Определение местоположения без GPS: как устроен Яндекс.Локатор

Сейчас всё больше мобильных приложений становятся геозависимыми. Одни просто не имеют смысла без знаний о местоположении пользователя, другие становятся с ним удобнее. Это так называемые Location Based Services (LBS): навигаторы, форскверы, инстаграмы с геотегами фотографий и даже приложения-напоминалки, которые срабатывают около конкретного места, например, рядом с офисом или магазином.  
  
Для сервисов и приложений Яндекса разработчики создали собственную реализацию метода определения местоположения без GPS — [Яндекс.Локатор](http://api.yandex.ru/locator/). Он экономит время пользователя и делает их приложения чуточку умнее. В Навигаторе и Картах она избавляет от ввода начальной точки маршрута, даже если вы на крытой парковке. А при выборе фильма в Киноафише или товара в мобильном Маркете помогает сразу показать, где их найти именно в вашем районе города. Ну и, разумеется, при поиске кафе и банкоматов — позволяет показывать вам сразу ближайшие, даже когда вы в метро.

#### Почему без GPS и как иначе

Спутниковые системы навигации ([GNSS](http://en.wikipedia.org/wiki/GNSS)), в нашем случае это GPS и ГЛОНАСС, — самый точный на сегодняшний день метод геоопределения. Соответствующие модули есть практически во всех современных смартфонах. Но не всегда и не везде он может решить задачи LBS.  
  
Во-первых, поиск спутников иногда занимает несколько минут, а бывают ситуации, в которых скорость определения важна даже с потерей точности. Например, когда нужно построить предварительный маршрут в навигаторе или [зачекиниться](http://ru.wikipedia.org/wiki/%D0%93%D0%B5%D0%BE%D1%81%D0%BE%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5_%D1%81%D0%B5%D1%82%D0%B8" \l ".D0.9E.D1.81.D0.BD.D0.BE.D0.B2.D0.BD.D1.8B.D0.B5_.D0.BF.D1.80.D0.B8.D0.BD.D1.86.D0.B8.D0.BF.D1.8B_.D0.B8.D1.81.D0.BF.D0.BE.D0.BB.D1.8C.D0.B7.D0.BE.D0.B2.D0.B0.D0.BD.D0.B8.D1.8F_.D0.B3.D0.B5.D0.BE.D1.81.D0.BE.D1.86.D0.B8.D0.B0.D0.BB.D1.8C.D0.BD.D0.BE.D0.B9_.D1.81.D0.B5.D1.82.D0.B8). Во-вторых, спутники обычно не «видны» в помещениях или под землёй. В-третьих, GPS-модули есть не в каждом мобильном телефоне или планшете, и их почти нет в ноутбуках. То есть для LBS нужны альтернативы.  
  
И альтернативы, конечно, есть — определять местоположение можно по ближайшим GSM-вышкам, сетям Wi-Fi и даже по IP-адресу. Точность определения у каждого из этих способов гораздо хуже, чем у GPS. Но если их скомбинировать, они вместе дадут приемлемое качество. При этом какие-то недостатки одного нейтрализуются возможностями другого. GSM-вышки есть практически везде, а Wi-Fi сети — нет. При этом по Wi-Fi точность определения лучше. Поэтому комбинированный способ по полноте и точности лучше, чем каждый в отдельности. Менее известен факт, что у двух роутеров в разных частях города может оказаться одинаковый MAC-адрес. Совмещение GSM и Wi-Fi решает такие коллизии. У этих роутеров, скорее всего, рядом будут находиться вышки с разными идентификаторами — ведь вероятность совпадения в пределах квартала гораздо меньше, чем в масштабах всего города.  
  
В мире есть несколько реализаций такого комбинированного способа геоопределения. И кажется, первый вопрос, с которым сталкивались все разработчики, — где же взять информацию о местоположении сетей Wi-Fi и сотовых вышек?

#### База местоположений сетей

В дилемме «купить или создать» Яндекс в конечном счёте предпочли второе. Основная причина — что с собственными данными и алгоритмами гораздо легче контролировать качество результата. В сборе информации им помогли пользователи мобильных Яндекс.Карт.  
  
Когда разработчики начинали разрабатывать Локатор, на улицах городов были уже сотни тысяч людей с включёнными в телефонах Яндекс.Картами. С согласия пользователя приложение постоянно передаёт его GPS-координаты — на основе этой информации строятся Яндекс.Пробки. Они подумали, что вместе с этим приложение может отмечать, какой базовой станцией обслуживается телефон в этих координатах, какие видны сети Wi-Fi (при этом, конечно, к самим сетям не подключаясь — чтобы не создавать privacy-рисков).  
  
Человеку для участия в таком краудсорсинге ничего специально делать не нужно — просто пользоваться приложением. Как и о координатах, данные об окружающих Wi-Fi сетях и станциях GSM обезличены. Они практически ничего не «весят», и батарейка от их передачи, соответственно, быстрее не садится.  
  
Таким образом, пользователи стали помогать друг другу:

Одни, с GPS-приёмником в телефоне, узнают точное расположение сетей и передают информацию в Яндекс. Другие, у которых GPS-модулей нет, присылают список сетей, которые видят в данный момент, и получают в ответ своё примерное местоположение на карте.  
  
