

**ANÁLISIS EXPLORATORIO DE DATOS PARA EL MODELAMIENTO
PROBABILÍSTICO PARA EL MEJORAMIENTO Y OPTIMIZACIÓN DE REDES
INALÁMBRICAS**

**LAURA LIESEL TORRES ACEVEDO
YEFERSSON ADRIÁN ROJAS REAL**

**DIRECTOR:
HECTOR IVAN REYES MONCAYO**

**UNIVERSIDAD DE LOS LLANOS
FACULTAD CIENCIAS BÁSICAS E INGENIERÍA
INGENIERÍA DE SISTEMAS
VILLAVICENCIO-META**

**ANÁLISIS EXPLORATORIO DE DATOS PARA EL MODELAMIENTO
PROBABILÍSTICO PARA EL MEJORAMIENTO Y OPTIMIZACIÓN DE REDES
INALÁMBRICAS**

**LAURA LIESEL TORRES ACEVEDO
JEFERSSON ADRIAN ROJAS REAL**

Presentado para optar al título de: INGENIEROS DE SISTEMAS

**DIRECTOR
HÉCTOR IVÁN REYES MONCAYO Ph.D.**

**UNIVERSIDAD DE LOS LLANOS
FACULTAD DE CIENCIAS BÁSICAS E INGENIERÍA
INGENIERÍA DE SISTEMAS
VILLAVICENCIO – META**

NOTA DE ACEPTACIÓN

Héctor Iván Reyes
Director

DEDICATORIA

*A Dios, a mis padres, a mi hermana y a mi novio
que siempre han estado junto a mí apoyándome
Incondicionalmente, y en especialme a Mí por no darme por vencida
en seguir adelante cuando todos me decían lo contrario.*

Laura L. Torres Acevedo

*A Dios, a mis padres, a mis hermanos por su ayuda
incondicional y apoyo, a mi novia por su paciencia y amor,
para ellos este trabajo.*

Yefersson Adrián Rojas Real

AGRADECIMIENTOS

A Dios por permitirnos realizar nuestros sueños.

*A nuestro director Ing. Héctor Iván Reyes M.
Que nos guio y ayudó en el proceso para
Culminar nuestra tesis.*

*A la universidad de los llanos por brindarnos,
una educación de calidad para poder
realizar nuestros estudios*

*A nuestras familias por su ayuda
incondicional, su apoyo fue
fundamental para el desarrollo
de nuestra tesis.*

CONTENIDO

NOTA DE ACEPTACIÓN	3
DEDICATORIA.....	4
AGRADECIMIENTOS	5
CONTENIDO.....	6
CONTENIDO DE GRÁFICAS.....	8
RESUMEN DEL PROYECTO	14
DESCRIPCIÓN DEL PROBLEMA.....	15
JUSTIFICACIÓN	16
INTRODUCCIÓN	17
4. OBJETIVOS.....	19
4.1 Generales.....	19
4.2 Específicos	19
5. MARCO TEÓRICO	20
5.1 Wi-Fi.....	20
5.1.1 Trama 802.11	20
5.1.3 Tramas de gestión.....	22
5.1.4 Trama de control	22
5.1.5 Trama de datos	23
5.2 Análisis exploratorio de datos.....	23
5.2.1 Tipos de variables	24
5.3.1 Gráficas.....	26
5.4.1 Método Pearson.....	26
5.4.2 Método Spearman.....	26
5.4.3 Kendall	27
6 ANÁLISIS EXPLORATORIO DE DATOS APLICADO A CAPTURAS DE WIRESHARK DE UNA RED WI-FI.....	28
6.1 Definición de variables	28
6.2 Proceso de obtención de datos y variables	29
6.3 Obtención de variables.....	29

6.4	Caracterización.....	33
6.4.1	Frecuencias esperadas	33
6.4.2	Gráficas.....	34
6.5	Relación entre variables	37
6.5.1	Gráficas.....	37
6.5.2	Correlación.....	38
7	RESULTADOS Y DISCUSIÓN	40
7.1	Caracterización.....	40
7.1.1	Caracterización de la variable access points	40
7.1.2	Caracterización de la variables estaciones	44
7.1.3	Caracterización de la variable retry rate.....	47
7.1.4	Caracterización de la variable potencia.....	51
7.1.5	Caracterización de la variables longitud.....	56
7.2	Correlación	63
7.2.1	Correlación de la variable longitud	63
7.2.2	Correlación de la variable potencia	69
	Correlación tasa de retransmisión y potencia de varias redes	77
7.2.3	Correlación de la variable estaciones.....	79
7.2.4	Correlación de la variables access points	84
8	LISTA DE ACRÓNIMOS.....	89
9	CONCLUSIONES	91
10	ANEXOS	93
10.1	Captura de datos por medio de la herramienta Wireshark	93
10.2	Procesamiento de los datos capturados	96
10.3	Importación de librerías.....	97
11	BIBLIOGRAFÍA.....	99

CONTENIDO DE GRÁFICAS

Gráfica 1 Histograma v.AP conectividad eficiente y conectividad deficiente	41
Gráfica 2 KDE conectividad eficiente y conectividad deficiente	42
Gráfica 3 CDF V.AP conectividad eficiente y conectividad deficiente	42
Gráfica 4 Histograma de V. número de AP conectividad deficiente con respecto a tramas retransmitidas, no retransmitidas	43
Gráfica 5 KDE de número de AP conectividad deficiente con respecto a tramas retransmitidas y no retransmitidas.....	43
Gráfica 6 CDF del número de AP conectividad deficiente con respecto a tramas retransmitidas y no retransmitidas.....	44
Gráfica 7 Histograma V.STA conectividad eficiente y conectividad deficiente	44
Gráfica 8 KDE V.STA conectividad eficiente y conectividad deficiente	45
Gráfica 9 CDF V.STA conectividad eficiente y conectividad deficiente	45
Gráfica 10 Histograma V.STA conectividad deficiente con respecto a tramas retransmitidas, no retransmitidas	46
Gráfica 11 KDE V.STA conectividad eficiente y deficiente con respecto a tramas retransmitidas, no retransmitidas	46
Gráfica 12 CDF V.STA conectividad eficiente y deficiente con respecto a tramas retransmitidas, no retransmitidas	47
Gráfica 13 Histograma V.retry rate conectividad eficiente y conectividad deficiente	49
Gráfica 14 KDE V.retry rate conectividad eficiente y conectividad deficiente.....	49
Gráfica 15 CDF V.retry rate conectividad eficiente y conectividad deficiente.....	50
Gráfica 16 PDF V.retry rate conectividad eficiente y conectividad deficiente	50
Gráfica 17 Histograma V.Potencia conectividad eficiente y conectividad deficiente	51
Gráfica 18 KDE V.potencia conectividad eficiente y conectividad deficiente	52
Gráfica 19 PDF V.potencia conectividad eficiente y conectividad deficiente.....	52
Gráfica 20 CDF V.potencia conectividad eficiente y conectividad deficiente	53
Gráfica 21 Potencia de número de tramas retransmitidas y no retransmitidas	54
Gráfica 22 Histograma de tramas retransmitidas y no retransmitidas con cierta potencia.....	54
Gráfica 23 KDE de tramas retransmitidas y no retransmitidas con cierta potencia.....	55
Gráfica 24 PDF de tramas retransmitidas y no retransmitidas con cierta potencia.....	55
Gráfica 25 CDF tramas retransmitidas y no retransmitidas con cierta potencia.....	56
Gráfica 26 Longitud conectividad eficiente y deficiente	58
Gráfica 27 Histograma V.longitud con conectividad eficiente y conectividad deficiente	59
Gráfica 28 KDE V.Longitud con conectividad eficiente y conectividad deficiente..	59

Gráfica 29 CDF Longitud conectividad eficiente y conectividad deficiente.....	60
Gráfica 30 Longitud tramas retransmitidas y no retransmitidas.....	60
Gráfica 31 Histograma de tramas retransmitidas y no retransmitidas con cierta longitud.....	61
Gráfica 32 KDE de tramas retransmitidas y no retransmitidas con cierta longitud	61
Gráfica 33 CDF de tramas retransmitidas y no retransmitidas con cierta longitud	62
Gráfica 34 Dispersión de la variable longitud máxima respecto al porcentaje de retransmisión.....	67
Gráfica 35 Dispersión de la variable longitud mínima respecto al porcentaje de retransmisión.....	67
Gráfica 36 Dispersión de la variable longitud promedio respecto al porcentaje de retransmisión.....	68
Gráfica 37 Dispersión de la variable longitud frecuente respecto al porcentaje de retransmisión.....	68
Gráfica 38 Dispersión de la variable potencia máxima respecto al porcentaje de retransmisión.....	73
Gráfica 39 Dispersión de la variable potencia mínima respecto al porcentaje de retransmisión.....	73
Gráfica 40 Dispersión de la variable potencia promedio respecto al porcentaje de retransmisión.....	74
Gráfica 41 Dispersión de la variable potencia frecuente respecto al porcentaje de retransmisión.....	74
Gráfica 42 Hexagonal binning de la variable potencia máxima respecto al porcentaje de retransmisión	75
Gráfica 43 Hexagonal binning de la variable potencia mínima respecto al porcentaje de retransmisión	75
Gráfica 44 Hexagonal binning de la variable potencia promedio respecto al porcentaje de retransmisión	75
Gráfica 45 Hexagonal binning de la variable potencia frecuente respecto al porcentaje de retransmisión	76
Gráfica 46 Histograma de potencia	76
Gráfica 47 Cuartil de potencia	77
Gráfica 48 Correlación tasa de retransmisión y potencia de varias redes.....	77
Gráfica 49 Hexagonal-binning potencia promedio donde se presenta mayor porcentaje de tasa de retransmisión en una red	78
Gráfica 50 Dispersión número de estaciones respecto al porcentaje de retransmisión.....	83
Gráfica 51 Hexagonal binning número de estaciones respecto al porcentaje de retransmisión.....	83
Gráfica 52 Dispersión número de access points respecto al porcentaje de retransmisión.....	88

Gráfica 53 Hexagonal binning número de access points respecto al porcentaje de retransmisiones 88

ÍNDICE DE ILUSTRACIONES

Ilustración 1 formato de trama de datos	21
Ilustración 2 proceso de análisis exploratorio de datos	29
Ilustración 3 tabla de frecuencias	34
Ilustración 4 gráfica de histograma.....	35
Ilustración 5 gráfica del PDF	35
Ilustración 6 gráfica de CDE	36
Ilustración 7 gráfica de KDE	36
Ilustración 8 gráfica de dispersión hexagonal.....	37
Ilustración 9 gráfica de dispersión	38
Ilustración 10 tabla de correlaciones Spearman rank.....	39
Ilustración 11 tabla de correlaciones Normal y Spearman	39
Ilustración 12 inicialización Wireshark.....	93
Ilustración 13 interfaz de la herramienta Wireshark	93
Ilustración 14 configuración de modo de captura inalámbrica.....	94
Ilustración 15 captura con la herramienta Wireshark	95
Ilustración 16 exportar un archivo a formato csv	95
Ilustración 17 guardar archivo en formato csv.....	96
Ilustración 18 iniciar notebook por terminal	96
Ilustración 19 interfaz jupyter notebook.....	97
Ilustración 20 hoja de trabajo ipython.....	97

CONTENIDO DE TABLAS

Tabla 2 tipos de tramas estándar 802.11	21
Tabla 3 Conteo de número de access points por cada 100 tramas	40
Tabla 4 frecuencias de la variable retry rate.....	47
Tabla 5 tabla de frecuencia de la variable potencia	51
Tabla 6 tabla de frecuencia de la variable longitud	56
Tabla 7 Correlación de la variable longitud con conectividad eficiente	63
Tabla 8 Correlación de la variable longitud con conectividad deficiente	64
Tabla 9 Correlación de la variable longitud entre retry rate con conectividad eficiente.....	65
Tabla 10 Correlación de la variable longitud entre retry rate con conectividad deficiente.....	65
Tabla 11 Correlación Spearman con p-value de la variable longitud entre retry rate con conectividad eficiente	66
Tabla 12 Correlación Spearman con p-value de la variable longitud entre retry rate con conectividad deficiente	66
Tabla 13 Correlación de la variable potencia con respecto a retry rate con conectividad eficiente.....	69
Tabla 14 Correlación de la variable potencia con respecto a retry rate con conectividad deficiente	70
Tabla 15 Correlación Spearman con p-value de la variable potencia entre retry rate con conectividad eficiente	71
Tabla 16 Correlación Spearman con p-value de la variable potencia entre retry rate con conectividad deficiente	71
Tabla 17 Correlación Spearman de la variable potencia entre retry rate con conectividad eficiente.....	72
Tabla 18 Correlación Spearman de la variable potencia entre retry rate con conectividad deficiente.....	72
Tabla 19 Correlación de la variable estaciones con respecto a retry rate, conectividad eficiente.....	79
Tabla 20 Correlación de la variable estaciones con respecto a retry rate, conectividad deficiente.....	80
Tabla 21 Correlación Spearman variable estaciones respecto a retry rate con conectividad eficiente.....	82
Tabla 22 Correlación Spearman variable estaciones respecto a retry rate con conectividad deficiente.....	82
Tabla 23 Correlación Spearman con p-valúes variable estaciones respecto a Retry rate con conectividad eficiente	82

Tabla 24 Correlación Spearman con p-values variable estaciones respecto a retry rate con conectividad deficiente	82
Tabla 25 Correlación de la variable access points con respecto a retry rate, conectividad eficiente	84
Tabla 26 Correlación de la variables access points con respecto a retry rate, conectividad deficiente	85
Tabla 27 Correlación Spearman de la variable access points con respecto a retry rate, conectividad eficiente	86
Tabla 28 Correlación Spearman de la variable access points con respecto a retry rate, conectividad deficiente	87
Tabla 29 Correlación Spearman de con p-value la variable access points con respecto a retry rate, conectividad eficiente	87
Tabla 30 Correlación Spearman de con p-value la variable access points con respecto a retry rate, conectividad deficiente	87

RESUMEN DEL PROYECTO

Las redes y sistemas inalámbricos están sujetos a situaciones que determinan su desempeño eficiente, tales como interferencia, ruido, pérdidas de propagación, escasez de espectro radioeléctrico, entre otros. Debido al acontecimiento de situaciones como las mencionadas, existen diversos desarrollos que permiten monitorear el funcionamiento de las diferentes redes, fundamentalmente midiendo y reportando la calidad en el intercambio de paquetes. No obstante persiste una necesidad por contar con herramientas para analizar la información reportada utilizando métodos estadísticos y probabilísticos, que faciliten la toma de decisiones relativas al diseño y configuración de redes inalámbricas, en función de la calidad de la transmisión de los datos.

Este proyecto busca realizar el análisis exploratorio de datos siendo el primer paso para la construcción de modelos probabilísticos y estadísticos con base a los datos capturados a través de la herramienta Wireshark, siguiendo una metodología de investigación. El análisis que se obtiene en paralelo con los objetivos del proyecto concibe una serie de variables con los que se determinaran el desempeño de una red seleccionada de acuerdo a unas características predeterminadas. La mayor parte del proceso de extracción y transformación de los datos se construye en Python y consigo una serie de librerías especializadas en big data. La finalidad del análisis e identificación de variables que contiene este proyecto es dar el desarrollo de la creación de una información visible y entendible, resolviendo las problemáticas presentes en las redes Wi-Fi, al igual la observación de análisis con sus respectivos resultados en los diferentes ambientes en los que encontramos las redes inalámbricas realizando una comparación entre estas, e inicios de nuevos estudios.

Se espera que los resultados sirvan para la creación de nuevo diseño en la construcción de redes y sistemas inalámbricos que mejoren el desempeño. Igualmente, el entendimiento de la dinámica de funcionamiento de las redes inalámbricas.

DESCRIPCIÓN DEL PROBLEMA

Planteamiento del problema de la investigación: las redes inalámbricas de comunicación, están sujetas a una serie de fenómenos que impactan su operación y rendimiento. Por ejemplo, la interferencia y ruido electromagnéticos afectan la señal transmitida, inalámbricamente provocando errores que causan retransmisiones de paquetes, que causan aún más congestión en las redes inalámbricas. Este problema es más evidente visible en redes Wi-Fi, que usan espectro no licenciado y donde los dispositivos que trabajan bajo este estándar, deben compartir porciones del espectro entre ellos y con dispositivos de otras tecnologías.

Otro problema, es que con el crecimiento de aplicaciones inalámbricas, el espectro radioeléctrico se hace cada vez más escaso, lo cual implica un replanteamiento a la forma como se viene gestionando este recurso. Actualmente se cuenta con algunas herramientas, que permiten monitorear el funcionamiento de redes inalámbricas, tales como el software Wireshark y las estadísticas entregadas por enrutadores inalámbricos tarjetas inalámbricas Wi-Fi. Sin embargo, estas herramientas simplemente entregan datos sin interpretación. Debido a que estos datos corresponden a variables aleatorias, como por ejemplo potencia de señal recibida, número de retransmisiones, tiempos de respuesta, entre otros. En este proyecto se propone, extraer la información contenida en estos por medio de inferencia estadística y modelos probabilísticos. Por lo tanto, este proyecto busca identificar las variables capturadas en herramientas la Wireshark.

Como se sabe en si Wireshark no arroja las variables como tal, se debe hacer un análisis más afondo para poder realizar su debido estudio, de esta manera escoger las variables que afectan más en una red revisarlas cada una, observar el comportamiento hacer diferentes preguntas para poder ver más allá de lo que nos arrojan las diferentes herramientas, en grupo o frente a una variable específica, de tal manera poder verificar el error de la red y darle solución a este.

JUSTIFICACIÓN

La mayoría de las herramientas como Wireshark no entrega datos visibles para la solución de problemáticas que se presenta comúnmente en redes inalámbricas, esto conlleva a realiza una investigación con base a estos datos, teniendo como resultado redes con mejores desempeños. Se necesita analizar los datos que se obtienen, de esta manera podemos empezar a resolver las preguntas o dudas que el investigador deba tener. Es importante no solamente tener datos, si no también saber formular las preguntas adecuadas, que se puedan responder. Esta investigación, quiere llegar al cómo plantear las preguntas que se puedan responder, con base a los datos que se obtuvieron por medio de Wireshark. Se sabe inicialmente que las preguntas están relacionadas con el rendimiento de la red; sin embargo, durante la investigación se buscará, expresar las preguntas de la manera apropiada para ser respondidas por procedimientos probabilísticos y de inferencia estadística.

Por lo tanto, se debe tener en cuenta a la hora de realizar el estudio, variables como son el medio, el tiempo, entre otros aspectos donde se realiza la captura, dado que afecta los resultados de la red la cual se va analizar.

INTRODUCCIÓN

Las redes inalámbricas de comunicación, están sujetas a una serie de fenómenos que impactan su operación y rendimiento. Este problema es más evidente en redes Wi-Fi, como es la interferencia que usan espectro no licenciado y donde los dispositivos que trabajan bajo este estándar, deben compartir porciones del espectro entre ellos y con dispositivos de otras tecnologías.

Actualmente se cuenta con algunas herramientas, que permiten monitorear el funcionamiento de redes inalámbricas tales como el software Wireshark [1] y las estadísticas entregadas por enrutadores inalámbricos tarjetas inalámbricas Wi-Fi. Sin embargo, estas herramientas simplemente entregan datos sin interpretación. Debido a que estos datos corresponden a variables aleatorias, como por ejemplo potencia de señal recibida, número de retransmisiones, tiempos de respuesta, entre otros.

En este proyecto se propone extraer la información contenida, en estos por medio de inferencia estadística y modelos probabilísticos. Por lo tanto, este proyecto busca identificar las variables capturadas por herramientas como Wireshark, y analizarlas de tal manera, que se pueda encontrar el error o lo que afecta a la red, poder ver más allá de lo que arroja la herramienta a utilizar. Para poder realizar esto se debe coger variable por variable, hacerle la respectiva caracterización, luego determinar la relación entre variables por medio de su correlación y de gráficos de dispersión. También se deben generar diferentes gráficos donde se explique el comportamiento de las variables, según su promedio, máximo, mínimo y moda, lo cual nos ayudara a ver mejor la situación en el momento de lo que sucede en la red, de esta manera podemos ver que es lo que afecta a la hora de la transmisión de datos.

El análisis de correlación, consiste en medir la relación que tiene las variables entre ellas, se debe ver la relación de las variables dependientes e independientes. El análisis de correlación múltiple muestra el grado de asociación entre dos o más variables independientes y dependiente. Existen dos tipos de correlaciones la simple y múltiple. La correlación múltiple consiste en mostrar el grado de asociación que se encuentra entre dos o más variables independientes o dependientes. La correlación simple consiste en mostrar la cantidad da variación conjunta que tiene dos variables aleatorias de una distribución, en si cuantifica la dependencia lineal, por lo tanto, recibe el nombre de correlación lineal, también conocida como coeficiente de correlación de Pearson designado [2].

