The Need for an Intelligent Measurement Plane: the Example of Time-Variant CDN Policies

The Need for an Intelligent Measurement Plane: the Example of Time-Variant CDN Policies

A. Finamore, V. Gehlen, M. Mellia, M. M. Munafò
Politecnico di Torino, Italy
Email: lastname@tlc.polito.it

S. Nicolini
NEC Laboratories Europe
Email: niccolini@nw.neclab.eu

***Abstract*—In this paper we characterize how web-based services are delivered by large organizations in today's Internet. Taking advantage of two week-long data sets separated in time by 10 months and reporting the web activity of more than 10,000 ADSL residential customers, we identify the services offered by large organizations like Google, Akamai and Amazon. We then compare the evolution of both policies used to serve requests, and the infrastructure they use to match the users' demand.**

Results depict an overcrowded scenario in constant evolution. Big-players are more and more responsible for the majority of the volume and a plethora of other organizations offering similar or more specific services through different CDNs and traffic policies. Unfortunately, no standard tools and methodologies are available to capture and expose the hidden properties of this in constant evolution picture. A deeper understanding of such dynamics is however fundamental to improve the performance of current and future Internet. To this extend, we claim the need for a Internet-wide, standard, flexible and intelligent measurement plane to be added to the current Internet infrastructure.

I. Introduction

In the past few years, the Internet has witnessed an explosion of cloud-based services and video streaming applications. Services such as Google Maps, Facebook or DropBox are used by millions of people every day. Similarly, video streaming services are very popular and account for the majority of the web traffic, with YouTube and Netflix leading the group [1]. To meet both scalability and availability requirements, these services rely on Content Delivery Networks (CDN) and cloud computing services such as Amazon EC2. Some recent studies focused on systems such as YouTube [2], [3] and Akamai [4], exposing some of the adopted internal mechanisms. Unfortunately, web-based services usually exploit different CDNs and are governed through secret algorithms and policies not easy to characterize. This is even more exacerbated by the fact that recently popular organizations like Twitter, Facebook, and Google have started adopting encryption (TLS/SSL) by default to deliver content to the users[1]. This trend is expected to gain more momentum in the next few years. If on one side this helps to protect end-users' privacy, it can be a big impediment for effective security and network management operations given the higher complexity of the traffic classification.

The combination of all these effects leads to a very "tangled" picture, overcrowded of services and organizations which serve them adopting mechanisms difficult to study and understand. While system policies (and their secrecy) are one of the key of success for company such as Google and Akamai, knowing these mechanisms is fundamental for operators willing to optimize their own network. In fact, given the rich set of services available, operators are facing questions such as (i) What are the services/applications contributing to the traffic mix on my network? (ii) How to guarantee performance to select services? (iii) Is there any advantage in deploying a CDN caching node in my network? At the same time, end-users are interested in understanding which services offer the best Quality of Experience (QoE): (i) Which is the best performing Video on Demand (VoD) service? (ii) Is Dropbox more reliable than GoogleDrive?

Given the complexiy of today's Internet, answering those questions is far from being trivial. Indeed, there are no comprehensive solutions available that can offer visibility to what is happening in the network. In this paper, we aim at showing how complex the picture can be. In particular, taking advantage of large data sets of measurements collected from an European Internet Service Provider (ISP), we focus on web traffic (that is, on HTTP and HTTPS/SSL traffic), and we look at "big players", i.e., the top *organizations* serving the largest amount of traffic. This work is in the same spirit of [5]. However, in this paper we consider two data sets, each of one week-long and collected at a distance of 10 months. After overviewing on the top players and services, we consistently compare and quantify the traffic in the two data collections focusing on the evolution of both the systems architecture and performance. Despite being armed by advanced monitoring tools which expose a lot of valuable information [6], the picture results largely fuzzy, incomplete and in constant evolution. In particular, we observe that:

- The traffic handled by the big players is constantly increasing, with about 75% of web traffic being served by the top 15 organizations as of April 2012.
- HTTPS/SSL traffic is gaining momentum, with the percentage of encrypted flows that went from 6% to more than 20% in only one year.
- To meet the traffic demand faced by big players some new data centers have been added and the number of IP addresses contacted increased of 30-50% while the volume served by the top 5 /24 subnets decreased of 2-13%.

