Check for updates

EDUM: Classroom Education Measurements via Large-scale WiFi Networks

Mengyu Zhou[†], Minghua Ma[‡], Yangkun Zhang[†], Kaixin Sui[‡], Dan Pei[‡]*, Thomas Moscibroda[§]
[†]Institute for Interdisciplinary Information Sciences, Tsinghua University
[‡]Department of Computer Science and Technology, Tsinghua University
[‡]Tsinghua National Laboratory for Information Science and Technology (TNList)
[§]Microsoft Research

ABSTRACT
Behavior in classroom-based courses is hard to measure at large-scale. In this paper, we propose the **EDUM** (EDUcation Measurement) system to help characterize educational behavior through data collected from WLANs (WiFi networks) on campuses. EDUM characterizes students' punctuality (attendances, late arrivals, and early departures) for lectures using longitudinal WLAN data, and further characterizes the attractiveness of lectures using mobile phone's interactive states at minute-scale granularity. EDUM is easy to deploy and extensible for new types of data. We deploy EDUM at Tsinghua University where ∼700 volunteer students' data are measured during a 9-week period by ∼2,800 APs and two popular mobile apps. Our results show that EDUM makes it possible to obtain large-scale observations on punctuality, distraction and study performance, and quantitatively confirm or disprove numerous assumptions about educational behavior.

ACM Classification Keywords
H.1.2 Models And Principles: User/Machine Systems; I.5.2 Pattern Recognition: Design Methodology—*Pattern analysis*; J.1 Administrative Data Processing: *Education*

Author Keywords
Classroom Education Measurements; WiFi Networks; Lecture Punctuality; Course Attractiveness; Mobile Devices.

INTRODUCTION
As the major and formal way for compulsory and higher education, traditional classroom based courses are still irreplaceable for students. In recent years, its on-line counterpart MOOCs (Massive Open On-line Courses) provide a brand new opportunity to measure education at large scale with fine-grained web user behaviors [4]. In contrast, measuring the effectiveness of traditional classroom teaching is notoriously hard. Most studies on classroom education are still done through traditional methods by recruiting a small number of volunteers, or conducting costly, intrusive and sometimes subjective observations, surveys and tests. For example, in [26] smartphone sensing data collected from 30 undergraduate students of a course are used to inspect traditional education performance; while surveys, small-scale tracking and lab experiments are common in educational psychology research [9]. In order to better understand classroom education of a dense population and to compare with other forms of education, large-scale in-situ measurements of accurate classroom behavior are essential.

With wireless technologies – especially 802.11 WiFi networks (WLANs) increasingly becoming ubiquitous [7], and mobile devices becoming widely carried by students [19], there is a tremendous opportunity to gain a deeper insight into behaviors in classroom courses. First, lecture attendance can be derived from mobility traces of the devices. The mobility of carry-on mobile devices is a good approximation to the mobility of their owners, and can thus help us estimate re-occurring patterns in traditional courses. This is especially true given the ever-dense deployments of access points (APs) on campuses, yielding richer traces, better coverage and finer details to derive the mobility of the dense population on campus. Second, with the ease of accessing Internet through WLAN during lectures, the usage of wireless mobile devices has become a common sign for student distractions. To understand distractions during courses, Internet accessing traces from WLAN and sensor data from smartphones can lead to new measurable indicators.

In this paper, we realize the above ideas in practice, and use WiFi measurements and mobile data to study the lecture punctuality (*i.e.*, skip, late arrival, early departure) and attractiveness at large-scale. Through the WLAN and mobile data, we determine *whether* a student attended a registered course, *from when to when*, and *how much* the student is using his smartphone during the lecture. To the best of our knowledge, this problem has not been previously tackled in depth and scale. Several non-trivial challenges exist: First, educational ground truths — *e.g.*, the schedule (classroom venue and time span) of courses and their registered students — are fundamentally hard to collect. Not all school offices and

*Dan Pei is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
© 2016 ACM. ISBN 978-1-4503-4461-6/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971657

Table 1. Data Sources and Intermediate Measurements of EDUM.

	Measurements	Data sources at Tsinghua University	Example alternative data sources
(1)	The mapping from a *sid* (student ID or campus account) to his smartphone (identified by its MAC address)	Crowd-sourcing TUNet and TUNow	Retrieved from web-based authentication
(2)	The registered courses (with coarse-grained schedule) of a *sid*	Crowd-sourced from TUNow	From university registration office
(3)	AP closeness	Derived from SNMP polls	
(4)	A smartphone's location at a specific time (mobility)	Derived from (3)+SNMP polls and traps	Vendor provided indoor localization services
(5)	A course's venue (as WiFi RSSI fingerprint)	Derived from (1)+(2)+(3)+(4)	
(6)	A course's fine-grained schedule	Derived from (1)+(2)+(3)+(4)+(5)	
(7)	A smartphone's ON/OFF states at a specific time (usages)	TUNet	Other usages, *e.g.* WiFi traffic statistics from SNMP polls

students are willing nor able to provide data easily. Second, using WLAN data to determine a course's venue and whether a student is at the scheduled location draws another challenge. Typically on campuses, there exists neither a mapping from physical locations to WiFi data, nor (usually expensive and immature) indoor localization services.

To tackle the above challenges, we propose the **EDUM** (EDUcation Measurement) system integrated for schools, institutes and universities that have public WLAN infrastructures. EDUM addresses the challenges through either data analyses (*e.g.*, automatically detect course schedule) or crowd sourcing (*e.g.*, inquiring students to contribute timetables), and outputs educational metrics and reports. We deploy EDUM on the large campus of Tsinghua University where 2,786 APs provide essentially complete coverage across a diverse set of 114 buildings. Our crowd-sourcing mobile apps TUNet and TUNow also attract thousands of users. From November 2015 to January 2016, WiFi traces of ~700 mobile devices (out of 201,230 appeared client devices on the campus) of mobile app volunteers are tracked by EDUM using the campus WLAN.

By analyzing outputs of EDUM, numerous interesting findings are observed, such as: 1) Attendance ratio and late arrival ratio to courses both show that Wednesday is the most hard-working day. 2) Class attendance is at its highest in the morning, and gradually drops as the day progresses. Meanwhile, fewer students arrive late to classes as the day progresses. 3) The more years a student stays at school, the lower his/her attendance ratio becomes, and the more frequently s/he arrives late to classes or leaves early from classes. Also, the ratio of "night owls" in the 2^{nd} and 4^{th} year is higher than those among 1^{st} and 3^{rd} year students. 4) On average students with higher GPA attend class more. However, they are also more likely to be late compared to low-performance students. 5) Students are more easily distracted as the day progresses. Device usage is highest at the beginning of a lecture, then drops, and then slowly increases as the lecture progresses.

In summary, our main contributions are:

- We design a scalable, non-intrusive, extensible and easy-to-deploy classroom education measurement system EDUM.
- To the best of our knowledge, the deployment of EDUM at Tsinghua is one of the largest-scale classroom education measurements via WLAN on a densely-populated campus.
- Multiple punctuality and attractiveness metrics are derived from WiFi and mobile data to characterize courses and the behavior of students.
- New observations and confirmation of some common senses are done by analyzing the metrics and correlating them with other properties of courses and students.

