Cell Phones and Microsimulation: Advancing the State of the Practice

Smith Siromaskul, P.E., Steve Speth, P.E., P.T.O.E

Abstract

Cell phone tracking technology is currently used to estimate travel times and average speeds along major corridors in major cities. Cell phone signals are anonymized, aggregated and processed to provide traffic information which can be consumed by a variety of applications such as Google Maps. The accuracy of the data collected in this fashion has been verified by further studies, but its use has been fairly restricted to real-time travel monitoring.

In traditional traffic engineering, travel time, origin-destination (OD) and speed studies are often collected for various purposes, including the calibration of micro and macrosimulation models. The collection of this data often requires a significant financial expenditure for relatively meager results such as a handful of travel time runs along a given corridor. Hardware or software must be purchased and man-hours must be invested just to obtain the data. Processing the data is also time-consuming and expensive.

Traditional data collection procedures may be replaced, saving time and money while providing a larger set of data for use in traffic studies and traffic simulation calibration. Replacing traditional data collection procedures will result in more accurate models that would allow agencies to explore more alternatives. This larger dataset may also lead to entirely new ways of modeling or understanding the true impacts of changes to a network on a far greater scale. This paper will discuss traditional traffic engineering practices and how the manner in which we do business may change with an expanded use of this technology.

Introduction

Traffic engineering and transportation planning requires data collection, often significant amounts of data collection. Data is currently collected in a variety of ways, some automated, others a slightly modernized version of a clipboard and a stop watch.

While not always obvious, the types of analyses that have become the norm for different applications is driven by what data can be practically and economically collected. Often, our analyses incorporate assumptions to bridge these data gaps caused by data that is not or cannot be collected.

Cell phone tracking has been used for a number of years. Current applications have been largely limited to real time congestion monitoring disseminated through the Internet, radio, or television. Tapping into this pool of data for use in traffic engineering and planning has been limited, to date. This paper discusses past experiences with cell phone tracking data, lessons learned, and the future of this technology and its potential long range impacts on the state of the practice.

Overview of Cell Tracking Technology

Cellular phones broadcast location based data on a regular basis. Wireless carriers need to know where a phone is to properly route communications to and from the device. Most of the time, particularly in urban areas, a phone is within range of multiple towers. A few examples of such towers are shown in Figure 2 and Figure 1.

Figure 2 - Cell Tower Example 1

Figure 1 - Cell Tower Example 2

The more towers a phone is in range of, the better the precision. In urban areas where tower density is high, location accuracy can be accurate within a hundred feet. In more remote areas, the accuracy can be off by up to a mile or more. Once a phone is located, it needs to be mapped to a roadway network. Triangulation is not quite an accurate term for how this is done. Triangulation assumes that you have a bearing from known points. The distance from each point is not important.

Note the towers in the figures above. A tower can tell which array a phone is located within. (Figures 1 and 2 show cell towers with 3 arrays, however other tower can and do contain more.) The tower also knows how far a phone is away from the tower. Knowing the distance and quadrant allows the determination of a location, as shown in Figure 3. Figure 4 shows a potential issue if only one tower is available. The base map that provides the roadway network on which vehicles are mapped is in GIS format.

Figure 3 - Locating a device with two towers

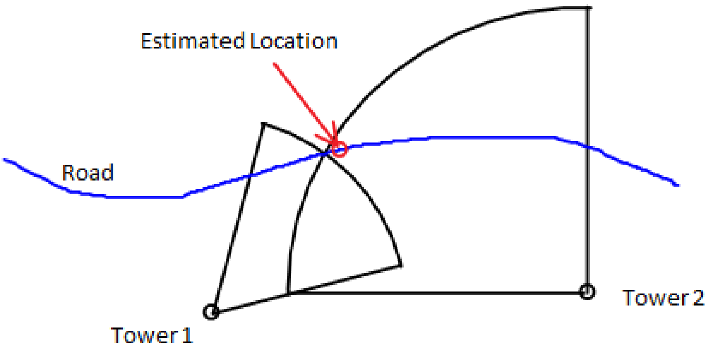
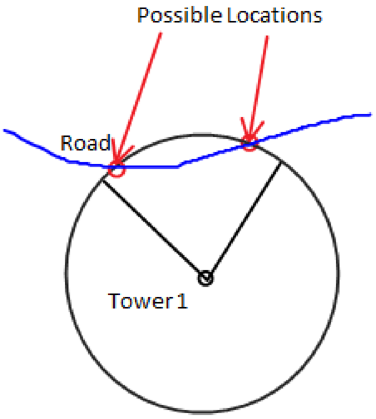


