Characterizing Web Services Provisioning via CDNs: The Case of Facebook

Pierdomenico Fiadino, Alessandro D'Alconzo, and Pedro Casas
Telecommunications Research Center Vienna - FTW
surname@ftw.at

***Abstract*—Today's Internet consists of massive scale web services and Content Delivery Networks (CDNs). This paper sheds light on the way major Internet-scale web services content is hosted and delivered. By analyzing a full month of HTTP traffic traces collected at the mobile network of a major European ISP, we characterize the paradigmatic case of Facebook, considering not only the traffic flows but also the main organizations and CDNs providing them. Our study serves the main purpose of better understanding how major web services are provisioned in today's Internet, paying special attention to the temporal dynamics of the service delivery and the interplays between the involved hosting organizations. To the best of our knowledge, this is the first paper providing such an analysis in mobile networks.**

Keywords*—*Content Delivery Networks, HTTP traffic, Facebook, Akamai, Mobile Networks.

I. Introduction

A big share of today's Internet ecosystem is shaped by the success and influence of the most popular web services (e.g., video and audio streaming, social networking, on-line gaming, etc.). With more and more services being offered on the Internet, the network and server infrastructures are now extremely complex. The very last few years have seen an astonishing development in CDNs' technology, and today's Internet content is largely delivered by major CDNs like Akamai or Google CDN [1], [2]. In this evolved scenario, content and services are no longer located in centralized delivery platforms, owned by single organizations, but rather are distributed and replicated across the Internet and handled by multiple players. Understanding issues such as traffic delivery behavior, content location, hosting organizations, and addressing dynamics is highly valuable for network operators. Consequently, the study and characterization of the Internet traffic hosted and delivered by the top content providers has gained important momentum in the last few years [1]–[4]. In addition, several studies have addressed the characterization of large CDNs, focusing on the analysis of well known CDNs such as Google CDN, Akamai, and Limelight among others [5]–[7].

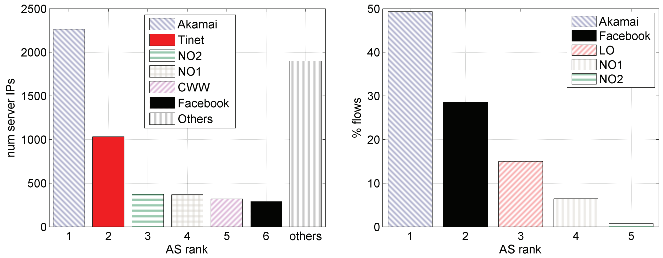
In this paper we focus on the characterization of the traffic and the delivery infrastructure of a highly popular web-service: the Facebook Online Social Network (OSN). Facebook is the most popular and widely spread OSN, with hundreds of millions of users worldwide sharing and accessing content on a daily basis[1]. Facebook content is mainly hosted by the well known Akamai CDN, which represents the most dynamic and widely deployed CDN today, with more than 137,000 servers in more than 85 countries across nearly 1,200 networks[2]. Facebook content is additionally hosted by the Facebook organization itself, with servers present both in the US and Europe. Finally, the intensive usage of transparent caching at the edge of ISP networks [3] and the deployment of large CDN caches inside the ISPs makes that a large fraction of the Facebook content accessed by the users is hosted at the premises of multiple network operators. Our study permits to better understand how the Facebook content is hosted and served by the aforementioned organizations. We show that the way these normally serve Facebook contents is very dynamic and complex to characterize, even revealing in some cases unexpected and interesting load balancing events.

The analyzed dataset corresponds to one month of HTTP flow traces collected at the 3G mobile network of a major European ISP by mid 2013. Traffic is captured at the well known Gn interface, and flows are analyzed through the stream data warehouse DBStream [8]. Only Facebook flows are kept for this study, using the HTTPTag traffic classification tool [9] to filter them out from the rest of the HTTP traffic. To preserve user privacy, any user related data (e.g., IMSI, MSISDN) are removed on-the-fly, whereas any payload content beyond HTTP headers is discarded. Using the origin (server) IP addresses of the flows, the dataset is complemented with the name of the organization hosting the content, extracted from the MaxMind GeoCity databases [11]. In the following analysis, dates are not disclosed and flow/volume counts are obfuscated (i.e., weighted by an unknown constant value) to preserve business privacy.