DATA COLLECTION AND MEASUREMENTS

As discussed in the previous section, a fundamental task for EDUM is to determine whether and how a student attended a registered lecture, based on which more metrics can be derived. To this end, at a high level, EDUM needs to conduct several intermediate measurements as listed in Table 1 – which also lists the data sources and their alternatives. Using the data in (1), (2), (4), (5), (6), we can measure the lecture punctuality; by adding the data in (7), we can measure lecture attractiveness; (3) AP closeness is an automatically generated metric to filter out the noise in the raw device RSSI.

In the rest of this section, we first briefly overview the three primary data sources used in our deployment of EDUM as shown in Table 1: the readily available SNMP polls and traps of the existing operational WLAN, and two crowd-sourcing mobile Apps TUNet and TUNow. We then present some necessary data details for each of the measurements in Table 1, except for (5) and (6) whose details will be discussed later in §Course Schedules with Students' Mobility.

Overview of Data Sources

From November 2015 to January 2016, 11 weeks of data (including 9 normal weeks and 2 exam weeks out of 18 weeks of the whole autumn semester) are collected from WLAN and TUNet. In this paper we mainly focus on the observation period of the 9 normal weeks. Non-time-sensitive static data (*e.g.* device MAC addresses and course time tables) are also collected from both TUNet and TUNow.

SNMP Polls and Traps

The campus of Tsinghua University covers an area of $\sim 4.4 km^2$ on which ~45,000 students and ~12,000 faculty and staff members are living. By January 2016, there are 2,786 Cisco enterprise APs in 114 buildings (9 of them are dedicated classroom buildings while tens more are department buildings which also have some classrooms) on the campus, providing a dense deployment in most areas. At peak, there are ~20,000 devices concurrently connected to the campus WLAN. The total number of unique devices surpass 60,000 each day, which means on average everyone uses at least one wireless device. WLAN data, namely polled SNMP (Simple Network Management Protocol) objects and SNMP trap messages, from all APs are provided by the network administrators. With the fast expansion of WiFi infrastructures, these data are readily available at the wireless controllers of most vendors [22, 28].

Mobile Apps: TUNet and TUNow

In addition, our mobile client WLAN tool app **TUNet** (Tsinghua University Network, developed by a student interest group led by the authors) has been installed on more than 8,600 Android devices and 6,500 iOS client devices until

January 2016. (The first Android version was released in October 2013, the first iOS version was released in June 2015.) TUNet helps users manage their network account on their mobile phones, and also automatically login onto the campus WLAN in the background for a smooth Internet experience.

TUNow (Tsinghua Now, developed by the same team) is another mobile app which helps students view their course announcements and homework at the E-learning platform of Tsinghua University. GPA calculator, course timetable viewer and a few other plugins are also provided in the app. Since the first Android release in December 2015, it has been installed on more than 1,200 smartphones until March 2016.

(1) The mapping from a *sid* to smartphone MAC address

To track the students and conduct further analysis, we need to fill the gap between device unique IDs — such as MAC (media access control) addresses from WLAN and mobile data — and student identities such as campus accounts or student IDs. This can be done in multiple ways. *E.g.*, for WLAN that requires login with account, network operators can easily track the mapping between MAC addresses and campus accounts. For schools and universities that have mobile apps based on campus accounts, crowd-sourcing is another solution.

At Tsinghua, we crowd-sourced the mapping between 2,483 MAC addresses and 2,363 campus accounts from volunteers of our popular apps TUNet and TUNow. There are two possible problems to determine the mapping through mobile data: 1) More and more Android and iOS devices block the access to MAC address for mobile apps [1]; 2) Multiple accounts can be used duplicately on multiple devices, which no longer yield one-to-one relation. For the two problems, we adopt the solutions in [28]: 1) Develop a web API for devices to query their MAC address through SNMP according to the SSID and BSSID of connected AP (of campus WLAN) and the obtained IP address; 2) Assign each MAC address to the most frequently used account on the device. Thus for a student that add new devices during the observation period, old and new devices will all be mapped to the same account.

Together with the mapping, from the volunteers we also collect several personal attributes attached to campus accounts — gender, grade, department, class (undergraduates are grouped into classes for each department), category (undergraduate, master and PhD students), living apartment and dorm. Attributes like these are useful but not necessary for deeper analysis on education metrics produced by EDUM.

(2) Registered courses (with coarse schedule) of a *sid*

Besides the device to account mapping, for EDUM to work, we need additional information about the attended courses of the sampled account and the basic schedules of the courses. A direct solution is to get detailed course schedules and accounts of a sampled population of each course from the school office. At Tsinghua, we adopt another approach: we crowd-source course timetables of volunteers from the TUNow app. By March 2016, a total of 723 students contributed their course timetables of the autumn semester to us. In this paper, we focus on the tracking results — mobility (4) and usages (7) — of the smartphones of these 723 accounts.

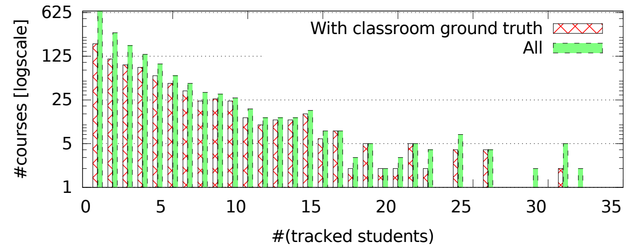


Figure 1. Number of Courses *v.s.* Number of Tracked Students.

From the school website, we crawled public curricula information of all 4,233 courses of autumn semester. Additional course information (not necessary in EDUM) including name, teacher, department, capacity, credit, category, *etc.* are thus available for further analysis of education metrics. The 723 volunteers of TUNow registered to 1,721 different courses out of the 4,233. On average each student takes about 10 courses. As shown in Fig. 1 (“All” rectangles), we track ≥ 5 students in 444 courses, and ≥ 10 students in 179 courses.

(3) AP closeness

Before deriving (4) mobility, (5) course venue and (6) schedule, we first introduce the AP closeness metric as defined in [28]. As we will discuss in upcoming sections, the mobility of devices and course venue are derived through WiFi data. In other words, locations are represented and compared through RSSI fingerprints. To map real-world location to virtual WiFi fingerprints, to smooth the flucationation in raw RSSI data, and to avoid the costly and prone-to-error process of manual labelling, we need an automatic generated metric to characterize how close are the nearby APs.

The AP closeness metric is calculated based on the fact that nearby APs can hear probe request scans of individual devices at almost the same time consistently over a long period of time. The more scans of common devices can be heard simultaneously by two APs, the closer these two APs are. From probe scan history of all devices in SNMP polls, a closeness threshold is chosen to determine whether two APs are “close” based on the overall distribution [28].

(4) Mobility: a smartphone’s location at a specific time

To track regular indoor events, mobility information of each device is useful [12]. The mobility of a device can be used to derive the lecture punctuality behaviors of its owner.

There are multiple ways to derive mobility traces of devices. One straightforward approach would be to employ indoor WiFi localization techniques [15, 27] to locate a student’s position, and compare it with the course venue. However, indoor positioning techniques are still not mature and typically not available in university campus.

In fact, to infer whether a student attends a lecture, we only need to know if his/her devices are *close* (as measured by the WiFi data) to the lecture venue (also as measured by the WiFi data) at the scheduled time. It is quite rare that a student stays around the APs near the course location but does not actually attend the class. Thus we adopt a relatively coarse but simple algorithm instead of excessively accurate, complex and costly localization methods.

Labelling-free Mobility Detection

With the assumption that no location labels (*e.g.* 3D coordinates, nearby rooms, manual fingerprinting) are available on APs (since they are hard to collect and prone to variations), we utilize an easy-to-adopt and cost-effective mobility detection algorithm [28] which is based on WiFi traces from relatively dense-deployed APs.

