Pervasive and Mobile Computing 29 (2016) 95–112

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Passive profiling of mobile engaging behaviours via user-end application performance assessment[☆]

Xiaming Chen[a,*], Siwei Qiang[a], Jianwen Wei[b], Kaida Jiang[c], Yaohui Jin[a,c]

[a] State Key Lab of Advanced Optical Communication System and Network, PR China

[b] Center for High Performance Computing, Shanghai Jiao Tong University, PR China

[c] Network and Information Center, Shanghai Jiao Tong University, PR China

ARTICLE INFO

Article history:
Received 22 October 2014
Received in revised form 1 August 2015
Accepted 9 December 2015
Available online 24 December 2015

Keywords:
Mobile measurement
Engaging behaviour
Quality of experience
Application performance
Structured correlation analysis

ABSTRACT

Benefiting from the booming mobile industry, our daily connections have become instant and ubiquitous. In the meanwhile, user experience in mobile services becomes an important consideration for service providers to cultivate customers' loyalty, or for network operators to seek profit chances from existing infrastructure. To meet these requirements, we specify *engaging behaviours* to characterize the dynamics of user participation in mobile applications in real contexts. Leveraging a five-month collection of backbone traffic in an operational WiFi mesh network, we make comprehensive insights on the variation of engaging behaviours, as well as their non-linear and counter-intuitive interactions with user-perceived application performance. With the Hidden Markov Modelling of individual's engaging trajectory, two distinctive behavioural clusters are identified and investigated. These achievements will be contributed to the development of multiple pioneer R&D areas, such as mobile network simulation and user experience optimization.

1. Introduction

Mobile markets are driven to be explosive by the prosperity of portable devices and well-designed software applications. This trend also stimulates the research areas of ubiquitous computing [1] and human behaviour science [2] on behalf of convenient access of human generated mobile content. For example, the success of subscription and advertisement revenue models relies heavily upon the knowledge of user preference and participation learned from fragmented digital traces. In mobile world, an accurate profiling of user experience (UX) and engaging behaviour is of benefit to the resource optimization for service providers and even the hacking-perspective enhancement of user privacy protection.

User experience is a key aspect in designing human-centred technologies [1,3–5] and service models [6]. It usually focuses on the positive manner of user interactions with available resources, and particularly the phenomena associated with being captivated by technology or services. From end-users' view, a successful design of target products not only stimulates their interests to use it immediately, but also engage them with a solid endurability. However, the original designer begins to lose

☆ This work was supported by NSFC under Grant No. 61371084 and the Major State Basic Research Development Program of China (973 Program) under Grant No. 2010CB328200.

* Corresponding author.
E-mail addresses: chen_xm@sjtu.edu.cn (X. Chen), qiangsiwei@sjtu.edu.cn (S. Qiang), weijianwen@sjtu.edu.cn (J. Wei), kaida@sjtu.edu.cn (K. Jiang), jinyh@sjtu.edu.cn (Y. Jin).

http://dx.doi.org/10.1016/j.pmcj.2015.12.005
1574-1192/

control with the way customers use the product, once it is shipped to the market in the deployment phase. The information of customer usage and experience becomes a precious source to enhance product features in the next iteration.

Since the birth of Internet, the collection of experience information about network applications has gained enthusiasm in multiple R&D areas. One of the modern and evolving concepts is *engagement*, once proposed by the Advertising Research Foundation (ASF) [7] and widely used in marketing policies. O'Brien initially borrowed the concept into web measurement and generalized it to cultivate user-technology interactions [3]. The framework is a participation-based model which characterizes different temporal stages (e.g., engagement and disengagement) when users are involving in an application. As a complementary view, a functionality-based model with multiple engagement dimensions (e.g., endurability) is discussed in [8] which connects theoretical notations with practical metrics.

Despite the generality of these two frameworks, they suffer from two main weaknesses when we evaluate mobile user engaging behaviours: the deficiency of contextual adaptation and a lack of quantitative correlative interpretation. In mobile networks, contextual factors impact user engaging behaviours in multiple facets (such as temporal effect [9], applications [2], geographic locations [10], and public events [11]), while the quantitative correlation interpretation sheds light on developing diagnostic and actionable policies for mobile user experience. The correlative interactions are also addressed in recent video community [12–14], and yet user behaviours in mobile networks have not gained comprehension in a large-scale and systematic manner. In this paper, we focus on the profiling and modelling of user participating behaviours in mobile services, especially the interactions with user-perceived application performance in varying contexts.

For the assessment of user engaging properties and perceived application performance, we have two board alternative methods: subjective and objective [8]. The former recording user experience via interviews or questionnaires [3], represents real user experience but suffers from drawbacks such as lab-setting bias and individual cognitive prejudice (a.k.a. "halo effect" [15]). The latter based on passively collected metrics from communication data or management logs, suppresses the drawbacks and branches out conveniently into large-scale characterization. In this sense, we perform the passive measurement in our studies.

Although passive measurements of mobile traffic have been conducted in previous literature [16–18], they mostly address protocol and traffic properties, rather than responsive user behaviour consequences. For example, the same user under similar environment is observed having distinctive behavioural patterns against the degradation of network performance (Section 4.4). Furthermore, measuring protocol performance solo is far away from user experience, because low-level metrics (e.g., flow byte rate) explicate the efficiency of data transmission experienced by networks rather than by users [18]. In a different manner, we develop a set of objective user-centred metrics of user engaging behaviours and application performance. The profiling substance comes from an augmented behavioural model of mobile traffic, which gains a higher accuracy than existing models as for explicit distinction of engaging activities from network perspective. With this model, we make three contributions from following perspectives:

- Perform a characterization of mobile traffic and engaging behaviours from end-user's view. The proposed concurrence index equipped by the model is more powerful to capture delicate difference of user-perceived application performance than previous volume-based metrics (Section 3.3).
- Profile the behavioural dynamics of user participation in mobile usage and its interaction with user-perceived application performance. We observe high skewness of engaging session lengths and large diversity of user engagement on differing device platforms. The contextual factors also put significant impacts on the interaction (Section 3.4).
- Perform a unique modelling of individual engaging trajectories and a model-based clustering to explore user behavioural patterns. We find that user engaging behaviour is primarily governed by a small portion of latent states, and the behavioural patterns regarding principle engaging states illustrate distinctive properties in discovered user clusters (Section 3.5).

The remaining parts are structured as follows. Section 2 reviews previous literature on related topics. Section 3 describes the global methodology to analyse and model mobile engaging behaviours passively. Section 4 shows the experimental results, and we finalize the paper with a discussion on potential applications and limitations in Section 5.

2. Related work

A brief literature review is made on the general user engagement frameworks and mobile traffic measurement. Relative achievements are classified into three main categories:

General frameworks of user engagement: Several conceptual frameworks have been proposed on user engagement in different domains, including the participation-based [3] and functionality-based [8] models for technologies. Authors in [3] decompose user engagement into four distinct stages according their occurrence order (i.e., point of engagement, period of sustained engagement, disengagement, re-engagement), and discuss a wide range of attributes attached to each stage. Researchers in [8] address the functionality of engagement properties as well as their measurement in front-end web technology. Beyond technology domain, considering customer engagement in business models [6] gives assistance to understanding customer-brand value exchange and improving customer loyalties.

Measurement of user engagement: Subjective methods evaluate user experience directly via, e.g., questionnaires [13] or MOS metric [19], while objective methods suppress the subjectivity bias of the former and are suitable for quantitative analysis. In related domains, a batch of objective analytics have been proposed to study user engagement in computer

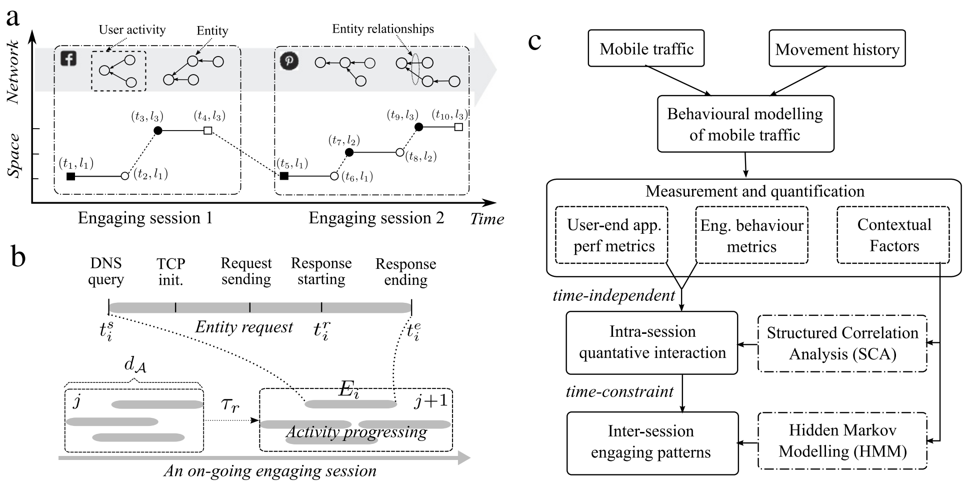


Fig. 1. (a) Illustration of user engaging behaviour in mobile networks from multiple dimensions. (b) User-oriented measurement of perceived application performance (each activity $\mathcal{A}_j$ consisting of one or multiple triggered entities E_i is the atomic unit to derive features). (c) The profiling and modelling framework for mobile engaging behaviour with user-end application performance assessment in contexts.

games [1], web analytics [20,4], and video applications [5,14,12]. However, the characteristics of user engaging behaviours in mobile services are less comprehended, which motivates our study to fill this gap.

Characterization of mobile traffic: This line of work covers the measurement of mobile traffic compositions [10], protocol performance [16,17], and service characteristics [18]. Huang et al. [16] utilize both crowd-sourcing and controlled platforms to analyse TCP and web application performance. Chen et al. [17] declaim that modern handheld devices in a campus WiFi network generate biased protocol mix while Gember et al. [18] report that handheld devices often send smaller and shorter TCP flows, along with narrower server distribution than non-handheld devices. These pioneer studies become the foundation of our understanding user perceived application performance in mobile networks.

Modelling mobile behaviours in context: This category endeavours to understand mobile user behaviours in contextual environment. The usage of Google's commercial WiFi networks is characterized in [10] where user activities (e.g., traffic volumes, mobilities) present substantial patterns with reference to the temporal and geographical span. Researchers in [2] probe the serendipity between user mobilities, semantic surroundings and application preferences. Gong et al. [11] investigate users' behaviour dynamics and the change of their mobility patterns in a campus Wi-Fi network. Another in-context measurement [9] focuses on the impacts of use states and physical environments upon the network performance. Built upon existing studies, our work makes it distinctive by characterizing and modelling user engagement considering user perception and a wide range of contextual factors. Our discoveries not only complement aspects of previous conclusions, but also present new properties of mobile user behaviours.