The remainder of the paper is organized as follows: Section II presents a thorough characterization of the Facebook traffic, both in terms of flow properties, as well as the underlying hosting/delivery infrastructure. Section III analyzes the temporal dynamics of the Facebook content delivery. In Section IV we further investigate the delivery of Facebook flows, particularly focusing on the interplays between the main hosting Autonomous Systems (ASes), additionally pointing out some unexpected events observed in our traces. Finally, Section V concludes this work.

The research leading to these results has received funding from the European Union under the FP7 Grant Agreement n. 318627, "mPlane".

[1]http://newsroom.fb.com/key-facts
[2]http://www.akamai.com/html/about/facts_figures.html



(a) Server IPs per AS. (b) Share of flows hosted per AS.

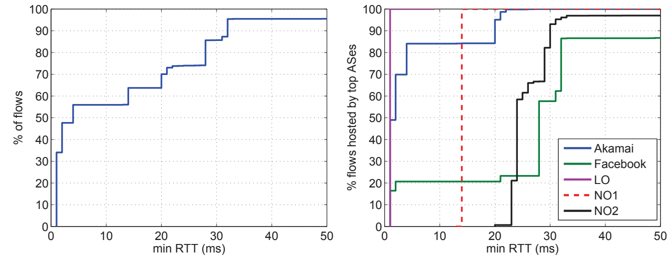
Figure 1. Unique server IPs used by the top hosting organizations/ASes and flow shares per hosting AS, considering the complete dataset. Akamai is clearly the key player in terms of Facebook content hosting.

II. Traffic and Content Delivery Infrastructure

We start by characterizing the Facebook traffic as seen in our traces, with a special focus on its underlying hosting/delivery infrastructure. Due to the high number of daily users and the high volumes of served traffic, Facebook follows a sophisticated content delivery strategy. Indeed, we observed more than 6500 server IPs hosting Facebook contents in our traces, distributed across 20 countries and more than 260 different ASes. This confirms the wide-spread presence of several organizations hosting Facebook contents, turning the service provisioning into a very tangled scenario. Figure 1 shows the main organizations/ASes hosting Facebook content, both in terms of number of unique server IPs observed and share of delivered flows. Akamai is clearly the key player in terms of Facebook content hosting, delivering almost 50% of the flows in our traces, using more than 2260 different server IPs. Interesting enough is the large number of server IPs observed from two organizations which actually deliver a negligible share of the flows: the Tiscali International Network (Tinet) and Cable & Wireless Worldwide (CWW). We believe these organizations are only caching spurious Facebook contents. In the remainder of the study we focus on the top 5 organizations/ASes in terms of served flows, depicted in figure 1(b): Akamai, Facebook AS, the Local Operator (LO) which hosts the vantage point, and two neighbor operators, Neighbor Operator 1 (NO1) and Neighbor Operator 2 (NO2).

A. *Geographical Diversity of Facebook Hosting Servers*

Table I provides an overview of the geographical diversity of the Facebook hosting infrastructure, listing the top countries where servers are located in terms of volume. The servers' location is extracted from the MaxMind GeoCity database, which is highly accurate at the country level [12]. "Europe (generic)" refers to a generic location within Europe for which MaxMind did not return a more accurate information. Almost 99% of the traffic comes from servers and data centers located in Europe, close to our vantage point, while only 1% of the traffic comes from other continents. This is due to three factors: (i) Akamai, the biggest Facebook content provider, has a very geographically distributed presence, pushing contents as close as possible to end-users [6]; (ii) operators heavily employ local content caching, and large CDNs like Akamai tend to deploy caches inside the ISPs networks, explaining the amount of traffic coming from the local country as well as neighboring countries to the vantage point; (iii) the rest of the traffic is handled directly by Facebook, which has servers split between Ireland (headquarter of Facebook International) and the US.



(a) Min RTT to all server IPs. (b) Min RTT to top hosting ASes.

