Characterizing Service Access Patterns under Heterogeneous Clients

Shuyu Zheng[1], Fuqi Lin[1], Xuan Lu[2], Yulian Yang[3], Hongfei Deng[3], Jun Zhang[3], Yun Ma[4*], Xuanzhe Liu[1]
[1]Key Lab of High-Confidence Software Technology (Peking University), Ministry of Education, Beijing, China
[2]University of Michigan, USA; [3]Alibaba Group, Beijing, China; [4]Tsinghua University, Beijing, China
{zhengshuyu,linfuqi}@pku.edu.cn; luxuan@umich.edu; {yihan.yyl,hongfei.denghf,zj157077}@alibaba-inc.com;
yunma@tsinghua.edu.cn; liuxuanzhe@pku.edu.cn

***Abstract*—Service providers usually maintain multiple kinds of service clients, such as desktop Web pages, mobile Web pages, and mobile native apps, to satisfy different user requirements. Understanding how users consume services via these heterogeneous clients is important for service providers to efficiently manage back-end resources. However, little has been known about such facts. To bridge the knowledge gap, in this paper, we conduct the first empirical study on service access patterns via heterogeneous clients. Our study is based on two large-scale real-world datasets of launching events from both app and Web clients of 986 services, involving a sample of 60 million users within 7 days. We analyze the characteristics of service access patterns from both spatial view and temporal view. We explore how users consume online services through different service clients, and to what extent their usage behavior patterns vary when consuming the same service. Our findings provide better understandings for heterogeneous service clients that can facilitate the development, configuration, deployment and maintenance of service back-end.**

***Keywords*-Service Access, Service Provider, Heterogeneous Clients, Web and App, Empirical Study**

I. Introduction

The prevalence of Web services is bringing a number of facilities to us and shaping our daily life. People are used to consuming a variety of online services through heterogeneous service clients.

The fact is, service providers often maintain several forms of clients as their service consumers are on different hardware. These heterogeneous clients can be divided into two categories: native apps (hereinafter called App) and Web apps (hereinafter called Web). Conceptually, both of the two forms of service clients fulfill specific tasks and features for end-users. However, they are quite different for service providers and developers which brings a lot of challenges for the development, configuration, deployment and maintenance of service back-end. Previous work [1] [2] revealed that the request semantics of Web service vary among different forms given the same features. Choosing proper back-end architectures according to the characteristic of client usage can improve the quality of service (QoS) for service providers.

However, little has been known about how different service clients are accessed by users, which is meaningful. If the user behavior patterns differ across heterogeneous clients, service providers can adjust the processing logic or configuration of service back-end accordingly to improve QoS and reduce the cost. More specifically, this paper tries to answer the following two research questions:

- **RQ1: How different are services consumed by users from heterogeneous service clients?**
- **RQ2: How are services revisited by users from heterogeneous service clients?**

We conduct an empirical study based on two large-scale real-world datasets of service launching events collected from both Web and App form of service clients. We focus on a sample of 986 popular online services, and analyze 2 billion service launching events of 60 million sampled users within a week from both spatial view and temporal view.

II. Methodology

A. Dataset

We use two datasets, one App usage dataset and one Web usage dataset, to indicate users' usage behavior of both App and Web. The two usage datasets are collected by *Umeng+*, a third-party data intelligence service provider.

To bridge the data collected from different service clients, we map the mobile app and the web page of the same service into a service identifier with manually collected packagename-domain correspondence. Next, we match users in the two datasets with their anonymized identifiers.

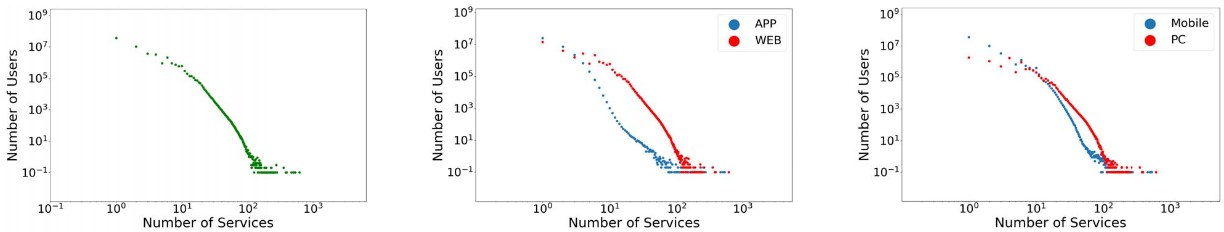
We sampled 986 services which have both App clients and Web clients. Then we extracted a sample of 60 million users who have used at least one of the sample services during the study. We verified that the sample distribution and population distribution are consistent. The combined dataset captured more than 2 billion launch events from both forms of service clients during a one week period starting from November 14, 2019. The percentage of App launch events and web launch events are 59.7% and 40.3%, respectively. Roughly 53.1% of users use only App clients, and 43.0% of users use only Web clients. Only a small amount (3.9%) of users use both App and Web clients.

