



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES

CREACIÓN DE ANALÍTICOS PARA ENTENDER EL COMPORTAMIENTO DE LOS CLIENTES Y MEJORAR LA GESTIÓN EMPRESARIAL, UTILIZANDO SENsoRES DE MACs EN AMBIENTES INTERIORES

Por

Br. Sarait de Jesús Hernández Ruiz

Tutor: Prof. José Luis Paredes, Ph.D

Febrero, 2020

©2020 Universidad de Los Andes Mérida, Venezuela

C.C. Reconocimiento

Creación de analíticos para entender el comportamiento de los clientes y mejorar la gestión empresarial, utilizando sensores de MACs en ambientes interiores

Br. Sarait de Jesús Hernández Ruiz

Proyecto de Grado — Sistemas Computacionales, 68 páginas

Resumen: En la actualidad, la mayoría de las personas cuentan con dispositivos móviles que tienen tecnologías wifi integradas. Su uso globalizado ha hecho que los investigadores se interesen en el análisis de los datos que pueden ser obtenidos a través de estos dispositivos. Los dispositivos Meshliums de la empresa Libelium, son parte de los sensores de señales inalámbricas (wifi, Bluetooth...) más populares en el mercado. En un centro comercial, estos sensores pueden ser instalados con el fin de capturar todas las señales emitidas por los dispositivos móviles. Este proceso es de especial interés para un experto en mercadotecnia ya que los datos capturados asociados a dispositivos móviles, están íntimamente relacionados con el comportamiento del consumidor dentro de cualquier ambiente interior. El presente trabajo se enfoca en aplicar analítica a los datos capturados por los sensores wifi Meshliums, con el fin de generar información de utilidad que sirva de apoyo, a un analista de negocios, en la toma de decisiones para la mejora de gestión en un comercio y, posiblemente, un incremento de sus ventas.

Palabras clave: Analítica de datos, Meshlium, wifi, comportamiento del consumidor, mercadotecnia

C.C. Reconocimiento

Índice general

Índice de Tablas	VII
Índice de Figuras	VIII
Agradecimientos	X
1. Introducción	1
1.1. Antecedentes	2
1.2. Planteamiento del problema	4
1.3. Justificación	4
1.4. Objetivos	5
1.4.1. General	5
1.4.2. Específicos	5
1.5. Alcance	6
1.6. Metodología	6
1.6.1. Fase de contextualización	6
1.6.2. Fase de pruebas locales	7
1.6.3. Fase de análisis de datos	7
1.6.4. Fase de generación de analíticos	7
2. Marco Teórico	8
2.1. Redes inalámbricas wifi	8
2.1.1. Solicitudes de sondeo	9
2.1.2. Sensores wifi	10
2.1.3. Dirección MAC	11

2.1.4. Aleatorización de MAC	11
2.1.5. Indicador de fuerza de la señal recibida (RSSI)	13
2.1.6. Propagación de señales inalámbricas en ambientes interiores	13
2.2. Localización en ambientes interiores	14
2.3. Mapas térmicos	16
2.4. Diagramas de Sankey	17
2.5. Mercadotecnia	18
2.6. Análisis de datos	19
2.7. Analítica de datos	20
3. Equipo de adquisición de datos	21
3.1. Meshlium Xtreme Scanner 4G AP	21
3.1.1. Meshlium Manager System V4.0	22
3.1.2. Detección de dispositivos móviles	25
3.1.3. Valores de RSSI medidos por los Meshliums	27
3.2. Limitaciones de los equipos Meshliums	27
3.3. Pruebas locales	28
3.3.1. Estudio de la calidad del dato	28
3.3.1.1. Descripción del escenario de prueba	29
3.3.1.2. Conclusiones y resultados	31
3.3.2. Estudio de la aleatorización	35
3.3.2.1. Descripción del escenario de prueba	36
3.3.2.2. Conclusiones y resultados	38
4. Análisis de los datos capturados en un ambiente de producción real	41
4.1. Descripción del ambiente de producción real	41
4.2. Análisis de datos	42
4.2.1. Análisis exploratorio y visualización de datos	42
4.2.1.1. Reporte de estabilidad de datos	43
4.3. Limpieza de datos	51
4.3.1. Detección de valores atípicos	51
4.3.2. Filtros aplicados	52

5. Analíticos generados a partir de la base de datos capturada	54
5.1. Estadísticos de interés	54
5.2. Mapas térmicos	58
5.3. Flujo de clientes	60
6. Conclusiones y trabajos futuros	63
A. Consulta al soporte técnico de Libelium	65
Bibliografía	66

www.bdigital.ula.ve

C.C. Reconocimiento

Índice de Tablas

3.1. Escenario de prueba.	30
3.2. Escenario de prueba.	37
4.1. Clasificación de las MACs según el tiempo de permanencia.	45
4.2. Resumen del tiempo de escaneo para los Meshliums.	47

www.bdigital.ula.ve

Índice de figuras

2.1. Arquitectura de una red WLAN. Tomado de Kurose y cols. (2010).	9
2.2. Solicitud de sondeo por medio de la exploración activa. Tomado de Kurose y cols. (2010).	10
2.3. Dirección física o MAC de un dispositivo.	11
2.4. Comunicación en redes inalámbricas: desde A hasta B sin línea de vista y desde B hasta C con línea de vista. Adaptado de Kurose y cols. (2010). .	14
2.5. Diagrama funcional de un sistema de posicionamiento. Traducido de Pahlavan y cols. (2002).	15
2.6. Ejemplo de mapa térmico. Tomado de Goulart (2014).	17
2.7. Ejemplo de diagrama de Sankey. Tomado de Schmidt (2006).	18
3.1. Meshlium Scanner 4G AP. Tomada de Libelium (2019b).	22
3.2. Sensores Wasp mote interconectados a través del equipo Meshlium como puerta de enlace. Tomada de Libelium (2019b).	23
3.3. Diagrama para los conectores a la nube. Tomada de Libelium (2019b). .	24
3.4. Meshlium Manager System. Tomada de Libelium (2019b).	25
3.5. Extracto de la base de datos capturada por un Meshlium.	26
3.6. Escenario de prueba para el estudio de la calidad del dato.	30
3.7. Frecuencia de las solicitudes de sondeo para los casos de estudio.	32
3.8. RSSI para diferentes posiciones con línea de vista.	33
3.9. RSSI para diferentes posiciones sin línea de vista.	34
3.10. Representación en diagramas de cajas del RSSI.	35
3.11. Escenario de prueba para el estudio de la aleatorización de MACs. . . .	38
3.12. Solicitudes de sonda para el estado asociado.	39

3.13. Solicitudes de sonda para el estado no asociado.	39
4.1. Plano y ubicación de los AP y sensores wifi en el CCA.	42
4.2. Diagrama entrada-proceso-salida: Programa para la generación del reporte de estabilidad de datos.	43
4.3. Tiempo total de adquisición de datos por Meshlium.	44
4.4. Clasificación de las MACs capturadas según su tiempo de permanencia.	46
4.5. Estabilidad del tiempo de escaneo.	47
4.6. Variabilidad del RSSI capturado.	48
4.7. Promedio del RSSI capturado.	49
4.8. Representación del rango de cobertura de los equipos Meshliums.	50
4.9. Ejemplo de coincidencias de MACs encontradas.	51
4.10. Ejemplo de coincidencias de MACs encontradas.	51
5.1. Cantidad estimada de clientes para un día.	55
5.2. Cantidad estimada de clientes para todos los días.	56
5.3. Distribución de los clientes por horas del día.	56
5.4. Porcentaje de visitas de clientes por día.	57
5.5. Porcentaje de visitas: semana vs fin de semana.	58
5.6. Mapa térmico generado para un lunes dentro del CCA.	59
5.7. Mapa térmico generado para un sábado dentro del CCA.	59
5.8. Muestra de los colores seleccionados para cada AP en el CCA.	61
5.9. Diagrama de Sankey para un día sábado en el CCA.	61
5.10. Diagrama de Sankey para un día lunes en el CCA.	62
A.1. Respuesta del soporte técnico de Libelium por el RSSI = -1dBm.	65

Capítulo 1

Introducción

Con los grandes avances tecnológicos de los últimos años, las empresas se encuentran en un proceso de transformación constante lleno de incertidumbre y dinamismo. Por esta razón, para un empresario, estudiar el comportamiento del consumidor dentro de sus tiendas, asegura que este proceso de transformación sea exitoso para poder seguir satisfaciendo las necesidades de los clientes de forma óptima y mantenerse en el mercado (Zeng y cols., 2015).

La gran cantidad de usuarios que poseen dispositivos móviles con tecnología wifi, permite que los comerciantes hagan seguimiento a los posibles clientes dentro de una tienda, con el despliegue de sensores wifi que capturan los paquetes que emiten estos dispositivos.

Los Meshliums son sensores inalámbricos desarrollados por la empresa Libelium, los cuales han obtenido gran popularidad los últimos años por su amplia gama de aplicaciones. Una de sus aplicaciones es fugir como sensor wifi en ambientes interiores (Libelium, 2019a); como por lo general, estos ambientes están caracterizados por la presencia de obstáculos que hacen que las señales inalámbricas fluctúen y sean capturadas con ruido (Long Vu, 2011), se requiere un análisis de los datos que los Meshliums capturan para poder determinar si son aptos para hacerle seguimiento a usuarios dentro de una tienda, y determinar qué analíticos de utilidad se pueden generar a partir de ellos; con el fin de que un experto en mercadotecnia logre proponer soluciones factibles para mejorar la gestión empresarial dentro de un comercio.

Entre los problemas principales que se encuentran al desplegar sistemas de seguimiento a usuarios usando sensores de redes wifi, está la asignación de direcciones MACs aleatorias de los dispositivos móviles (Rutermann y cols., 2019). El análisis del comportamiento de los dispositivos móviles mientras realizan este proceso permite comprender las implicaciones sobre los analíticos generados a partir del despliegue de las tecnologías de sensores wifi dentro de un centro comercial.

1.1. Antecedentes

Según C. Wu y Xi (2013), la localización en ambientes interiores (centros comerciales, teatros, hospitales, etc.) usando sistemas de posicionamiento wifi, ha tenido un gran auge en los últimos años, debido a que presentan una amplia gama de aplicaciones, como son el comercio, las ciencias sociales, la planificación urbana, el transporte, la vigilancia y monitoreo, entre otros, y por lo rentable que es su despliegue, ya que existe una gran cantidad de dispositivos que cuentan con esta tecnología.

Los sistemas de posicionamiento wifi utilizan sensores de señales inalámbricas (emitidas por una gran cantidad de dispositivos electrónicos), para determinar la posición de equipos. Actualmente, los dispositivos Meshliums son uno de los sensores más populares en el mercado, han sido usados en una gran cantidad de proyectos de interés comercial, por ejemplo, en 2018 se realizó un estudio para identificar el número de visitantes y su comportamiento en una feria de Reino Unido, también se usaron para monitorear la actividad de los pasajeros en el aeropuerto de Manchester y como sensores para una aplicación del Internet de las Cosas que automatiza procesos agrícolas (Libelium, 2019a).

Estos sensores capturan datos, cuyo estudio, permite comprender el comportamiento del consumidor a través del análisis de su desplazamiento dentro de un centro comercial. Estos datos proporcionan información crucial para un comerciante en términos de efectividad de las promociones, organización de productos y eficiencia de los servicios.

Trabajos recientes han propuesto distintos métodos para analizar el comportamiento de los consumidores a partir de los datos obtenidos por sensores wifi,

específicamente, en Zeng y cols. (2015) reconocen algunos estados representativos del comportamiento del comprador, en la entrada y dentro de la tienda, por medio del perfilado de la información del estado del canal de wifi. El principal problema de este enfoque es que suponen que hay un solo comprador en el rango de cualquier enlace wifi, lo que no sucede en un ambiente de producción real. Otro estudio previo Yaeli y cols. (2014), busca, por medio del análisis y la visualización de los datos escaneados, obtener información sobre los patrones de comportamiento de los clientes y comprender, por ejemplo, cuánto tiempo pasan los clientes en diferentes áreas de la tienda y cuáles son las rutas que toman, usando mapas *Choropleth* y de flujo respectivamente.

Existen varios enfoques propuestos en la literatura para los sistemas de posicionamiento en ambientes interiores a partir de la captura de señales wifi y bluetooth emitidas por dispositivos. En Jeril Kuriakose y Nandhini (2014) usan los puntos de acceso como nodos wifi que actúan como sensores para detectar ubicaciones. Analizar el comportamiento y los patrones de movimiento del consumidor en un ambiente interior sigue siendo un desafío. Los investigadores han concluido que entre los problemas principales de este estudio y del posicionamiento de dispositivos en interiores, está la fluctuación de la señal wifi capturada (aunque un teléfono celular se encuentre en una ubicación fija) y la aleatorización de MACs de los dispositivos móviles (Long Vu, 2011).

En Martin y cols. (2017) proponen una serie de especificaciones sobre las diferentes marcas de dispositivos móviles celulares y bajo qué condiciones éstos realizan aleatorización de MACs, así como también mencionan una serie de ataques que pueden ser aplicados para obtener la MAC real de los dispositivos. Aunque obtienen muy buenos resultados en su investigación, los fabricantes mejoran significativamente sus algoritmos de aleatorización con las nuevas actualizaciones de software, además, muchos de los ataques requieren un gran poder de computo que hace poco viable su uso. Por otra parte en Rutermann y cols. (2019) realizan una serie de experimentos que permiten tomar una idea general de la cantidad real de dispositivos móviles en un ambiente interior aún cuando éstos realizan aleatorización de MACs.

1.2. Planteamiento del problema

Los dispositivos Meshliums pertenecen a un conjunto de sensores inalámbricos que permiten capturar la señales emitidas por cualquier dispositivo que cuente con tecnologías como wifi, Bluetooth y del Sistema de Posicionamiento Global (GPS). En los últimos años estos dispositivos han obtenido un gran interés a nivel comercial, ya que los datos que capturan están íntimamente relacionados a usuarios y a partir de su estudio se puede generar información para descubrir tendencias de consumo, lugares más visitados dentro de una tienda, cantidad de visitantes, entre otros.

Investigaciones pasadas sobre el funcionamiento y aplicaciones de los dispositivos Meshliums, no se han ocupado de hacer un análisis exhaustivo de la calidad de los datos que escanean, su aplicabilidad para la localización de dispositivos móviles en ambientes interiores y los analíticos que se pueden generar a partir de ellos con el fin de mejorar la gestión empresarial.

Aunado a esto, en ambientes interiores sistemas para la localización como el GPS no funcionan de forma correcta. Sin embargo, la localización en ambientes interiores puede hacerse a partir del rastreo de dispositivos móviles capturando las señales wifi que éstos emiten. Los ambientes interiores están caracterizados por la presencia de obstáculos que impiden una buena línea de vista, fluctuación de señales y ruido, por lo que usar señales wifi para hacer seguimiento a usuarios es una tarea compleja. Además, a partir de iOS 8 y Android 6, los fabricantes agregaron al software de los dispositivos móviles una funcionalidad para la asignación de direcciones MACs aleatorias, por ello identificar de manera única y hacer seguimiento a dispositivos móviles dentro de cualquier entorno puede tornarse sumamente complicado.

1.3. Justificación

Los equipos Meshliums se posicionan cada vez más, en el mundo del Internet de las cosas como los principales escáneres de dispositivos móviles orientados a monitorear la actividad en comercios, ciudades y tráfico vehicular, funcionando a su vez como una puerta de enlace para otros sensores. Aunque existe una gran variedad de dispositivos en

el mercado capaces de realizar estas tareas, ninguno ofrece el total de las funcionalidades de los Meshliums, haciéndolos populares en el ámbito.

Dado que el mundo empresarial es un ambiente dinámico, es sumamente importante llevar a cabo una adecuada evaluación de los potenciales clientes y buscar satisfacer sus necesidades. Encontrar patrones de consumo y mejorar la calidad del servicio puede ser lo que marque la diferencia del comportamiento del mercado en los próximos años, es por ello que en la actualidad, los emprendedores en el mundo comercial, buscan hacer seguimiento a los usuarios dentro de sus tiendas. Los equipos Meshliums presentan una buena opción para implementar esta tecnología dentro de centros comerciales. Además, la gran cantidad de dispositivos móviles en la actualidad, hace que desplegar sistemas de monitoreo de clientes en ambientes interiores con señales wifi, sea rentable sin la necesidad de una gran inversión adicional en infraestructura.

1.4. Objetivos

1.4.1. General

Crear analíticos para entender el comportamiento de los clientes y mejorar la gestión empresarial, utilizando los sensores wifi Meshliums en ambientes interiores.

1.4.2. Específicos

- Realizar una revisión bibliográfica sobre los analíticos dirigidos a mejorar la gestión empresarial a partir de la localización de dispositivos móviles en ambientes interiores.
- Comprender el funcionamiento y detalles técnicos de los dispositivos Meshliums.
- Ejecutar pruebas en un ambiente controlado para medir el alcance, la robustez y la reproducibilidad de los datos capturados por los equipos Meshliums.
- Analizar la estabilidad de los datos capturados por los dispositivos Meshliums en un ambiente de producción real.

- Generar analíticos para entender el comportamiento de los clientes y mejorar la gestión empresarial, a partir de los datos capturados por los Meshliums en un ambiente de producción real.

1.5. Alcance

En esta investigación se pretende realizar un amplio estudio de la calidad de los datos wifi capturados por los dispositivos Meshliums. Los datos capturados para su posterior análisis pertenecen a dispositivos propuestos para una serie de pruebas locales y del despliegue de 4 Meshliums en un ambiente de producción real.

A partir del análisis de los datos, se aspira determinar qué analíticos e información útil se puede generar para la mejora de la gestión empresarial, ya que en ambientes reales, los datos capturados por los Meshliums corresponden a dispositivos móviles que pueden ser asociados a clientes.

1.6. Metodología

1.6.1. Fase de contextualización

- Revisión y estudio de los conceptos fundamentales asociados al análisis de datos y la analítica de datos.
- Revisión de artículos científicos relacionados con mejorar de la gestión empresarial.
- Revisión de artículos científicos relacionados con la generación de analíticos a partir de la localización de dispositivos móviles en interiores.
- Revisión de artículos científicos relacionados con la aleatorización de MACs en dispositivos móviles.
- Revisión de manuales y soporte técnico brindados por Libelium para comprender el funcionamiento, configuraciones y limitaciones de los dispositivos Meshliums.

1.6.2. Fase de pruebas locales

- Diseñar, implementar y analizar los resultados de una prueba local que mida el alcance, la robustez y la reproducibilidad de los datos capturados por los equipos Meshliums.
- Diseñar, implementar y analizar los resultados de una prueba local para estudiar la aleatorización de MACs en dispositivos móviles.

1.6.3. Fase de análisis de datos

- Obtener una base de datos escaneada por los dispositivos Meshliums en un ambiente de producción real (entorno de estudio).
- Comprender las ventajas y limitaciones de la distribución física del entorno de estudio.
- Determinar, a partir del procesamiento y análisis de la base de datos de estudio, los filtros a ser aplicados para obtener datos útiles para generar analíticos.
- Estudiar la influencia que puede tener un punto de acceso, sobre el funcionamiento de los dispositivos Meshliums.

1.6.4. Fase de generación de analíticos

- Calcular estadísticos de interés para comerciantes por medio de la clasificación de los dispositivos móviles capturados según su tiempo de permanencia en el entorno de estudio.
- Determinar las zonas más visitadas por los clientes dentro del entorno de estudio usando mapas térmicos.
- Precisar las rutas más transitadas por los visitantes dentro del entorno de estudio usando diagramas de flujo tipo Sankey.