Similar to traditional *syslog* events [13] which report the AP association process of client devices, SNMP traps messages are sent in UDP packages in real-time. Prior studies [5, 28] has shown that the raw AP association trace *alone* is a bad estimator of device location and mobility. Thus for mobility detection we also utilize records of packet signal strengths.

Signal strength records of three types of packets — probe scan request packets, data packets sent from device to a campus AP or a rogue AP[1] — can be polled from SNMP objects (which include timestamps and MAC addresses). Probe and rogue RSSI records allow tracking of a device even when it is not connected to the WLAN. Association events and RSSI records are complementary to each other. They together lead to higher coverage and finer granularity in mobility detection [28].

The "**mobility**" of a device is defined as the interval (start time, end time) and location (RSSI fingerprint, *i.e.* a set of APs with corresponding packet RSSI and heard time) where it appeared. The detection algorithm takes the following steps:

1. To smooth the fluctuations and flip-flops in raw WiFi data, fingerprint snapshots of each device are generated from its packet RSSI records and association events. The snapshots are generated by using a sliding window and continuously kicking out deprecated records and records of far-away (*w.r.t.* the AP closeness metric) APs.

2. Continuous similar fingerprint snapshots are merged into large fingerprints to represent the *location* of where the device appeared. Two RSSI fingerprints are considered "similar" if most APs in the records are close to each other and the fluctuation on RSSI values of each AP is limited.

3. The first and last recorded time of each merged fingerprint defines the *appeared interval*. One can further simplify the merged fingerprint by accumulating the RSSI value and packet count of each AP, and only keep the top APs to represent the location.

During the 9 weeks of observation at Tsinghua, $201,230$ client devices appeared in the campus WLAN. The granularity of the mobility detection depends on the density of APs. At Tsinghua, on average a client device can be heard by 3.59 APs, which means that there are enough APs to continuously monitor the location of each device. For most buildings on the campus, APs divide the indoor space into regions with granularity finer than $10{\sim}17m$ (region diameter) — in other words, at room level. This is sufficient for fingerprint comparison in classroom education, where courses are separated by signal-blocking obstacles such as walls and corridors.

[1]Rogue APs are WiFi access points that are installed not by network operators, *e.g.* by graduate students in their labs through wired network. Thus they are not directly accessible for data collection.

(5) Course venue and (6) fine-grained schedule

Education resources are often scarce and set with access limitations. For example, at Tsinghua, the classroom location of each course is only accessible to registered students, not to the general public, thus we cannot download all lectures' venue information in a batch unless working directly with the registration office. Together with the requirement that EDUM does not depend on AP location information (see §Labelling-free Mobility Detection), in EDUM we do not assume prior knowledge on the classroom location of each course.

We will discuss in §Course Schedules with Students' Mobility how EDUM derives (5) course locations as RSSI fingerprints using data (1)+(2)+(3)+(4), and further (6) fine-grained course schedules using data (1)+(2)+(3)+(4)+(5).

(7) Smartphone's interactive state at a specific time

From the volunteer users of the TUNet Android app, their phone interactive states (`SCREEN_ON` and `SCREEN_OFF` in Android [2], indicating interactive and asleep/doze modes), WiFi supplicant events [3], and corresponding timestamps are uploaded. By comparing the WiFi supplicant event `COMPLETED` with SNMP trap event `association` and `associated` at the moment of successful association, we align the phone epoch time to the server epoch time.

During the 11 weeks, in total $12,568,138$ interactive state intervals are derived from the $\sim 2,500$ devices. The total duration of interactive mode is $319,856$ hours and the total duration of asleep mode is $861,185$ hours.

COURSE SCHEDULES WITH STUDENTS' MOBILITY

Before generating and analyzing the educational metrics, EDUM tries to infer accurate schedules on (5) **course location (as RSSI fingerprint)** and (6) **start and ending time** (row (5) and (6) as in Table 1) based on the WiFi and mobility traces. At the same time, whether a student did appear at a course location is also derived by comparing (5) course location with (4) the mobility trace of the student.

For detection of course schedules and students' attendance to lectures, currently EDUM assumes the following common characteristics of traditional courses:

- Time-space uniqueness: A course or a student can only occur at one place at a time.
- Fixed time spans: Normally, each day in a school is divided into chunks of non-overlapping timeslots. Then all courses are scheduled to fit into the timeslots.
- Fixed location: Each course has a designated venue (classroom for majority cases) which is rarely changed.
- Fixed participants: From the longitudinal view, the group of people attending the lectures of a course is relatively stable. Although there are occasional changes (e.g. when a student skips a lecture or quits the course), the majority of the group remains largely unchanged.
- Repeating patterns: Students of a course regularly return to the course location. The re-occurring pattern is in accordance with the schedule of the course.

Table 2. Sessions (S.), timeslots (T.) and durations at Tsinghua.
In each timeslot the number of courses that have valid fingerprint location is shown.
(The number of tracked students of the courses are also shown in parentheses.)

S.	T.	Duration	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	1	08:00-08:45	33 (312)	47 (297)	39 (282)	44 (289)	48 (318)	2 (34)	1 (5)
	2	08:50-09:35	33 (312)	47 (297)	39 (282)	44 (289)	47 (314)	2 (34)	1 (5)
2	3	09:50-10:35	59 (437)	75 (443)	71 (475)	70 (393)	61 (390)	6 (47)	1 (5)
	4	10:40-11:25	57 (409)	75 (443)	70 (443)	67 (369)	59 (336)	6 (47)	1 (5)
	5	11:30-12:15	36 (305)	30 (235)	34 (257)	23 (139)	34 (189)	6 (47)	1 (5)
3	6	13:30-14:15	38 (195)	46 (292)	48 (301)	4 (24)	24 (165)	6 (35)	1 (4)
	7	14:20-15:05	37 (177)	44 (265)	47 (286)	4 (24)	23 (120)	6 (35)	1 (4)
4	8	15:20-16:05	42 (303)	55 (289)	45 (265)	9 (35)	28 (177)	7 (39)	2 (63)
	9	16:10-16:55	35 (276)	51 (255)	43 (260)	9 (35)	23 (134)	1 (4)	1 (59)
5	10	17:05-17:50	10 (43)	5 (16)	8 (51)	1 (2)	3 (11)	0 (0)	1 (4)
	11	17:55-18:40	3 (7)	4 (11)	2 (4)	1 (2)	0 (0)	0 (0)	1 (4)
6	12	19:20-20:05	19 (121)	33 (201)	24 (159)	23 (115)	9 (89)	0 (0)	1 (4)
	13	20:10-20:55	19 (121)	33 (201)	22 (155)	23 (115)	9 (89)	0 (0)	1 (4)
	14	21:00-21:45	6 (42)	12 (87)	12 (63)	7 (39)	2 (19)	0 (0)	1 (4)

As shown in Table 2, at Tsinghua University, each day during a semester is divided into 14 45*min*-long timeslots. These 14 timeslots are merged into 6 sessions. Each course takes several scheduled timeslots each week. In this paper, we call each series of continuous timeslots of a course a "**lecture**". At Tsinghua, lectures always start at the first timeslot of a session, *i.e.* the 1^{st}, 3^{rd}, 6^{th}, 8^{th}, 10^{th} and 12^{th} timeslot. Each course has at most one lecture in each day.