[1] http://googleblog.blogspot.it/2011/10/making-search-more-secure.html

TABLE I
WEB TRAFFIC DATA SETS.

Name	Volume (%)	Flows (%)	Clients (%)	Servers (%)
April-12	15.8 TB (58.7)	149.4M (51.3)	14,484 (98.2)	216 k (4.2)
June-11	10.7 TB (56.7)	92.3M (58.2)	11,784 (98.3)	189 k (7.1)

- Long-term shifts due to changes in the CDNs architecture (e.g., the activation of new data centres) can cause abrupt changes in the paths typically used to fetch some content. Similarly, some other policies cause traffic shifts on short-term scale in the order of hours.

The combination of these effects poses serious questions about the performance experienced by end-users. For operators, this is even more critical since they do not have control on these policies or, even worst, they are not aware of the effect these policies have on their network. Based on these observations, we advocate the creation of a standard, holistic solution to offer visibility on the Internet obscure dynamics: the introduction of a *measurement plane (mPlane)*, a measuring architecture which, from users' devices to networks core, allows to measure the traffic in a cooperative and flexible way.

II. DATA SET

The data sets considered in this work have been collected using Tstat [6], the Open Source packet sniffer developed in the last 10 years by the Telecommunication Network Group (TNG) at the Politecnico di Torino. Tstat rebuilds TCP connections by monitoring the traffic sent and received by hosts. The connections are then monitored to provide several types of Layer-4 statistics. Using Deep Packet Inspection (DPI) and statistical techniques [7], each connection is further classified based on which application has generated it.

In this work we focus on two data sets collected at the same Point of Presence (PoP) of an European ISP. A *probe* consisting of an high-end PC running Tstat has been installed in the PoP to monitor all the traffic generated and received by more than 11,000 residential ADSL subscribers. The two data sets considered correspond to the traffic of two different weeks, the first (June-11) starting from 12:00 AM of June 20th, 2011 and the second (April-12) starting from 12:00 AM of April 2nd, 2012. Both data sets are composed of TCP flow-level logs where each line reports the statistics of a different TCP connection and the columns detail the specific indexes measured[2].

In this paper we study web-based services. Thus, we focus only on HTTP and HTTPS/SSL traffic, referred as web traffic in the following. Tab. I summarizes the data sets size reporting their name, the volume of bytes and flows due to web traffic, the number of monitored subscribers and the number of distinct servers contacted during the two different weeks. To highlight the importance of web traffic, in brackets we report volumes as percentages with respect to the total traffic observed at the vantage point. Considering for example April-12, 15.8 TB are due to web traffic corresponding to 149.4 millions TCP connections. This traffic accounts for 58.7% of the total volume exchanged by monitored hosts, and 51.3% TCP connections. The remaining part of the traffic is due to other applications like email, chat, and, most of all, peer-to-peer (P2P) applications. The traffic has been generated by 14,484 subscribers which have contacted more than 216,000 web servers. Almost all the users have generated web traffic while the web servers represent only 4.2% of the IP addresses contacted given the presence of P2P traffic.

Comparing the two data sets, two considerations hold. First, the volume of web traffic increased from 56.7% to 58.7% during the 10 months. This is related mainly to the decrease of P2P traffic which however is still accounting for the majority of the non-web traffic. Second, the number of customers monitored by the probe increased. This is due to some modifications in the ISP network and the subscription of new customers. We argue this is not affecting the measurements reported in this work since all the users are located in the same geographical area.