3. Methodology

3.1. Terminologies and overview

As stated in the general engagement model, user engagement is a multifaceted concept including "challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control" [3]. Our research complements the abstract concept with objective profiling from user behavioural perspective, which is, however, still complicated and impacted by alternative network choices, objective service quality and environmental factors. E.g., users in a heterogeneous network is able to switch service providers to obtain better network performance. This freedom of choices makes it harsh to quantify user engaging behaviour uniformly. To simplify the modelling procedures and refine our experience-centred goals, we merely consider homogeneous networks where a single technology is deployed. Additionally, our experimental network covers almost the entire observed area and is free for its users, which guarantees a higher priority in user's choosing list. Formally, we define the user engaging behaviour as *the overall observed user participation and interactions in mobile network services, especially under specific user-perceived application quality and contextual influences*.

As for the high dynamics of user behaviour in mobile scenario,[1] a robust behavioural model is still desired to gather user engaging states in a large-scale network. In Fig. 1(a), a multi-dimensional behavioural model is illustrated. On the network

[1] Please note that we refer the mobile network to not only the cellular 3/4G technology, but also other deployed wireless techniques such as WiFi and WiMAX mesh networks which provide users with continuous data services on the move.

dimension, atomic user activities are identified (with semantic category) from raw traffic traces by an augmented detection algorithm. On the spatial dimension, each activity is labelled with user whereabout, which is practically represented by the coordinates of cell base stations (BSs) in cellular networks, or access points (APs) in WiFi mesh networks. If the user activity is across an instant handover, we record the first location as triggered. On the temporal scale, a remarkable transmission silence (e.g., 15 min) is employed to indicate a separation between individual engaging sessions.

With this flexible behavioural model, we quantify engaging behaviours and application performance from the user side (Fig. 1(b)), which provides a fresh view against existing packet and flow level evaluation [9,18]. Different from the client-based paradigms [16], our approach employing passively collected network traces is convenient to expand into large-scale scenarios. The last but not least advantage comes from the revelation of intuitive connection between objective user experience and behavioural feedback [19].

Fig. 1(c) gives a bird's-eye view of the profiling and modelling framework. We perform analysis on user engaging behaviour from *time-independent* and *time-constraint* aspect respectively. For the former within an engaging session, user behaviour is characterized by individual dimensional metrics and analysed for the interaction between dimensions. One potential method is mutual information analysis [21] where the mutual information measures the uncertainty reduction for a random variable X if the variable Y is known. As the information of dependent direction is absent in this technique, we involve the structured correlation analysis (SCA) instead to quantify the detailed dimensional interactions, as well as the impacts of contextual factors.

From the time-constraint perspective, we are curious about the clustered engaging patterns under varying perceived application performance. By transforming an engaging session into a multidimensional data object, one direct method is multivariate time series clustering [22], which is, however, challenging because there are still no efficient methods to calculate time series distance, especially when both continuous and nominal features are present. For this reason, we involve a Hidden Markov Model (HMM) based clustering method to derive significant patterns in user engaging behaviour. We use HMM rather than the multivariate stochastic models [23], because the latter extracts time series into multiple components (containing stochastic trends, seasonal patterns and random errors) that are less interpretive of human behaviour. By introducing an efficient calculation of HMM distance, consistent user clusters are observed over multiple runs in our data-driven experiment.

3.2. Behavioural modelling of mobile traffic

Since the birth of Web technology, remarkable efforts have been made to understand HTTP traffic. There exist two typical categories developed based on timestamps [24] and web page structure [25,26] respectively. The former is identifying individual web pages with the time constraint between individual web objects, and the latter with content organization and page structures. Nevertheless, there are two limitations when we export existing algorithms into mobile scenario. First, modern mobile operating systems (OS) are popular with native applications[2] that mostly implement service logic via RESTful interfaces and result in a blur boundary of user behaviour in traffic. Second, the HTTP reference relationship (that records the source of a new coming web request) is rarely observed in mobile traffic, possibly for administrative considerations such as user privacy and system security. E.g., more than 70% of requests in the groundtruth dataset and 63% in experimental WiFi dataset are observed carrying empty referrer in HTTP headers. To capture these fresh properties, an enhanced behavioural model is proposed to analyse user behaviour.

Empirically, while interacting with mobile services, a user clicking activity may trigger one more network requests determined by underlying application implementation (Fig. 1(b)). E.g., when a user clicks the PLAY button in music apps, related requests for lyric and album data will follow the primary music streaming connection. Besides this user-aware portion, there is also some background traffic automatically generated by daemon services such as ads or analysis services. As the background traffic takes little connection with user consciousness, it is initially wiped out before our analysis and modelling.

In the behaviour model, an *activity* (denoted $\mathcal{A}$) is defined to represent an atomic user-triggered operation in mobile services, and an *entity* (denoted E) to be an individual request to remote service. A relationship of $E \in \mathcal{A}$ is qualified if the entity E is right issued by activity $\mathcal{A}$. Then the engaging behaviour of user can be modelled by a series of activities in chronological order, i.e., $\{\mathcal{A}_t\}_{T_l:T_h}, T_l \leq t \leq T_h$. In a broader sense, this paradigm is also compatible with the modelling of web traffic. That is, each web object [26] becomes an entity of a page-requesting activity. The core enhancement of our model is the consideration of logic relationships between entities, which tackles the problem that non-referrer traffic constitutes a great proportion of mobile data. Specifically, four typical relationships are considered in this paper: (i) Parallel ($E_i \parallel E_j$): Entities are associated in an independent way and can be requested theoretically in parallel; (ii) Conjunction ($E_i \wedge E_j$): There is a dependency between two entities and they must be requested simultaneously; (iii) Serial ($E_i \rightarrow E_j$): One entity is dependent on another, but they must be fetched in a successive order; (iv) Relayed ($E_i \mapsto E_j$): One entity takes over the role of previous one to perform the same functionality. Among these relationships, the parallel and serial represent loose

[2] Native applications stand for standalone (usually third-party) utilities installed into the mobile operating system. Most of them leverage representational state transfer (REST) interfaces to communicate with remote servers [17], but own a weaken structure than traditional websites from network perspective.

Algorithm 1 $\mathcal{AID}$ algorithm

Input: A chronologically ordered entities **E** extracted from mobile traffic, for a specific combination of {user, application}; the inter-activity idleness threshold τ_L.

for all $E_i \in \mathbf{E}$ **do**:

Peering: Match E_i to its precedent E_{i-1} *iff* $\tau_{i-1,i} \leq 2\tau_L$ and $LR(i-1, i) \in \{\|, \wedge, \rightarrow, \mapsto\}$;

Linking: If E_i has empty referrer, it is linked to E_{i-1} directly; otherwise, E_{i-1} and E_i are linked to their nearest referrer in time;

end for

Cutting: Traverse previous linking tree and perform a temporal separation (erase the link between entities) where $\tau_{ref,i} > \tau_L$.

Output: Isolated entity clusters as detected user activities.

constraints between entities in most applications, while the conjunction and relayed present strict constraints that cover a small group of applications, yet conveying significant volumes (e.g., HTTP-based streaming).

In the activity detection ($\mathcal{AID}$) algorithm, we use the linked tree data structure to store entity relationships and properties. The key idea is to restore the tree link by employing both explicit (the temporal and reference constraints [26]) and implicit (the logical relationships) rules. Afterwards, the constructed linked tree is cut into several subtrees, each of which represents an identified user activity. Algorithm 1 gives main sketch of the detection process. Specifically, by eliminating background traffic, a series of entities **E** under a combination of 'user' and 'application'[3] are extracted from the traffic and fed into $\mathcal{AID}$ algorithm. Before being added onto a linked tree, an entity must undergo a two-stage processing: At the peering stage, a new coming entity E_i is peered to its precedent if both the temporal and logical constraints are met, otherwise un-peered; at the linking stage (for peered entities), we check the referrers of E_i and its peer E_{i-1} to link them both to their nearest referrer entity in time. If E_i is un-peered and has no referrer information, a new linked tree will be created for next iteration. After the construction of linked trees, we traverse the trees and perform a temporal separation when E_i and its linked peer span over a time slot larger than a threshold τ_L. In the algorithm, parameter $\tau_{i,j}$ and τ_L estimate inter-entity and inter-activity idleness respectively. Function $LR(*)$ evaluates the type of logical relationships between entities. As we can see below, the logical relationships dominate both temporal and reference constraints in the linking stage. This kind of design emphasizes the critical quality of logical rules in the detection performance of $\mathcal{AID}$ algorithm.[4]

3.3. Passive measurement of user engaging behaviour in context

3.3.1. Behavioural metrics

We quantify user engaging behaviour in mobile networks with two primary considerations: (1) Measures are intuitive to interpret user behaviour about service usage; (2) they are actionable to direct quantitative analysis and optimization in practice. According to the proposed model, behavioural metrics are derived at the granularity of individual engaging sessions, i.e. $Q = \{\mathcal{A}_i\}_m$ being a session with m activities.

Engaging session duration: This metric is utilized to characterize user active participation in an engaging session, i.e., $D_e = \max\{T_i^e\} - \min\{T_i^s\}$, $i \in [1, m]$, where T^s and T^e are the starting and ending times of user activity (Fig. 1(b)). Practically, encouraging users for longer engaging sessions is beneficial for potential resource optimization and business profitability.

User visit frequency: Use effort is a vital aspect about user endurability, interactivity and pleasure [3]. We use this metric to characterize the activeness and smoothness of user engaging behaviours. Given $m-1$ pause intervals τ_r between activities, the visit frequency is quantified by the expected number of user activities within a time unit, i.e., $f_v(H) = \mathbf{E}[\tau_r]_{m-1}{}^{-1}$. In web community, a related metric for web browsing is the number of page views [20] to represent user click depth in a conversation.

Ratio of interruption activities: Interruption activities are a category of abnormal behaviour observed when users endeavour to recover their usage from performance fluctuations. This is a signal of user positive recovery of patience. From network level, the real interruption activities are easily confused by TCP control signals and distinctive browser implementations [27]. Here we employ a payload estimation criteria proposed by Rossi and colleagues [27] to eliminate the misreport (see Definition 1). Finally, the interruption ratio r_{int} is defined by the proportion of interruption activities within an engagement session.

Definition 1. Interruption Activity: An activity with a flow meeting two conditions simultaneously: 1. $\neg(FIN_s \vee RST_s) \wedge RST_c$, 2. $\frac{t_{FIN}}{\mu_{RTT}+\sigma_{RTT}} \leq 1$, where t_{FIN} measures the idle time between the last data packet and flow termination; μ_{RTT} and σ_{RTT} give the expectation and variance of RTT estimations.

[3] A unique user is recognized by device's MAC address and an application by its user-agent string in HTTP header.

