Information Technology

Recent Changes in the Internet Landscape

Prof. Dr. Steve Uhlig: Queen Mary University of London, Mile End Road, E14NS London, UK
Tel: +44-20-78825823, Fax: +44-20-78827997, E-Mail: steve@eecs.qmul.ac.uk
Steve Uhlig is the Professor of Networks and Head of the Networks Research group at Queen Mary University of London, UK. He obtained a Ph.D. degree in Applied Sciences from the University of Louvain, Belgium, in 2004. From 2004 to 2006, he was a Postdoctoral Fellow of the Belgian National Fund for Scientific Research (F.N.R.S.). His thesis won the annual IBM Belgium/F.N.R.S. Computer Science Prize 2005. Between 2004 and 2006, he was a visiting scientist at Intel Research Cambridge, UK, and at the Applied Mathematics Department of University of Adelaide, Australia. Between 2006 and 2008, he was with Delft University of Technology, the Netherlands. Prior to joining Queen Mary, he was a Senior Research Scientist with Technische Universitt Berlin/Deutsche Telekom Laboratories, Berlin, Germany. His research interests include on the large-scale behavior of the Internet, especially through network measurements, as well as software-defined networking and content delivery.

Dr. Felix Cuadrado Latasa: Queen Mary University of London, Mile End Road, E14NS London, UK
Tel: +44-20-78825819, Fax: +44-20-78827997, E-Mail: felix@eecs.qmul.ac.uk
Félix Cuadrado is a lecturer (assistant professor) in the School of Electronic Engineering and Computer Science, Queen Mary University of London. He received his degree in telecommunications engineering from Universidad Politécnica de Madrid (UPM), Spain, in 2005, and his Ph.D. degree from UPM in 2009. He received an award to the best PhD thesis in telecommunications in 2011 by the Spanish Engineering Association. His research interests include autonomic computing, services engineering, cloud computing and big data processing.

Keywords: Internet, content delivery, IXP, SDN, Cloud

Schlagworte: keywords in german

MS-ID: Heft: 53/* (2012)	steve@eecs.qmul.ac.uk	December 10, 2012

Abstract

The Internet is in a state of constant flux. This property of the Internet, that stems from the pressure from its continuously reinvented usage, makes it a lively environment, which must be re-assessed fundamentally from time to time. In this article, we review recent trends that have radically changed our perception of what the Internet is.

Popular applications such as Youtube and Facebook have demonstrated the massive scale at which services can be delivered across the Internet. This has been possible only thanks to the massive deployment in content delivery infrastructures, leading to the rise of the Cloud paradigm.

Recent studies from the research community have observed the consequences of this growth in the Internet infrastructure, through structural changes in the Internet topology and traffic patterns. The increasingly dynamic nature of Internet traffic patterns asks for improved capabilities to engineer traffic, that require a completely different level of programmability from the network. Software-defined networking is expected to fill this gap. Overall, the picture of the Internet that we draw in this article is the one of a very dynamic, challenging, and exciting playground.

Zusammenfassung

abstract in german

1 The Changing Internet Ecosystem

Today's Internet [1] differs significantly from the one observed a decade ago. The early commercial Internet had a strongly hierarchical structure, with large transit Internet Service Providers (ISPs) providing global connectivity to a multitude of national and regional ISPs [18].

During the early times of the commercial Internet (mid-90's), most of the content was delivered by client-server applications that were largely centralized. At the time, content was coming mainly in the form of websites hosted in enterprise data-centers. The tremendous growth of the World Wide Web, video streaming, and user-generated content has transformed hosting and content distribution into commodity services. A large fraction of today's content is hosted and delivered from data-centers and content distribution networks [1].