Course Fingerprint Location v.s. Students' Mobility

To determine the RSSI fingerprint location of a course ((5) in Table 1), we merge all the RSSI packet records of the students' devices during the approximate lecture time. The merging of RSSI records into fingerprints is the same as the method used in mobility detection (§Labelling-free Mobility Detection). We only keep the top 6 records (ranked by the accumulated RSSI value $\sum_i (200 + RSSI_i)$). To ensure the merged fingerprint is valid, we skip the courses with too few samples (<3 tracked students or <4 days they did appear in WLAN range). Furthermore, inconsistent RSSI records are eliminated from merged fingerprint. (Two RSSI records are considered consistent if their APs are close — same definition as in mobility detection — with each other.) Only the fingerprints with enough (≥ 3) nearby APs are kept.

Finally, fingerprint locations of 775 courses are derived. We check the fingerprints by comparing the top one AP (The deployed or nearest room is marked on each AP at Tsinghua.) with the actual classroom location of 495 of these courses (collected from part of TUNow users, as shown in Fig. 1.). 86.3% top APs are the exact ones in the classrooms. 12.3% are APs deployed near the course venue — common cases for areas with low AP density and some outdoor physical courses.

With (5) the course fingerprint location available, (4) the mobility trace of a student can be easily converted into a sequence of "appear at course location" and "appear at other locations" labels. Since it is rare that a student appears near the course location but does not attend the lecture, we can approximate students' attendances using also the fingerprint determination method in mobility detection.

Time Span of a Lecture

The students' digital timetables at Tsinghua is at the granularity of a session. The actual number of timeslots of a lecture is decided by the course teacher. Thus we need to derive accurate schedules at timeslot granularity ((6) in Table 1).

After knowing whether students appear at their course locations, we can heuristically derive the timeslot-level schedules of each lecture of a course. A dramatic drop (40% in our case) of the number of appeared students in a timeslot indicates the lecture is finished. Data can be further accumulated since in most courses the lectures are repeated weekly. With starting timeslot directly known, the time span of the lecture is derived.

For all the lectures of the 775 courses that have fingerprint locations, 15 last for 1 timeslot, 673 last for 2 timeslots, 233 last for 3 timeslots, 27 last for 4, and 12 last for 5 timeslots, which is in line with the actual lecture length distribution on the campus. Only 2 lab courses have 2 lectures on Monday and Wednesday take ≥ 6 timeslots. In total, these 775 courses cover 691 students with devices being tracked by EDUM.

One can extend our algorithms in this section to deal with the cases that course time schedules are not available. For example, regular co-location of a group of devices during lecture timeslots of different weeks indicates a course is scheduled there, leading to various group event detection algorithms [6, 21]. However, this is out of the scope of this paper and we currently assume approximate time schedule information is available to EDUM.

LECTURE PUNCTUALITY

With information about the students' attendance at their courses, we can answer several important questions on education. To better understand educational behavior, to improve teaching and to give improvement suggestions, in this section we first focus on problems related to punctuality, including:

1. When do students choose to skip a lecture? Morning, afternoon or evening? How does the time schedule of a course impact attendance and late arrival to its lectures?
2. Besides time variations, do other properties of a course also influence the punctuality of students?
3. When focusing on individuals, are there patterns in different groups of students? *E.g.*, do graduate students skip classes more than undergraduates?
4. Do punctuality patterns correlate with study performance?

In EDUM we try to tackle these punctuality measurement problems by defining automatically generated metrics, including lecture attendance ratios and late arrival / early departure ratios. From essentially the server-side WLAN data, EDUM can evaluate students' punctuality more effectively than manual processes such as prone-to-cheat check-ins.

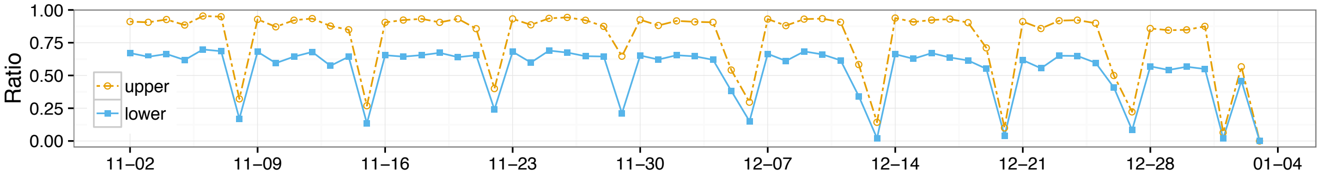
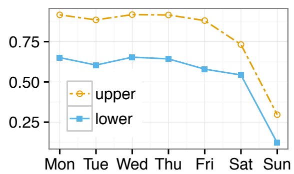
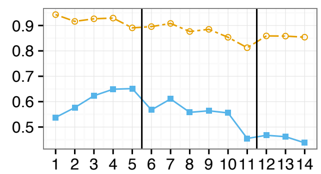


Figure 2. Time Series of Lecture Attendance Ratios. (Equation (1) aggregated daily on both numerator and denominator; Mondays are labelled.)

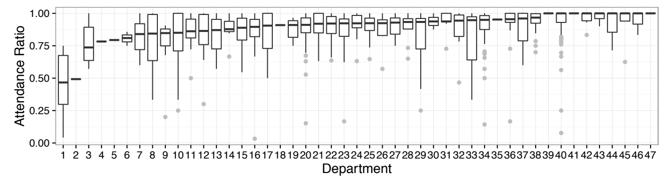


(a) Week Pattern.

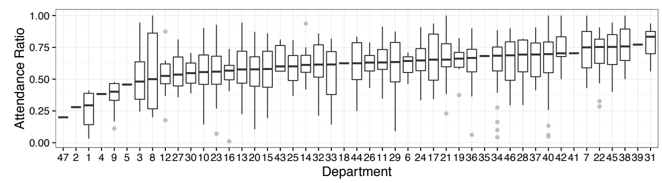


(b) Day Pattern (in Timeslots).

Figure 3. Aggregated Patterns of Attendance Ratios.



(a) Upper Bound (ranked by median).



(b) Lower Bound (ranked by median).

Figure 4. Course Attendance Ratios *v.s.* Departments.
Lower and upper hinges are 1^{st} and 3^{rd} quartiles. Whiskers are observations within range of $1.5*$IQR (inter-quartile range) to hinges.

Attendance Ratios

We define the attendance ratio of lectures and timeslots as:

$$\frac{\#(\text{attended students})}{\#(\text{appeared students on campus})}. \qquad (1)$$

Here "attended" means the students' devices appeared at the course location during the scheduled duration. However, to measure whether a student appeared on the campus is not a trivial task. A straightforward problem is that the student may stay away at places with no campus WLAN coverage. *E.g.*, at Tsinghua, APs are not deployed in the apartments of undergraduate students by January 2016. Sometimes students may choose to turn off WiFi and use the cellular data network, thus disappearing from the campus WLAN. So it is not a good approach to directly divide the number of attended students by the total number of tracked students of the course. Instead, we define an **upper** bound approach and a **lower** bound approach to the actual attendance ratio by two definitions of the number of appeared students on campus: 1) count all the students whose devices appeared anywhere on the campus during only the lecture or the timeslot (leading to the upper bound); 2) count all the students whose devices appeared anywhere during the whole day (leading to the lower bound). The first definition of the denominator gives a smaller value than the second one, and also approaches an optimistic attendance ratio while the second one is much more conservative.