Wireshark, es una herramienta que nos facilita la recolección de los datos, este programa nos facilita las cosas un poco, teniendo en cuenta que podemos ver los datos y nos muestran más información de la que pensamos, se necesita ver más allá de lo que se ve para poder hacer un buen análisis, esta herramienta es muy buena, su entorno gráfico es muy amigable, es libre cualquier persona la puede utilizar, se dará cuenta de todas las opciones que tiene para brindar su servicio.

Por otra parte el Wi-Fi es una tecnología que está siendo muy utilizada últimamente, por la facilidad que ofrece a los usuarios de estar conectados en cualquier lugar, momento sin estar en un solo sitio para poder acceder a este, esta nueva conectividad tiene sus ventajas y desventajas, pero sin importar esto se utiliza mucho, teniendo en cuenta esto nos ofrece mucha información, que a simple vista no podemos ver ni entender, pero es muy importante para nosotros, lo cual podemos realizar varios análisis para mejorar o descubrir los errores que puede tener una red, de esta manera podemos solucionar todo lo que sucede y se puede descubrir muchas cosas que no sabíamos que se podían ver.

Al hacer una captura con el Wireshark y las redes Wi-Fi que nos suministran mucha información, se debe revisar con mucha precaución la información que nos da cada uno de los campos, se deben mirar que variables nos ofrecen información más relevante, para así poder hacer un análisis más afondo de lo que en si nos sirve, para observar que es lo que sucede en la red, de esta manera poder hacer las diferentes preguntas que nos pueden llevar, a sus diferentes respuestas o conclusiones de lo que está sucediendo, por lo tanto el análisis que se realiza, debe ser por ítems teniendo en cuenta que se debe incluir los días, horas, lugares, redes donde se realiza la captura para el estudio de esta, todo influye a la hora del análisis por que no es lo mismo una red muy utilizada a una que no se utiliza, o de un sitio público a uno privado, una hora donde muchas personas están trabajando a una donde no hay nadie, todo se debe tener en cuenta para el análisis, porque si no se hace no se obtiene el resultado que se espera.

4. OBJETIVOS

4.1 Generales

Este proyecto busca identificar las variables capturadas por la herramienta Wireshark y los drivers de tarjetas inalámbricas Wi-Fi, y a partir de ellas, realizar un análisis exploratorio de datos.

4.2 Específicos

- Identificar las variables capturadas por herramientas de monitoreo Wireshark.
- Identificar la interacción entre estas variables.
- Plantear las preguntas específicas que pueden proveer información acerca del rendimiento y funcionamiento de la red.

5. MARCO TEÓRICO

5.1 Wi-Fi

En los últimos años, podemos darnos cuenta que todo se ha vuelto móvil, ya muchas tecnologías se están volviendo insuficientes, teniendo en cuenta que cada día evolucionan más las cosas, por consecuente hay que ir a la par para poder crecer, de igual manera y así adaptarnos al nuevo estilo de vida, que nos propone esta nueva era. Esta tecnología Wi-Fi, le permite al usuario conectarse sin necesidad de estar en un solo sitio específico para poder navegar, esto le facilita a muchas personas que se deben trasladar en todo momento, realizar sus diferentes labores que no pueden dejar para más adelante, sino que se deben hacer en ese mismo instante. [10] Es una tecnología inalámbrica que se utiliza para conectarse, cambiar información entre diferentes dispositivos, la cual no necesita cableado para poder transmitir datos entre computadores.

Esta tecnología, cada día tiene más fuerza porque es muy práctica, teniendo en cuenta que no va hacer a un lado, al Ethernet por que se llevan muchos años de diferencia, el Wi-Fi cada día tiene más avances y más cobertura, la ventaja que tiene esta tecnología es que varios ordenadores se pueden conectar al mismo tiempo a internet, de tal manera también se pueden interactuar entre sí inalámbricamente. Wi-Fi utiliza el protocolo TCP/IP el cual utiliza internet [11].

5.1.1 Trama 802.11

El estándar 802.11 contiene diferentes tipos de tramas donde cada una tiene un objetivo. Estas tramas atienden funciones específicas del funcionamiento de las redes inalámbricas, a parte de las tramas dedicadas a la transmisión de datos. Estas tramas las Podemos clasificar las tramas según la función que desempeñan. Tenemos tramas de datos, las que transportan la información de capas superiores, tramas de gestión que permiten mantener las comunicaciones y tramas de control

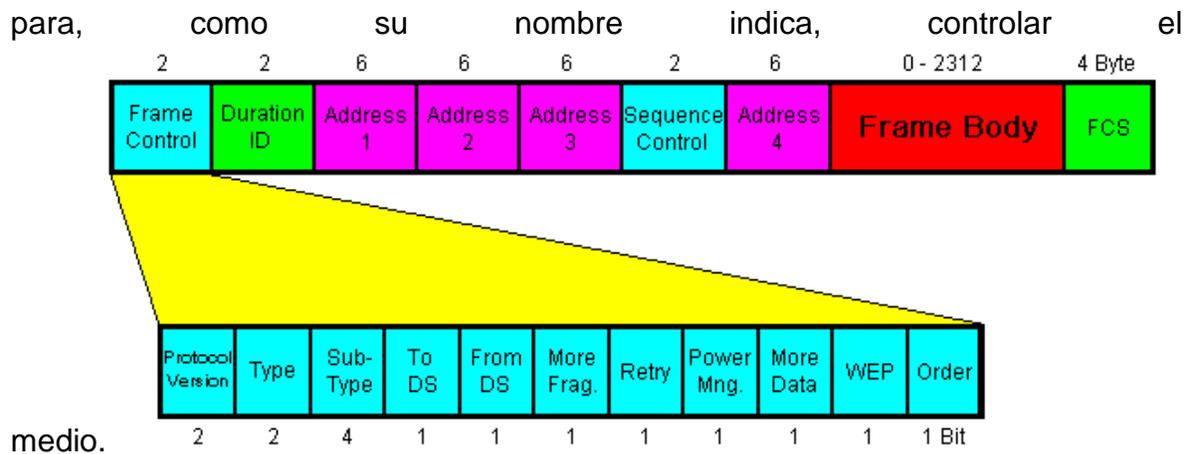


Ilustración 1 formato de trama de datos (Tomado de <http://einstein.informatik.uni-oldenburg.de/rechnernetze/frame.htm>)

5.1.2 Tipos de tramas

La siguiente tabla muestra algunos tipos y subtipos de tramas que son utilizadas en las redes inalámbricas.

Tabla 1 tipos de tramas estándar 802.11

Tipo	tipo(binario)	Subtipo	Algunos nombres
gestión	00	0	Association request
		1	Association response
		100(4)	probe request
		101(5)	probe response
		1000(8)	beacon
control	01	1011	Rts
		1100	Cts
		1101	Ack
datos	10	0	Datos

5.1.3 Tramas de gestión

Las tramas de gestión se encargan de establecer y terminar conexiones entre los access points y las estaciones. Algunas de las tramas de gestión más comunes son:

Trama beacon

Los access points transmiten tramas beacon frecuentemente para mostrar su presencia a las diferentes estaciones en su radio de cobertura. Las estaciones obtienen access points disponibles por medio de las tramas beacon en los diferentes canales, estas tramas también contienen características de las redes.

Trama Association Request

Es la trama usada por la estación para solicitar asociación a un access point. [10, p. 127]. Después de estar asociada, la estación empieza a recibir todos los servicios de red.

Trama Association Response

Es la trama usada por el access point para responderle a una estación que su solicitud de asociación ha sido otorgada. Después de estar asociada, la estación empieza a recibir todos los servicios de red. Como parte de la respuesta, el punto de acceso asigna una ID de asociación. La forma en que un punto de acceso asigna el ID de asociación depende de la implementación [10, p. 141].

Trama Probe Request

Es usada por las estaciones para escanear el entorno con el fin de descubrir access points. Esta trama es enviada como un broadcast y los access point disponibles responden a esta trama con la trama probe response. [10, p. 138].

Trama Probe Response

Es usada por los access points para responder a tramas probe request. Estas tramas contienen la información con los detalles de operación del access point. [10, p. 138]

5.1.4 Trama de control

La trama 802.11 de control la cual se utiliza para administrar el envío de los datos entre las diferentes tramas. Algunos ejemplos de tramas de control se explican a continuación [1, p. 70].

Formato de trama request to send (RTS)

Es utilizada para minimizar las colisiones en el caso de dos estaciones asociadas a un mismo punto de acceso mutuamente pero fuera del rango de cobertura. La estación transmite una trama RTS para empezar el dialogo de inicio de transmisión de una trama [3, p. 102].

Formato de la trama clear to sent (CTS)

Las estaciones usan las tramas CTS para dar respuesta a una trama RTS para desocupar el canal de transmisión. Las tramas CTS tiene un valor de tiempo el cual el resto de estaciones deja de enviar paquetes por una determinada fracción de tiempo hasta que termina de transmitir [5, p. 104].

Tramas acknowledgement (ACK)

Las tramas ACK confirman la llegada de una trama. En caso de no llegar el ACK el emisor reenvía la trama de datos [5, p. 105].

5.1.5 Trama de datos

Esta trama es la encargada de transportar la información de las capas superiores la información se envía por trozos de información, en la cual contiene dirección, protocolos e información, en esta trama se tiene en cuenta el número de bytes que compone un paquete de esta manera se delimita lo que se va a enviar.

5.2 Análisis exploratorio de datos

¿Qué es?

Corresponde al análisis estadístico que se realiza a los datos que son recolectados, mediante una investigación que se hace durante un determinado tiempo, por un grupo de personas de un tema específico. La información recolectada se representa a través de gráficos, sobre los cuales se realiza dicho análisis. En concordancia se extraen datos estadísticos que permiten identificar mejor las distribuciones de probabilidad, que permiten determinar a través de la estadística descriptiva el comportamiento de la información analizada conjuntamente o de forma separada. Cuando se analizan las variables independientemente, es posible identificar si los datos arrojan información importante o relevante y en consecuencia dar conclusión adecuada al proceso realizado [3, 4].

¿Para qué sirve?

Para dar respuesta a las diferentes preguntas, que cada individuo tiene a la hora de empezar un proyecto, y orientar las diferentes respuestas al mundo real para que pueda ser entendida, no por las personas que realizan los diferentes estudios sino también a personas del común. Por lo tanto, cada dato, o información que nos pueda dar la herramienta a utilizar se debe revisar con detalle, para decidir qué información de desempeño se va utilizar.

De acuerdo con lo anterior surgen las siguientes inquietudes: ¿Al analizar las redes se puede determinar que tan bien estas están funcionando?, ¿Hay algunas fallas a la hora de transmitir?, ¿Cuáles son las interrupciones que se presentan?, ¿Qué información adicional arrojan los diferentes datos que son capturados en las diversas redes, tiempos y lugares? ¿Existe relación entre las variables? ¿Existe dependencia entre variables?

5.2.1 Tipos de variables

Variables cualitativas

También llamadas variables categóricas o atributos, son aquellas que no necesitan números para expresarse; es decir toman valores no susceptibles de realizar operaciones aritméticas sobre ellas. Son las que te muestran las cualidades o atributos de tu variable.

Variables cuantitativas

Cuantitativas o numéricas, son aquellas que necesitan números para ser expresadas, permiten hacer cálculos aritméticos. Estas variables pueden discretas o continuas.

Variables discretas

Número finito de valores entre dos puntos. Esta variable discreta habitualmente es un número entero.

Variables continuas

Hacen referencia a aquellas que solamente pueden expresarse en el campo de los números reales.

5.2.2 Caracterización de variables

Histogramas

El histograma es una gráfica de barras que muestra la frecuencia de un determinado conjunto de datos, donde es usada para obtener una percepción del comportamiento de la variable, o también ver la tendencia central. El histograma es aplicable a variables continuas y discretas.

PDF (función de densidad de probabilidad)

Esta función proporciona la probabilidad de que la variable aleatoria tome algún valor de cierto intervalo. Esta función debe cumplir con las siguientes propiedades: Debe ser un número positivo, ≥ 0 . La integral de todo el dominio debe ser equivalente a 1.

CDF (función de distribución acumulada)

Esta función permite determinar la probabilidad acumulada de una variable aleatoria, proporciona la probabilidad de que la variable aleatoria tome un valor menor o igual a cierto valor.

Esta función toma un valor de probabilidad entre 0 y 1.

PMF (función de masa de probabilidad)

Se aplica a variables discretas. Es la probabilidad de que una variable aleatoria tome un valor determinado.

KDE (kernel density estimation)

Estimadores de densidad de Kernel es una forma de mostrar la función de densidad de probabilidad de una variable aleatoria, la cual no muestra una figura fija, sirve para suavizar los datos donde se presenta mayor número, se realiza en una determinada cantidad de datos. En pocas palabras es una forma suave de mostrar el histograma.

5.3 Relación entre variables

5.3.1 Gráficas

Dispersión

Diagrama de dispersión sirve para mostrar gráficamente la relación que existe entre dos variables cuantitativas.

Hexagonal Binning

Binning es una técnica alternativa para visualizar la densidad cuando se trabaja con grandes conjuntos de datos, de otra forma es el número de puntos que caen dentro de un rectángulo o hexágono en una superficie con rejilla. Hay otros métodos para realizar binning y otras formas de interpolar los valores entre puntos, pero el binning hexagonal es una forma eficaz de agregar y visualizar sus datos.

5.4 Correlación

5.4.1 Método Pearson

Este coeficiente de correlación se utiliza para calcular el vínculo que puede existir entre dos variables cuantitativas que tienen distribución normal [2].

Se debe tener en cuenta las siguientes propiedades:

Si es 1, existe una correlación perfecta positiva.

Si $0 < x < 1$, existe una correlación positiva.

Si $-1 < x < 0$, existe una correlación perfecta negativa [3].

Si $x = 0$, no existe relación lineal. Esto no implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.

5.4.2 Método Spearman

Este coeficiente de correlación se utiliza para calcular el vínculo que puede existir entre dos variables cuantitativas continuas, los cuales deben de estar ordenados por orden para utilizar los rangos o intervalos [4]. No requiere que las variables tengan distribución normal.

Se debe seguir las siguientes propiedades para especificar qué tipo de correlación es:

Si $x < 1$, existe una asociación positiva.

Si $-1 < x$, existe una asociación negativa

Si $x = 0$ se dice que no existe correlación, ni dependencia. [5]

5.4.3 Kendall

El coeficiente de Kendal se utiliza cuando los datos no tienen una distribución normal, de esta manera poder observar si tiene alguna relación lineal entre sí.

El coeficiente de Kendall puede variar entre -1 y 1 el valor positivo indica una relación directa y si es negativa indica relación inversa, entre más alta sea la magnitud, más fuerte es la asociación.

Si $0 < x < 1$, existe una relación directa.

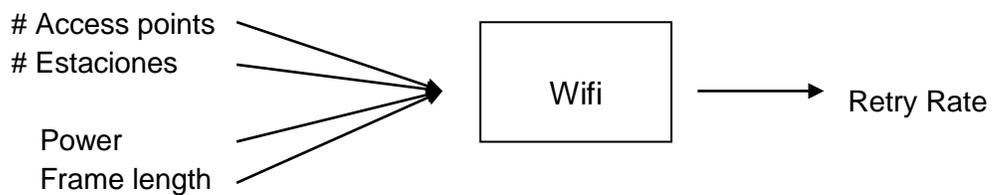
Si $-1 < x < 0$, existe una relación inversa.

Este coeficiente y sus valores son utilizados para elegir entre dos hipótesis opuestas, basándose en sus datos.

6 ANÁLISIS EXPLORATORIO DE DATOS APLICADO A CAPTURAS DE WIRESHARK DE UNA RED WI-FI

6.1 Definición de variables

En el proceso exploratorio se definen cinco variables para el análisis a una red inalámbrica, variables que se puede relacionar directa e indirectamente con la tasa de retransmisión. En la siguiente ilustración da a conocer las variables que interfieren en una red Wi-Fi.



6.2 Proceso de obtención de datos y variables

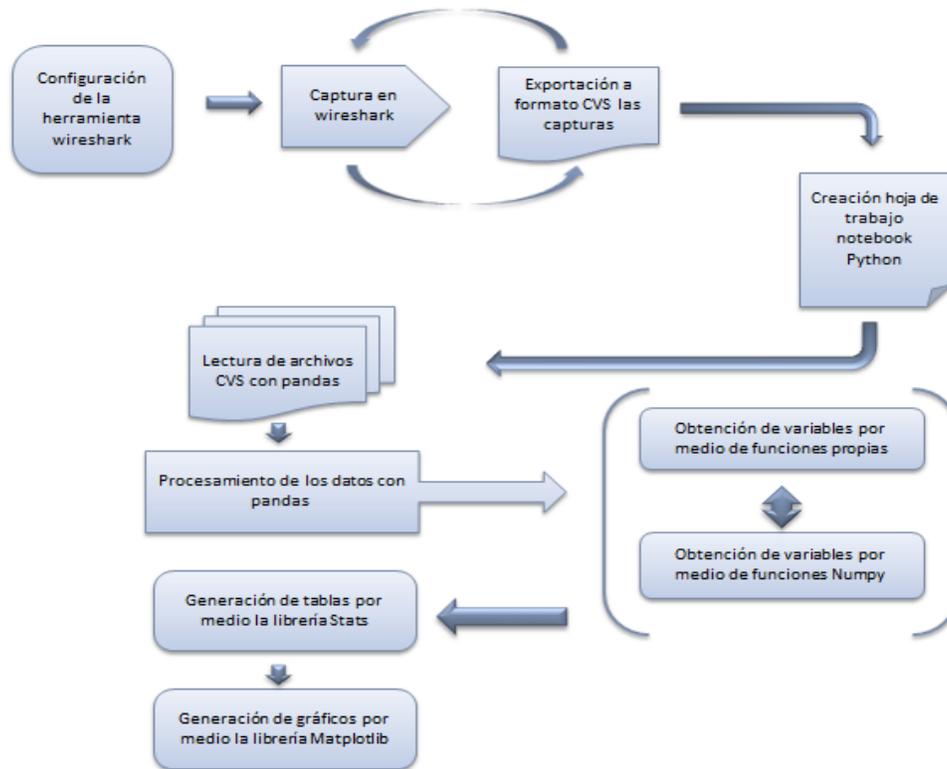


Ilustración 2 proceso de análisis exploratorio de datos

En la gráfica anterior se describe el proceso que se lleva a cabo para el análisis exploratorio de los datos en la obtención de las variables, desde la captura a la información hasta generación de gráficos para la lectura de resultados.

