

A Technical Review of Cisco's Wi-Fi-Based Location Analytics

Authors:

Dr. James Little and Dr. Brendan O'Brien

White Paper

July 2013

Overview

The rapid increase of mobile devices is making Wi-Fi the preferred method of network access, predominantly within buildings, but also outside. This trend presents businesses with both tremendous potential and unique challenges. Location analytics, in conjunction with the Cisco® Mobility Services Engine (MSE), enables businesses to realize unprecedented benefits from location-based services:

- Location analytics: Estimates the number of visitors, the amount of time they spend, and the frequency of their visits within the site
- Advanced analytics: Provides knowledge of movement patterns by these visitors while they are in the building

Together, these analytics provide detailed insights into general behavior patterns of people moving and interacting within a venue or open space. All that is required is Cisco's wireless infrastructure and Wi-Fi being enabled on the visitors' smartphone or tablet devices.

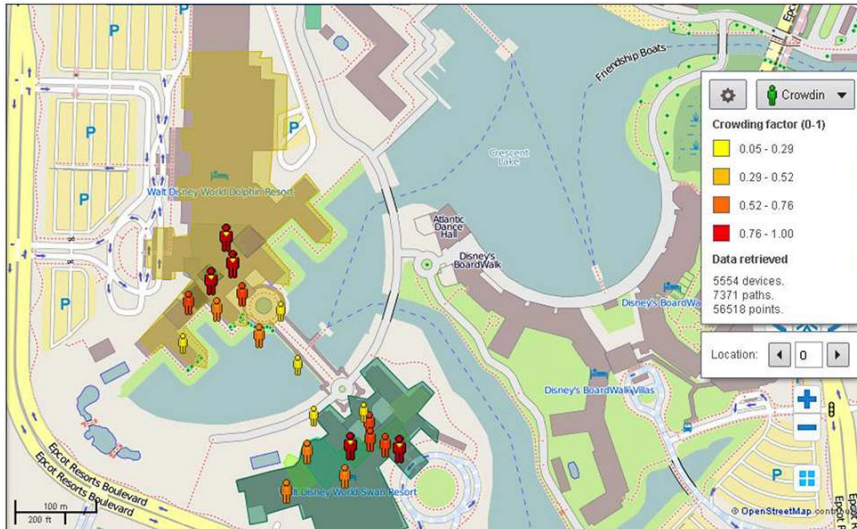
Location analytics uses advanced data mining techniques tuned to the type of data Wi-Fi location provides. This data is based on potentially millions of position triangulations made per day by Cisco's MSE.

This white paper discusses the technology and capabilities behind Cisco's location analytics.

Introduction

The data from the Cisco MSE calculates discrete time, location, and MAC address of devices detected within the coverage area of the Cisco access points in the network. A device need not be associated to be identified and have its location estimated. This data is what allows us to estimate behaviors throughout the different parts of the building. The data can be assembled into a set of chronological points per device, or paths, on which movement analytics can also be applied. However, with the potential huge quantities of data, to get valuable insights requires innovative application of data mining techniques and data filtering.

The venues where this technology is being first adopted have a range of functions: retail stores, hotels, conference facilities, shopping malls, schools, and even city centers - anywhere that there is Wi-Fi coverage. Yet location analytics is flexible enough to model the characteristics of any of these venues and measure many parameters within it, such as flow, speed, dwell, and penetration. The results related to the devices located over time have value not only to the marketing and retail function of the venue, but also for operational efficiency and improved security.