Capítulo 2

Marco Teórico

2.1. Redes inalámbricas wifi

El wifi es una tecnología que le permite a los dispositivos electrónicos conectarse entre sí o a Internet por medio de un punto de acceso (AP, por sus siglas del inglés *Access Point*), se basa en los estándares IEEE 802.11 y es ampliamente utilizado a través del mundo. Según Pozuelo (2009) “Actualmente, es la tecnología más extendida para la transmisión inalámbrica de datos en el interior de edificios y está presente en la mayor parte de los edificios tanto públicos como de empresas privadas”.

Un AP es un dispositivo que crea una red de área local inalámbrica (WLAN, por sus siglas del inglés *Wireless Local Area Network*) al conectarse a un router por un cable Ethernet y proyecta una señal wifi en un área específica (ver Fig. 2.1). En una red doméstica por lo general existirá un punto de acceso y un router, normalmente integrados en una misma unidad y vendidos comercialmente como routers wifi (Kurose y cols., 2010). En general, los routers toman cada paquete que llega a sus enlaces de entrada y lo reenvían a uno de sus enlaces salida, eligiendo la ruta que debe tomar el paquete para llegar a su destino.

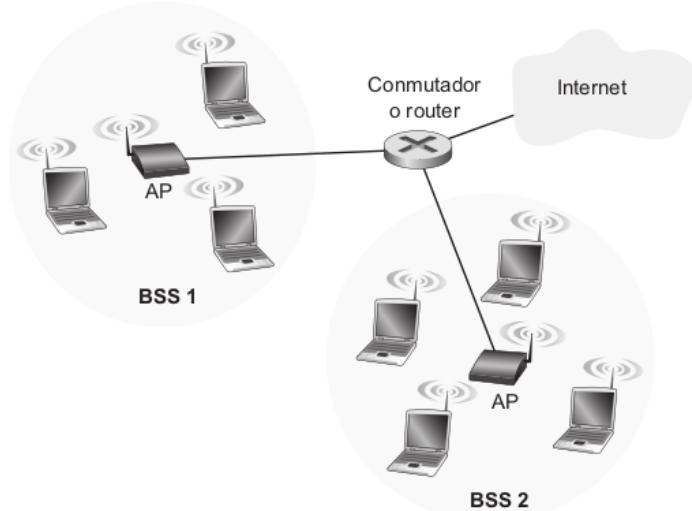


Figura 2.1: Arquitectura de una red WLAN. Tomado de Kurose y cols. (2010).

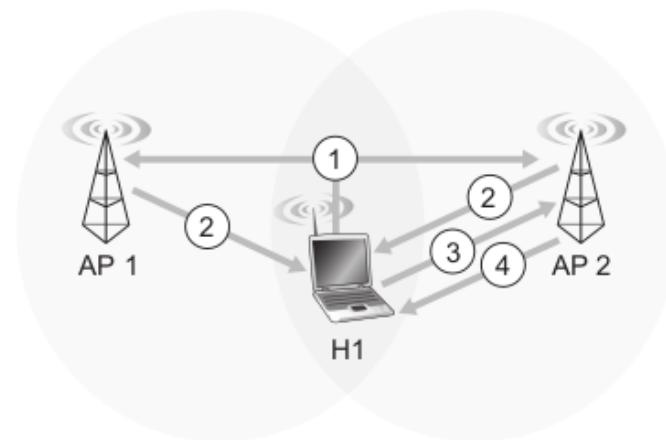
2.1.1. Solicitud de sondeo

En las redes wifi cada dispositivo inalámbrico necesita asociarse con un punto de acceso antes de poder enviar o recibir datos de la capa de red, es decir, sólo el punto de acceso puede enviar las tramas (paquetes de datos de la capa de enlace) del dispositivo inalámbrico hacia Internet.

El estándar 802.11 requiere que un punto de acceso envíe periódicamente tramas baliza (del inglés, *beacon frames*), cada una de las cuales incluye la dirección MAC (ver sección 2.1.3) y el identificador del punto de acceso. Los dispositivos interesados en asociarse a un punto de acceso exploran los once canales de transmisión wifi buscando las tramas baliza (proceso denominado exploración pasiva). Cuando un dispositivo determina los puntos de acceso disponibles a través de las tramas baliza, selecciona uno para llevar a cabo la asociación (Kurose y cols., 2010).

Otra forma en la que un dispositivo puede asociarse a un punto de acceso, es difundiendo una trama denominada solicitud de sondeo (del inglés, *probe request*) que será recibida por todos los puntos de acceso que se encuentren en el rango de cobertura del dispositivo (proceso denominado exploración activa). Cada punto de acceso responde a la solicitud de sondeo con una trama denominada respuesta

de sondeo, en este momento, el dispositivo puede elegir a qué punto de acceso, en el conjunto de los que respondieron, asociarse (Kurose y cols., 2010). Después de seleccionar el punto de acceso al cual se va a conectar, el dispositivo envía una solicitud de asociación a ese AP, el cual responde con una trama de respuesta de asociación. En la Fig. 2.2 se ve de forma explícita este proceso.



b. Exploración activa

1. Difusión desde H1 de una trama de solicitud de sondeo.
2. Envío de tramas de respuesta de sondeo desde los AP.
3. Envío de trama de solicitud de asociación desde H1 al AP seleccionado.
4. Envío de trama de respuesta de asociación desde el AP seleccionado a H1

Figura 2.2: Solicitud de sondeo por medio de la exploración activa. Tomado de Kurose y cols. (2010).

2.1.2. Sensores wifi

En redes de computadores el modo monitor o modo promiscuo, es una configuración que soportan algunas tarjetas de red que permite en una red de uso compartido, capturar todo el tráfico de paquetes que circulan por ella (Aspyct.org, 2019). De este modo, un sensor wifi es un dispositivo que tiene un conjunto de tarjetas de red integradas que permiten ser configuradas en modo monitor.

Del mismo modo, los sensores wifi emiten constantemente solicitudes de sondeo (ver sección 2.1.1) con la idea de detectar la presencia de todos los dispositivos móviles que

se encuentre en su rango de cobertura.

2.1.3. Dirección MAC

Cada dispositivo que cuente con una tarjeta de red integrada, posee una dirección única física denominada dirección MAC (por su siglas en inglés, *Media Access Control*) y son (en teoría) globalmente únicas (dos dispositivos no pueden contar con la misma dirección MAC). La MAC está almacenada en el firmware de la tarjeta adaptadora y cada dirección cuenta con 48 bits, divididos por bloques de 6 bytes o octetos, estando a su vez, cada bloque conformado por dos caracteres hexadecimales.

El conjunto de direcciones MAC que poseen las organizaciones son asignados por el Instituto de Ingeniería Eléctrica y Electrónica (IEEE). Cada empresa posee un identificador único de organización (OUI) de 24 bits, el cual corresponde a los primeros 24 bits de todas las MACs pertenecientes a los dispositivos que venden estas organizaciones, de esta forma al observar los primeros octetos de una MAC se puede identificar el fabricante del dispositivo. La Fig. 2.3 muestra un ejemplo de una dirección MAC, los octetos pertenecientes al OUI de la organización y la parte asignada a la tarjeta de red.



Figura 2.3: Dirección física o MAC de un dispositivo.

2.1.4. Aleatorización de MAC

Es posible hacer seguimiento a dispositivos móviles y por lo tanto a usuarios, usando su dirección MAC como un identificador único. Algunos proveedores de dispositivos móviles han tomado la decisión de aleatorizar las direcciones MACs de los teléfonos

con el fin de mantener la privacidad de los clientes.

La aleatorización de MACs es un proceso que se lleva a cabo mientras el dispositivo móvil tiene activa su tarjeta de red inalámbrica y no se ha conectado a un router o punto de acceso a Internet. Los algoritmos de aleatorización hacen que cuando un dispositivo emite solicitudes de sondeo no revele su dirección física (como es de esperarse), en cambio emite direcciones MACs falsas y aleatorias. De este modo, cuando un sensor de redes inalámbricas escanea un dispositivo, no detecta su MAC real garantizando que un usuario no pueda ser rastreado fácilmente. Mientras que la relación entre el dispositivo y la dirección MAC real es 1:1, la relación entre el dispositivo y las MACs aleatorias es 1:N, es decir, un dispositivo móvil sólo cuenta con una dirección MAC real, pero un dispositivo móvil pueda generar muchas direcciones MACs aleatorias (Martin y cols., 2017).

El proceso de aleatorización se detiene en el momento en que un dispositivo móvil establece conexión con un punto de acceso, a partir de ese momento cada solicitud de sondeo emitida por el dispositivo tendrá su MAC real. Según Martin y cols. (2017) el proceso de aleatorización de MAC varía de acuerdo al fabricante del producto. En este contexto, la MAC de los dispositivos con sistema operativo Android varía solamente los últimos 3 octetos, manteniendo el OUI de la empresa fabricante; de forma contraria, la MAC de los dispositivos con sistema operativo iOS varía completamente, es decir, los 8 octetos de la MAC cambian. Esto ocasiona que los dispositivos fabricados por la empresa Apple puedan simular ser dispositivos fabricados por otras empresas.

Intentos realizados por terceros (Martin y cols., 2017) proponen una serie de ataques que pueden ser aplicados para obtener la MAC real de los dispositivos y de esta manera poder hacerles seguimiento. Aunque obtienen muy buenos resultados en su investigación, en las nuevas actualizaciones de software de los dispositivos móviles, los fabricantes mejoran significativamente sus algoritmos de aleatorización, haciéndolos más fuertes para combatir los ataques o incluso eliminando las debilidades existentes, además, muchos de los ataques requieren un gran poder de computo que hace poco viable su uso.

2.1.5. Indicador de fuerza de la señal recibida (RSSI)

El indicador de fuerza de la señal recibida (RSSI por las siglas del inglés *Received Signal Strength Indicator*), es un valor numérico que sirve como indicador para medir el nivel de potencia de las señales recibidas por dispositivos móviles en redes inalámbricas (Barbolla, 2005). Este valor se obtiene durante la etapa de preámbulo en la comunicación inalámbrica, es decir, en el momento que se realiza una solicitud de sondeo.

Las unidades de medida que indican la intensidad de la señal recibida son el RSSI y el dBm. El dBm es una relación de la potencia medida con respecto a un mW (milivatio). Mientras que el dBm es un índice absoluto, el RSSI es relativo.

El RSSI es un número entero, mientras más alto el valor, más fuerte la señal. En general, cada fabricante o vendedor de chips define su propio valor máximo de RSSI. Esto porque el estándar 802.11 no da ninguna definición de cómo se relacionan el RSSI y el nivel de potencia en mW o dBm. Por ejemplo, Symbol usa un RSSI máximo de 100, mientras que Atheros lo fija en 60. Cuando se mide en dBm el RSSI toma números negativos, cuyo rango está comprendido entre -35 dBm y -90 dBm; el RSSI cuyo valor se encuentre más cercano a cero indica una mayor fuerza en la señal recibida (NetSpot, 2019).

2.1.6. Propagación de señales inalámbricas en ambientes interiores

A diferencia de los ambientes exteriores, en ambientes interiores existe una gran cantidad de obstáculos que impiden una buena propagación de las señales de radiofrecuencia, es por ello que la señal emitida por una antena experimenta múltiples transformaciones a lo largo de la trayectoria en la que debe propagarse hasta llegar a un receptor. Esto hace muy difícil predecir la intensidad de la señal recibida en un determinado punto (Olivares, 2009).

En entornos cerrados, existe una gran cantidad de objetos u obstáculos que impiden tener visión directa o línea de vista (ver Fig. 2.4) entre los emisores y receptores de señales inalámbricas. Los materiales (vidrio, madera, cemento...) con los que están

construidos los objetos que crean interferencias, en muchos casos causan atenuación al ser atravesados, reflexión o absorben la señal inalámbrica, haciendo que los niveles de la misma varíen en mayor medida que en entornos abiertos.

Es por esto que según Olivares (2009) los modelos de propagación en ambientes interiores difieren de los modelos de propagación tradicionales en dos aspectos:

- Las distancias cubiertas son mucho más pequeñas.
- El componente variable del entorno es mucho mayor para separaciones más pequeñas entre transmisor y receptor.

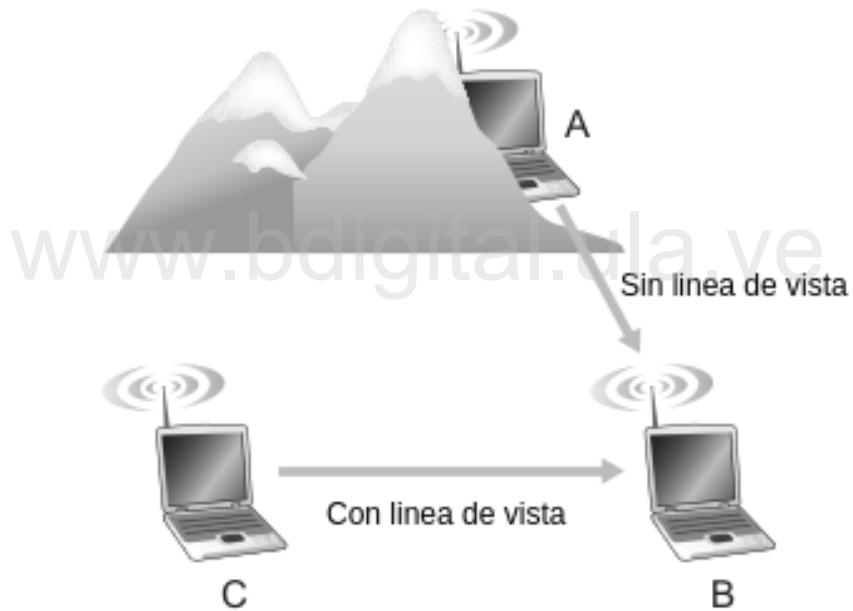


Figura 2.4: Comunicación en redes inalámbricas: desde A hasta B sin línea de vista y desde B hasta C con línea de vista. Adaptado de Kurose y cols. (2010).

2.2. Localización en ambientes interiores

Los servicios de localización en ambientes interiores surgen como una solución alternativa al uso del Servicio de Posicionamiento Global (GPS) en espacios cerrados, ya que éste no penetra bien en estas zonas. La localización en ambientes interiores basada

en la captura de las señales emitidas por dispositivos móviles, según Nuño Barrau (2006) puede ofrecer información útil en distintas aplicaciones como:

- En aplicaciones comerciales para analizar tráfico de clientes en tiendas.
 - Gestión del funcionamiento interno de instalaciones: control de temperatura dependiendo de los usuarios que se encuentran en el área, manejo del encendido y apagado de luces, máquinas, equipos de sonido, entre otros.
 - Sistemas de control de accesos: puertas que se abren solo para ciertos usuarios, acceso automático en cines, pago directo al transporte público, entre otros.
 - Estadísticas: información sobre rutinas de empleados u ocupación de una oficina.
 - En sistemas para la seguridad en la red, al añadir una verificación de ubicación física del usuario además del proceso de encriptación de archivos electrónicos.

Con el fin de ayudar al desarrollo de un marco científico para la localización en ambientes interiores Pahlavan y cols. (2002) plantea un diagrama de bloques funcional de los sistemas de posicionamiento inalámbricos.

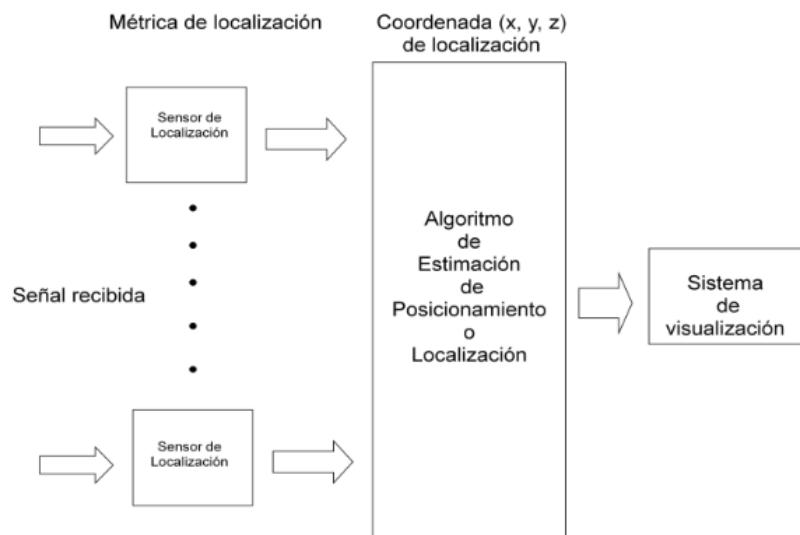


Figura 2.5: Diagrama funcional de un sistema de posicionamiento. Traducido de Pahlavan y cols. (2002).

Como se puede observar en la Fig. 2.5, los elementos iniciales del sistema, los cuales se encargan de capturar las señales emitidas por los dispositivos móviles, son los sensores. En líneas generales, las métricas relacionadas con la posición que se pueden medir a partir de la captura de señales son: el ángulo de llegada para los sistemas basados en dirección, el tiempo de llegada o retardo de la señal, la señal portadora de llegada y la potencia de la señal recibida, todas éstas usadas para la estimación de la distancia. A medida que las mediciones de las métricas se vuelven menos confiables, aumenta la complejidad del algoritmo de posicionamiento (Pahlavan y cols., 2002).

El segundo elemento de la Fig. 2.5 hace referencia al algoritmo de estimación de posicionamiento o localización, el cual procesa la métrica de localización y estima la ubicación del dispositivo usando enfoques tales como el procesamiento de señales, el uso de enfoques probabilísticos, métodos basados en modelos físicos o modelos predictivos que usen algoritmos inteligentes para extraer patrones de movimiento. Finalmente, el sistema de visualización convierte la información procesada por el algoritmo de estimación de posicionamiento en un formato apropiado para el usuario final.

2.3. Mapas térmicos

Los mapas térmicos son representaciones gráficas de datos que utilizan sistemas codificados por colores. Tienen como propósito principal representar gráficamente la relación volumétrica que existe dentro de un conjunto de datos, siendo las áreas con mayor volumen de datos las más notorias (Goulart, 2014).

Debido al sistema de codificación de colores que los mapas térmicos usan para transmitir información, son altamente usados para representar grandes volúmenes de datos. Por lo general, este sistema de colores, va desde rojo intenso, pasando por naranja y amarillo hasta verde y distintas degradaciones de azul, siendo los colores cálidos (rojo, naranja, amarillo) los que representan mayores concentraciones de datos, y los colores fríos (verde, azul) menores concentraciones de datos (ver Fig. 2.6). Dependiendo del contexto, estos datos hacen referencia a concentraciones de clientes, usuarios, lugares más usados dentro de un sitio web, movimientos físicos, flujos de vientos térmicos, entre otros.