6.3 Obtención de variables

Access points

En la obtención de la variable access points se inicia con la lectura de la captura, en este se hace uso de la función `read_csv` de pandas, dado que este formato es procesable en Python y se puede exportar en Wireshark.

```
datosM=pd.read_csv('C:\Users\yearr\Documents\capturas\hora', index_col=0)
```

Para calcular el número de access points, se hace uso del índice de la tabla para crear los intervalos.

```
datosB=datosB.reset_index(level=1)
```

Al generar los intervalos se procede a realizar la agrupación de los datos, por cada cien tramas por medio del siguiente código.

```
cantidad= len(datosB)
Tintervalos = np.arange(0,cantidad, 100)
datosB['rangos'] = pd.cut(datosB['index'], Tintervalos)
```

La función cantidad () para el llenado de la tabla access points, la función recibe como parámetro un data frame y retornar el número de access points, esto lo realiza agrupando por dirección “groupby” y por ESS, con el objetivo de tomar solo las direcciones de los AP, el conteo de este con count().

```
def cantidad(x):
    x1=x.groupby(['Transmitter address','ESS capabilities']).count()
    x1=x1.reset_index()
    x1=x1.groupby('ESS capabilities').count()
    ap=x1['Transmitter address'].sum()
    return ap
```

Con los datos agrupados y la función cantidad creado se calcula el porcentaje de retransmisión, primero se crea una lista AP[] cual va a contener el resultado de la función de cantidad por cada dataframe de los datos agrupados por cada cien tramas, y al tener el porcentaje de todos los intervalos se convierte la lista en un dataframe con el nombre de la columna “Número_AP” y el index de la tabla original.

```
AP=[]
indice=datosB.groupby('rangos').count()
t=datosB.groupby('rangos')
for k, gp in t:
    AP.append(cantidad(gp))
estacionesB=pd.DataFrame(AP,columns=['Numero_AP'], index=indice.index)
```

Estaciones

En la obtención de la variable estaciones se inicia con la lectura de la captura, en este se hace uso re “read_csv” como parámetro el directorio del archivo de la librería de pandas y se asigna al dataframe datosM.

```
datosM=pd.read_csv('C:\Users\yearr\Documents\capturas\hora', index_col=0)
```

Para calcular el número de estaciones, se hace uso del index de la tabla para crear los intervalos.

```
datosB=datosB.reset_index(level=1)
datosB=datosB.reset_index(level=1)
```

Al generar los intervalos se procede a realizar la agrupación de los datos, por cada cien tramas por medio del siguiente código.

```
cantidad= len(datosB)
Tintervalos = np.arange(0,cantidad, 100)
datosB['rangos'] = pd.cut(datosB['index'], Tintervalos)
```

Para el conteo se crean algunos métodos, primero el método cantidad () para el llenado de la tabla estaciones, la función recibe como parámetro un data frame y retornar el número de estaciones, esto lo realiza agrupando por dirección “groupby” y el conteo de este.

```
def cantidad(x):
    sta=x.groupby('Transmitter address').count()
    station=len(sta)

    x1=x.groupby(['Transmitter address','ESS capabilities']).count()
    #x1=x.groupby(['Source','ESS capabilities']).count()
    x1=x1.reset_index()
    x1=x1.groupby('ESS capabilities').count()
    ap=x1['Transmitter address'].sum()

    estaciones=station-ap

    return estaciones
```

Con los datos agrupados y la función cantidad creado se calcula el porcentaje de retransmisión, primero se crea una lista ST[] cual va a contener el resultado de la función de cantidad por cada dataframe de los datos agrupado por cada cien tramas, y al tener el porcentaje de todos los intervalos se convierte la lista en un dataframe con el nombre de la columna “Número_estaciones” y el index de la tabla original.

```
ST=[]
indice=datosB.groupby('rangos').count()
t=datosB.groupby('rangos')
for k, gp in t:
    ST.append(cantidad(gp))
estacionesB=pd.DataFrame(ST,columns=['Numero_estaciones'],
index=indice.index)
```

Potencia

Se extraer la columna signal strength (dBm) dado que estas contienen las potencias con la que fue recibida la trama, para observar su comportamiento. Teniendo en cuenta la potencia mínima y la máxima se realiza el cálculo para el tamaño de los intervalos con la fórmula de Sturges

Para la distribución de frecuencia, primero creamos un rango para las categorías. Por medio de la función `arange()`, con el valor mínimo y máximo.

```
Tintervalos = np.arange(-100, 0., 10) (6)
```

Al tener los intervalos se asignan los datos en cada contenedor con la función `cut()`

```
pot = pd.cut(hola['Signal strength (dBm)'], Tintervalos) (7)
```

Por último, hacemos el recuento de los contenedores, para la tabla de frecuencia.

```
tabla_pot = pd.value_counts(pot) (8)
```

Longitud

Se hace la lectura de los datos con el método `read_csv` y se guarda en el dataframe en `datosM`

```
datosM=pd.read_csv('C:\Users\yearr\Documents\capturas\hora', index_col=0)
```

Creación de los intervalos de acuerdo al frame body dado que el máximo es 2312 bytes y el mínimo es de 0 bytes, a partir de la formula sturges se determina el número de intervalos en los que se pueden agrupar las diferentes longitudes.

Se toma la longitud máxima y se divide en el número de intervalos, para determinar el tamaño de la longitud del intervalo.

```
Tintervalos = np.arange(0, 2312., 48)
```

Luego cortamos los datos en cada contenedor

```
potenciaB = pd.cut(datosB['Length'], Tintervalos)  
potenciaM = pd.cut(datosM['Length'], Tintervalos)
```

Por último, hacemos el recuento de los contenedores para armar la tabla de frecuencia.

```
tablaB = pd.value_counts(potenciaB)
```

```

tablaM = pd.value_counts(potenciaM)
tablaPB=pd.DataFrame(tablaB)
tablaPB=tablaPB.rename(columns={'Length':'Longitud-conectividad
eficiente'})
tablaPM=pd.DataFrame(tablaM)
tablaPM=tablaPM.rename(columns={'Length':'Longitud-conectividad
deficiente'})

```

Retry

En la obtención del porcentaje de retransmisión se inicia con la lectura de la captura, en este se hace uso re “read_csv” de pandas.

```

datosM=pd.read_csv('C:\Users\yearr\Documents\capturas\hora', index_col=0)

```

Para calcular el porcentaje de retransmisión, se hace uso del index de la tabla para crear los intervalos.

```

datosB.Retry=datosB.Retry.replace(['Frame is not being
retransmitted','Frame is being retransmitted'],[0,1])
datosB=datosB.reset_index(level=1)
datosB=datosB.reset_index(level=1)

```

En la creación del data frame de la variable retry se es necesario calcular que porcentaje de retransmisión por dataset dado el total calculado con la función total (), para ello se crea la función porcentaje ().

```

cantidad= len(datosM)
Tintervalos = np.arange(0,cantidad, 100)
datosM['rangos'] = pd.cut(datosM['index'], Tintervalos)
finalPorcentaje=datosM.groupby('rangos').sum()

tablaRM=pd.DataFrame(finalPorcentaje.Retry)

```

6.4 Caracterización

6.4.1 Frecuencias esperadas

Para la creación de la tabla de valores esperados se realiza por medio del método *stats.contingency.expected_freq* de la librería *scipy* de *Python*, la siguiente fracción de código genera la tabla de valores esperados en el estudio de las correlaciones

```

table =stats.contingency.expected_freq ( tablafinal)

```

```

table=pd.DataFrame(table, index=tablafinal.index, columns= tablafinal.
columns )
table

```

La generación de la tabla de frecuencias, luego de realizar los conteos de los datos agrupados en cada una de las variables, se calcula las respectivas frecuencias relativas y absolutas, el siguiente código describe la generación y cálculo de las frecuencias para la variable potencia.

Potencia Absoluta

```
tablaPB["potenciaA"]=tablaPB['Potencia-conectividad eficiente'].cumsum()
```

frecuencia relativa

```
tablaPB['frecuenciaR']=tablaPB['Potencia-conectividad eficiente']/len(datosB)
```

frecuencia absoluta

```
tablaPB["frecuenciaA"]=tablaPB.frecuenciaR.cumsum()
```

porcentaje relativo

```
tablaPB["PorcentajeR"]=tablaPB['Potencia-conectividad eficiente']/((tablaPB['Potencia-conectividad eficiente'].sum())*1.0)*100
```

porcentaje absoluto

```
tablaPB["PorcentajeA"]=tablaPB.PorcentajeR.cumsum()
```

```
tablaPB
```

	SSI Signal	potencia A	frecuenciaR	frecuencia A	PorcentajeR	PorcentajeA
(-70, -60]	46331	289329	0.138099	0.862407	14.022785	87.569840
(-50, -40]	22157	311486	0.066044	0.928451	6.706154	94.275994
(-60, -50]	12394	323880	0.036943	0.965394	3.751233	98.027228
(-30, -20]	3735	327615	0.011133	0.976527	1.130455	99.157683
(-100, -90]	2656	330271	0.007917	0.984444	0.803879	99.961562
(-20, -10]	127	330398	0.000379	0.984822	0.038438	100.00000

Ilustración 3 tabla de frecuencias

6.4.2 Gráficas

Histogramas

Una de las mejores formas de describir una variable es por medio de un reporte de un conjunto de datos, los valores que aparecen y de veces que aparece cada valor en un determinado tiempo, llamado de otra forma distribución de la variable. La mejor forma de representar una distribución es un histograma, gráfica que muestra la frecuencia, cantidad de veces que la variable toma un valor. Para las caracterizaciones de las variables seleccionadas se realiza uso del histograma,

por medio de la función hist() que ofrece la librería matplotlib, la siguiente línea de código realiza la gráfica del data frame access points.

```
accespoint.plot.hist()
```

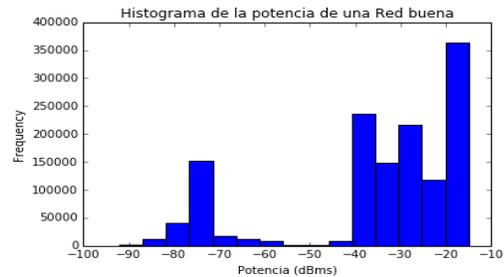


Ilustración 4 gráfica de histograma

PDF

La función de densidad de probabilidad se define sólo para variables aleatorias continuas. En el PDF lo que es más se resalta es saber que los valores que se dan son un rango de posibles valores que da la probabilidad de la variable aleatoria que cae dentro de ese rango, el siguiente código es el utilizado para la generación de las gráficas PDF.

```
estacionesB.Numero_estaciones.hist(cumulative=True, normed=1, bins=30),  
p.title('PDF del número STA de una Red buena')
```

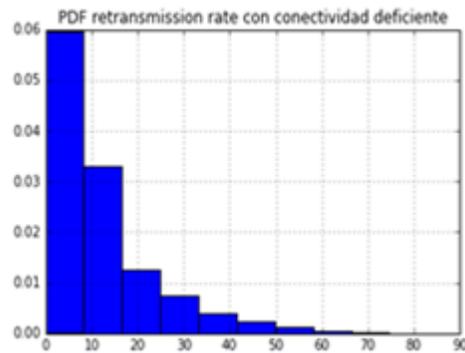


Ilustración 5 gráfica del PDF

CDF

Para determinar la probabilidad de que una respuesta sea menor o igual que cierto valor. Se usa la función de probabilidad acumulada, esta función es aplicable tanto a las variables discretas como continuas. Para la caracterización de las variables se hace uso de esta función, el siguiente código es el utilizado para la generación de las gráficas CDF.

```
datosB['SSI Signal'].hist(cumulative=True, normed=1,
bins=30)
p.title('CDF de la potencia de una Red buena')
p.xlabel('Potencia (dBm)')
```

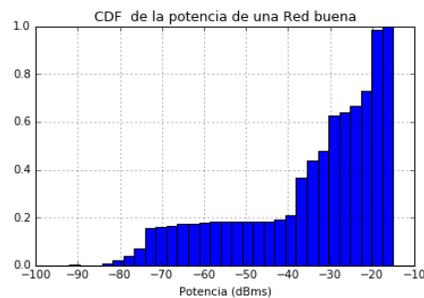


Ilustración 6 gráfica de CDE

KDE

Para la estimación del PDF, que describe la probabilidad de cada valor específico que puede tener una variable, se hace uso del método de estimación de kernel KDE por medio de la función `kde()`, Scipy proporciona la implementación del KDE

```
accespoint.plot.kde() (16)
```

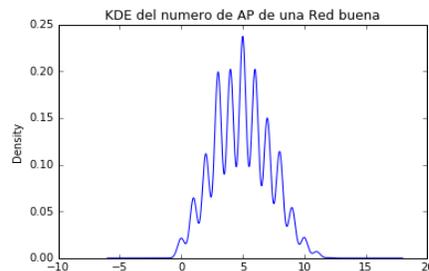


Ilustración 7 gráfica de KDE

6.5 Relación entre variables

6.5.1 Gráficas

Hexagonal binning

Binning es una gráfica con la que podemos visualizar la densidad cuando se trabaja con grandes conjuntos de datos, de otra forma es el número de puntos que caen dentro de un rectángulo o hexágono en una superficie con rejilla. Hay otras formas de interpolar los valores entre puntos, pero el binning hexagonal es una forma eficaz de agregar y visualizar sus datos, y Python nos ofrece esta visualización con el parámetro `kind` igual a `hexbin` de la función `plot()`

```
tablafinal.plot(kind='hexbin', x='Retry', y='min-potencia', gridsize=25);  
p.title("LONGITUD MINIMA DONDE SE PRESENTA MAYOR PORCENTAJE DE  
(22) RETRASMISION")  
p.ylabel("LONGITUD MINIMA (DBM) ")
```

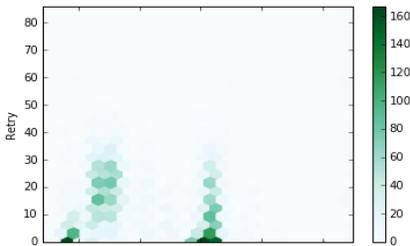


Ilustración 8 gráfica de dispersión hexagonal

Dispersión

En la generación de las gráficas en Python se hace uso de la propiedad `“alpha”` de la función `plot`, esta se le asignan valores para el grado de tonalidad de acuerdo a la frecuencia del valor. En la correlación se le asigna un valor de 0.2, esto significa que tendremos 5 niveles de tonalidad, de tal manera que cuando un dato no se superpone a otro, lo veremos el punto en un tono claro, cuando dos datos se superpongan el punto se visualizará más oscuro (en teoría el doble de oscuro que el anterior), si cinco o más datos se superponen entre sí, ese punto será visualizado como el más oscuro en el gráfico. Y para la generación de la gráfica de dispersión se realiza por medio de la propiedad `“kind”` asignándole `“scatter”`.

```
tablafinal.plot(kind='scatter', x='Retry', y='promedio-potencia',  
alpha=0.2);
```

```

p.title("LONGLITUD PROMEDIO RESPECTO AL PORCENTAJE DE RETRASMISION
PRESENTE EN CADA INTERVALO")
p.ylabel("LONGLITUD PROMEDIO (DBM)")
p.xlabel("PORCENTAJE DE RETRASMISION DE CADA 100 TRAMAS ")

```

(21)

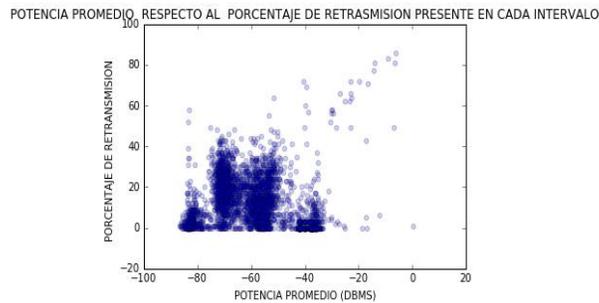


Ilustración 9 gráfica de dispersión

6.5.2 Correlación

El siguiente código genera la tabla que describe la correlación de Spearman Rank que es realizada a dos conjuntos de datos, en el análisis de correlación se es aplicado a la retransmisión “retry”, longitud y potencia. Para la correlación se utiliza el método de la librería Scipy stats.spearmanr() que recibe como parámetros las dos matrices que contiene las variables y como resultado se obtiene la matriz de correlación de Spearman o coeficiente de correlación y p-valor para la prueba de hipótesis de la existencia de una correlación entre dos variables.

```

tablasr=tablafinal.Retry
tablasm=tablafinal['moda-length']
tablasx=tablafinal['max-length']
tablasn=tablafinal['min-length']
tablasp=tablafinal['promedio-length']
tablaspearman={'Corr Spearman Rank(co-
corr,pvalue)':pd.Series([stats.spearmanr(tablasr,tablasm),
stats.spearmanr(tablasr,tablasx),
stats.spearmanr(tablasr,tablasp),
stats.spearmanr(tablasr,tablasp)],
index=[ 'Retry-moda_length',
'Retry-max_length',
'Retry--min_length',
'Retry--promedio_length'])}
tablaspearman=pd.DataFrame(tablaspearman)
tablaspearman

```

(25)

	Corr Spearman Rank(co-corr,p-value)
Retry-moda_potencia	(-0.306460761112, 1.32881136816e-288)
Retry-max_potencia	(-0.242685504572, 1.93193892178e-178)
Retry--min_potencia	(-0.489064913928, 0.0)
Retry--promedio_potencia	(-0.489064913928, 0.0)

Ilustración 10 tabla de correlaciones Spearman rank

Código para el cálculo de la correlación por el método Pearson, método Spearman y correlación normal

```
corrPearson=tablafinal.corr(method='pearson', min_periods=1)
corrPearson=corrPearson.drop('Retry',1)
corrPearson=corrPearson.loc['Retry']
corrPearson=pd.DataFrame(corrPearson)
corrPearson.rename(columns={'Retry':'Retry - metodo Pearson'},      (20)
inplace=True)
correlaciones=corrPearson.join(corrNormal)
correlaciones=correlaciones.join(corrSpearman)
```

	Retry- metodo Normal	Retry - metodo Spearman
max-potencia	-0.018585	-0.058598
min-potencia	0.036935	0.027970
promedio-potencia	0.040367	0.159020
moda-potencia	0.041441	0.159560

Ilustración 11 tabla de correlaciones normal y Spearman

7 RESULTADOS Y DISCUSIÓN

7.1 Caracterización

En la siguiente sección se describe la caracterización de las variables access points, estaciones, potencia, longitud, retry.

7.1.1 Caracterización de la variable access points

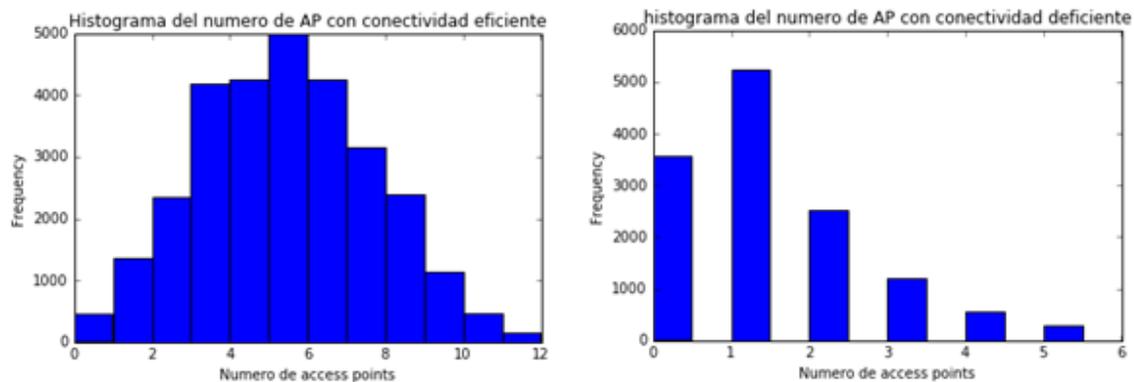
La siguiente tabla es una parte del conteo del número de access points por cada cien tramas.

Tabla 2 Conteo de número de access points por cada 100 tramas

	Numero_AP
rangos	
(0, 100]	4
(100, 200]	4
(200, 300]	4
(300, 400]	4
(400, 500]	5
(500, 600]	4
(600, 700]	5
(700, 800]	6
(800, 900]	4
(900, 1000]	5
(1000, 1100]	5
(1100, 1200]	4
(1200, 1300]	5
(1300, 1400]	6
(1400, 1500]	5
(1500, 1600]	5
(1600, 1700]	5
(1700, 1800]	4
(1800, 1900]	5

	Numero_AP
rangos	
(1900, 2000]	4
(2000, 2100]	4
(2100, 2200]	4
(2200, 2300]	4
(2300, 2400]	4
(2400, 2500]	4
(2500, 2600]	4
(2600, 2700]	4
(2700, 2800]	4
(2800, 2900]	4
(2900, 3000]	5
...	...

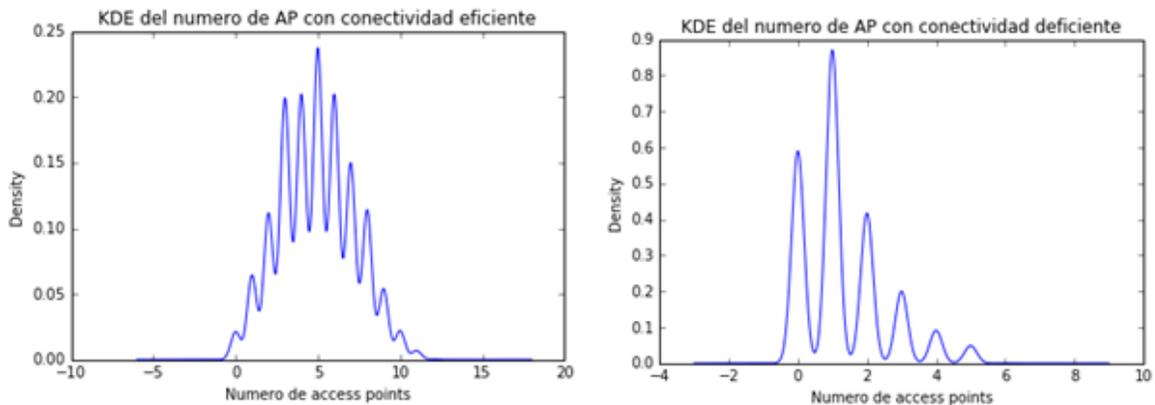
A continuación de muestra las gráficas generadas a partir del análisis exploratorio a una red con conectividad eficiente y una red con conectividad deficiente.



