Measuring and Modeling Car Park Usage: Lessons Learned from a Campus Field-Trial

Thanchanok Sutjarittham*, Gary Chen*, Hassan Habibi Gharakheili*, Vijay Sivaraman*, and Salil S. Kanhere†
*School of Electrical Engineering and Telecommunications, †School of Computer Science and Engineering
University of New South Wales, Sydney, Australia
Emails: {t.sutjarittham@, gary.chen@student. h.habibi@, vijay@, salil.kanhere@}unsw.edu.au

***Abstract*—Transportation is undergoing significant change due to the growth of ride-sharing, electric cars, car-sharing, and self-driving cars. Organizations that have significant real-estate dedicated to on-premise employee car parking are therefore looking to adapt the use of this space, motivated by the opportunity to become greener, improve sharing, and pursue new revenue opportunities. In this paper, we outline our experiences from instrumenting, measuring, and analyzing car-park usage in our University's multi-storey parking lot, and building a model that explores its use for multiple purposes in the near future.**

Our specific contributions are as follows: (1) We begin by describing experiences and challenges in measuring car-park usage on our campus and cleaning the collected data; (2) We analyze data collected over 23 weeks (covering teaching and non-teaching periods) and draw insights into the usage patterns, including occupancy patterns by times-of-day and days-of-week, and identifying various user groups based on attributes such as arrival time and duration of stay; (3) We develop a queuing model to optimize the use of parking space for generating revenue from shared cars with minimal impact on private car users. We believe our study guides campus managers wanting to generate more value from their existing parking resources.

I. Introduction

University campuses in much of the Western world are experiencing a surge in student enrollments [1], accompanied by an expansion in staff numbers, which are jointly putting pressure on demand for on-campus parking. Many college campuses are large and spread-out, meaning that even though the use of private vehicles to commute to campus is increasing [2], as many as 10-45% of available spaces are unused because they are distributed across the campus [3]. This problem has also been observed at our campus in UNSW Sydney, where one of the multi-story parking lots fills up by 10am while the other often has availability. Commuters lack real-time visibility into parking availability in various parts of the campus, and therefore often have to spend time driving around. Indeed, information on parking space utilization will not only benefit users, but will also provide evidence and statistics to campus Estate Manager in planning for critical decisions such as pricing policy and the need for new parking spaces. Further, with new trends in transportation, ranging from electric cars (Tesla) and ride-sharing (Uber) to car sharing (GoGet) and ultimately self-driving cars, campus estate managers need quantitative data to inform their future strategies that optimize available space while becoming greener, encouraging sharing, and pursuing new revenue opportunities.

The medium-term future is likely to evolve around shared transport, leading to autonomous vehicles in the long-term. These will not only enhance commuter experience, but equally importantly cut down the use of fuel, alleviate the number of cars on the road, and improve overall traffic congestion. Car sharing (offered by companies such as GoGet, ZipCar, Car2Go, etc.) is projected to grow at a rate of over 20% between 2018 and 2024 [4] and is becoming a more mainstream option. Universities such as UNSW can leverage such trends to go green by encouraging car sharing scheme, reduce parking congestion on campus, and avoid the need to build new multilevel parking structures as the university grows. In order to prepare for such trends in future transport, the University needs visibility into current and predicted usage of their car parking facilities so as to inform the future decision making process.

This paper describes our experiences in instrumenting and building a system to monitor car-park usage on campus, including the real-time collection of data and cleansing it for analysis. We then comprehensively analyze the car park usage data that spans 23 weeks, covering both teaching and non-teaching periods, and highlight interesting insights into car arrival and departure patterns, and user parking behavior. Finally, we develop continuous-time non-homogeneous Markov models based on historical usage data, and showcase how it can be used for space-allocation planning for future car-sharing schemes.

The rest of this paper is organized as follows: §II describes relevant prior work. §III outlines our experiences in implementing the car park monitoring system and collecting data while §IV presents interesting insights obtained therein. §V shows how we develop car park behavior models to aid car-park dimensioning decision and the paper is concluded in §VI.

II. Related Work

Sensing Technologies: A number of technologies are available on the market to monitor parking spaces. Many smart parking deployments used various sensor networks including ultrasonic, light, temperature, acoustic, and magnetic sensors to detect the presence of vehicle at each parking spot in the car-park [5]. However, monitoring every spot individually is expensive, especially for a large parking lot with hundreds

978-1-7281-0270-2/19/$31.00 ©2019 IEEE

EVENT_DESC	DATE	TIME	PLATE_NOT_READ
Ocr Read	7/10/18	11-12-58-306	READ
PLATE_STRING	**PLATE_COUNTRY**	**PLATE_REGION**	**OCRSCORE**
AB83JN	AUS	NS	86
OCRSCORE_CHAR		**SPEED**	**DIRECTION**
086 085 080 087 089 087		14794	APPROACH

Fig. 1. LPR camera outputs for author's car entering into the campus car-park.

of parking bays, so the solution is typically employed in commercial parking areas such as shopping centers. A more cost-effective solution is to use existing CCTV camera to acquire images or videos of a car-park's view and apply image processing to obtain car occupancy data [6]. However, continuously recording images (or videos) of users' vehicles may raise privacy concerns, especially when images are collected without users consent. RFID technology is another solution for car park monitoring, where RFID readers are installed at the entrance/exit of the car park [7]. This method requires a RFID tag, that can be scanned whenever a car enters/exits the parking lot, to be incorporated into parking permit of individual user, and thus does not accommodate for casual visitors of a university campus. The simplest solution (which we use in this paper) is to install a license plate recognition camera at each entrance and exit point of the parking lot to automatically record the license plate number of incoming and outgoing cars.

Use-cases: There are various applications of parking management system worldwide to improve user experience, reduce congestion, and increase revenue from parking services. Real-time data on car-park usage allows realization of parking guidance information (PGI) system where users are informed on the availability and location of parking spaces, and thus have the ability to reserve a parking spot in advance (reducing/removing search time) [8]. Furthermore, historical data from car park monitoring system can be used to predict future demand and dynamic parking fee can be applied based of times-of-day and location of parking. For example, a large-scale dynamic pricing scheme (SFpark) was implemented in San Francisco [9] to overcome traffic congestion.

Parking Usage Modeling: Several researchers attempted to model parking dynamics using queuing theory, incorporating cars arrival and departure processes. Authors of [10] modeled parking activities using $M/M/m/m$ queue to predict parking space occupancy in a vehicular ad-hoc network. Work in [11] applied $M/G/c/c$ queuing model for a parking lot of occupancy considering a Poisson arrival process and a constant average parking time. Blocking probability was calculated and disseminated to users via vehicular communications to prevent more users from entering the parking lot. Similarly, Gongjun et el. [12] modeled the parking process as a birth-death stochastic process to predict the parking revenue. Existing works essentially assumed a fixed rate for either arrival or service time (parking duration) which is not realistic. We instead use varying rate of arrival and departure obtained from real data to model the car-park usage behaviour.

