November 13, 2017

A Method For Detecting Scheduled-Service Vehicles By Crowdsensing Of On-Board Beacons

Lin Chen

Alexander Fabrikant

Rachael Morgan

A METHOD FOR DETECTING SCHEDULED-SERVICE VEHICLES BY CROWDSENSING OF ON-BOARD BEACONS

ABSTRACT

The technology detects scheduled-service vehicles using a model generated based on onboard beacon location data. A trajectory of the onboard beacon may be determined by correlating the beacon detection data with the location data of the detecting devices. The trajectory may be analyzed to derive features of the beacon trajectory that correlate with a periodic motion typical of scheduled-service vehicles. Such features may be used to train a machine capable of learning patterns from data to generate a model that classifies whether a beacon trajectory is that of a scheduled-service vehicle. Various beacon trajectories may be classified by the generated model as either a scheduled-service vehicle or not. Once certain beacons are identified as onboard scheduled-service vehicles, further trajectory data may be gathered on such beacons to map out accurate public-transit routes and schedules, as well as to provide real-time locations and changes to routes and schedules.

BACKGROUND

Transit agencies around the world design various routes and schedules, and provide the information on the routes and schedules to their customers. Some transit agencies also collect and provide real-time updates about changes in routes and schedules. Collecting such information from the transit agencies directly on a global scale involves complex partnership agreements and engineering done separately for each transit agency around the world, which may take many years to accomplish, if at all feasible. Further, the reliability of the information collected directly from the transit agencies may depend greatly on the particular transit agency.

DETAILS

The technology generally relates to detecting scheduled-service vehicles by detection of onboard beacons. The technology may include receiving beacon detection data and detecting device location data from various devices, correlating the detecting device location data with the beacon detection data to obtain a trajectory for each individual beacon, deriving features of the beacon trajectory that correlate with episodes of spatiotemporally periodic behavior having frequencies typical of a public-transit vehicle, and training a machine capable of learning patterns from data to generate a model that classifies whether a beacon trajectory is that of a scheduled-service vehicle. Using the generated model, various beacon trajectories may be classified as either that of a scheduled-service vehicle or not. Once certain beacons are identified as located onboard scheduled-service vehicles, information regarding such beacons may be gathered to determine public-transit routes and schedules, as well as to provide real-time locations and changes to routes and schedules.

Many vehicles are equipped with a device that has a beacon capable of broadcasting a unique identifier, which may be detected by nearby devices in the course of the devices' normal operation. For example, a WiFi Internet Access Point (AP) may be detected by a phone that is trying to geolocate itself with a GPS. Similarly, many Internet of Things (IoT) devices may broadcast their unique identifiers as if they are an Internet AP. For example, a Bluetooth Low Energy (BLE) beacon might be detected by a phone scanning for a hardware to peer with. Even if the phone does not connect to or pair with the broadcasting device, the phone's detection of the broadcasting device, together with the phone's location, may provide an estimate of the vehicle's location at the time of the detection. Other example beacons may be a wireless fare payment terminal, a dashboard camera, etc.

The detecting devices may be any device capable of detecting unique identifiers broadcasted by a beacon, such as the beacons exemplified above. For example, the detecting devices may be BLE detectors attached to bus stops, whose locations are already known. As another example, the detecting devices may be satellites capable of receiving signals from a beacon on a vehicle. As another example, the detecting devices may also be user devices having a location determination system, such as a mobile phone, a wireless-enabled PDA, a tablet PC, a netbook, a smart watch, a head-mounted computing system, etc. The location determination systems may determine location based on GPS, wireless access signal strength, images of geographic objects such as landmarks, semantic indicators such as light or noise level, etc.

Therefore, location data of a beacon may be collected in a number of ways. Most directly, location data of a beacon onboard a public-transit vehicle may be uploaded directly from the public-transit vehicle itself. Indirectly, for example, location data of various beacons may be collected from satellites. As another example, the location data of various beacons may be collected from detectors attached to bus stops. Still further, with permission, detection data of various beacons may be collected from user devices, which, when correlated with location data of the user devices themselves, may provide a trajectory of the vehicles on which the beacons are attached.

