

Анализ транспортных потоков дорожной сети мегаполиса по данным оператора сотовой связи

А.Б. Дерендяев, В.Г. Гитис

Институт проблем передачи информации им. А.А. Харкевича, Российская академия наук, Москва, Россия

Поступила в редколлегию 30.05.2013

Аннотация—Рассматриваются основные положения метода анализа транспортных потоков автомобильной сети мегаполиса по данным оператора сотовой связи. Предложена модель, позволяющая приближенно вычислять скорости движения по участкам дорожной сети мегаполиса и оценивать загруженность этих участков. Метод реализован на платформе ГИС ГеоТайм 3. Обсуждаются экспериментальные результаты для дорожной сети г. Москвы и Московской области.

Ключевые слова: станции сотовой связи, транспортные потоки, геоинформационная система ГеоТайм 3

1. ВВЕДЕНИЕ

В настоящее время практически каждый человек обладает мобильным сотовым телефоном. Оператор сотовой связи фиксирует базовые станции (БС), в области действия которых находится телефон. Опираясь на последовательность смены БС, можно, с некоторой степенью достоверности, определить маршрут пользователя с телефоном.

В работах [1] и [2] предложены методы анализа дорожных пробок для отдельного шоссе, вдоль которого расположены станции сотовой связи. В статьях [3] и [4] рассматриваются методы анализа скоростей движения автотранспорта по участкам пересечения нескольких магистральных трасс.

В данной работе предлагается модель, с помощью которой на основании данных оператора сотовой связи можно приближенно вычислять скорости движения по участкам дорожной сети мегаполиса и оценивать загруженность этих участков. Снижение скорости движения ниже определенного уровня на участках транспортной сети может указывать на транспортные заторы и пробки. Метод экспериментально исследовался на данных одного из операторов сотовой связи для дорожной сети г. Москвы и Московской области.

2. ИСХОДНЫЕ ДАННЫЕ

Для решения задачи имеются два типа данных: выписка из базы данных CDR (Call Data Record) оператора сотовой связи и географически локализованный граф автомобильных дорог.

В базе данных CDR хранятся события, связанные с привязкой мобильных телефонов к базовым станциям, рис. 1. Выписка из базы данных оператора CDR включает в себя следующие поля:

1. Момент события.
2. ID телефона – хеш от номера телефона.
3. CID - Номер БС в подсети.
4. LAC – номер подсети. Любая БС идентифицируется по паре CID/LAC.
5. Тип события.
6. Географические координаты БС.
7. Тип БС (OUTDOOR, INDOOR, METRO).
8. Углы направленности антенны БС.

Пример записи: 2012-07-02

00:00:00.006;5068feb0;1612;50031;1;37.56268044;55.39734577;OUTDOOR;210;178.25;

Исследуемая в работе база данных CDR для Москвы и Московской области содержит около 230 млн. событий за одни сутки. В ней фиксируется от 44 тысяч событий в минуту (ночью) до 311 тысяч (вечером).