III. Data Collection and Cleansing

In this section, we begin by outlining our experience with license plate recognition (LPR) technology to measure car-park usage. Note that license plate information is private and confidential, and therefore we obtained appropriate ethics clearances for this study (UNSW Human Research Ethics Advisory Panel approval number HC171044). We then briefly explain our system architecture for collecting, storing, and analyzing data collected by the LPR cameras. Lastly, we discuss measurement challenges, quantify errors and present our method for cleaning data.

A. License-Plate-Recognition Camera

We investigated several commercial sensors with two objectives [13]: we wanted to: (a) have complete ownership of the data and not risk it leaving our campus infrastructure; and (b) not be beholden to a vendor to access our own data, hence freeing us from ongoing service costs. In other words, we wanted a "sale" model of the device so we could have unlimited access to our data without any ongoing "service" fees. This model allows us to integrate data into a centralized repository as part of an overarching smart campus project to facilitate better analytics across the many data feeds we have on campus [14].

We chose Nedap automatic number plate recognition camera model ANPR Access HD [15]. The camera uses license plate recognition (LPR) technology to read license plate from the captured images. The technology consists of two main stages: (a) locating license plate in the captured image by isolating a rectangle area (of the license plate number), using physical characteristics such as the shape, symmetry, width to height ratio and alphanumeric characters; and (b) character separation and recognition inside the isolated image of the license plate number [16].

The Nedap camera unit consists of several components including a high definition camera, infrared (IR) illumination, and automatic number plate recognition engine allowing the camera to read complex number plate at various conditions, dark time as well as day time. The camera relies on libraries

Fig. 2. A real picture showing entry/exit of our campus car-park and a pair of LPR cameras installed side-by-side close to the ceiling of the car-park ground floor.

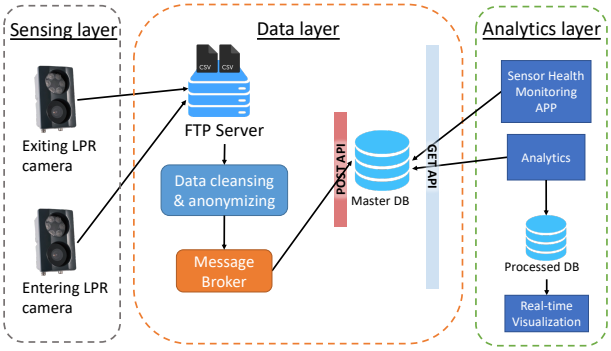


Fig. 3. System architecture of collecting, analyzing and visualizing data from LPR cameras.

that are supporting license plates from specific countries. Each country is using its own characters, colors, and designs.

The camera provides a management console (a web-page accessible via its IP address) that allows users to configure various parameters such as shutter time, strobo time (activation time of IR illuminator), and gain. We configured the camera to use the default "Autoiris" mode which will automatically determine the best parameters to use in its working environment; this mode is recommended by the camera's documentation.

For every vehicle that passes through the frame of the camera, it runs an on-board image processing optical character recognition (OCR) algorithm to output two types of data; an JPEG image (with adjustable quality value between 0 and 100, where 100 is the highest resolution 1080p) of the vehicle, and a data record (in a CSV file) containing parameters such as timestamp, license plate string, OCR score, speed, country, state/region, type of vehicle and up to 50 more fields, in which most of them are substring of another field. For example: fields for hour, minute and second are also available in an aggregate field called "TIME". Fig. 1 shows an example of real license plate (for the private vehicle of an author of this paper) as recognized by the camera. On the left is the JPEG captured image showing the isolated license plate, and on the right is a list of selected key data fields generated by the OCR algorithm. The definition of the fields are as follows:

- `EVENT_DESC`: The value "OCR Read" means that the algorithm was able to recognize individual characters in the isolated license plate.
- `DATE` and `TIME`: time-stamp of the record.
- `PLATE_NOT_READ`: This field indicates whether the license plate was successfully isolated or not – the value could be either "READ" or "NOTREAD".
- `PLATE_STRING`: This is the output string of the recognized license plate. Australian cars and motorcycles respectively have 6 and 5 alphanumeric characters on their license plate.
- `PLATE_COUNTRY` and `PLATE_REGION`: country and state of the license plate – AUS for Australia, and NS for New South Wales.
- `OCRSCORE`: Overall confidence value (between 0 and 100) given by the camera on how accurately the entire license plate number is recognized.
- `OCRSCORE_CHAR`: OCR score for individual characters of the string – in this example, character "J" has the highest score 89 and character "8" has the lowest score 80.
- `SPEED`: Speed of the vehicle in 100× actual speed (km/h) – we found that this data field is unreliable since reported values ranged from 1000 to 10,000,000.
- `DIRECTION`: Direction of the vehicle relative to the camera, *i.e.,* "APPROACH" for entry camera and "GO-AWAY" for exit camera.

For our field trial, we only used text outputs of the camera and intentionally deactivated the JPEG image recording for two reasons: (a) maintaining users privacy, and (b) managing volume of data collected. Note that we temporarily recorded image captures for tuning the camera parameters and collecting ground-truth data for quantifying the accuracy of measurement. The camera supports both local storage on an SD card (if inserted) and an external FTP data collection. We used the latter option for continuous data collection, where the camera creates and updates the CSV file in real-time when a vehicle is detected.

For our experiment, we (in consultation with our Estate Management) chose a 5-story car park of our university campus with 895 parking spaces serving students, staff, visitors, and contract workers. The first 4 levels are reserved for permit holders and the top level for hourly-based paid parking. The car-park has one entry and one exit at its ground floor, therefore we installed two LPR cameras to capture cars entering to and exiting from the parking lot, as shown in Fig. 2. The camera requires 24 VDC power supply, and communicates via Ethernet port. Hence, we (with help form our campus Estate Management) supplied new power points and provisioned Ethernet ports for the cameras.

Measurement System Architecture: Fig. 3 shows the system architecture of data collection from the car-park using LPR cameras. It comprises of three main layers: (a) "sensing layer" is where LPR cameras captures license plate numbers of entering (arriving) and exiting (departing) cars – cameras are connected to the campus wired network over a private VLAN; (b) "data layer" is the core of our system that hosts an FTP server, a software engine for data cleansing and anonymization, message broker and multiple databases for backup and load balancing. Upon cleansing, the data is passed onto the

TABLE I
MEASUREMENT ACCURACY OF LPR CAMERA FROM GROUND-TRUTH OF CARS AND THEIR LICENSE-PLATES.