[2]A description of all statistics is available from http://tstat.tlc.polito.it/measure.shtml

TABLE II
ORGANIZATION RANKING.

	%Bytes			%Flows		
Orgname	April-12	June-11	Diff	April-12	June-11	Diff
Google	32.52	22.73	9.78	14.20	12.74	1.46
Akamai	17.26	11.70	5.56	18.46	16.99	1.47
Level3	5.08	4.61	0.46	1.89	1.93	-0.04
Limelight	5.07	3.83	1.23	1.29	1.64	-0.35
Netload	4.32	2.74	1.58	0.04	0.01	0.03
Leaseweb	1.69	1.04	0.64	1.52	1.52	0.00
VideotimeSPA	1.68	0.53	1.15	0.49	0.40	0.10
Facebook	1.29	0.86	0.43	5.16	4.31	0.85
Amazon	1.12	0.63	0.49	3.09	3.99	-0.91
OVH	1.03	1.09	-0.06	1.23	0.73	0.49
Zynga	0.14	0.01	0.13	2.37	0.12	2.26
Edgecast	2.00	0.94	1.05	1.25	0.74	0.51
Webzilla	0.98	2.89	-1.91	0.23	0.30	-0.07
Megaupload	-	10.81	-10.81	-	0.24	-0.24
PSINET	-	3.16	-3.16	-	0.23	-0.23
Total	74.18	67.57	6.58	51.21	45.9	5.31

III. VOLUMES

We start our analysis identifying the most important organizations and comparing the ranking obtained from the two data sets. Table II reports the percentage of bytes and flows for both data sets sorted by the percentage of bytes according to April-12 data set. To better highlight differences between the two week-long data collections, we report also the difference between the shares of each organization. For example, Google accounted for 22.73% of the web traffic in June-11 and 32.52% in April-12 corresponding to an increase of 9.78%.

The table lists only 15 organizations which are further split in three groups. The group on the top corresponds to the 10 organizations consuming the majority of the bytes in April-12. With very few exceptions, this top 10 is the same in June-11. Nevertheless, we can notice some differences. First, all the organizations increased their volumes during the 10 months,

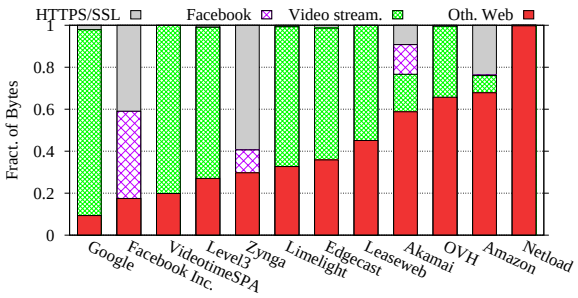


Fig. 1. Breakdown of the volume downloaded from each organization with respect to the type of content. Results refer to April-12.

TABLE III
FRACTION OF HTTPS/SSL TRAFFIC FOR EACH ORGANIZATION.

	%Bytes			%Flows		
Orgname	April-12	June-11	Diff	April-12	June-11	Diff
Zynga	59.35	8.95	50.40	43.27	5.18	38.08
Facebook	40.93	10.64	30.30	50.24	14.00	36.23
Amazon	23.73	6.58	17.15	15.70	1.61	14.09
Akamai	9.23	2.93	6.30	28.33	8.53	19.80
Google	2.11	1.01	1.10	18.56	8.56	10.00

with Google and Akamai having the largest variations both in bytes and flows. Second, while most of the organizations in the top 10 are well known, others are less popular such as VideotimeSPA, a company offering video streaming services.

The group at the bottom of the table shows some organizations having strong importance in June-11 but negligible volume in April-12. In particular, Megaupload was responsible for more than 10% of web traffic in June-11 but, after the service has been shutdown in January 2012[3], it completely disappeared. Similarly, PSINET and Webzilla, other two organizations offering file hosting services, have nearly vanished. Conversely, Netload, another file hosting service, was already prominent in June-11 and it gained 1.58% of share of volume possibly because of users' migration after the death of Megaupload. This shows that file hosting services continue to represent a significant portion of the web traffic even after the shutdown of Megaupload, and that they are still evolving.