Data may be collected only from users who agree to share the location and detection data of their devices. For example, data may be collected only from users who agree to upload the location and detection data to a cloud service. In addition to providing the users the option to supply data regarding recognition of beacons at any time, provisions are made for protecting privacy and security. For example, only generic non-personal data may be collected, and the data provided in response to requests may be an aggregation of generalized non-personal information. Collected data is transmitted from the collection devices to the server over secure channels to prevent interception or eavesdropping. Relative and absolute thresholds are established across all collection data types to prevent association of particular data with particular individuals or devices with certainty or statistical near-certainty. Additional provisions include anonymization of personally identifiable information, aggregation of data, filtering of personal information, encryption, hashing or filtering of personal information to remove personal attributes, time limitations on storage of information, or limitations on data use or sharing.

An example system to implement the technology may include at least one server computing device capable of learning patterns from data. The server computing device may be in communication with one or more detecting devices capable of detecting beacons, as well as a storage system, through a network. The server computing device may contain one or more processors, memory and other components typically present in general purpose computing devices.

Data from a user device over some period of time may be collected in two sets: a location data set indicating the location of each user device at various times, which may include a latitude and a longitude from the GPS receiver of the user device and the timestamp for each GPS reading, and a detection data set indicating the detection of various beacons by each user device, which, for example, may include the WiFi AP MAC address of each beacon and the timestamp of each detection.

A trajectory for each beacon may be determined from the location and detection data collected. As shown in Fig. 1 below, detection data from one user device may be aggregated for each unique identifier. Each detection data point (shown as a MAC address and a detection timestamp) may be matched with a location data point (shown as a GPS coordinate with a localization timestamp) having a location timestamp closest to the detection timestamp. The resulting aggregated data points for each beacon may each include a unique identifier of the detected beacon, a timestamp of the detection, a latitude of the detecting user device, a longitude of the detecting user device, and a timestamp of the localization of the detecting user device.

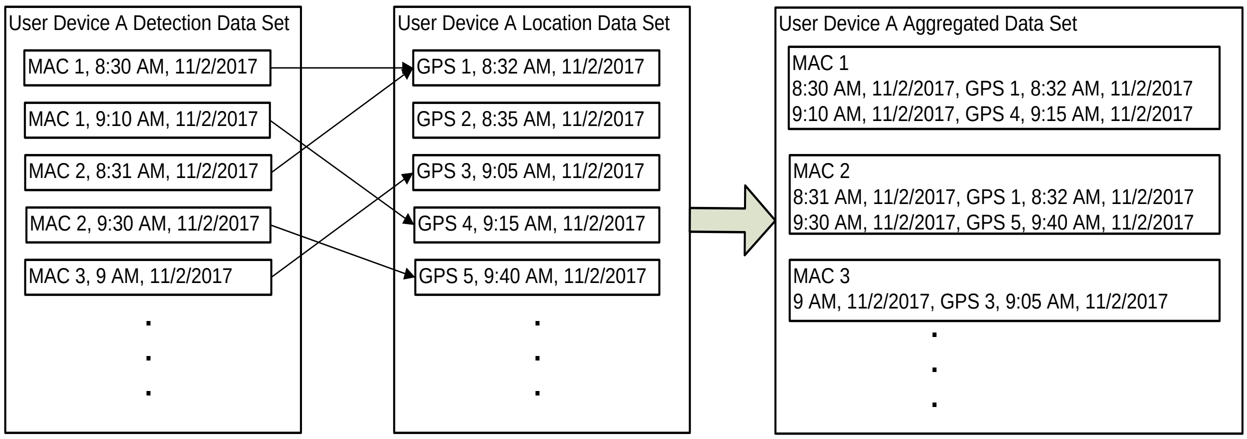

Fig. 1

As shown in Fig. 2 below, the aggregating and matching may be repeated for each user device for each unique identifier. One of either the detection or the localization timestamps may be removed, so that each data point in the record may contain a unique identifier, a timestamp, a latitude, and a longitude. The resulting record is a list of all the data points available for each beacon, representing a trajectory of the beacon. Therefore, for each unique identifier, a series of 3-tuples of timestamp, latitude, and longitude maps out a trajectory of the beacon.

In addition, though not shown, accuracy of the trajectories may be improved by removing certain bad data points from the record. For example, in some cases the timestamp of the detection may be significantly different from the timestamp of the localization, by setting a threshold difference, such poor matches may be removed from the record. As another example, data points that appear to contain device error, e.g., having a timestamp in the future, may also be removed from the record.