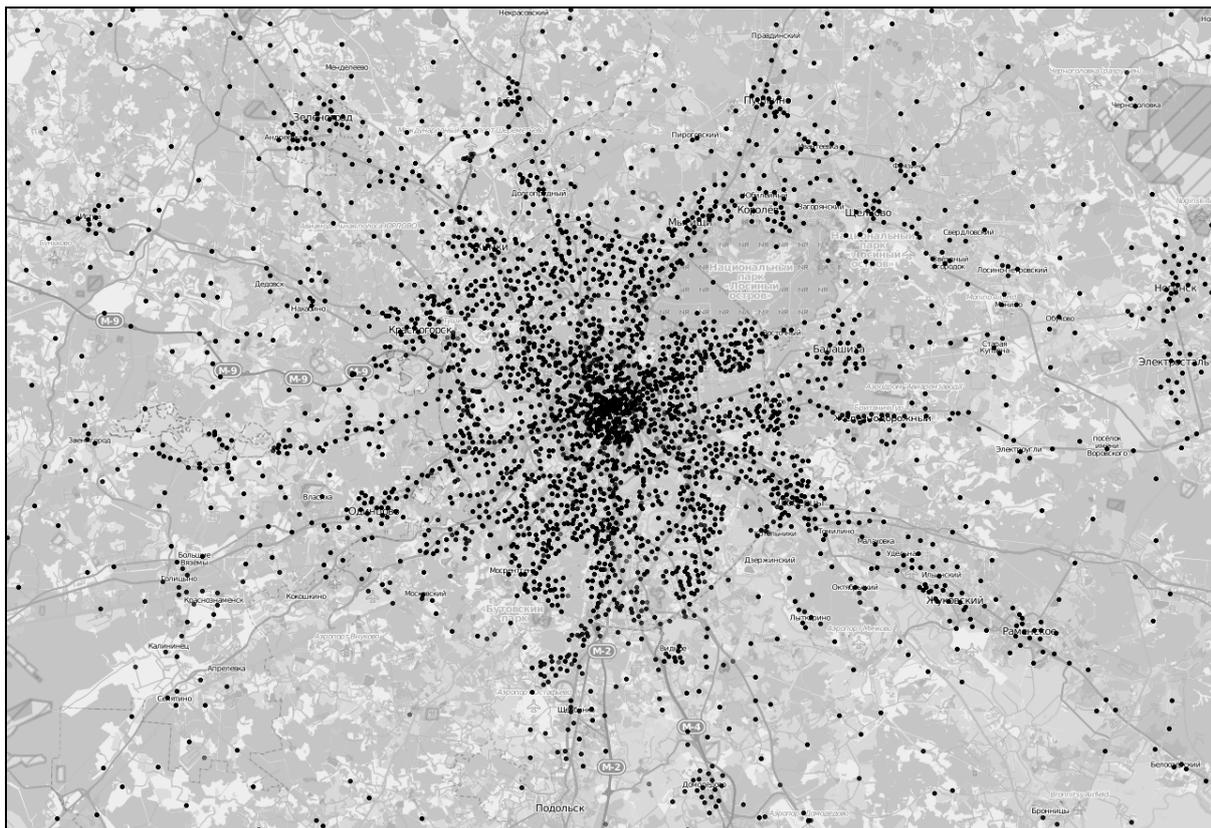


Рис.1. Распределение БС оператора сотовой связи на территории Москвы и Московской области.

В базе данных CDR существует множество различных типов событий. Для упрощения в данной работе рассматриваются только события LocationUpdate, которые происходят при смене соты, включении-выключении телефона или по таймауту. Доля таких событий в CDR составляет 69%. Кроме того, в выписке из CDR убраны все

записи, касающиеся БС в метро. Поэтому маршруты телефонов через метро в работе не рассматриваются. В дальнейшем, если не сказано иного, мы будем предполагать, что маршруты телефонов располагаются только вдоль автомобильных дорог.

Локализованный граф дорог получен с сайта OpenStreetMap (OSM), рис. 2. Вершинами графа являются пересечения и концы дорог.

Данные OSM состоят из двух shape-файлов (для Москвы и Московской области), в которых одна и та же дорога может быть разрезана на несколько участков. Атрибутивная информация файлов имеет вид:

1. ID на сайте OSM.
2. Направление движения (одностороннее или двустороннее).
3. Тип дороги (пешеходная, сервисный проезд, дорога 1-го, 2-го и 3-го класса, магистраль и др.)
4. Максимальная скорость.
5. Число полос.
6. Некоторые другие атрибуты.



Рис. 2. Граф дорог из OSM. Толщина линии соответствует классу дороги.

3. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ

Предварительная обработка данных включает в себя два этапа.

На первом этапе предварительной обработке подвергается граф автомобильных дорог: участки дорог Москвы и Московской области «сшиваются» с учетом ID по

данным OSM, и из всего массива полученных дорог убираются лишние (в том числе пешеходные тропы, сервисные подъезды и внутридомовые дороги). Поскольку в OSM дорога с одним ID может иметь несколько пересечений с другими дорогами (в том числе и не в конечных точках), то требуется выполнить топологическую регуляризацию данного графа, т.е. требуется разделить такие дороги на независимые участки. Кроме того, дороги с двухсторонним движением надо разделить на пары участков с односторонним движением. Каждому такому участку присваивается новый ID.

На втором этапе выполняется привязка БС к участкам дорог.