	Accuracy	
Camera	Capture rate	Read rate
Entry	94%	85%
Exit	88%	42%

message broker which unifies the data into a JSON format, each record is tagged with timestamp and sensor UUID, and is posted via a RESTful API to our master database (where data from various sensors of our campus network is stored); and (c) "analytics layer" that includes health check monitoring (to ensure the cameras are active and functioning), data analysis and visualization modules – this layer retrieves data from the master DB via RESTful API and writes processed data (*i.e.,* real-time occupancy, stay duration) into another DB that is used as the backend for visualization.

B. Measurement Accuracy

The accuracy of the LPR camera depends on various factors including placement (*e.g.,* height/angle at which it is installed), lighting conditions, speed of vehicles, angle of the license plate (on vehicles), and also the physical condition of the license plate. We quantify the accuracy by two metrics: (a) "capture rate" which is the fraction of cars detected, and (b) "read rate" which is the fraction of license plate numbers that are recognized correctly by the OCR algorithm running inside the camera. According to the ANPR's manufacturer [17], it is expected to have capture rate and read rate of 98% and 95%, respectively.

For capture rate, we performed several spot measurements over 5 days (*i.e.,* Monday to Friday) each for a period of 5 hours, (*i.e.,* 11am to 4pm), to extract ground-truth data of car-park usage. We used an GoPro camera to record video logs of a total 1400 vehicles entering/exiting to/from the car-park. For read rate, we first enabled the JPEG recording on both cameras for a week, and then we manually inspected 700 images for each of the ANPR cameras, the first 100 images per day of that week.

Table I summarizes the accuracy of two cameras. It is seen that the exit camera under-performs by both accuracy metrics, especially with a very low read rate of 42%. This means that the license plate of more than half of exiting vehicles are not correctly recognized. The poor read rate of the exit camera is mainly due to its non-ideal placement which causes it to see the license plate of departing vehicles at a further distance and also at a slight angle, and thus affecting the performance of the OCR algorithm. Re-positioning the exit camera was a nontrivial task due to difficulty (and cost) of provisioning Ethernet port and power outlet for the new position.

C. Measurement Errors

We observed three types of errors from cameras' outputs: (a) multiple recognitions of the same license plate, (b) incorrect recognition of license plate location, and (c) incorrect recognition of license plate characters.

TABLE II
DISTRIBUTION OF ERROR TYPES FOR EACH CAMERA (GROUND-TRUTH DATASET).

	Error types		
Camera	Multiple recognitions	Incorrect locating	Incorrect recognition
Entry	59.3%	4.2%	36.4%
Exit	13.5%	34.9%	51.6%

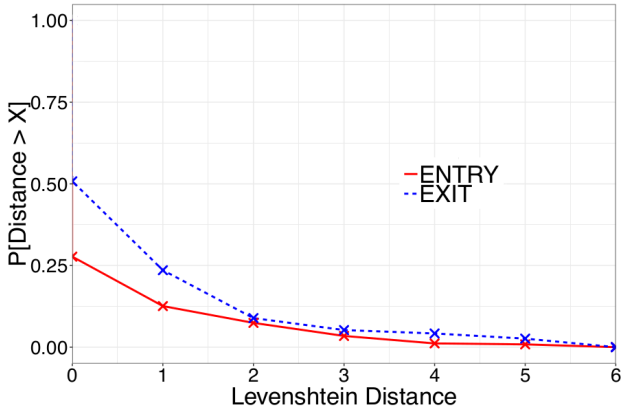


Fig. 4. CCDF of Levenshtein distance for ground-truth dataset.

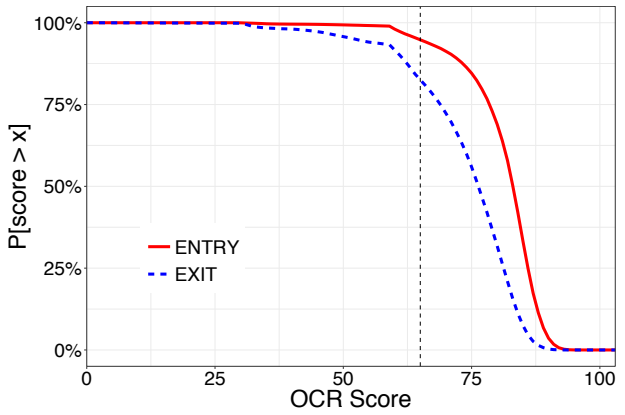


Fig. 5. CCDF of OCR score for entry and exit cameras.

Multiple recognitions: Multiple recognitions occur when the camera takes multiple images of a vehicle (possibly because of its speed or moving pattern), and thus triggers the OCR algorithm multiple times, generating multiple data records for the vehicle in the CSV file. This results in over-counting the number of vehicles entering/exiting. Note these multiple records do not necessarily have the exact same license plate string – it may output a slightly different string due to the angle of the moving car and its distance to the camera in a sequence frames captured.

Incorrect locating: This type of error occurs when the OCR algorithm incorrectly locates the license plate in an image and attempts to recognize the characters within that area of the image. The output license plate string from these errors are almost always "`OCR NOT READ`" as the recognized plate does not match any of known formats of license plates available in its embedded library. There are rare cases where a non-license

TABLE III
PERCENTAGE OF OF RECORDS REMOVED FROM THE ENTIRE DATASET AT EACH STAGE OF CLEANSING.

	Error types		
Camera	Multiples (redundant)	low OCR score	OCR-Not-Read
Entry	9.7%	12.9%	2.5%
Exit	7.9%	8.5%	8.2%

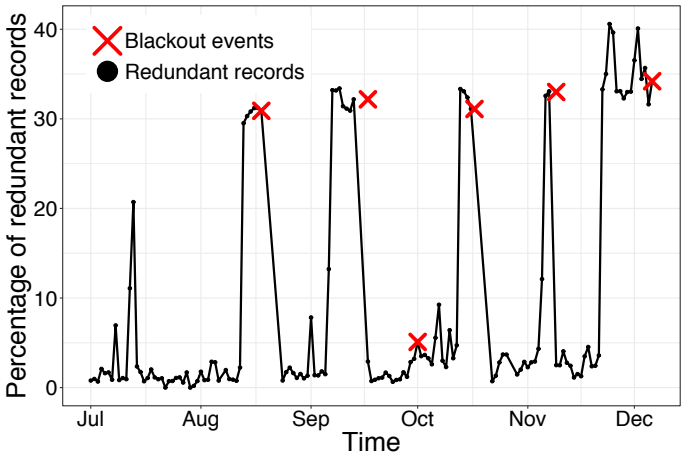


Fig. 6. Rate of error due to redundant (multiple) records rises prior to camera black-outs.

plate object gets partially recognized that their output appears as incomplete strings with low OCR scores of below 60.