Gráfica 1 Histograma V.AP conectividad eficiente y conectividad deficiente

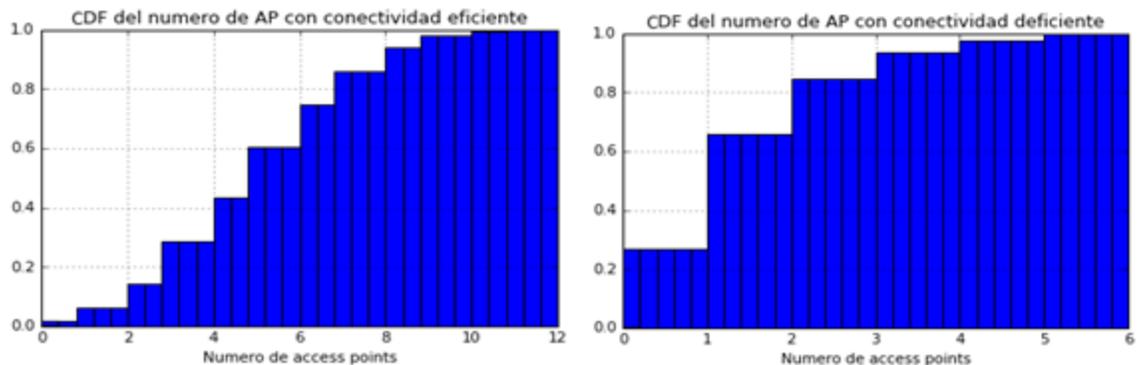
La gráfica de histogramas (gráfica 1), tanto de una red con conectividad eficiente y deficiente hace parte de los análisis realizados en la tabla de frecuencia, esta gráfica aplica para observar el número de access points presentes en las redes de la universidad, en el histograma de una red con conectividad eficiente se muestra un rango de access points de 0 a 12, donde existe un mayor frecuencia

en 5, como resultado obtenemos que existe una mayor presencia durante las transmisiones de 4 a 6 access points, a comparación de la red con conectividad deficiente que maneja un rango de access points de tan sólo de la mitad de la red eficiente con mayor presencia de un AP durante la transmisión.



Gráfica 2 KDE conectividad eficiente y conectividad deficiente

En las gráficas de densidad para el número de access points en una red con conectividad eficiente tiende a ser una distribución de probabilidad normal con una densidad de 0.25 con 5 access points, a diferencia de la red con conectividad deficiente que se asemeja a una distribución de probabilidad de Poisson, con una densidad de 0.9 para un acces point mayor que la red eficiente.

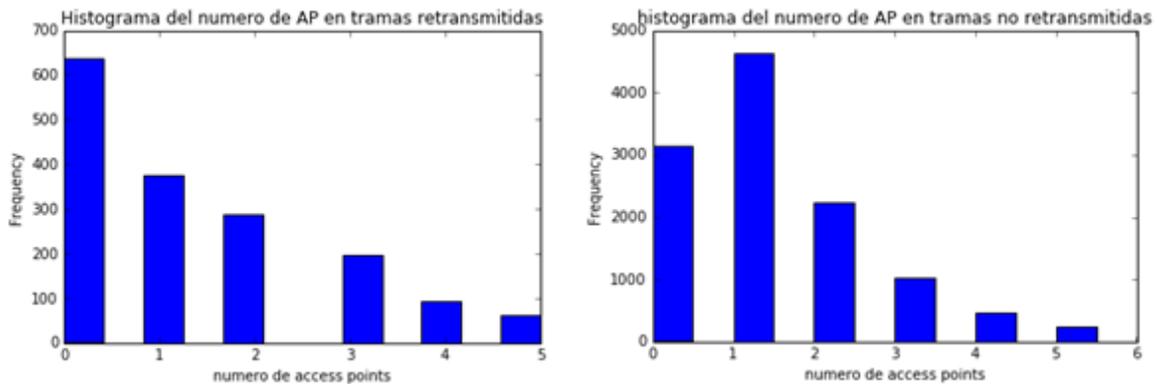


Gráfica 3 CDF V.AP conectividad eficiente y conectividad deficiente

En la gráfica del CDF de la red con conectividad eficiente la probabilidad de que el número de access points sean menor a 11 es de 1 y menor a 2 AP la probabilidad es baja menos del 0.1, a comparación de la red deficiente donde la probabilidad de 1 se da cuando el máximo número de access points es de 5, y de que sean menores o iguales a dos es más del 0.6.

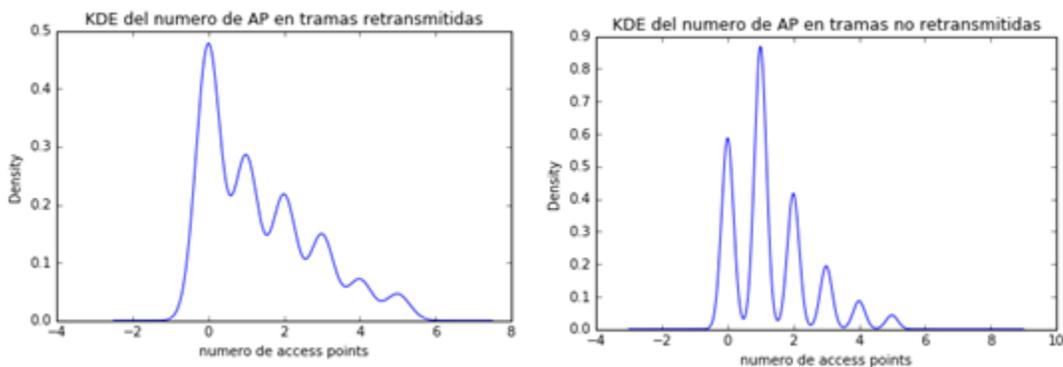
Comportamiento de la variable access points con respecto a las tramas retransmitidas y no retransmitidas

Las siguientes gráficas son generadas a partir de las tramas no retransmitidas y retransmitidas de una captura.



Gráfica 4 Histograma de V. número de AP conectividad deficiente con respecto a tramas retransmitidas, no retransmitidas

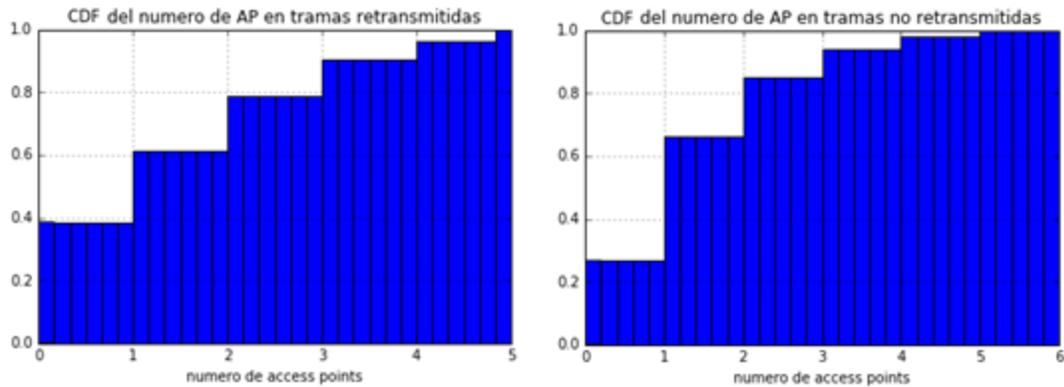
El histograma del número de access points con respecto a las tramas retransmitidas nos muestra la mayor frecuencia en 0 access points esto nos indica que la presencia de access points es baja, a comparación de las tramas no retransmitidas, donde no muestra la presencia de un acces point continuamente, en los dos casos las frecuencias más altas se dan para menos de 2 access points.



Gráfica 5 KDE de número DE AP conectividad deficiente con respecto a tramas retransmitidas y no retransmitidas

La densidad que se tiene entre 0 y 2 estaciones tanto para las tramas no retransmitidas como las retransmitidas son las más altas, en las tramas

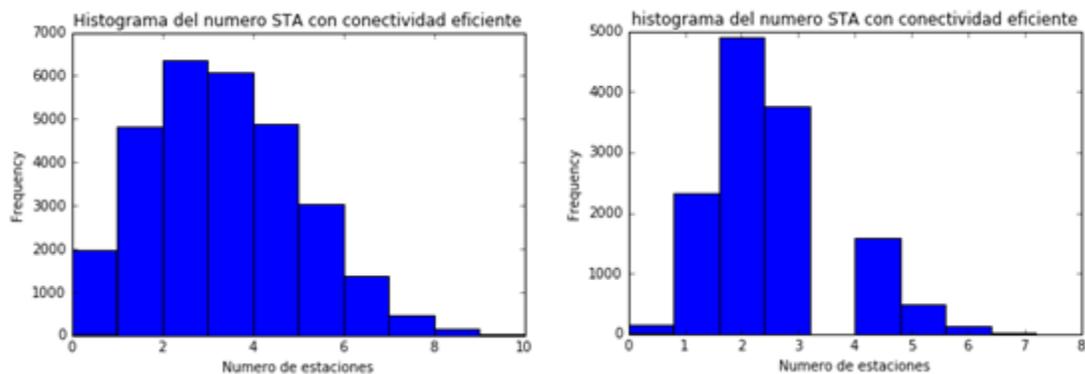
retransmitidas la densidad se acerca a 0.5 y en las tramas no retransmitidas se acerca a 0.9, la distribución de probabilidad a la que se asemejan es a la Poisson.



Gráfica 6 CDF del número DE AP conectividad deficiente con respecto a tramas retransmitidas y no retransmitidas

Las gráficas anteriores muestran una probabilidad similar, esto se da con access points mayores a 2, dado que la probabilidad de encontrar menos de 2 access points es mayor a 0.6 para ambos casos al igual la probabilidad de encontrar menos de 3 y 4 access points, la gran diferencia se presenta en encontrar 0 access points ya que en las tramas retransmitidas la probabilidad es cercana a 0.4 y en las no retransmitidas la probabilidad es menos de 0.3.

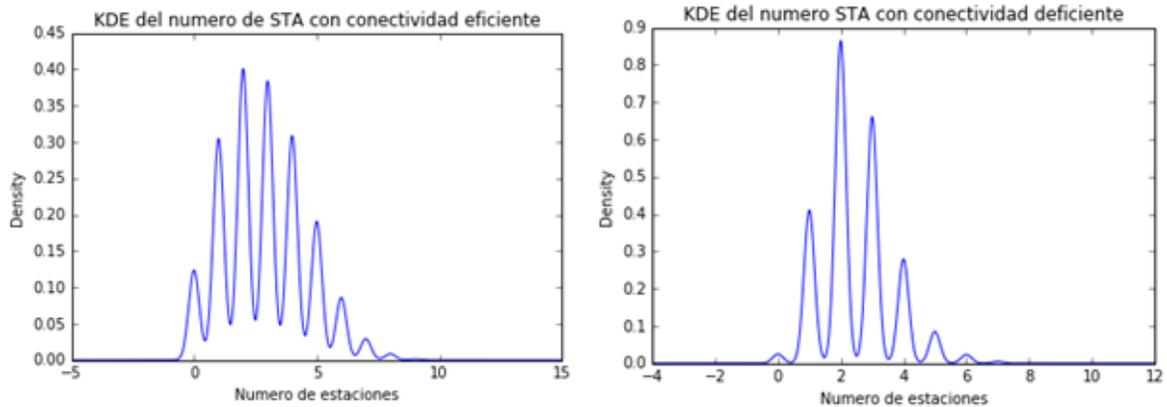
7.1.2 Caracterización de la variables estaciones



Gráfica 7 Histograma V.STA conectividad eficiente y conectividad deficiente

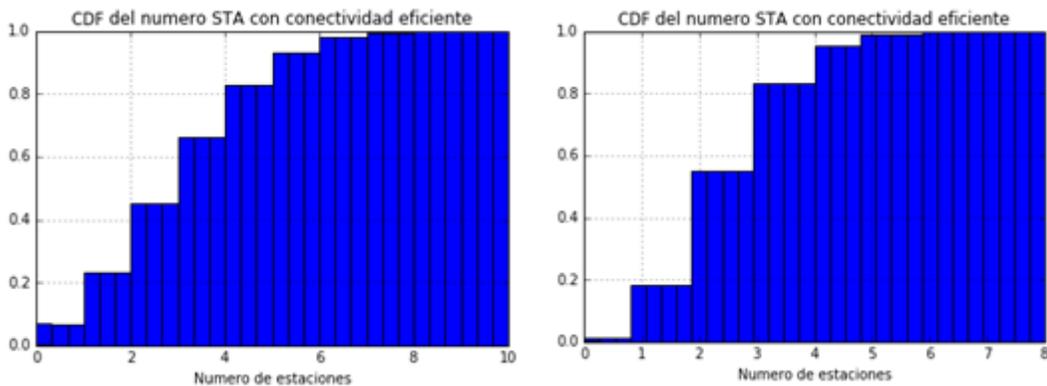
En los histogramas anteriores se observa que el número de estaciones para ambas redes son similares, teniendo en la red eficiente un máximo de 10 estaciones y en la red deficiente 8 estaciones, a igual el número de estaciones

constantes durante toda las transmisiones, con un intervalo de 2 a 4 estaciones con mayor frecuencia en las dos redes, destacando que 2 estaciones es la cantidad con más frecuencia.



Gráfica 8 KDE V.STA conectividad eficiente y conectividad deficiente

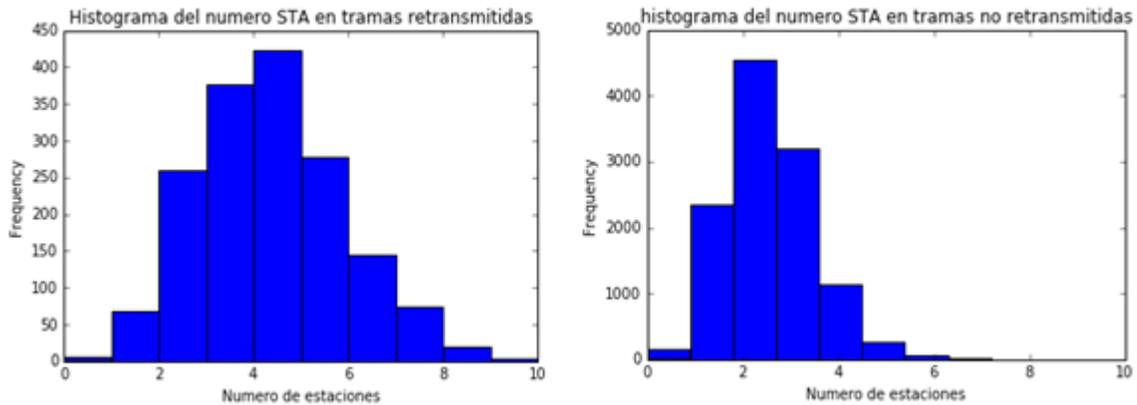
Las gráficas anteriores muestran similitud al tender a ser una distribución de probabilidad binomial, para el caso de la red con conectividad eficiente se tiene la mayor densidad en 0.4 para un número de estaciones comprendidos entre 2 y 3, igual a la red con conectividad deficiente, pero con una densidad mayor.



Gráfica 9 CDF V.STA conectividad eficiente y conectividad deficiente

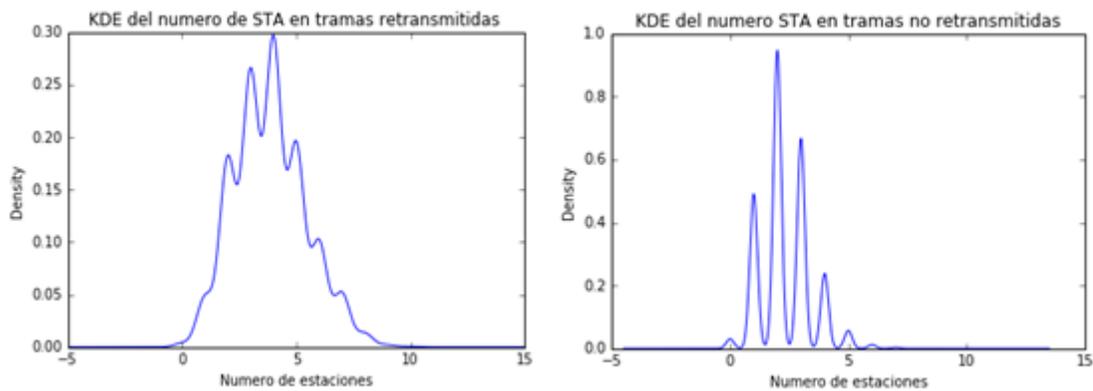
La probabilidad que se muestran en los CDF existe una gran variación al número de estaciones, en la red eficiente la probabilidad de que el número de estaciones sean menor a 4 es mayor de 0.6 y en la red deficiente es mayor al 0.8, la probabilidad máxima se da similar en las dos redes donde el número de estaciones casi siempre es menor a 8

Comportamiento de la variable estaciones con respecto a las tramas retransmitidas y no retransmitidas



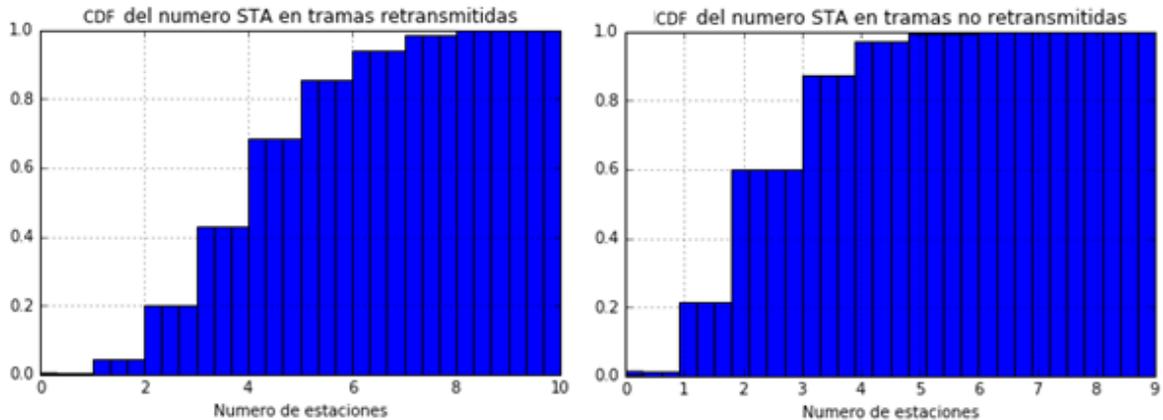
Gráfica 10 Histograma V.STA conectividad deficiente con respecto a tramas retransmitidas, no retransmitidas

En el histograma muestra que la mayor parte de las retransmisiones se dieron entre tres y cinco estaciones a comparación de las no retransmitidas en donde se mantienen 2 estaciones la mayor parte, aunque se ve una diferencia en el número máximo de estaciones, en las tramas retransmitidas se llegan hasta 10 estaciones y en las no retransmitidas solo se llega hasta 7, teniendo en cuenta que las tramas pertenecen a la misma red.



Gráfica 11 KDE V.sta conectividad eficiente y deficiente con respecto a tramas retransmitidas, no retransmitidas

Las gráficas anteriores muestran una curva similar a la distribución binomial, con la diferencia al número de estaciones donde se presenta la mayor densidad, para el caso de las tramas retransmitidas se encuentra entre 3 y 4 estaciones, mientras en las tramas no retransmitidas están entre 2 y 3 estaciones, con una densidad cerca de 1.



Gráfica 12 CDF V.STA conectividad eficiente y deficiente con respecto a tramas retransmitidas, no retransmitidas

En las tramas retransmitidas 8 estaciones es donde se tiene la probabilidad máxima, aunque encontrar menos de 5 estaciones es bastante alto teniendo una probabilidad por encima de 0.8 y en las no retransmitidas encontraremos máximo 5 estaciones. Algo a recalcar es que es más frecuente encontrar menos de tres estaciones de las tramas no retransmitidas a las retransmitidas con probabilidades de 0.6 a 0.4 respectivamente.