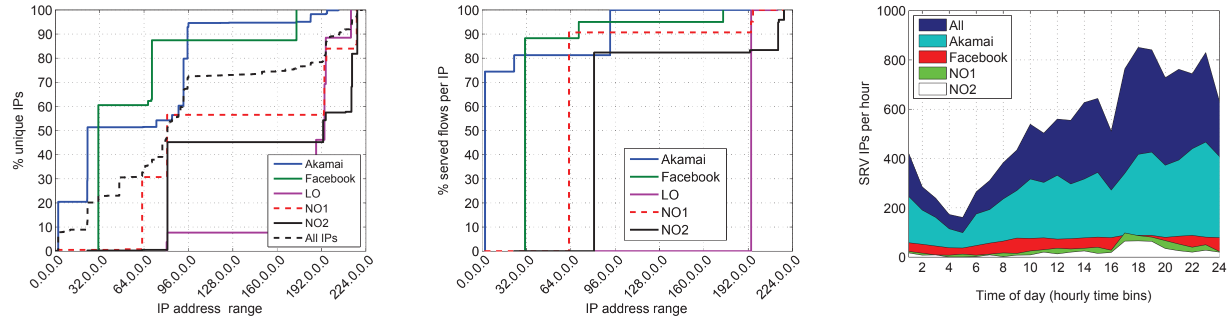
Figure 2. Distribution of overall min RTT and min RTT per top hosting ASes to server IPs, weighted by the number of flows hosted.

Country	% hosted volume
Europe (generic)	46.8%
Local country	37.2%
Ireland	12.7%
Neighbor country	2.1%
United States	1.1%
Unclassified	0.1%

Table I. Top Facebook hosting countries by volume.

To complement this picture, we investigate the location of these servers from a network topology perspective, considering the distance to the vantage point in terms of Round Trip Time (RTT). The RTT to any specific IP address consists of both the propagation delay and the processing delay, both at destination as well as at every intermediate node. Given a large number of RTT samples to a specific IP address, the minimum RTT values are an approximated measure of the propagation delay, which is directly related to the location of the underlying server. It follows immediately that IPs exposing similar min RTT are likely to be located at a similar distance from the source, whereas IPs with very different min RTTs are located in different places. RTT values are obtained from active measurements, performed during the span of the dataset, using a standard ping tool. Every unique IP hosting Facebook content is pinged with trains of 100 IMCP echo request packets every 10 minutes, resulting in a total of 6 individual values of min RTT per hour and per IP.

Figure 2 plots the cumulative distribution of the minimum RTT to (a) all the server IPs hosting Facebook, and (b) the aforementioned top orgs./ASes. Values are weighted by the number of flows hosted at each IP, to get a better picture of where the traffic is coming from. As a further confirmation of the geographical diversity, the distribution of min RTT presents some steps or "knees", suggesting the existence of different data centers and/or hosting locations. The largest majority of flows are served by close serves, located at less than 5 ms from the vantage point. As we mentionned, Akamai deploys its servers following the "enter deep into ISPs" approach [7], placing content distribution servers inside ISP POPs, which explains the short latency to the vantage point. The LO is the one with shortest delays for all the flows it serves and, along with the NO1, is the one with the least geographical diversity, with only one visible location. Three main steps appear in the CDF of the Facebook servers, which correspond to the headquarters in Ireland (min RTT about 30ms), the servers in the US (min RTT $>$ 100ms), and some servers located at only few milliseconds from the vantage point. Traceroutes to those



(a) IP range of unique server IPs. (b) Fraction of served flows per unique server IP. (c) # hourly unique server IPs per AS.

Figure 3. Distribution of the server IP range per AS. Akamai shows the most diverse IP range, but most of the flows hosted by Akamai come from a single subnet.

AS/Organization	# IPs	#/24	#/16
All	6551	891	498
Akamai	2264	132	48
Facebook AS	294	57	5
LO	26	8	6
NO1	368	26	14
NO2	374	33	9

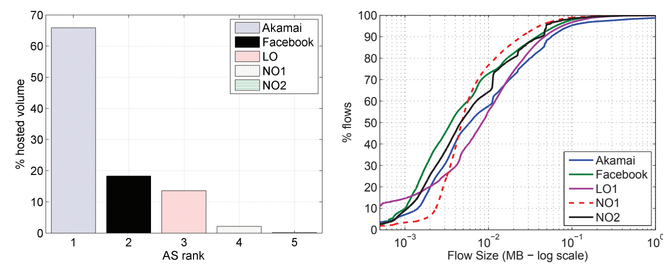
Table II. Number of IPs and blocks hosting Facebook.

servers revealed a direct connection to the Internet eXchange Point (IXP) of the local country, explaining the so low delays.