[4] These logical rules are keyword-based and are integrated into the algorithm with static configurations; e.g., a serial relationship is identified when the keywords (e.g., 'auth', 'login') exist in the preceding URL to access authorized content.

3.3.2. Metrics of perceived application performance

In this part, we define objective metrics about user perceived application quality. This group of metrics is derived from intuitive user behaviour based on the granularity of individual activities.

Perceived activity duration: This measure gives the global latency before user finishing an activity. Given an activity $\mathcal{A} = \{E_i\}_n$, each entity E_i is decorated by a tuple of three time-stamps, (t_i^s, t_i^r, t_i^e), denoting the request emission, response start and finish respectively (Fig. 1(b)). The activity duration experienced by a user is calculated by $d_{\mathcal{A}} = \max\{t_i^e\} - \min\{t_i^s\}, i \in [1, n]$.

Perceived waiting time: We also choose the entity-level quality to investigate subtle user experience. The perceived waiting time gives an intuitive evaluation of user patience consumption during service engagement. By decomposing an entity into two segments, zero-byte waiting ($|t_i^r - t_i^s|$) and data transmission ($|t_i^e - t_i^r|$), the former indicates the passage of time before the first byte of response, while the latter the receiving process. From user end, the zero-byte waiting part, $w_{\mathcal{A}} = \frac{1}{n}\sum_i |t_i^r - t_i^s|$, is a purely patience-consuming process as addressed in [27]; thus it makes sense to be an important perceptual metric.

Perceived throughput: Temporal metrics are not sufficient to quantify user-perceived performance, as for the diversity of mobile applications and usage scenarios. E.g., a streaming activity lasting for one minutes may introduce 'equivalently' negative user perception when compared to a twitter activity lasting for 30 s. To evaluate this difference, we define the perceived throughput to measure the application data rate from user side, i.e., $b_{\mathcal{A}} = I_c \sum_i r(i)$, where I_c stands for the concurrency index (CI for short) given by Definition 2 and $r(i)$ the average data rate of ith entity.

Definition 2. Concurrency Index: Given an activity $\mathcal{A} = \{E_i\}_n$, the concurrence index of $\mathcal{A}$ is defined as $I_c = \frac{t_{\mathcal{A}}}{n \cdot d_{\mathcal{A}}} \in (0, 1]$, and $t_{\mathcal{A}} = \sum_i |t_i^s - t_i^e|$ indicating the accumulated intermission of all entities with a *relayed* relationship assumed.

3.3.3. Contextual factors

Previous engagement analysis in web community [8,20] considered more about page content inner structures rather than external factors, because traditional web users usually browsed websites through a linked network. However, this is another situation in mobile networks: users with high mobility are facing dynamic network conditions and changeable external contexts. It is of great importance to explore the contextual impacts on mobile user engaging behaviour and perceived application quality. In this part, we evaluate four typical dimensions of contextual factors concerning engaging behaviour: *user personality* [28], *temporal effect* [28], *location preference* and *application semantic*.

User personality encodes the diversity of use habits, historical experience and behavioural patterns. This factor brings about significant distinctions of user activities in three other dimensions. While the temporal effect has been explored tremendously in previous traffic characterization [10,16,18], we complement previous observations with user behavioural properties and introduce two variables: time of the day (ToD) and time of the week (ToW). For the spatial dimension, *location familiarity* (or PlaceRank) is defined to represent the spatial preferences of individuals. By summing up the total dwelling time of each user at each observed location, the location familiarity is calculated by the reverse rank of accumulated time the user being observed at that place. That means, the smallest PlaceRank gains most visitation and is considered to be with highest familiarity. For application semantics, we classify mobile traffic into several categories by checking the host and user-agent data carried in HTTP headers. Specifically, the application category of each activity is determined by the information of its first requested entity.

In following analyses and modelling, contextual factors are evaluated at the granularity of individual engaging sessions. By assigning each engaging session into time slots, measures in each slot are normalized by the maximum over whole day. For those sessions containing multiple label candidates, the dominant one with the largest number of activities is selected. For example, an engaging session with 35 activities in 'news' category and 10 activities in 'music' category will obtain a label of 'news'.

3.4. Correlation profiling of behavioural and perceptual metrics

Quantitative analysis [4,5,12,14,20] is an effective way to integrate derived data models into existing productive systems and operational policies. In this section, we introduce the correlation analysis of user engaging behaviour and underlying application performance with Structured Correlation Analysis. The core spirit is shared with the theory of Reconstructability Analysis [29] in which a distribution is decomposed (or compressed, simplified) into projected distributions or margins. Assuming a set of observations I and a feature space $O = \{m_i\}_q$, we have each observation consisting of a q-dimensional feature vector, within which *structured features* (denoted $S \subseteq O$) are a subset to explore their subtle impacts on the profiling target. Given structured feature variables $X, Y \subseteq S$, the conditional distribution is considered:

$$R_{ij}(Y|X) = \{r_{ij}(x, y) | \forall y \in Y, X = x\}, \tag{1}$$

where $r_{ij}(x, y)$ gives the correlation of m_i and $m_j \notin X \cap Y$, with X and Y being the structured feature dimensions. For a distribution, odds of significant correlations, $\phi = \frac{|R_{p<0.05}|}{|R_{p\geq 0.05}|}$, are also calculated to represent the probability that a significant instance occurs over the insignificant ones.

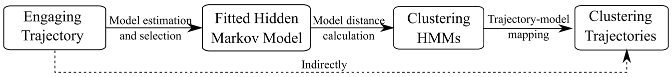


Fig. 2. A model-based clustering for user engaging trajectories.

Afterwards, we choose Spearman's correlation coefficient (r_s) to quantify the connection between behavioural and perceptual metrics. As the available data can be viewed as a sample of underlying stochastic process, the real correlation coefficient is approximated by a Bayesian estimation [30] from the empirical correlation coefficient:

$$P(r_s|m_i, m_j) \propto P(r_s)\frac{(1-r_s^2)^{(n-1)/2}}{(1-r_s \times r_s')^{n-3/2}}, \tag{2}$$

where r_s' defines the empirical correlation. By making substitution $r_s = \tanh \xi$ and $r_s' = \tanh \eta$, ξ is found to be a normal distribution with mean η and variance $1/n$. Concerning the prior $P(r_s)$, an easier inference may be $P(r_s) \propto (1 - r_s^2)^c$, in which a common choice of $c = 0$ will be reasonable if we have limited prior information [30]. Finally, positive r_s means an increasing monotonic tendency, and vice versa. An extreme value of 1 or -1 represents a strictly monotonic relationship while zero means no obvious dependency.

3.5. *Contextual modelling of user engaging behaviour*

In previous studies, it has been shown that human behaviour in mobile networks exhibits strong temporal dependency and group similarity, as a result of regular commuting routines and stable personal interests. For example, Trestian et al. [2] find that the mobile applications accessed by users are correlated with their movement patterns, and the probability of two users sharing similar cyber interest is proportional to the density of hot spot. To complement these observations, we are curious about two relative questions: (1) How does mobile user engaging behaviour change over time? (2) What patterns about user behaviour are observed at a population level?

3.5.1. *HMM-based modelling of user engaging trajectory*

Different from the intra-session analysis above, we introduce the temporal constraint on user sessions and give the representation of engaging trajectories in Definition 3. Each observation of an engaging trajectory is characterized by ten features, i.e., $\mathbf{O}_e = \{D_e, f_v, r_{int}, d_{\mathcal{A}}, w_{\mathcal{A}}, b_{\mathcal{A}}\}$ and $\mathbf{O}_c = \{App, PR, ToD, ToW\}$, where *PR* means the location familiarity (or PlaceRank).

Definition 3. Engaging Trajectory: Given an q-variate vector $\mathbf{O}_t = (O_t^1, \ldots, O_t^q)$ for each session, the engaging trajectory of an individual is represented by a sequence of observations in chronological order $\mathbf{O}_{1:T} = \langle \mathbf{O}_1, \ldots, \mathbf{O}_T \rangle$.

For these multivariate time series, Hidden Markov Model is a powerful tool to capture underlying rules governing the generation of time series. Its family has accepted widespread use in speech recognition [31], bioinformatics [32], etc. The mechanism of HMM is that observations at any time are distributed as a mixture of *finite* latent engaging states satisfying certain Markov properties over time [31]. Intuitively, contextual factors usually represent the confounding parts that impact user engaging behaviour and experience potentially. In this regard, contextual factors are considered separately from the behavioural and perceptual metrics. Let $\mathbf{S}_{1:T}$ and $\mathbf{c}_{1:T}$ be the hidden states and covariates of engaging trajectory $\mathbf{O}_{1:T}$ respectively. An HMM (λ) is estimated to determine transition probabilities between hidden states, emission probability at a given state, and the distribution of initial states.

We perform the Baum–Welch criteria [31] to fit model parameters, and the likelihood $P(\mathbf{O}|\lambda, \mathbf{c})$ is maximized via iterative calculation of estimated model $\hat{\lambda}$ with $P(\mathbf{O}|\hat{\lambda}, \mathbf{c}) \geq P(\mathbf{O}|\lambda, \mathbf{c})$. To determine the number of hidden states, Akaike's Information Criteria (AIC) [33] is involved to balance the *goodness-of-fit* and *model complexity*, which is a penalty-based measure offering relative information loss to use an estimated model to represent the true model. Because the engaging trajectory contains a small number of observations ($\bar{T} = 50.87$) in our scenario, a correction version of AIC for small samples, i.e., $AIC_c = AIC + \frac{2k(k+1)}{T-k-1}$ (k being the parameter number), is preferred to reduce the overfitting risk of selecting models with too many parameters [34]. Finally, we define *principle engaging states* as the ones identified with occurrence frequency in the hidden state sequence above a given threshold p_0.

3.5.2. *Extraction of significant engaging behaviour patterns*

Now we tend to answer the second question by exploring the patterns of user engaging behaviour in large populations. As we discussed in Section 3.1, a model-based clustering is preferred to the direct clustering of multivariate time series, due to the mixture of nominal and continuous features in engaging trajectories. As shown in Fig. 2, two engaging trajectories are considered to be similar when their model parameters or remaining residuals after fitting the model are similar [22].