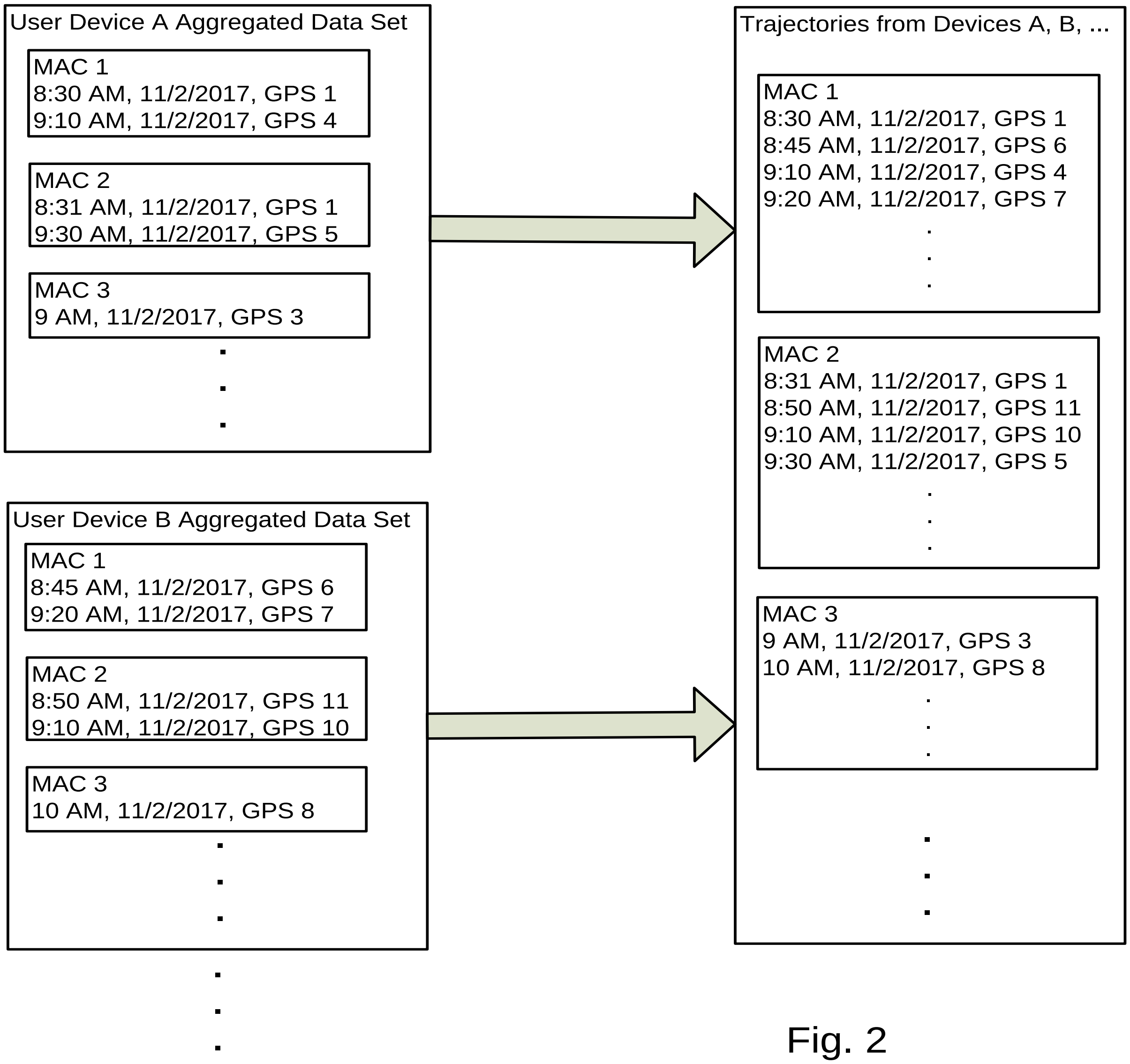

Fig. 2

The trajectories may be further processed to reduce noise. For instance, detection data may be dense at a location where many user devices are available (e.g., the timestamps of the data points may be overlapping or very close to each other), and sparse at another location where few user devices are available (e.g., the timestamps of the data points may be spaced far apart). Such a variation in data density does not correspond to a frequency of the vehicle and should be corrected by linear interpolation and sampling. In addition, if the detection data is not evenly spaced in time, it may be more difficult to perform autocorrelation, such as correlation of the detection data with a time-delayed copy of itself as a function of the time-delay. One example solution may be to perform a linear interpolation on the trajectory to produce latitude and longitude that are piecewise linear functions of time. Then, by sampling these functions at a predetermined sampling interval, e.g., 20 seconds, the resulting trajectory T may be temporally evenly spaced, i.e. spaced by the sampling interval of 20 seconds. Because after sampling, the trajectory T is temporally evenly spaced, the timestamps may no longer be needed, so the trajectory T may simply be a set of data where each data point is a pair of latitude and longitude, i.e., a time series.