B. Facebook IP Address Space

Table II provides a summary on the number of unique server IPs observed in the traces, and the /24 and /16 IP blocks covered by the top orgs. hosting Facebook. Akamai and Facebook together account for about 2560 servers scattered around almost 200 /24 IP blocks, revealing again their massively distributed infrastructure. However, we shall see next that only a few of them are actually hosting the majority of the flows.

Figure 3 depicts the distribution of the IP address ranges associated to the top orgs. during the observation period, as well as the daily utilization of IPs per top org.. Figure 3(a) considers the distribution of IPs itself, whereas 3(b) weights each of the server IPs by the number of flows delivered. Despite the high number of /24 IP blocks, only few of them are responsible for the largest majority of the flows per org.. In particular, 75% of Akamai flows are served by only one single address range, covering a small number of /24 IP blocks. The same observation is valid for Facebook AS and the two Neighbor Operators, with 89%, 91% and 82% of their flows hosted at one single range respectively. Finally, the LO serves almost all the flows from a small range of IPs. Figure 3(c) shows the daily usage of these IPs on a single day, considering the number of unique server IPs per hour, per org. The number of active IPs (i.e., IPs serving flows in the corresponding time slot) used by Akamai follows the daily utilization of the network, peaking at the heavy-load time range. Interestingly, the IPs exposed by Facebook AS are constantly active and seem loosely correlated with the network usage. This comes from the fact that Facebook AS servers normally handle all the Facebook dynamic contents [2], which include the user sessions keep-alive.



(a) Shares of hosted volume per org./AS. (b) Distribution of flow sizes.

Figure 4. Hosted volume and distribution of flow sizes per organization.

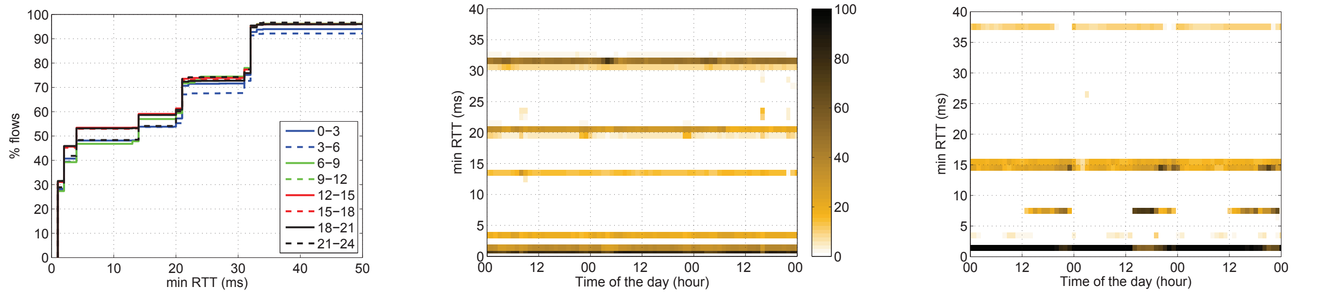
C. Facebook flow sizes

Figure 4 depicts the volume share of Facebook contents hosted by each org./AS, as well as the flow size distributions. Akamai hosts more than 65% of the total volume observed in our traces, followed by Facebook AS itself with about 19%. Comparing the volume shares in figure 4(a) with the flow shares in figure 1(b) evidences a clear distinction on the content sizes handled by both Akamai and Facebook AS: while Akamai hosts the bigger flows, Facebook AS serves only a small share of the service content. Indeed, as previously flagged by other studies [2], Akamai serves the static contents of the Facebook service (e.g., photos, songs, videos, etc.), whereas the Facebook AS covers almost exclusively the dynamic contents (e.g., chats, tags, session information, etc.).

To further explore this distinction, figure 4(b) reports the distribution of the flow sizes served per org.. The CDF reveals that Akamai clearly serves bigger flows than Facebook AS. The remaining ASes tend to host bigger flows than Facebook AS, which is coherent with the fact that ISPs caching is generally done for bigger objects, aiming at reduce the load on the core network.