Generally, a matrix about distance (or similarity) of data objects is the basis to feed a clustering algorithm to discover potential patterns [35]. A key insight is that the likelihood $P(\mathbf{O}|\lambda_i)$[5] can be interpreted as the probability distribution conditioning on model λ_i over the whole trajectory space. We hence employ the divergence measure (i.e., Hellinger distance [36]) for probability distributions that defines the distance between two HMMs. For a specific engaging trajectory, the optimal state sequence that most probably generates the trajectory is considered, i.e., $\mathbf{S}_m := \arg\max_{\mathbf{S}} P(\mathbf{O}|\mathbf{S}, \lambda_i)P(\mathbf{S}|\lambda_i)$. The HMM can be represented by the probability distribution of this optimal state sequence, viz. $f(\lambda_i) = P(\mathbf{O}, \mathbf{S}_m|\lambda_i)$. Interpretively, the optimal state sequence corresponds to user behaviour patterns that generate observed trajectory with the highest probability. Given HMM λ_i and λ_j, their distance is quantified by the Hellinger distance of corresponding probability distributions, whose square is given by

$$H^2(f(\lambda_i), f(\lambda_j)) = 2\int [\sqrt{f(\lambda_i)} - \sqrt{f(\lambda_j)}]^2 d\mathbf{O}. \tag{3}$$

However, the integral cannot be calculated directly in practice, as we hardly enumerate all possible engaging trajectories in the parameter space. According to the law of large numbers, one possible strategy is involving Monte Carlo sampling which draws a large number of trajectories ($\sum_i T_i \to \infty$) *uniformly* from the space, and the integral can be replaced with a finite sum over the sample space. We argue that this approach is still compute-intensive, because there are no prior constraints on model parameters and the drawn number could be prohibitively large to achieve a good approximation error of order $n^{-1/2}$ [37].

For these reasons, we propose a fast and approximated solution. Assuming underlying true distribution of observations is highly centralized around trajectories observed in our dataset, we can perform sampling on the small observed space to lower computing cost. This assumption is reasonable because user behaviours in similar network and physical environments exhibit high homogeneity [2]. With the increment of sampling number of engaging trajectories, the model distance $H_{i,j}$ could converge to a stable value that lies right close to the true value [37]. In the data experimental part, we exhibit that the model distance can be figured out with a small group of trajectory samples, which is very efficient to evaluate a large group of users.

In the clustering procedure, the distance matrix calculated above is fed into the algorithms to find cluster structures. The family of k-means algorithms are not suitable in our problem, because the centres of generated clusters in each iteration are calculated via Euclidean distance that is very difficult to derive from models. The algorithm of Partitioning Around Medoids [35] endeavours to represent various aspects of the structure of data with k medoids, and thus less interpretive if we change the analysing standard. For these reasons, we choose hierarchical clustering [35] which remains global data characteristics and one can often decide on the optimal value of k through graphical representations.

4. Data-driven analyses and modelling

4.1. Data collection

In this study, we involve two real datasets to profile user engaging behaviour in mobile networks:

Mobile phone dataset. This dataset is collected as the ground truth to evaluate our $\mathcal{AID}$ algorithm. All volunteers from our colleagues (12 students and 3 staff) use smart phones with a light-weight sniffing tool running on the mobile device directly. The collected data (including network packets, application status, location, and user touch actions) are stored locally and uploaded to data server periodically. The collection lasts for four weeks and roughly contributes one trace (recording 5–10 min on-line behaviour) every two days per user. Comparing with the second dataset collected from network links, our client-based data records the right time when users click the screen and forms an accurate baseline to evaluate the behavioural modelling of mobile traffic.

Campus WiFi dataset. This dataset is the primary experimental part to profile mobile user engaging behaviour, which is collected from an operational WiFi mesh network [11] that serves ∼80k campus customers with 2.7k access points. Regarding the scale of this dataset, we leverage Hadoop-based utilities (e.g., Pig Latin) to facilitate data manipulation. Initially, the raw network packets are constructed into flows with a five-tuple (source and destination addresses, source and destination ports, as well as the protocol type). Each flow is recorded with transmission statistics of bidirectional communication and application types via deep packet inspection techniques. Especially, the HTTP flows gain a finer analysis to extract header meta data and delicate performance metrics. Beyond network traffic data, we also record user mobility history leveraging management logs reported by a event-driven WiFi network controller, and each flow is labelled by an associated AP location and user identify. To differentiate the characteristics of mobile and non-mobile groups, we classify each user by combining the software (i.e., the user-agent keyword in HTTP header) and hardware (i.e., the device Organizationally Unique Identifier, OUI) signatures, similar to the method used in [18]. Specifically, as the tablet users tend to stay in a small number of places ($\bar{n} = 2.3$), we classify them into the non-mobile group correspondingly.

To evaluate the data quality, Fig. 3 visualizes the raw network traffic from protocol composition. We can see that the traffic records a continuous picture of user activities over five months from network aspect. In addition, Table 1 gives basic statistics as well. We observe comparable magnitude of both popularity and traffic volume for two groups. However, traffic composition shows difference: mobile flows are dominated by HTTP (82.5%), network management (6.5%) and instant

[5] The presence of covariates $\mathbf{c}$ in conditions is suppressed hereinafter for brief.

Table 1
The breakdown of raw campus WiFi dataset (04/20/2013–09/30/2013).

User types	Population	Traffic		Engaging sessions		
		Byte	Flow	Sessions/user	Bytes/session	Duration/session (s)
Mobile	66.9k	49T	687M	91.3	15.7M	1395
Non-mobile	52.1k	62T	1388M	31.8	106.6M	4200

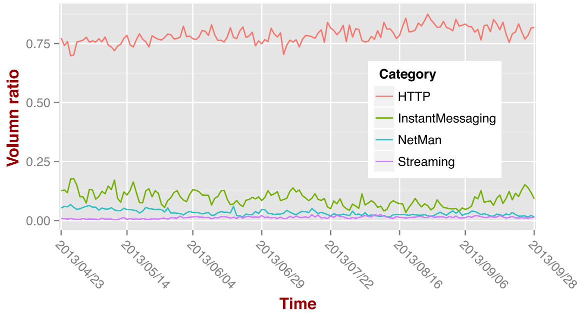


Fig. 3. The protocol time series of our experimental campus WiFi dataset.

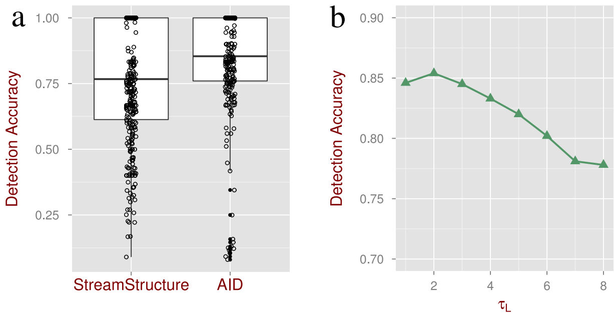


Fig. 4. Evaluating $\mathcal{AID}$ algorithm: (a) A comparison with state-of-the-art StreamStructure algorithm; (b) The estimation of the parameter τ_L used in $\mathcal{AID}$.

messaging (4.3%), while the non-mobile group relies most on HTTP (78.2%), P2P (7.3%) and network management (3.9%). As for the predominance of HTTP traffic in volume, we mainly concentrate on user engaging behaviour observed from this portion of traffic in following analyses.

4.2. Evaluation of $\mathcal{AID}$ algorithm

Our $\mathcal{AID}$ algorithm is evaluated on mobile phone dataset and compared with the state-of-the-art StreamStructure algorithm [26], which was originally proposed to recover web pages from network traffic. The groundtruth is a finite set of identified user real activities (clicks) after time alignment, data filtering and human double-check. As shown in Fig. 4(a), the $\mathcal{AID}$ obtains a precision improvement of $\sim$10% than the StreamStructure algorithm and takes stabler performance over all instances (a smaller converged zone of data points). These improvements are mostly contributed by the logical constraints (recall that more than 70% of HTTP requests are reference-free in mobile phone dataset). Performing $\mathcal{AID}$ algorithm against varying parameter τ_L from 1 to 8 s (Fig. 4(b)), we observe an optimal selection of 1–3 s with close performance; after the peak point ($\tau_L = 2$), the accuracy degrades monotonically. This implies that mobile users are pursuing instant data-consuming experience and active engagement on mobile devices.

4.3. Basic characteristics

4.3.1. User perceived application performance

In this part, we characterize user perceived application performance on differing platforms. In Fig. 5, three representative application categories are exhibited for brevity. We find that email and microblog earn shorter activity duration on mobile devices in spite of smaller perceived throughput. This mainly results from content adaptation and compression in mobile scenarios [16]. From application aspect, music outperforms three categories with perceived bandwidth, and the email activities are more possible to suffer from a longer duration (Fig. 5(a)). Steps in the curve imply that they are probably

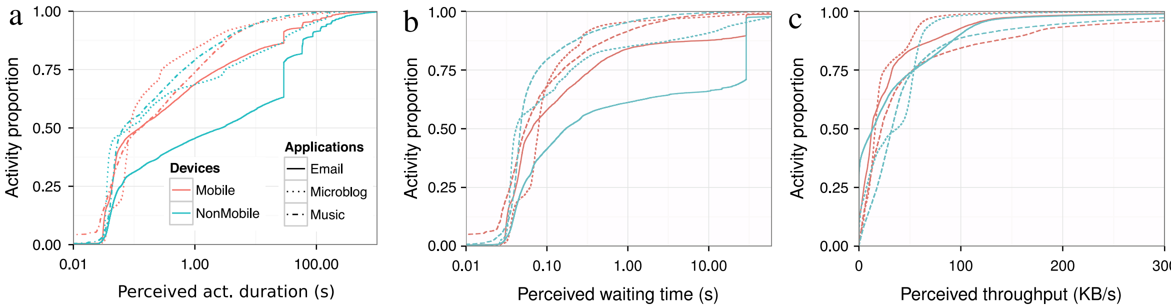


Fig. 5. Visualizing the impact of platforms and application diversity on user-perceived application performance.

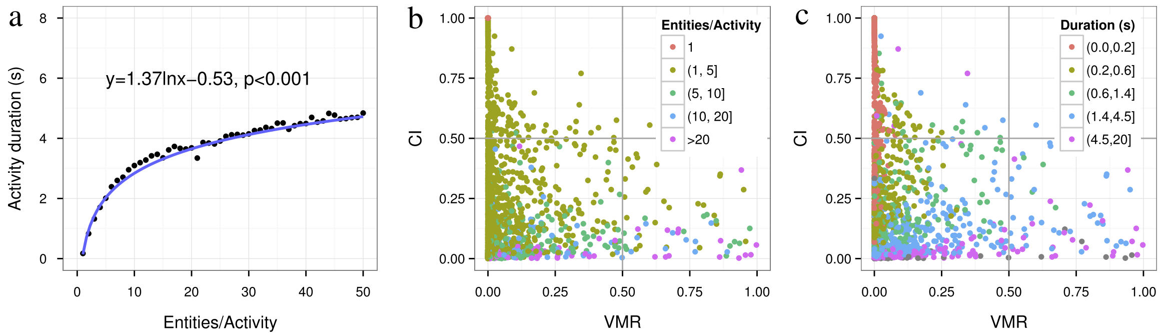


Fig. 6. The proposed concurrency index (CI) captures more detailed information about activity duration than the traditional activity volume metric (the number of entities in an activity).

caused by TCP implementations of client/server libraries. For a rapid verification of our guess, several popular browsers on a laptop are tested and similar timeouts are obtained, i.e., IE-10:60s, Firefox-28:115s, Opera-22:120s. A specialized analysis of mobile timeouts is left for future work. These findings imply that perceived mobile performance results from a mix of platform and application protocol behaviour.