Figure 4 - Location using one tower



AirSage, a company based in Atlanta, Georgia, is the leading company in the industry at this time that is focused strictly on the use of cellular phone location data. Their primary product is travel time data that is either collected, aggregated, and transmitted in a live stream in real-time, or is post-processed months or minutes after the fact. There is a substantial difference in the algorithms that are used for the real-time data aggregation versus the post-processed, or historical, data aggregation. The historical algorithms are much more robust and promise a far greater accuracy. Initial studies performed in Virginia in 2005 and 2007 were focused on the output of the real-time data stream.

Hampton Roads, Virginia – 2005

Algorithms are required to filter out "noise" from phones that aren't moving, are being used by pedestrians, and other miscellaneous activities. At the start of tracking efforts, the data stream was relatively small. Phones would only talk to the tower sporadically. In 2005, the quantity of data points, the quality of data points, and the maturity of the algorithms had not reached beyond a state of infancy. Not surprisingly, the evaluation report developed by the University of Virginia (Smart Travel Laboratory Center for Transportation Studies, 2006) was not kind. It found that the test project in Hampton Roads, Virginia did not perform well. The report found that the system "cannot provide the data of sufficient quality to support operations within VDOT".

Northern Virginia – 2007

Over the course of the 2006, AirSage continued to refine their algorithms for noise filtering. This filtering also includes removing outliers in the data stream that include things like vehicles stopping in mid-corridor, not progressing through an entire segment, and travelling far faster or slower that the majority of traffic. Location accuracy is very dependent on tower density. In some cases, location accuracy issues may cause frontage road traffic to be included in the freeway traffic stream, for example.

In addition to algorithm updates, the data stream from AirSage's only wireless partner at the time continued to increase. Text messaging and smartphone use was increasing, and with those data services, phones were now communicating with towers more often.

A second evaluation study complete by the University of Virginia (Guo, 2008) found that while the data quality had not yet reached the point where it was recommended for use on dynamic message signs, it was indeed capable of providing useful travel time trends on both arterials and freeways and could be used by the DOT to monitor and manage congestion. However, the technology was recommended for use as a supplement to traditional sources but not firm enough to stand on its own quite yet.

A separate study performed by GeoStats (Geostats, LP, 2007) found a systemic bias[1] in AirSage data that could be "controlled through calibration". The technology was beginning to mature[2].

---

[1] AirSage speeds were consistently higher than the field-measured data. The offset was consistent through all congestion and volume levels.

[2] Noise reduction filters were expanded to account for the systemic error found in the study. Further refinement of the post-processing algorithms also continued during this time.

Next Steps

Further validation studies are expected to be completed in late 2010 and early 2011. Since the 2007 studies, the amount of data flowing into the AirSage servers has increased by 500%. AirSage has continued to refine both their real-time and historical algorithms. The segment models upon which the phone locations are mapped has improved thereby also significantly improving the accuracy of the positioning data. It should be expected that this data source will come to fruition in the very near future.

Sample Data Types and their Current Uses

Travel Time Data

Travel time is currently collected in many ways, all of which boil down to two primary methods: extrapolation of spot speeds or probe vehicles. Spot speeds may be collected by loop detector, Side-Fire radar, or other means, and probe vehicles include onboard GPS logging devices, automatic vehicle identification (AVI), Bluetooth (or MAC address) tracking, and the like. Past research has shown that spot speed data quality is dependent on the location, quantity, and spacing of detection. Travel time is extrapolated from the spot speeds but is not measured directly. This extrapolation produces less reliable data than direct measurement of travel time via probe vehicles in high congestion (Jackson, 2007).

Of the probe vehicle technologies, all but cell phone tracking requires the deployment of equipment in the field. Bluetooth tracking is the most similar in concept to cell phone tracking but is hampered by the relatively short range of Bluetooth radios, the pulse rate, and acquisition times for data exchange.