Incorrect recognition: Incorrect recognition occurs when the camera successfully locates the license plate, but fails to recognize the characters in the license plate correctly. To quantify the severity of the read errors, we manually inspected 200 images per camera to obtain ground-truth license plate string which we use to compare against the output string of cameras. We employ Levenshtein distance [18] for measuring difference between two string words, *i.e.,* the minimum amount of single character addition, substitution or deletion required to make two strings identical. For example, the Levenshtein distance between the string "`ABC123`" and "`AC123`" is 1, two strings will be identical by inserting a character "`B`" into the string "`AC123`". We compute the Levenshtein distance between ground-truth and recognized strings, and show the CCDF plot of the distance in Fig. 4. It is seen that up to 28% of records for the entry camera and 52% for the exit camera have at least 1 character being mis-recognized (*i.e.,* distance of more than 0 character). We also observe that the majority of these errors were caused by 1 misread character, accounting for 15% for entry and 30% for exit of the total observed records. Note that incorrect recognitions also result in lower OCR scores for the output record. In Fig. 5 we show the CCDF of the overall OCR score of each camera. It is seen that 83% of records for the exit camera have an OCR score of higher than 65, and 85% of records for the entry camera have an OCR score of above 75 – we will use these thresholds to filter out "low OCR score" records in our data cleansing process.

We summarize the distribution of error types for each camera in Table II. It is observed that errors from the entry records

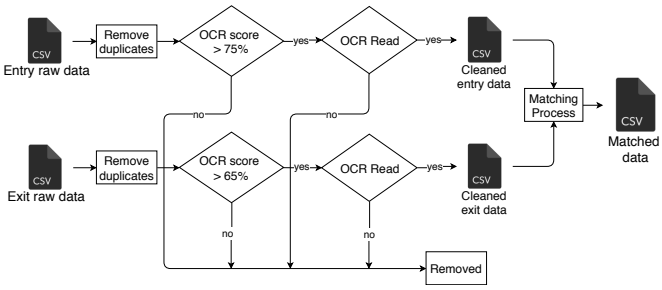


Fig. 7. Data cleaning process.

are largely dominated by multiple-recognitions (59.3%). This relates to an issue for our entry camera we faced several times where the camera view unexpectedly turned black and stopped collecting data (*i.e.,* black-out events). To understand this relationship, we plot in Fig. 6 the time-trace of daily error rate due to redundant records, overlaid by black-out events which required camera reboot. It is interesting to observe that the rate of redundant records steeply rises a few days prior to black-out events (evidenced by red cross markers in the middle of August to middle of December). On the other hand, we can see from Table II that errors of the exit records are mostly from incorrect recognition (51.6%). This is due to the fact that the exit camera is placed at a non-ideal location, causing the camera to capture the license plate at an angle and hence resulting in poor performance of the OCR algorithm. We can also see that 34.9% of errors are from incorrectly locating license plates for the exit camera while this measure is only 4.2% for the entry camera. By inspecting the collected images, we found that moving grass (close to the exit of the car-park and is visible in Fig. 2) gets occasionally detected as a moving object by the exit camera. This does not impact our entry camera, and thus it displays a much lower rate of incorrect locating error.

D. Data Cleansing

We now clean our raw data collected from the two cameras with the following objectives: (a) removing multiple records to obtain the correct count of arrival/departure, (b) removing records of non-vehicle objects incorrectly captured by cameras, and (c) matching license plates captured by both cameras to deduce the distribution of stay-duration in our campus car-park.

Fig. 7 shows our cleansing process with various stages involved. The raw data, first undergoes a function that removes multiple recognitions to remove multiple records that have the same or slightly different strings. Since multiple records of a vehicle appear consecutively in our dataset, a license plate is considered as "redundant" if it re-appears in the next 5 records of the CSV file after its first appearance and the Levenshtein distance between the plates is 2 or less. We note that the distance threshold of 2 is selected because our ground truth observation has shown that this filtering only eliminates 8% and 10% of entry and exit records (Fig. 4), and increasing the threshold to 3 or more will only improve the coverage by less than 10% which is very low considering the trade-off

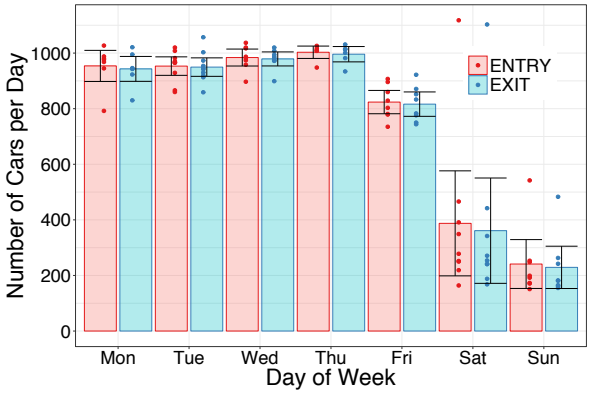


Fig. 8. Number of cars per day, arriving/departing to/from university campus car-park, during teaching period.

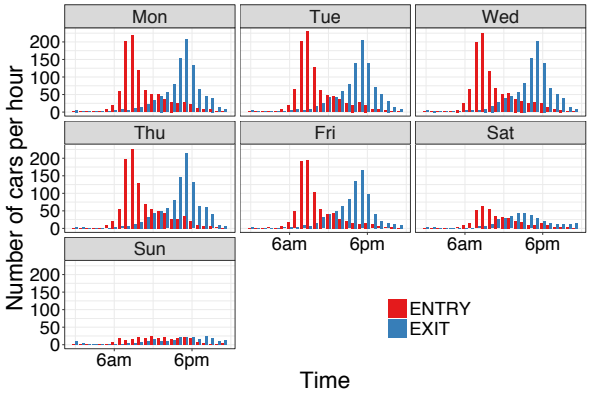


Fig. 9. Average hourly arrival and departure rate of cars for each day-of-week.

for a high chance of plates getting mismatched (two different plates get matched and treated as the same plate). We compare the OCR score of license plates identified as redundant, and the license plate with the highest score is kept and others are removed.

Next, we remove records with low OCR scores – these records again correspond to redundant (multiple) captures of one vehicle. As discussed earlier in Fig. 5, we use filtering threshold of 75 and 65 for OCR scores for the entry and exit cameras respectively. Lastly, we remove all records with `OCR NOT READ` value – those with incorrectly located license plate in the captured image. The pair of cleaned data will be used for arrival/departure counts. A summary of records removal due to each error type is shown in Table III.

As mentioned earlier, we apply the last stage of cleansing to extract vehicles whose records are matched in both entry/exit datasets – the output of this stage will enable us to compute stay duration. Again for matching, we use the Levenshtein distance of 2 or less. In rare cases we found one to many matches in the two cleaned datasets for which we chose the pair with the lowest Levenshtein distance and the highest OCR score. By running the matching process on our cleaned data from July to end of December, the matching process was able to match on average 86% of the cleaned records from both cameras. Remaining unmatched records correspond to (a) vehicles stayed over-night (we process daily dataset individually), (b) vehicles that are captured by one camera (mostly entry camera) but not the other camera. We found from our spot measurements that only a small number of vehicles stayed overnight (*i.e.,* 23 cars on average for a sample size of 11 nights), and thus they would not have significant impact on the car-park usage patterns. We, therefore, analyze our dataset on a per-day basis (*i.e.,* midnight-midnight).