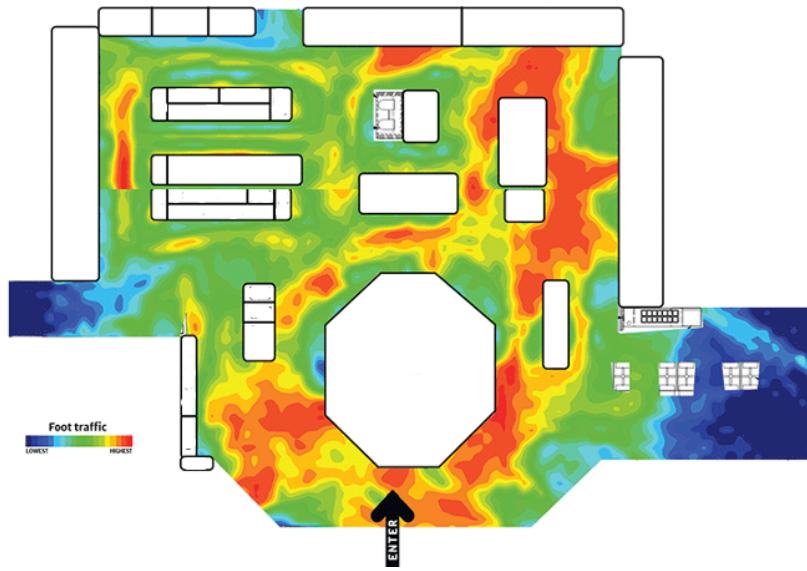


Figura 2.6: Ejemplo de mapa térmico. Tomado de Goulart (2014).

Según Zenko (2018) entre las aplicaciones que tienen los mapas térmicos están:

- Ubicación de multitudes de clientes dentro de centros comerciales.
- Determinar las áreas más visitadas y con mayor atención de los usuarios dentro de un sitio web.
- Para representar la densidad de una población, su ingreso per cápita, etc.
- En mapas geográficos para mostrar temperaturas promedios anuales.

2.4. Diagramas de Sankey

Un diagrama de flujo es la representación gráfica de un conjunto de relaciones dinámicas en un sistema. Los diagramas de flujo se utilizan para estructurar y ordenar un sistema complejo, o para revelar la estructura subyacente de los elementos, su interacción y como están relacionados entre ellos (Meyer y cols., 2006).

Los diagramas de Sankey son un tipo de diagrama de flujo en el que el ancho de las flechas es proporcional al caudal o cantidad de datos relacionados con ellas (ver Fig. 2.7). Originalmente los diagramas de Sankey representaban todos los flujos de energía

primaria en una fábrica, siendo el ancho de las flechas linealmente proporcional a la producción, utilización y pérdida de energía. Éstos diagramas enfatizan las principales transferencias o flujos dentro de un sistema ayudando a localizar los flujos de mayor caudal (Schmidt, 2006).

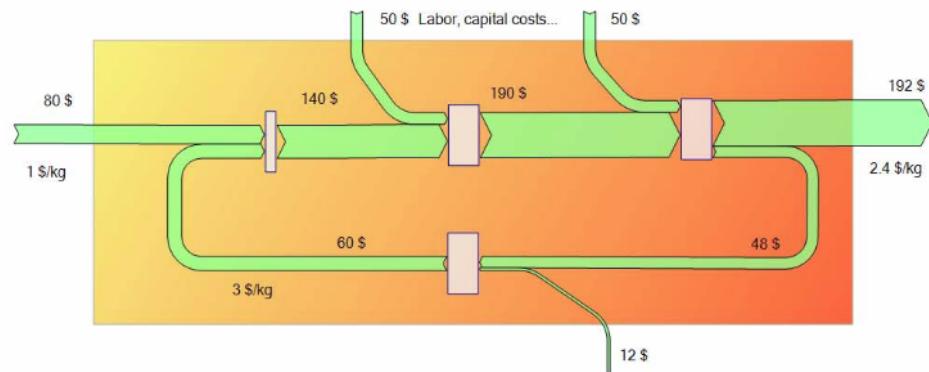


Figura 2.7: Ejemplo de diagrama de Sankey. Tomado de Schmidt (2006).

En la actualidad, además de representar flujos de energía dentro de una fábrica, los diagramas de Sankey son usados para representar el flujo de clientes en sitios web. Los analistas usan estos diagramas para comprender qué páginas pertenecientes a un sitio web obtienen más tráfico y cómo se mueven los usuarios en ellas (Schaffer, 2017).

2.5. Mercadotecnia

La mercadotecnia (del inglés, *Marketing*) se considera como “un conjunto de instituciones y procesos para crear, comunicar, entregar e intercambiar ofertas que tienen valor para los clientes, socios y la sociedad en general” (Kotler y cols., 2002).

Tiene como objetivo principal favorecer el intercambio de valor entre el comprador y el vendedor, de forma que ambas partes resulten beneficiadas, asegurando que dicho intercambio se repita en futuras ocasiones. Busca fidelizar clientes posicionando en sus pensamientos productos o marcas, con el fin de ser su opción principal al momento de que éste decida realizar una compra (Kotler y cols., 2002).

Además, la mercadotecnia no busca solamente favorecer las necesidades del cliente o consumidor y de la organización a corto plazo, busca que la organización sea percibida

en el mercado, con una visión de rentabilidad a largo plazo.

2.6. Análisis de datos

Según Runkler (2012) “El análisis de datos es un proceso de inspección, limpieza, transformación y modelado de datos con el objetivo de descubrir información útil, para sugerir conclusiones y apoyo en la toma de decisiones”.

El análisis busca dividir un todo en pequeños componentes para disminuir la complejidad del estudio. Es un proceso que a partir de datos sin procesar encuentra información útil para la toma de decisiones por parte de los usuarios. Los datos se recopilan y analizan para responder preguntas del tipo cómo y por qué sucedió un evento, probar hipótesis o refutar teorías (Careers, 2017).

El proceso de análisis de datos según Judd y cols. (2011) cuenta con 6 fases:

1. Recolección de datos: los datos se pueden capturar de diversas fuentes, como sensores en el entorno, cámaras, entrevistas, etc.
2. Procesamiento de datos: por lo general los datos recolectados no vienen de forma estructurada, por ende éstos en principio deben ser procesados u organizados para su posterior análisis.
3. Limpieza de datos: una vez procesados y organizados, los datos pueden estar incompletos, tener duplicados o errores. La limpieza de datos es el proceso de prevenir y corregir estos errores.
4. Análisis exploratorio de datos: es un primer paso para comprender los mensajes contenidos en los datos a partir de la generación de estadísticos básicos, este proceso puede generar nuevas propuestas de limpieza de datos o la necesidad de recolectar nuevos datos, por lo que el análisis de datos es un proceso iterativo.
5. Modelado y algoritmos: se busca desarrollar modelos que generen a partir de los datos, estadísticos más precisos para su comprensión.
6. Generación de productos: un producto de datos puede ser una aplicación informática donde se pueda visualizar la información generada a partir de los

datos analizados, por lo general esto se hace a través de gráficos de barras o histogramas.

2.7. Analítica de datos

La analítica de datos se define como la aplicación de sistemas informáticos para el análisis de grandes conjuntos de datos que brinden soporte a la toma de decisiones. Es un campo interdisciplinario que ha adoptado aspectos de disciplinas científicas como: estadística, reconocimiento de patrones, inteligencia computacional, aprendizaje automático e investigación operativa (Runkler, 2012).

Según Careers (2017) en contraste al análisis de datos, la analítica busca predecir el futuro en vez de explicar eventos pasados. Es la aplicación del razonamiento lógico y computacional a las partes o componentes obtenidos del análisis de datos y busca patrones para determinar qué información se puede generar a partir de ellos. Por ende, el análisis de datos es una parte de la analítica de datos.

La analítica de datos está conformada por dos clases: la analítica cualitativa que busca usar la experiencia o intuición más los resultados obtenidos de la etapa de análisis de datos y la analítica cuantitativa que a los resultados de la etapa de análisis de datos le aplica fórmulas y algoritmos, ambos para predecir eventos futuros (Runkler, 2012).

Por lo tanto, la analítica requiere el uso de varias herramientas cuantitativas, desde álgebra y cálculo, hasta estadísticas y economía, con implementaciones en varios lenguajes de programación y software. Requiere experiencia técnica, buen juicio, la capacidad de formular preguntas perspicaces y desplegar datos para responder a estas preguntas.

Capítulo 3

Equipo de adquisición de datos

En este capítulo, se pretende ahondar en los detalles técnicos, funcionamiento y limitaciones del Meshlium Xtreme Scanner 4G AP, equipo usado en esta investigación como escáner de redes inalámbricas wifi, para la detección de dispositivos móviles que cuentan con esta tecnología. Aparte, en la sección 3.3 se dan a conocer las experiencias prácticas obtenidas con el Meshlium a través de una serie de pruebas locales. Estas pruebas se realizaron con el fin de estudiar la calidad de los datos que captura dicho sensor, su susceptibilidad en presencia de interferencias y el comportamiento de los dispositivos móviles que realizan aleatorización de MACs.

3.1. Meshlium Xtreme Scanner 4G AP

El dispositivo Meshlium es una puerta de enlace para el IoT (por sus siglas en inglés, *Internet of Things*) fabricado por la empresa multinacional tecnológica Libelium (2019b). Posee 4 diferentes interfaces de radio:

1. Wifi a una frecuencia de 2.4GHZ para funcionar como AP.
2. Una entrada 4G / 3G / GPRS / GSM y 2 radios XBee / RF para conectarse a Internet.
3. Un módulo GPS para aplicaciones móviles y vehiculares.

4. Un módulo de configuración de las tarjetas en modo monitor: con esta configuración la tarjeta de red integrada en el Meshlium es capaz de capturar todos los tipos de paquetes que se están enviando a través de la red, de este modo le es posible detectar no sólo a los AP sino también a todo tipo de dispositivos que emitan señales wifi y Bluetooth.

La información recopilada por el Meshlium se puede enviar a Internet usando 4G, 3G, GPRS, GSM o conectado a un AP. La Fig. 3.1 ilustra el proceso de escaneo del Meshlium y las tecnologías que usa para conectarse a Internet.

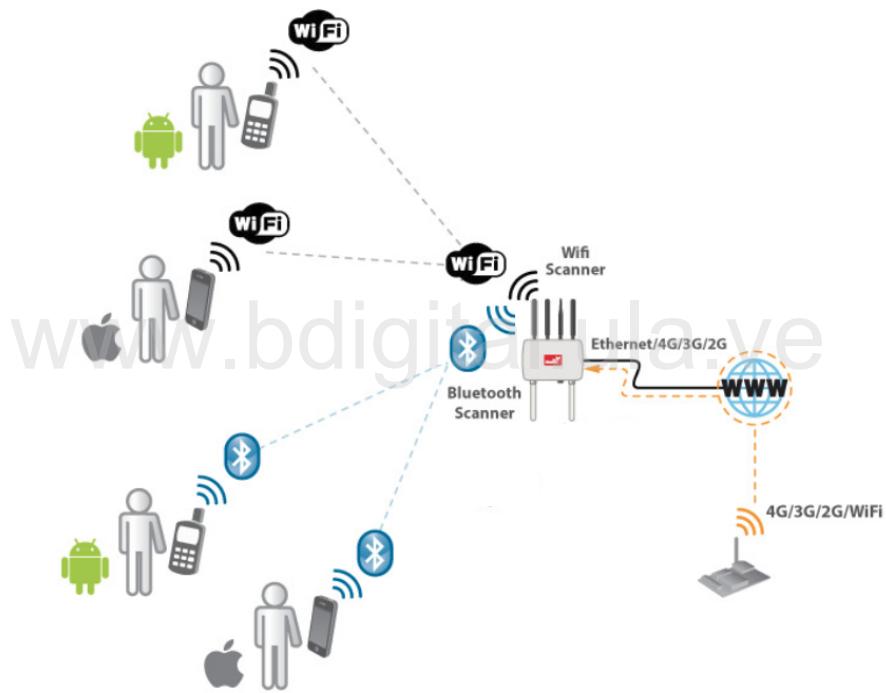


Figura 3.1: Meshlium Scanner 4G AP. Tomada de Libelium (2019b).

3.1.1. Meshlium Manager System V4.0

Es un sistema que le permite al usuario controlar la configuración del dispositivo de forma gráfica, según Libelium (2019b) los módulos de configuración con los que cuenta son:

1. Interfaces: este módulo permite acceder a las configuraciones de las interfaces de red como la interfaz de red alámbrica (Ethernet), la interfaz de red inalámbrica

- (wifi) para configurar el equipo Meshlium como AP, conexión del módem 4G y las configuraciones Proxy.
2. Red de sensores (del inglés, *Sensor Networks*): una de las principales funcionalidades del equipo Meshlium es funcionar como una puerta de enlace para redes de sensores inalámbricos (del inglés, *Wireless Sensor Networks*) del tipo Wasp mote de Libelium (ver Fig. 3.2). Este módulo es un entorno de desarrollo integrado (IDE) completo y de código abierto, que permite una forma fácil para comenzar a trabajar y configurar la red de sensores Wasp mote interconectados a través del equipo Meshlium.

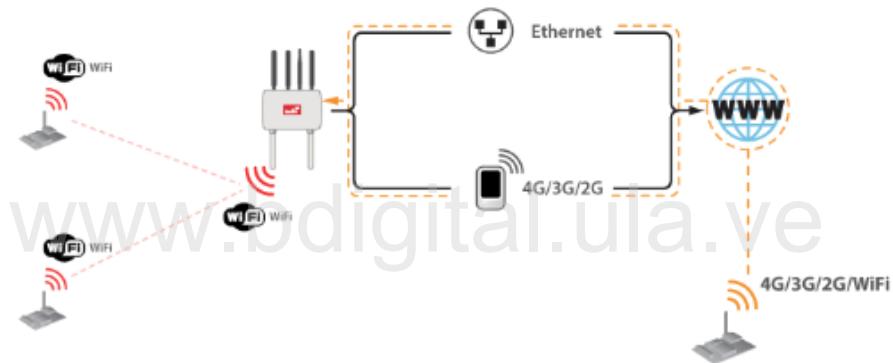


Figura 3.2: Sensores Wasp mote interconectados a través del equipo Meshlium como puerta de enlace. Tomada de Libelium (2019b).

3. Conectores a la nube (del inglés, *Cloud Connector*): uno de los más importantes servicios de la nube es que le permite a los desarrolladores correr aplicaciones, almacenar bases de datos, configurar y manejar recursos a distancia, de forma estable y con mejores prestaciones de las que puede tener en su hardware local. El equipo Meshlium posee un conjunto de scripts implementados para sincronizar la base de datos local a la “la nube”, en otras palabras, estos scripts envían los datos a los servidores web que le brindan servicios al usuario (ver Fig. 3.3). Además el equipo Meshlium ya está afiliado a un grupo de servidores en la nube por lo que puede brindar soporte para ellos, como son: Arrow, ElementBlue, Ericsson DMM,

Libelium Cloud Bridge, Telit, ThingWorx, Microsoft Azure, Ensura, infiswift, Ubicamovil, entre otros.

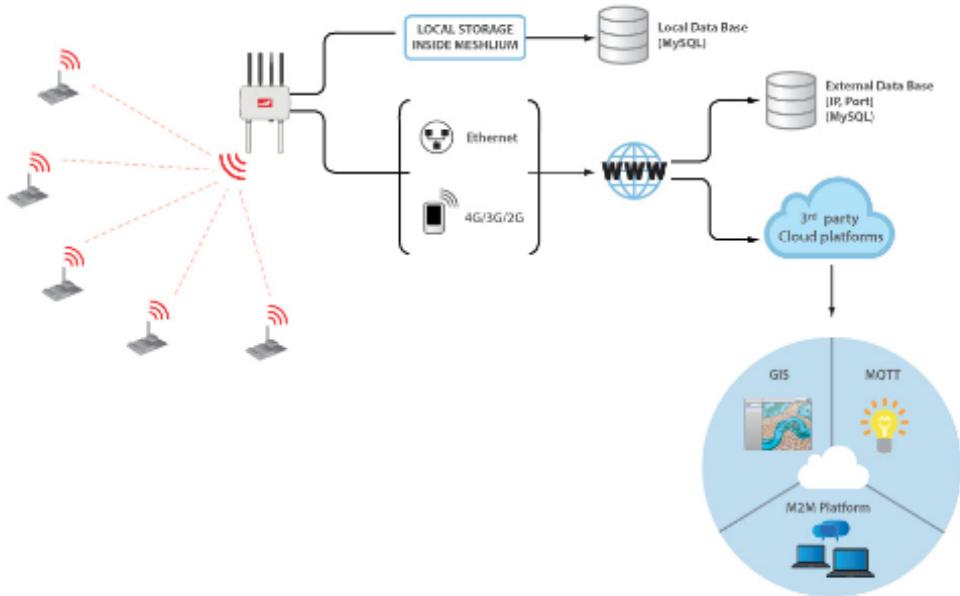


Figura 3.3: Diagrama para los conectores a la nube. Tomada de Libelium (2019b).

4. Herramientas: el equipo Meshlium trae integrado un conjunto de herramientas para la gestión de redes, que permiten verificar el correcto funcionamiento del mismo, entre éstos:

- Calculador de la zona Fresnel: la zona Fresnel es el espacio que debe estar libre de objetos entre la conexión punto a punto de dos dispositivos en redes inalámbricas.
- Iperf: esta herramienta permite ver el ancho de banda real entre el equipo Meshlium y un servidor.
- Ping y Traceroute: usados para verificar la conexión a un equipo en específico de la red.
- Neststat: descubrir que puertos IPv4 e IPv6 están activos.
- GPS: el equipo Meshlium puede integrar un receptor GPS que comparta su localización exacta en cualquier momento.

5. Sistema: configuraciones generales del sistema como nombre del equipo, configuración de los parámetros para el escáner wifi (tiempo de escaneo, bases de datos), cambio de clave, entre otros.
6. Manejador de actualizaciones: para instalar las nuevas versiones del Manager System.
7. Ayuda: despliega el manual del sistema.

La Fig. 3.4 muestra la página de inicio del administrador del sistema (del inglés, *Manager System*) integrado en los Meshliums.



Figura 3.4: Meshlium Manager System. Tomada de Libelium (2019b).

3.1.2. Detección de dispositivos móviles

Según Libelium (2019b) Meshlium permite detectar dispositivos iPhone y Android y en general cualquier dispositivo que funcione con wifi o Bluetooth. Estos dispositivos pueden detectarse sin necesidad de estar conectados a un punto de acceso específico, lo que le permite la detección de cualquier dispositivo móvil que ingrese al área de cobertura del Meshlium.

Los usuarios finales no deben hacer ninguna configuración en sus dispositivos para poder ser detectados. Cuando intentan conectarse a un punto de acceso o simplemente

tienen la antena wifi encendida, los Meshliums los capturan inmediatamente.

Para el monitoreo y captura de los datos en redes inalámbricas wifi, los Meshliums usan el software de código abierto Airodump-ng, diseñado para la auditoría y seguridad inalámbrica (ver Aspyct.org (2019)). Cabe destacar, que los dispositivos Meshliums no realizan ningún tipo de preprocesamiento de los datos que capturan antes de subirlos a la base de datos.

La Fig. 3.5 muestra un extracto tomado de una base de datos escaneada por un equipo Meshlium, el contenido de cada campo contiene lo siguiente:

- ID_frame: número de registro.
- Tiempo de escaneo (del inglés, *TimeStamp*): tiempo en el que se subieron los datos.
- Dirección MAC (ver sección 2.1.3).
- SSID (del inglés, *Service Set Identifier*): identificador de paquetes de red inalámbricos.
- Potencia de la señal recibida RSSI (ver sección 2.1.5).
- Vendedor: fabricante del producto.
- Tipo: tipo A para puntos de acceso y tipo C para dispositivos móviles.
- AP: MAC del punto de acceso al que el dispositivo está conectado.
- MeshliumID: identificador del equipo Meshlium que captura la señal.