In Fig. 2, the upper and lower bounds are plotted for the 9 studied weeks. One may expect an increasing trend of attendance ratio towards the final weeks[2] because of more check-ins, addressing of exams and important summaries during lectures. However, the opposite occurs. The overall class attendance is decreasing for both upper and lower bounds. Possible reasons for this can be: First, after mid-terms students are not as hard-working as before; Second, towards the ending of a long semester and the upcoming of a brand new year, students are distracted and join more parties and go-outs. Notice that Friday, January 1, 2016 is New Year's Day. All courses are cancelled on this holiday, leading to an abnormal drop of attendance ratios towards zero.

[2] At Tsinghua, the mid-term week is the 8^{th} week starting on 11-02, and the last normal week is the 16^{th} week starting on 12-28. Starting from 01-04 are two weeks dedicated for exams.

Weekly patterns are also interesting (Fig. 2). We can see an obvious drop of attendance ratio for lectures on weekends compared to weekdays. At Tsinghua, a number of unimportant secondary courses for minor and second degrees are scheduled on weekends. Also students may be more relaxed on weekends and choose to skip lectures easily. A common saying is that on Mondays and Fridays people are not in the state of working compared to middle week. To further understand weekly attendance patterns, we aggregate the attendance samples of 9 weeks into Fig. 3(a). It is clearly visible that attendance ratio drop from Wednesday to Friday. However, it seems that on Monday students are in the state of hard studying — Monday has the 2^{nd} highest ratio on average, just after Wednesday. It is surprising to see a clear drop on Tuesday, rather than a continuous increase to the climax on Wednesday, in both Fig. 3(a) and Fig. 2. The reasons for this somewhat mystifying behavior should be studied further.

We also study the intra-day patterns of attendance. Based on our past experience as college students, the authors would have expected that attendance might be higher in the afternoon, rather than in the morning. However, as Fig. 3(b) shows that the overall attendance ratio steadily decreases from morning to afternoon to evening (ratios are calculated for each timeslot, not lecture). Even for the earliest 1^{st} timeslot starting at 8AM, the average attendance ratio is much higher than the evening timeslots. It would be interesting to conjecture what might cause this effect. Clearly, there must be a counter-force to the "stay- in-bed"-effect in the morning. Maybe students gradually lose self-control as the day progresses, eventually leading them to skip classes in the process.

Beside time variations, other course properties can also be analyzed with attendance ratios. As an example, we consider the *offering department* of the course (*i.e.* the department

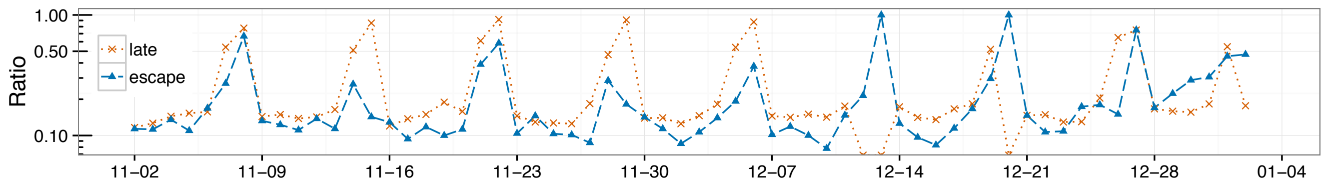
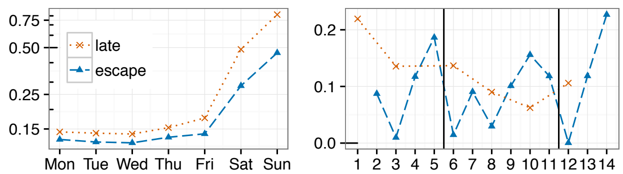


Figure 5. Time Series of Late Arrival (late, Equation (3)) and Early Departure (escape, Equation (4)) Ratios. [log-scale]



(a) Week Pattern (log-scale). (b) Day Pattern (in Timeslots).

Figure 6. Aggregated Patterns of Late and Escape Ratios.

Table 3. Punctuality of different Groups of Students.

All values are shown in percentage as mean(sd).

		upper	lower	late	escape	owls
Grade	2012	67.6(23.2)	37.9(23.5)	28.8(31.3)	28.9(23.4)	21.1
	2013	86.8(15.0)	58.0(20.9)	21.9(18.6)	14.8(14.2)	12.2
	2014	86.4(14.2)	55.8(19.1)	24.0(19.4)	17.1(15.3)	22.0
	2015	91.5(9.9)	62.7(19.0)	16.0(14.1)	14.9(14.8)	13.0
Type	Under.	87.7(14.2)	58.1(20.1)	20.4(18.4)	15.9(15.1)	16.0
	Master	93.3(11.6)	71.6(22.8)	16.1(15.8)	25.2(27.9)	18.2
	Ph.D.	86.8(19.2)	63.7(25.9)	23.7(24.5)	24.6(23.1)	NA
Sex	Female	87.1(12.4)	61.6(16.9)	19.0(14.2)	13.9(13.5)	18.1
	Male	88.0(14.3)	57.8(20.8)	20.4(18.7)	16.8(16.0)	15.5

or institute that offers the course). The attendance ratio of a course is aggregated for all 9 weeks based on Equation (1). The basic distribution of attendance ratios of each department is shown in Fig. 4 where department numbers are ranked according to upper bound attendance ratio. The ranks by upper bound and lower bound are consistent to a large extent. Out of 1081 department pairs, 778 are ranked in the same order by both upper and lower bound. The two definitions of attendance ratio share 10 common departments from bottom-15 ones of both, and 11 from of top-15 ones. To the authors' knowledge, the rank observed in the data fits what we would expect at Tsinghua. For example, the median attendance ratio of courses opened by the Department of Mathematics (numbered 34) and the Department of Foreign Languages (numbered 40) are ranked high in both definitions. Student interactions are common in language courses and hard efforts are required in math courses. From this point of view, attendance ratios can be considered as a first approximation to a course attractiveness measure, which we will discuss in §Lecture Attractiveness.

Yet another view is to measure the attendance ratio for individual students. We calculate a student's attendance ratio during a period as:

$$\frac{\#(\text{lectures attended})}{\#(\text{lectures that s/he appeared on campus})} \tag{2}$$

Same as Equation (1), there are two definitions of time span on "appeared on campus": 1) appeared anywhere in campus WLAN range during the lecture (or similarly, during a timeslot), which leads to the upper bound; 2) appeared anywhere during the day, which leads to the lower bound.

Ratios are calculated for each tracked student from the whole 9-week period mobility traces. In the "upper" and "lower" columns of Table 3, personal attributes including grade (the year entering school), type (undergraduate, master and PhD student) and gender (female and male) are considered for comparing different groups of students. A clear trend of decreasing attendance ratio is shown from lower grade of 2015 (freshmen) to higher grade 2012 students (4^{th} year). Master students have the highest attendance ratio, which can be caused by the course-based qualifying requirement for them. No large gap exists between female and male students.

Late Arrival and Early Escape

When taking a closer look at the mobility traces, another set of punctuality metrics can be derived to characterize the late arrival and early escape of students in a lecture. We say a student has a "**late**" arrival to a lecture when his device appeared at the lecture later than 15*min* (one third of a timeslot length at Tsinghua University) after the beginning of the lecture. Similarly, we define an early "**escape**" from a lecture when the device disappears from the course location 15*min* before the ending of the lecture. Thus the late ratio and escape ratios can be defined as:

$$\text{late ratio} = \frac{\#(\text{late arrived students})}{\#(\text{attended students})} \tag{3}$$

and

$$\text{escape ratio} = \frac{\#(\text{early escaped students})}{\#(\text{attended students})}. \tag{4}$$

The overall pattern of late and escape ratios are shown in Fig. 5. Similar to what we have discovered in Fig. 2, towards the end of both year and term, more students are late to lectures or escape early from lectures. We also look at the weekly pattern closer in Fig. 6(a), here we can see the interesting pattern that both late and escape ratios drop a little from Monday to Wednesday, and then rise during the rest of the week.