III. Content Delivery Temporal Dynamics

The characterization performed in previous section only considers the static characteristics of Facebook during the complete duration of the dataset. In this section we focus on the temporal dynamics of the Facebook content delivery. To start with, we focus on the temporal evolution of the min RTT



(a) min RTT daily dynamics in 2013. (b) min RTT dynamics in 2013 (c) min RTT dynamics in 2012

Figure 5. Temporal variations of the min RTT to Facebook servers. The temporal patterns in 2012 show a strongly periodic load balancing cycle, focused in a small number of hosting regions. Results in 2013 suggest that Facebook content delivery is becoming more spread and load balancing cycles are less evident. In the heat maps of figures (a) and (c), the darker the color, the bigger the fraction of flows served from the corresponding min RTT value.

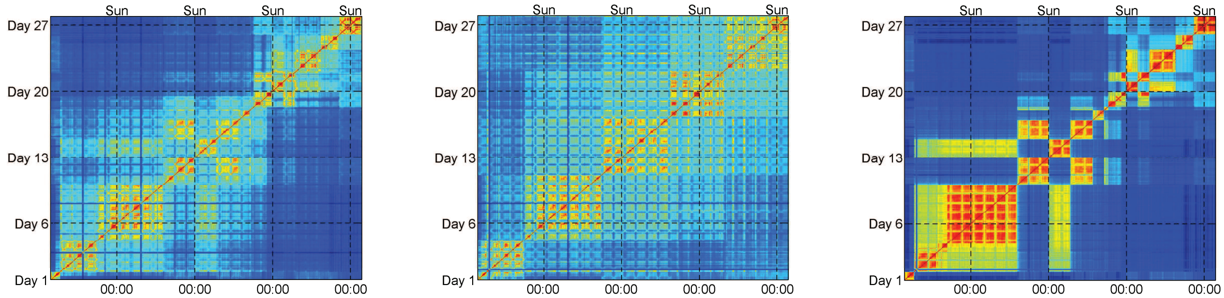
reported in figure 2. Figure 5(a) depicts the temporal variation of the CDF for all the flows and for a complete day, considering a single CDF every three hours period. The CDFs are rather stable during the day, but present some slight variations during the night and early morning. To get a better picture of such dynamics, figure 5(b) depicts the hourly evolution of the min RTT for all the Facebook flows during 3 consecutive days, being the first day the one analyzed in figure 5(a). Each column in the figure depicts the PDF of the min RTT for all the served flows, using a heat map-like plot (i.e., the darker the color, the more concentrated the PDF in that value). The flagged variations are observed during the first day, with some slight shifts between 6am and 12am from servers at 14ms and 20ms. The heat map also reveals some periodic flow shifts between 9pm and midnight from servers at 20ms, but impacting a small fraction of flows. Figure 5(c) presents the same type of heat map for Facebook flows, but considering a dataset of 2012 from the same vantage point [4]. The temporal patterns in 2012 show a much stronger periodic load balancing cycle, focused in a small number of hosting regions at 7ms, 14ms, and 37ms. Comparing the results from 2012 with those in 2013 suggests that Facebook content delivery is becoming more spread in terms of hosting locations, and load balancing cycles are becoming a-priori less marked. However, when deeply analyzing the complete dataset of 2013, conclusions are rather different.

To drill down deeply into this issue, we analyze the dynamics of the content delivery for the complete dataset, spanning 28 consecutive days. Instead of considering the variations of the min RTT, we consider now the variations on the number of flows served by the observed IPs. Changes in the distribution of the number of flows coming from the complete set of 6551 server IPs reflect variations in the way content is accessed and served from the hosting infrastructure observed in our traces. For this analysis, we consider a time granularity of one hour, and therefore compute the distribution of the number of flows provided per server IP in consecutive time slots of one hour, for the complete 28 days. This results in a time-series with a total of $24 \times 28 = 672$ consecutive distributions. To quantify how different are two distributions in the resulting time-series, we use a symmetric and normalized version of the Kullback-Leibler divergence described at [10].