Since the diverse impacts of client and application behaviours, is there still a consistent way to evaluate mobile user perception? In web measurement, the number of embedded objects in a web page is considered the primary factor that impacts web performance [26,20]. Thus the optimization such as caching is explored considerably to share data among web pages. We argue that these techniques are helpful for user activities with large volumes but with limited contribution for those smaller ones; e.g., in Fig. 6(a), the saturated increment within the large range of activity volumes implies a rapid limitation of the traditional content-sharing optimization in reducing activity latency, thus potential opportunities for the smaller range. With our behavioural model, user activities involve a parallelism of requested entities (captured by CI) as well as the balance of response progressing (captured by the variance-to-mean ratio, or VMR, of entity duration). In Fig. 6(b) and (c), we observe that the proposed CI (with VMR) can efficiently reflect subtle difference of activity duration, which implies that the entity relationships make major contributions to delicate perceptual diversity. Practically, backend services with inconsistent quality can be optimized via: (i) guaranteeing consistent qualities of intra-activity services; (ii) adjusting the entity fetching mechanism to obtain a highly converged loading process. Similar optimization strategies at the TCP layer were also addressed in [16].

4.3.2. User engaging behaviour in mobile context

In mobile context, high mobility and ad-hoc data requests usually lead to the deviation of user engaging behaviour. In Fig. 7(a), the individual difference is explored: the distribution of session count for a user can be well fitted by truncated log-normal models ($p < 0.01$). Their high skewness ($\sigma = 3.79 \pm 0.20$ for mobile and $\sigma = 4.14 \pm 0.40$ for non-mobile) implies that a small portion of active users contributes a considerable part of mobile traffic. For the activity number within an engaging session, Fig. 7(b) presents a large gap (with K–S test value $ks = 0.694$) on different platforms and Fig. 7(c) confirms this distinction from another dimension (similarly $ks = 0.642$) where the session duration gets a peak around 12 and 78 min for mobile and non-mobile groups respectively. These behavioural diversities may root in the platform characteristics, i.e., mobile devices are born with high mobility but limited power and computing capability. In a different manner, the visit frequency (Fig. 7(d)) encodes some behavioural nature that reading habits are consistent across multiple platforms ($ks = 0.387$).

Temporal observations of mobile traffic are of great importance for administrative policies and security monitoring. Fig. 8(a) demonstrates the diurnal variation of engaging session duration, where three explicit spikes are observed around {9 h, 12 h, 18 h} that tally with human rest and meal schedules exactly. The relatively smaller bursts of mobile users indicate

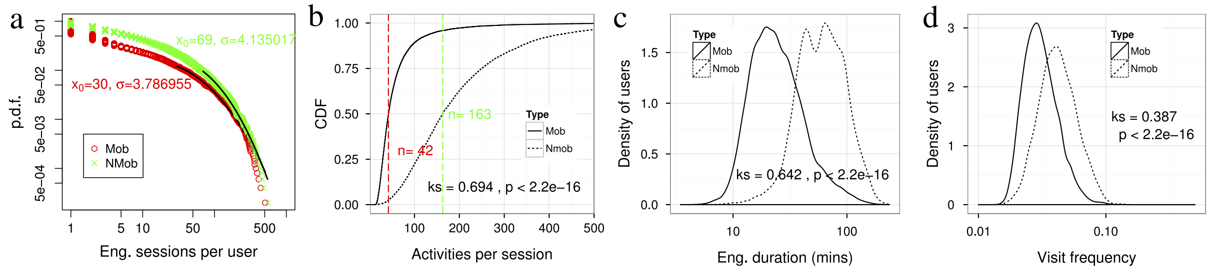


Fig. 7. The diversity of user engaging behaviour concerning platform difference (Mob: mobile, NMob: non-mobile).

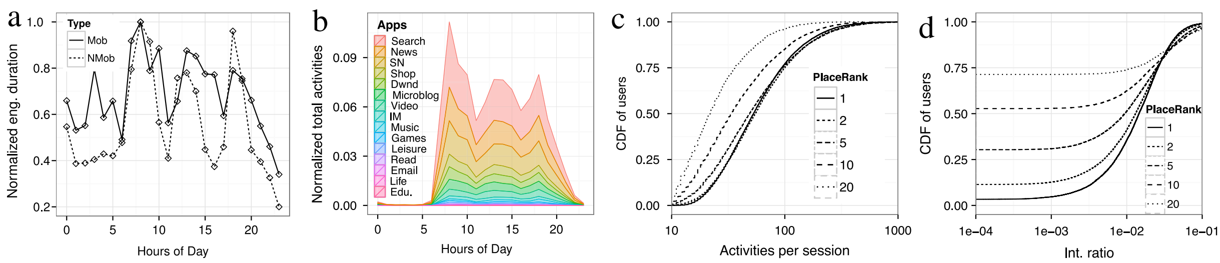


Fig. 8. Contextual impacts (temporal dynamics, application diversity and spatial preference) on mobile engaging behaviour.

the limited spatial constraint on mobile usage. We also check the service semantics[6] as shown in Fig. 8(b). Specifically, heavy hitters are observed with *search, news, social networking, video* and *microblog* categories, while *search* and *video* dominate the left regarding traffic bytes (not shown for brevity). These diurnal patterns and traffic composition suggest inherent use preference in mobile services. As for the dynamic context, another interesting question would be asking: How does user engaging behaviour adapt to locations with different familiarity? Fig. 8(c) gives the distributions of five place ranks. We observe that the number of activities in an engaging session is positively correlated with the location familiarity. This is consistent with our intuition that users tend to engage in longer sessions if they are longer to stay there. Also, we expect the interruption ratio to be lower (users having more patience) at a more familiar place. On the contrary, it is observed that users pay more patience to network fluctuations when they are at a *less* familiar place (Fig. 8(d)). This counter-intuitive observation may be connected with a reduced user expectation of network quality in a foreign area.

4.4. Interactions of engaging behaviour and perceived application performance

4.4.1. Qualitative analysis

Now we attend to profile the interaction between behavioural and perceptual metrics. As for the similar tendency of perceived activity duration and waiting time, we merely illustrate the former for brevity (Fig. 9(a)–(f)). The locally smoothed mean and variance are also presented. Overall, both linear and non-linear correlative relationships are observed between corresponding construct pairs. The engaging session duration takes a truncated increment against the perceived activity duration within $d_{\mathcal{A}} \in (0, 10]$ (Fig. 9(a)), while the interruption ratio takes a monotone increment correspondingly (Fig. 9(b)). This implies that mobile user engaging behaviour is sensitive to performance fluctuations via reducing the activity number in a visit. Meanwhile, the visit frequency gains *insignificant* impacts from perceived activity duration (Fig. 9(c)), and takes a long-ran positive correlation with user perceived throughput.

As addressed in [27], the growth of interruption ratio reports negative signals of user experience. However, we find that the interruption ratio increases with perceived throughput within the lower range (Fig. 9(e)). This counter-intuitive result is confirmed by the heterogeneity of mobile user activities: For small activities (with low averaged perceived throughput due to TCP slow start), the throughput of early-terminated activities is $\sim$1.42 times larger than that of completed ones, which is consistent with our observation that larger perceived throughput leads to higher interruption ratio. To evaluate the impact of early latency of activities on the interruption ratio, we plot the perceived waiting time of both mobile and non-mobile users in Fig. 9(g) and (h) respectively. Although mobile users more probably encounter an interruption,[7] they show more patience against network latency than non-mobile users. For example, there exist striking spikes in a range of [35, 60] for the mobile group and [15, 45] for the non-mobile. These observations are inspiring for mobile developers to optimize user experience with flexible adaptation strategies for network states.

6 We have labelled 15 subtle categories: *downloads, education, email, games, instant message, leisure, life, music, microblog, news, reading, search, social networking, video*. Please note that we separate *reading* from *news* because some e-book applications (e.g., kindle) are really popular on mobile platforms.

7 The 75th percentile of interruption ratio being 6.25% (4.49%) for mobile (non-mobile) users.

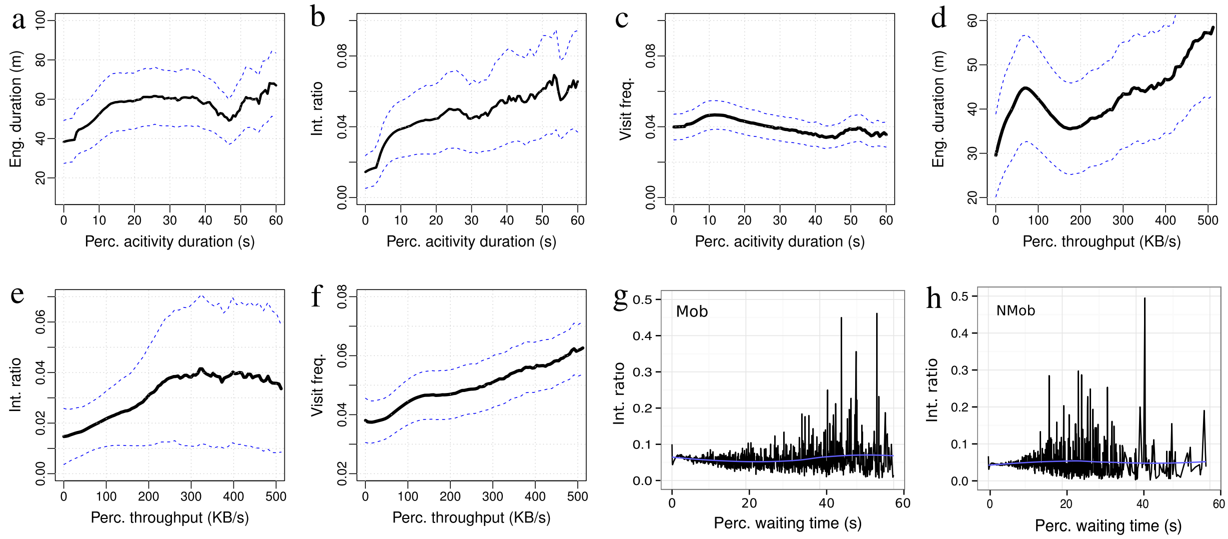


Fig. 9. Qualitative relationships between mobile user engaging behaviour and perceived application performance. Blue slash lines indicate the half variance of behavioural measures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

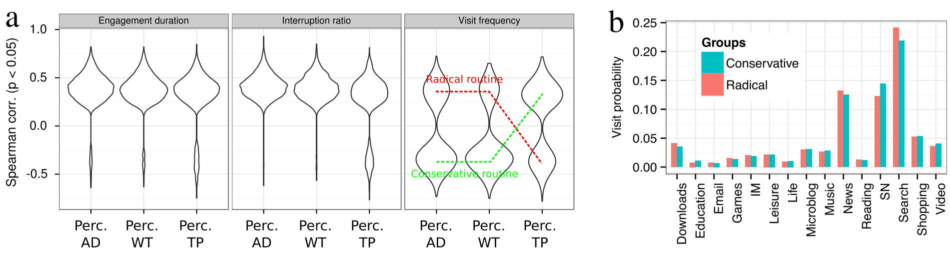


Fig. 10. (a) The distribution of Spearman coefficient ($p < 0.05$) for mobile users ($n \geq 20$). User densities are illustrated by the violin shape. (b) The probability distribution of visiting different applications by users in identified two groups.

