an increase of 500 ms in search page generation time can drop Google's traffic by 20%.

TABLE I. Examples of SLAs for popular cloud providers.

Provider	Promise	Violation Policy
Amazon EC2	99.95 % availability	The service is unavailable if *all* customer's instances have no connectivity in more than one Availability Zone.
Google Apps	99.9 % uptime	Uptime is accounted in minutes per month. A service is down if it has 5 % of "user error rate" in 1 min.
Windows Azure	Per product	Each functionality has its own policy, with specific metrics.
Dropbox	Best effort	None

It is often assumed that customers are backed by SLAs. However, current SLAs of cloud services are weak at best and, in general, written to protect the providers. Table I gives examples of SLAs for popular cloud providers. This limited list shows how customers have very little protection when accepting the standard contracts of large cloud providers. The table shows that some cloud providers do not offer any guarantees. Others include terms to make it harder for customers to request refunds, for example. Amazon calculates violations on a monthly basis and only refunds a customer when (i) the customer has instances in more than one Availability Zone;[2] and (ii) all customer's instances in at least two Availability Zones have no external connectivity. Google has a similar strict policy, accounting downtime only if the service has more than 5% of "user error rate" (which is poorly defined in the contract) in a 1-minute interval. While these terms might not be a problem to individuals who use the cloud for non-critical tasks, enterprise customers need assurance before migrating any essential application or content. In the next section we argue how cloud providers can monitor their networks as a first step towards ensuring the performance of their cloud services.

IV. Traffic Monitoring & Measurements

Monitoring is of utmost importance for network operators to learn about the status of their network; not only for monitoring uptime and diagnosing problems, but also for monitoring all the aspects covered in SLAs. The widespread use of cloud services puts even higher constraints on uninterrupted availability of both the service as a whole, and the facilitating networks.

Key to network monitoring is performing measurements, such as the utilization of network links or the round-trip time between two devices. A widely used manner of monitoring traffic, especially in high-speed networks, is by means of passively measuring flows. In this method, traffic flows are exported, collected, and analyzed, rather than individual packets. A flow is defined in [5] as *a set of IP packets passing an observation point in the network during a certain time interval; All packets belonging to a particular flow have a set of common properties.* In more practical terms, flow data usually provides an overview of who communicated with whom, when, how (using which protocol), and how much (number of packets and bytes).

[2]Availability Zones are independent and physically isolated parts of the Amazon Web Services (AWS) infrastructure.

Fig. 3. Typical flow monitoring architecture.

A typical flow monitoring architecture is shown in Fig. 3. A *flow exporter*, which is often part of a packet forwarding device, exports flow data to a *flow collector* using technologies as Cisco NetFlow or IPFIX. It is the task of the *flow collector* to store and process this data, after which it can be made available to analysis *applications*. Besides providing great advantages in terms of processing requirements and low hardware costs, flow export technologies are widely available in packet forwarding devices such as routers and switches (as they may be used in data center networks – Fig. 1). This makes flow monitoring a relatively simple and cost-effective solution for large-scale monitoring of cloud networks.

Flow export technologies have received much attention mainly because of their major advantages in terms of scalability and wide availability in packet forwarding devices. As a result, a large spectrum of monitoring applications has been developed in the field of performance, security, and accounting, among others. However, our experience in the area of flow-based measurements also has shown that flow data should always be considered with great care, before deriving any (potentially misleading or even invalid) conclusions from it.

In the case of flow measurements, we have researched how implementation details and protocol design choices might lead to artifacts that affect the accuracy of monitoring data. For example, timing errors have been identified in NetFlow data that limit the accuracy of the data to the level of seconds (rather than the advertised millisecond accuracy) [20]. As a result, no meaningful conclusions can be derived about the response time of cloud services, which should normally be in the order of tens of milliseconds [10]. Also, some flow export devices were found to export no flag information about TCP flows [17]. In situations where the transport-layer is monitored, rather than the cloud service or the application itself, it is a common practice to rely on TCP flags for monitoring TCP connections. For example, many reset connections (*i.e.,* TCP flag RST) often indicate an overloaded application. When TCP flags are not present in the monitoring data, however, this cannot be observed by network managers and signals of performance problems might remain unseen.