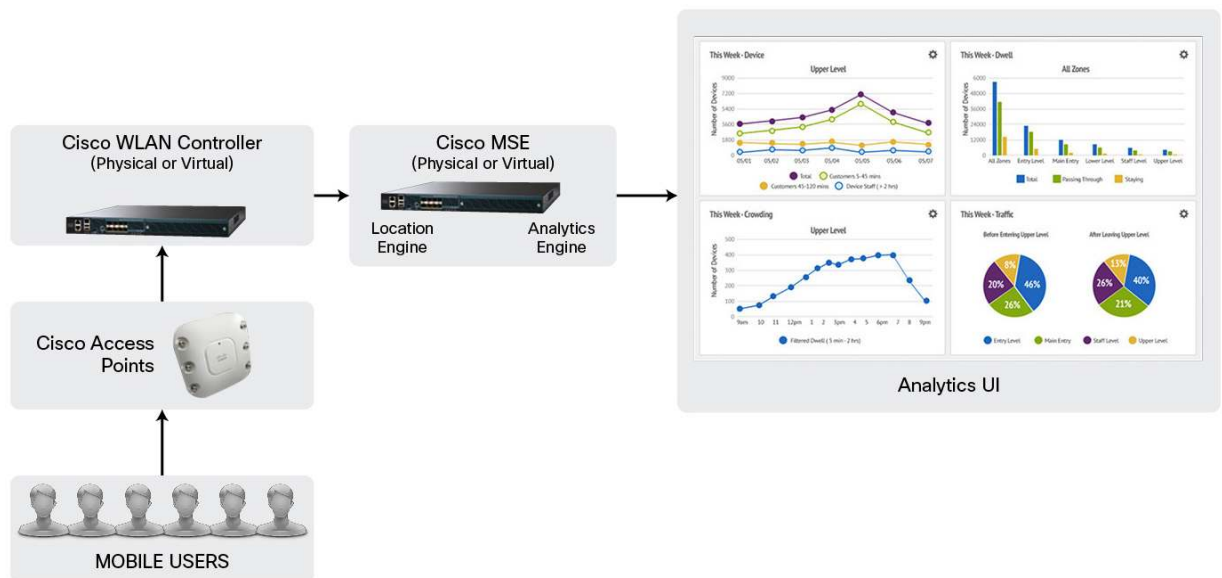
Addressing Privacy

There have been a lot of conversations over the past few months about in-store analytics and end-user privacy. Analytics are designed to capture aggregate information, rather than to track a particular visitor. No personal profile information about the visitor is collected - rather, trends and patterns of collective behavior are gathered based on the discrete time, locations, and MAC addresses of devices. Organizations can use this information to better serve customers from an operational perspective by rearranging floor layouts, or from a marketing perspective by allowing opt-in users to receive location-based mobile coupons or promotions. Participation in Wi-Fi location services is completely up to the customer or visitor - they can opt out simply by turning off their Wi-Fi signal.

Generation of the Data

When a Wi-Fi client device is enabled, it transmits 802.11 Probe Request packets to identify the wireless networks in its neighborhood and also to find the received signal strength indication (RSSI) associated with the identified Service Set Identifiers (SSIDs). Even after becoming associated with a particular access point in the WLAN, the client device will continue to transmit 802.11 Probe Request packets to identify and potentially roam for other access points, for better quality of service. The various Cisco access points in the WLAN gather these probe requests and the associated RSSI from the wireless devices in the network and forward them, along with their MAC address, to the wireless LAN controller (WLC) managing the access points. The WLC then forwards this information to the MSE, which uses the data collected from different access points to triangulate the location of the wireless device by translating the access point RSSIs to X, Y, and Z coordinates in space. As the wireless device moves through the network of access points, the MSE continuously tracks the location of the device in time. However, in practice most wireless devices have been designed to save battery life, and hence the periodicity of the probe requests to other access points may be on the order of several seconds. This indirectly translates to the MSE having discrete time location snapshots instead of a continuous stream. Since the identifier is unique to each device, this means that a chronological trail of points is available, describing the movement or path of an individual through the venue (Figures 1 and 2).

Figure 2. Flow of Location Data



These snapshots may not give a complete picture of an individual's movement. For example, the occupant may switch on his or her phone only at certain areas in the venue. However, from the millions of points recorded each week, it is possible to identify a sufficiently good set of user data on which to base analysis and derive actionable, valuable information.