The group in the middle of the table highlights two organizations that gained importance during the 10 recent months. Zynga, negligible in June-11, accounts for 2% of the flows in April-12. Notice that it is also the organization with the largest increase of share of flows. Instead, Edgecast is known to offer some video streaming services.

Considering the total volume, we can notice a tendency to concentrate more and more the volume in very few organizations. In particular, in June-11, 60.57% of the traffic was due to the top 10 (and Megaupload); the same set of organizations accounts for 71.06% in April-12. In both data set we found that the remaining volume is associated to more than 25,000 different organizations. This confirms results reported in [8], [5] and shows that, if it is true that big players have

[3]http://www.nytimes.com/2012/01/20/technology/indictment-charges-megaupload-site-with-piracy.html

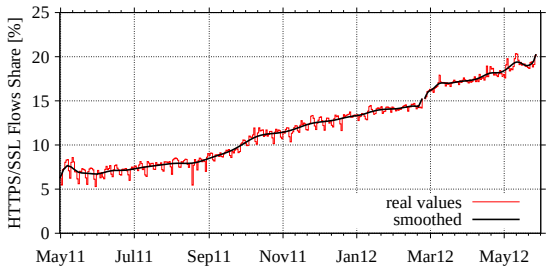


Fig. 2. Evolution of the percentage of HTTPS/SSL connections over 1 year.

a prominent role, a plethora of smaller organizations offer similar or more specific services. Thus, the global picture results overcrowded and difficult to understand.

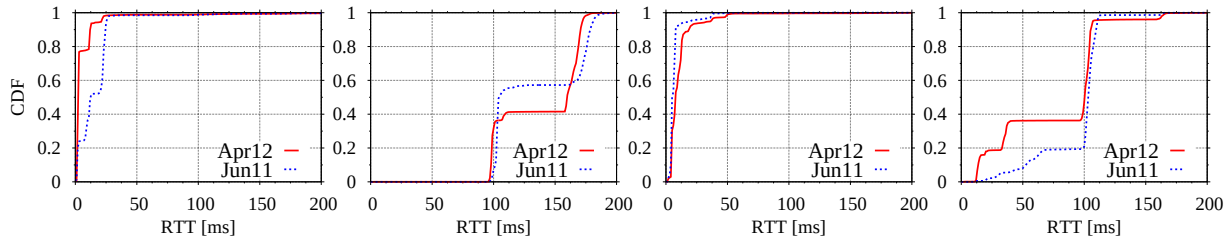
A. Type of services

Considering each of the top 12 organizations reported in Table II, Fig. 1 shows the breakdown of the volume they handle with respect to four classes of traffic obtained relying on the traffic classification capabilities of Tstat: HTTPS/SSL, Video streaming, Facebook and Other Web. The figure refers to April-12 and organizations are sorted by increasing fraction of Other Web. Almost all the organizations offer video streaming services. For Google, that is YouTube, and VideotimeSPA video content accounts for more than 80% of their volume. Instead, for other organizations such as Akamai and Amazon it tops 20%. This underline the big momentum of video streaming services. However, there are still some organizations for which the majority of the volume is related to other services. For example, Facebook and Zynga do not offer any video streaming service at all, while Netload serves some video content but only through file hosting (aggregated in Other Web in the figure).