B. Revisitation Analysis

The revisitation curve was proposed in [3] to perform web usage analysis. Jones et al. [4] then applied the method to analyze the users' revisitation patterns of Mobile Apps. We

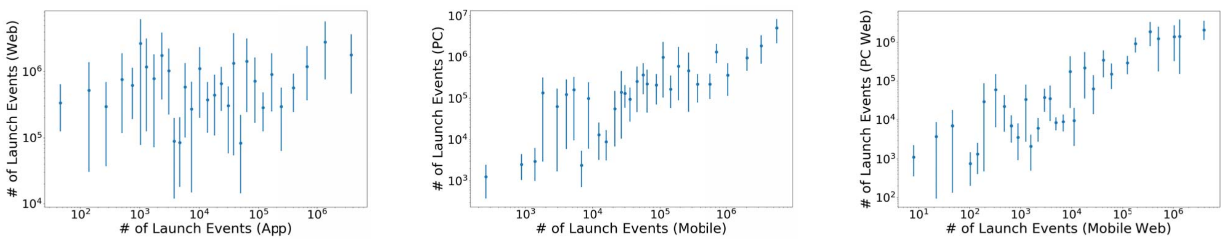
*Corresponding author.

2474-2473/20/$31.00 ©2020 IEEE
DOI 10.1109/SCC49832.2020.00016



(a) (b) (c)

Fig. 1: Statistics of the App usage dataset and Web usage dataset.



(a) (b) (c)

Fig. 2: Aggregated statistics of service launch events, each point representing a service.

use the revisitation curve to compare and evaluate revisitation behavior for different services. The curve of a service indicates the probability of this service being revisited by the same user within a period of time. Revisitation curves are generated by the following steps: 1) Calculate all inter-arrival time between consecutive pairs of revisits. 2) Assign the inter-arrival time into 15 exponential bins: 2, 4, 8, 16, 32 minutes; 1, 2, 4, 8, 16, 32 hours; 2, 4, 6 days and above. 3) Aggregate data across all users and calculate the frequency of revisits within the corresponding bin.

We build up the Web revisitation and App revisitation vectors corresponding to the 15 exponential bins of users' inter-arrival time. We treat the hybrid vector as features of the services and apply K-Means clustering algorithm to identify services with similar revisitation curves.

III. Findings and Implications

A. Basic Statistics

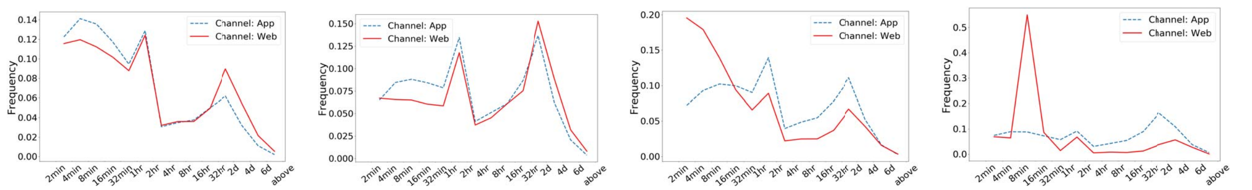
In this section, we provide a description analysis of the consumption of services from different channels and different devices.

Figure 1 shows the distribution of the number of services consumed per user. The x-axis represents the number of unique services a user has consumed during the experiment period, and the y-axis represents the number of such users. Figure 1a conveys to us that most users consume very few services, which is consistent with previous work. As is shown in Figure 1b, the number of services a user consumes through Web clients is more than that of App clients. This finding implies that web service is more abundant than app service for most users. Figure 1c shows that users consume more services on the PCs than on mobile devices. However, the usage difference between devices is smaller than the usage difference between App and Web.

Implication*: Since users consume more services through Web clients, it is better for service providers to develop Web clients if they want to attract more users.*

Figure 2 shows the correlation between the launch events of services from different forms of clients and different devices. We divide the data points into 30 discrete bins, then estimate the central tendency. As is shown in Figure 2a, the number of service launch events concentrates between 1,000 and 10 million for both forms of clients. This implies that although a few services are consumed by a single form of clients, a majority of services are consumed by both forms of clients in balance. However, Figure 2b shows an inconsistent result: the more a service is launched on the mobile devices, the more it is launched on PCs. This conveys to us that popular services attract visits from both PCs and mobile devices. The Web client is OS-independent, and can be accessed through web browsers on any device. Figure 2c shows the number of service launch events of Web clients on different devices. We can observe a similar result as Figure 2b. A majority of services are consumed by Web clients through PCs and mobile devices in balance.