Travel time data is commonly used for calibration of microsimulation models as well as travel demand models. (These models form the basis for corridor studies, transit options, and multi-million dollar design decisions.) In the ITS realm, travel time is used by agencies for monitoring and managing traffic and disseminated to the public through a variety of different channels. (AirSage is a primary travel time information provider for Clear Channel Communications and Google Maps.)

Unless existing equipment, such as loop detectors or AVI technology, is already deployed in the field, travel time data collection is often limited to a peak hour snapshot where a minimum number of data points is required in a given one or two hour window on a random day during a year. Data collection via GPS logging, Bluetooth tracking, or AVI tracking (with non-permanent equipment) can be cost prohibitive, if not from the direct expenses for data collection, then certainly in the cost for post-processing. This is not the case where robust, permanent detection, data storage, and dedicated data processing have been set up for specifically this purpose. Many of the larger agencies in major metropolitan areas have either the basic framework in place, or have a more robust system online. (PeMS in California[i], GCM in the Chicago area[ii], and PORTAL in the Portland, Oregon area[iii] are good examples.) These systems are also limited to freeways or other limited-access facilities.

Travel time data is also gathered regularly by agencies (typically MPOs) on arterial roadways as a manner in which to monitor long term system performance. Again, this is often a one day snapshot meant to represent an average day in a year.

Origin-Destination Data

Travel demand modeling uses a combination of public surveys and land use assumptionsto develop the regional origin-destination information used to formulate travel demand models. Land use assumptions are based on permitted developments and lane use zoning and provide a way to estimate future trips.

Existing trips are based on survey responses, existing land use, assumptions tying existing residential areas with existing commercial areas, and educated guesses to adjust the model volumes to match existing conditions.

Smaller-scale origin-destination (O-D) data is often required for microsimulation purposes, particularly in roadway networks where merging, weaving, and lane utilization is imbalanced and such information is critical to properly calibrating a model. The current method primarily used to collect this data is license plate tracking. Fully automated, partially automated, or completely manual, this is a tedious data gathering technique that quickly becomes expensive. O-D is often *the* major data collection cost on projects where such data is required.

The Future State of Data and the Practice

Over the past five years, a significant amount of work has been put into refining cell phone tracking data for use in measuring travel time. With cell phone data as a reliable source, data collection for straightforward travel time data becomes simple. The cost of deploying GPS probe vehicles in a floating car study for a single days' worth of morning and evening peak hours will be roughly equivalent to a week's worth of continuous data. In all aspects, travel time datacollection*cost* will be vastly reduced allow the same data collection budget to collect *far greaterquantities (vastly increased sample size)* of data for each time period as well as the collection of data over a longer time period. What does this mean?

Implications of increased travel time data availability

Calibration of Microsimulation Models

Travel time is frequently utilized as criterion for calibration of microsimulation models. This data is often limited solely to the peak hours and is based on a handful of runs per direction in a single peak hour. Having a week's worth of peak hours will help mitigate day to day fluctuations in traffic operations. An accident occurring somewhere nearby that increased or decreased traffic volumes in the study area will no longer invalidate all of the data collected.

In the near future, automated count data collection will allow simulation models to include more and more of the shoulder hours leading to and following the peak hours. These shoulder hours can be calibrated as well as the travel time would already be available.

Determining the Capacity of a Roadway

Peak spreading refers to a phenomenon that occurs on heavily congested roadways and networks. Capacity is finite. As demand exceeds capacity, the peak period begins to spread. One peak hour becomes two, then three, then four. Agencies have begun to undertake projects to study the impacts and implications of peak spreading and to try to estimate the extent to which the peak hours will spread over time. Combined with robust count data, such as that provided by a roadway network will full detection, cell phone tracking data would allow measurements of capacity under a variety of different conditions and during different seasons. (This would be useful particularly in areas where the time year makes substantial differences in driver population due to large influxes of tourists, for example.) Capacities can be more closely determined for different segments, times of day, and days of the week.

Before and After Studies

Before and after studies often focus on the operations of a single facility that was altered or improved. While such a study may show the impacts of a change on a target roadway, they do not show the big picture of the impact on the entire network. For example, imagine a congested urban arterial is the subject of an improvement and a before and after study. If substantial latent demand existed prior to improvements, it is likely that the improvements measured, be it travel time, speed, or delay, may not show much in the way of a substantial improvement over time. But what actually happens?