ID	Tiempo de escaneo	MAC	SSID	RSSI	Vendedor	Tipo	AP	MeshliumID
1	2019-11-11 18:01:02	00:18:E7:7A:5E:26		-80	Cameo Communications	C	84:16:F9:E6:E5:18	183580815
2	2019-11-11 18:01:02	F0:D7:AA:41:7A:5C		-62	Unknown	C	(not associated)	183580815
3	2019-11-11 18:01:02	D4:11:A3:9D:66:D7		-81	Unknown	C	64:66:B3:AB:6E:34	183580815
4	2019-11-11 18:01:02	00:15:6D:60:ED:7B	meshliuma6c8	-35	Ubiquiti Networks Inc.	A		183580815
5	2019-11-11 18:01:02	90:F6:52:3E:9C:EC	MeridaTech2	-48	TP-LINK TECHNOLOGIES	A		183580815
6	2019-11-11 18:01:02	64:66:B3:AB:6E:34	TP-LINK	-60	TP-LINK TECHNOLOGIES	A		183580815

Figura 3.5: Extracto de la base de datos capturada por un Meshlium.

Estos registros pueden ser almacenados en:

1. Base de datos local: siempre está activa en la configuración, el Meshlium permite ver, filtrar, borrar y sincronizar registros con bases de datos en la nube directamente desde el dispositivo.
2. Base de datos externa: los datos se sincronizan con una base de datos externa desde la base de datos local. El servicio de sincronización se ejecuta cada 60 segundos y sincroniza hasta 100 datos en cada ciclo comenzando por los datos más nuevos, ya que son más relevante para la toma de decisiones.

3.1.3. Valores de RSSI medidos por los Meshliums

En los chipsets Atheros, existe una conversión especial para traducir valores de RSSI a dBm. El rango dBm usando esta conversión es de -35 dBm al 100 % (nodos más cercanos) y -95 dBm al 0 % (los más lejanos). Los Meshliums cuentan con estos chipsets y según su manual técnico los valores que capturan están comprendidos en ese rango (Libelium, 2019b). Sin embargo, el valor RSSI no tiene una proporción directa con la distancia, y es una medida sumamente sensible a los efectos cambiantes del ambiente de propagación, existencia o inexistencia de línea de vista entre el Meshlium y los dispositivos móviles, e incluso el RSSI puede cambiar ligeramente para 2 teléfonos a la misma distancia y con las mismas condiciones del canal propagación entre los mismos.

3.2. Limitaciones de los equipos Meshliums

Para mantener una alta tasa de detección se deben respetar ciertas condiciones, en Libelium (2019b) se mencionan:

- Los dispositivos que se detecten deben estar a unos metros del escáner y deben permanecer algunos segundos en su zona de cobertura.
- La configuración predeterminada para el tiempo de escaneo del Meshlium es de 40 segundos más el tiempo de sincronizar la base de datos (alrededor de 1 segundo).

Este parámetro (tiempo de escaneo) puede ser cambiado en un rango de 20 a 90 segundos.

- Los dispositivos Android y iOS tienen una opción especial para desactivar la conexión wifi cuando el usuario bloquea la pantalla para ahorrar batería. Esto ocasiona que el dispositivo sea invisible en la mayoría de los casos para los Meshliums.
- Por lo general, los AP siempre se detectan, ya que transmiten su SSID. También se detectan SSID ocultos. Los únicos AP que pueden ser difíciles de detectar son los AP que no difunden su presencia. Estos AP solo pueden detectarse cuando hay tráfico de dispositivos conectados.
- El comportamiento individual de los dispositivos definirá si son detectables. Como regla general, todos los dispositivos que generen tráfico en la red serán detectados.
- Dispositivos con el wifi apagado nunca serán detectados.
- Los dispositivos Meshliums no proporcionan mecanismos para combatir la aleatorización de MACs. El Meshlium detectará la dirección aleatoria y le será posible rastrear el dispositivo durante el período en que la MAC cambie, que es normalmente hasta que se conecta a una red wifi (Libelium, 2019b).

3.3. Pruebas locales

Para evaluar la efectividad de ciertos métodos, primero se debe definir cómo evaluar la calidad de los datos, para ello se realizaron una serie de pruebas locales que permitieron medir el alcance, la robustez y la reproducibilidad de los datos capturados por los equipos Meshliums.

3.3.1. Estudio de la calidad del dato

Este experimento se llevo a cabo en un ambiente controlado, para disminuir la cantidad de dispositivos móviles externos que el equipo de adquisición de datos

Meshlium puede capturar por su amplio rango de cobertura (200 m con línea de vista). De esta forma, se garantizan conclusiones veraces con respecto a la estabilidad de las medidas capturadas por el Meshlium (como la variabilidad del RSSI) y la frecuencia de solicitudes de sondeo de dispositivos móviles bajo condiciones predefinidas.

Para la adecuación del escenario de prueba se marcaron 6 puntos de referencia, cada punto a una distancia de 5 m, 10 m y 15 m, con y sin línea de vista (ver Fig. 3.6). El ciclo de prueba por punto de referencia es de 5 minutos.

Para medir la cantidad de solicitudes de sondeo, se plantea un modelo de interacción con las pantallas de los dispositivos de prueba para contrastar un incremento o disminución de las mismas. Para un ciclo, sí y solo sí la configuración deseada es “con interacción”, se simula el uso del dispositivo móvil al cerrar y abrir aplicaciones, en caso contrario el dispositivo se mantiene en reposo o “sin interacción”. Este modelo de interacción también se propuso cuando los dispositivos están o no asociados a un punto de acceso, con la idea comprobar si las solicitudes de sondeo en cada configuración son consistentes.

3.3.1.1. Descripción del escenario de prueba

- Lugar: Centro Comercial Alto Prado, Mérida-Venezuela.
- Duración: 2 h.
- Configuración del Meshlium:
 - Tiempo de escaneo: 40 s.
 - Almacenamiento: bases de datos local.
 - Filtro por tipo: ninguno.
- Dispositivos móviles de prueba:
 - Motorola MotoG5, con sistema operativo Android 8.1.
 - Xiaomi Redmi Note 5, con sistema operativo Android 9.
- Parámetros del escenario de prueba (ver Tabla 3.1):

- Estado de asociación del dispositivo móvil a un punto de acceso wifi: asociado, no asociado.
 - Nivel de interacción del usuario con la pantalla del dispositivo móvil: con interacción, sin interacción.
 - Distancia del dispositivo móvil al Meshlium: 5 m, 10 m y 15 m.
 - Nivel de obstrucciones en la línea de vista del Meshlium y el dispositivo móvil: con línea de vista, sin línea de vista.
 - Tiempo de permanencia en cada punto de referencia: 5 min.

Tabla 3.1: Escenario de prueba.

Caso	Estado del wifi	Interacción	Distancia (m)	línea de vista
1	Asociado	Con interacción	5, 10, 15	Con línea de vista
2	Asociado	Sin interacción	5, 10, 15	Con línea de vista
3	No asociado	Con interacción	5, 10, 15	Con línea de vista
4	No asociado	Sin interacción	5, 10, 15	Con línea de vista
5	Asociado	Con interacción	5, 10, 15	Sin línea de vista
6	Asociado	Sin interacción	5, 10, 15	Sin línea de vista
7	No asociado	Con interacción	5, 10, 15	Sin línea de vista
8	No asociado	Sin interacción	5, 10, 15	Sin línea de vista

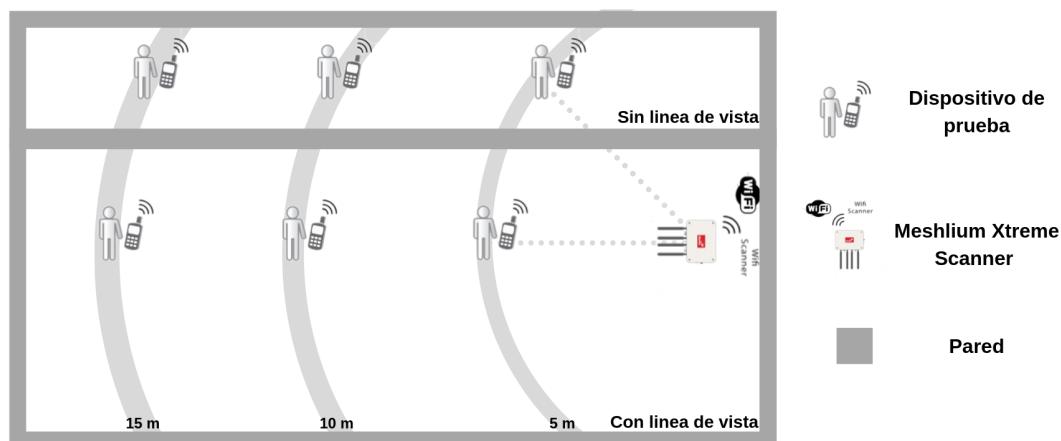


Figura 3.6: Escenario de prueba para el estudio de la calidad del dato.

3.3.1.2. Conclusiones y resultados

Frecuencia de las solicitudes de sondeo

La posición de un usuario dentro de un ambiente interior puede variar constantemente, si se desea hacer seguimiento a un dispositivo móvil es importante determinar la frecuencia en la que el Meshlium lo captura. Por consiguiente medir la frecuencia de solicitudes de sondeo que emiten los móviles da una idea clara de las condiciones que hacen posible hacerle seguimiento a los usuarios.

En el experimento expuesto, se espera que el Meshlium capture como máximo para cada ciclo de prueba 30 solicitudes de sondeo (1 solicitud por dispositivo cada 20 s por 5 min). La cantidad total de solicitudes de sondeo medidas por cada caso de prueba (ver Tabla 3.1) se exponen en el histograma de la Fig. 3.7, donde se observan los siguientes resultados:

1. La solicitudes de sondeo disminuyen claramente en el caso 2 respecto al caso 1, esto debido al estado sin interacción de los dispositivos en el caso 2.
2. Para los casos 3, 4, 7 y 8 los dispositivos móviles se encuentran en un estado no asociado a un punto de acceso, por ende se observan escasas solicitudes de sondeo capturadas por el Meshlium. La causa principal de este comportamiento es atribuida a la aleatorización de MAC que realizan los dispositivos móviles (ver sección 2.1.4), es decir, mientras éstos no reconocen a un punto de acceso como “confiable” (establecen conexión), por lo general no añaden su MAC real a la trama de la solicitud de sondeo, por ende, son capturados escasamente.
3. Los dispositivos con línea de vista directa al Meshlium son detectados con mayor probabilidad, es decir, ya que las obstrucciones pueden causar que las tramas emitidas por los móviles no sean capturadas por el Meshlium, en la mayoría de los casos con linea de vista (casos 1, 2 , 3 y 4) se capturan más solicitudes de sondeo.

Ya que el estado de asociación e interacción del usuario al dispositivo móvil puede aumentar o disminuir la frecuencia de solicitudes de sondeo, la detección de dispositivos depende en gran medida de su contexto ambiental.

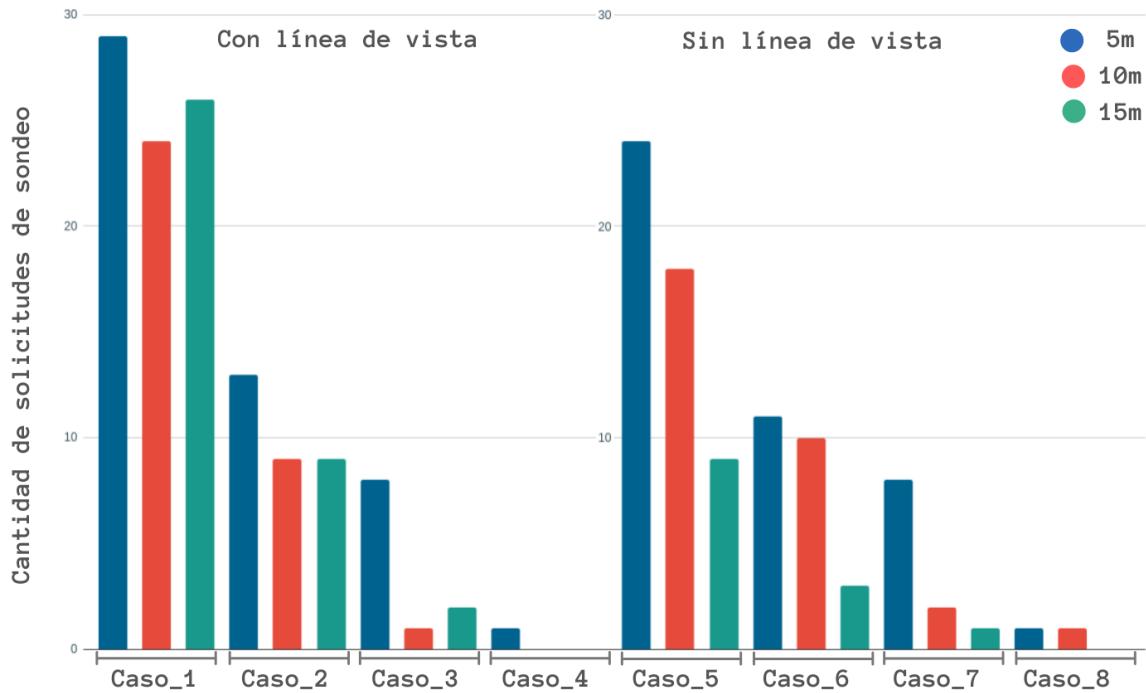


Figura 3.7: Frecuencia de las solicitudes de sondeo para los casos de estudio.

Estudio del RSSI

Para determinar si el RSSI es una medida confiable para calcular la distancia de los dispositivos móviles respecto al Meshlium, se realizó un análisis exhaustivo de la calidad, estabilidad y reproducibilidad en la captura de este dato.

Las mejores condiciones para evaluar la estabilidad de este dato, son presentadas cuando el dispositivo móvil está asociado a un punto de acceso, ya que de esta forma los móviles emiten una mayor cantidad de solicitudes de sondeo con su MAC real (ver Fig. 3.7). Estas condiciones se presentan en los casos 1, 2, 5 y 6 y a partir del estudio del RSSI capturado por los Meshlium se presentan los siguientes resultados:

1. Las medias de RSSI (líneas punteadas en las figuras 3.8(a) y 3.8(b)) son menores en el caso 2 con respecto al caso 1 por alrededor de -3 dBm. En este caso, cuando los dispositivos cuentan con linea de vista hacia el Meshlium, la media de RSSI entre los estados con interacción y sin interacción varía muy poco.
2. En los casos 2 y 6 (ver figuras 3.8(b) y 3.9(b)) se perciben menor cantidad de muestras con respecto a los casos 1 y 5 respectivamente (ver figuras 3.8(a)

y 3.9(a)). La razón de este comportamiento para cada caso es el estado de interacción del usuario hacia los dispositivos móviles.

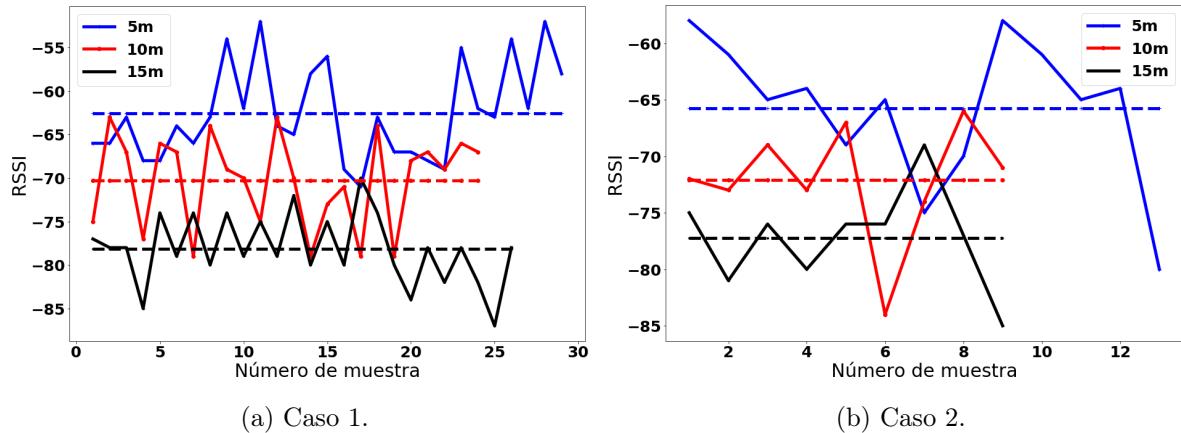


Figura 3.8: RSSI para diferentes posiciones con línea de vista.

3. La disminución de la cantidad de muestras en el caso 5 respecto al caso 1 (ver figuras 3.8(a) y 3.9(a)), es debida a la falta de linea de vista de los dispositivos móviles con el Meshlium.
4. Las medias de RSSI (lineas punteadas en las figuras 3.9(a) y 3.9(b)) son menores en el caso 5 con respecto al caso 6 por alrededor de -10 dBm. En este caso, cuando los dispositivos no cuentan con linea de vista hacia el Meshlium, la media de RSSI entre los estados con interacción y sin interacción tiene un amplio rango de variabilidad.
5. Aunque el rango de cobertura del RSSI es amplio (de -35 dBm a -90 dBm), en la Fig. 3.8 para una distancia de 5 m, se puede observar que el RSSI toma valores cercanos a -65 dBm. Es de esperar, que las medidas alrededor de -35 dBm sean capturadas para dispositivos muy cercanos al Meshlium (de 1 m a 3 m), lo que no sucede en un ambiente de producción real.

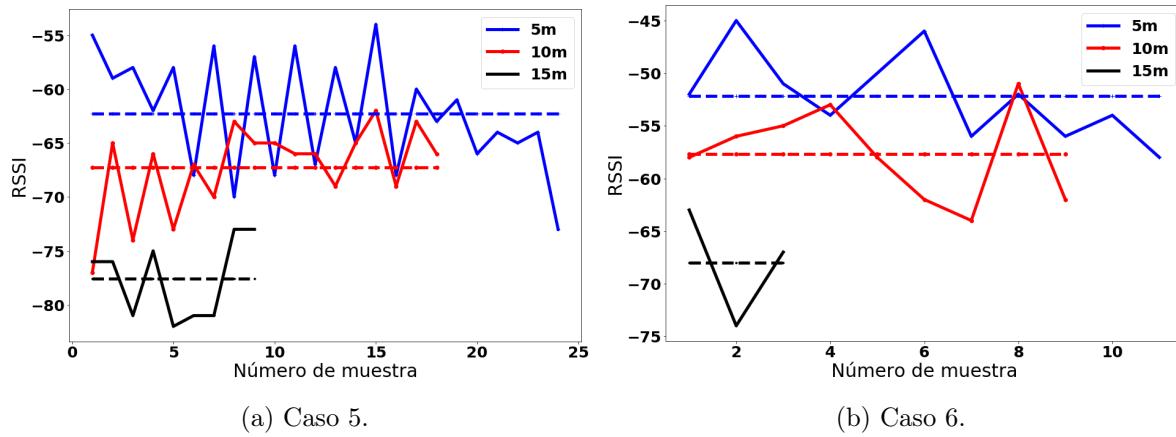


Figura 3.9: RSSI para diferentes posiciones sin línea de vista.