База собрана и регулярно обновляется. И тут мы сталкиваемся со следующей проблемой.

#### «Переезд» сетей

Опыт показывает, что идентификаторы сотовых вышек постоянно меняются — номер, который вчера был в центре города, завтра может оказаться на окраине. Переезжать могут и Wi-Fi-роутеры — вместе со своими владельцами. И получается, что с каждым переездом нужно инвалидировать заметную часть данных.  
  
Вот как удалось решить одновременно проблемы с переездом и вышек, и роутеров. От пользователя поступает запрос на определение местоположения вместе с данными о том, какие сети он видит. Если в списке сетей есть та, что была замечена в разных частях города, алгоритм учитывает, сколько сигналов от неё накоплено в каждом районе и возраст последнего. Каждое плотное скопление сигналов от Wi-Fi сети или сотовой вышки называется «облаком». Чем больше сигналов в облаке и чем они свежее, тем больше оно заслуживает доверия. Ответом будет, соответственно, самое большое и свежее. А облако, в котором нет сигналов больше месяца, считается устаревшим — даже если для этой сети не появилось более свежего облака в другом районе.

#### Радиус облака

Поскольку положение определяется примерно, нельзя показать точку — нужно нарисовать круг (ведь радиосигнал в отсутствие помех распределяется во все стороны равномерно). Хотя, если посмотреть на фактическую картину сигналов, чаще всего это эллипс. Ведь больше всего пользуются мобильными Картами автомобилисты. Их GPS-следы остаются на дорогах, а из дворов и, тем более, из зданий сигналов практически не поступает.

Чтобы ответ был предельно точным, радиус круга должен быть минимальным. Если просто обвести окружность вокруг всех точек сигналов конкретной сети, радиус получится слишком большим. Уменьшить его помогла мат. статистика. Плотность сигналов подвержена нормальному распределению, то есть применимо [правило трёх сигм](http://ru.wikipedia.org/wiki/%D1%F0%E5%E4%ED%E5%EA%E2%E0%E4%F0%E0%F2%E8%F7%E5%F1%EA%EE%E5_%EE%F2%EA%EB%EE%ED%E5%ED%E8%E5#.D0.9F.D1.80.D0.B0.D0.B2.D0.B8.D0.BB.D0.BE_.D1.82.D1.80.D1.91.D1.85_.D1.81.D0.B8.D0.B3.D0.BC). В окрестность такого радиуса попадает 99,7% точек.

#### Улучшение точности определения…

##### **…по GSM-сетям**

Когда-то приложениям была доступна информация лишь об одной базовой станции, хоть телефон видит чаще всего несколько. После появления платформы Android приложения смогли научиться [видеть их все](http://developer.android.com/reference/android/telephony/TelephonyManager.html#getNeighboringCellInfo%28%29) (кроме подключения в стандарте 3G, который позволяет узнать только одну сотовую вышку). Местоположение стало определяться точнее — уже не по одному облаку, а по совокупности нескольких. Оказалось, что для множества облаков можно использовать тот же подход, что и для одного. Радиус считается по среднеквадратичному отклонению сигналов, входящих в совокупность облаков, а центр вычисляется по среднему их координат.

##### **…по Wi-Fi-сетям**

Когда смартфон находится в радиусе действия нескольких Wi-Fi-сетей, он может сообщить не только их список, но и мощность сигнала каждой. Знание об этой мощности мы и использовали для уточнения центра окружности, в которой находится пользователь. К центрам наблюдаемых облаков мы начали подвешивать воображаемые пружинки — тем туже, чем сильнее сигнал. А их свободные концы — соединять. Точка, в которой эти пружинки уравновешиваются, и есть уточнённый центр.

#### Получившееся качество

Сначала несколько слов о том, как мы оцениваем качество нашего решения. Как уже говорилось, от пользователей, у которых есть в устройствах GPS-модуль, Локатор получает и координаты, и список сетей, которые видят устройства. Для оценки качества он сначала определяет примерное местоположение, ориентируясь только на эти сети. А затем проверяет, попали ли истинные координаты от пользователя в предположенную Локатором окружность.

Используя эту методику, мы получили следующие цифры:

* для 83% запросов в сутки местоположение определено правильно — GPS-координаты устройства попали в область, названную Локатором
* 14% сигналов — с ошибкой:
  + 7% — ошибка меньше 100 метров
  + 5,6% — от 100 метров до нескольких километров
  + 1,4% — Локатор ошибается городом
* оставшиеся 3% запросов получают ответ «Местоположение не найдено»

Можно ли добиться лучшего качества? Да. Преимущество метода в том, что при определённой зрелости алгоритмов достаточно лишь собирать больше данных, чтобы определять местоположение точнее. А это достаточно легко, потому что растёт и количество Wi-Fi сетей, и количество пользователей наших приложений.  
  
Но есть технологические пределы:

* если телефон сообщает только об одной GSM-вышке — минимальный радиус составит несколько сотен метров в городе, и несколько километров за городом
* если телефон видит несколько вышек — центр можно определить точнее, но радиус уменьшить вряд ли получится
* если видна Wi-Fi сеть — минимальный радиус будет 10 метров