With the commodification of content hosting and delivery, the Internet traffic and topology landscape has been fundamentally reshaped. Today, a limited number of content contributors are responsible for a large fraction of the traffic. Indeed, recent work [1] has observed a consolidation of the traffic at the level of network organizations, as well as at the application-level. These consolidations took place together with changes in the interconnection structure between networks. Nowadays, large content contributors often have direct peerings with large ISPs or are even co-located within ISPs to deliver their content to the users. The Internet structure is not as strongly hierarchical as it used to be, and has been described as a flattening of its structure.

These changes in content delivery and Internet structure have serious implications on the extent to which we know the Internet, both its topology and traffic. The current mental model and map of the Internet misses fundamental aspects of todays Internet, e.g., the content delivery infrastructure [6] and Internet exchange points [23].

1.1 Content delivery

Driven by the tremendous demand from end-users for content, a diversity of hosting and content delivery infrastructures has emerged during the last years. These infrastructures have multiple choices on how and where to place their servers. As described by [7], the main aproaches are (1) centralized hosting, (2) data-center-based content distribution network (CDN), (3) cache-based CDNs, and (4) peer-to-peer (P2P) networks. Approaches 2 and 3 allow scaling content delivery by distributing the content onto a dedicated infrastructure. This infrastructure can be composed of a few large data-centers, a large number of caches, or any combination. In many cases, DNS is used by the infrastructure to select the server from which a user will obtain content [39, 14]. Approach 4, P2P, can be seen as a fully distributed way to deliver content. Despite the widespread popularity of P2P traffic in the early 2000, recent measurements have observed a decline of P2P traffic [1]. Still, in specific parts of the world P2P is still popular and infrastructures to distribute P2P content have been identified [36].

1.1.1 Massively distributed content infrastructures

To cope with the increasing demand for content, content delivery infrastructures deploy massively distributed server infrastructures [7] to replicate content and make it accessible from different locations in the Internet [15]. For example, Akamai operates more than 60,000 servers in more than 5,000 locations across nearly 1,000 networks [7]. Google is reported to operate tens of data-centers and front-end server clusters worldwide [4]. Microsoft has deployed its CDN infrastructure in 24 locations around the world. Amazon maintains at least 5 large data-centers and caches in at least 21 locations around the world. Limelight operates thousands of servers in more than 22 delivery centers and connects directly to more than 900 networks worldwide. As more and more infrastructures are being deployed in the Internet, and these serve an overwhelming fraction of the Internet traffic, it is imperative to take them into account to obtain a timely and accurate map of the Internet.

1.2 Cloud data centers

Today, we are living in the age of the cloud. Popular Internet-scale services are provided through cloud data centers, exploiting virtualization to provide scalability with the required level of agility. The use of a utility computing model has reduced costs thanks to the economies of scale, and has enabled a significant reduction in the deployment time of new services [8, 9]. The current model is primarily based on simple client-server interactions, with most applications hosted either at internal premises (private clouds), or at data centers from the main players that maximize the economies of scale (public clouds).

The current model has been a tremendous success for maximizing cost effectiveness, achieving agility in the deployment of innovative services, as well as elastically coping with the demand. However, a growing number of users want to access content with progressively higher expectations: Latency times must be minimized, services must be always available, with no expected down-time, and mobility must be seamlessly supported.

1.2.1 Warehouse-scale Computing

The last years have seen the consolidation of several major Internet services with millions of active users (e.g., Google services, Facebook, Amazon). Services delivery at this unprecedented scale presents a new wave of challenges, including the already mentioned content delivery. The tasks have mostly complicated because in order to process a single user request a massive amount of data needs to be processed. This phenomenon is known as the Big Data challenge, forcing storage and processing systems to adopt parallel computing techniques in order to scale up with the demand and perform their tasks at the required performance levels. This way, services are not delivered by single servers, but by large-scale data centers, in what is known as Warehouse-scale Computing [13]. The scale of these systems, both in user traffic and data center investment, have a significant impact in the overall Internet landscape.