An appropriate trajectory length may be chosen such that the trajectory is long enough to observe the periodic motion of a public-transit vehicle, yet not unnecessarily long to complicate computation. For example, a trajectory length of one day may be chosen.

As may be expected, some of the devices detected may be stationary, and among those devices that are moving in some trajectory, many may not be moving in a periodic fashion. The trajectory data of known beacons, i.e., ones known to be onboard a public-transit vehicle or not, may be used to train one or more computing devices to learn how to classify whether a beacon's trajectory shows that it is onboard a public-transit vehicle.

Any of a variety of different types of models may be used to generate the classification model, for example, a linear model, a boosting tree model, a random forest model, a neural network model, or a recurrent neural network model. In one example, a neural network with two one-dimensional convolutional layers followed by three fully-connected layers may be used.

Features that may be important in training the computing devices may be ones that indicate the periodicity of a beacon trajectory. For example, hotspots may be determined from the trajectory of the beacon. Hotspots are locations where a beacon is frequently detected, e.g., for a public-transit vehicle, hotspots may be bus stops. In one example, the Earth may be discretized into S2 cells of a predetermined level, and a subset S of S2 cells containing at least one data point in the trajectory T may be collected. A set of distances D between a location t in the trajectory T to each S2 cell in S may be determined, and a Gaussian function H with a predetermined smoothing parameter ɤ may be computed of the distances D. The S2 cells corresponding to a predetermined top % of the Gaussian function H may be chosen as candidate spots. A hotspot may be determined as a set of neighboring candidate spots. In another example, the S2 cells corresponding to a top number of data points in trajectory T may be chosen as candidate spots. In this case, a smaller smoothing parameter, e.g., 700 meters, may be chosen.

For each hotspot, a binary sequence may be computed from the trajectory T of the beacon. Fig. 3 below shows an example binary sequence β for a hotspot C. The binary sequence represents whether the beacon is at the hotspot C at any given time, i.e., when the value is 1 at a certain time, that means the beacon is at the hotspot C. As may be expected from a public-transit vehicle, the binary sequence shown in Fig. 3 is periodic.