Con respecto a la variabilidad del RSSI en la Fig. 3.10 se muestran una serie de diagramas de cajas donde se percibe con exactitud la dispersión de los datos capturados, a partir de su análisis se concluye:

1. Por lo general los valores de RSSI no se solapan a partir de una distancia de 10 m, por lo que el radio de cobertura para la ubicación de dispositivos móviles puede ser de este valor.
2. A menor distancia (5 m) se percibe un mayor rango de variabilidad de los datos. El peor de los casos (ver Fig. 3.10(b)) presenta un rango de variabilidad hasta de -25 dBm.
3. A mayor distancia (15 m) se percibe un menor rango de variabilidad y una mayor concentración de los datos cerca de la mediana.
4. Se detecta una clara correlación del RSSI con la distancia (5 m, 10 m, 15 m), con pocos valores atípicos (fuera de los bigotes).
5. El caso ideal de variabilidad en el RSSI es cuando éste abarca un rango de -1 dBm a -3 dBm. Ninguno de los diagramas expuestos en la Fig. 3.10 tienen este comportamiento, por ende el RSSI no es una medida confiable para ser usado como referencia de la ubicación de dispositivos móviles.

6. Por su alto rango de variabilidad (incluso cuando los dispositivos están fijos y en condiciones ideales) los valores capturados no son reproducibles, es decir, al repetir el experimento con los mismos parámetros y bajo las mismas condiciones las medidas capturadas por el Meshlium serán diferentes.

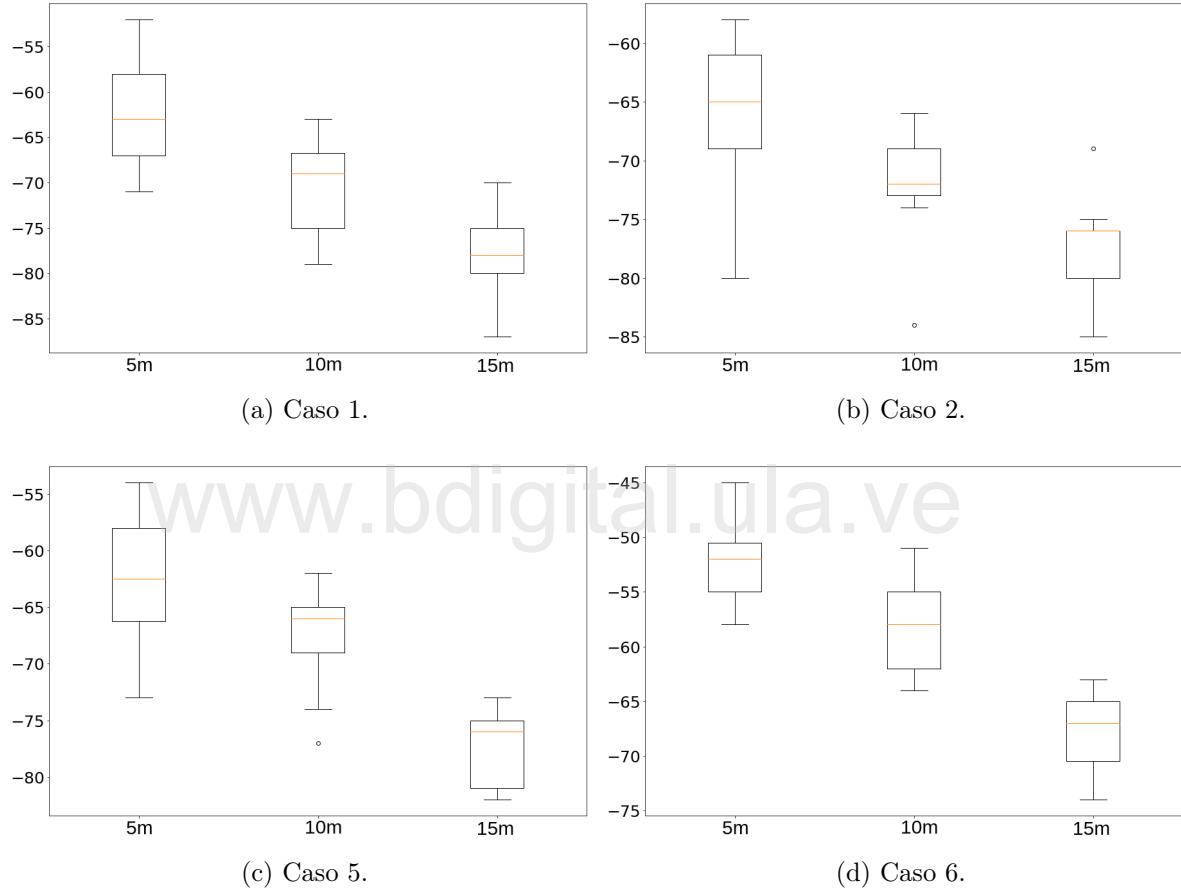


Figura 3.10: Representación en diagramas de cajas del RSSI.

3.3.2. Estudio de la aleatorización

El objetivo de este experimento es mostrar cómo el conteo de dispositivos móviles se ve afectado por la aleatorización de MACs y la frecuencia de solicitudes de sondeo de cada uno de ellos, usando una cantidad limitada de dispositivos para resaltar cómo un pequeño subconjunto puede influir en la precisión de la estimación.

Siguiendo la metodología presentada por Rutermann y cols. (2019), cada 5 minutos 2 dispositivos móviles entran al escenario de prueba al activar su antena wifi. Si los dispositivos tienen registrados los datos de conexión al punto de acceso, se conectan inmediatamente, bajo esta configuración no realizan aleatorización de MACs, por lo tanto, cada cierto tiempo son escaneados por el Meshlium con su MAC real. De forma contraria sucede si los dispositivos no tienen registrados los datos de conexión al punto de acceso, lo que ocasiona que cada vez que sean escaneados por el Meshlium realicen aleatorización de MACs.

Como en el experimento 3.3.1 para medir la cantidad de solicitudes de sondeo, se plantea un modelo de interacción con la pantalla del dispositivo móvil que permita contrastar un incremento o disminución de las mismas. En este experimento, al iniciar un ciclo de entrada de un celular al escenario de prueba, sí y solo sí la configuración deseada para la muestra es “con interacción”, se simula el uso del dispositivo móvil al cerrar y abrir aplicaciones, durante 1 minuto de los 5 minutos de duración del ciclo de prueba, los 4 minutos restantes el dispositivo permanecerá en modo reposo o “sin interacción”. Esta es la representación de un escenario real, cuando un usuario que visita una tienda, revisa el celular por una fracción de tiempo y luego vuelve a dejarlo sin uso.

Después de que la cantidad total de dispositivos móviles se encuentren dentro del escenario de prueba, cada 5 minutos 2 dispositivos son retirados del mismo al desactivar su antena wifi, lo que hace que sean “invisibles” para el Meshlium, este proceso se repite hasta que no quede ningún dispositivo activo.

3.3.2.1. Descripción del escenario de prueba

- Lugar: Centro Comercial Alto Prado, Mérida-Venezuela.
- Duración: 1 h.
- Configuración del Meshlium:
 - Tiempo de escaneo: 20 s.
 - Almacenamiento: bases de datos local.

- Filtro por tipo: ninguno.
- Dispositivos de prueba:
 - Motorola MotoG5, con sistema operativo Android 8.1.
 - Motorola MotoG7, con sistema operativo Android 9.
 - Xiaomi Redmi Note 5, con sistema operativo Android 9 (2 dispositivos).
 - Xiaomi Redmi Note 7, con sistema operativo Android 9.
 - iPhone 6, con sistema operativo iOS 12.4.1.
- Con línea de vista.
- Parámetros del escenario de prueba (ver Tabla 3.2):
 - Estado de asociación del dispositivo móvil a un punto de acceso wifi: asociado, no asociado.
 - Nivel de interacción del usuario con la pantalla del dispositivo móvil: con interacción (durante el primer minuto de inicio del ciclo correspondiente), sin interacción.
 - Distancia para las medidas (ver Fig. 3.11): 5 m.
 - Eventos de entrada y salida de dispositivos al área de monitoreo: cada 5 min.

Tabla 3.2: Escenario de prueba.

Medida	Estado del wifi	Interacción
1	Asociado	Con interacción
2	No asociado	Con interacción
3	Asociado	Sin interacción
4	No asociado	Sin interacción

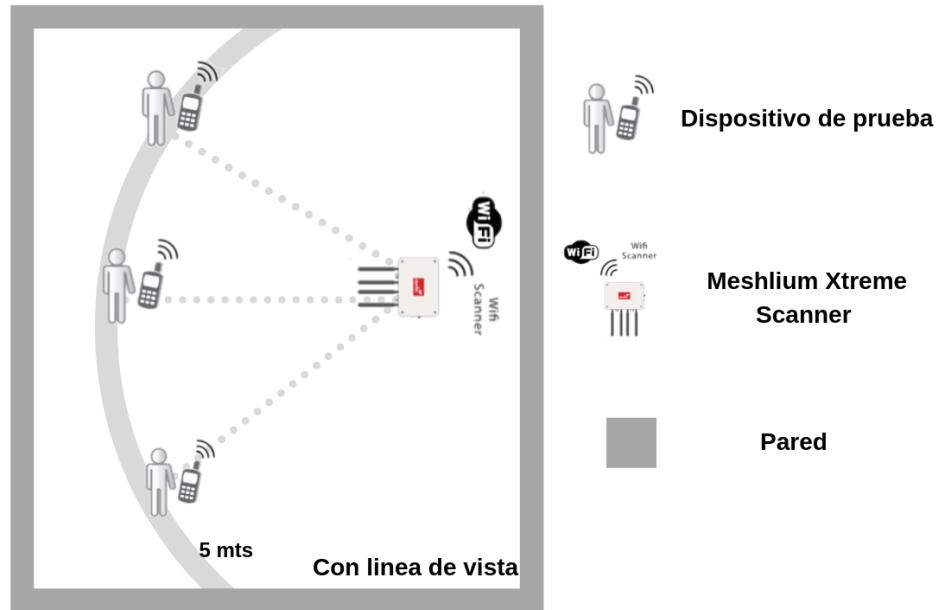


Figura 3.11: Escenario de prueba para el estudio de la aleatorización de MACs.

3.3.2.2. Conclusiones y resultados

Tomando como modelo las gráficas expuestas por Rutermann y cols. (2019) en sus resultados, las figuras 3.12 y 3.13 muestran una relación entre la cantidad de dispositivos reales (línea base en verde) y las solicitudes de sondeo capturadas por el Meshlium (en azul).

La Fig. 3.12(a) presenta el caso ideal para el estudio (dispositivos asociados a un punto de acceso y con interacción), como se observa, la cantidad de solicitudes de sondeo son prácticamente iguales a la cantidad de dispositivos por ciclo del experimento.

A pesar que en la Fig. 3.12(b) los dispositivos se encuentran asociados a un punto de acceso, su estado sin interacción hace que a diferencia del caso anterior, sólo en algunos momentos, la cantidad de solicitudes de sondeo sea igual a la cantidad de dispositivos móviles que se encuentran activos en el experimento. Esto ocasiona que el Meshlium no detecte la cantidad real de dispositivos móviles; un ejemplo claro se observa alrededor del minuto 12 de dicha gráfica, donde el Meshlium sólo detecta 2 dispositivos móviles cuando en realidad hay 4.

Además, para ambos casos de la Fig. 3.12, se observa un pequeño retraso entre el

tiempo que le toma al Meshlium detectar y dejar de detectar a los móviles, este retraso es consistente durante la duración del ciclo de prueba.

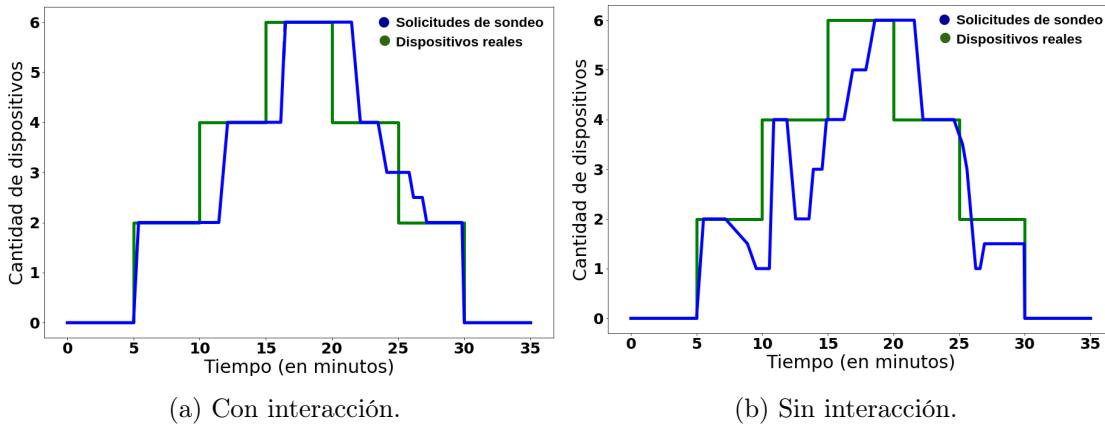


Figura 3.12: Solicitudes de sonda para el estado asociado.

En la Fig. 3.13 se muestra el comportamiento que tienen los dispositivos móviles al no estar conectados a un punto de acceso, esto hace que incrementen de forma extraordinaria la cantidad de solicitudes de sondeo que envían. Como se observa, para ambos casos (con interacción y sin interacción) las solicitudes de sondeo emitidas por los móviles con MACs falsas, sobrepasa la cantidad real de dispositivos de prueba; siendo este comportamiento más notable cuando los dispositivos se encuentran con interacción.

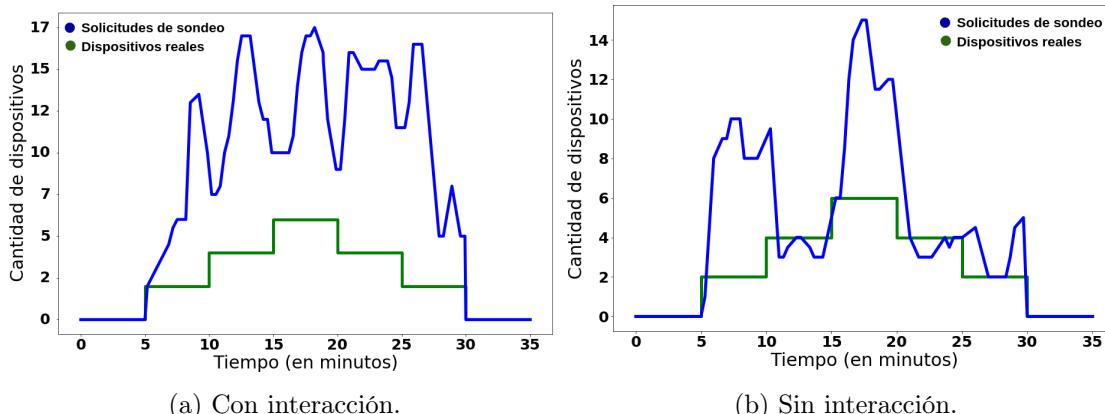


Figura 3.13: Solicitudes de sonda para el estado no asociado.

Es importante destacar que el comportamiento que tienen los dispositivos a la hora de aleatorizar sus MACs no se puede predecir y está definido por los fabricantes de dispositivos móviles, por lo que cambia entre un dispositivo a otro. Por ejemplo, los dispositivos Motorola sólo aleatorizan los últimos 3 octetos de sus MACs, mientras que los iPhone aleatorizan sus MACs completamente. También es aleatoria la cantidad de solicitudes de sondeo que envían durante un ciclo de escaneo; en un caso de estudio se detectaron hasta 8 solicitudes de sondeo enviadas por un dispositivo Motorola durante 20 s.

www.bdigital.ula.ve

C.C. Reconocimiento

Capítulo 4

Análisis de los datos capturados en un ambiente de producción real

En este capítulo, se busca comprender la naturaleza de los datos capturados por los Meshliums en un ambiente de producción real. En la sección 4.1 se expone una idea general sobre la distribución física del establecimiento donde se capturaron los datos para este estudio y la cantidad de dispositivos Meshliums que se usaron para el escaneo de dispositivos móviles. Además en las secciones 4.2 y 4.3 se detalla el tipo de preprocesamiento, análisis de datos y filtros, que fueron aplicados sobre la base de datos total capturada por los Meshliums, con el fin de obtener datos útiles capaces de generar analíticos.

4.1. Descripción del ambiente de producción real

Por razones de confidencialidad, no es posible mencionar el ambiente de producción real donde fueron instalados los 4 Meshliums que capturaron las bases de datos para este estudio. En lo sucesivo el entorno de estudio estará referido como Centro Comercial A (CCA), este cuenta con una superficie para el área de ventas de $5.000\ m^2$ y 16 routers inalámbricos que funcionan como puntos de acceso wifi cuyo rango de cobertura conjunto abarca la totalidad del espacio. La Fig. 4.1 muestra el plano del CCA, la ubicación de los 16 routers, los 4 Meshliums instalados y las zonas que representa cada

uno de ellos.

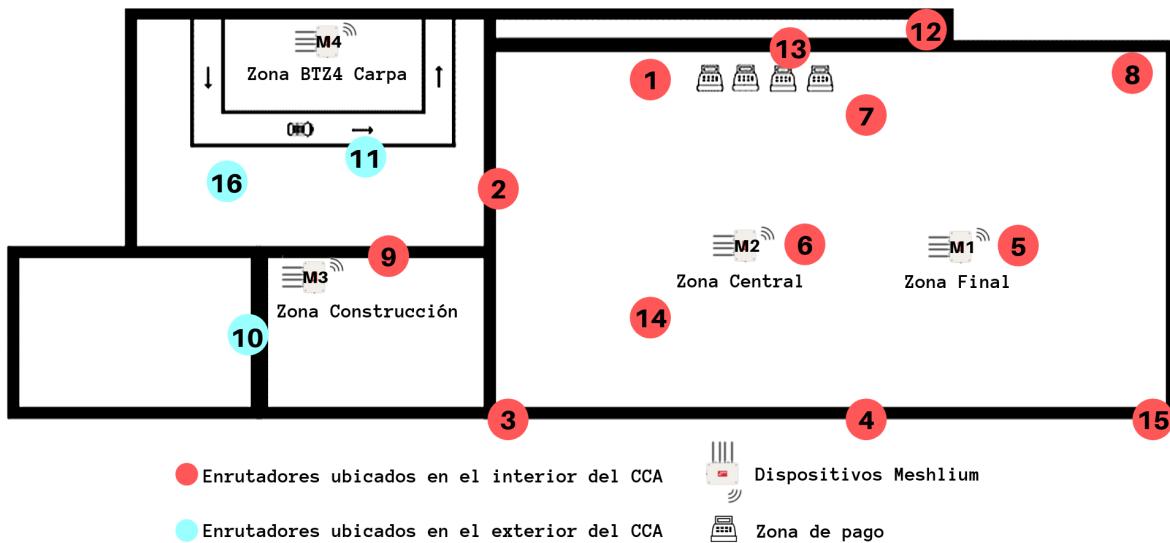


Figura 4.1: Plano y ubicación de los AP y sensores wifi en el CCA.

4.2. Análisis de datos

Para generar teorías y propuestas en base a los datos capturados, se realizó un análisis de los mismos, con el fin de extraer del total de las capturas de los Meshliums las relevantes para alcanzar los objetivos del proyecto.