B. Impact of HTTPS

As previously reported, some important organizations have started to adopt HTTPS/SSL as default protocol. Our measurements confirm this change. Fig. 1 shows that 40.9% and 59.3% of the volume of Facebook and Zynga respectively is due to HTTPS/SSL in April-12. We can also notice that HTTPS/SSL is prominent only for few organizations. Table III details the percentages of bytes and flows related to HTTPS/SSL in both data sets considering only the organizations having more than 1% of volume related to HTTPS/SSL. Comparing the two data sets, we can notice that all the organizations considered have increased their share of flows due to HTTPS/SSL by more than 10%. Zynga presents the largest changes and Facebook is the organization having the highest share of flows in April-12, with 50% of its connections due to HTTPS/SSL. Surprisingly, only 18.56% of the Google's connections run over HTTPS/SSL in April-12.

As to better highlight the impact of HTTPS/SSL in today's traffic, Fig. 2 reports the evolution of the percentage HTTPS/SSL connections during 1 year, starting from 1st May, 2011. Measurements are collected at the same vantage



(a) Google (b) Facebook (c) Akamai (d) Amazon

Fig. 3. Comparison of CDF of minimum RTT in June-11 and April-12 data sets.

TABLE IV
COMPARISON OF NUMBER OF IPS AND /24 SUBNETS.

(a) April-12

	IPs		SNet /24	
Orgname	No.	%	No.	Top5
Google	4666	2.16	150	92.43
Akamai	11106	5.15	1708	82.87
Amazon	10992	5.09	2481	60.02
Facebook	391	0.18	36	61.51

(b) June-11

	IPs		SNet /24	
Orgname	No.	%	No.	Top5
Google	2699	1.43	119	94.30
Akamai	6246	3.31	787	88.18
Amazon	7716	4.09	1373	50.52
Facebook	314	0.16	24	75.27

point of June-11 and April-12 and the figure reports both the actual values (average value per day) and the Bezier curve interpolation. Results are astonishing: the number of HTTPS/SSL connections increased of 4 times during the year and 20% of the web connections are due to HTTPS/SSL as of June 2012. This shows the on going tendency of "securing the web" as to increase the protection of end-users.

Despite Google claims that SSL is not computationally expensive anymore[4], the adoption of this protocol raises some questions. First of all, it reduces the visibility on the traffic so that improving the network security and traffic management is more complicated. Second, SSL can impact the perceived QoE given the higher latency required to complete the initial connection handshake.

IV. ORGANIZATIONS INFRASTRUCTURE

In this section we take a look at the infrastructure of the most popular organizations, namely Google, Akamai, Facebook and Amazon. We aim at giving an high-level overview obtained from users/ISP point of view rather then performing an in depth analysis of the whole system of these organizations. To this purpose, we use simple metrics as the number of servers contacted and the minimum RTT. The first gives a raw indication of the size of the organizations while the latter allows to identify the data centers position. We conclude the analysis studying the evolution of the bulk download rate as simple qualitative metric to express the performance.

A. Organization size

For the four considered organizations, Table IV reports the absolute number and the percentage of the IPs with respect to the total number of web servers contacted, the number of different /24 subnets related to these addresses, and the percentage of volume served by the top 5 subnets for each organization. Table IV(a) refers to April-12 while Table IV(b) refers to June-11. Considering for example April-12, 4,666 different Google servers have been contacted, corresponding to 2.16% of the whole set of web servers in the data set. These servers belong to 150 different /24 subnets, with the top 5 subnets responsible for 92.43% of the volume served by Google to the monitored ISP subscribers.

Comparing the number of servers, the four organizations considered have clearly different sizes. In particular, Facebook is the smallest one having only 391 IP addresses contacted, while more than 10,000 addresses have been contacted for Akamai and Amazon. As reported in Table II, the volume of Google doubles the volume of Akamai, but interestingly it is served by less than half of the addresses. This different "concentration" of volume is visible also comparing the volume served by the top 5 /24 subnets. In facts, the top 5 Google /24 subnets account for 92.43% of its volume while for Akamai the percentage is 82.87%. Amazon and Facebook present instead a lower concentration of volume in the top 5 /24 subnets possibly suggesting the adoption of different load balancing and caching policies.