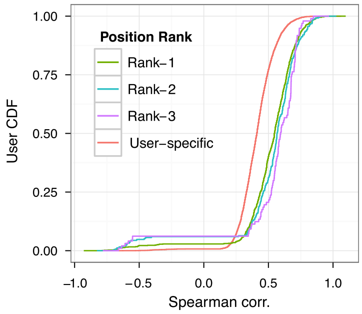
Table 2
Statistics of users with significant correlation ($p < 0.05$).

	Eng. session duration	Int. ratio	Visit freq.
Perc. activity duration	6144 (0.37)	12 590 (1.29)	7139 (0.46)
Perc. waiting time	7099 (0.46)	9562 (0.75)	4438 (0.25)
Perc. throughput	4731 (0.27)	5340 (0.32)	7757 (0.53)

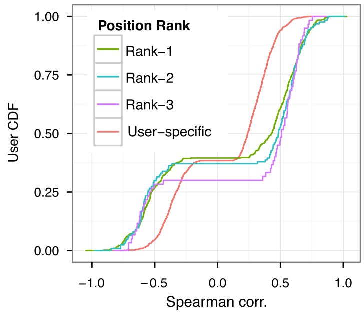
4.4.2. Quantitative analysis

We perform structured correlation analysis on user engaging behaviour ($m_i \in \{D_e, f_v, r_{int}\}$) and perceptual metrics ($m_j \in \{d_{\mathcal{A}}, w_{\mathcal{A}}, b_{\mathcal{A}}\}$) while considering the structured feature set $S = \{usr, app, loc\}$.

User difference. Using methods in Section 3.4, we analyse the distribution of $R_{ij}(\{usr\})$ and find that correlative interactions are really *user-specific*. Fig. 10(a) plots the spread and density of user population for each construct pair. An interesting pattern about two use routines is identified; i.e., most users tend to slow down their visit frequency when experiencing longer latency. There are two candidate hypotheses to explain this phenomenon: (1) the application preference of users, or (2) the inherent distinction between user types. We evaluate the first hypothesis by calculating the probability distribution of visiting applications for two groups of users (Fig. 10(b)). The consistent characteristics across groups help us reject it. Alternatively, we interpret this divergence as two use routines from empirical sense: *conservative* and *radical*. The former group tends to leave the application for other things when the performance is not satisfied and come back when it does, vice versa. This produces the observation that conservative users adapt to network conditions while the radical users give more trials to recover from performance fluctuations. To evaluate the significance of our observations, Table 2 presents concrete user count and odds (in parentheses). These numbers are directive for service providers to optimize performance regarding specific requirements; e.g., optimizing activity duration may reduce the interruption ratio of more than a half portion of users ($\phi = 1.29$), comparing with the optimization of perceived bandwidth mostly benefiting 29% users ($\phi = 0.32$).



(a) IntRatio vs. PercAD.



(b) VisitFreq. vs. PercTP.

Fig. 11. The impact of location familiarity on engaging behaviour ($p < 0.05$). The user-specific distribution without spatial information is present for comparison.

Application diversity. In this part, two distributions, $R(\{app\})$ and $R(\{usr\}|\{app\})$ are considered; the first captures the correlation at the whole population level while the second at an individual granularity. Table 3 presents the application-specific and user–application-specific correlations respectively (cells with significant negative coefficients being coloured by light grey). We find that the visit frequency exhibits more complicated behaviours than the engaging session duration and interruption ratio. The visit frequency of entertainment applications (such as *games, leisure, news*) exhibits completely negative correlation with the perceptual metrics, while the social applications (such as *IM, life, shopping, SN*) have positive correlation with perceived activity duration and waiting time, but negative correlation with perceived throughput. Interestingly, the interruption ratio of *IM, reading and video* illustrates negative correlation ($-0.25 \sim -0.37$) with perceived throughput when comparing with others. This implies different policies should be used when the service provider wants to increase user activeness in their applications.

Another important question is how do the application and user factors interact with each other? Comparing two values in each cell, we observe that the application factor obtains relatively smaller correlation over most observations, while a combination with user factor gives enhanced correlative relationships between user engaging behaviour and perceived application performance. To check the independence of these two factors, we also make a comparison with the user-specific distribution by involving $R(\{usr\})$ and $R(\{app\}|\{usr\})$ and the latter takes larger correlative coefficients than the former across all construct pairs. This suggests that user and application factors strengthen the correlative relationships in an independent but complementary way. However, from the statistics of visit frequency in Table 3, we may derive a conflictual conclusion comparing Fig. 10 about the connection between user groups and their application preference. Actually, this is just a fake confusion as we are viewing the correlative statistics at different scales. In other words, the engaging behaviour of an individual is really possible to be observed with distinctive correlative relationship when we check different parts (i.e., application categories) of the observations.

Location familiarity. The spatial factor is another vital factor on user engaging behaviour because people often show different interests with displacement [2]. Fig. 11 exhibits the impact of location familiarity on user engaging behaviour. Engaging sessions of *microblog* are involved to estimate the correlation. Here we have two basic discoveries: First, a combination of user factor and location familiarity reinforces the correlative effect comparing with single user factor. For example, the median of user-specific distribution of $(r_{int}, d_{\mathcal{A}})$ pair is 0.43 while the counterpart of Rank-1 distribution is 0.51. Second, users at places with less familiarity present stronger correlations, which means that mobile user engagement tends to be more sensitive to perception at a less familiar place. To evaluate our conclusions in a quantitative way, the K–S test is also preformed on user specific and space-constraint distributions. Specifically, we obtain the test statistics of $ks_1 = 0.384$, $ks_2 = 0.472$, $ks_3 = 0.578$ for the correlation of interruption ratio and activity duration, and of $ks_1 = 0.363$, $ks_2 = 0.444$, $ks_3 = 0.513$ for the correlation of visit frequency and perceived bandwidth. More interestingly, by combining Figs. 8(d) and 11(a), we find that mobile users tend to generate less interruption activities (with more patience) at a less familiar place, although the interruption ratio is more relative to activity duration at those places.

4.5. Modelling user engaging behaviour in context

4.5.1. HMM-based modelling of engaging trajectory

In previous sections, we have quantified the characteristics and interactions of mobile engaging behaviour and perceived application performance. We now endeavour to understand user engaging behaviour across multiple sessions. In our experiment, each observation of an engaging trajectory is characterized by behavioural and perceptual metrics $\{D_e, f_v, r_{int}, d_{\mathcal{A}}, w_{\mathcal{A}}, b_{\mathcal{A}}\}$, as well as the contextual variables $\{App, PR, ToD, ToW\}$. As for the inherent personality and diurnal mobility patterns (e.g., commuting), human behaviours show strong regularity in both spatial and applications [10,2]. Fig. 12 exemplifies the application and position distributions of two common users and illustrates the feature entropy across the whole population as well. Despite the diverse number of unique values, the entropy density exhibits significant convergence about mobile user engaging behaviour and the peak distant from simulated random behaviours.

Table 3
Spearman correlations ($p < 0.05$) for specific application (cell left) and user–application combination (cell right).

	Engaging session duration			Interruption ratio			Visit frequency		
	Perc. A.D.	Perc. W.T.	Perc. T.P.	Perc. A.D.	Perc. W.T.	Perc. T.P.	Perc. A.D.	Perc. W.T.	Perc. T.P.
Downloads[a]	0.34 / 0.49[b]	0.25 / 0.46	0.18 / 0.44	0.40 / 0.48	0.32 / 0.46	0.14 / 0.45	−0.01 / 0.26	−0.03 / −0.42	– / −0.49
Games	0.25 / 0.40	0.24 / 0.37	0.18 / 0.35	0.35 / 0.44	0.24 / 0.42	0.17 / 0.40	– / −0.26	0.02 / −0.29	−0.14 / −0.28
IM	0.27 / 0.47	0.23 / 0.45	0.05 / 0.46	0.25 / 0.47	0.25 / 0.47	0.04 / −0.37	−0.02 / −0.40	−0.04 / −0.43	– / 0.33
Leisure	0.18 / 0.38	0.18 / 0.42	0.13 / 0.40	0.28 / 0.46	0.31 / 0.37	0.16 / 0.37	−0.06 / −0.18	−0.04 / −0.20	0 / −0.36
Life	0.29 / 0.52	0.23 / 0.57	0.26 / 0.42	0.35 / 0.48	0.21 / 0.37	0.15 / 0.40	−0.07 / −0.43	−0.11 / −0.50	−0.04 / 0.51
Microblog	0.24 / 0.44	0.20 / 0.40	0.30 / 0.44	0.38 / 0.43	0.31 / 0.42	0.14 / 0.40	0.17 / 0.38	0.12 / 0.35	0.13 / 0.38
Music	0.25 / 0.44	0.21 / 0.43	0.21 / 0.41	0.43 / 0.59	0.35 / 0.53	0.15 / 0.45	0.03 / −0.58	0.05 / 0.02	– / –
News	0.31 / 0.46	0.27 / 0.45	0.20 / 0.43	0.33 / 0.44	0.30 / 0.42	0.18 / 0.41	−0.02 / −0.30	−0.02 / −0.37	−0.02 / −0.39
Reading	0.07 / 0.42	0.11 / 0.55	0.20 / 0.41	0.30 / 0.54	0.26 / 0.41	– / −0.25	−0.05 / 0.43	−0.05 / −0.33	– / −0.66
Search	0.20 / 0.44	0.29 / 0.43	0.27 / 0.44	0.33 / 0.44	0.28 / 0.42	0.21 / 0.41	−0.03 / −0.26	−0.02 / −0.35	−0.01 / −0.35
Shopping	0.31 / 0.50	0.15 / 0.43	0.16 / 0.46	0.33 / 0.49	0.24 / 0.44	0.13 / 0.41	−0.07 / −0.38	−0.11 / −0.40	0.06 / 0.36
SN	0.26 / 0.41	0.24 / 0.38	0.10 / 0.40	0.23 / 0.39	0.17 / 0.38	0.16 / 0.35	−0.03 / −0.37	−0.04 / −0.36	0.08 / 0.33
Video	0.15 / 0.48	0.25 / 0.49	0.24 / 0.49	0.29 / 0.48	0.22 / 0.49	−0.03 / −0.31	−0.02 / −0.23	0.05 / 0.44	0.11 / 0.25

[a] Applications and pairs with failed estimation (e.g., due to insufficient observations) of the real correlation are denoted by '–' in the table.
[b] For the right side of each cell, the median of zero means that roughly two halves of the whole population have opposite correlative tendencies.