Модель привязки предполагает, что принадлежность участков дорог к БС определяется положением БС и её диаграммой направленности, рис. 3. При этом к каждой БС относятся только ближайшие к ней участки дорог, попадающие в сектор её диаграммы направленности.

Рассмотрим привязку участков дорог более подробно. Разделим всю территорию диаграммой Вороного с центрами в БС. Будем считать, что данная БС относится к участку дороги, если этот участок дороги или его часть покрываются диаграммой направленности БС и попадает в соответствующую ячейку диаграммы Вороного. В результате одна БС может быть привязана к нескольким участкам дорог и несколько БС могут быть привязаны к одному и тому же участку дороги.

Более реальной привязкой БС к участкам дорог являлась бы карта распределения приема БС, полученная экспериментальным путем. Такая карта позволяет учесть все возможные поправки (на рельеф, переотражения, некорректно проставленные углы и прочее).



Рис. 3. Фрагмент расположения базовых станций оператора сотовой связи для участка дорог Подмосковья. Секторы станций соответствуют диаграммам направленности излучения антенн.

4. ИДЕЯ АЛГОРИТМА

Определим трек как последовательность событий, состоящих в смене БС для одного абонента. Будем предполагать, что автомобилист между двумя точками дороги всегда выбирает кратчайший (в географическом смысле) маршрут. Пара последовательных событий для одного абонента фиксирует моменты переключения определенных БС. Предполагается, что за этот временной интервал абонент перемещается вдоль дороги по кратчайшему пути между точками, расположенными на соответствующих участках по середине секторов диаграмм направленности БС. Скорость абонента вычисляется как отношение кратчайшего пути к временному интервалу. Это значение скорости вносится в статистики всех участков дороги, относящихся к кратчайшему пути абонента. Загруженность участка дороги для некоторого интервала наблюдений вычисляется как отношение числа абонентов, скорости которых внесены в статистику данного участка, к интервалу времени наблюдений.

Параметрами алгоритма анализа данных являются время начала и величина интервала наблюдений. Алгоритм строит треки перемещения всех сотовых телефонов, выбирает кратчайшие пути перемещения абонентов сотовой связи по связным участкам графа дорог, оценивает скорости абонентов, добавляет скорости в статистику участков дорог и оценивает загруженность участков.

Опишем алгоритм более подробно:

1. Для каждого поступающего события (по порядку из БД CDR), определяется наличие трека для данного абонента. Если трека нет, данное событие становится началом нового трека и повторяется п. 1 для следующего события.

2. Если трек есть и последняя пара событий из него удовлетворяет некоторым условиям (в частности, время между событиями меньше получаса, расстояние между БС меньше 20 км), то тогда жадным алгоритмом Дейкстры по графу дорог (вес ребер равен географической длине дорог) строится минимальный путь Φ между всеми дорогами, привязанными к первой БС и всеми дорогами, привязанными к второй БС (с учетом направления движения).

3. Вычисляется длина минимального пути Φ , и по разнице времен между событиями оценивается скорость прохождения маршрута Φ . Эта скорость записывается в статистику для всех участков дорог маршрута Φ .

4. Вернуться к п. 1 для анализа следующего события.

В результате получаем статистику скоростей абонентов для каждого участка дороги. По данной статистике для каждого участка дороги строится гистограмма с шагом в 10 км/ч и выбирается мода (минимальная, если их несколько), которая сохраняется в качестве атрибута данного участка дороги.

5. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Данный алгоритм реализован в виде набора плагинов для ГИС ГеоТайм 3 [5, 6]. Система ГеоТайм 3 позволяет пользователю интерактивно измерять гистограммы скоростей участков дорог, анализировать распределения скоростей для разных интервалов наблюдения, изучать треки прохождения участков дорог любым конкретным абонентом.

Для анализа результатов желательно иметь данные о реальных значениях скоростей на участках дорог. К сожалению, этими данными мы не располагаем. Поэтому для анализа используются качественные методы сравнения.