Comparing the results between the data sets we can appreciate some differences. First of all, both the number of IP addresses contacted and /24 subnets increased in April-12. However, not all the organizations present the same growth. In particular, Google and Akamai have doubled the number of contacted addresses; for Facebook the variation is smaller. It is also interesting to notice that, for all the organizations but Amazon, the top 5 subnets are responsible for a lower fraction of volume in April-12 than in June-11. This shows that the users' requests are now served by an higher number of servers possibly because of modification in the organizations' architecture or in the policies controlling the traffic.

B. Data Center Location

Geo-locate an IP is not a trivial task. State of the art techniques [9] require complex operations of triangulation performing several RTT measurements from different points in the network. Therefore, we limit our investigation to a qualitative measure of the distance between the organization's servers and the vantage point. We rely on the minimum RTT observed on each connection defined as the minimum

[4]http://www.imperialviolet.org/2010/06/25/overclocking-ssl.html

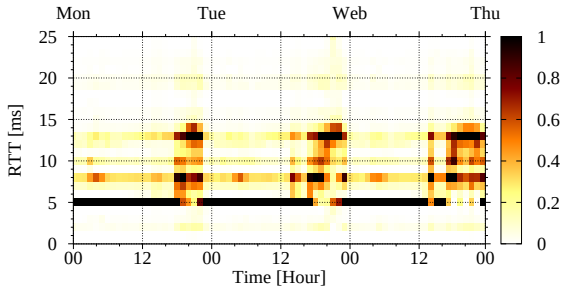


Fig. 4. Heatmap of the evolution of minimum RTT for Akamai during 3 days of April-12.

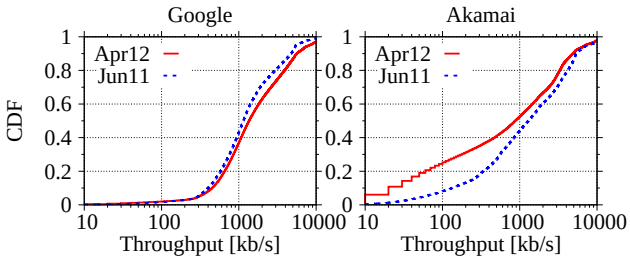


Fig. 5. Comparing the download rate of the flows with more than 1 MB in April-12 and June-11 data sets.

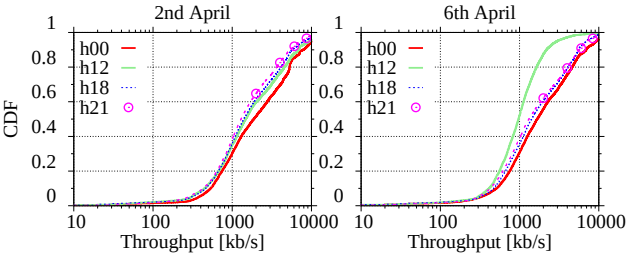


Fig. 6. Comparison of Google's throughput for flows with more than 10 MB over different days and hours in April-12.

time elapsed between a data packet sent by a client, and the reception of the associated TCP acknowledgement. Fig. 3 reports the Cumulative Distribution Function (CDF) of the minimum RTT for flows from Google, Akamai, Facebook and Amazon. Two line patterns are used to distinguish the data sets.

Steps in the CDFs suggest the presence of different data centers, which serve a fraction of the requests according to load balancing schemes. For example, in June-11 Google presents three data centers within 30 ms from the vantage point, i.e., all located in Europe. Facebook has only two data centers, both possibly in the U.S.

Comparing the data sets, we can see that there is no variation in the position of the data centers while the fraction of traffic each one handles has changed. In fact, 78% of the Google traffic is served by the closer data center in April-12 while only 22% was server by this data center in June-11. Conversely, for Facebook requests are more likely to be served by the further away data center than in April-12 than in June-11. Also for Akamai the distribution of the RTT is slightly increased but there are no strong variations neither in the servers location nor the traffic balance among the data centers. Amazon, as Google, presents an higher fraction of requests served by the closer data center in April-12 than in June-11. More in details, we can notice the presence of a knee around 20 ms in April-12 which was missing in June-11 indicating the activation of a new data center close to the vantage point.