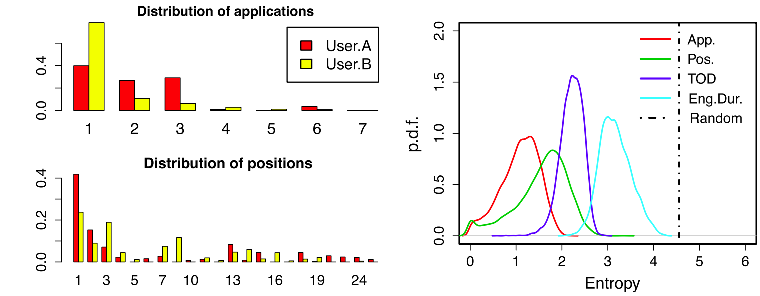
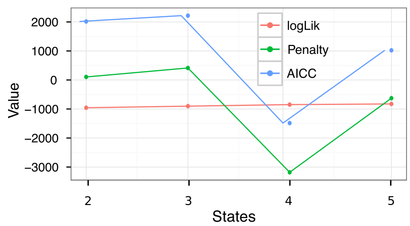
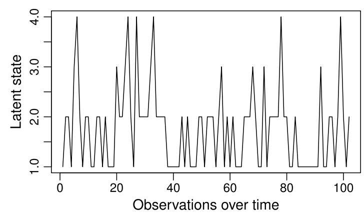


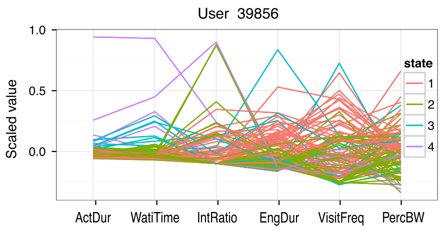
Fig. 12. The distribution of engagement observation features for two common users, as well as the feature entropy across all users. The dot–slash line simulates random user behaviours in mobile networks.



(a) Model selection.



(b) Engaging state dynamics.



(c) Features over estimated states.

Fig. 13. Illustration of model selection, series of hidden state, and feature distributions of an engaging trajectory (U39856).

The engaging trajectory of a user is derived from a five-month period. To make representative conclusions, we filter out users who generate less than 20 engagement sessions (one per week on average). In the learning process, an HMM model is learned for each user and contextual factors are involved as covariates in state transition probabilities, as described in Section 3.5. Fig. 13 illustrates the main parameters for deriving a functional HMM model. In the model selection procedure, the goodness-of-fit is well balanced by the complexity penalty, as integrated in the AICC measure (Fig. 13(a)). By setting the principle state threshold $p_0 = 0.3$, the states 1 and 2 are detected as principle ones for user U39856, while the states 3 and 4 are visited occasionally (Fig. 13(b)). To further inspect the composition of different user states, we plot the feature distribution (ordered by feature skewness) in a parallel coordinates graph (Fig. 13(c)). We can see that the activity duration and expected waiting time own relatively smaller variance than the non-principle states (3 and 4) who cover most portion of the abnormal observations. For the left four dimensions, the principle state 2 mostly owns smaller metric values and higher convergence than state 1. Specifically, the median of engaging session duration is 27.5 min for state 1 and 13.2 min for state 2.

Conclusively, we obtain two general observations over the whole population of experimental users: First, the engaging trajectory is dominated by a small number of engaging states; second, despite the heterogeneity of feature skewness, individual engaging state narrows the feature range and captures the stable part of user engaging behaviour.

4.5.2. Engaging behavioural pattern extraction and interpretation

With the HMM representation of engaging trajectory, we perform a model-based clustering mechanism to uncover the significant engaging patterns over a large population. As a basic ingredient, model distance calculation is compute-intensive if we address the integral solution given by Eq. (3). Here we evaluate our propose fast approximation of model distance by a finite sampling of engaging trajectories from observation space. Fig. 14(a) illustrates the convergence of model distance $H_{i,j}$ against the increase of trajectory count N for U39856 and U39874. The model distance becomes converged near $N = 100$ with the mean observations $\bar{T} * 100 \sim 2^{12}$ ($\bar{T} = 50.87$), which is consistent with Dugad et al.'s observation [37] that the distance between two HMMs converges to a stable value as the accumulated observation length increases to an order of

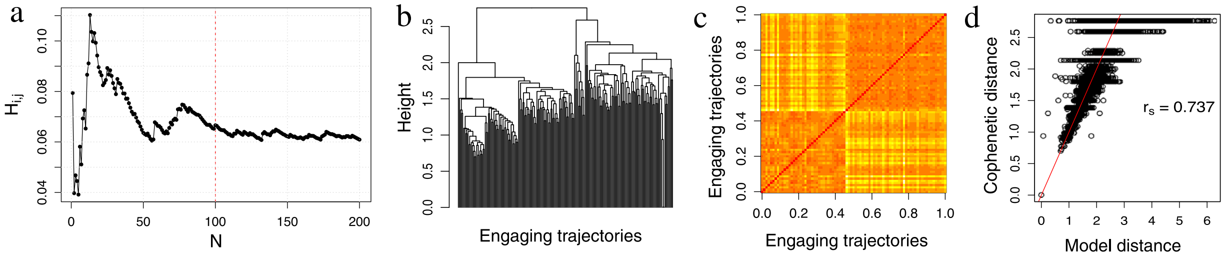


Fig. 14. HMM distance calculation and model-based clustering (200 users): (a) Investigating the impact of trajectory count N on estimated model distance (U39856 vs. U39874); (b) the dendrogram of engaging trajectories; (c) model distance matrix (scaled); (d) the relationship of model and cophenetic distances.

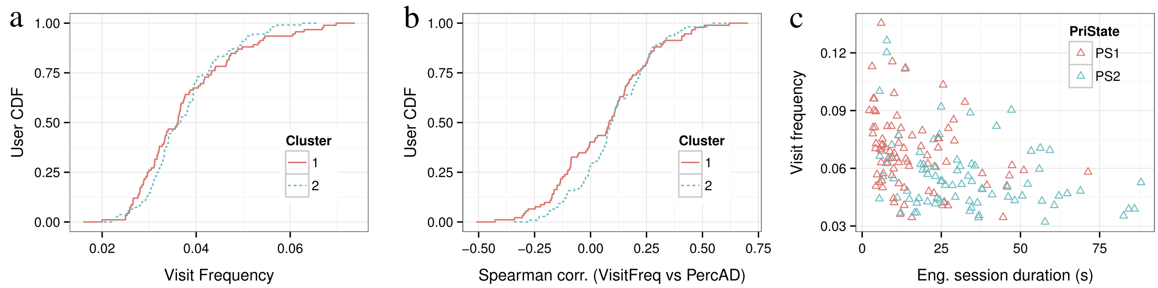


Fig. 15. Investigation of trajectory clusters: (a) the distribution of behavioural metrics in different clusters; (b) the distribution of behavioural and perceptual metric correlation across clusters; (c) the distribution of behavioural metrics for different principle engaging states.

$2^{10} \sim 2^{12}$. This criteria is also supported by other users' trajectories in the experimental dataset, and hence we use $N = 100$ to estimate the model distance in our study.

Afterwards, we feed the pair-wise distance matrix into the hierarchical clustering algorithm [35] to mine user engaging patterns. Fig. 14(b)–(d) exhibits the hierarchical clustering of 200 users randomly selected from the WiFi dataset. As shown in Fig. 14(b), the dendrogram encodes data structures in a hierarchical tree and the cophenetic distance measures the height of dendrogram where two branches merge into a single one. By varying the number of clusters k, the dendrogram can be cut into different number of branches. We manually change k from 2 to 8 and visually select the best option with $k = 2$ (see Fig. 14(c)). Finally, the Shepard diagram also confirms the clustering quality with a high correlation ($r_s = 0.677$) of model distance and cophenetic distance measures (Fig. 14(d)).

The mined clusters mean that some underlying patterns are governing user engaging behaviour in mobile networks. In order to interpret these patterns, we also extract the principle engaging states for each user with $p_0 = 0.3$. As shown in Fig. 15(a), the distribution of visit frequency exemplifies the individual metrics in different clusters. We find that there are no any clues about the underlying patterns when considering a single construct. Then we consider the interactions between the behavioural and perceptual metrics. All construct pairs are evaluated and an example of visit frequency and perceived activity duration in Fig. 15(b) suggests that the correlation relationships do not convey any significant difference either.

However, when we involve the principle engaging states in user engaging behaviour (Fig. 15(c)), some interesting patterns occur: the user engaging behaviour in mobile network mainly switches between two conditions, i.e., long engaging session duration with small visit frequency (e.g., PS2) and vice versa. This implies that the principle engaging states only dominate user behaviours temporally, but also encode some behavioural patterns in a large group of users. To verify this hypothesis, we present the distribution of metrics for different principle engaging states and clusters (Fig. 16). The visit frequency is merely shown because of a similar observation for other behavioural metrics. We find that the interaction between principle engaging states makes a significant distinction between clusters. The state consistence across different users within a cluster also implies the regularity of user engaging behaviour in mobile networks. Specifically, the visit frequency regarding perceived waiting time behaves in a distinctive manner for two principle engaging states in cluster 1, but shows similar trends in the other. These modelling results are also confirmed by our empirical knowledge in reality. For example, assuming a stable use preference in a person, his engaging behaviour is mainly impacted by a small number of physical states (such as at work or home) in the real world [2]. When the physical states are similar, the user tends to show similar mobile engaging behaviour, and vice versa. Conclusively, the principle engaging states encode the regularity of user engaging behaviour while the cluster uncovers underlying engaging patterns across a large population. These cluster properties are beneficial for network operators to implement personalized service provision and revenue strategies. From a diagnostic perspective, identification of the abnormal principle engaging states or behavioural patterns becomes a new way to optimize user experience in a transparent way.