IV. Analysis and Insights into Usage Pattern

In this section, we analyze our cleaned data (obtained from the previous section) spanned 23 weeks of teaching and non-teaching periods in 2018, to highlight the usage pattern of the campus car-park across various temporal dimensions including time-of-day, day-of-week, week-of-semester, and semester break/exam periods.

A. Arrival and Departure Pattern

We begin with Fig. 8 that depicts the average daily number of cars entering to (red bar) and exiting from (blue bar) the car-park for each day-of-week, during 13-weeks of teaching period. Red and blue dots represent actual measurements and error bars represent 95% confidence interval of data-points. It is seen that about 950 to 1000 cars use (*i.e.,* enter/exit) the car-park during a weekday (*i.e.,* midnight-to-midnight), except Friday for which the number slightly drops to about 830 due to fewer lectures. During weekends, the number of cars using the car park goes below 400. Also, we observe narrower error bars for weekdays compared to weekends, suggesting predictable usage patterns for weekdays. In contrast, the car-park usage varies significantly on weekends ranging from 200 to 600. Interestingly, it goes beyond 1000 cars for one particular Saturday (*i.e.,* 1 September 2018). After checking the University event calendar, we found that this corresponds to the University Open Day which typically attracts a large number of high-school students and their families. The University provides free parking for all Open Day Visitors.

Fig. 9 illustrates the average hourly count of arriving cars (red bars) and departing cars (blue bars) by times-of-day, for each day of the week (separate graphs per day). It is seen that during weekdays, the arrival rate starts rising steeply from 6am, peaks at 9am-10am, and falls slowly afterwards. The departure process displays a similar bell-shape pattern but shifted in time by about 8 hours, *i.e.,* rising in the afternoon, peaking at 5pm-6pm, and falling afterwards. During peak times, we see about 200 cars per hour enter/exit the parking lot – these numbers are slightly lower for Friday (*i.e.,* 194 and 165 cars per hour). We also observe an irregular pattern for weekends where car-park usage heavily depends on events hosted on campus. Our findings of weekday arrival/departure pattern corroborate with other studies [19].

Fig. 10 depicts the distribution of arrival and departure probability for the five weekdays (in a stacked representation with Friday on top and Monday at the bottom). To better illustrate the pattern, we have color-coded five time intervals: orange for prior-sunrise (12am-6pm), yellow for morning-peak (6am-11am), green for afternoon-offpeak (11am-4pm), blue for evening-peak (4pm-8pm), and purple for night-time (8pm-

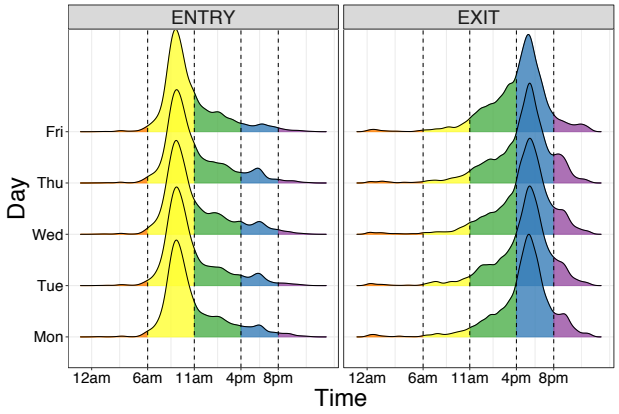


Fig. 10. Distribution of arrival (Entry) and departure (Exit) time across weekdays.

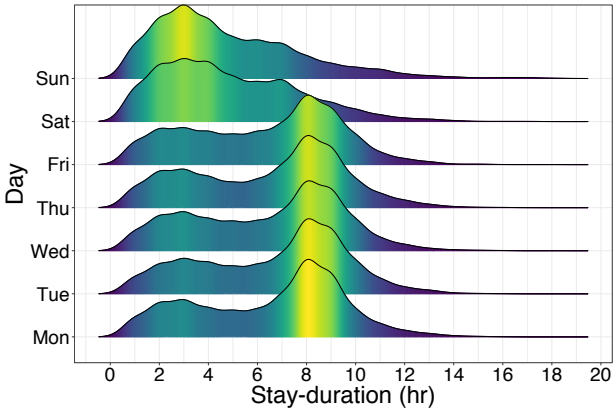


Fig. 11. Density distribution of stay-duration for each day of week.

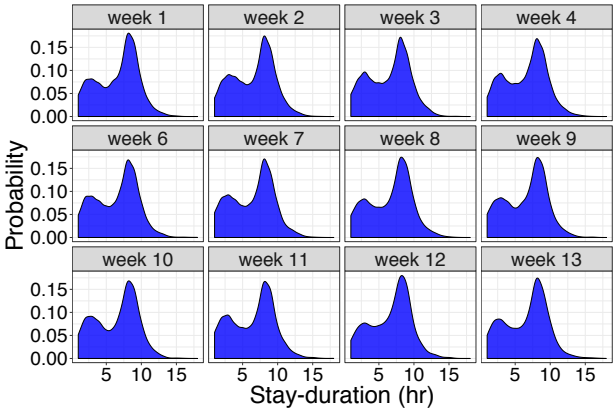


Fig. 12. Density distribution of stay-duration for each week of semester.

12am). Note that the difference in absolute numbers is not seen here for Friday due to normalization. Unsurprisingly the distribution plot shows that both patterns of arrival and departure are very consistent for most time-slots across weekdays, with the strongest similarity observed during peak hours. As before, we observe some different trends for Friday. Note that the blue part of the arrival curve (4-8pm) shows a less pronounced peak compared to other weekdays. This suggests fewer arrivals to campus on Friday evening as opposed to other weekdays evening time. This can be attributed to the fact that there are very few evening lectures running on Friday. A closer examination of the exit curve for Friday reveals that the area in green is larger than the corresponding regions of other weekdays. The purple part of the curve is also considerably flatter than the other days. This suggests higher percentage of cars leaving during early afternoon time and lower percentage of cars exiting the car park after 8 pm on Fridays compared to Mondays-Thursdays.