However, in Fig. 6(b) (late arrival ratios are only counted for starting timeslots of lectures, while early escape ratios are only aggregated in the ending timeslots of lectures; zero values are omitted), we can see clearly a reduction of the late arrival ratio as the day progresses. Together with Fig. 3(b), this further confirms the "hardness of getting up early". Hard to wake up in morning, easy to be distracted as day progress. The data suggests that this is the common tragedy for today's college students. The escape ratio of timeslots in Fig. 6(b) shows a more complex pattern. We can see that the escape ratio is lowest at the starting timeslot of each session — namely the 1^{st}, 3^{rd}, 6^{th}, 8^{th}, 10^{th} and 12^{th} timeslot. Otherwise as the lectures progress, or comes near lunch (5^{th}) and dinner ($10^{th}, 11^{th}$) time, more students escape from lectures.

So a natural question arises: Why and what students are late for lectures or escape from them? We try to partially answer the question by looking at the overall late and escape ratios of different categories of students. As shown in the "late" and "escape" columns of Table 3, we can see again the less punctual trend of students of higher grades — the late and escape ratios become highest in the group of 4^{th}-year students. A conflicting trend of low late ratio and high escape ratio is shown for master students. By comparing with the trends in attendance ratio, one possible explanation is that a lot of master courses have check-ins or in-class quizzes early during lectures, but students are still reluctant to finish them. Finally, compared to female students, male students are slightly more frequently late for or escape from classes.

In the above analysis of attendance, late and escape ratios, we find the counter-intuitive trend that in general class attendance is higher in the morning. Another observation which meets common sense is that students face heavy late arrival problems in the morning. Despite one possible explanation that students wake up too late for classes, an alternative possible fact is that different life patterns are mixed in our aggregated metrics in Fig. 3(b) and Fig. 6(b). Besides the morning people and those who can well fit the schedule of the university, there are night owls who sleep late and become efficient in the afternoon and evening. We try to identify these students by student punctuality metrics.

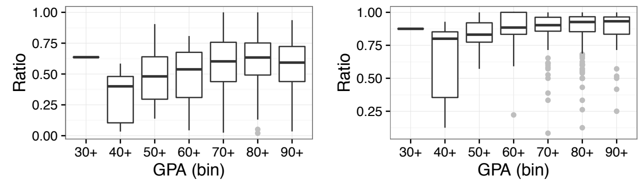
Attendance ratios of a student are already defined in Equation (2). Similarly we define the late and escape ratios of a student over a period as:

$$\text{student late ratio} = \frac{\#(\text{lectures that arrived lately})}{\#(\text{lectures attended})} \tag{5}$$

and

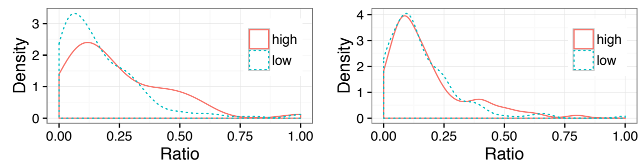
$$\text{student escape ratio} = \frac{\#(\text{lectures that departured early})}{\#(\text{lectures attended})}. \tag{6}$$

We consider the aggregated daily patterns of attendance and late ratios of a student. A "**night owl**" student should have a decreasing trend of late ratio and increasing trend of attendance ratio. To determine the trend within a day, attendance and late ratios are aggregated over 14 timeslots of the day. Then we apply linear regression to the sequence of timeslot ratios. The slope of the fitted line is then an indicator of an increasing trend (when the slope $\geq 0.5\%$, which means 6.5% increase in ratio from first to last timeslot) or a decreasing trend (when the slope $\leq 0.5\%$, which means 6.5% decrease in ratio from first to last timeslot). We classify a student as a night owl if the trend of the upper and lower bound attendance ratios are both increasing and the late ratio trend is decreasing. Out of 639 students whose ratio sequences has length ≥ 3, 103 are identified as night owls. In Table 3, the percentage of night owls in each category of students is shown. We can see that 2^{nd} and 4^{th} year students are more likely to be night owls. Maybe surprisingly, the fraction of night owls in female students is higher than that in male students.



(a) Lower Bound. (b) Upper Bound.

Figure 7. Students' Attendance Ratio *v.s.* GPA.
Lower bound and upper bound are compared with students' GPA respectively.
X-axis of GPA is discretized into bins of size 10 for aggregation.
Lower and upper hinges of plotted boxes present the 1^{st} and 3^{rd} quartiles.



(a) Student Late Ratio. (b) Student Escape Ratio.

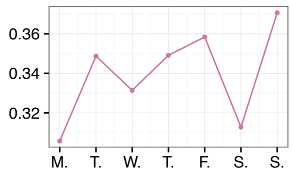
Figure 8. Estimated Density Distributions of Students' Late Arrival and Early Escape Ratios with High and Low Performances.

Study Performance

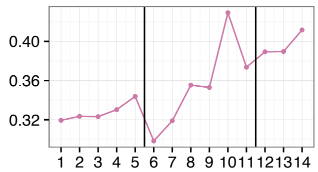
The results on punctuality metrics of EDUM can be used to get deeper insights into student performance. Giving scores is a common evaluation method for student performance in most courses. Thus here we calculate the GPA (grade point average) of the autumn semester as an overall study quality indicator for each tracked student (635 TUNow volunteers) from 718 scored courses out of the 775.

In general, *students with higher GPA attend lectures more*. As shown in Fig. 7, we can see the increasing trend of attendance ratio for groups of students that have higher GPA. The Pearson correlation between student punctuality ratios (Equation (5) and Equation (6) aggregated over all 9 weeks) and their GPA are: upper bound attendance ratio has 0.167 (p-value < 0.001), lower bound attendance ratio has 0.133 (p-value $= 0.001$), late ratio has 0.075 (p-value $= 0.078$), escape ratio has -0.027 (p-value $= 0.536$). It is *not so clear whether late arrival and early escape correlates with performance*.

To further understand the relationship between punctuality and GPA, we look at the distributions of late and escape ratios from students with high performance (GPA > 90) and low performance (GPA < 80). (Tsinghua University adopts a hundred-point grading policy.) As shown in Fig. 8(a), estimated kernel density [23] distributions are plotted. We can see a clear difference of late ratio between high-performance and low-performance students. We also conduct the K-S test (Kolmogorov-Smirnov Test) with α set to level 0.05. The test rejects the null hypothesis that the late ratios of low-performance students is larger than those of high-performance students (with p-value $= 0.033$). In other words, while high-performance students have higher attendance ratios, when appearing at class, they are more likely to be late than appearing low-performance students. Finally, in Fig. 8(b) we find no clear difference regarding the escape ratios between high-performance and low-performance students.



(a) Week Pattern.



(b) Day Pattern (in Timeslots).

Figure 9. Aggregated Patterns of Phone ON Ratio.