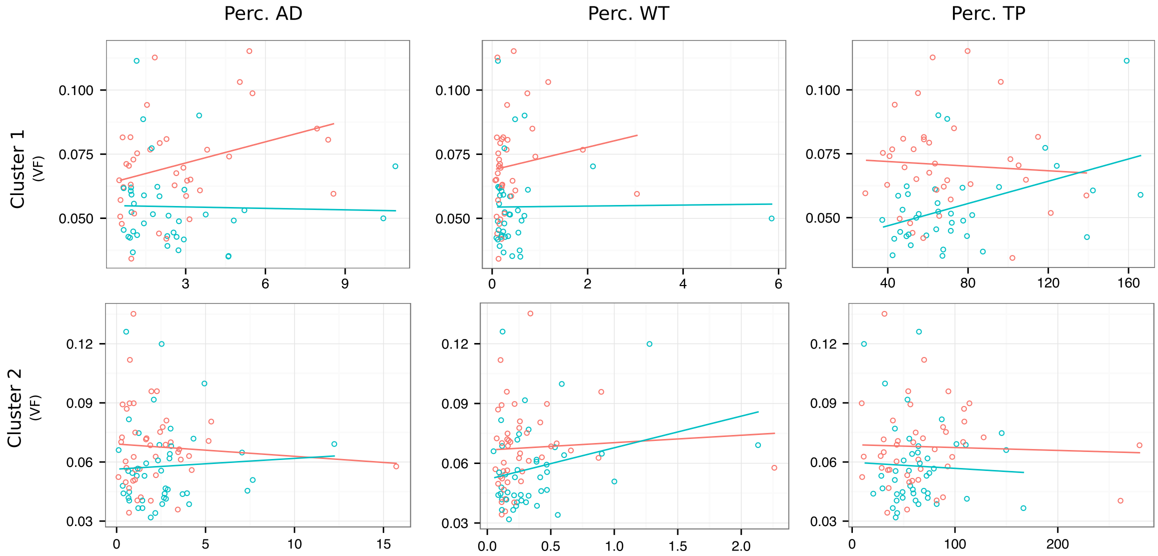


Fig. 16. Demonstration of the interaction between principle engaging states (red for PS1 and green for PS2) when considering behavioural and perceptual metrics in different clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Discussions and conclusions

In this paper, a passive and large-scale profiling of user engaging behaviour is performed, mainly focusing on its interactions with user-perceived application performance in real context. Our primary contribution is extending the conceptual engagement frameworks with an operational measurement and modelling. The proposed traffic model enhances our capability to understand mobile traffic at a behavioural level than the traditional transport layer. Especially, the embedded measure of concurrency index is helpful to uncover the performance details of small activities. Despite the divergence of mobile platforms and contextual factors, some aspects of engaging behaviour such as reading habits, diurnal patterns and location preference are still consistent across multiple users. With the assessment of perceived application performance, mobile users are observed sensitive to performance fluctuations by say reducing the activity number in a visit. Interestingly, user engaging behaviour tends to be more sensitive to the application quality at a less familiar place. In the modelling of engaging trajectory, significant convergence about mobile engaging behaviour is observed. By mining behaviour patterns with HMM-based clustering, we find that user engaging behaviour is mainly impacted by the interactions of principle engaging states.

Our insights and discoveries about user engagement dynamics imply wide applications in practice. The proposed behavioural model of mobile traffic paves a way towards *behaviour-* and *context-aware* traffic generators in network simulation and optimization strategies for the design of next-generation mobile operating systems. Bridging the gap between user engaging behaviour and perception, our analysis and modelling make it possible to develop an objective quantification standards of user experience in mobile networks. Although our experiments are conducted in campus WiFi networks, the structured correlation analysis and modelling of user engagement can easily be extended to other scenarios such as 3G and LTE networks [9].

There are also some limitations in our methodology. While we are motivated by uncovering interactions between user engaging behaviour and perception, there lies a potential assumption that it is a cognitive process that mobile users will adapt their engaging behaviours to perceptual quality and specific context, consciously or unconsciously. We hardly address this hypothesis directly because there are a large number of factors impacting user behaviour, some of which can be measured objectively and some cannot, e.g., the unconscious part. Although some supportive evidences are obtained for the validation of underlying assumption leveraging multiscale analysis and modelling, further confirmative analysis may require a cooperation of other metrics (e.g. physiological sensors [8]) or subjective measurements.

As the future work, we are also interested in developing an objective behaviour-oriented index of user experience in the future, which obtains practical benefit to diagnose the problem of mobile networks at a large scale and optimize network resources by considering user engagement in mobile services. Moreover, cooperating cross-domain metrics to investigate the impact of potential factors on user engaging behaviour is another interesting direction for the future.

Acknowledgements

The authors would like to thank the support of Network and Information Center, Shanghai Jiao Tong University in collecting the experimental datasets. We would also like to thank Dr. Xuan Luo, OMNILab members, and other sincere anonymous reviewers for the improvement of this paper. The analysing code and datasets used are free to access from http://omnilab.sjtu.edu.cn/beperf.

References

[1] A. Febretti, F. Garzotto, Usability, playability, and long-term engagement in computer games, in: CHI'09 Extended Abstracts on Human Factors in Computing Systems, 2009, pp. 4063–4068.
[2] I. Trestian, S. Ranjan, A. Kuzmanovic, A. Nucci, Measuring serendipity: connecting people, locations and interests in a mobile 3g network, in: Proc. of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, 2009, pp. 267–279.
[3] H.L. O'Brien, E.G. Toms, What is user engagement? A conceptual framework for defining user engagement with technology, J. Am. Soc. Inf. Sci. Technol. 59 (6) (2008) 938–955.
[4] E.T. Peterson, J. Carrabis, Measuring the Immeasurable: Visitor Engagement, Web Analytics Demystified, 2008.
[5] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, H. Zhang, Understanding the impact of video quality on user engagement, ACM SIGCOMM Comput. Commun. Rev. 41 (4) (2011) 362–373.
[6] UK Report, Engagement 3.0: a new model for customer engagement, Tech. rep, 2014.
[7] A. Wang, Advertising engagement: A driver of message involvement on message effects, J. Adv. Res. 46 (4) (2006) 355.
[8] S. Attfield, G. Kazai, M. Lalmas, B. Piwowarski, Towards a science of user engagement (position paper), in: WSDM Workshop on User Modelling for Web Applications, 2011.
[9] A. Gember, A. Akella, J. Pang, A. Varshavsky, R. Caceres, Obtaining in-context measurements of cellular network performance, in: Proc. of the 2012 ACM Conference on Internet Measurement Conference, 2012, pp. 287–300.
[10] M. Afanasyev, T. Chen, G.M. Voelker, A.C. Snoeren, Analysis of a mixed-use urban wifi network: when metropolitan becomes neapolitan, in: Proc. of the 8th ACM SIGCOMM Conference on Internet Measurement, 2008, pp. 85–98.
[11] W. Gong, X. Chen, S. Qiang, Y. Jin, Trajectory pattern change analysis in campus WiFi networks, in: Proc. of the Second ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, 2013, pp. 1–8.
[12] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, H. Zhang, Developing a predictive model of quality of experience for Internet video, in: Proc. of the ACM SIGCOMM 2013 Conference on SIGCOMM, 2013, pp. 339–350.
[13] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong, A. Dey, Factors influencing quality of experience of commonly used mobile applications, IEEE Commun. Mag. 50 (4) (2012) 48–56.
[14] M.Z. Shafiq, J. Erman, L. Ji, A.X. Liu, J. Pang, J. Wang, Understanding the impact of network dynamics on mobile video user engagement, in: The 2014 ACM International Conference on Measurement and Modeling of Computer Systems, 2014, pp. 367–379.
[15] R.E. Nisbett, T.D. Wilson, The halo effect: Evidence for unconscious alteration of judgments, J. Pers. Soc. Psychol. 35 (4) (1977) 250–256.
[16] J. Huang, Q. Xu, B. Tiwana, Z.M. Mao, M. Zhang, P. Bahl, Anatomizing application performance differences on smartphones, in: Proc. of the 8th International Conference on Mobile Systems, Applications, and Services, 2010, pp. 165–178.
[17] X. Chen, R. Jin, K. Suh, B. Wang, W. Wei, Network performance of smart mobile handhelds in a university campus WiFi network, in: Proc. of the 2012 ACM Conference on Internet Measurement Conference, 2012, pp. 315–328.
[18] A. Gember, A. Anand, A. Akella, A comparative study of handheld and non-handheld traffic in campus wi-fi networks, in: Passive and Active Measurement, Vol. 6579, 2011, pp. 173–183.
[19] ITU-t p.800. methods for subjective determination of transmission quality - series p: Telephone transmission quality; methods for objective and subjective assessment of quality.
[20] J. Lehmann, M. Lalmas, E. Yom-Tov, G. Dupret, Models of user engagement, in: User Modeling, Adaptation, and Personalization, 2012, pp. 164–175.
[21] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert, N. Veyrat-Charvillon, Mutual information analysis: a comprehensive study, J. Cryptol. 24 (2) (2011) 269–291.
[22] T. Warren Liao, Clustering of time series data—a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.
[23] G.M. Jenkins, A.S. Alavi, Some aspects of modelling and forecasting multivariate time series, J. Time Series Anal. 2 (1) (1981) 1–47.
[24] B. Mah, An Empirical Model of HTTP Network Traffic, Vol. 2, 1997, pp. 592–600.
[25] H.-K. Choi, J.O. Limb, A behavioral model of web traffic, in: Proc. of Seventh International Conference on Network Protocols, 1999, pp. 327–334.
[26] S. Ihm, V.S. Pai, Towards understanding modern web traffic, in: Proc. of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, 2011, pp. 295–312.
[27] D. Rossi, M. Mellia, C. Casetti, User patience and the web: a hands-on investigation, in: Global Telecommunications Conference, Vol. 7, 2003, pp. 4163–4168.
[28] M.Z. Shafiq, L. Ji, A.X. Liu, J. Wang, Characterizing and modeling Internet traffic dynamics of cellular devices, SIGMETRICS Perform. Eval. Rev. 39 (1) (2011) 265–276.
[29] M. Zwick, M.S. Johnson, An overview of reconstructability analysis, in: Proc. of 12th International World Organization of Systems and Cybernetics and 4th International Institute for General Systems Studies Workshop, 2004, pp. 877–905.
[30] E. Schisterman, K. Moysich, L. England, M. Rao, Estimation of the correlation coefficient using the Bayesian approach and its applications for epidemiologic research, BMC Med. Res. Methodol. 3 (1) (2003) 5.
[31] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, in: Proceedings of the IEEE, 1989, pp. 257–286.
[32] V. De Fonzo, F. Aluffi-Pentini, V. Parisi, Hidden Markov models in bioinformatics, Curr. Bioinform. 2 (1) (2007) 49–61.
[33] M. Xu, M. Golay, Survey of model selection and model combination, published January 1.
[34] C.M. Hurvich, C.-L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (2) (1989) 297–307.
[35] L. Kaufman, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley-Interscience, 2005.
[36] R. Beran, Minimum hellinger distance estimates for parametric models, Ann. Statist. 5 (3) (1977) 445–463.
[37] R. Dugad, U. Desai, A tutorial on hidden Markov models, Indian Institute of Technology (1996) 16.