Transformation of the Data

The raw Wi-Fi location data is first processed into a path for each device. Given that a device may appear several times within a day, a default cutoff of one hour with no data points indicates that a path or visit is over. Additional characteristics are introduced at this time to the devices and paths, such as whether the device was associated or probing, the zones through which it passed, and identification of a member of the staff or conference delegate. Many of these characteristics are deduced by the time and location combination in the data. For example, airports need to differentiate arriving from departing passengers; this information can be deduced by understanding the order of movement between the air-side and land-side areas of the venue. Many more such logical tags can be associated with the devices, and these can be used for later reporting, such as reports on delegates, staff, spectators, and arriving passengers. Although little information can be gleaned directly from the MAC address, except for the manufacturer, there is the option to hash the addresses to ensure another level of privacy for the user of the device.

Deriving Information on Areas and Movements

Location analytics provides two orthogonal ways to look at the processed data. This first is from the standpoint of the venue and what happens in different parts of it. Therefore, we may be interested in the waiting times in security, the number of devices seen inside a shop compared to those outside, the crowding factor at midday in a restaurant, etc. The second way is from the viewpoint of the device passing through the different parts of the venue. Here we are interested in the typical paths followed (Figure 3), where visitors start their visit, the flow through alternative paths at some junction, or the location visitors have reached after an elapsed time, etc. This information is providing different industries new ways of measuring their audience.

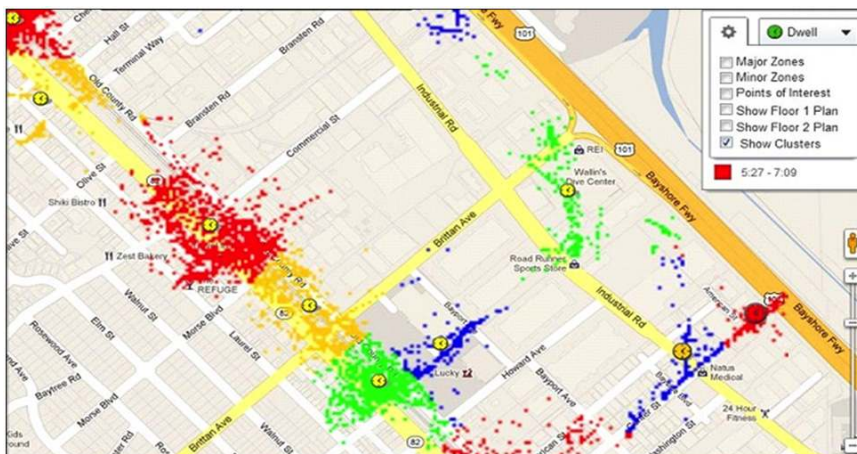
Figure 3. A Typical Path Denoted by a Sequence of Points Observed



Area and Zone Measurements

Once the parts of the venue have been established, a set of measurements can be made describing what happens within them. Location analytics provides two methods for defining the different parts of a venue. Explicit zones can be defined by the user as named polygons that represent known defined spaces. Each space can be busy or empty, thin or wide. The important factors are that it is identifiable by its name, and information pertaining to it can be put into a known context. The other method involves the use of more dynamic areas, which are generated mathematically based on where points are located. This process breaks the venue up into cluster areas using a location-based k-means algorithm (Figure 4). This option has the advantage of providing the user with immediate feedback on different parts of the venue, including when zones are either not clear or not immediately available. Indeed, it can be used to “break up” an explicit zone into areas representing different behaviors, such as parts with more and less crowding or with slow and fast movement.