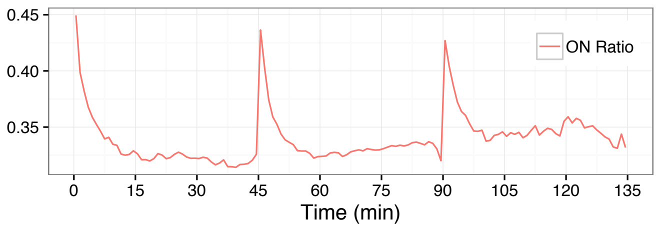


Figure 10. Phone ON Ratio as Lecture Progress.

Further prediction of study performance and support for bad study behaviors are possible based on metric modules of EDUM. Our study is scratching the surface only at this point. For example, it would be interesting to try and predict course score ranking based on punctuality and other EDUM metrics.

LECTURE ATTRACTIVENESS

It is hard to measure how good or attractive a course is. Nowadays course ratings are collected via a combination of surveys, anecdotal feedbacks of students, occasional audit of school officers, and teacher peer reviews. In this section, we show how to extend EDUM system by new data to contribute measurements that indicate the attractiveness of courses.

We approximate attractiveness as *the absence of distractions* in EDUM. Usage of mobile devices is a good indicator for distractions during a class. Collected from TUNet users, the device interactive and asleep mode intervals [2] are added to EDUM system. Then interactive distraction ratio (ON ratio) can be straightforwardly defined as:

$$\frac{\text{Total ON duration}}{\text{Total ON duration} + \text{Total OFF duration}} \tag{7}$$

which is intuitive and can be interpreted as the device usage ratio during the measured time span. ON ratio is accumulated only for students who appeared at class.

The weekly and daily patterns of distraction is shown in Fig. 9 by aggregating all data on the numerator and denominator of Equation (7) in each weekday and timeslot. As Fig. 9(a) shows, distraction metrics increase from Wednesday to Friday, and drop on Saturday. In Fig. 9(b), the overall pattern of ON ratio is increasing as the day progresses, which aligns with our observations regarding the drop of attendance ratio we have observed in §Lecture Punctuality. Similar to the escape ratio in Fig. 6(b), the ON ratio also drops around lunch and dinner (6^{th} and 11^{th}) timeslots, and increases as the day progresses.

Compared to short videos of on-line educations, classroom courses are sometimes criticized for their long length and low efficiency. Thus we further aggregate data to lecture scale to see how distraction varies minute by minute as a lecture progresses. In Fig. 10 the metric is shown for the most common 2-to-3-timeslot (each timeslot lasts 45*min*) lectures. The lecture break time is removed that only class time is left in the plot. We can see a clear trend of *high device usage at the start of each time slot*, which quickly drops in the first 10*min* of a timeslot. The implications are clear. During the first few minutes of a timeslot, students are attracted by teachers and stop using their devices. In general, from the first to the last timeslot, the distraction ratio increases, indicating that students gradually lose attention during the class. However, interestingly near the end of the lectures, *i.e.* usually near 90*min* or 135*min*, the device usage again slightly decreases.

Table 4. Correlations (p-value) among All Course Metrics in EDUM.

	upper	lower	late	escape
ON	-0.079 (0.032)	-0.093 (0.011)	0.067 (0.069)	0.026 (0.482)
upper		0.715 (<0.001)	-0.183 (<0.001)	-0.250 (<0.001)
lower			-0.123 (<0.001)	-0.208 (<0.001)
late				0.277 (<0.001)

Finally, we look at the Pearson correlations among all the course metrics. As shown in Table 4, we see clear correlations among punctuality metrics. Attendance ratios negatively correlate with late and escape ratios. But there is no significant correlation between ON ratio and the punctuality metrics.

DISCUSSION AND FUTURE WORK

Limitations of EDUM: First, one assumption of measurements in EDUM is that device owners leave traces in WLAN. However, students may turn WiFi off during a lecture. Further challenges are that some devices lazily do WiFi scan and connect; bad WiFi conditions of too little APs, too many devices and outdoor situations; varying density of APs. For deeper and finer analysis exceeding our work, these concerns may become critical and require further data analysis.

Second, without prior knowledge of course schedules, EDUM requires longitudinal data to seek accurate course schedules. Thus our current implementation does not support dynamic outputs as courses progress. However, classroom fingerprints can be generated based on historical data from past semesters. In the next version of EDUM we will introduce real-time measurements and thus allow real-time monitoring.

Third, our deployment is currently limited at Tsinghua University by tracking volunteers from Android users of our mobile apps. This may introduce unexpected biases and noises in our results. In other environments where the access to additional information — such as the device-account mapping and timetable samples — is limited, further development on algorithms might be necessary.

Fourth, it is fundamentally hard to get ground truths of educational behavior. In EDUM we try to derive metrics for the behavior to improve the foundation of education studies. On the other hand, our metrics currently lack evaluations from manually collected data. As future work we will verify the accuracy and variation of our metrics.

Extensions of EDUM: First, we design EDUM to be extensible and scalable. EDUM could be ported to other settings, *e.g.* large organizations where WLAN data is available, to conduct similar metrics as in education scenarios.

Second, more interesting data sources can enable additional measurements, and more kinds of educational metrics. For example, semantic knowledge from deep packet inspections (DPI) could be used to characterize the content of each course. More detailed behavior can be also measured through various sensors of mobile devices.

Third, better and more accurate results could be obtained if EDUM has access to detailed alternative data (as shown in Table 1), *e.g.* the accurate location of APs, more education ground truths from school office.

Fourth, the tracking and observation could include more than just the lectures. Behavior out of class, such as group-studying, could be analyzed with new techniques such as socio-physical networks.

More applications: In this paper we only show a basic set of applications based on EDUM's output. There are numerous more potential applications, *e.g.* feedbacks to students, course recommendations and personalization, optimization of course scheduling, *etc.* EDUM could also enable new research opportunities based on its in-situ large-scale measurements. Finally, EDUM could ultimately help us reflect on the effectiveness of traditional education methods compared to MOOCs.

RELATED WORK

There are generally three ways to collect mobility traces [5]: monitoring location (e.g. GPS, RFID based, Bluetooth, GSM and 802.11 beacons, *etc.*) [15, 14], monitoring communication (signal strength of base station/access point and the connectivity events of the device) [13] and monitoring contacts (use mobile devices with Bluetooth, WiFiDirect, *etc.* to sniff other nearby devices) [20, 17]. [18] uses WiFi sniffers to aid instructors in identifying who is in the classroom. We choose to monitor WiFi communication of devices in EDUM through mainly server-side SNMP data, because this is non-intrusive and easy to scale based on existing infrastructures.

The Ubicomp community has shown a great interest in understanding human behaviors through data related with mobile devices, such as mobility, online *v.s.* offline social networks, group behaviors, event detection, phone usage [8, 24, 16]. Recently, the StudentLife project [25, 26] and LiveLabs testbed [11, 12] have advanced our understanding of individual and group behaviors in campus environment settings. The StudentLife project involved 48 graduate and undergraduate students. Fine grained mobile data and psychological surveys are collected directly from their devices. Correlations and predictions with study performance is well-studied in the StudentLife project. We adopt a different approach in this paper (based on WLAN), and introduce several new metrics from the view of both courses and students for different purposes. The LiveLabs testbed is more similar to our setting — but it did not focus on education and was at a much smaller scale (156 smart phones) than ours.

Large-scale educational measurements have recently become popular for MOOCs (Massive Open On-line Courses) which are hosted by web platforms like EdX [10], Coursera [4], Udacity, *etc.* Many web interactions and Internet metrics have been studied for on-line courses, including video watching behaviors, on-line homework and quiz performance, forum participations [4], *etc.* In contrast, studies under classroom settings are mostly done through intrusive methodologies at small scale, preventing the possibility of a comparably large-scale analysis as in MOOCs. We try to design EDUM to work automatically, non-intrusively, scalably and extensibly, thereby showing the possibility to do educational research for traditional classroom courses at scale.