1.2.2 Cloud as a Utility Challenges

A significant part of the cloud users landscape consists of service providers using the major public clouds for hosting their services. Acquiring computing and storage resources following a utility model brings significant advantages for these stakeholders. However, cloud platforms from different vendors have limited interoperability, putting service providers at risk of a vendor lock-in. Moreover, this greatly complicates relying on multiple providers simultaneously for resilience purposes.

Cloud computing platforms are often classified into multiple layers (IaaS - Infrastructure as a Service, PaaS - Platform as a Service,...), each offering different levels of abstraction to developers. Infrastructure clouds have the most homogeneous set of capabilities because of their lower level nature – they provide base computing, storage and networking services. However, up to this stage interoperability among the available solutions is severely limited. Standardization efforts cannot match the speed of industrial evolution, and de facto industrial standards are becoming the norm. The situation is even more complex with PaaS clouds. As they aim to provide higher-level APIs and interfaces to developers, their competitive advantage lies in the richness of the platform services, with differentiation instead of interoperability being the competitive advantage.

Cloud interoperability should enable these heterogeneous applications to be developed over high-level APIs and be deployed on a wide selection of platforms, managing their runtime dependencies without further adaptation/configuration effort from the application deployers. In order to achieve that, models that abstract from the differences between these platforms, and techniques for automatic adaptation among the different types of cloud platform of platforms need to be provided for achieving seamless interoperability and opening up all the marketplace resources [5].

Application data physical storage is another important driver for federated cloud applications, and stronger interoperability. There are numerous concerns, including reservations on cloud security, and regulations on data privacy that push for cloud applications to adopt a hybrid deployment. These problems are already one of the main research challenges in single provider cloud environments, and the increased number of stakeholders in the ecosystem exacerbates the problem, while also requiring compatible models and standards that allow to guarantee those requirements with multiple tenants hosting the data and functionality.

The shift of applications towards public clouds also has an impact on the network. The increased elasticity in computing and storage resources eases supporting the variations in client workload, but they also have to be supported by the network.

Additionally, the shift toward federated application deployments greatly increases the importance of interdatacenter communications, which can have a significant impact in the end user quality of service. As part of these challenges, there is increasing awareness for integrating resilience as an internal part of the service design and deployment.

In the case of Warehouse-scale services, some of the latest disclosed innovations orchestrate the access to multiple datacenters in order to provide real-time Internet scale query results [41], although in those cases inter-data center traffic is routed through the internal network.

1.3 Internet Exchange Points

The basic role of Internet eXchange Points (IXPs) dates back to the establishment of Network Access Points (NAPs) as part of the decommissioning of the National Science Foundation Network (NSFNET) around 1994/95, a carefully orchestrated plan for transitioning the NSFNET backbone service to private industry. The vehicle that evolved in support of this transition was a set of four NAPs (i.e., MAE-East, Sprint NAP, PacBell NAP, and Ameritech NAP) that acted as connection points for the commercial carriers that were vying for offering backbone services (e.g., MCInet, Sprintlink, AGIS) and ensured that the network would remain connected at the top level once the NSFNET was retired.

1.3.1 Internet infrastructure growth

Over the past 15 years, as the Internet grew by leaps and bounds by any imaginable metric, the original four NAPs were replaced by a steadily increasing number of modern IXPs. Originally providing largely just the bare necessities for supporting easy interconnection between their member ASes (e.g., physical space, caches, cabling, power, A/C, or secure access), IXPs themselves have evolved over time. Numbering now more than 300 worldwide, many of these IXPs are offering an array of different services that rely on advances in networking technology (e.g., VLANs or MPLS), exploit existing routing protocols in innovative ways (e.g., use of BGP for prefix-specific peering), or provide the economic incentives for an ever-increasing number of networks to join as paying members (e.g., remote peering offerings, support for IXP resellers).