C. Load balancing policies

Given that the Internet is in constant evolution, it is expected that the organizations update their infrastructure and the traffic policies as to cope with new requirements. This generates long-term variations as the one reported in Fig. 3. It is interesting to investigate if there are also variation during the day. To depict this, Fig. 4 reports an heatmap showing the variation of the minimum RTT over three consecutive days in April-12 for Akamai. To obtain the picture, for each hour, we created an histogram of the minimum RTT considering bins of 1 ms and we normalized the values by the maximum in each hour. The obtained fractions are then mapped to a color scale where darker colors correspond to higher fractions. As we can see, the figure reports a dark horizontal line around 5 ms corresponding to the closer data center to the vantage point reported in Fig 3(c). However, during the afternoon and the evening, the traffic presents some shifts indicating the presence of either some regular network congestion issues or, more likely, some load balancing policies which tend to serve requests from further away data centers.

It is important to underline that all these events are completely handled by each organization "behind the scenes" and both operators and end-users are not aware of them. If it is true that these policies are strictly related to internal optimizations of each organization, at the same time they might impact the performance experienced when using the services offered by these organizations. However, the lack of proper monitoring tools being able to depict these behaviours limit the validity of any QoS and QoE policy.

D. Performance

Fig. 5 shows the CDF of the throughput of connections downloading more than 1 MB from Google and Akamai. Different line patterns are used for the two data sets. We can see that the throughput obtained from the two data sets is very similar, even if Google presents an higher fraction of traffic served by the closer data center as seen in Fig. 3(a). Conversely, Akamai presents lower performance in April-12 even if the data centers location is unchanged during the 10 months. This shows that, even if it reasonable to expect worst performance when downloading a content from far away servers, the physical distance is not the only metric impacting the bulk download rate. Internal policies or unexpected events may impair the download rate. For example, in Fig. 6 we compare the CDF of the Google's throughput considering 4 hours in two different days in April-12. We can see that all the distributions overlap during the 2nd of April, showing no

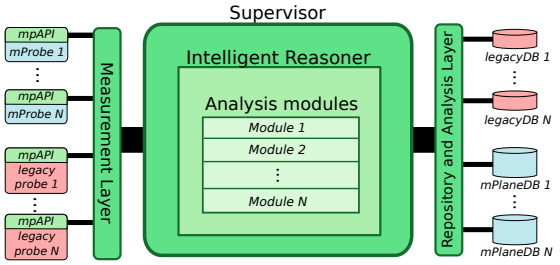


Fig. 7. Schematical representation of the mPlane architecture.

major changes during the daytime. Instead, on 6th of April some unexpected events happened so that the download rate during h12 significantly degrade (notice the logarithmic scale).

V. Revamping the State of Affairs: mPlane

All the examples reported in the previous sections show that several factors have to be considered when studying current web-based services. Take for example the case where a user is experiencing choppy playback when watching a YouTube video. Debugging such an issue is complex: is the users home router overrunning its buffer? Are the ISP network experiencing the same issue? Did the Content Delivery Network (CDN) correctly choose the right server for the users location in the network? These are just few possible problems on which investigate.

While results show that a general characterization is possible, there is a lack of standard and automatic methods to analyze and understand the implication of the traffic evolution. We thus advocate the introduction of a *measurement plane (mPlane)* that, next to the *data plane* (which carries packets) and the *control plane* (which manages the network) of the today's Internet architecture, allows a coherent and standard solution for measuring and monitoring the network status. We claim that only by disentangling the maze of complex relationships among layers and actors, we can understand the root causes behind availability and performance issues, and work to remedy them by enabling effective network management and operational procedures in today and future Internet.