Figure 4. Detected Locations Across a City, Grouped into Clusters and Measured by the Median Dwell Time Within Each



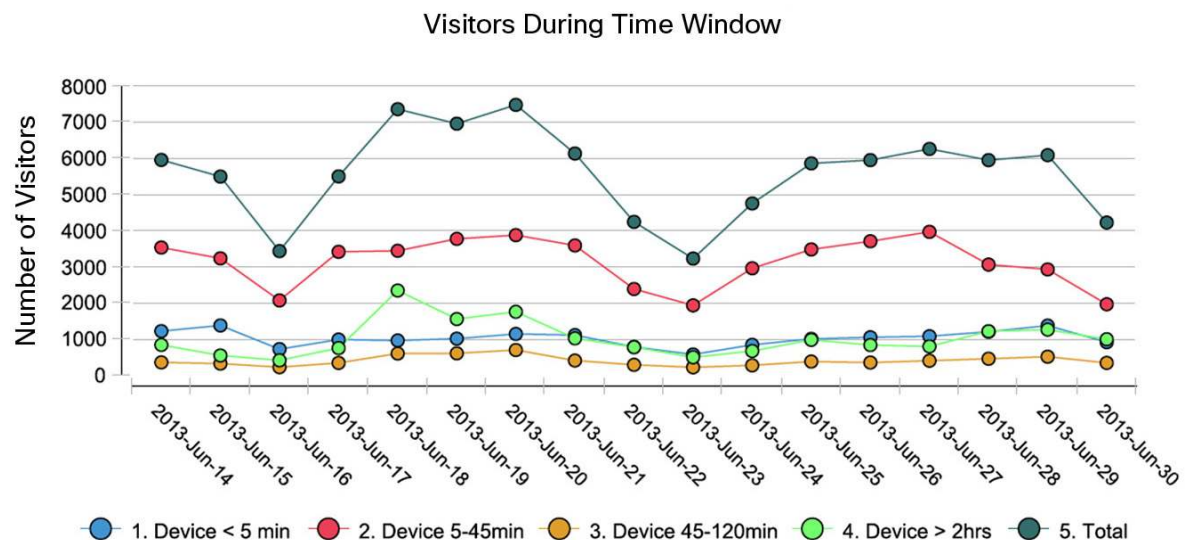
The types of measurements available for areas and zones are listed in Table 1.

Table 1. Descriptions of Area and Zone Parameters

Measure	Description	Application
Number of devices	The number of Wi-Fi devices (smartphones, tablets, or laptops) identified within an area during a time window	Determine the number of customers or potential customers by date and time
Dwell/waiting time	An estimation of the duration during which a device is present in a particular space	Determine the amount of queuing or the average time spent in a shop or facility
Crowding factor	A normalized value determined by the number of devices seen within an area during any time window	Helps determine potential bottlenecks or where extra resources are needed
Dominant direction	The most typical directions of travel through an area	Knowing the flow of pedestrians across a hallway at various times promotes better safety precautions
Speed	The straight-line distance between two points divided by the duration between them	Being able to differentiate between pedestrians, cyclists, and cars, for example, is important in a city context
Churn	The relative number of devices that appear and disappear during a time window	How well restaurants process clients at different times of the day is crucial to their profitability
Retention	The opposite of churn, measuring the relative numbers of devices remaining in the area over time	Measures how well certain venues attract and keep an audience
Frequency	The number of times a device is detected within a specific time window	Measures the loyalty of the visitor, or how many need to be informed of the layout if they are first-time visitors

Many of these measures can be combined with different zones and times to provide higher-level information, such as opportunity gap (the number of visitors outside a retailer compared to inside); retention (the number of visitors to a shop who stay more than 5 minutes, compared to passing through in less than 5 minutes); and customer loyalty (the frequency at which the same device is seen, impact of a promotion, etc.) (Figure 5).

Figure 5. The Number of Visitors to a Particular Zone and How Long They Spend There



Movement Measurements

Visitors moving through a venue can be analyzed in different ways. “Common paths” is a way to group and rank the routes taken by a set of devices. Movement is also measured by the sequence of zones a device passes through. Finally, flow analytics set up a framework of capture areas through which paths may pass.