1.3.2 IXPs: the missing piece?

In fact, large IXPs such as AMS-IX, situated in Amsterdam, and DE-CIX, in Frankfurt, offer high-end Service Level Agreements (SLAs) to their members that cover not only the initial provisioning and daily availability of a members port(s) but also the level of performance of key service parameters. Such innovation on parts of the IXPs has enabled them to compete more directly with the traditional carriers and has led to todays environment where some of the largest IXPs worldwide (e.g., AMS-IX, DE-CIX, LINX, MSK-IX) reportedly carry on a daily basis similar amounts of traffic as some large ISPs (e.g., AT&T, Deutsche Telekom). Some 300-500 networks, covering the whole spectrum of players in todays Internet marketplace, generate the traffic volumes at those IXPs. While there may be regional differences in how extensive in coverage or aggressive in the uptake of new members IXPs are, the critical role they have played in the Internet ecosystem has until recently gone largely unnoticed by the research community [23] whose focus has traditionally been on large carriers and large content [1].

A main finding of [23] is that the number of actual interconnections between networks at this single IXP exceeds 25% of the total number of interconnections in the entire Internet known as of 2010. Taking IXPs into account is necessary in order to obtain an accurate view of the Internet, and will require a significant amount of work to map this crucial piece of the Internet infrastructure.

2 Internet traffic

After more than a decade of work on understanding the properties of Internet traffic [40, 48, 24], there is a consensus that scaling models provide a concise and relevant description of traffic dynamics. However relevant, those models are limited by the fact that they describe traffic observed on an individual link, i.e., an access link connecting a University to its Internet provider or even a high-speed backbone network link. These works give us a view of Internet traffic purely focused on time dynamics, i.e., the traffic is seen as a mathematical function or a signal that varies over time.

Temporal models do not account for the correlations between different flows [44]. Gravity models, first used networking to model the volume of telephone calls in a network [3], go one step beyond and allow capturing spatial properties of traffic flows. More recently, most algorithms to estimate the volume of traffic between ingress-egress points inside large networks have incorporated some type of spatial model, such as the gravity models [45, 43, 46].

Unfortunately, existing network traffic models do not apply to the whole Internet [16]. Indeed, given the highly competitive nature of todays Internet ecosystem, network-wide traffic constitutes highly sensitive data that is rarely shared and studied by the research community. [1, 23] have helped better understand the current limitations to our knowledge of the traffic of the Internet. Despite existing efforts have moved towards large-scale characterization of specific services, e.g., Youtube [49], we are still far from a sufficient understanding of the global traffic pattern at this stage of the Internet lifetime.

2.1 Traditional traffic engineering

Traffic engineering consists of all the available techniques whose purpose is to directly or indirectly modify the behaviour of the traffic to achieve certain objectives. Traffic engineering has received a lot of attention during the last decade. Initially, traffic engineering was considered as a solution to allow large ISPs to optimize the utilization of their network. To achieve low network utilization, the traffic should be spread evenly among all the available links. Several techniques have been proposed to better spread the load throughout the entire network. One solution is to select appropriate link metrics based on a known traffic matrix [19]. Another solution is to rely on a connection-oriented technology [30, 25]. In this case, connections can be established statically or dynamically between distant routers and the layout of these connections can be optimized to achieve an even distribution of the traffic inside the network.

A significant fraction of the traffic engineering literature has focused on traffic engineering inside ISP networks, especially large ISP networks. However, most of the traffic in the Internet is exchanged between different networks [1]. The requirements for traffic engineering between networks are diverse and often motivated by the need to balance the traffic on links with other networks and to reduce the cost of carrying traffic on these links [22]. Networks that originate a lot of content, e.g., Google, connect directly with a large number of other networks [1], and need to optimize how content leaves their own network. Networks that provide Internet access to broadband or mobile users typically wish to optimize how the Internet traffic enters their network, as most users still download more than they upload content. In the middle, the transit networks will try to balance the load of the traffic exchanged between the networks they interconnect. Engineering traffic between networks relies on the specific constraints imposed on which traffic can flow on the connectivity between different networks, as well as on tweaking the routing protocol used to exchange reachability information between networks, BGP [22]. As different networks in the Internet have different economical models and technical constraints, traffic engineering between networks is complex and highly depends on the specifics of the objectives to be achieved.