To visualize the results of the comparison for the complete time span of 28 days, we use a graphical tool proposed in [10], referred to as *Temporal Similarity Plot* (TSP). The TSP allows pointing out the presence of temporal patterns and (ir)regularities in distribution time-series by graphical inspection. In a nutshell, a TSP is a symmetrical heatmap-like plot, in which the value $\{i, j\}$ reflects how similar are the two distributions at time t_i and t_j. We refer the interested reader to [10] for a detailed description of the TSP tool. Figure 6 gives an example of TSP for the distributions of all the Facebook flows across all the server IP addresses providing Facebook content, over the 28 days. Each plot is a matrix of 672×672 pixels; the color of each pixel $\{i, j\}$ shows how similar are the two distributions at times t_i and t_j: blue represents low similarity, whereas red corresponds to high similarity. By construction, the TSP is symmetric around the 45° diagonal, and it can be interpreted either by columns or by rows. For example, if we read the TSP by rows, for every value j in the y-axis, the points to the left [right] of the diagonal represent the degree of similarity to past [future] distributions.

The three TSPs in figure 6 represent the distribution variations for (a) all the observed IPs, (b) the Akamai IPs and (c) the Facebook AS IPs. Let us begin by the TSP for all the observed server IPs in figure 6(a). The regular "tile-wise" texture within periods of 24 hours evidences the presence of daily cycles, in which similar IPs are used to serve a similar number of flows. The lighter zones in these 24 hour periods correspond to the time of the day, whereas the dark blue zones correspond to the night-time periods when the traffic load is low. The low similarity (blue areas) at night (2am-5am) is caused by the low number of served flows, which induces larger statistical fluctuations in the computed distributions. This pattern repeats almost identical for few days, forming multiple macro-blocks around the main diagonal of size ranging from 2 up to 6 days. This suggests that during these periods, the same sets of IPs are used to deliver the flows, with slight variations during the night periods, similarly to what we observed in figure 5(a). However, the analysis of the entire month reveals the presence of a more complex temporal strategy in the (re)usage of the IP address space. For example, there is a reuse of (almost) the same address range between days 10-12 and days 15-16. Interestingly, we observe a sharp discontinuity on days 18-19, as from there on, all the pixels are blue (i.e., all the distributions are different from the past ones).

To get a better understanding of such behaviors, figures 6(b) and 6(c) split the analysis for Akamai and Facebook AS



(a) All IPs

(b) Akamai

(c) Facebook AS

Figure 6. TSP of hourly flow count distributions over 28 days for all the observed IPs, Akamai IPs, and Facebook AS IPs. A blue pixel at $\{i, j\}$ means that the distributions at times t_i and t_j are very different, whereas a red pixel corresponds to high similarity.

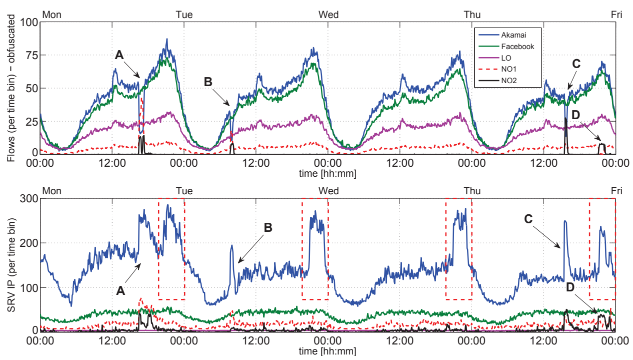


Figure 7. Flow counts (up) and server IPs (down) per AS, 5-min aggregation.

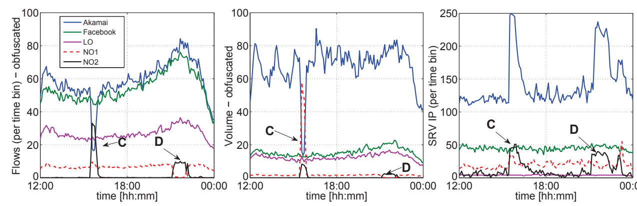


Figure 8. Flow counts, volume and server IPs per AS, for 12 hours.

IPs only. The figures reveal a different (re)usage policy of the IPs hosting the contents. In particular, Akamai uses the same servers for 4 to 7 days (see multi-days blocks around the main diagonal). When it changes the used addresses, the shift is not complete as we can observe the macro-blocks slowly fading out over time. This suggests a rotation policy of the address space of Akamai, on a time-scale of weeks. However, we cannot prove this conjecture because of the limited duration of the analyzed dataset. On the other hand, Facebook AS does not reveal such a clear temporal allocation policy. It alternates periods of high stability (e.g. between days 4 and 10) with highly dynamic periods (e.g., from day 18 onward). It is interesting noticing that Facebook AS is the responsible for the abrupt change in the distributions observed from the 18th day on, in the TSP of the overall traffic.