Each of these analysis techniques gives a different interpretation of movement and hence measurement. For instance, with zone movement, we can determine the most common zones leading to a part of the venue and away from it. Common paths show the most detailed view of the precise route commonly taken through a zone. Flow analysis highlights the patterns of decision making when moving through a venue and choosing one route over another (Figure 6).

Figure 6. Flow Analytics Results of Visitors Moving in Different Directions from a Central Area

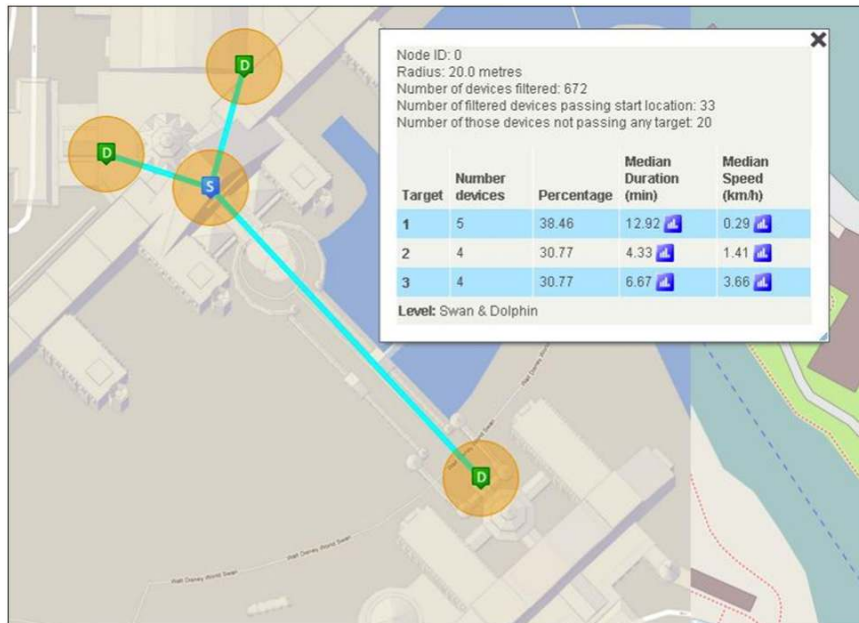


Table 2 shows the types of measurements available for movement.

Table 2. Descriptions of Movement Parameters

Measure	Description	Application
Popular paths	Using an adapted k-means mathematical clustering technique, sets of similar paths can be identified and ordered by the number of times seen. These can be further enhanced by estimating the actual path taken along corridors and through open space, without crossing barriers	Being able to identify the common routes people take to get to a shop allows identification of the type of customer and placement of advertising
Number of devices	The number of devices moving from one zone to another or between source and destination areas	Knowing the busiest parts of the venue, in terms of routes people take, gives an understanding of which parts of the venue are reached and from where. Venue penetration is a measure of its rental value
Duration	The time it takes for a device to move through a venue. This gives a measure of dwell time along a route	How long people take on the route helps provide an understanding of the opportunities to engage with passing traffic
Speed	The speed between two areas or points, calculated as a straight-line distance	This gives a safety measure for the flow of visitors down corridors

Measure	Description	Application
Percentage	A measure of how many visitors at different times of the day chose one direction over another	An easy way of seeing if any marketing initiative is having an effect on the route or destination people take
Zones before and after Zones starting and finishing Zones immediately before and after Zones after first and before last	A description of where devices have been in terms of zones and the frequency with which the devices have been seen in them	Allows shops to understand the relationships to other shops and the environment they are in, helping them learn more about their customers

Filtering

The key to the analysis of large amounts of data is to be able to focus easily on the target set of visitors and measure their behavior. Many venues have a mixed population, and trying to measure them together provides unclear results. Venues may have shoppers mixed with office workers, people walking through, and people having coffee. Outdoors it is even more heterogeneous, with different forms of transport appearing. The system provides many ways to slice the information in order to isolate the desired set of devices.