2.2 Content-aware traffic engineering

Traditional traffic engineering is important to reduce the likelihood that bottlenecks arise due to a mismatch between the network provisioning and the expected demand. However, recent changes in the way popular applications deliver their content, through distributed infrastructures, have significantly limited the relevance of traditional traffic engineering techniques as a way to alleviate network bottlenecks. Traditional traffic engineering works on timescales of hours or more [20], as obtaining accurate demand estimates in large ISP networks is challenging [12]. Popular content on the other hand generates bursts in demand over much smaller timescales, e.g., minutes.

CDN-ISP collaboration Recently, CDN-ISP collaboration approaches [42, 11, 38] have been proposed to augment traditional traffic engineering techniques to help deal with the traffic generated by content delivery networks. Portals have been proposed to allow peer-to-peer applications and users communicate with ISPs and get an updated view of their networks [37]. Collaborative approaches require sensitive information to be exchanged between two parties and hands control over traffic engineering to the client side, which makes it hard to be adopted by ISPs.

3 Software-defined networking

The success of open-source routing software is being paralleled with increasing virtualization, not only on the server side, but also inside network devices. Server virtualization is now followed by network virtualization, which is made possible thanks to recent advances in software-defined networking (SDN), e.g., OpenFlow [34]. SDN exposes the data path logic to the outside world. The outside world is usually a network controller such as NOX [31]. Contrary to the incumbent practice of network devices controlled by proprietary software tied to specific hardware, SDN allows software running anywhere in the network to have direct access to the network equipment, e.g., to obtain traffic statistics [33, 35].

3.1 Middleboxes

Middleboxes are an interesting example of the potential impact of SDN. The last decade has seen a proliferation of middleboxes [10, 47], i.e., network devices that support specific aspects of the network protocols and applications. While considered detrimental [17], especially to transport protocols [28], middleboxes serve many important purposes in today's Internet, such as address translation (NAT), security, media gateways, proxies. Currently, most middleboxes are hardware appliances deployed across the network, e.g., in enterprise [29] and mobile networks [47], as well as datacenters [27]. The wide deployment of middleboxes is an opportunity to demonstrate the potential of SDN, e.g., through the consolidation of software-defined middleboxes [10], or by running them in the cloud [26].

3.2 Towards network programmability

More generally, SDN has the potential of fundamentally changing how specific network functionalities are implemented in the Internet, by not requiring deploying physical appliances dedicated to address a very specific functionality. One of the advantages that SDN brings is the flexibility in how functionalities can be distributed across the network. Indeed, through a SDN controller [31], it is possible to either directly to control SDN-enabled network equipment, or alternatively to have the traffic dynamically redirected to close-by equipment that is able to implement the required functionality.

4 Summary

In the main part of this article, we have reviewed fundamental changes in the Internet ecosystem. The popularity of rich media has pushed content delivery infrastructures and Cloud to become the hottest part of the Internet infrastructure. The scale at which planetary scale services need to be delivered is putting tremendous pressure on evolving the management of computing and storage warehouses. The network has also reacted to this, with the advent of Internet exchange points that have allowed high-speed local interconnection of data producer and consumer networks.

The second part of this article described the current situation in Internet traffic. Our current understanding of the flow of traffic in the Internet has become very limited, due to the sheer speed in the deployment of new services and computing and storage infrastructure. As a consequence, the current traffic engineering tools are mostly obsolete, and do not address the needs of network operators. One answer to this is to enable content aware traffic engineering, through collaboration between infrastructure providers.

The third part of this article introduces a new trend in the Internet: software-defined networking (SDN). SDN has the potential to improve the programmability of the network. This would enable a more flexible and dynamic management of the network, to support the sometimes stringent requirements of popular applications.