Monitoring networks is a task that is far from trivial. However, once monitoring systems have been calibrated and set up properly, the monitoring data – flow data in our case – can be used for many more applications, such as intrusion detection [15] and link provisioning. The latter is discussed in the next section.

V. Link Capacity Provisioning

An important task of network operators is to properly provision their links such that QoS metrics defined by SLAs are met. In practice, the *rule-of-thumb* operators use to provision their links is based on 5 to 15-minute traffic averages obtained from SNMP counters. Clearly, such approach lacks accuracy since traffic averages over large periods completely overlook fluctuations that happen at shorter timescales (*e.g.,* 1 s or shorter). Although alternative provisioning approaches, such as our proposal in [19], are much more accurate because they can capture short-term traffic fluctuations, they often require continuous packet capturing. However, packet capturing is neither operationally nor financially scalable when considering high-speed links or large networking infrastructure.

Aiming at efficiency and practicality, recent work has proposed link provisioning approaches that use passive traffic measurements found at today's operators networks. These measurements are more scalable than continuous packet capturing and subtler than SNMP counters. For example, in [8] we propose a procedure for link provisioning using flow-level measurements, and in [7] we demonstrate the feasibility of using sFlow (packet sampling) to compute the required link capacity from sampled packets. These approaches are able to timely calculate accurate estimations of required capacity at timescales as low as 1 ms. The main goal of these two works is to inherit the accuracy from the provisioning approach from [19] while minimizing efforts on traffic measurements.

As mentioned in Section III, dependability of cloud services rely directly on the performance of the cloud data center network. Failure on completing flows within deadlines may strongly impact cloud providers credibility and revenue. Considering the generic topology presented in Fig. 1, devices misconfiguration or improper allocation of link resources may cause congested links in the interconnection between tiers. Therefore, provisioning approaches could be used to dimension links for the entire traffic aggregate or even to dimension capacity of Virtual LANs established to transfer traffic of specific (and priority) applications inside the data center.

Only few recent work addresses the problem of resource allocation and usage within data center networks. For example, [21] proposes a congestion control mechanism for data centers that focuses, among others, on high utilization of network links to maximize throughput of flows with deadlines. Briefly, this approach allocates network resources according to the rates of incoming flows. Hence, it behaves quite similarly to the above mentioned rules-of-thumb.

To allow for more flexibility on the allocation of network resources, efficient and practical provisioning approaches from, for example, our proposals [8], [7] could be brought into the context of data center networks. Furthermore, with the advent of Software Defined Networks (SDN) and the increasing adoption of tools such as OpenFlow, we envision that scalable per-flow-based traffic measurements will be feasible even at large network infrastructures. Initiatives, such as [22], show the viability of using SDN-based tools to monitor and measure network traffic. However, although promising these works still lack the accuracy of link provisioning approaches such as [19], [8], [7].

VI. Future of Clouds

Fig. 4 illustrates what we believe to be a direction to which clouds will evolve. Today's clouds depend on the

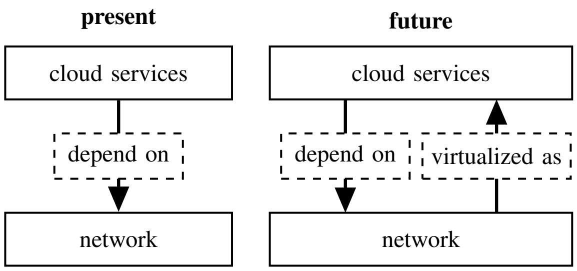


Fig. 4. Present and future relation between network and cloud operators.

infrastructure solely provided by network operators. However, due to the fast-growing popularity of virtualized services in clouds, we expect that in the future the network infrastructure itself will also be part of the application mix offered by cloud providers. This means that several network operators may have their infrastructure virtualized on top of a common underlying physical network. This will bring many advantages to the operators, such as improved elasticity and flexibility. However, we also foresee additional challenges on the control and management of virtualized networks in order to guarantee good performance of the whole system. In this respect we can point out SDN as a potential direction for cloud providers to monitor and manage the virtualized infrastructure. SDN enables the decoupling of control and data planes, facilitating the management of virtualized services on top of the physical networks, paving the way for the highly dynamic scenarios of future clouds.