The measurement plane must be flexible to naturally adapt to the Internet evolution, open to allow share of information, and intelligent to provide already processed data and not just raw measurements. To this extent, we envision that mPlane has to be composed by several elements as sketched in Fig. 7. The *measurement layer* combines a set of new (software and hardware) programmable mPlane probes *(mProbe)* with legacy probes adapted to the mPlane measurement layer interface *(mpAPI)* into a common, large-scale, distributed measurement infrastructure. The *repository and analysis layer* has to provides an efficient framework to store and process large volumes of data collected by the measurement layer leveraging on distributed parallel computing framework. The system is globally monitored by the *supervisor*, which controls the actions and synthesizes the results of the far-flung probes and repositories, and iterates on these results to drill down to the root cause of a specific issue and/or to investigate the relationship underlying a general phenomenon. This iterative analysis, supported and automated by an *intelligent reasoner* and several *analysis modules*, is sorely missing in present measurement systems, and is one key element of the mPlane.

The design, implementation, deployment and application of measurement plane will be the main objectives of the mPlane project, a 3 years long European project starting from November 2012.

VI. Conclusions

In this work we took advantage of two week-long data sets of web traffic and we overviewed the characteristics of the main organizations, showing their evolution over 10 months. The picture obtained is very complex. The majority of the volume is more and more concentrated in few big players while the remaining is related to thousands of other smaller organizations. The infrastructures are periodically updated so that long-term evolutions arise, for example because of the activation of new data centers or to cope with other needs. At the same time, short-term policies are used to impose a fine-grained control on the traffic during a single day.

If on the one hand, all these characteristics allow to have the huge variety of currect web-based services, on the other hand everything is under the control of these organizations. Given the lack of standard tools and methodologies, operators have to face hard times trying to understanding the tangled scenario of web-based services. At the same time, understanding the services performance is not trivial. As such, without a better knowledge on the hidden dynamics of the today's Internet, any QoS/QoE optimization is naive. We claim that there is a urge for a *measurement plane (mPlane)*, i.e., a standard, holistic solution capable of characterizing both the services and the network from the user to the backbone.

References

[1] Sandvine, "Global Internet Phenomena Report." http://www.sandvine.com/downloads/documents/Phenomena_1H_2012/Sandvine_Global_Internet_Phenomena_Report_1H_2012.pdf, 2012.

[2] S. Alcock and R. Nelson, "Application Flow Control in YouTube Video Streams," *SIGCOMM Comput. Commun. Rev.*, vol. 41, pp. 24–30, April 2011.

[3] R. Torres, A. Finamore, J. Kim, M. Mellia, M. Munafò, and S. Rao, "Dissecting Video Server Selection Strategies in the YouTube CDN," in *IEEE ICDCS*, 2011.

[4] E. Nygren, R. K. Sitaraman, and J. Sun, "The Akamai Network: a Platform for High-performance Internet Applications," *SIGOPS Oper. Syst. Rev.*, vol. 44, pp. 2–19, August 2010.

[5] V. Gehlen, A. Finamore, M. Mellia, and M. M. Munafò, "Uncovering the Big Players of the Web," in *TMA*, pp. 15–28, 2012.

[6] A. Finamore, M. Mellia, M. Meo, M. M. Munafò, and D. Rossi, "Experiences of Internet Traffic Monitoring with Tstat," *IEEE Network*, vol. 25, no. 3, pp. 8–14, 2011.

[7] A. Finamore, M. Mellia, M. Meo, and D. Rossi, "Kiss: Stochastic Packet Inspection Classifier for UDP Traffic," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1505–1015, 2010.

[8] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," in *Proceedings of the ACM SIGCOMM 2010 Conference*, (New York, NY, USA), pp. 75–86, ACM, 2010.

[9] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-based Geolocation of Internet Hosts," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1219–1232, 2006.