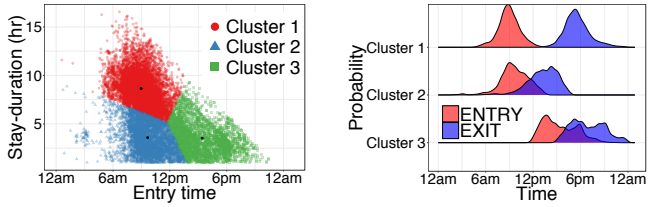
Additionally, we looked at the arrival/departure pattern during different periods of the academic calendar including orientation week (O-Week) which is largely geared for new entrants to get acquainted with the university, regular teaching weeks, a week long mid-semester break, a study-break week right before final exams, exam periods and a lull period before the end of year holiday shutdown during which undergraduate students are away and campus attendance for everyone else progressively reduces. The general patterns are similar to Fig. 9 for all periods. One main difference observed was the morning arrival peak occurred one hour earlier (*i.e.,* 8am-9am) during the O-Week and other non-teaching periods. The former can be attributed to the fact that a portion of students and admin staff arrive early during the O-week for setting up various activities. The latter can be ascribed to the fact that students are less likely to be on campus during the non-teaching period and thus the morning peak time is largely determined by the arrival pattern of staff who typically arrive earlier (between 8-9am) than students. Also, the actual count of arrival/departure slightly drops (*i.e.,* around 50 cars) during mid-semester break and other non-teaching period. This suggests that the majority of car-park users are university staff members which is not unexpected due to high price of parking permits at our campus.

B. Stay Duration Pattern

Our dataset allows us to obtain insight into stay-duration of car-park users given that the license plate numbers are captured at entry and exit points. Note that we rather to analyze the distribution of these patterns across various user groups, instead of individuals. For this analysis we used the cleaned dataset after matching process mentioned in Section III-D.

Fig. 11 shows the distribution of stay duration (in hours) across days-of-week. It is clearly seen that the distribution is maximized at around 8 hours for Mondays to Fridays – this is consistent with the standard working hours in Australia, which is 7.6 hours a day [20]. We also observe that users tend to use the car-park for about 2-4 hours during weekends - these users are likely to attend events hosted on or near-by the campus or Postgraduate students attending Saturday lectures.

We further look at the stay duration pattern for each week of the semester in Fig. 12. A consistent pattern of bi-modal (double-peaked) distribution is observed for each week. The peak stay duration, as expected, centers at 8 hours, a typical full-time work day. The second peak centered between 2 to 3 hours, highlighting usage patterns for weekends as well as students/visitors.



(a) stay-duration versus entry-time for each cluster.

(b) distribution of entry-time and exit-time for each cluster.

Fig. 13. Clustering results.

C. Parking Behaviour Users

We now cluster car-park users using k-means algorithm [21] to identify parking behaviour of certain user groups on weekdays during teaching period. For each car we extract three features including arrival time, departure time, and stay duration. Three clusters were selected based on the elbow method. Fig. 13 shows the result of our clustering. We show in Fig. 13(a) the scatter plot of stay-duration versus entry time for individual vehicles in our dataset – clusters are color-coded. It is seen that "Cluster-1" (shown by red region) corresponds to users who enter the car-park early and stay more than 8 hours (center of cluster-1 is located at 9am entry time and 8.8 hours of stay duration) – this cluster represents full-time staff. "Cluster-2" users enter the car-park at about the same time as Cluster-1 and stay shorter (*i.e.,* less than 5 hours) – morning visitors and typical students. Lastly, "Cluster-3" users are those who enter late and stay for a short period (with center at 3pm entry time and 3.5 hours of stay) – this cluster is likely to denote afternoon visitors and postgraduate coursework students who attend evening classes.

In Fig. 13(b), we show the distribution of arrival and departure time for each cluster. We observe that the distributions for Clusters-2 and -3 are relatively wider than of Cluster-1. Again, it is seen that full-time staff in Cluster-1 typically enter at about 9am and exit at about 5pm. Similar to scatter plot, Cluster-2 users arrive early and leave early too. Lastly, as expected Cluster-3 users enter in the afternoon and exit in the evening. We also observe a significant overlap in the exit and entry distributions for these users, which suggests that they tend to stay on campus for a short period (about 3 hours).

V. Modeling Car-Park Usage

We observed in the previous section that car-park usage varies by times-of-day, days-of-week, and weeks-of-semester which leads to under-utilization of car-park. This presents an opportunity for campus managers to dynamically offer car-park spaces to other users such as car-sharing service providers with minimal impact on private car users. In this section, we develop a model for the usage of car-park to determine the potential of new offerings. We assume that the car-park with arriving/departing vehicles represents a queuing system where system states are captured by a continuous-time Markov chain and arriving/departure are modeled as non-homogeneous Poisson processes.

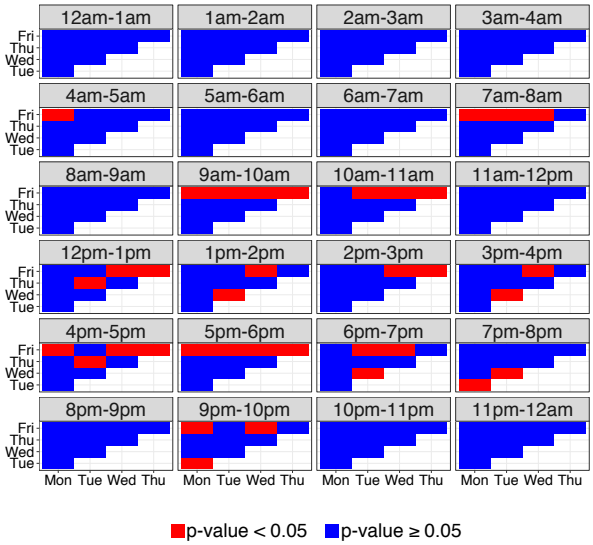


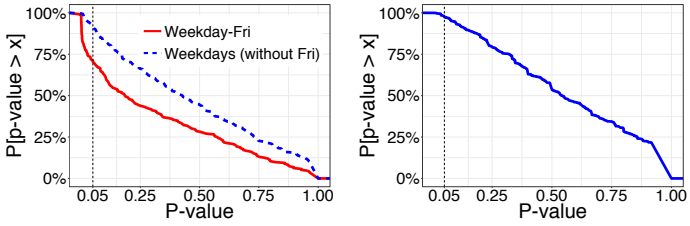
Fig. 14. Wilcoxon test results of hourly arrival rate (teaching period).

A. Fitting a Poisson Distribution

We observed in §IV-A that arrival and departure rates of vehicles highly depend on time-of-day, where morning and afternoon peaks are expected at around 9am and 5pm respectively. This suggests a non-homogeneous Poisson process where the average rate of arrivals is allowed to vary with time. We also observed that this general hourly pattern of arrival/departure seems to be consistent across weekdays as well as weeks of semester. However, this does not mean that the same hourly rate can be used to capture incoming/outgoing processes for all days and weeks. Hence, statistical tests to identify the similarity or variation of the average hourly rate is needed to be performed.