The initial improvement may have increased capacity and reduced delays and travel times. Latent demand may have flowed on to the improved arterial causing operations to return to what appear to be the pre-project state. One could measure volumes to determine if flow rates have increased allowing one to say the project increased capacity by x percent.

With the ability to collect more data in a cost efficient manner, it would be possible to measure the before and after impacts on the surrounding roadway network as well. A traditional study may miss a dramatic improvement in the operations of a parallel route that suddenly experiences less demand as the subject facility is able to absorb the pre-existing latent demand.

Smart Work Zones

Most current smart work zone technologies are based on relaying travel time information in real time to the travelling public. While FHWA requires work zones of significant size to have some “smart” technology applied for monitoring, such monitoring is not applied as it is can be cost prohibitive. Shifting traffic control plans also present a hurdle to temporary detector placement and providing the communication between field equipment and an offsite server. With cell phone tracking, no equipment needs to be deployed and communications with an offsite server are not an issue. Due to its low cost, this may vastly increase the number of work zones nationwide that provide real-time monitoring.

Travel Time Reliability

As traffic volumes continue to increase and outstrip our capabilities to increase capacity, travel time reliability is becoming a more important metric for the quality of operations on a facility. Cell phone tracking offers the first cost effective option for long term or more frequent data collection along arterial roadways as well as freeways.

Origin-Destination Data

The concept of latent demand currently relies quite heavily on a travel demand model, which itself is dependent on O-D data. As mentioned previously, the state of the practice relies on land use data, educated guesses, or on surveys of the travelling public which are costly to create and process and often yield a poor response rate. This technology will allow the **directcollection** of accurate origin/destination data and the indirect measurement of latent demand.With this new data in hand, travel demand models will incorporate both more accurate base data and fewer far-reaching assumptions.

Microsimulation models require vehicular inputs and routing of some kind as base data for the model. Static routing assignments are often used in existing conditions models as O-D data is too costly to acquire. Static routing assignments do not allow vehicles to change their path, for example, to a nearby parallel, uncongested roadway, to avoid massive delays on their static path. This is not the way drivers in the real world behave and it should not be the way we code our models. Most, if not all, high-end microsimulation models offer a variation of dynamic assignment. This allows the model to iteratively reach equilibrium where not all vehicles that go from point A to point B travel along the same path. Incorporating real world O-D data would allow for more realistic existing conditions models that should yield more accurate future conditions models.

On a smaller scale yet, this technology would also allow a real world select link analysis. For example, we could determine the starting and ending point for all of the vehicles that cross a bridge during a specific time period, such as the evening rush hour.

Conclusion

Cell phone tracking technology has been maturing over the past five years and will soon be reaching a point where its uses can extend into the realms of traffic engineering and transportation planning. Our analyses are often limited by the data that we are capable of collecting. As this technology matures, we, as an industry, will be challenged with coming up with new ways to apply the vast quantities of data that will soon be available. Data collection will continue to decline in cost, increase in quality, and move in directions we cannot yet conceive. In terms of travel time data alone, cost effective access to such quantities of data will allow reductions in the costs of doing what we already do, and allowing us to do more than we ever have as engineers.

Authors' Information

Smith Siromaskul
Professional Associate
HDR Engineering
1001 SW 5th Avenue, Suite 1800
Portland, OR 97204
503-449-1524
smith.siromaskul@hdrinc.com

Steve Speth
Professional Associate
HDR Engineering
24954 Hwy 395 South
Canyon City, OR 97820
541-575-0434
steve.speth@hdrinc.com

Bibliography

Geostats, LP. (2007). *AirSage Data Quality Assessment: A Comparison to GPS probe vehicle data on I-95 in the Washington DC Area.*

Guo, B. S. (2008). *Evaluation of an Early Generation Travel Time Data Service - Northern Virginia.* University of Virginia.

Jackson, J. (2007). A Comparative Analysis of Side-Fire Radar and AVI Technology for Monitoring Work Zone Travel Times. *ITE District 6 Annual Meeting.* Portland, OR.

Smart Travel Laboratory Center for Transportation Studies. (2006). *Evaluation Final Report.* University of Virginia.

---

[i] https://pems.eecs.berkeley.edu/?redirect=%2F%3Fdnode%3DState

[ii] http://www.gcmtravel.com/gcm/home.jsp

[iii] http://portal.its.pdx.edu/