7.1.3 Caracterización de la variable retry rate

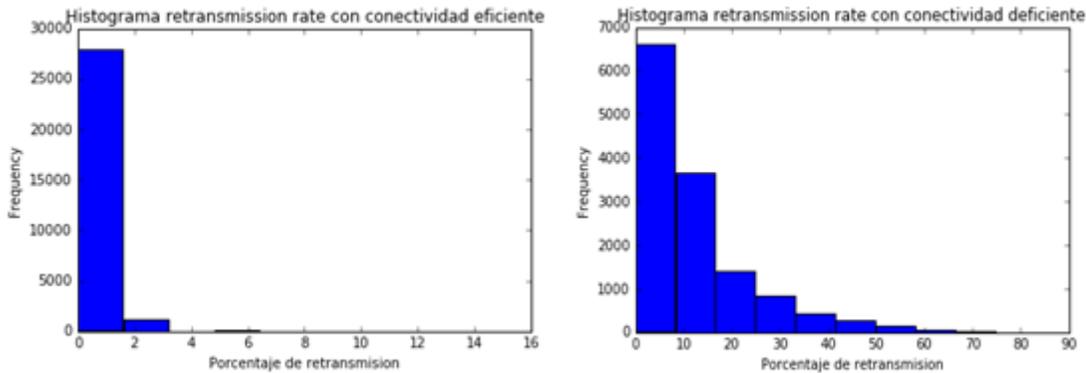
La siguiente tabla es una parte de la tabla de frecuencias de la variable retry rate

Tabla 3 frecuencias de la variable retry rate

	Retry	RetryA	frecuenciaR	frecuenciaA	PorcentajeR	PorcentajeA
Rangos						
(0, 100]	42.0	42.0	0.000932	0.000932	0.093236	0.093236
(100, 200]	10.0	52.0	0.000222	0.001154	0.022199	0.115435
(200, 300]	16.0	68.0	0.000355	0.001510	0.035518	0.150953
(300, 400]	22.0	90.0	0.000488	0.001998	0.048838	0.199791
(400, 500]	22.0	112.0	0.000488	0.002486	0.048838	0.248629

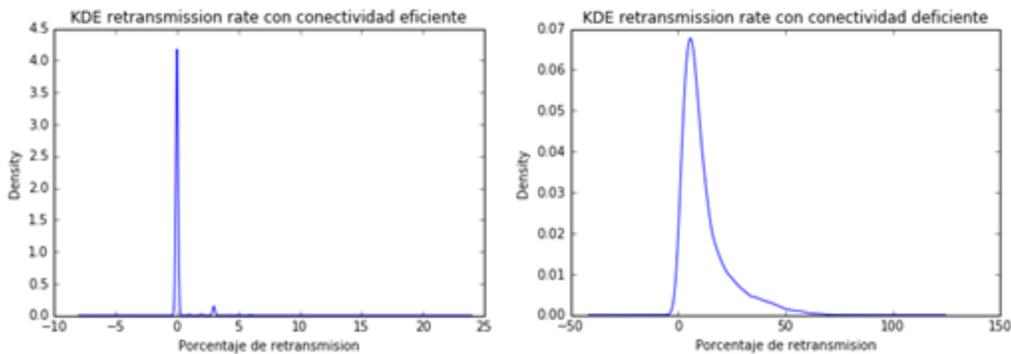
	Retry	RetryA	frecuenciaR	frecuenciaA	PorcentajeR	PorcentajeA
Rangos						
(500, 600]	24.0	136.0	0.000533	0.003019	0.053278	0.301907
(600, 700]	32.0	168.0	0.000710	0.003729	0.071037	0.372944
(700, 800]	31.0	199.0	0.000688	0.004418	0.068817	0.441761
(800, 900]	4.0	203.0	0.000089	0.004506	0.008880	0.450640
(900, 1000]	4.0	207.0	0.000089	0.004595	0.008880	0.459520
(1000, 1100]	1.0	208.0	0.000022	0.004617	0.002220	0.461740
(1100, 1200]	1.0	209.0	0.000022	0.004640	0.002220	0.463960
(1200, 1300]	5.0	214.0	0.000111	0.004751	0.011100	0.475059
(1300, 1400]	11.0	225.0	0.000244	0.004995	0.024419	0.499478
(1400, 1500]	3.0	228.0	0.000067	0.005061	0.006660	0.506138
(1500, 1600]	2.0	230.0	0.000044	0.005106	0.004440	0.510578
(1600, 1700]	3.0	233.0	0.000067	0.005172	0.006660	0.517238
(1700, 1800]	20.0	253.0	0.000444	0.005616	0.044398	0.561636
(1800, 1900]	11.0	264.0	0.000244	0.005861	0.024419	0.586055
(1900, 2000]	12.0	276.0	0.000266	0.006127	0.026639	0.612693
(2000, 2100]	19.0	295.0	0.000422	0.006549	0.042178	0.654872
(2100, 2200]	23.0	318.0	0.000511	0.007059	0.051058	0.705929
(2200, 2300]	11.0	329.0	0.000244	0.007303	0.024419	0.730348
(2300, 2400]	5.0	334.0	0.000111	0.007414	0.011100	0.741448
(2400, 2500]	20.0	354.0	0.000444	0.007858	0.044398	0.785846
(2500, 2600]	10.0	364.0	0.000222	0.008080	0.022199	0.808045
(2600, 2700]	27.0	391.0	0.000599	0.008680	0.059937	0.867982

En la tabla de frecuencias muestra el porcentaje de retransmisión por cada 100 tramas y en las que podemos observar que hay retransmisiones por encima del 50 por ciento.



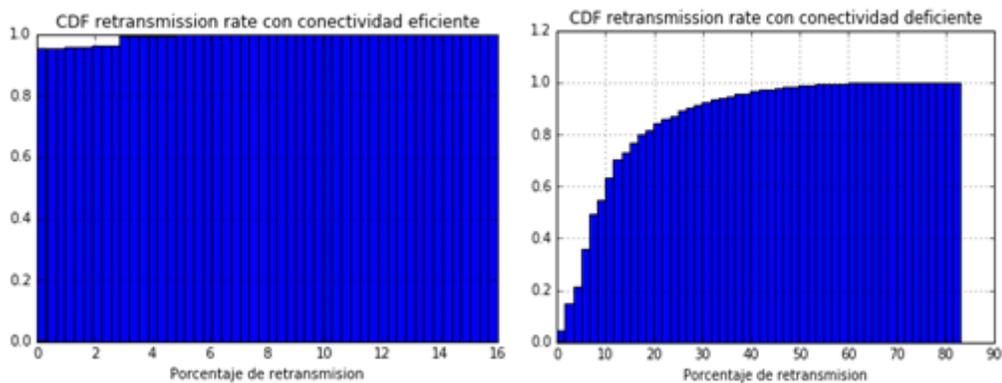
Gráfica 13 Histograma V.retry rate conectividad eficiente y conectividad deficiente

Los histogramas basados en las tablas de frecuencia muestra una gran diferencia entre las dos redes, donde la red con conectividad eficiente tiene un máximo retransmisión del 6 por ciento a cambio de la red deficiente que se tiene retransmisiones de hasta 80 por ciento, en la red con conectividad eficiente centrando su tasa de retransmisión por debajo del 2 por ciento, muestra que no existe retransmisión, a diferencia de la red con conectividad deficiente, que se tiene diferentes porcentaje, pero predomina las retransmisiones menores al 20 por ciento.



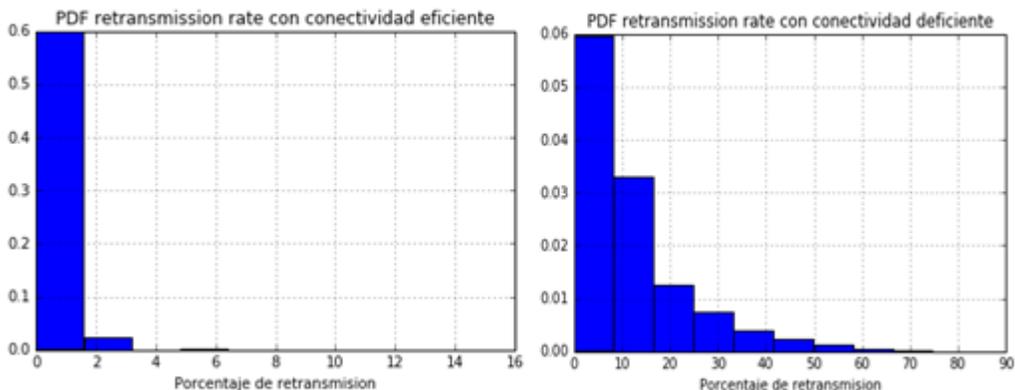
Gráfica 14 KDE V.retry rate conectividad eficiente y conectividad deficiente

En las estimaciones de densidad obtenida en las gráficas anteriores muestran una curva de densidad de distribución exponencial, donde el porcentaje de retransmisión en la red con conectividad eficiente forma una curva más suave a comparación de la red eficiente donde se encuentra casi centralizada en 0 con una densidad de 4



Gráfica 15 CDF V.retry rate conectividad eficiente y conectividad deficiente

El CDF de una red eficiente muestra que la probabilidad de que el porcentaje de retransmisión sea cero es mayor al 0.9 a diferencia de una conectividad deficiente donde la probabilidad de que el porcentaje de retransmisión sea mayor a cero es bastante alta.



Gráfica 16 PDF V.retry rate conectividad eficiente y conectividad deficiente

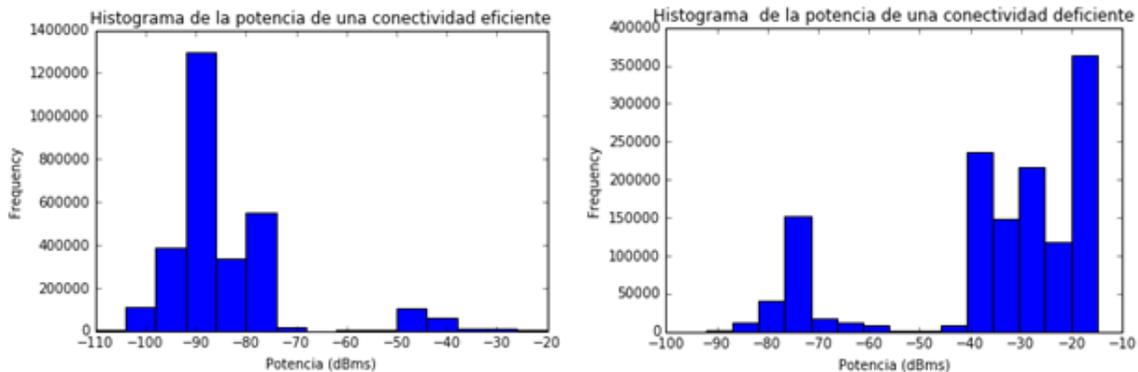
Las gráficas anteriores nos muestran una gran diferencia, dado de que en la red eficiente con una densidad de probabilidad de 0.6 los contiene el valor 0, dicho de otra forma existe una probabilidad de 0.4 de que el porcentaje de retransmisión sea diferente de 0 por ciento, y en una red deficiente la probabilidad máxima es de 0.06 en donde el porcentaje de retransmisión sea menos del 10 por ciento, pero en general las probabilidades son bajas para cada uno de los porcentajes de retransmisión.

7.1.4 Caracterización de la variable potencia

Tabla 4 tabla de frecuencia de la variable potencia

	SSI Signal	potenciaA	frecuenciaR	frecuenciaA	PorcentajeR	PorcentajeA
(-80, -70]	110806	110806	0.330281	0.330281	33.537128	33.537128
(-40, -30]	83177	193983	0.247927	0.578208	25.174789	58.711917
(-90, -80]	49015	242998	0.146100	0.724308	14.835138	73.547055
(-70, -60]	46331	289329	0.138099	0.862407	14.022785	87.569840
(-50, -40]	22157	311486	0.066044	0.928451	6.706154	94.275994
(-60, -50]	12394	323880	0.036943	0.965394	3.751233	98.027228
(-30, -20]	3735	327615	0.011133	0.976527	1.130455	99.157683
(-100, -90]	2656	330271	0.007917	0.984444	0.803879	99.961562
(-20, -10]	127	330398	0.000379	0.984822	0.038438	100.00000

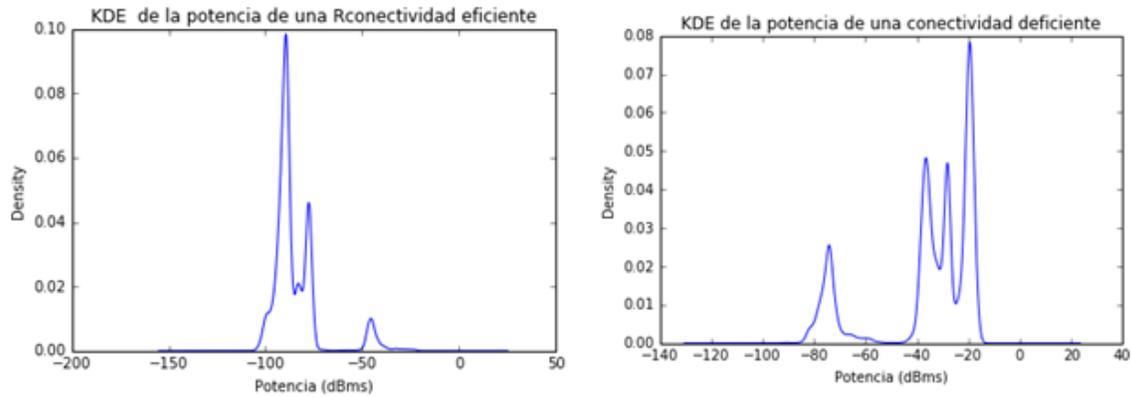
La tabla anterior contiene las frecuencias de la variable potencia donde nos muestra que el 33.5 por ciento de las tramas se recibieron con una potencia baja comprendida entre -70 y -80 dBm, y un 25.1 por ciento entre -40 y -30 dBm. También se puede observar que más del 80 por ciento de las tramas se retransmitieron en solo 4 intervalos de potencias, en las que se comprenden 3 intervalos de potencias medianamente bajas y una alta.



Gráfica 17 Histograma V. Potencia conectividad eficiente y conectividad deficiente

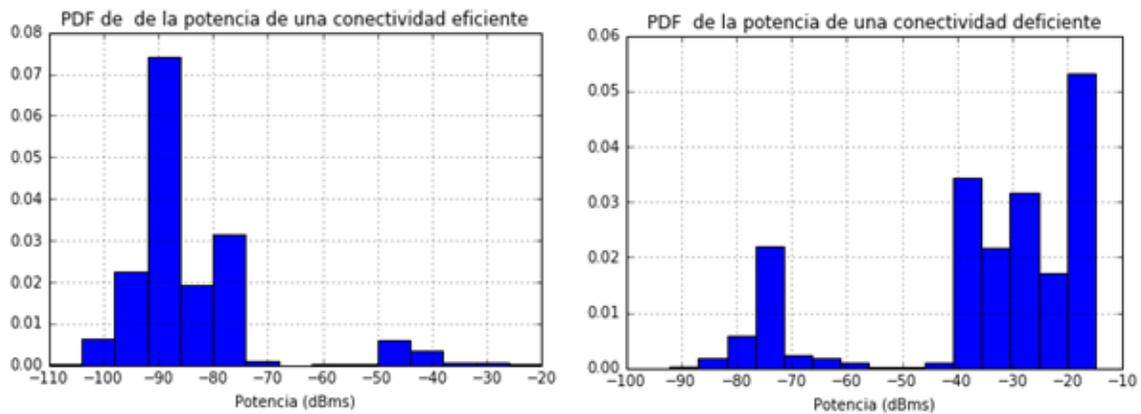
La gráfica nos muestra los datos, tal cual como se capturaron. Se presentan diferentes potencias, al ver la gráfica a simple vista podemos darnos cuenta los

rangos de potencias donde se transmitieron más tramas (-100 dBm, -70 dBm) en la red con conectividad eficiente y un rango de (-40 dBm, -20 dBm) para una red con conectividad eficiente, se transmitieron aproximadamente 350000 tramas, (-10 dBm, -20 dBm) en la red eficiente y 1200000 de tramas en -90 dBm.



Gráfica 18 KDE V. Potencia conectividad eficiente y conectividad deficiente

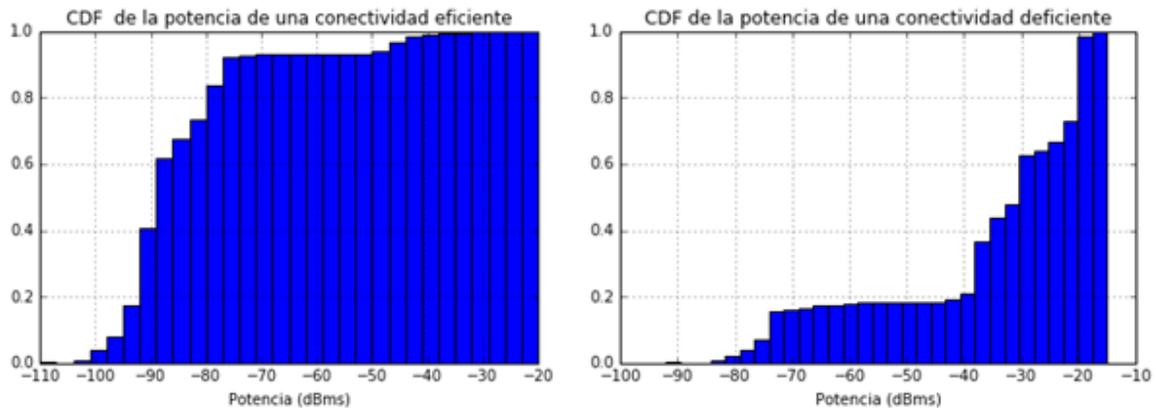
Las gráficas anteriores, la estimación de densidad no tienden a tener una curva de distribución de probabilidad semejante a las ya mencionadas anteriormente, aunque se destaca las densidades mayores que para ambos casos se acerca a 1, pero para el caso de la red con conectividad eficiente se tiene en una potencia aproximada de -90 dBm, mientras en la red con conectividad deficiente se encuentra con una potencia de -20 dBm.



Gráfica 19 PDF V. Potencia conectividad eficiente y conectividad deficiente

En los PDF anteriores de las dos redes ninguna potencia llega a tener una probabilidad mayor a 0.1, la potencia con probabilidad más alta en de -90 dBm en una red eficiente, seguido de -20 dBm de la red deficiente con una probabilidad de 0.05 en la red con conectividad eficiente se llega a tener potencias de hasta -110

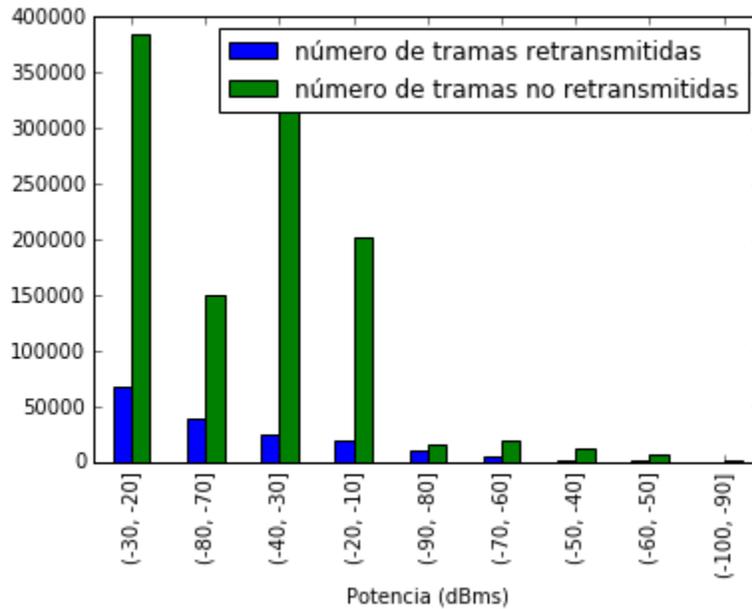
dBm con una probabilidad menor al 0.01 a comparación de la red con conectividad deficiente donde su potencia más baja es -90 dBm.



Gráfica 20 CDF V. Potencia conectividad eficiente y conectividad deficiente

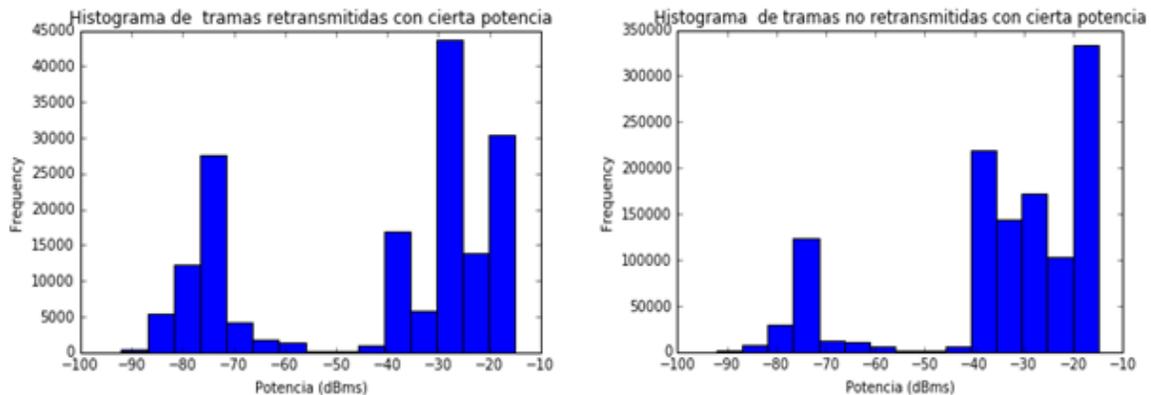
En las gráficas anteriores, existen una gran diferencia en la probabilidad en cada una de las potencias entre la red eficiente y la red con conectividad deficiente, una de las diferencias notables es que en la red con conectividad eficiente la probabilidad de encontrar potencias menores a -40 dBm es mayor a 0.8 al igual que -80 dBm, mientras que en la red con conectividad deficiente la probabilidad de encontrar potencias menores a -40 dBm es menor a 0.2, esto nos indica que en la red eficiente es más común encontrar potencias bajas y en la deficiente potencias altas.