We employ *two-sample Wilcoxon rank sum test* [22] to determine whether the difference in means of two independent samples is significant. For instance, the test allows us to determine if there is any significant difference between the hourly rate observed on Monday and the rate observed on Friday during a certain hour-of-day. The Wilcoxon test generates an output "p-value" (ranged between 0 and 1) indicating the risk of concluding that a difference exists when there is no actual difference. The p-value is compared to a significance level of 0.05 to accept "null hypothesis" (*i.e.,* there is no difference between the tested samples). The p-value smaller than 0.05 means that we safely conclude that the two samples differ.

Fig. 14 shows the Wilcoxon test results of arrival rate for individual hours-of-day in facets, each comparing a pair of two different weekdays (x-axis versus y-axis) – we omit results of the departure rate due to space constraint. We use color-coded cells (red for p-value < 0.05, and blue for p-value ≥ 0.05) to indicate statistical difference (*i.e.,* red cells) or similarity (*i.e.,* blue cells) in distribution of data for a given pair. Focusing on results of the arrival rate in Fig. 14, red-color cells are seen when Friday is tested against other weekdays, especially during morning and afternoon peak hours (*i.e.,* 9-10am and 4-6pm). On the other hand, pairing non-Friday weekdays typically results in blue cells for most of the time, highlighting



(a) Daily pairs. (b) Weekly pairs.

Fig. 15. CCDF of p-value from Wilcoxon test for both arrival and departure rates (during teaching period).

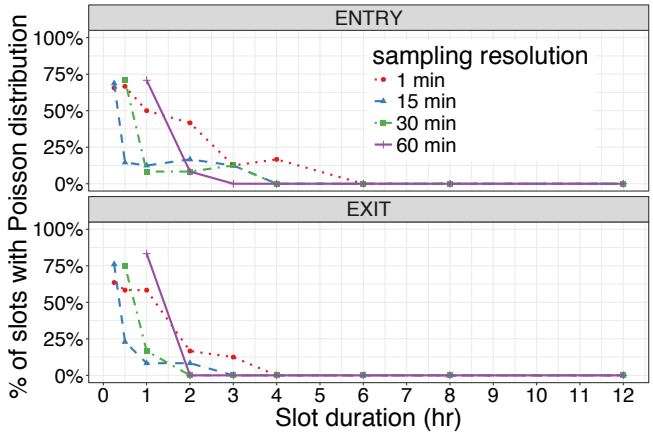


Fig. 16. Fraction of slots following Poisson distribution (p-value ≥ 0.05) for varying slot duration and sampling resolution.

the strong correlation. We made a similar observation in Fig. 8 when aggregate daily count was considered. Also, Friday's hourly pattern deviates from other weekdays by departure rate too – Friday rows are full in red during 5-8pm.

We show in Fig. 15 the CCDF plot of p-value obtained from Wilcoxon test. In Fig. 15(a) two daily cohorts are considered: (a) pairs of non-Friday weekdays (shown by dashed blue lines), and (b) pairs of Friday and other weekdays (shown by solid red lines). It is seen that 92% of pairs among non-Friday weekdays have p-value > 0.05, whereas this fraction is less than 75% for pairs of Friday-weekdays. Also, for aggregate weekly cohorts (*i.e.,* pairs are chosen from weeks-of-semester) shown in Fig. 15(b), 97% of week pairs display p-value>0.05, indicating no (or very few) difference in distribution across weeks of semester. The overall results of Wilcoxon-test suggest that two models of Poisson process are needed for the arrival/departure rate of cars during weekdays of teaching period: one model for Mondays-Thursdays, and one for Fridays.

We next attempt to fit Poisson distribution to each of the two cohorts in our dataset. To achieve this we employ Chi-Square Goodness of Fit Test [23] and consider two parameter choices: (a) slot duration over which the arrival/departure rate is relatively constant, (b) sampling resolution for the rate (*e.g.,* per-minute or per-hour). We tune our parameters of choice by varying slot duration (15-min, 30-min, 1-hr, 3-hr, 6-hr and 12-hr) and sampling resolution (1-min, 15-min, 30-min, and 60-min). In each slot, arrival and departure rates are sampled over

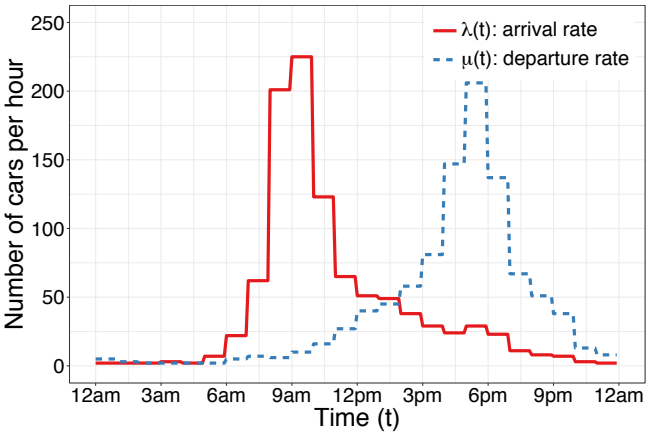


Fig. 17. Hourly rate of arrival and departure for a day from Mon-Thu cohort.

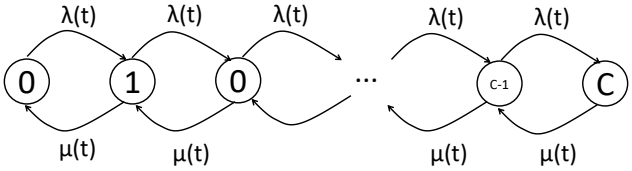


Fig. 18. Continuous-time Markov chain of the car-park.

a moving window (*i.e.,* equals to the sampling resolution) with a step of one minute. This method of data sampling conforms to the memoryless property of the Poisson process, where the future arrival (or departure) is independent of the event in the past.

For each combination of parameters we quantify the fraction of time that our data follows a Poisson distribution. Fig. 16 shows our fitting experiment results. Each line corresponds to a sampling resolution value, shown on the legend. It is seen that the best result (*i.e.,* highest fraction of fitting) obtained when the slot duration is 1 hour and the rate is computed on a per-hour basis. We, therefore, compute the hourly rate of arrival (λ) and departure (μ) – Fig. 17 shows the results for the cohort of Mon-Thu.

B. Modeling Car-park Usage as a Queuing System

Given the rate of arrival and departure, we model the car-park as an M/M/1/C queue whose operation is visualized as a continuous-time Markov chain shown in Fig. 18. Each state of the chain represents the number of occupied space in the car-park with total capacity C, and $\lambda(t)$ and $\mu(t)$ are the rates of arrival and departure respectively. If the car-park becomes full (*i.e.,* at state C), the arriving vehicle is rejected.

We developed a discrete-event simulator in R, to quantify the probability of rejection considering a futuristic scenario whereby the campus Estate Management allocates a fraction of the car-park spaces to car-sharing services and the remaining portion gets allocated to existing private cars. The primary aim for the Estate Management would be to optimize the use of parking space for generating revenue from shared cars with minimal impact on private car users.