4.2.1. Análisis exploratorio y visualización de datos

El primer paso fue observar la integridad de los datos capturados por los equipos Meshliums para tomar decisiones sobre los posteriores filtros que serán aplicados sobre éstos. Como el análisis de datos es un proceso iterativo, se realizó un programa para generar un reporte que permita visualizar el comportamiento de los datos, sobre cualquier base de datos escaneada en el entorno de estudio. Este reporte tiene el fin de presentar los datos para su posterior análisis de forma clara y automática.

4.2.1.1. Reporte de estabilidad de datos

El diseño del programa para la generación del reporte, como diagrama entrada-proceso-salida, se muestra en la Fig. 4.2, donde es notable la existencia de tres subprocesos principales encargados del procesamiento de los datos. El proceso principal denominado en el diagrama “Generar reporte de estabilidad de los datos capturados” cuenta con tres módulos, el primero crea todas las gráficas y métricas necesarias que muestran el comportamiento de los datos, el segundo inserta todos los datos e imágenes procesadas dentro de un documento, en este caso una plantilla de Latex y el tercero compila esta plantilla y produce el reporte en formato pdf para su posterior análisis. Estos módulos fueron programados usando el lenguaje de programación Python y su librería Jinja2.

Cada reporte que se genera (normalmente con los datos escaneados de una semana en el CCA) cuenta con 4 capítulos, 287 imágenes y una tabla, todo esto en 80 páginas de una presentación en formato pdf.



Figura 4.2: Diagrama entrada-proceso-salida: Programa para la generación del reporte de estabilidad de datos.

Los capítulos que contiene el reporte se explican a continuación:

- Capítulo 1. Resumen: el objetivo de este capítulo es mostrar que tan constante es la captura de los datos por parte de los Meshliums. Para ello se realizó una gráfica que muestra el tiempo en el que permanecen activos cada uno de los equipos.

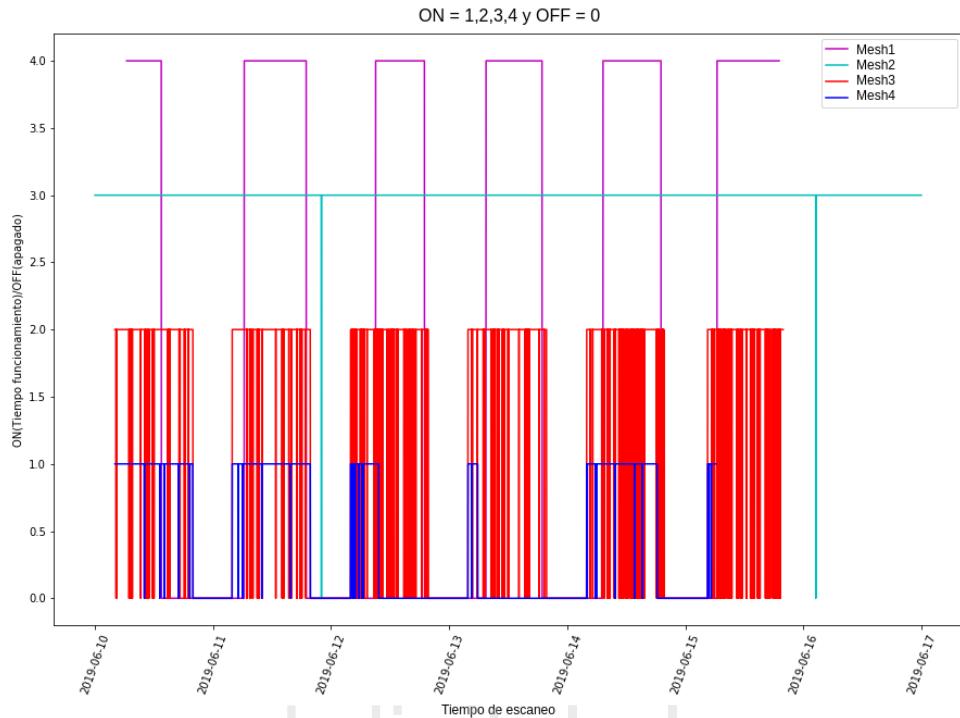


Figura 4.3: Tiempo total de adquisición de datos por Meshlium.

Este estudio es importante ya que para generar analíticos veraces se necesita que la captura de los datos por parte de los Meshlium sea constante y estable. De forma contraria las conclusiones a partir de los analíticos estarían sesgadas por los Meshlium que si capturen datos.

Para determinar el tiempo de funcionamiento de los dispositivos se planteo un modelo en donde cada Meshlium es representado con un número entero mayor a cero (1, 2, 3 y 4) para indicar que el equipo esta encendido o en funcionamiento. En caso contrario el valor del Meshlium se representa con cero.

Un comportamiento ideal se detecta para la gráfica azul turquesa de la Fig. 4.3, esta gráfica corresponde al Meshlium número 2. Como se observa, éste Meshlium trabajó de forma constante durante toda la semana, excepto por dos momentos (el 12 y 16 de junio) donde este equipo dejó de capturar datos.

Para los demás Meshlium en la Fig. 4.3, se detectan patrones de apagado de los

equipos en horas de la noche. Además, se percibe inestabilidad en la captura de datos para los Meshliums representados en la gráfica con los colores rojo y azul oscuro (Mesh3 y Mesh4), donde se observan intervalos de tiempo constantes en los que estos dispositivos dejan de adquirir datos sin explicación.

- Capítulo 2. Clasificación de las MACs según el tiempo de permanencia: cada vez que se captura una MAC se registra el tiempo en el que fue escaneada, en función de estos tiempos se puede determinar el tiempo de permanencia de un dispositivo dentro del CCA. Libelium (2019b) propone la siguiente clasificación de los dispositivos según su tiempo de permanencia:

Tabla 4.1: Clasificación de las MACs según el tiempo de permanencia.

Tiempo	Clasificación	Clasificación
$t < 30$ s	Visitante fugaz	Dispositivos capturados sólo una vez.
30 s $< t < 1$ min	Tasa de abandono	Dispositivos que pasan alrededor del CCA.
1 min $< t < 90$ min	Visitantes	Clientes en la tienda.
> 90 min	Tasa de olvido	Puntos de acceso o empleados.

Como el tipo en la base de datos indica si el dispositivo es un AP o un móvil, se agregó un nuevo clasificador denominado “punto de acceso” con el fin de compararlo con la cantidad de dispositivos pertenecientes a la tasa de olvido.

En las figuras 4.4(a) y 4.4(b) se detallan dos histogramas que indican la cantidad de MACs únicas capturadas y pertenecientes a cada uno de estos conjuntos.

Las cantidades de MACs únicas identificadas como Visitante fugaz en la Fig. 4.4(a), son debidas a la aleatorización de MACs en los dispositivos móviles. Cada dispositivo móvil que no esté asociado a un punto de acceso wifi en el ambiente de producción real, cambia dinámicamente su MAC n-veces por periodo de escaneo del Meshlium. Durante este proceso existen pocas probabilidades de que una MAC se repita, de este modo, un dispositivo móvil con MAC aleatorizada tiene una relación uno muchos (un dispositivo muchas MAC). Es por ello que se detectan tantos valores asociados con este clasificador.

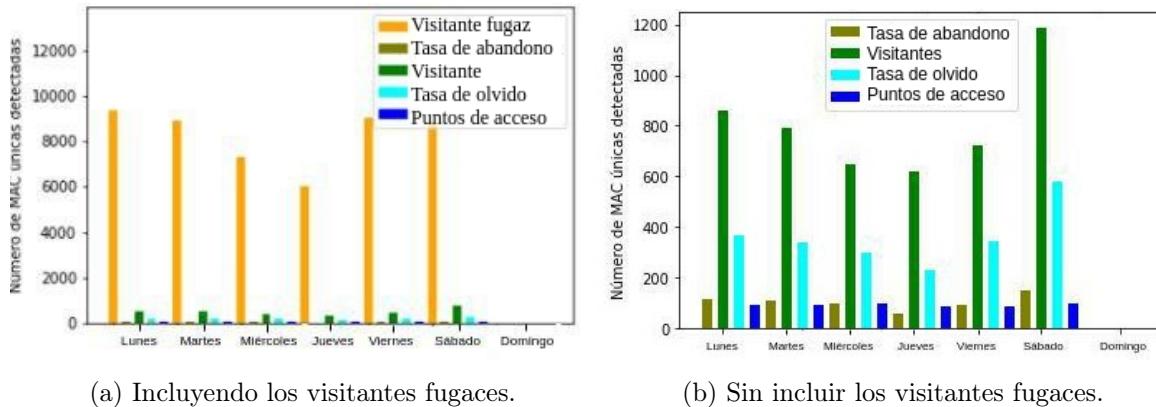


Figura 4.4: Clasificación de las MACs capturadas según su tiempo de permanencia.

Por otra parte, las cantidades de MACs únicas identificadas en la tasa de olvido y puntos de acceso en la Fig. 4.4(b), son consistentes a lo largo de la semana, es decir, son capturados aproximadamente el mismo número de veces.

- Capítulo 3. Estabilidad en la adquisición de datos: la idea principal de este capítulo es determinar si el tiempo de escaneo configurado en los equipos Meshliums es constante. Este estudio se hizo por inspección visual, en donde cada punto graficado corresponde al delta del TimeStamp capturado para la MAC más frecuente.

La Fig. 4.5 muestra un conjunto de gráficas que representan cada cuánto tiempo un Meshlium dentro del CCA adquiere datos. De ser totalmente estable la adquisición de datos, los tiempos de escaneo siempre deben ser iguales como en la Fig. 4.5(b). Esta figura muestra un tiempo de escaneo de 20s a 21s, el valor de 1s de más corresponde al tiempo que tarda un Meshlium en subir la base de datos local a la nube.

Cada pico que sobresale del cuadro rojo en la Fig. 4.5(a) representa un tiempo de escaneo entre muestras diferente al configurado en el equipo. En la mayoría de los casos, el Meshlium no realizó escaneos por lapsos de tiempo mayores al debido, representando inestabilidad en la adquisición de datos.

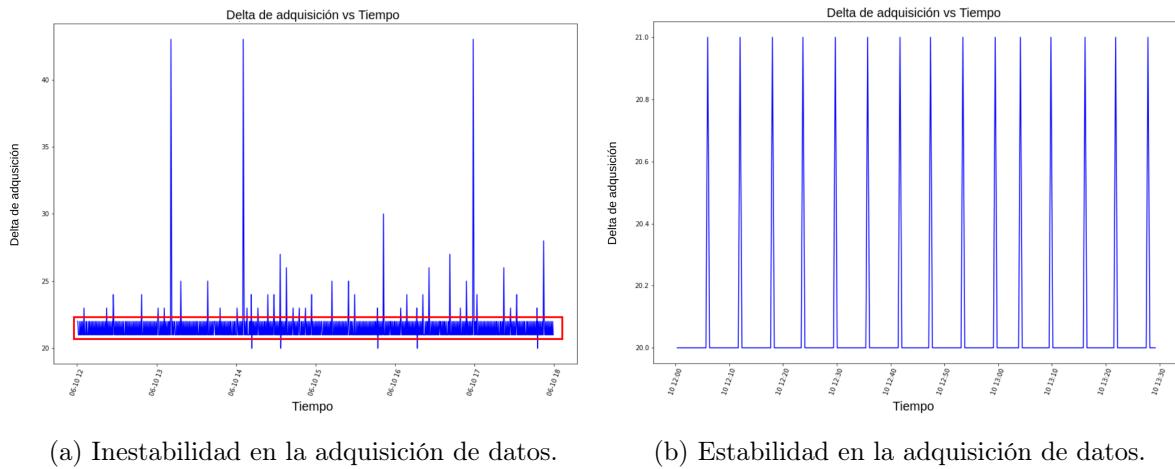


Figura 4.5: Estabilidad del tiempo de escaneo.

Otra representación de la estabilidad en la captura de los datos, se puede apreciar en la Tabla 4.2 donde se observan los tiempos de escaneo totales capturados y el porcentaje de aparición de cada uno de ellos con respecto al total de las capturas. De allí se concluye que el tiempo de escaneo configurado en los Meshliums es de 20s.

Los tiempos de escaneo de 21s y 19s son debidos a la demora o anticipación de los Meshlium para subir los datos a la base de datos externa.

Tabla 4.2: Resumen del tiempo de escaneo para los Meshliums.

Tiempo de escaneo (s)	Total de apariciones	Porcentaje (%)
20	11034	93.890
21	683	5.811
19	30	0.255
60626	1	0.008
50727	1	0.008
45017	1	0.008
44568	1	0.008
40993	1	0.008

- Capítulo 4. Cambios en el ambiente de propagación para dispositivos fijos: es de esperar, que para dispositivos cuya ubicación es fija como los puntos de acceso, la ausencia de comportamientos atípicos en la captura de datos, como una gran variabilidad de RSSI o valores de RSSI que se encuentren fuera del rango permitido.

Para detectar estos comportamientos, se hizo un estudio sobre los valores de RSSI capturados por los Meshliums y pertenecientes a la MAC más frecuente en la base de datos (esta MAC pertenece a un punto de acceso dentro del CCA). Una muestra de los resultados obtenidos se puede apreciar en las figuras 4.6(a) y 4.6(b) donde se reflejan los grados de varianza (menor y mayor respectivamente) de los valores de RSSI escaneados.

El mejor escenario posible para este estudio, se evidencia cuando los valores de RSSI presentan pequeños rangos de variabilidad, de éste modo el RSSI puede ser asociado como un valor aproximado de la distancia entre el dispositivo móvil hasta el Meshlium, en un rango de cobertura limitado.

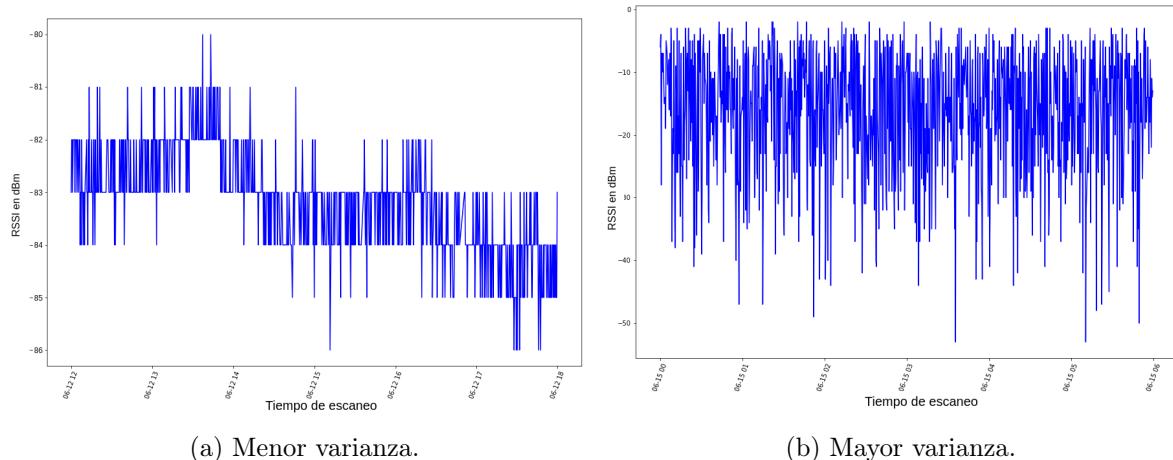


Figura 4.6: Variabilidad del RSSI capturado.

En la Fig. 4.6(a) podemos observar este comportamiento deseable del RSSI, presentándose para equipos cuya medida de este valor es baja (cercaos al límite inferior medido por los Meshliums -90dBm). Se concluye que todos los equipos

cuyos valores de RSSI son bajos, también tienen un grado de variabilidad de RSSI menor en todos los casos, estos valores bajos corresponden a equipos que se encuentran alejados de los Meshliums.

De forma contraria, el peor escenario posible se presenta cuando los valores de RSSI capturados tienen grandes rangos de variabilidad, este comportamiento se observa en la Fig. 4.6(b), donde el rango de variabilidad es hasta de -40dBm, además se perciben valores de RSSI mayores al límite superior permitido por los Meshliums (-35dBm) indicando altas tasas de saturación de los equipos. Se concluye que todos los equipos cuyos valores de RSSI son altos tienen un grado de variabilidad mayor en todos los casos, estos valores altos de RSSI pertenecen a equipos cercanos a los Meshliums.

Como se puede observar en la Fig. 4.7 los valores de RSSI promedio para MACs cuya ubicación es fija, comparten el mismo comportamiento explicado con anterioridad, es decir, las MACs con menores valores promedio de RSSI (ver Fig. 4.7(a)), presentan rangos de variabilidad pequeños (de -2 a -6 dBm) y las MACs con mayores valores promedio de RSSI (ver Fig. 4.7(b)), presentan altas tasas de variabilidad (hasta de -50 dBm).

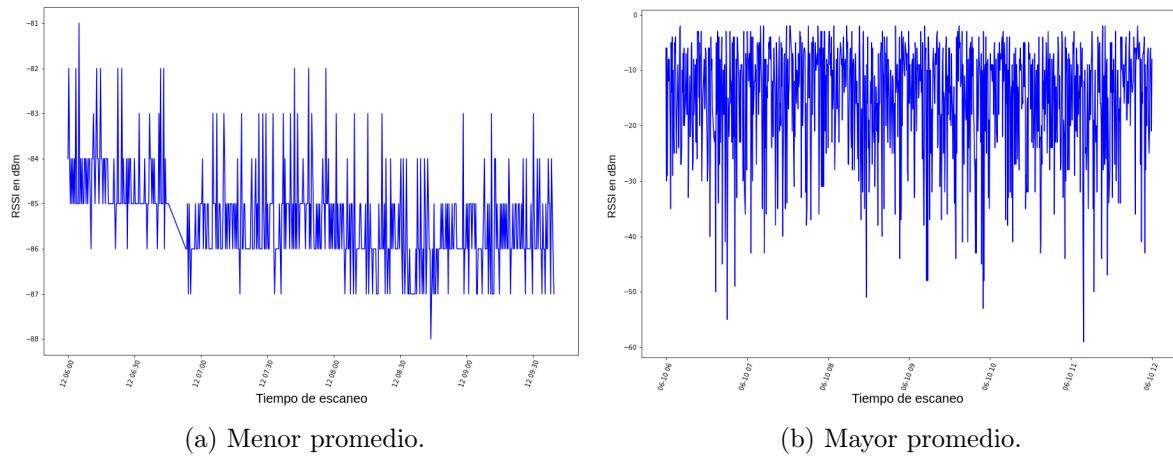


Figura 4.7: Promedio del RSSI capturado.

Este estudio da una idea del rango de cobertura que tienen los equipos Meshliums dentro del CCA. Para los Meshliums de la Zona Final y la Zona Central, los

equipos fijos que presentan una alta tasa de saturación y de variabilidad en la mayoría de los casos, son los puntos de acceso 5 y 6 respectivamente. Los equipos con menor variabilidad y que dan una idea del máximo rango de cobertura de los Meshliums son los puntos de acceso 12 y 3 respectivamente.

Sus ubicaciones se pueden apreciar en la Fig. 4.8, donde es notable la distancia de estos puntos de acceso con respecto a cada Meshlium; los routers con bajos valores de RSSI se representan con color azul cielo y los que tienen altos valores de RSSI con naranja. Es evidente, que los grados de saturación que presentan los routers 5 y 6 son debidos a la cercanía física que tienen éstos respecto a los Meshliums 1 y 2 respectivamente.