Comportamiento de la variable estaciones con respecto a las tramas retransmitidas y no retransmitidas



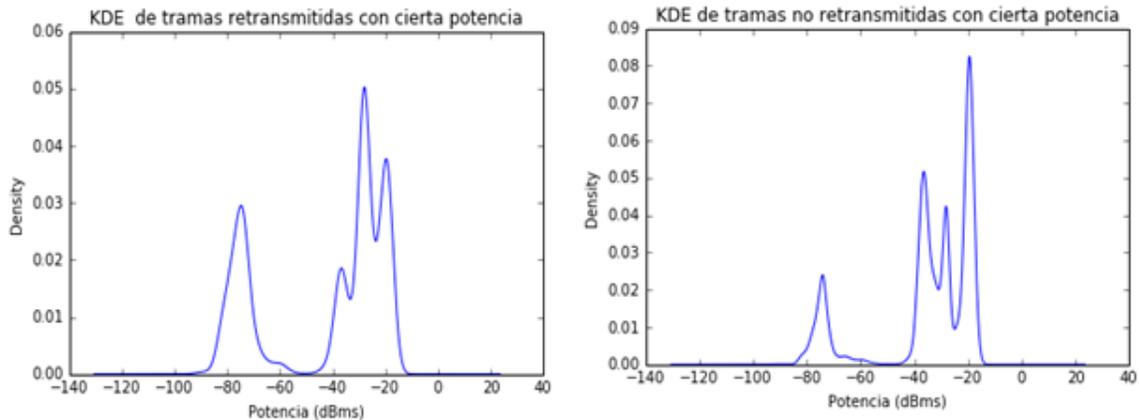
Gráfica 21 Potencia de número de tramas retransmitidas y no retransmitidas

La gráfica anterior nos muestra una clara diferencia del número de tramas que fueron retransmitidas y no retransmitidas con cierta potencia en este caso el número de tramas no retransmitidas es mayor en cada uno de los intervalos de la potencia, siendo la potencia de -30 a -20 dBm la más frecuente para ambos casos, para las tramas no retransmitidas tiende a tener una potencia recibida media alta con -30 y -40 dBm y las tramas no retransmitidas una potencia alta y baja comprendiendo la potencia baja entre los -70 y -80 dBm.



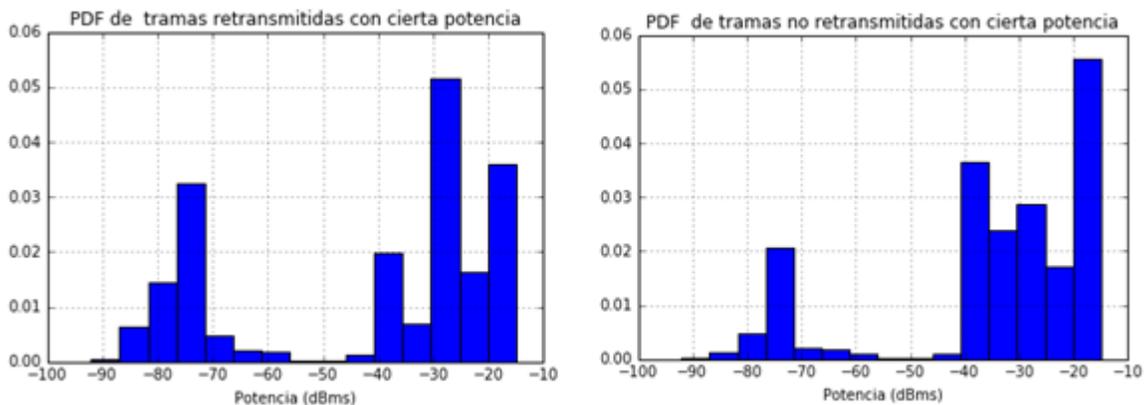
Gráfica 22 Histograma de tramas retransmitidas y no retransmitidas con cierta potencia

El histograma de las tramas retransmitidas y no retransmitidas da visión más clara de las potencias en que se presentó las transmisiones de las tramas, como resultado se tiene una simetría en las potencias recibidas con la diferencia que en las tramas retransmitidas son más frecuentes las potencias bajas y en las no retransmitidas potencias más altas, con respecto en las potencias altas las tramas retransmitidas se tiene una mayor frecuencia en -30 dBm y en las no retransmitidas mayor a -20 dBm, en la potencia altas si se mantienen la misma potencia comprendida entre los -70 dBm y -80 dBm.



Gráfica 23 KDE de tramas retransmitidas y no retransmitidas con cierta potencia

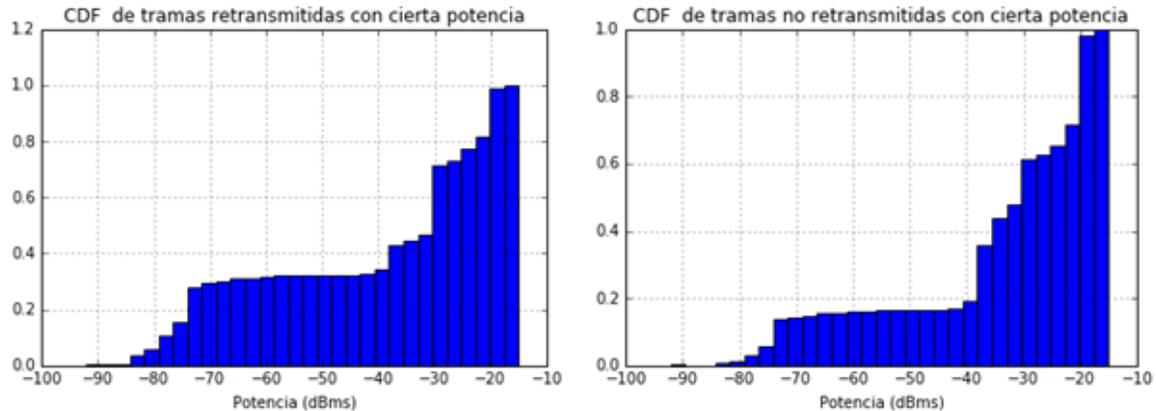
Al igual que en la estimación de densidad para la potencia en una red deficiente y eficiente, se tiene una similitud en cuanto la curva de distribución de probabilidad, pero se muestra en este caso una simetría en las densidades tanto para las potencias de las tramas retransmitidas como las no retransmitidas, esta potencia se comprende entre los -20 dBm y -40 dBm.



Gráfica 24 PDF de tramas retransmitidas y no retransmitidas con cierta potencia

En el PDF de las tramas retransmitidas se tienen tres casos de probabilidades altas a comparación de las demás siendo la probabilidad 0.05 la más alta para la potencia de -30 dBm. En las probabilidades de las tramas no retransmitidas, la

probabilidad es un poco más de 0.05 para las potencias mayores a -20 dBm, a diferencia de las tramas retransmitidas. Varias potencias en ambos casos manejan probabilidades cercanas a 0.02



Gráfica 25 CDF tramas retransmitidas y no retransmitidas con cierta potencia

La distribución acumulada para las tramas retransmitidas nos muestra que la probabilidad de que la potencia sea menor a -15 dBm es de 1, caso que nos afirma que no existen potencias mayores, al igual que en las tramas no retransmitidas. La diferencia de las tramas retransmitidas con respecto a las no retransmitidas es que se maneja una probabilidad mayor en cada una de las potencias de las tramas no retransmitidas.

7.1.5 Caracterización de la variables longitud

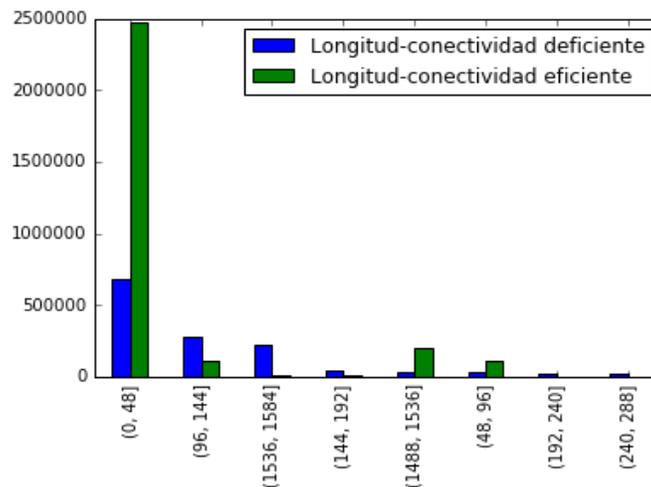
Tabla 5 tabla de frecuencia de la variable longitud

	Length	LengthA	frecuenciaR	frecuencia	PorcentajeR	PorcentajeA
(0, 48]	179756	179756	0.375010	0.375010	37.500965	37.500965
(240, 288]	90287	270043	0.188358	0.563368	18.835809	56.336774
(192, 240]	76102	346145	0.158765	0.722133	15.876513	72.213286
(96, 144]	50349	396494	0.105039	0.827172	10.503883	82.717170
(48, 96]	37248	433742	0.077707	0.904879	7.770733	90.487903
(144, 192]	14771	448513	0.030815	0.935695	3.081548	93.569451
(384, 432]	13593	462106	0.028358	0.964052	2.835792	96.405243

	Length	LengthA	frecuenciaR	frecuencia	PorcentajeR	PorcentajeA
(1536, 1584]	4987	467093	0.010404	0.974456	1.040395	97.445638
(1488, 1536]	4456	471549	0.009296	0.983753	0.929617	98.375256
(1440, 1488]	1721	473270	0.003590	0.987343	0.359038	98.734293
(336, 384]	1420	474690	0.002962	0.990305	0.296243	99.030536
(624, 672]	1083	475773	0.002259	0.992565	0.225937	99.256473
(288, 336]	841	476614	0.001755	0.994319	0.175451	99.431924
(480, 528]	396	477010	0.000826	0.995145	0.082614	99.514538
(720, 768]	373	477383	0.000778	0.995924	0.077816	99.592354
(576, 624]	256	477639	0.000534	0.996458	0.053407	99.645761
(528, 576]	237	477876	0.000494	0.996952	0.049443	99.695204
(768, 816]	187	478063	0.000390	0.997342	0.039012	99.734216
(432, 480]	183	478246	0.000382	0.997724	0.038178	99.772394
(1200, 1248]	121	478367	0.000252	0.997976	0.025243	99.797637
(1296, 1344]	120	478487	0.000250	0.998227	0.025035	99.822672
(1152, 1200]	111	478598	0.000232	0.998458	0.023157	99.845829
(1104, 1152]	99	478697	0.000207	0.998665	0.020654	99.866482
(672, 720]	93	478790	0.000194	0.998859	0.019402	99.885884
(1056, 1104]	83	478873	0.000173	0.999032	0.017316	99.903200
(1008, 1056]	67	478940	0.000140	0.999172	0.013978	99.917177
(1392, 1440]	64	479004	0.000134	0.999305	0.013352	99.930529
(960, 1008]	61	479065	0.000127	0.999433	0.012726	99.943255
(864, 912]	58	479123	0.000121	0.999554	0.012100	99.955355
(1248, 1296]	47	479170	0.000098	0.999652	0.009805	99.965160
(1344, 1392]	46	479216	0.000096	0.999748	0.009597	99.974757
(816, 864]	45	479261	0.000094	0.999841	0.009388	99.984145
(1680, 1728]	44	479305	0.000092	0.999933	0.009179	99.993324
(912, 960]	32	479337	0.000067	1.000000	0.006676	100.000000
(1920, 1968]	0	479337	0.000000	1.000000	0.000000	100.000000
(2160, 2208]	0	479337	0.000000	1.000000	0.000000	100.000000
(2112, 2160]	0	479337	0.000000	1.000000	0.000000	100.000000

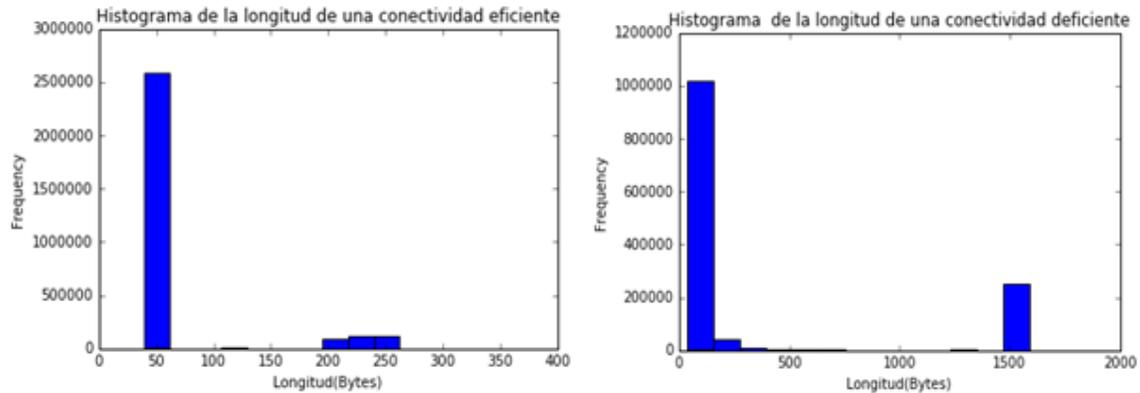
	Length	LengthA	frecuenciaR	frecuencia	PorcentajeR	PorcentajeA
(2064, 2112]	0	479337	0.000000	1.000000	0.000000	100.000000
(2016, 2064]	0	479337	0.000000	1.000000	0.000000	100.000000
(1968, 2016]	0	479337	0.000000	1.000000	0.000000	100.000000
(1632, 1680]	0	479337	0.000000	1.000000	0.000000	100.000000
(1872, 1920]	0	479337	0.000000	1.000000	0.000000	100.000000
(1824, 1872]	0	479337	0.000000	1.000000	0.000000	100.000000
(1776, 1824]	0	479337	0.000000	1.000000	0.000000	100.000000
(1728, 1776]	0	479337	0.000000	1.000000	0.000000	100.000000
(2208, 2256]	0	479337	0.000000	1.000000	0.000000	100.000000
(1584, 1632]	0	479337	0.000000	1.000000	0.000000	100.000000
(2256, 2304]	0	479337	0.000000	1.000000	0.000000	100.000000

En la tabla de frecuencias de la variable longitud nos da a conocer que el 90.3 por ciento de las tramas fueron transmitidas en cinco intervalos y estos siendo longitudes pequeñas, el rango en el que se transmitieron la mayoría de las tramas fue de 0 a 48 bytes con el 37.5 por ciento del total de las tramas, entre 1536 a 1584 bytes es el rango de longitud más grande con mayor frecuencia que en este caso es de 1 por ciento



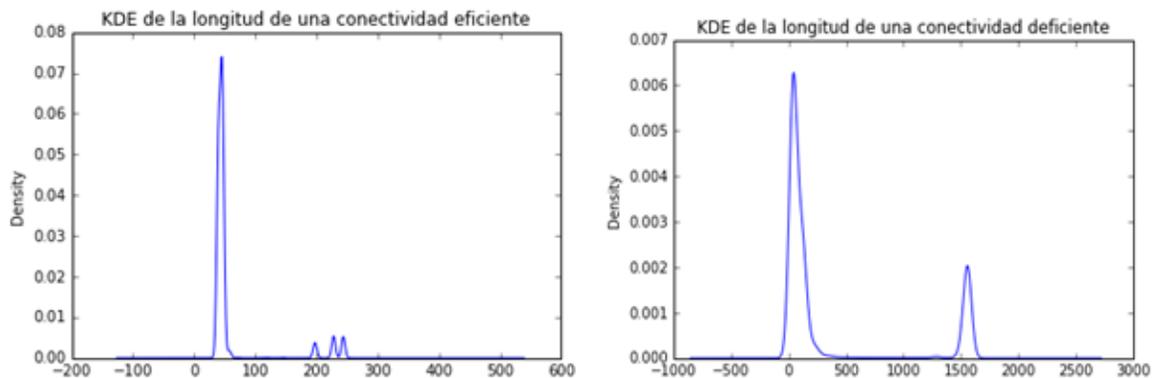
Gráfica 26 Longitud conectividad eficiente y deficiente

En la gráfica de barras tanto para la red con conectividad eficiente como deficiente la longitud con que más se transmite se encuentran entre 0 y 48 bytes, siendo los picos más altos.



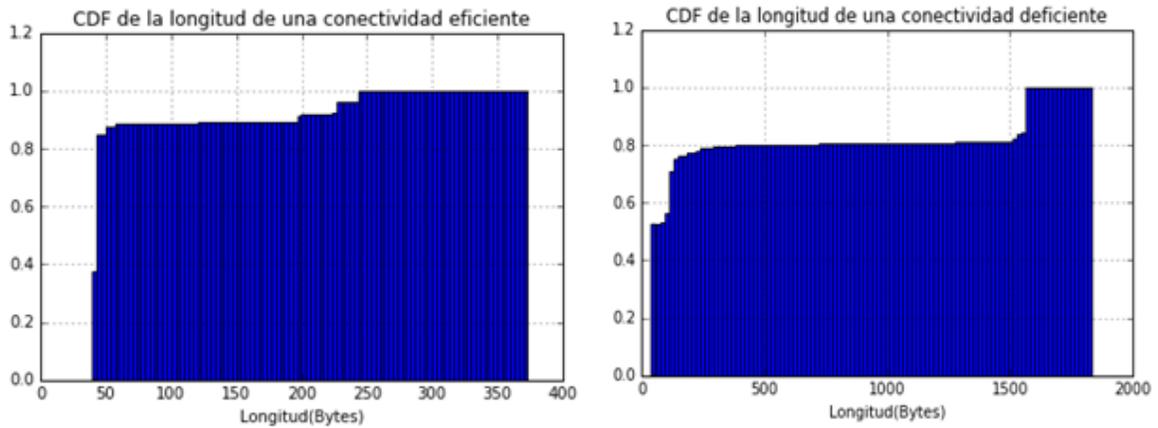
Gráfica 27 Histograma V. Longitud con conectividad eficiente y conectividad deficiente

Comparando los histogramas con las gráficas de barras se observa se tiene transmisiones más frecuentes con longitudes pequeñas, pero en el histograma nos muestra una diferencia entre la red eficiente con la deficiente y es que para el caso de una red con conectividad eficiente se transmitieron con longitudes menores a los 300 bytes mientras que en la red con conectividad deficiente más de 200000 tramas se transmitieron con una longitud un poco mayor a los 1500 bytes.



Gráfica 28 KDE V. longitud con conectividad eficiente y conectividad deficiente

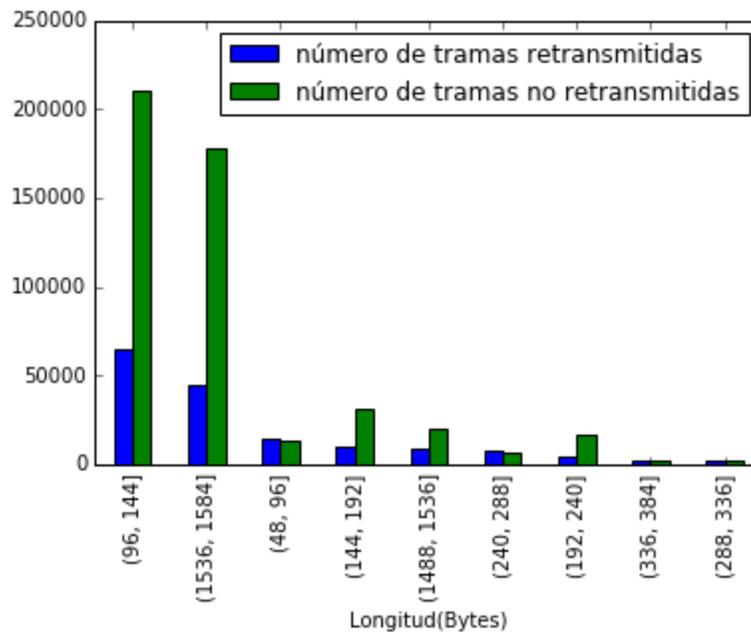
Para las estimaciones de densidad anteriores se maneja una similitud donde se concentra la densidad en dos puntos, para el caso de la red con conectividad eficiente se concentra en una longitud aproximada de 50 bytes con una densidad cercana a 0.08, y en la red con conectividad deficiente la densidad se centra en las longitudes aproximadas de 50 y 1500 bytes



Gráfica 29 CDF longitud conectividad eficiente y conectividad deficiente

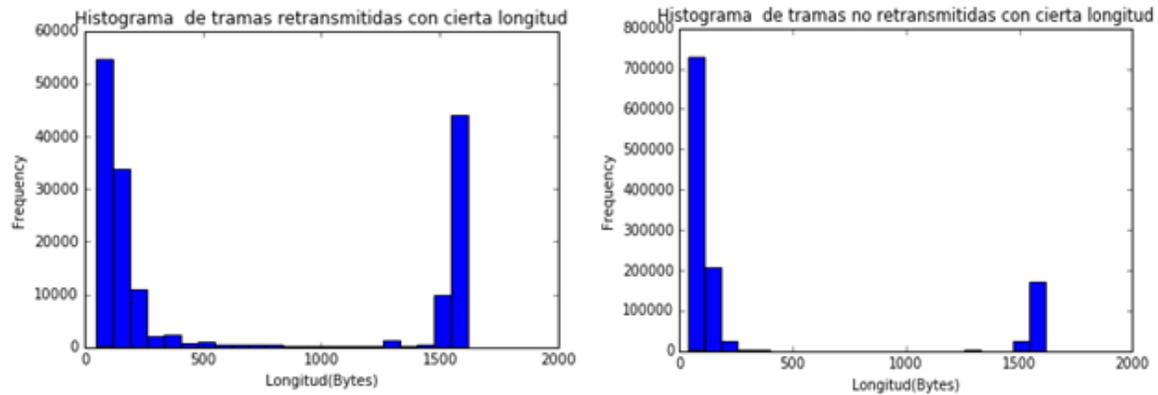
Al hacer la comparación de las distribuciones acumuladas para la red con conectividad eficiente y la red con conectividad deficiente se muestra una gran simetría, pero en realidad la red eficiente maneja una probabilidad 1 de que las longitudes sean menores a 250 mientras la red deficiente se tiene la probabilidad de 1 en la una longitud mayor a 1500 bytes.