Рассмотрим гистограммы скоростей абонентов в направлении с востока на запад для Кремлевской набережной (рис. 4). Гистограммы скоростей показаны на рис. 5. Они относятся к 30-минутным интервалам наблюдений: 3:00 – 3:30, 7:00 – 7:30, 15:00 – 15:30, 19:00 – 19:30, 23:00 – 23:30. На рисунке можно видеть, что число абонентов и моды распределений скоростей абонентов (т.е. характерные скорости) сильно зависят от времени суток. В ночные часы от 3:00 до 3:30 загруженность дороги очень мала. Загруженность резко возрастает утром от 7:00 до 7:30. Типичная скорость устанавливается около 40 км/час. Дневное и вечернее время от 15:00 до 15:30 и от 19:00 до 19:30 показывают дальнейшее увеличение автотранспортного потока, которое приводит к затору движения со скоростью 10-20 км/час. Последний временной интервал от 23:00 до 23:30 показывает меньшую загруженность дороги и совпадает по форме с гистограммой для утренних часов от 7:00 до 7:30.

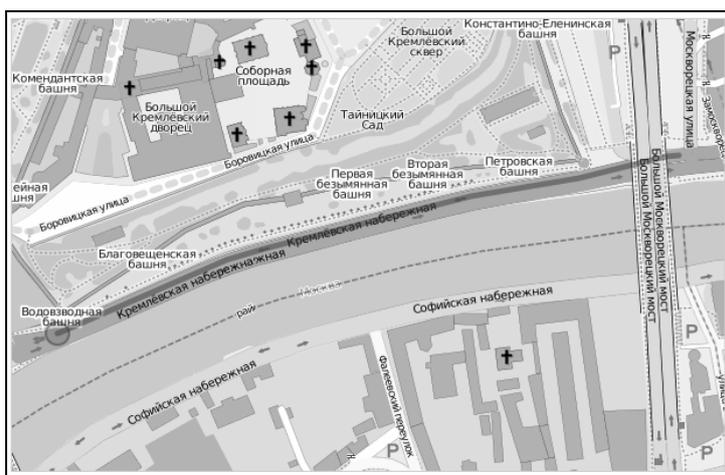


Рис. 4. Анализируемый участок Кремлевской набережной (жирная синяя линия).

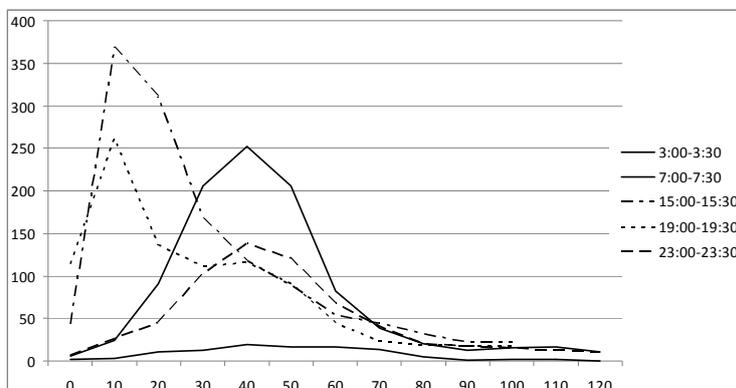


Рис. 5. Гистограммы скоростей абонентов (км/ч) на участке Кремлевской набережной в 30-минутные интервалы времени наблюдения: 3:00 – 3:30, 7:00 – 7:30, 15:00 – 15:30, 19:00 – 19:30, 23:00 – 23:30.

На рис. 6 на картах Москвы и Московской области показаны моды распределений скоростей абонентов на участках дорог для двух интервалов наблюдения. Можно видеть, что ночью от 3:00 до 3:30 (рис. 6А) на основных магистралях достаточно свободно и скорости поддерживаются высокими. На рис. 6Б видно, что вечером от 19:00 до 19:30 значения мод скоростей на ряде участков существенно уменьшаются. Эти участки соответствуют заторам движения (пробкам). Такое распределение пробок согласуется с реальностью.