CONCLUSION

By tracking devices through WiFi traces, our proposed EDUM system provides a new way to measure student behavior and the effectiveness of classroom-based courses. Through a large-scale deployment, we show that EDUM is scalable, non-intrusive and extensible for new types of data and measurements. Our measurement results show that EDUM enables new observations on aspects such as punctuality, study performance and lecture attractiveness (or student distraction), and quantifies aspects of education that have up to now been notoriously hard to measure.

We believe this paper makes an important first step towards *automatic*, *data-driven*, *quantitative*, and *objective* classroom education measurements. As our future work, we plan to further explore the directions sketched in the §Discussion and Future Work section.

ACKNOWLEDGEMENTS

We thank the network center[3] of Tsinghua University for their kind support to our researches on the campus WLAN. Zimu Li *et al.* helped us a lot on building the MobiCamp testbed [28] on which our experiments are taken. We strongly appreciate team members of the student interest group Lab μ[4] at Tsinghua University, including Guang Chen, Haoyu Hu, *et al.*, who put great amount of efforts into the development of mobile Apps. Thorough comments and valuable feedbacks from the reviewers also helped us improve the work.

This work was partly supported by the National Basic Research Program of China (973 Program) under grant 2011CBA00300 & 2011CBA00301 & 2013CB329105, the National Natural Science Foundation of China (NSFC) under grant 61033001 & 61361136003 & 61472214 & 61472210, the Key Program of the National Natural Science Foundation of China under grant 61233007 & the National High Technology Research and Development Program of China (863 Program) under grant 2013AA013302, the Tsinghua National Laboratory for Information Science and Technology key projects, the Global Talent Recruitment (Youth) Program, and the Cross-disciplinary Collaborative Teams Program for Science & Technology & Innovation of Chinese Academy of Sciences-Network and system technologies for security monitoring and information interaction in smart grid.

[3] http://www.itc.tsinghua.edu.cn
[4] http://www.lab.mu

REFERENCES

1. 2016a. Android 6.0 Changes on Access to Hardware Identifier. http://developer.android.com/about/versions/marshmallow/android-6.0-changes.html#behavior-hardware-id. (2016). Accessed: 2016-03-29.
2. 2016b. Android phone interactive broadcast action. http://developer.android.com/reference/android/content/Intent.html#ACTION_SCREEN_ON. (2016). Accessed: 2016-03-29.
3. 2016c. Android Supplicant States from wpa_supplicant. http://developer.android.com/reference/android/net/wifi/SupplicantState.html. (2016). Accessed: 2016-03-29.
4. Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2014. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*. ACM, 687–698.
5. Nils Aschenbruck, Aarti Munjal, and Tracy Camp. 2011. Trace-based mobility modeling for multi-hop wireless networks. *Computer Communications* 34, 6 (2011), 704–714.
6. Chloë Brown, Neal Lathia, Cecilia Mascolo, Anastasios Noulas, and Vincent Blondel. 2014. Group colocation behavior in technological social networks. *PloS one* 9, 8 (2014), e105816.
7. I Cisco. 2013. Cisco visual networking index: Forecast and methodology, 2013–2018. *CISCO White paper* (2013), 2013–2018.
8. Nathan Eagle, Alex Sandy Pentland, and David Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* 106, 36 (2009), 15274–15278.
9. Noel J Entwistle. 2013. *Styles of learning and teaching: An integrated outline of educational psychology for students, teachers and lecturers*. Routledge.
10. Andrew Dean Ho, Isaac Chuang, Justin Reich, Cody Austun Coleman, Jacob Whitehill, Curtis G Northcutt, Joseph Jay Williams, John D Hansen, Glenn Lopez, and Rebecca Petersen. 2015. HarvardX and MITx: Two years of open online courses fall 2012-summer 2014. *Available at SSRN 2586847* (2015).
11. Kasthuri Jayarajah, Youngki Lee, Archan Misra, and Rajesh Krishna Balan. 2015a. Need accurate user behaviour?: pay attention to groups!. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 855–866.
12. Kasthuri Jayarajah, Archan Misra, Xiao-Wen Ruan, and Ee-Peng Lim. 2015b. Event Detection: Exploiting Socio-Physical Interactions in Physical Spaces. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 508–513.
13. Minkyong Kim, David Kotz, and Songkuk Kim. 2006. Extracting a Mobility Model from Real User Traces.. In *INFOCOM*, Vol. 6. 1–13.
14. Ilias Leontiadis, Antonio Lima, Haewoon Kwak, Rade Stanojevic, David Wetherall, and Konstantina Papagiannaki. 2014. From Cells to Streets: Estimating Mobile Paths with Cellular-Side Data. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*. ACM, 121–132.
15. Liqun Li, Guobin Shen, Chunshui Zhao, Thomas Moscibroda, Jyh-Han Lin, and Feng Zhao. 2014b. Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 459–470.
16. Ming-Xia Li, Zhi-Qiang Jiang, Wen-Jie Xie, Salvatore Miccichè, Michele Tumminello, Wei-Xing Zhou, and Rosario N Mantegna. 2014a. A comparative analysis of the statistical properties of large mobile phone calling networks. *Scientific reports* 4 (2014).
17. Shu Liu and Aaron D Striegel. 2013. Exploring the potential in practice for opportunistic networks amongst smart mobile devices. In *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 315–326.
18. Clemens Martin and Heba Zakaria. 2015. Sensing student presence augmented reality for the lecture hall. In *2015 9th International Conference on Sensing Technology (ICST)*. IEEE, 844–849.
19. Allison Mooney and Jordan Rost. 2014. A Report Card on Back to School 2014: The Season's Trends and What They Mean for Holiday. https://www.thinkwithgoogle.com/articles/a-report-card-on-back-to-school.html, (2014). Accessed: 2016-07-17.
20. Vedran Sekara and Sune Lehmann Jørgensen. 2014. The strength of friendship ties in proximity sensor data. *PL o S One* 9, 7 (2014).
21. Rijurekha Sen, Youngki Lee, Kasthuri Jayarajah, Archan Misra, and Rajesh Krishna Balan. 2014. Grumon: Fast and accurate group monitoring for heterogeneous urban spaces. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM, 46–60.
22. Kaixin Sui, Youjian Zhao, Dan Pei, and Li Zimu. 2015. How Bad Are The Rogues' Impact on Enterprise 802.11 Network Performance? 361–369. DOI: http://dx.doi.org/10.1109/INFOCOM.2015.7218401
23. William N Venables and Brian D Ripley. 2013. *Modern applied statistics with S-PLUS*. Springer Science & Business Media.
24. Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1100–1108.

25. Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 3–14.

26. Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell. 2015. SmartGPA: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 295–306.

27. Chenshu Wu, Zheng Yang, Yunhao Liu, and Wei Xi. 2013. WILL: Wireless indoor localization without site survey. *IEEE Transactions on Parallel and Distributed Systems* 24, 4 (2013), 839–848.

28. Mengyu Zhou, Kaixin Sui, Minghua Ma, Youjian Zhao, Dan Pei, and Thomas Moscibroda. 2016. MobiCamp: a Campus-wide Testbed for Studying Mobile Physical Activities. In *Proceedings of the 3rd Workshop on Physical Analytics*. ACM.