In order to simulate the dynamics of shared vehicles, we begin by assuming that 20% of the existing users subscribe

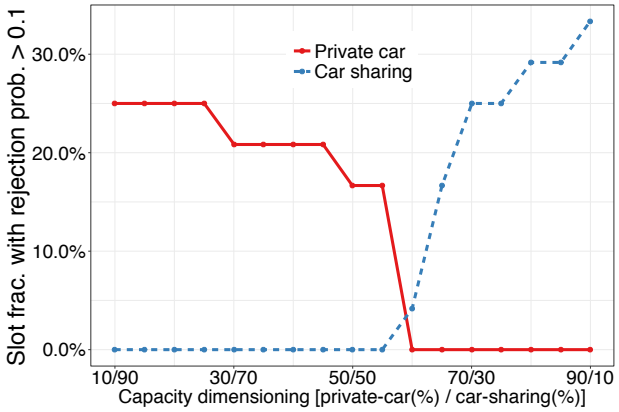


Fig. 19. Rejection probability of private/shared cars varies by car-park dimensioning structure.

and use car-sharing services on a regular basis (the reasons can be due to the potential reduction in travel cost, avoiding parking fee, and eliminating the need to own a car) and 80% of the current users continue commuting to the university by a private car. We also assume that there will be new users of shared vehicles – these additional users are modeled by random sampling of a binomial distribution, $B(100, 0.4)$, over each slot of 60 minutes – meaning a total of 100 potential vehicles with a 40% chance for each vehicle to use the car-park within the slot. Given these assumptions, we compute the rates of arrival and departure for both private and shared vehicles.

Our campus car-park has a total capacity of 895 vehicle spaces. In our simulation we split the car-park into two queues (*i.e.,* one for private cars and one for shared cars) for various dimensioning configuration. We start with 10% of spaces allocated to private cars and 90% to shared cars, and increase/decrease the allocation for private/shared cars by 5% at each step untill dimensioning scenario 90%/10% is achieved. For each run of simulation, we quantify the probability of rejection (*i.e.,* no space left) during each slot for both groups of private and shared cars. Fig. 19 shows the fraction of slots that a private/shared car experiences a rejection probability of 0.1 or more. As we expect, increasing the portion allocated to private cars would reduce the rejection probability (shown by sold red lines) for this users group and increase the rejection for shared car users (shown by dashed blue lines). The optimal dimensioning can be obtained by considering the minimum rate of rejection tolerable for private users traded with additional revenue from car-sharing service providers.

VI. Conclusions

Digital transformation in transport industry demands large organizations such as universities to revisit the operation of their expensive on-campus parking facilities. In this paper we have outlined our experiences in designing and deploying a monitoring system for a real car-park at our university campus. We collected data over 23 weeks (covering both teaching and non-teaching periods) and cleaned it for analysis. We then analyzed the usage data and highlighted insights into car arrival and departure patterns as well as users parking behaviour. Finally, we developed continuous-time non-homogeneous Markov models using historical data and showcased (via simulation) its use for space allocation planning.

References

[1] B. Council, "The shape of things to come: higher education global trends and emerging opportunities to 2020," British Council, Tech. Rep., 01 2012.

[2] L. dellOlio, R. Cordera, A. Ibeas, R. Barreda, B. Alonso, and J. L. Moura, "A methodology based on parking policy to promote sustainable mobility in college campuses," *Transport Policy*, Apr 2018.

[3] A. Filipovitch and E. F. Boamah, "A systems model for achieving optimum parking efficiency on campus: The case of minnesota state university," *Transport Policy*, vol. 45, pp. 86–98, Jan 2016.

[4] (2018) Car Sharing Market Size By Model. https://www.gminsights.com/industry-analysis/carsharing-market.

[5] E. Polycarpou, L. Lambrinos, and E. Protopapadakis, "Smart parking solutions for urban areas," in *Proc. IEEE WoWMoM*, Madrid, Spain, June 2013.

[6] D. Bong, K. Ting, and K. Lai, "Integrated approach in the design of car park occupancy information system (COINS)," *IAENG International Journal of Computer Science*, vol. 35, no. 1, Feb 2008.

[7] Z. Pala and N. Inanc, "Smart parking applications using rfid technology," in *Proc. IEEE Annual RFID Eurasia*, Istanbul, Turkey, Sep 2007.

[8] S. Shaheen, "Smart parking management field test: A bay area rapid transit (bart) district parking demonstration," *UC Davis: Institute of Transportation Studies*, 2005.

[9] (2017) SFpark. http://sfpark.org/.

[10] M. Caliskan *et al.*, "Predicting parking lot occupancy in vehicular ad hoc networks," in *Proc. IEEE Vehicular Technology Conference*. Dublin, Ireland: IEEE, Apr 2007.

[11] R. Lu *et al.*, "An intelligent secure and privacy-preserving parking scheme through vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2772–2785, July 2010.

[12] G. Yan, W. Yang, D. B. Rawat, and S. Olariu, "Smartparking: A secure and intelligent parking system," *IEEE Intelligent Transportation Systems Magazine*, vol. 3, no. 1, pp. 18–30, Apr 2011.

[13] T. Sutjarittham, H. Habibi Gharakheili, S. S. Kanhere, and V. Sivaraman, "Experiences with IoT and AI in a Smart Campus for Optimizing Classroom Usage," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[14] T. Sutjarittham, H. Habibi Gharakheili, S. S. Kanhere, and V. Sivaraman, "Realizing a Smart University Campus: Vision, Architecture, and Implementation," in *Proc. IEEE ANTS*, Indore, India, Dec 2018.

[15] (2019) ANPR Access HD - HD license plate camera for vehicle access control. https://bit.ly/2HixxU5.

[16] S.-L. Chang, L.-S. Chen, Y.-C. Chung, and S.-W. Chen, "Automatic license plate recognition," *IEEE transactions on intelligent transportation systems*, vol. 5, no. 1, pp. 42–53, Mar 2004.

[17] V. Lyons, *Guidance on ANPR Performance Assessment and Optimisation*, 1st ed., Home Office of UK government, Centre for Applied Science and Technology, Sandridge, St Albans, AL49HQ United Kingdom, Mar 2014.

[18] S. Takahashi *et al.*, "Travel Time Measurement by Vehicle Sequence Matching Method," in *Proc. SICE-ICASE*, Bexco, Busan, Korea, Oct 2006.

[19] T. Rajabioun, B. Foster, and P. Ioannou, "Intelligent parking assist," in *Proc. IEEE Mediterranean Conference on Control and Automation*, Chania, Greece, 2013.

[20] (2019) Maximum weekly hours. https://bit.ly/1I14M9L.

[21] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[22] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[23] V. Ricci, "Fitting distributions with r," *Contributed Documentation available on CRAN*, vol. 96, 2005.