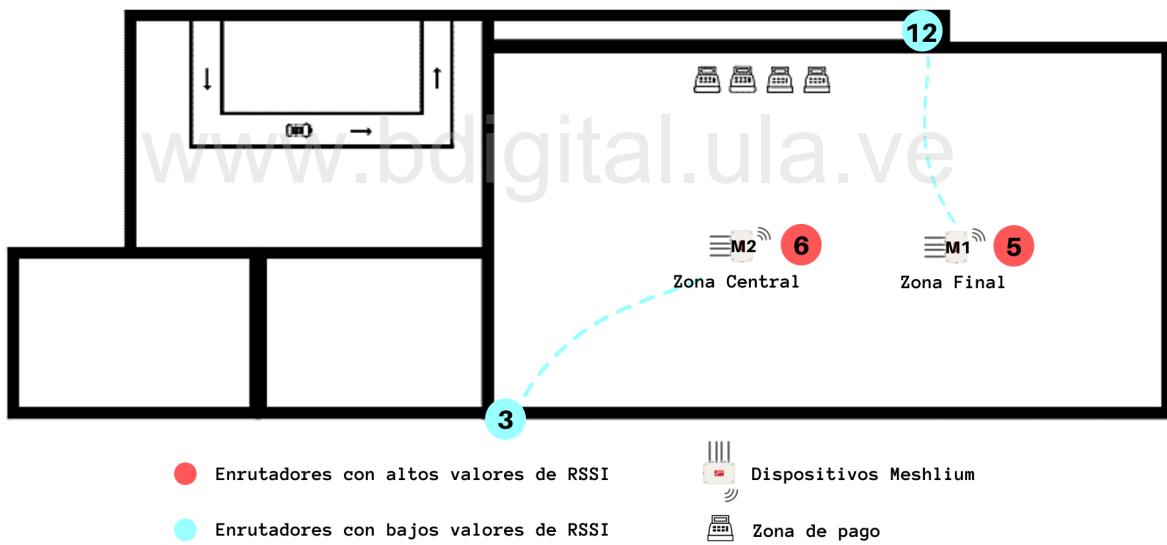


Figura 4.8: Representación del rango de cobertura de los equipos Meshliums.

Como de forma general los valores de RSSI captados tienen una alta tasa de variabilidad, incluso para dispositivos cuya ubicación es fija, este valor no es confiable y no se puede usar para estimar la distancia de los dispositivos móviles con respecto a los Meshliums.

4.3. Limpieza de datos

El segundo paso fue realizar un proceso de limpieza a los datos capturados, con el fin de eliminar datos incompletos, duplicados y errores.

4.3.1. Detección de valores atípicos

Después de observar los registros capturados se encontraron valores “atípicos” respecto a los esperados para algunos campos de la base de datos como:

1. Campo MAC: el equipo técnico del CCA facilitó una lista con las MACs de cada uno de los 16 puntos de acceso ubicados en sus instalaciones, cada una de estas MACs varía el valor de su último octeto de la siguiente forma:
 - Si la MAC termina en 0 se encontraron coincidencias para 1 y 2. En la Fig. 4.9 se muestra un ejemplo de esto.



Figura 4.9: Ejemplo de coincidencias de MACs encontradas.

- Si la MAC termina en D se encontraron coincidencias para E y F. En la Fig. 4.10 se muestra un ejemplo de esto.

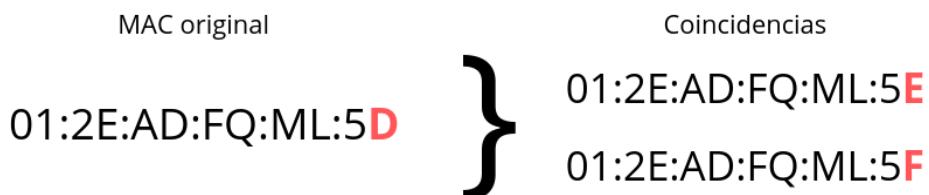


Figura 4.10: Ejemplo de coincidencias de MACs encontradas.

Por el alto grado de coincidencia de las MACs y la cantidad de dispositivos registrados como punto de acceso en el CCA, se concluyo, que cada uno de estos

subconjuntos de MAC están asociados a un único punto de acceso y corresponden a cada una de sus antenas e interfaces de red.

2. Campo RSSI:

- Valores de RSSI iguales a -1 dBm: alrededor de un 2.5 % de la base de datos total capturada corresponden a valores de RSSI iguales a -1 dBm. Después de una consulta al soporte técnico de Libelium (ver anexo A), aclararon que estos valores corresponden a paquetes enviados por el Meshlium cuando el dispositivo móvil se encuentra fuera de su cobertura, por lo tanto el Meshlium solo registra la mitad de la comunicación, de este modo, al no completarse el proceso de comunicación se registra el RSSI con un valor de -1 dBm.
- Valores de RSSI fuera del rango dinámico permitido por el Meshlium: en un 11 % del total de la base de datos se observaron muestras mayores a -35 dBm, indicando grados de saturación en los equipos, estos valores se asociaron a capturas hechas por los Meshliums a puntos de acceso.

4.3.2. Filtros aplicados

Las bases de datos capturadas por los 4 Meshliums del CCA cuentan en conjunto, aproximadamente con 2.600.000 registros semanales. Después de observar y analizar estos datos se realizaron los siguientes filtros:

- Tipo: para el caso de estudio, los registros de la base de datos que son relevantes, pertenecen a dispositivos móviles, por lo tanto todos los registros con Tipo “A” fueron eliminados, ya que corresponden a puntos de acceso. Al aplicar este filtro se eliminaron un total del 80.23 % de los registros de la base de datos, estos datos incluyen las capturas de MACs atípicas relacionadas a puntos de acceso.
- RSSI: todos los valores atípicos de RSSI (mayores a -35 dBm) fueron eliminados. Luego de aplicar este filtro la base de datos cuenta con 481.819 registros.
- AP: para superar los desafíos de la aleatorización de MACs, se decidió tomar en consideración solo los registros pertenecientes a dispositivos conectados a puntos

de acceso conocidos, de esta forma se cuenta con la verdadera MAC del dispositivo móvil y se pueden hacer estudios sobre la misma. Luego de aplicar los filtros anteriores, de los 481.819 registros restantes, el 87% pertenece a dispositivos cuyas direcciones MACs son aleatorias.

Después de extraer todos estos registros de la base de datos original, los datos útiles para generar información, son en promedio el 2,4 % (aprox. 62.636 registros semanales).

www.bdigital.ula.ve

C.C. Reconocimiento

Capítulo 5

Analíticos generados a partir de la base de datos capturada

El enfoque principal de este capítulo es mostrar qué analíticos pueden ser generados a partir de los datos capturados por los Meshliums. Además, que éstos sirvan como base de estudio para los analistas de negocios que buscan incrementar las ventas mejorando la experiencia del consumidor dentro de una tienda.

Con el fin de que cualquier usuario pueda entender la información suministrada por estos analíticos, los datos se presentan en un formato de alto nivel (tablas, gráficas, mapas...), y representan el comportamiento grupal de los clientes dentro del CCA.

5.1. Estadísticos de interés

En busca de mejorar las condiciones actuales de servicio al cliente en el CCA, se establecieron distintos estadísticos de interés que ayuden a un analista de negocios a mejorar el funcionamiento de dicho comercio, con el fin de aumentar las ventas y optimizar la distribución del personal dentro de las instalaciones.

La base de datos fue fraccionada semanalmente (de lunes a domingo) para disminuir la complejidad en comprensión de los analíticos. El reporte generado (similar al proceso de la sección 4.2.1) cuenta con los siguientes capítulos:

- Capítulo 1. Distribución de los clientes por horas del día: para un comerciante

es muy importante saber qué horas del día son las más transitadas por los clientes. Ésta información le permite al mismo planificar de forma óptima la distribución y cantidad de empleados necesarios dentro de la tienda, con el fin de brindar la mejor atención posible a los potenciales compradores, así como también desarrollar estrategias de mercadotecnia que le ayuden a incrementar las ventas en las horas menos transitadas.

Para todos los días de la semana, en la Fig. 5.2 se puede observar, los días y horas en los cuales los clientes visitan con mayor frecuencia el CCA, cada día es representado con un color diferente. Por lo general, se percibe un comportamiento similar entre días, siendo el sábado el día con mayor visitas de clientes.

Las horas con mayor cantidad de visitas son de 9am a 11am y de 3pm a 5pm, con una clara deserción de visitantes a horas del medio día. Un ejemplo más explícito de este comportamiento se detalla en la Fig. 5.1, donde para un día de la semana en específico (en este caso viernes) se muestra con un histograma la cantidad estimada de visitantes que hay por cada hora del día en la tienda.

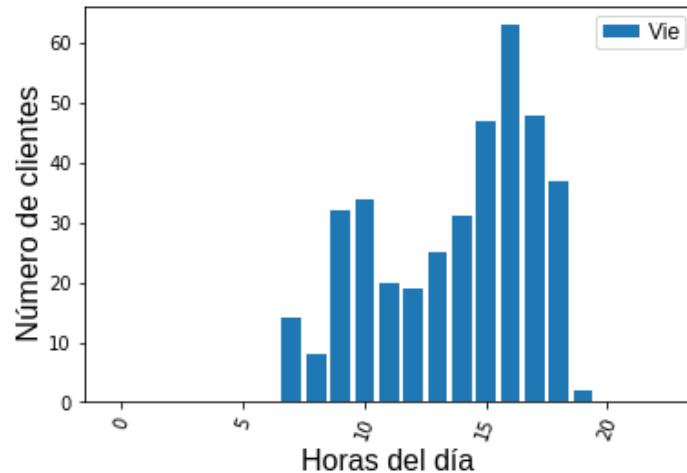


Figura 5.1: Cantidad estimada de clientes para un día.

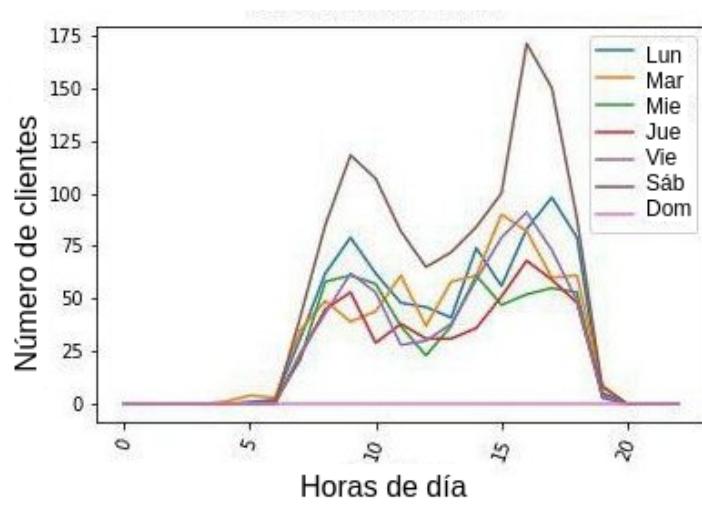


Figura 5.2: Cantidad estimada de clientes para todos los días.

La Fig. 5.3 responde a preguntas del tipo: porcentaje de visitantes de último minuto (después de las 6pm) y porcentaje de visitantes en la mañana y en la tarde. Es claro que la mayor cantidad de clientes (52.7 %) visita el CCA en la tarde y en cambio existe una pequeña porción (8.3 %) de clientes que son visitantes de último minuto.

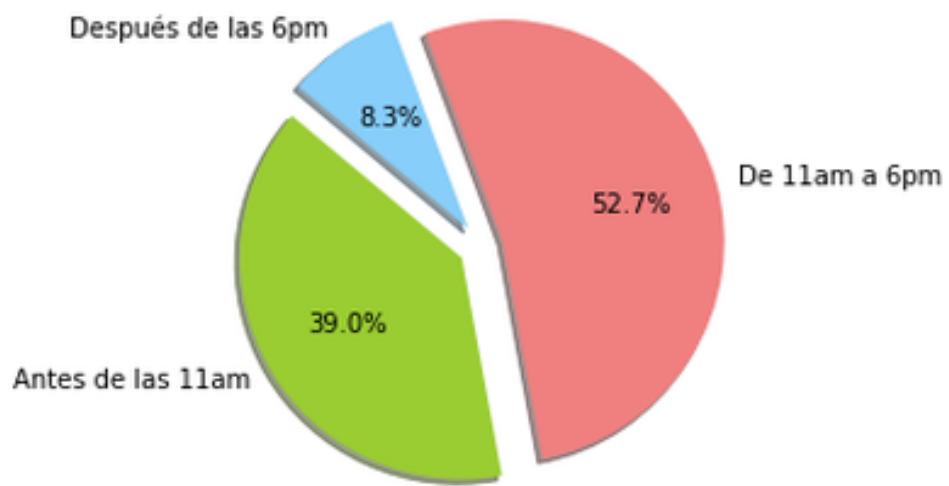


Figura 5.3: Distribución de los clientes por horas del día.

- Capítulo 2. Distribución de los clientes por días de la semana: una segunda forma de englobar información útil, en busca de que el comerciante tome decisiones óptimas sobre el manejo de sus recursos (humanos y tecnológicos), es determinar la frecuencia de visitas de clientes a lo largo de los días. La Fig. 5.4 muestra el porcentaje de visitas diarias dentro del CCA. Es evidente que el día con mayor cantidad de clientes es sábado (como ya fue expuesto en la Fig. 5.2) y que el día menos frecuentado es el jueves. El día domingo el CCA no labora por ende no hay clientes.

Al observar la Fig. 5.4, un analista de negocios podría exponerse algunas de las siguientes interrogantes: ¿Cómo incrementar las ventas los días miércoles y jueves?, ¿Cuántos empleados se necesitan el día sábado para cubrir el alto porcentaje de clientes?, ¿Es necesario tener todo el personal de ventas durante todos los días de la semana?. Preguntas como estas permiten plantear propuestas de mercadotecnia que incrementen la cantidad de visitas a lo largo de la semana y la calidad del servicio dentro del CCA.

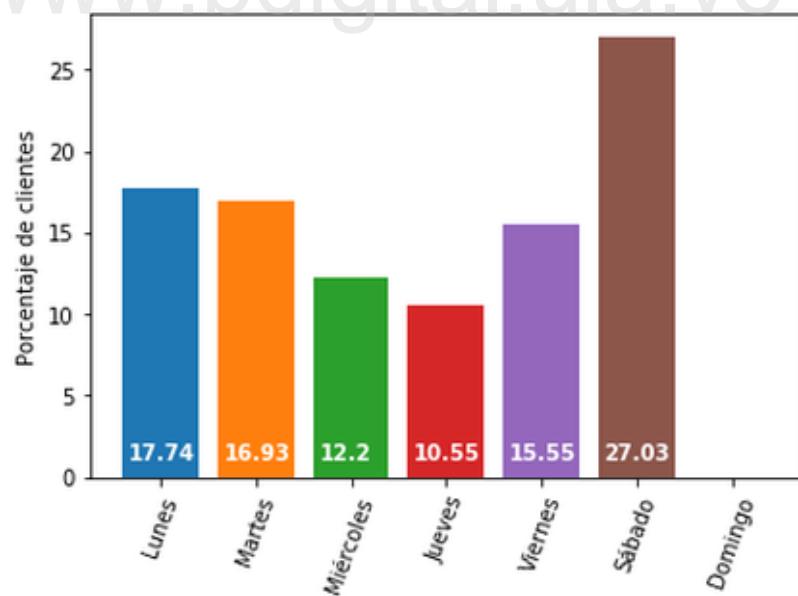


Figura 5.4: Porcentaje de visitas de clientes por día.

En la Fig. 5.5 se observa otra forma de representar esta información. En ésta figura se presenta una relación del porcentaje de visitas entre semana contra el

porcentaje de visitas para el fin de semana. Aunque el porcentaje de visitas entre semana es mayor al del fin de semana, teniendo en cuenta que para el fin de semana sólo se considera el día sábado (el domingo no laboran) su porcentaje de visitas es alto.

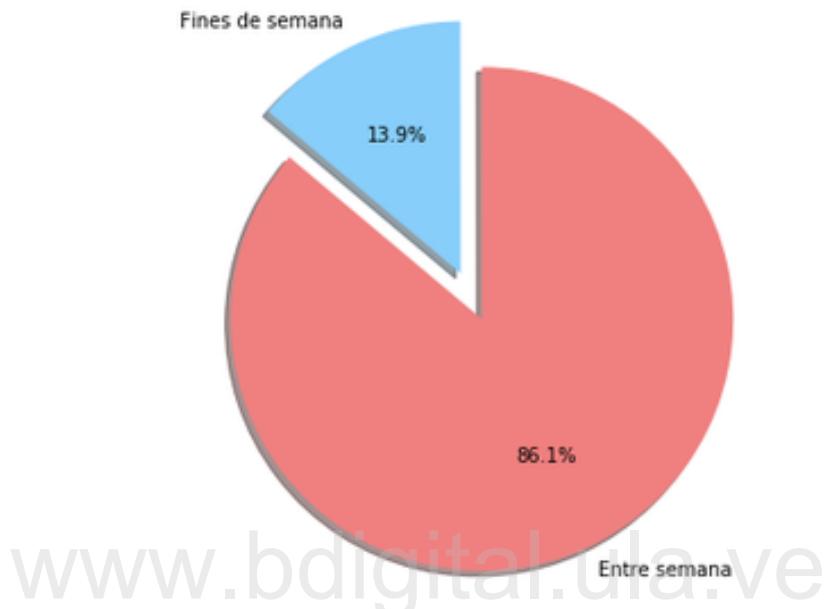


Figura 5.5: Porcentaje de visitas: semana vs fin de semana.

5.2. Mapas térmicos

Como fue mencionado con anterioridad, los datos captados por los Meshliums, no cuentan con las mejores características para hacer seguimiento de clientes a través del CCA. Por ejemplo, el RSSI no es una medida confiable por su gran variabilidad y la aleatorización de MACs dificulta identificar de forma única a los clientes. Es por ello que contando con los datos de clientes asociados a puntos de acceso en la tienda, se generaron mapas térmicos (ver sección 2.3) que indican las zonas con mayor ocupación de los clientes a lo largo del día.

Las figuras 5.6 y 5.7 son unas muestras de los mapas térmicos generados para el CCA, cada zona está representada por un punto de acceso cuya ubicación es conocida (ver Fig. 4.1). Los colores cálidos (en naranja intenso) representan las zonas más transitadas y los fríos (en azul oscuro) las menos transitadas por los clientes.

Para un día lunes (ver Fig. 5.6) las zonas con mayor afluencia de clientes son la zona Central, final, BTZ4 carpa y la de pago; y para un día sábado (ver Fig. 5.7) las zonas con mayor afluencia de clientes son la BTZ4 carpa y la de pago (ver Fig. 4.1).