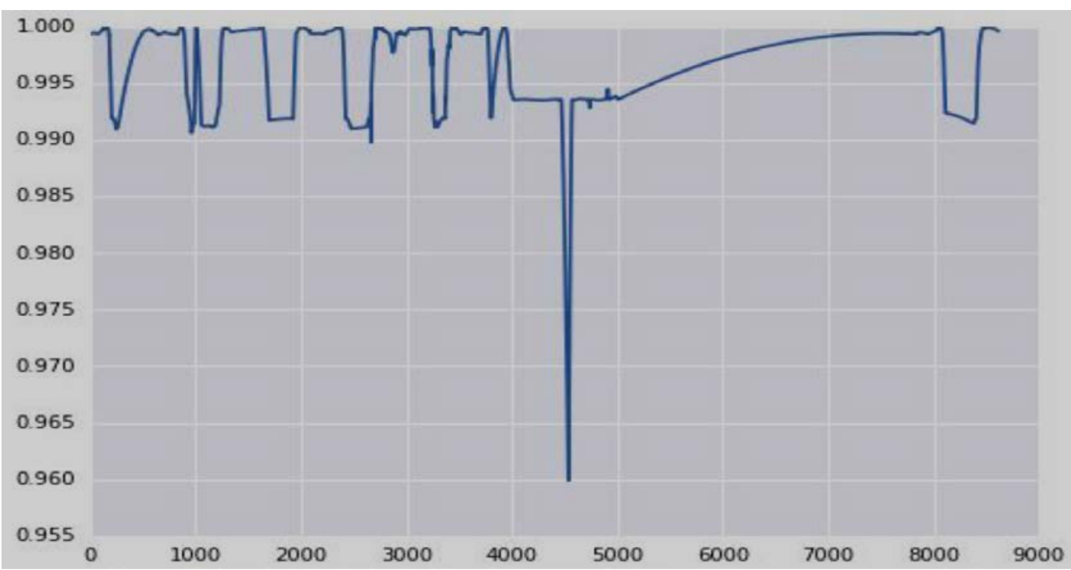

Fig. 3

For each hotspot, autocorrelation of its binary sequence may be performed. For example, Fig. 4 below shows an autocorrelation function of the binary sequence β shown in Fig. 3. Each peak in Fig. 4 represents a period. For example, in Fig. 4 below, the first peak occurs at around 700, since the sampling period is 20 seconds in this example, this means that the first peak occurs at approximately 14000 seconds or approximately 4 hours. The location of such peaks for one or more hotspots may be used as features of the model. The autocorrelation function may be any of a variation of autocorrelation functions, for example, a cyclic autocorrelation function.

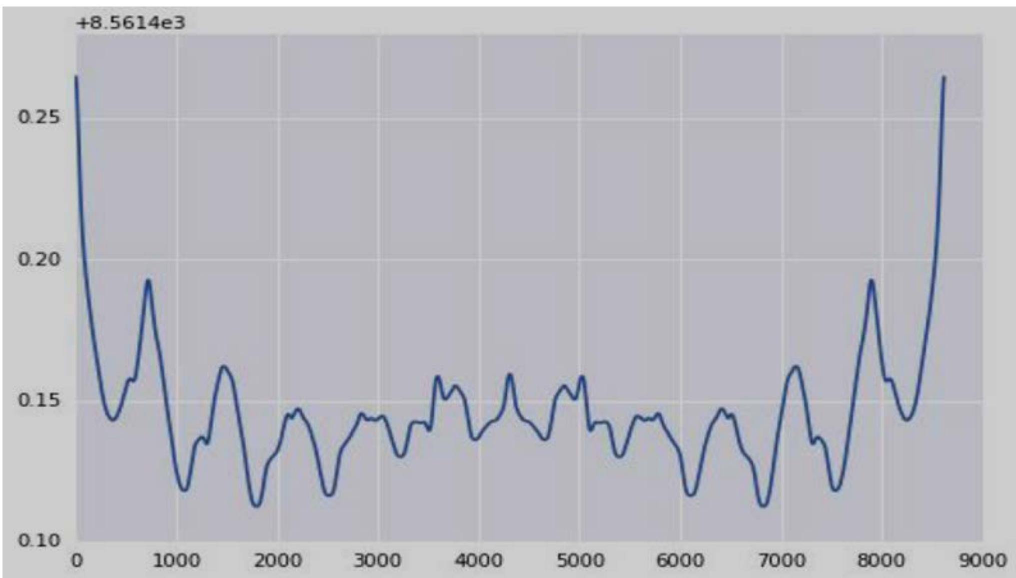

Fig. 4

The binary sequence for an improperly chosen hotspot may not show regular periods, in such cases, a threshold may be applied to the binary sequence. Filtering may also be applied to the binary sequence before autocorrelation, for example, a sliding window maximum filter.

As an alternative or in addition to computing autocorrelation functions of binary sequences at hotspots, Fourier transforms may be performed directly on the trajectories of the beacons. The resulting frequency data may be features used to generate the model. The Fourier transform may be any of a variety of Fourier transforms, including a non-uniform Fourier transform, a discrete Fourier transform, a fast Fourier transform, etc.

Additional features may also be used to generate the model. For example, the number of distinct devices that report observing a beacon may indicate whether the beacon is onboard a public-transit vehicle, since such vehicles may have many passengers. As another example, whether the trajectory forms loops may be another feature likely indicating whether the beacon is onboard a public-transit vehicle.

Using the generated model, trajectory data of various beacons may be classified as either that of a public-transit vehicle or not. Once certain beacons are identified as onboard scheduled-service vehicles, data on such beacons may be gathered to map out public-transit routes and schedules, as well as to provide real-time locations and changes to routes and schedules.

The technology described here provides ways to automating the data collection process on public-transit routes and schedules. The technology described are robust to a variety of noises. Further, low location history sampling rates does not drastically reduce the quality of the model. Still further, the technology has a low dependence on frequent transit usage, i.e., the method operates well even in areas where very few users use public-transit, such as the case in North American suburbs.