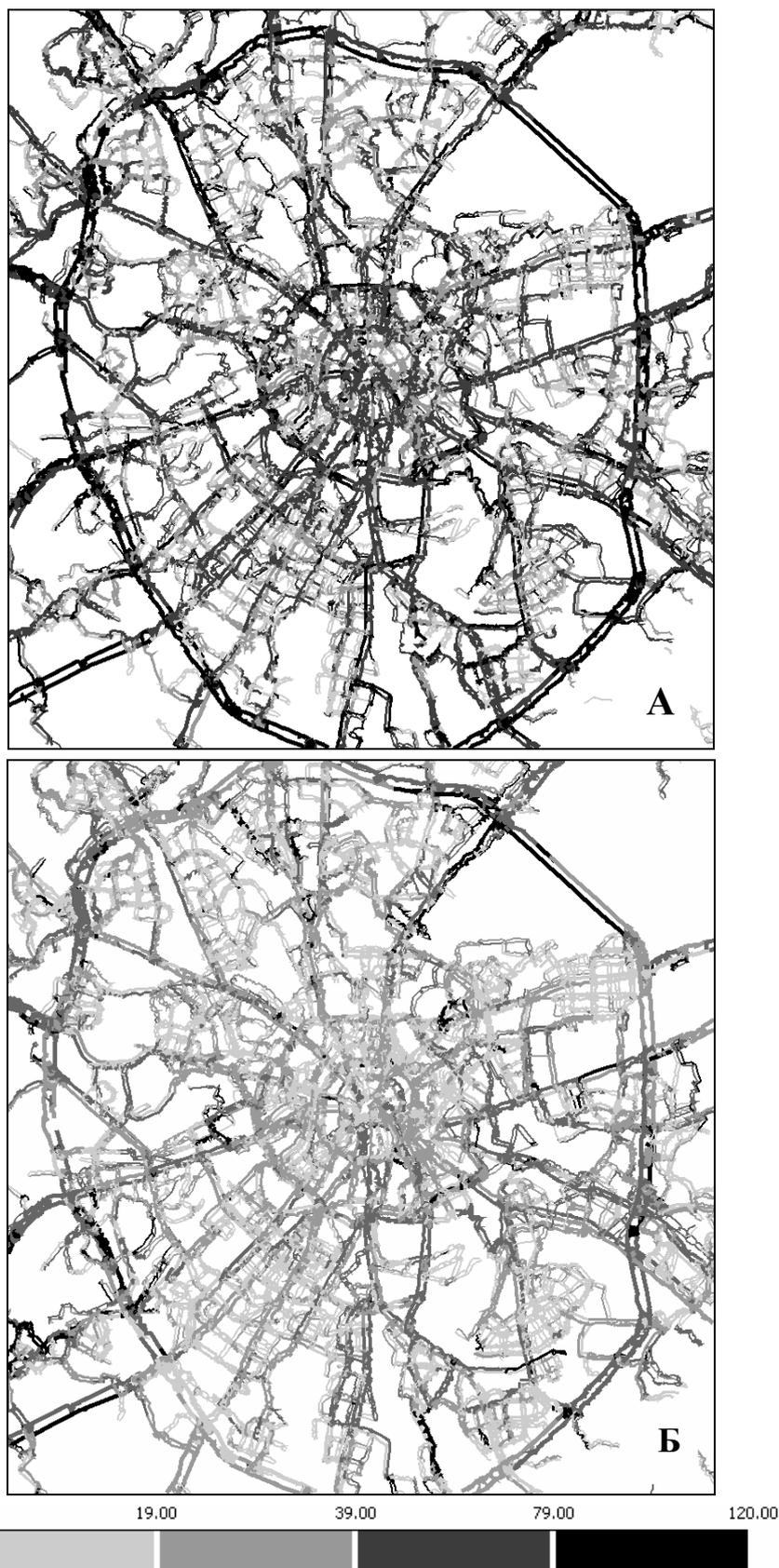


Рис. 6. Карта транспортных потоков в разные интервалы времени наблюдений: А) 3:00 – 3:30, Б) 19:00 – 19:30, В) легенда значений скоростей, км/час.

Для неосновных дорог, в первую очередь – в спальнях кварталах, алгоритм дает заниженные скорости. Более подробное исследование маршрутов на этих дорогах показывает, что заниженные скорости создают пешеходы, идущие к метро или к остановкам общественного транспорта (ОТ). Эти скорости можно было бы отсекаать по некому порогу. Но, из-за того, что пешеходы, как правило, идут не вдоль автомобильных дорог и из-за недостаточной точности определения их координат, скорости пешеходов попадают в диапазон от 0 до 15 км/ч. Эта скорость часто превышает скорость транспорта в пробках. В дальнейшем для отделения пешеходов от автотранспорта планируется использовать данные маршрутов ОТ, в том числе метро.