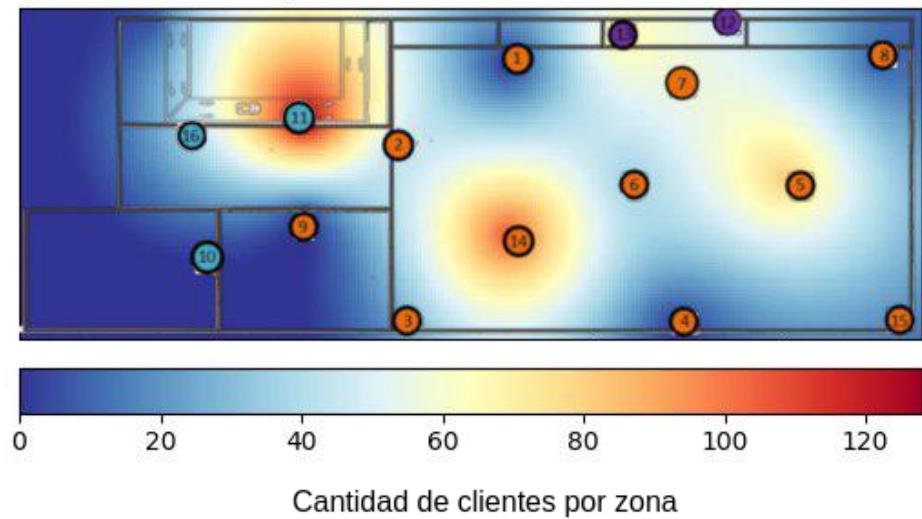


Figura 5.6: Mapa térmico generado para un lunes dentro del CCA.

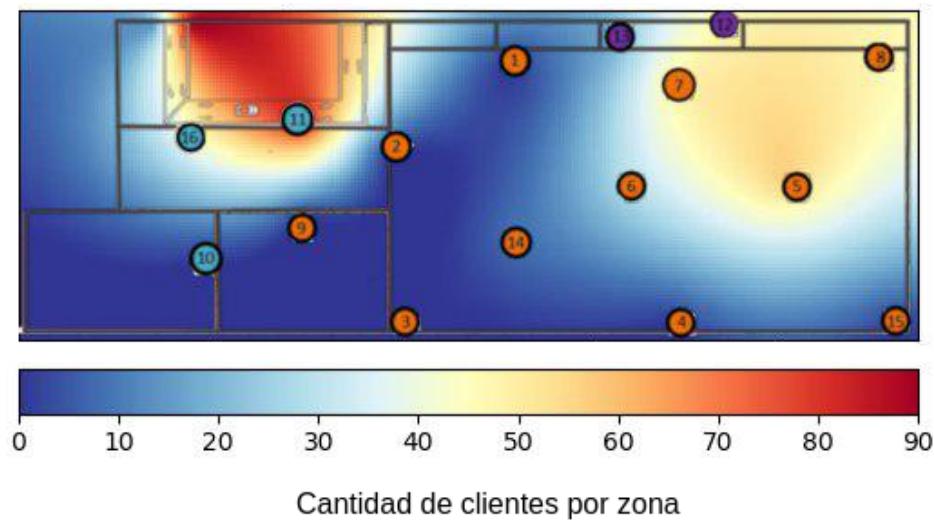


Figura 5.7: Mapa térmico generado para un sábado dentro del CCA.

Los mapas térmicos proporcionan información de sumo interés para un comerciante; a través de su análisis se puede determinar por día y si se quisiera por horas, las zonas

con mayor y menor popularidad dentro de un centro comercial. Esto permite saber cuándo y en qué lugares exponer publicidad a los usuarios, cómo distribuir el personal de la tienda a lo largo del día para brindar un mejor servicio y descubrir qué mejorar en las zonas poco visitadas para incrementar las ventas, todo esto como algunos ejemplos.

Para generar los mapas térmicos se usó el módulo de interpolación de datos de la biblioteca SciPy de Python. La razón principal de usar funciones de este tipo es porque sólo se conoce la cantidad de clientes por cada punto de acceso, por lo tanto los únicos píxeles cuyo color es conocido para la generación de los mapas térmicos son los asociados a la ubicación de los puntos de acceso (ver Fig. 4.1). Con la función de interpolación este problema se elimina ya partiendo del conjunto discreto de puntos conocidos se pueden generar nuevos, de esta forma se obtuvieron las degradaciones de colores observadas en las figuras 5.7 y 5.6.

5.3. Flujo de clientes

Para obtener una idea clara de como distribuir los productos en el interior de una tienda, es importante conocer los caminos que recorren los usuarios a través ella, ésto permite lograr un diseño ideal del interior del comercio y enfocar las campañas de mercadotecnia de forma óptima.

Para el estudio del flujo de clientes dentro del CCA, se generaron diagramas de Sankey (ver sección 2.4). Aunque en la actualidad éstos diagramas comúnmente son usados para seguir el tráfico de clientes en una página web, su uso es análogo al problema planteado (Schaffer, 2017). En una página web cada pestaña representa un nodo, de esta manera se pueden obtener relaciones del tipo origen→destino. De forma análoga, en el CCA cada punto de acceso representa un nodo cuya numeración ya es conocida (ver Fig. 5.8), lo que permite obtener el mismo tipo de relaciones.

La ventaja principal de usar estos diagramas es que se pueden determinar las rutas más populares seguidas por los clientes, de una forma sencilla y atractiva para el lector.

Los diagramas de Sankey en Python se producen usando el módulo `graph_objects` de la biblioteca Plotly. Del mismo modo que para los mapas térmicos, los diagramas de Sankey se construyeron con los datos de clientes asociados a puntos de acceso.

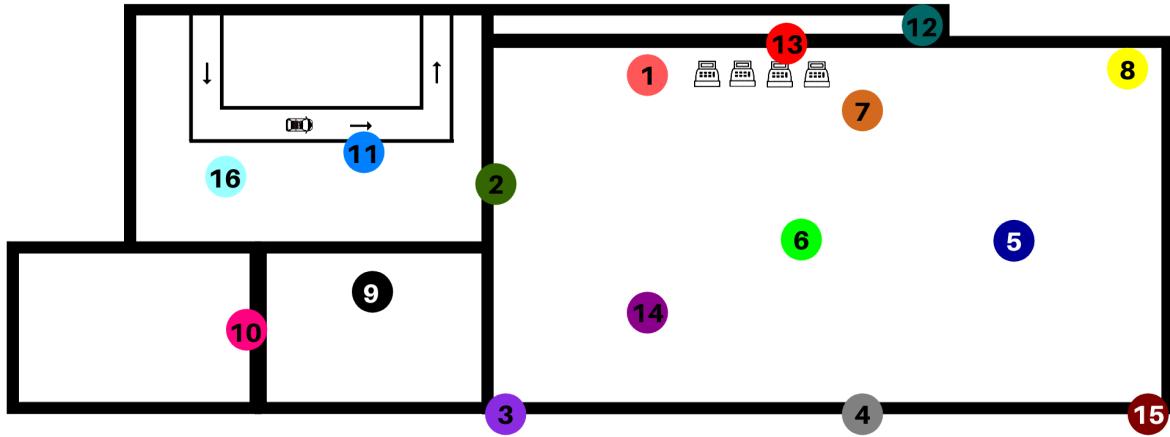


Figura 5.8: Muestra de los colores seleccionados para cada AP en el CCA.

En la Fig. 5.9 se muestra un diagrama de Sankey generado para el CCA. Cada nodo o punto de acceso es representado con un color diferente (ver Fig. 5.8), el flujo de los clientes entre un punto de acceso origen a un punto de acceso destino es representado con líneas del mismo color del punto de acceso de origen, de esta forma se conoce el sentido de la ruta. El grosor de cada línea describe la cantidad de clientes que la transita, siendo las más gruesas las que poseen un mayor flujo de clientes. Es importante mencionar que las posibles combinaciones de rutas entre puntos de accesos que no aparecen en el diagrama no existen.

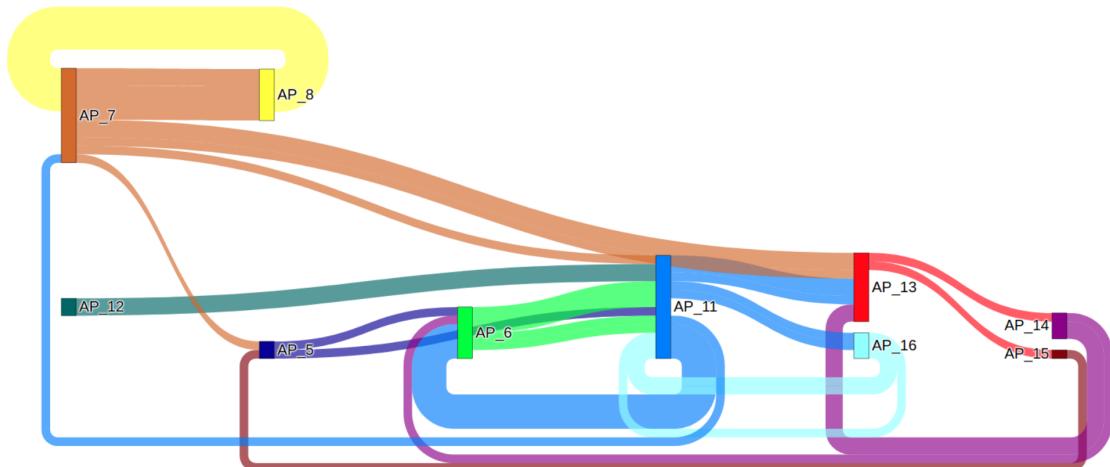


Figura 5.9: Diagrama de Sankey para un día sábado en el CCA.

Por ejemplo en la Fig. 5.9, tomando como nodo de origen el punto de acceso

AP_7 (marrón), el flujo de clientes va a los nodos de destino AP_5 (morado), AP_8 (amarillo), AP_11 (azul), AP_13 (rojo), donde el camino más transitado es el AP_7 → AP_8. En este contexto, esta información se puede interpretar como que los clientes pasan con frecuencia por la parte superior de la Zona Central para llegar a la parte superior de la Zona Final en el CCA (ver Fig. 4.1).

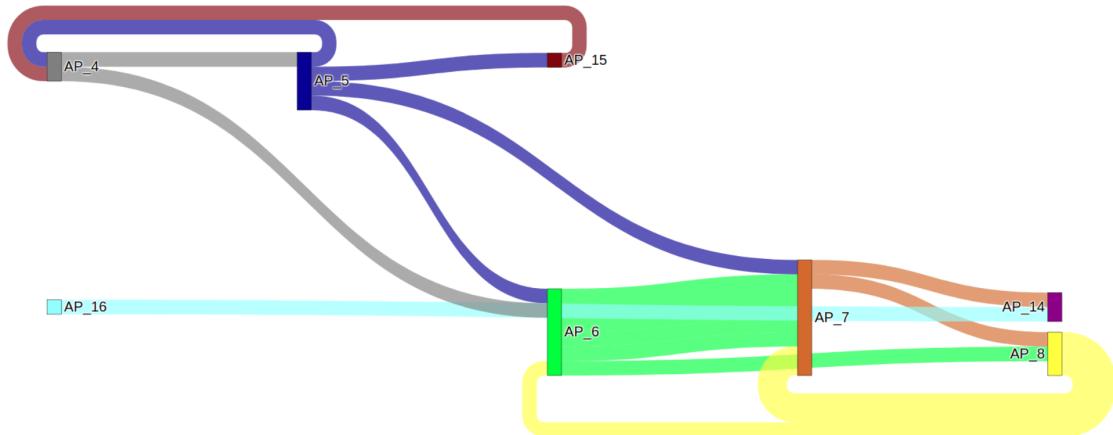


Figura 5.10: Diagrama de Sankey para un día lunes en el CCA.

Otro ejemplo de los diagramas de flujo generados para el CCA se observa en la Fig. 5.10. Por lo general los diagramas de Sankey toman formas totalmente diferentes al momento de ser graficados, debido a que el algoritmo que genera estos diagramas busca ubicar cada nodo de manera que éstos no se solapen entre sí. Ésto con el objetivo de que las relaciones que existen entre los nodos se entiendan claramente.

Capítulo 6

Conclusiones y trabajos futuros

Aunque los dispositivos Meshliums han adquirido una gran popularidad en los últimos años, los datos que escanean no son lo suficientemente estables para hacer seguimiento a los usuarios en ambientes interiores. En concreto, las medidas de RSSI que capturan estos sensores son altamente variables, incluso para dispositivos móviles que se encuentran en la misma posición por tiempos prolongados.

Se detectó que la variabilidad del RSSI está relacionada a la medida capturada, es decir, los equipos con valores de RSSI bajos presentaron bajas tasas de variabilidad y los equipos con valores de RSSI altos presentaron altas tasas de variabilidad.

La aleatorización de MACs que realizan los dispositivos móviles, hace muy difícil la identificación de usuarios de forma única dentro de una localidad. Aunado a esto, el gran incremento de solicitudes de sondeo que genera este proceso, imposibilita la oportunidad de tener un conteo real de usuarios sólo con los datos capturados por los equipos Meshliums, es por ello que los datos de los dispositivos que realizan aleatorización no se pueden usar para generar analíticos confiables.

Luego de hacer un análisis exhaustivo de los datos del caso de estudio, para la generación de analíticos sólo se usaron el 2 % de los mismos, debido a que la mayoría de los datos que los Meshliums capturaron dentro del CCA pertenecían a puntos de acceso o a dispositivos móviles que realizaban aleatorización de MACs.

Los mapas térmicos permiten estudiar las zonas con mayor ocupación de clientes, este estudio ayuda a un comerciante a optimizar el uso de sus recursos (humanos

y tecnológicos) con el fin de mejorar la calidad del servicio a los consumidores e incrementar las ventas. Además, un analista de negocios puede usar estos mapas para idear campañas de mercadotecnia y publicitarlas justo en los lugares donde existe una mayor concentración de clientes, incluso estas campañas pueden ser orientadas a publicitar las zonas menos visitadas.

El análisis de los diagramas de flujo ayuda a comprender cómo se desplazan los clientes dentro de una localidad, presentando información útil para mejorar la distribución de los productos y el diseño dentro de una tienda; ya que este diseño influye en el deseo de compra de los clientes. Además, un comerciante puede aprovechar esta información para presentar anuncios, promociones o publicidad en los recorridos más transitados por los clientes.

Como trabajo futuro se puede proponer usar otros tipos de sensores inalámbricos (de calor, movimiento, infrarrojos...) que obtengan un conteo real de la cantidad personas en un ambiente interior; con el fin de encontrar una relación entre la cantidad de MACs aleatorias con la cantidad de usuarios reales. Y eventualmente, calcular un estimado más exacto de la cantidad de usuarios en el comercio o la zonificación de los clientes en cada área de la tienda.

Anexo A

Consulta al soporte técnico de Libelium

Andres posted 03/25/2019 10:22 AM

Hi,

If the RSSI is -1 then this is for a packet which came from the AP to the client but the client transmissions are out of range for your card. Meaning you are hearing only 1/2 of the communication.

Regards

Figura A.1: Respuesta del soporte técnico de Libelium por el RSSI = -1dBm.

Bibliografía

- Akyildiz, I. F. y Wang, X. (2005). A survey on wireless mesh networks. *IEEE Communications magazine*, 43(9):S23–S30.
- Aspyct.org (2019). Software Airodump-ng. url: <https://www.aircrack-ng.org>. Accedido 08-06-2019.
- Barbolla, A. M. B., P. J. A. B. (2005). *Tecnologías de localización: Fundamentos y Aplicaciones*. Universidad Politécnica de Madrid.
- C. Wu, Z. Yang, Y. L. y Xi, W. (2013). WILL: Wireless indoor localization without site survey. *IEEE Trans. Parallel and Distributed Systems*, 24:839–848.
- Careers, . (2017). The Data Science Course 2019: Complete Data Science Bootcamp. url: <https://www.udemy.com/course/the-data-science-course-complete-data-science-bootcamp>. Accedido 05-04-2019.
- Goulart, V. (2014). Techniques for using a heat map of a retail location to disperse crowds. US Patent App. 13/715,635.
- J.E. van Engelen, J.J. van Lier, F. T. y Trautmann, H. (2018). Accurate WiFi-based Indoor Positioning with Continuous Location Sampling. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, p. 524–540.
- Jeril Kuriakose, A. V. y Nandhini, S. (2014). A Survey on Localization of Wireless Sensor Nodes. *In International Conference on Information Communication and Embedded Systems*, p. 1–6.

- Judd, C. M., McClelland, G. H., y Ryan, C. S. (2011). *Data analysis: A model comparison approach*. Routledge.
- Kotler, P., Armstrong, G., Saunders, J., y Wong, V. (2002). Capítulo 1: ¿Qué es Marketing? *Principles of Marketing (3^a edición europea edición)*. Essex (Inglaterra): Prentice Hall.
- Kurose, J. F., Ross, K. W., Hierro, C. M., y Pablo, Á. P. d. M., y Marrone, L. (2010). *Redes de computadoras: un enfoque descendente*, volumen 5. Pearson.
- Libelium (2019a). Case Studies. url: <http://www.libelium.com/libeliumworld/case-studies>. Accedido 01-04-2019.
- Libelium, C. D. (2019b). *Meshlium Xtreme Technical Guide*. Document version: v8.4 - 06/2019.
- Long Vu, Quang Do, K. N. (2011). Jyotish: A Novel Framework for Constructing Predictive Model of People Movement from Joint WifilBluetooth Trace. *IEEE International Conference on Pervasive Computing and Communications*, p. 54–62.
- Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E. C., y Brown, D. (2017). A study of MAC address randomization in mobile devices and when it fails. *Proceedings on Privacy Enhancing Technologies*, 2017(4):365–383.
- Meyer, U. B., Creux, S. E., y Marin, A. K. W. (2006). *Process oriented analysis: design and optimization of industrial production systems*. CRC Press.
- NetSpot (2019). What is RSSI and its relation to a Wi-Fi network. url: <https://www.netspotapp.com/what-is-rssi-level.html>. Accedido 02-03-2019.
- Nuño Barrau, G. y J. P. B. (2006). Algoritmos para localización en interiores mediante WLAN comerciales basados en funciones discriminantes lineales y modelos ocultos de Markov. "Tesis Doctoral ETSI Telecomunicación, Madrid".
- Olivares, J. L. C. (2009). *Modelo para la cobertura de redes inalámbricas en interiores*. Tesis Pregrado, Universidad de Sevilla.

- Pahlavan, K., Li, X., y Makela, J.-P. (2002). Indoor geolocation science and technology. *IEEE Communications Magazine*, 40(2):112–118.
- Pozuelo, F. (2009). *Algoritmos de aprendizaje estadístico aplicados a la radiolocalización en interiores*. Tesis Doctoral, Universidad Rey Juan Carlos, Fuenlabrada.
- Runkler, T. A. (2012). *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Springer Vieweg.
- Rutermann, T., Benabbas, A., y Nicklas, D. (2019). Know Thy Quality: Assessment of Device Detection by WiFi Signals. En *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, p. 639–644. IEEE.
- Schaffer, J. (2017). 4 interactive Sankey diagrams made in Python. url: <https://medium.com/plotly/4-interactive-sankey-diagram-made-in-python-3057b9ee8616>. Accedido 05-03-2019.
- Schmidt, M. (2006). Der Einsatz von sankey-diagrammen im stoffstrommanagement. Reporte técnico, Beiträge der Hochschule Pforzheim.
- Yaeli, A., Bak, P., Feigenblat, G., Nadler, S., Roitman, H., Saadoun, G., Ship, H. J., Cohen, D., Fuchs, O., Ofek-Koifman, S., y cols. (2014). Understanding customer behavior using indoor location analysis and visualization. *IBM Journal of Research and Development*, 58(5):3–1.
- Zeng, Y., Pathak, P. H., y Mohapatra, P. (2015). Analyzing shopper's behavior through wifi signals. En *Proceedings of the 2nd workshop on Workshop on Physical Analytics*, p. 13–18. ACM.
- Zenko, J. (2018). When (and Why) to use Heat Maps. url: <https://www.dundas.com/resources/dundas-data-visualization-blog/when-and-why-to-use-heat-maps>. Accedido 05-03-2019.