Date and time are the main parameters, as we filter on what is happening in a venue at a particular time of day, or across a week at particular time, or on the same day in successive weeks. However, even in the same time window, it may be necessary to differentiate between different types of people. Thus, the time parameters are often combined with a location filter to focus on one zone of a venue. The devices belonging to one group of people may be determined generally through tagging the device in advance through their location and time. This can be done, for example, with arrivals and departures, or with staff and conference delegates. Identifying people who are engaging with the environment as opposed to passing through is possible through the dwell time parameter, which identifies, for example, people waiting for their luggage as opposed to exiting directly, or people sitting down to eat in a busy thoroughfare. Many other ways are possible by combining the parameters listed in Table 3.

Table 3. Parameter Descriptions

Type	Description
Date	For specific days or ranges of dates
Time	For certain times of the day, such as lunchtimes, evenings, or peak times
Zone	Devices that are used only in one or more parts of the building
Start zone, finish zone, intermediate zone	Allows the analysis to focus on device paths starting, passing through, or finishing at a particular zone. Usually used in conjunction with popular paths analysis.
Quality	Data quality can fluctuate based on access point coverage, device type, and device usage. This parameter allows fuller sets of data to be used, such as long paths or confirmed dwell times
Associated or probing	Client devices can at any point be associated or probing
Advanced tagging	Devices can be tagged in advance according to certain user-defined location and time characteristics

Mobility Services API and Location Analytics

The Mobility Services Engine also provides the Mobility Services API with support for REST. This API data is presented to the processed analytics data, (points, paths, and devices) as well as to the result information (number of devices, average dwell time, movement, etc. - all indexed by zone and time/date). This API allows the use of industry-specific analytics and also enables users to import the results into existing reports or management information systems. This API is especially useful as it provides a robust interface to the MSE raw data. This data has been cleaned from stray devices and false locations, while aggregated points are created where a device is perceived to be stationary.

Real-Time and Predictive Analytics

The main difference between historical and real-time analytics is that in real time we have incomplete knowledge of what a visitor is doing. That is, paths have yet to terminate, and so we do not know how long the device will stay or where it will go before finishing the visit. Consequently, new algorithms for parameter estimation have been built. These either take what knowledge is available or use historical data to deduce likely outcomes. Now the user is able to ask questions such as, “What is the current waiting time?” “How many people are passing through this corridor?” and “What is the current flow between these two routes?” Actionable information can be delivered immediately to use in making a new set of decisions involving resource allocation, variable signage, or security engagement.

Rule-Based Notifications and Alerting

Analytics allow us to discover certain patterns, such as waiting times during the day, crowding in certain zones, etc. However, it can be the case that these patterns don't appear very often, and when they do appear it is important to realize this and take action. Location analytics provides the ability to describe conditions involving the parameters and send alerts when these conditions are met. For example, a gradual buildup of visitors across different parts of a venue can give early notification that certain access routes should be opened. In a more advanced scenario, the location of staff on a shop floor can be better distributed to match the presence of customers.

Conclusion

Each shopper, passenger, or office worker interacts in different ways with their environment. Some may stay in certain spots while others pass straight through or move from spot to spot, and some start off as shoppers and progress to being cinema goers. Only with a powerful, integrated set of locational analytics tools can we focus on the different groups and understand the general pattern of movement across the building at different times.

Having established a model of visitor behavior, location analytics then allows us to take a real-time view, a predictive view, and a reactionary view to provide higher-level capabilities for the owner of the shop, mall, or airport. These views provide knowledge about the occupants of the building to deliver better service, operation, and security for all.

Cisco's location analytics tools deliver these insights from the masses of data obtained through the MSE. For more information on Cisco's location analytics and Connected Mobile Experiences solution, visit

<http://www.cisco.com/go/cmz> and <http://www.cisco.com/go/mse>.



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)