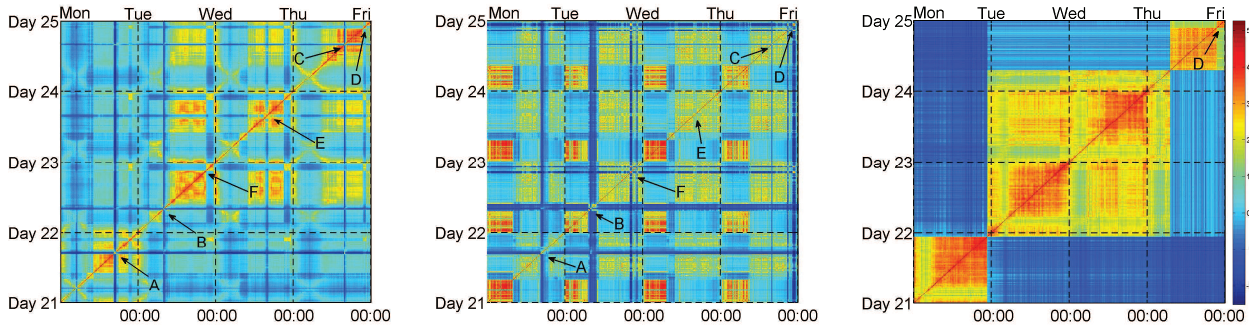
Our deeper analysis reveals that Akamai and Facebook AS actually employ periodical rotations of the servers they used to provide the contents, alternating periodic cycles of relatively low dynamics with more abrupt changes, especially as observed for the case of Facebook AS.

IV. Inter-AS Behavior & Unexpected Events

We devote the last section of the paper to study how the main hosting orgs./ASes interact with each other to serve the Facebook flows, and specially focus on the identification of some unexpected and worth to analyze events. We clarify to the reader that these events are assessed as "unexpected" with respect to the behavior observed in our traces, i.e., from the perspective of the ISP hosting the vantage point. In this study we do not have enough data (e.g., from multiple vantage points) to find the root causes of such behaviors, which might be the result of more complex and planned activities by the involved ASes.

To illustrate the temporal behavior of the different orgs. hosting Facebook, figure 7 plots the time series of the 5-min count of flows and server IPs per org., for four consecutive days (from day 21 to day 24). The figure shows that the flow share across the orgs. remains practically constant during the day, with a clear daily pattern in the number of active server IPs. It is worth noting that Akamai systematically doubles the number of servers during the peak hours (9pm - 10pm). Interesting is also the fact that, according to figure 7 (up), Facebook AS and Akamai serve a similar number of flows, which does not correspond a-priori with the flow shares depicted in figure 1(b) for the complete evaluation period. This is actually due to the duration of the active flows being served by each org.. Indeed, recall from Section II that Facebook AS servers host the dynamic contents of Facebook, especially those corresponding to chats and sessions keep-alive. This results in a big number of small yet long-lasting flows, explaining the difference in the overall share of served flows.

Fig. 7 additionally shows the occurrence of four unexpected events, identified as A, B, C and D, which break the normal traffic pattern. Events A and B have similar characteristics: even if the number of IPs steeply increases, the number of flows and traffic volume served by Akamai abruptly decreases. The number of flows served from NO1 and NO2 abruptly increase, and so does the number of active IPs in both orgs., at least in event A. This strongly indicates that the flows served by Akamai under normal operation (i.e., the majority of the time) are now served by neighboring ISPs. As we said before, Akamai actually deploys servers inside the ISPs [7], which



(a) Akamai (b) NO1 (c) Facebook AS

Figure 9. TSP of flow counts distributions at 5 min. time-scale. The TSP can also be used to detect abrupt changes. A transient anomalous event appears in the TSP as a full blue cross centered on the main diagonal, at the time of the event.

also explains the synchronized shift of flows. Figure 8 depicts a 12 hours zoom around the events C and D. During the event C, the Akamai drop is again compensated by NO1 and NO2 in terms of volume. However, unlike NO2, there is a limited increase in the number of flows served from NO1, suggesting that the latter takes over the largest flows from Akamai. Event D differs from the previous ones since it does not involve Akamai, and it is characterized by a swap in the number of flows between NO1 and NO2.