References

[1] C. Labovitz, S. Lekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, Internet Inter-Domain Traffic, in proc. of ACM SIGCOMM, 2010.

[2] L. Kleinrock and W. Naylor, On measured be-havior of the ARPA network, in proc. of the AFIS Na-tional Computer Conference, 1974.

[3] J. Kowalski and B. Warfield, Modeling traffic demand between nodes in a telecommunications network, in Proc. of ATNAC, 1995.

[4] R. Krishnan, H. Madhyastha, S. Srinivasan, S. Jain, A. Krishnamurthy, T. Anderson, and J. Gao, Moving Beyond End-to-end Path Information to Optimize CDN Performance, in Proc. of ACM Internet Measurement Conference, 2009.

[5] M.P. Papazoglou and W. van den Heuvel, "Blueprinting the Cloud," IEEE Internet Computing, 15(6):74-79, 2011.

[6] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig, Web content cartography, in Proc. of ACM SIGCOMM Internet Measurement Conference, 2011.

[7] T. Leighton, Improving Performance on the Internet, Commun. of the ACM, 52(2):44-51, 2009.

[8] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. of the ACM, 53(4):50-58, April 2010.

[9] L. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM Comput. Commun. Rev., January 2009.

[10] V. Sekar, N. Egi, S. Ratnasamy, M. Reiter, and G. Shi, The Design and Implementation of a Consolidated Middlebox Architecture, in proc. of USENIX/ACM NSDI, 2012.

[11] W. Jiang, R. Zhang-Shen, J. Rexford, and M. Chiang, Cooperative Content Distribution and Traffic Engineering in an ISP Network, in proc. of ACM SIGMETRICS, 2009.

[12] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True, Deriving traffic demands for operational IP networks: methodology and experience, IEEE/ACM Transactions on Networking, 9(3):265-280, 2001.

[13] L.A. Barroso, and U. Hölzle. The datacenter as a computer: An introduction to the design of warehouse-scale machines. Synthesis Lectures on Computer Architecture 4.1 (2009): 1-108.

[14] B. Ager, W. Mhlbauer, G. Smaragdakis, and S. Uhlig, Comparing DNS Resolvers in the Wild, in proc. of ACM Internet Measurement Conference, 2010.

[15] S. Triukose, Z. Al-Qudah, and M. Rabinovich, Content Delivery Networks: Protection or Threat?, in proc. of ESORICS, 2009.

[16] D. Alderson, H. Chang, M. Roughan, S. Uhlig, and W. Willinger, The many facets of Internet topology and traffic, Networks and Heterogeneous Media, 1(4): 569-600, 2006.

[17] M. Walfish, J. Stribling, M. Krohn, H. Balakrishnan, R. Morris, and S. Shenker, Middleboxes no longer considered harmful, in proc. of OSDI, 2004.

[18] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, Characterizing the Internet Hierarchy from Multiple Vantage Points, in Proc. of IEEE INFOCOM, 2002.

[19] B. Fortz and M. Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, in Proc. of IEEE INFOCOM, 2000.

[20] B. Fortz and M. Thorup, Optimizing OSPF/IS-IS Weights in a Changing World, IEEE Journal on Selected Areas in Communications, 20(4):756767, 2002.

[21] W. Fang and L. Peterson, Inter-AS traffic patterns and their implications, In Proc. of IEEE Global Internet Symposium, December 1999.

[22] B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure, Interdomain traffic engineering with BGP, IEEE Communications, 2003.

[23] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a Large European IXP. in Proc. of ACM SIGCOMM, Helsinki, Finland, August 2012.

[24] W. Willinger, V. Paxson, R. H. Riedi, and M. S. Taqqu. Long-Range Dependence and Data Network Traffic. In Long range Dependence: Theory and Applications", edited by Doukhan, Oppenheim and Taqqu, 2001.

[25] A. Elwalid, C. Jin, S. Low, and I. Widjaja, "MATE : MPLS adaptive traffic engineering," in proc. of IEEE INFOCOM, 2001.