There are many initiatives from academia and industry on the future of clouds, for example the EU FP7 Mobile Cloud Networking (MCN) project[3]. MCN conceptualizes that mobile network systems can be partially or fully virtualized in the cloud. That is, MCN proposes to join the two worlds of mobile cellular networks and clouds. The former exploits the virtualization capabilities from the latter to get additional on-demand applications and services to mobile customers without the need of setting up and operating additional physical infrastructure. In this view, as shown in Fig. 5, cloud services would span from isolated data centers to several network parts, such as core and radio access networks.

As an example of an MCN scenario, we can think of a situation in which a mobile operator (*e.g.,* KPN, AT&T, T-Mobile or Vodafone) needs to temporarily increase service capacity in a specific location due to an event that will result in an increase on the number of mobile users (*e.g.,* a festival). Such situations already happen, as reported in [11]. Today, the mobile operator's infrastructure need to be physically upgraded with new devices and connections to support the increasing number of users. In the future, however, this extra capacity can be provided by a virtualized infrastructure in the cloud.

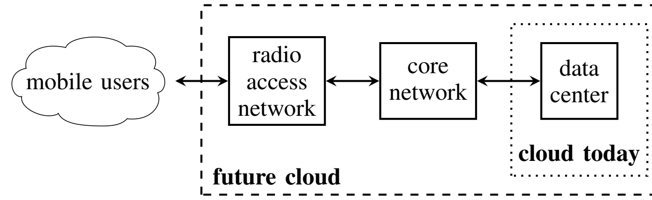


Fig. 5. A view of the future: mobile cellular networks and clouds.

VII. Final Considerations

Cloud services have gained a lot of popularity in recent years. Data centers behind the could services have inherited the same topology found in general purpose networks. We have discussed how cloud providers can take advantage of traffic monitoring technologies available in today's network devices, such as flow export technologies, to enhance performance of data center networks and, ultimately, improve dependability of the offered services. Furthermore, we have also discussed how cloud providers can profit from flow measurements to efficiently provision the internal network of cloud data centers.

We envision that future clouds will be comprised of even more complex services than today. Virtualized networks will be provided on demand, without the need for extending any physical infrastructure, for example, to mobile operators that need to dynamically increase their capacity. This will bring additional challenges to the control and to the management of the cloud infrastructure, so that performance is guaranteed and service dependability is achieved.

[3] http://www.mobile-cloud-networking.eu/

Acknowledgements

This work has been funded by the EU FP7 UniverSelf (#257513), the EU FP7 Mobile Cloud Networking (#318109), and EU FP7 Flamingo Network of Excellence (ICT-318488).

References

[1] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A View of Cloud Computing. *Communications of the ACM*, 53:50–58, 2010.

[2] Algirdas Avizienis, Jean-Claude Laprie, Brian Randell, and Carl E. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable Secure Computing*, 1(1):11–33, 2004.

[3] Theophilus Benson, Aditya Akella, and David A. Maltz. Network Traffic Characteristics of Data Centers in the Wild. In *Proceedings of the 10th ACM SIGCOMM Internet Measurement Conference*, IMC'10, pages 267–280, 2010.

[4] Benoit Claise. Cisco Systems NetFlow Services Export Version 9. RFC 3954 (Informational), 2004.

[5] Benoit Claise, Brian Trammell, and Paul Aitken. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information. RFC 7011 (Internet Standard), 2013.

[6] Roger Clarke. How Reliable is Cloudsourcing? A Review of Articles in the Technical Media 2005-11. *Computer Law & Security Review*, 28(1):90–95, 2012.

[7] Ricardo de O. Schmidt, Ramin Sadre, Anna Sperotto, and Aiko Pras. Lightweight Link Dimensioning using sFlow Sampling. In *Proceedings of the 9th International Conference on Network and Services Management*, CNSM'13, pages 152–155, 2013.

[8] Ricardo de O. Schmidt, Anna Sperotto, Ramin Sadre, and Aiko Pras. Towards Bandwidth Estimation using Flow-Level Measurements. In *Proceedings of the 6th IFIP WG 6.6 International Conference on Autonomous Infrastructure, Management, and Security*, AIMS'12, pages 127–138, 2012.