Детальное изучение маршрутов по всему региону показывает, что предложенный алгоритм привязки станций к дорогам в большинстве случаев работает адекватно. Исключение составляют нескольких случаев, обусловленные особенностями рельефа местности. В дальнейшем, для более точного анализа обстановки, необходимо учитывать карту высот при построении привязки дорог к БС.

6. ДАЛЬНЕЙШИЕ ИССЛЕДОВАНИЯ

1. Дополнение исходных данных картой маршрутов ОТ позволит определять абонентов, не пользующихся автомобилем.

2. Дополнение исходных данных картой высот позволит строить более точные привязки участков дорог к БС.

3. Анализ треков пользователей показывает, что использование только двух станций для построения маршрута не всегда оправдано и часто уводит с прямого и логичного маршрута на маленькие боковые дороги. При построении маршрутов использование более двух событий позволит убрать лишние съезды и более точно оценивать маршрут абонента.

4. Оценивание веса ребер графа дорог не по их географической длине, а по времени их прохождения (с учетом уже вычисленных скоростей) позволит учесть информированность водителей об уже существующих пробках.

5. Оценка скорости на участке дороги не по моде распределения, а с помощью других статистик (например, 70% квантиль) может дать более адекватную оценку движения.

6. Включение в анализ других типов событий, таких как смена соты при передаче данных или во время разговора, позволит увеличить количество статистических данных, доступных для обработки.

7. ВЫВОДЫ

В данной работе предложен новый метод анализа автомобильной сети мегаполиса по данным оператора сотовой связи. Показано, что предложенный метод позволяет строить динамические карты скоростей прохождения автотранспортом разных участков дорог в разные моменты времени. Дальнейшее исследование ограничивается недостатком данных, в том числе отсутствием реального графа пробок, с которым можно сравнивать полученные результаты.

Авторы благодарны А.П. Кулешову и Д.Н. Мацневу за ряд важных замечаний и предложений, высказанных при обсуждении работы.

Исследования частично поддержаны грантом 13-07-00224 РФФИ.

СПИСОК ЛИТЕРАТУРЫ

- [1] N. Caceres, J.P. Wideberg and F.G. Benitez, *Deriving origin–destination data from a mobile phone network*, IET Intell. Transp. Syst., 2007, 1, (1), pp. 15–26
- [2] A. Janecek, D. Valerio, K. Hummel, F. Ricciato, H. Hlavacs, *Cellular Data Meet Vehicular Traffic Theory: Location Area Updates and Cell Transitions for Travel Time Estimation*. UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA
- [3] H. Bar-Gera, *Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel*, Transportation Research Part C 15 (2007) 380–391
- [4] M. Alger, *Real-time traffic monitoring using mobile phone data*, Vodafone, available at <http://www.maths-in-industry.org/miis/30/1/TrafficMonitoring.pdf>
- [5] Gitis V., Derendyaev A., Metrikov P., Shogin A. 2012. *Network geoinformation technology for seismic hazard research*, Natural Hazards: Volume 62, Issue 3, pp. 1021-1036
- [6] Gitis, V., G., Derendyaev A., B. Geoinformation technology for spatio-temporal processes research // Proceedings of the ISPRS workshop on Dynamic and Multi-dimensional GIS, Shanghai, China, October, 2011. pp. 111-115.

Analysis of traffic flow for megalopolis road network by the data of mobile operator

A. Derendyaev, V.Gitis

Institute for Information Transmission Problems, Russian Academy of Sciences (Kharkevich Institute),

Abstract. The basic concepts of the method traffic flow for megalopolis road network by the data of mobile operator of road megalopolis network traffic flow analysis according to mobile operator database are considered. The model, which allows one to estimate the traffic speed and workload, is proposed. The method is implemented on the GIS GeoTime 3 platform. The experimental results for the Moscow and Moscow region road network are discussed.

Keywords: base transceiver station, traffic flows, geographic information system GeoTime 3