Comportamiento de la variable longitud con respecto a las tramas retransmitidas y no retransmitidas



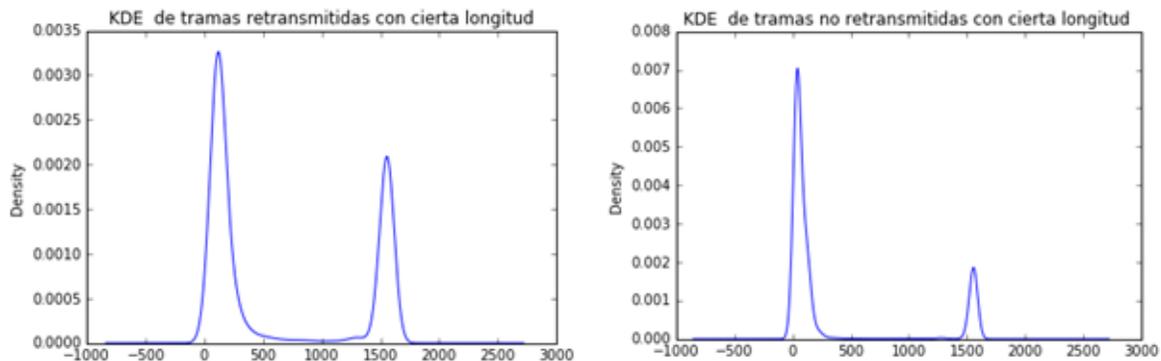
Gráfica 30 Longitud tramas retransmitidas y no retransmitidas

En la gráfica anterior se compara las longitudes de las tramas retransmitidas y no retransmitidas, en la cual se observa que las longitudes de 96 a 144 bytes y 1536 a 1584 bytes son las más frecuentes para ambos casos, aunque para las demás longitudes si se manejan cantidades diferentes de las tramas transmitida con cierta longitud.



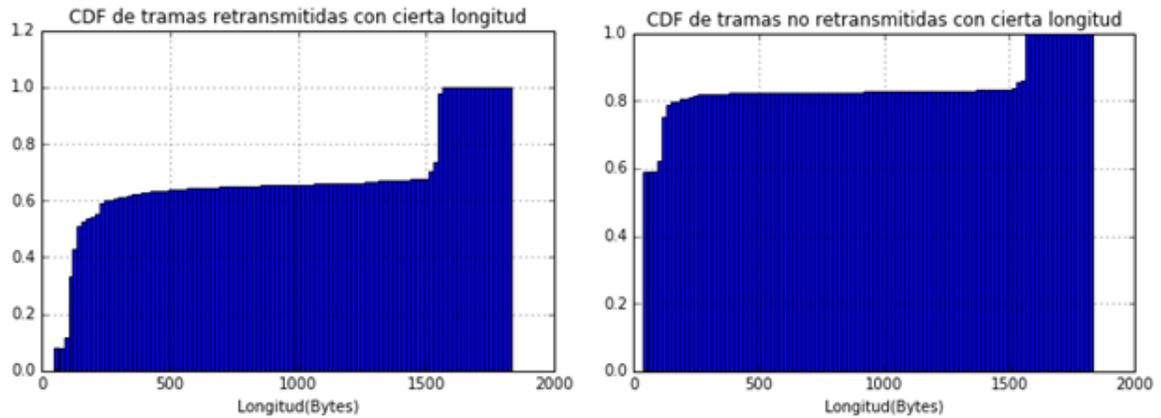
Gráfica 31 Histograma de tramas retransmitidas y no retransmitidas con cierta longitud

El histograma de las tramas retransmitidas muestra una gran similitud con el de las tramas no retransmitidas con la diferencia que existe más variedad de longitudes pequeñas con la que se retransmitieron las tramas.



Gráfica 32 KDE de tramas retransmitidas y no retransmitidas con cierta longitud

Las gráficas anteriores no se tiene una distribución de probabilidad asociada, pero se maneja una similitud en ambos casos, donde se concentra la densidad en dos puntos, para el caso de las tramas retransmitidas y en las tramas no retransmitidas se concentra en una longitud aproximada de 50 bytes y 1500 bytes, en las dos situaciones las densidades son muy bajas.



Gráfica 33 CDF de tramas retransmitidas y no retransmitidas con cierta longitud

En los CDF obtenidos tiende a ser igual con la probabilidad de 1 de que las longitudes sean menores a 2000 bytes con la diferencia en la probabilidad para la longitud 1500 bytes donde las tramas retransmitidas manejan una probabilidad aproximada de 0.6 de que las demás longitudes sean menores, en cambio las tramas no retransmitidas se encuentra la probabilidad aproximada de 0.8.

7.2 Correlación

7.2.1 Correlación de la variable longitud

Tabla 6 Correlación de la variable longitud con conectividad eficiente

	Retry	max-length	min-length	promedio-length	moda-length
rangos					
(0, 100]	0	244	40	162.35	244
(100, 200]	0	244	40	170.44	244
(200, 300]	0	244	40	145.73	200
(300, 400]	0	244	40	119.47	40
(400, 500]	0	244	40	161.34	200
(500, 600]	0	244	40	172.43	200
(600, 700]	0	244	40	176.68	200
(700, 800]	0	281	40	152.74	40
(800, 900]	0	244	40	188.96	200
(900, 1000]	0	244	40	159.87	200
(1000, 1100]	0	244	40	180.35	200
(1100, 1200]	0	273	40	177.65	200
(1200, 1300]	0	244	40	169.90	200
(1300, 1400]	0	244	40	185.17	200
(1400, 1500]	0	281	40	161.59	200
(1500, 1600]	0	244	40	175.22	200
(1600, 1700]	0	244	40	187.11	200
(1700, 1800]	0	273	40	176.53	200

Tabla anterior con los rangos de rata de retransmisión mínima, máxima, promedio y moda de la variable longitud con conectividad eficiente, todo esto se puede observar en la tabla con rangos de tramas de (0,100] así hasta llegar al rango en este caso hasta (1700,1800] este rango fue el escogido para mostrar en la tabla teniendo en cuenta que los rangos son muchos más.

Se observa que la longitud máxima tiende a ser 281 bytes, para la longitud mínima es de 40 bytes, para la longitud promedio máxima es de 187.11 bytes, con un promedio mínimo de 119.47 bytes y la longitud frecuente es de 200 bytes.

Tabla 7 Correlación de la variable longitud con conectividad deficiente

	Retry	max-length	min-length	promedio-length	moda-length
rangos					
(0, 100]	14	1564	40	263.97	40
(100, 200]	14	1564	40	260.83	40
(200, 300]	5	444	40	69.73	40
(300, 400]	10	243	40	73.02	40
(400, 500]	19	1564	40	162.98	40
(500, 600]	21	1564	40	191.91	40
(600, 700]	32	1564	40	225.37	40
(700, 800]	47	1564	40	532.07	40
(800, 900]	16	1564	40	332.11	40
(900, 1000]	21	1564	40	243.26	40
(1000, 1100]	27	1564	40	107.31	40
(1100, 1200]	43	414	40	129.92	40
(1200, 1300]	31	647	40	124.56	40
(1300, 1400]	65	324	40	125.12	127
(1400, 1500]	36	1562	40	173.61	40
(1500, 1600]	52	260	40	114.82	40
(1600, 1700]	54	315	40	105.97	40
(1700, 1800]	33	1512	40	190.59	40

La tabla anterior con los rangos (0,100] hasta un rango escogidos para ser mostrados (1700,1800] teniendo en cuenta que son muchos más los intervalos que nos dan los datos.

El porcentaje mínimos de retransmisión es de 5% y se presenta en el rango de (200,300] y el porcentaje máxima es de 54% el cual se presenta en el rango de (1600,1700], con una longitud máxima de 1562 bytes, longitud mínima de 40

bytes, una longitud promedio máximo de 532.07 bytes, promedio mínimo de 69.73 bytes y una longitud frecuente de 40 bytes.

Tabla 8 Correlación de la variable longitud entre retry rate con conectividad eficiente

	RetryNormal	RetrySpearman
max-length	0.022733	0.023412
min-length	0.077373	0.074445
promedio-length	0.000755	0.054349
moda-length	-0.001918	0.104746

Teniendo en cuenta la tabla de anterior a la cual se realizó los diferentes métodos como son el método de Spearman y el método normal para una conectividad eficiente. Con el método Spearman podemos decir que, en una longitud máxima de 0.023, longitud mínima 0.074, longitud promedio 0.054, longitud moda 0.104, tienden a tener una asociación positiva esto nos indica que hay una probabilidad baja de tener correlación.

Con el método normal si puede decir que, en longitud máxima con valor de 0.022, longitud mínima con valor de 0.077 y longitud promedio con un valor de 0.0007 ,no hay correlación, pero tienden a ser una asociación positiva teniendo en cuenta que los resultados tienden a 0, en longitud moda con un valor de -0.001 tiende a una correlación negativa pero la probabilidad es muy mínima.

Tabla 9 Correlación de la variable longitud entre retry rate con conectividad deficiente

	RetryNormal	RetrySpearman
max-length	-0.203407	-0.128871
min-length	-0.000464	0.003315
promedio-length	-0.124618	-0.243051
moda-length	0.418465	0.343878

Teniendo en cuenta la tabla anterior se le realizan los diferentes métodos como son: Método normal, Método Spearman para una conectividad deficiente.

De acuerdo con todos los métodos de correlación que se realizaron, existe una relación es directamente proporcional dado que algunos valores son positivos entre potencia y la tasa de retransmisión, el método Spearman es quien nos arroja

una mayor correlación entre la tasa de retransmisión respecto a la longitud moda, con valores 0.418465, 0.418465 y 0.343878 los demás valores se acercan a 0 indicando una correlación baja, esto quiere decir que se acerca a 0. Si fuera una correlación alta nos indica que se acerca a 1.

Tabla 10 Correlación Spearman con p-value de la variable longitud entre retry rate con conectividad eficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-moda_length	(0.10474614608, 7.96852825768e-72)
Retry-max_length	(0.0234122875629, 6.46093950348e-05)
Retry--min_length	(0.054348876557, 1.69108292881e-20)
Retry--promedio_length	(0.054348876557, 1.69108292881e-20)

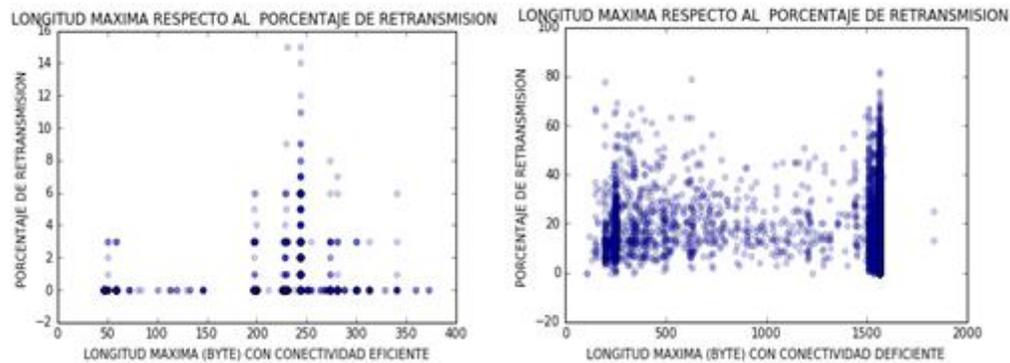
En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de los diferentes rangos de longitud con retry las cuales son moda, longitud máxima, mínima y promedio con una conectividad eficiente. Lo cual podemos decir que el coeficiente de correlación de retry con respecto a la moda longitud de (0.104, 7.968e-72), retry con respecto a longitud máxima de (0.023, 6.460e-05), retry con respecto a longitud mínima y retry con respecto a longitud promedio con un valor de (0.054, 1.691e-20), es positiva teniendo en cuenta que si retry aumenta y longitud aumenta podemos dar este resultado.

Tabla 11 Correlación Spearman con p-value de la variable longitud entre retry rate con conectividad deficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-moda_length	(0.343877592803, 0.0)
Retry-max_length	(-0.128871386864, 1.32848107772e-50)
Retry--min_length	(-0.243051061533, 5.46659860468e-179)
Retry--promedio_length	(-0.243051061533, 5.46659860468e-179)

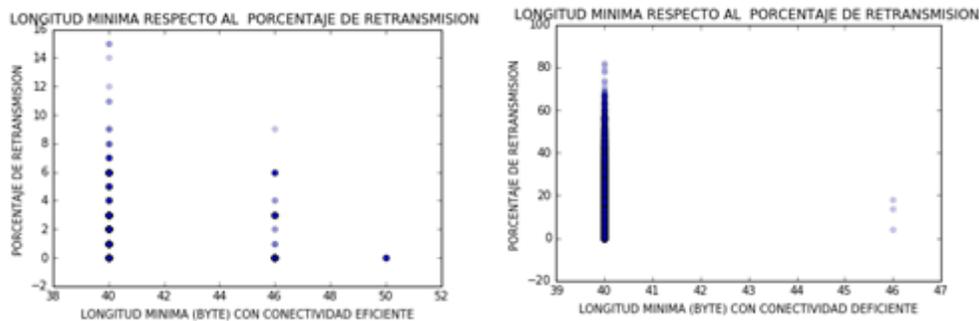
En esta tabla se observa los valores de coeficiente de correlación(co-corr), el valor p(p-value) de los diferentes rangos de longitud, las cuales son retry con respecto a moda longitud (0.343, 0.0) , retry con respecto a longitud máxima (-0.12, 1.32e-50), retry con respecto a longitud mínima y longitud promedio con un valor de (-0.243, 5.466e-179) que es positiva, teniendo en cuenta que si retry aumenta y

length aumenta podemos dar este resultado, en comparación con retry con respecto a longitud máxima, mínima y promedio con el coeficiente de correlación es negativo se puede decir que entre la variable longitud y retry existe una correlación inversa, partiendo de lo anterior si la retry aumenta la length tiende a disminuir. Al observar los datos de la tabla podemos decir que según el p-value de longitud máxima, mínima y promedio aceptamos la hipótesis, pero con moda longitud rechazamos la hipótesis.



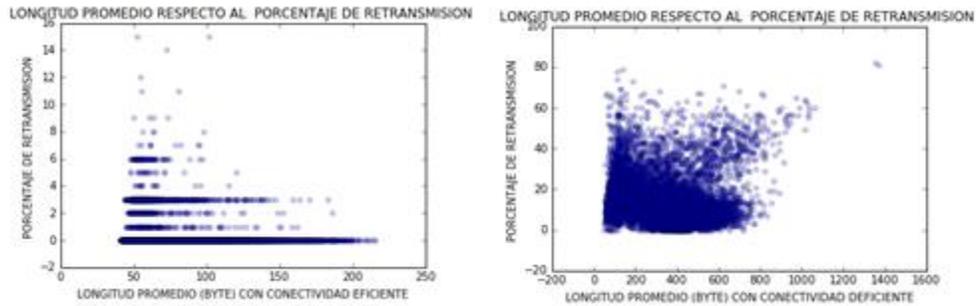
Gráfica 34 Dispersión de la variable longitud máxima respecto al porcentaje de retransmisión

La gráfica de dispersión del porcentaje de retransmisión con respecto a la longitud máxima muestra que en la red eficiente la mayor parte de las retransmisiones se realizaron con una longitud de 250 bytes, aunque se resalta más la no presencia de retransmisión en diferentes longitudes. A diferencia de la red con conectividad deficiente donde se ve que la mayor parte de las retransmisiones se realizaron con una longitud de 1500 bytes.



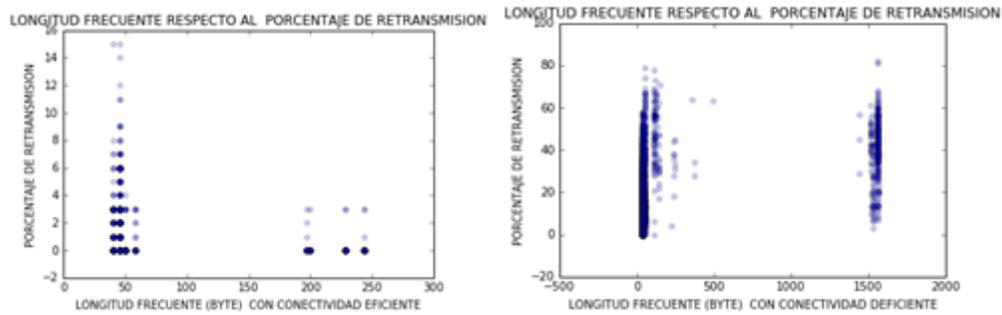
Gráfica 35 Dispersión de la variable longitud mínima respecto al porcentaje de retransmisión

En la red con conectividad eficiente como deficiente mantuvieron la misma longitud mínima con diferencia que el porcentaje de retransmisión en la red deficiente llega hasta el 80 por ciento y en la red eficiente tan solo al 16 por ciento.



Gráfica 36 Dispersión de la variable longitud promedio respecto al porcentaje de retransmisión

En la comparación de las longitudes promedio en ambas redes tienden a tener una mayor concentración en porcentajes de retransmisiones bajas al igual la longitud, para el caso de la red eficiente la longitud no supera los 250 y en la red deficiente la concentración se da mayormente en las longitudes menores a 800 bytes



Gráfica 37 Dispersión de la variable longitud frecuente respecto al porcentaje de retransmisión

En el grafico anterior se observa que La longitud más frecuente en las que se retransmitió fue 50 bytes para ambas redes y a diferencia de la red eficiente la red con conectividad deficiente tuvo una frecuente retransmisión las longitudes de 1500 bytes

7.2.2 Correlación de la variable potencia

Tabla 12 Correlación de la variable potencia con respecto a retry rate con conectividad eficiente

	Retry	max-potencia	min-potencia	promedio-potencia	moda-potencia
rangos					
(0, 100]	0	-91	-106	-96.44	-95
(100, 200]	0	-91	-104	-97.58	-95
(200, 300]	0	-48	-107	-95.42	-95
(300, 400]	0	-88	-102	-95.78	-98
(400, 500]	0	-90	-103	-95.50	-96
(500, 600]	0	-84	-103	-94.89	-91
(600, 700]	0	-82	-102	-95.18	-91
(700, 800]	0	-77	-102	-94.02	-92
(800, 900]	0	-90	-103	-94.41	-92
(900, 1000]	0	-83	-103	-95.14	-94
(1000, 1100]	0	-88	-101	-93.76	-92
(1100, 1200]	0	-84	-103	-94.26	-94
(1200, 1300]	0	-83	-101	-94.55	-93
(1300, 1400]	0	-89	-104	-94.86	-93
(1400, 1500]	0	-83	-103	-94.77	-95
(1500, 1600]	0	-51	-103	-93.13	-93
(1600, 1700]	0	-90	-104	-94.73	-94
(1700, 1800]	0	-46	-104	-92.18	-95

La tabla anterior con los rangos de rata de retransmisión por cada (0,100] tramas, se observa que el porcentaje de retransmisión es 0, con una potencia máxima de -91 dBm en el intervalo de (0,100], mínima potencia es de -101 dBm en el intervalo de (1200,1300], el promedio máximo es de -97.58 dBm en el intervalo de (100,200] y un promedio mínimo de -92.18 dBm en el intervalo de (1700,1800], con una potencia frecuente de -95 dBm. Los intervalos que se muestran en la tabla son una pequeña porción de los datos totales que se obtuvieron.