[9] Idilio Drago. *Understanding and Monitoring Cloud Services*. PhD thesis, University of Twente, 2013.

[10] Idilio Drago, Rick Hofstede, Ramin Sadre, Anna Sperotto, and Aiko Pras. Measuring Cloud Service Health using NetFlow/IPFIX: The WikiLeaks Case. *Journal of Network and Systems Management*, 2013. Online First Articles. http://dx.doi.org/10.1007/s10922-013-9278-0. Accessed Jun 2013.

[11] Jeffrey Erman and Kadangode K. Ramakrishnan. Understanding the Super-Sized Traffic of the Super Bowl. In *Proceedings of the 13th ACM SIGCOMM Internet Measurement Conference*, IMC'13, pages 1–7, 2013.

[12] Alessandro Finamore, Vinicius Gehlen, Marco Mellia, Maurizio M. Munafò, and Saverio Nicolini. The Need for an Intelligent Measurement Plane: The Example of Time-Variant CDN Policies. In *Proceedings of 15th International Telecommunications Network Strategy and Planning Symposium*, NETWORKS'12, pages 1–6, 2012.

[13] Vinicius Gehlen, Alessandro Finamore, Marco Mellia, and Maurizio M. Munafò. Uncovering the Big Players of the Web. In *Proceedings of the 4th International Conference on Traffic Monitoring and Analysis*, TMA'12, pages 15–28, 2012.

[14] Mohammad Hajjat, Xin Sun, Yu-Wei Eric Sung, David Maltz, Sanjay Rao, Kunwadee Sripanidkulchai, and Mohit Tawarmalani. Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud. *ACM SIGCOMM Computer Communication Review*, 40(4):243–254, 2010.

[15] Laurens Hellemons, Luuk Hendriks, Rick Hofstede, Anna Sperotto, Ramin Sadre, and Aiko Pras. SSHCure: A Flow-Based SSH Intrusion Detection System. In *Dependable Networks and Services. Proceedings of the 6th International Conference on Autonomous Infrastructure, Management and Security, AIMS'12*, volume 7279 of *Lecture Notes in Computer Science*, pages 86–97. Springer Berlin Heidelberg, 2012.

[16] Todd Hoff. Latency is Everywhere and It Costs You Sales – How to Crush It. http://highscalability.com/latency-everywhere-and-it-costs-you-sales-how-crush-it. Online. Accessed Oct 2013.

[17] Rick Hofstede, Idilio Drago, Anna Sperotto, Ramin Sadre, and Aiko Pras. Measurement Artifacts in NetFlow Data. In *Proceedings of the 14th International Conference on Passive and Active Measurement*, PAM'13, pages 1–10, 2013.

[18] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. Internet Inter-Domain Traffic. In *Proceedings of the ACM SIGCOMM 2010 Conference*, SIGCOMM'10, pages 75–86, 2010.

[19] Aiko Pras, Lambert Nieuwenhuis, Remco van de Meent, and Michel Mandjes. Dimensioning Network Links: A New Look at Equivalent Bandwidth. *IEEE Network*, 23(2):5–10, 2009.

[20] Brian Trammell, Bernhard Tellenbach, Dominik Schatzmann, and Martin Burkhart. Peeling away Timing Error in NetFlow Data. In *Proceedings of the 12th International Conference on Passive and Active Measurement*, PAM'11, pages 194–203, 2011.

[21] Christo Wilson, Hitesh Ballani, Thomas Karagiannis, and Ant Rowtron. Better Never than Late: Meeting Deadlines in Datacenter Networks. In *Proceedings of the ACM SIGCOMM Conference*, SIGCOMM'11, pages 50–61, 2011.

[22] Curtis Yu, Cristian Lumezanu, Yueping Zhang, Vishal Singh, Guofei Jiang, and Harsha V. Madhyastha. FlowSense: Monitoring Network Utilization with Zero Measurement Cost. In *Proceedings of the 14th International Conference on Passive and Active Measurement*, PAM'13, pages 31–41, 2013.

[23] Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud Computing: State-of-the-Art and Research Challenges. *Journal of Internet Services and Applications*, 1:7–18, 2010.