These unexpected events can also be identified through the TSP tool, applied to the traffic distributions. A transient anomalous event appears in the TSP as a full blue cross centered on the main diagonal, at the time of the event. Figure 9 shows the TSPs of the flow counts distributions between days 21 and 24 at a 5 minutes time-scale (i.e., the same period and aggregation depicted in figure 7), for Akamai, NO1, and Facebook AS respectively. The events A, B, and C are clearly visible in the TSPs of Akamai and NO1, and are totally absent from the Facebook AS TSP. These events are also clearly visible in the TSP of NO2 (not reported for space limitations), and are in total accordance with the analysis of the time-series in figures 7 and 8. Regarding the event D, it is observable in all the TSPs, even though it is completely invisible in the time-series of number of flows and volume for Facebook AS in figure 8. Furthermore, figures 9(a) and 9(b) pinpoint the presence of two more anomalous events in the Akamai and NO1 traffic, namely the events E and F, that are completely invisible in the flow and volume plots. This additionally justifies the usage of the TSP tool to identify such unexpected events.

We acknowledge that we do not know the ground truth or root causes of the aforementioned unexpected server selection events. We did not observe any abrupt variation in the total traffic, throughput, average RTT to the server IPs, nor in the number of erroneous HTTP responses during the events A-D, suggesting that the server selection did not impact the end-user QoE. However, we argue that these fast and significant traffic shifts might be highly costly for the LO. Indeed, we verified via traceroutes that Akamai, NO1, and NO2 are neighbors to LO. As reported in the Internet AS-level topology archive[3], the relation between LO and Akamai is peer-to-peer (P2P), whereas the relation between LO and both NO1 and NO2 is customer-to-provider (C2P). In a nutshell, the P2P relation results in no transit costs for the LO for the flows served by Akamai, whereas the C2P relation might represent additional transit costs for the LO for flows coming from NO1 and NO2. For this reason, such events are worth to be detected and further analyzed.

[3]Internet AS-level Topology Archive at http://irl.cs.ucla.edu/topology/

V. Conclusions

In this paper we have characterized the traffic of a highly popular, Internet-scale web service such as Facebook, as well as analyzed the dynamic behavior of the underlying hosting servers. We have shown how complex is the hosting infrastructure serving Facebook flows in a mobile network, in terms of involved organizations caching contents, the number of servers distributed at multiple locations, the complexity of the load balancing policies, and the temporal characteristics of the traffic. In addition, we have provided an analysis of some unexpected events in the way Facebook traffic is delivered, which might have a direct impact on the transport costs faced by the ISP providing the Internet access. We believe that the characterization provided in this paper offers a sound basis to network operators for understanding the dynamics behind Internet-scale web services delivered by large CDNs.

References

[1] C. Labovitz et al., "Internet Inter-domain Traffic", in *SIGCOMM*, 2010.

[2] V. Gehlen et al., "Uncovering the Big Players of the Web", in *TMA*, 2012.

[3] J. Erman et al., "Network-Aware Forward Caching", in *WWW*, '09.

[4] P. Casas et al., "IP Mining: Extracting Knowledge from the Dynamics of the Internet Addressing Space", in *ITC*, 2013.

[5] R. Krishnan et al., "Moving Beyond End-to-End Path Information to Optimize CDN Performance", in *IMC*, 2009.

[6] E. Nygren et al., "The Akamai Network: A Platform for High-Performance Internet Applications", in *SIGOPS*, 2010.

[7] C. Huang et al., "Measuring and Evaluating Large-Scale CDNs", in *IMC*, 2008.

[8] A. Bär et al., "DBStream: an Online Aggregation, Filtering and Processing System for Network Traffic Monitoring", in *TRAC*, 2014.

[9] P. Fiadino et al., "HTTPTag: A Flexible On-line HTTP Classification System for Operational 3G Networks", in *INFOCOM*, 2013.

[10] A. D'Alconzo et al., "Distribution-based Anomaly Detection in 3G Mobile Networks: from Theory to Practice", IJNM, vol. 20, John Wiley & Sons, 2010.

[11] MaxMIND GeoIP Databases at http://www.maxmind.com. [acc. on 15-01-2014]

[12] I. Poese et al., "IP Geolocalization Databases: Unreliable?", in *ACM CCR*, vol. 41(2), 2011