Implication*: When allocating back-end resources, service providers should separately consider the Web and App clients since the number of requests has no relationship between the two forms of clients. However, they could just consider either PCs or mobile devices to estimate the resource demand since both kinds of devices have similar amounts of requests.*



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3 (d) Cluster 4

Fig. 3: Centroid revisitation curves for the three clusters of applications.

TABLE I: Clusters of services based on their revisitation curve and their descriptions.

Cluster Group	Cluster Number	Description	Cluster Size (# Services)
Synchronous	1	Three synchronous revisitation peaks	414 (42%)
	2	Two synchronous revisitation peaks	255 (26%)
Semi-synchronous	3	Two synchronous revisitation peaks and one exclusive peak	283 (29%)
Asynchronous	4	Different revisitation peaks	34 (3%)

B. Joint Analysis of Revisitation Patterns

The revisitation clustering results are listed in Table I. We group the clusters according to the characteristics of their curves of centroid.

1) Synchronous: The 414 services in cluster 1 and 255 services in cluster 2 are grouped into *synchronous* because the revisitation pattern of App and that of Web are synchronous. The services in cluster 1 have three synchronous revisitation peaks (Figure 3a) at 4-8 minutes, 1-2 hours, and 32 hours-2 days, respectively, while the services in cluster 2 (Figure 3b) have two different synchronous revisitation peaks at around 1-2 hours, and 32 hours-2 days, respectively. It is worthy noting that the revisitation pattern of App and that of Web are synchronous for most services (68% of all services).

2) Semi-synchronous: As is shown in Figure 3c, the revisitation pattern of App and that of Web are not always synchronous for the services in this group. There are two synchronous revisitation peaks for the services in this group, which occur at 1-2 hours, and 32 hours-2 days, respectively. There is an exclusive peak for the revisitation pattern of Web, which locates at 2-4 minutes. 283 of 986 (29% of all services) services belong to *semi-synchronous*.

3) Asynchronous: In opposite to the *synchronous*, the revisitation pattern of App and that of Web have different revisitation peaks for the services in *asynchronous* (Figure 3d) For the revisitation pattern of Web, there are three different peaks, which occur at 8-16 minutes, 1-2 hours, and 2-4 days, respectively. While for the revisitation pattern of App, there are two different peaks, which locates at 1-2 hour, and 32 hours-2 days, respectively. The highest peak at 8-16 minutes indicates that the revisitation of Web channel is consistent with the fast revisit pattern mentioned in previous work [3]. While the revisitation of App channel shows the characteristics of slow revisit pattern.

Implication*: Services are more likely to be revisited by Web clients for short time period (i.e., after several minutes) so that service providers could maintain more cache resources for Web clients than for App clients. Based on revisitation patterns, service providers could do pre-loading or pre-processing before the users are likely to revisit.*

IV. Conclusion

This paper conducts an empirical study on the characteristics of service access patterns of heterogeneous service clients. We find that the number of services a user consumes through Web is more than that of App. A majority of services are consumed by both forms of clients in balance. Popular services attract visits from both PCs and mobile devices. We observe three revisitation patterns: synchronous, semi-synchronous, and asynchronous. Based on our findings, we draw several implications for service providers to efficiently manage back-end resources.

Acknowledgment

This work was supported by the Key-Area Research and Development Program of Guangdong Province under the grant number 2020B010164002, and the Beijing Outstanding Young Scientist Program under the grant number BJJWZYJH01201910001004.

References

[1] Y. Liu, X. Liu, Y. Ma, Y. Liu, Z. Zheng, G. Huang, and M. B. Blake, “Characterizing restful web services usage on smartphones: A tale of native apps and web apps,” in *2015 IEEE International Conference on Web Services*. IEEE, 2015, pp. 337–344.

[2] Y. Ma, X. Liu, Y. Liu, Y. Liu, and G. Huang, “A tale of two fashions: An empirical study on the performance of native apps and web apps on android,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 5, pp. 990–1003, 2017.

[3] E. Adar, J. Teevan, and S. T. Dumais, “Large scale analysis of web revisitation patterns,” in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2008, pp. 1197–1206.

[4] S. L. Jones, D. Ferreira, S. Hosio, J. Goncalves, and V. Kostakos, “Revisitation analysis of smartphone app use,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 1197–1208.