[26] J. Sherry and S. Hasan and C. Scott and A. Krishnamurthy and S. Ratnasamy and V. Sekar, Making Middleboxes Someone Else's Problem: Network Processing as a Cloud Service, in proc. of ACM SIGCOMM, 2012.

[27] D. Joseph, A. Tavakoli, and I. Stoica, A policy-aware switching layer for data centers, in proc. of ACM SIGCOM, 2008.

[28] A. Medina, M. Allman, and S. Floyd, Measuring interactions between transport protocols and middleboxes, in proc. of ACM SIGCOMM Internet Measurement Conference, 2004.

[29] T. Benson, A. Akella, and A. Shaikh, Demystifying configuration challenges and trade-offs in network-based ISP services, in proc. of ACM SIGCOMM, 2011.

[30] Y. Wang, Z. Wang, and L. Zhang, Internet Traffic Engineering Without Full Mesh Overlaying, in proc. of IEEE INFOCOM, 2001.

[31] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, Nox: towards an operating system for networks, ACM Computer Communication Review, 2008.

[32] K. Claffy, H. Braun, and G. Polyzos, Traffic characteristics of the T1 NSFNET backbone, in proc. of IEEE INFOCOM, 1993.

[33] L. Jose, M. Yu, and J. Rexford, Online measurement of large traffic aggregates on commodity switches, in proc. of the USENIX HotICE workshop, 2011.

[34] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, OpenFlow: Enabling Innovation in Campus Networks, ACM Computer Communication Review, 2008.

[35] A. Tootoonchian, M. Ghobadi, and Y. Ganjali, OpenTM: traffic matrix estimator for openflow networks, in proc. of Passive and Active Measurement conference, 2010.

[36] R. Cuevas, M. Kryczka, A. Cuevas, S. Kaune, C. Guerrero, and R. Rejaie, Is content publishing in BitTorrent altruistic or profit-driven?, in proc. of ACM CoNEXT, 2010.

[37] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, P4P: Provider Portal for Applications, in proc. of ACM SIGCOMM, 2008.

[38] I. Poese, B. Frank, B. Ager, G. Smaragdakis, and A. Feldmann, Improving Content Delivery using Provider-aided Distance Information, in proc of ACM Internet Measurement Conference, 2010.

[39] B. Krishnamurthy, C. Wills, and Y. Zhang, On the Use and Performance of Content Distribution Networks, in proc. of ACM Internet Measurement Workshop, 2001.

[40] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, On the Self-similar Nature of Ethernet Traffic (Extended Version), IEEE/ACM Transactions on Networking, 1994.

[41] J.C. Corbett, J. Dean, M. Epstein, A. Fikes, C. Frost, J. J. Furman, S. Ghemawat et al. "Spanner: Googles Globally-Distributed Database." In proc of USENIX OSDI ,2012.

[42] D. DiPalantino and R. Johari, Traffic Engineering vs. Content Distribution: A Game-theoretic Perspective, in proc. of IEEE INFOCOM, 2009.

[43] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot, Traffic matrices: balancing measurements, inference and modeling, in proc. of ACM SIGMETRICS, 2005.

[44] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, Traffic matrix estimation: Existing techniques and new directions, in proc. of ACM SIGCOMM, 2002.

[45] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, Estimating point-to-point and point-to-multipoint traffic matrices: An information-theoretic approach, IEEE/ACM Transactions on Networking, 13(5):947960, 2005.

[46] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, Spatio-temporal compressive sensing and internet traffic matrices, in proc. of ACM SIGCOMM, 2009.

[47] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, An untold story of middleboxes in cellular networks, in proc. of ACM SIGCOMM, 2011.

[48] K. Thompson, G.Miller, and R.Wilder, Wide-area internet traffic patterns and characteristics, IEEE Network, 11(6):1023, 1997.

[49] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in proc. of ACM Internet Measurement Conference, 2007.