Tabla 13 Correlación de la variable potencia con respecto a retry rate con conectividad deficiente

	Retry	max-potencia	min-potencia	promedio-potencia	moda-potencia
rangos					
(0, 100]	14	-20	-84	-61.15	-72
(100, 200]	14	-20	-85	-60.21	-72
(200, 300]	5	-19	-78	-55.52	-72
(300, 400]	10	-19	-92	-57.48	-72
(400, 500]	20	-19	-80	-55.66	-28
(500, 600]	21	-21	-82	-55.28	-28
(600, 700]	31	-20	-92	-56.36	-28
(700, 800]	47	-19	-82	-40.09	-28
(800, 900]	16	-28	-82	-59.12	-72
(900, 1000]	21	-28	-92	-62.30	-74
(1000, 1100]	27	-28	-82	-48.60	-28
(1100, 1200]	43	-19	-84	-43.01	-28
(1200, 1300]	31	-19	-82	-41.99	-28
(1300, 1400]	66	-28	-84	-61.37	-74
(1400, 1500]	35	-20	-80	-41.87	-28
(1500, 1600]	52	-20	-92	-39.81	-30
(1600, 1700]	54	-19	-92	-44.95	-28
(1700, 1800]	34	-19	-83	-42.90	-28

La tabla anterior con los rangos de rata de retransmisión mínima -92 dBm, máxima -19 dBm, promedio máximo -62.30 dBm en el rango de (900,1000], promedio mínimo -39.82 dBm en el rango de (1500,1600] y moda de la variable potencia -28 dBm, de cada intervalo con estos datos que se obtiene se puede observar la relación que puede haber entre estas dos variables, por cada 100 tramas, así hasta llegar al rango en este caso hasta (1700,1800] el cual fue escogido para la muestra, teniendo en cuenta que los rangos superan este valor.

Tabla 14 Correlación Spearman con p-value de la variable potencia entre retry rate con conectividad eficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-moda_potencia	(0.159560174755, 2.67147967696e-165)
Retry-max_potencia	(-0.0585984978392, 1.41889093488e-23)
Retry--min_potencia	(0.159019503484, 3.51172617644e-164)
Retry--promedio_potencia	(0.159019503484, 3.51172617644e-164)

En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de los diferentes rangos de potencia con retry las cuales son retry-moda potencia de (0.159, 2.671e-165], retry con respecto a potencia máxima de (-0.058, 1.418e-23] el coeficiente de correlación es negativo se puede decir que entre la variable potencia y retry existe una correlación inversa. Lo cual podemos decir que el coeficiente de correlación de retry con respecto a la potencia mínima y promedio (0.159, 3.511e-164] el coeficiente de correlación es positiva teniendo en cuenta que si retry aumenta y potencia aumenta podemos dar este resultado

Tabla 15 Correlación Spearman con p-value de la variable potencia entre retry rate con conectividad deficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-moda_potencia	(-0.306460761112, 1.32881136816e-288)
Retry-max_potencia	(-0.242685504572, 1.93193892178e-178)
Retry--min_potencia	(-0.489064913928, 0.0)
Retry--promedio_potencia	(-0.489064913928, 0.0)

En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de los diferentes rangos de potencia con retry las cuales son potencia frecuente (-0.306, 1.328e-288), potencia máxima (-0,242, 1.931e-178), potencia mínima y promedio con valor de (-0.489, 0.0) para una red con conectividad deficiente. Dado que el coeficiente de correlación es negativo, se puede decir que entre la variable potencia y la tasa de retransmisión existe una correlación inversa, partiendo de lo anterior si la retry aumenta la potencia tiende a disminuir. Según el valor que nos muestra la tabla de p-value con potencia

promedio y potencia mínima se rechaza la hipótesis, pero con potencia máxima y moda se acepta la hipótesis.

Tabla 16 Correlación Spearman de la variable potencia entre retry rate con conectividad eficiente

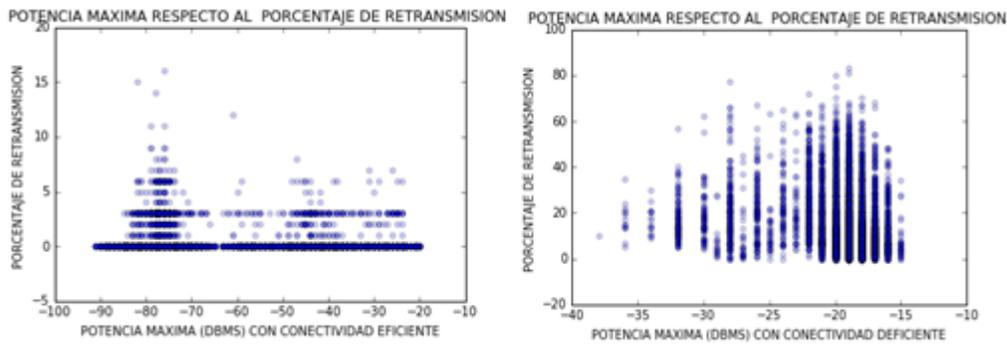
	Retry- metodo Normal	Retry - metodo Spearman
max-potencia	-0.018585	-0.058598
min-potencia	0.036935	0.027970
promedio-potencia	0.040367	0.159020
moda-potencia	0.041441	0.159560

Con el método Spearman y el método normal se puede decir la probabilidad de correlación es mínima, pero tienden a ser una asociación positiva teniendo en cuenta que los resultados tienden a 0 hacia la derecha, en potencia máxima (-0.018) Normal, (-0.058) Spearman tiende a una correlación negativa pero la probabilidad es muy mínima.

Tabla 17 Correlación Spearman de la variable potencia entre retry rate con conectividad deficiente

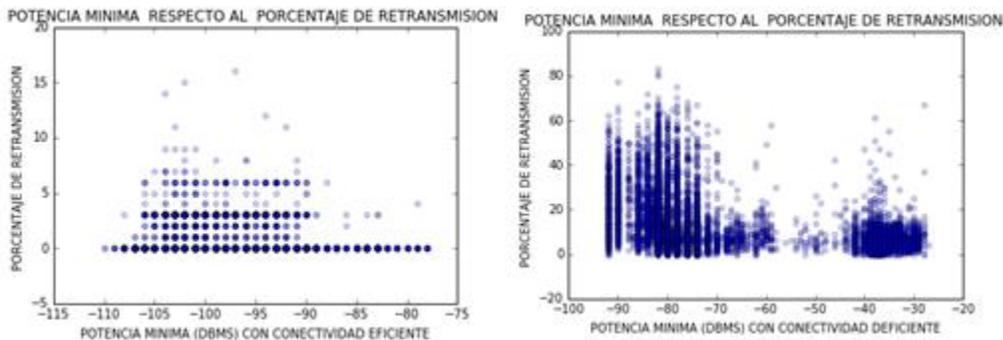
	Retry- metodo Normal	Retry - metodo Spearman
max-potencia	-0.214128	-0.242686
min-potencia	-0.338121	-0.472422
promedio-potencia	-0.363490	-0.489065
moda-potencia	-0.194042	-0.306461

Con el método Spearman y el método normal se puede decir que, en potencia máxima, mínima, promedio y moda, hay correlación, pero tienden a ser una asociación negativa teniendo en cuenta que los resultados tienden a 0 por la izquierda, tiende a una correlación negativa pero la probabilidad es muy mínima.



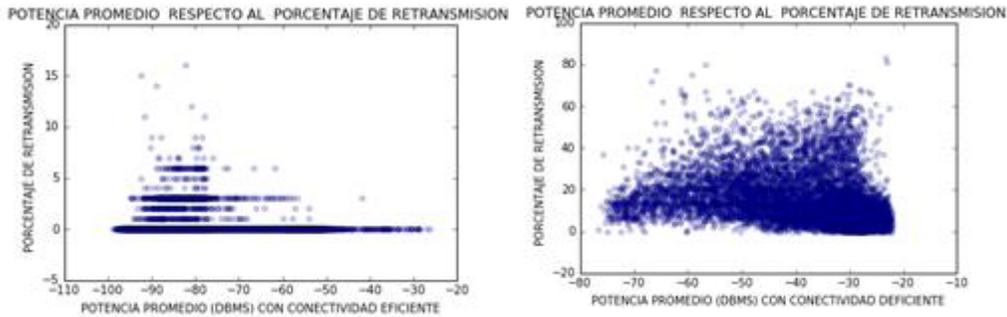
Gráfica 38 Dispersión de la variable potencia máxima respecto al porcentaje de retransmisión

La relación del porcentaje de retransmisión con la potencia máxima en la red con conectividad eficiente y deficiente es diferente de una a la otra, dado que para la red eficiente la potencia en la que se retransmitió se concentra más entre -70 y -90 dBm con un porcentaje menor al 10 por ciento y en la red con conectividad deficiente, el porcentaje de retransmisión se encuentra entre 0 y 80 por ciento entre -15 y -25 dBm mayormente.



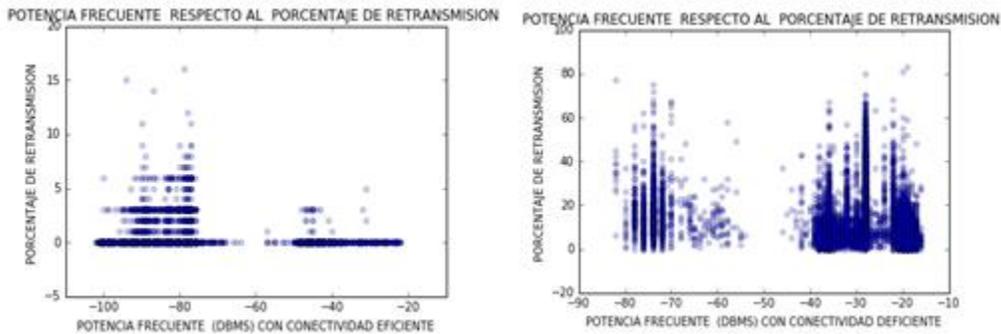
Gráfica 39 Dispersión de la variable potencia mínima respecto al porcentaje de retransmisión

En la gráfica de dispersión de la potencia mínima, el porcentaje de retransmisión se mantiene constante entre 0 y 5 por ciento en un rango de potencias entre -90 y -105 dBm a diferencia de la potencia que se maneja en una red con conectividad deficiente que se divide en dos partes, la primera un porcentaje de retransmisión de 0 a 20 por ciento entre menos -30 y -40 dBm y la segunda de -70 a -90 dBm con un porcentaje de retransmisión mayor al 20 por ciento



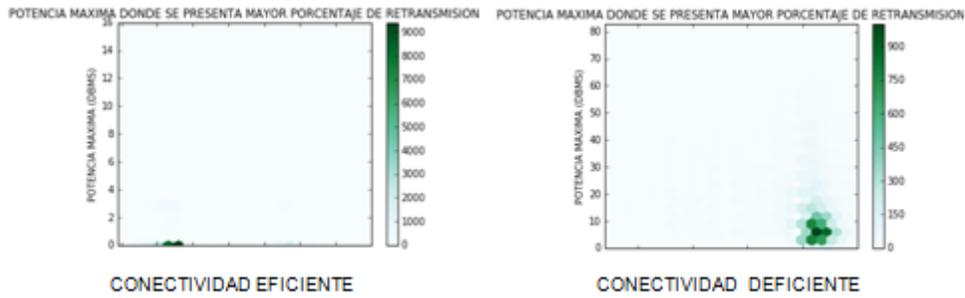
Gráfica 40 Dispersión de la variable potencia promedio respecto al porcentaje de retransmisión

Debido al número de tramas retransmitidas en la red con conectividad eficiente, no se logra una gráfica similar a la red con conectividad deficiente. Pero en este caso las potencias promedio tienden a no ser igual dado que en la red eficiente las potencias son bajas para porcentajes de retransmisión diferente a 0, mientras en la red con conectividad deficiente las potencias son altas en los porcentajes de 0 a 20 por ciento.



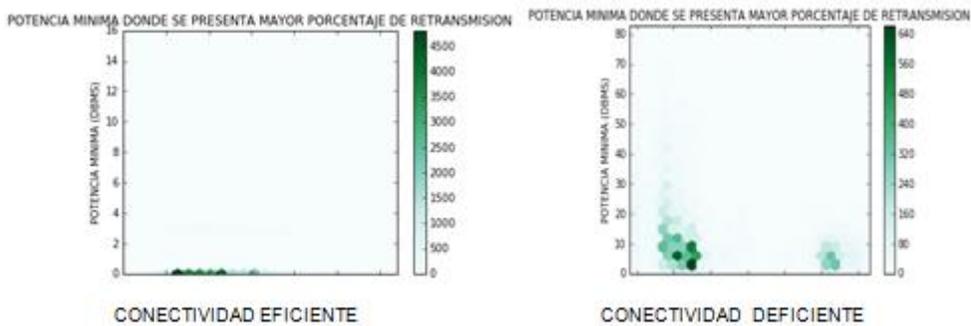
Gráfica 41 Dispersión de la variable potencia frecuente respecto al porcentaje de retransmisión

Gráficas de la potencia frecuente respecto al porcentaje de retransmisión presente en cada intervalo, tienden a tener un resultado inverso ya que potencias frecuentes de las tramas retransmitidas en la red eficiente son en potencias bajas mientras en la red con conectividad deficiente son con potencias altas, recalcando que el porcentaje de retransmisión de la red deficiente es mayor a la red eficiente

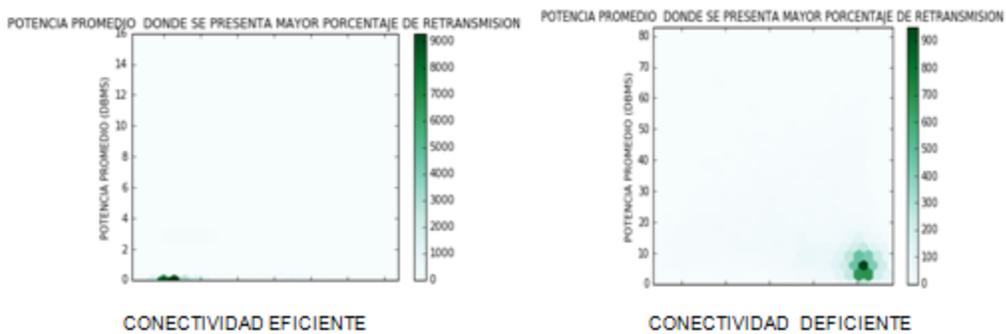


Gráfica 42 Hexagonal binning de la variable potencia máxima respecto al porcentaje de retransmisión

La gráfica binning nos muestra la mayor concentración que para el caso de la potencia máxima se presenta entre 0 y 10 dBm para la red deficiente para la red eficiente no se evidencia dado que el porcentaje de retransmisión es bastante bajo

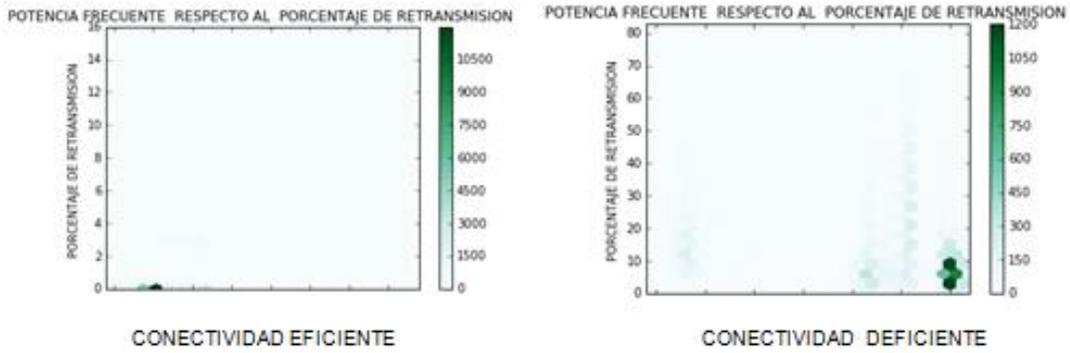


Gráfica 43 Hexagonal binning de la variable potencia mínima respecto al porcentaje de retransmisión



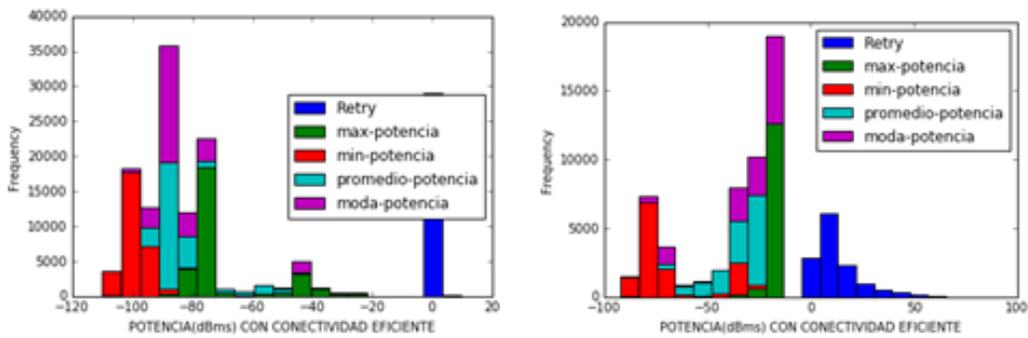
Gráfica 44 Hexagonal binning de la variable potencia promedio respecto al porcentaje de retransmisión

El resultado que se tiene con las gráficas de la potencia promedio son iguales a la gráfica de dispersión de la potencia mínima respecto al porcentaje de retransmisión, en el caso de la red con conectividad eficiente no se genera una concentración con un porcentaje de retransmisión diferente a cero ya que este valor es bajo.



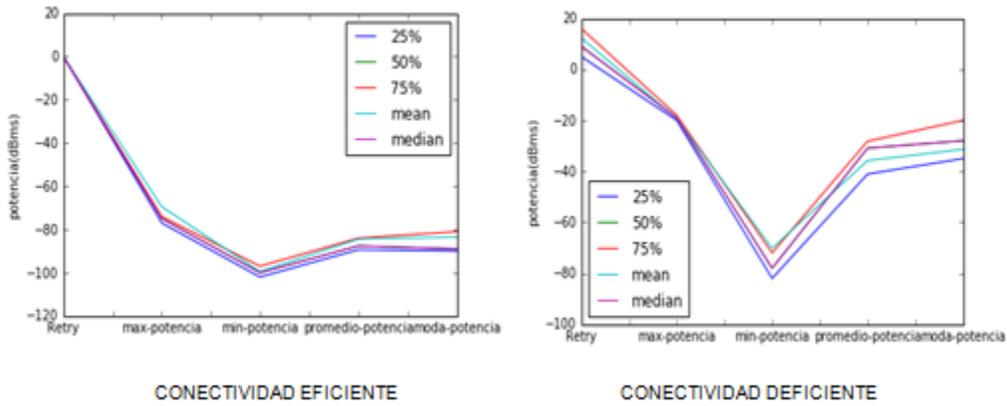
Gráfica 45 Hexagonal binning de la variable potencia frecuente respecto al porcentaje de retransmisión

Gráfica binning de la potencia frecuente de la red con conectividad eficiente y deficiente concentraciones que tienden a ser similares a las demás potencias calculadas.



Gráfica 46 Histograma de potencia

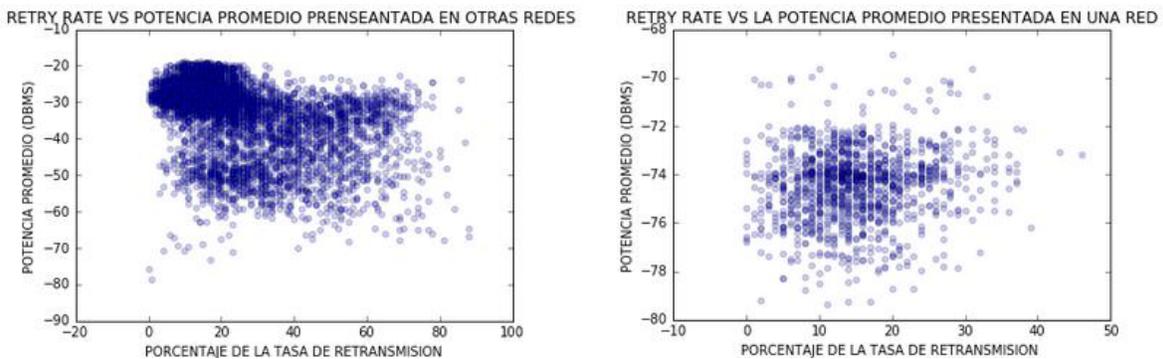
En la gráfica anterior nos muestra el comportamiento de la potencia tanto en la red eficiente como deficiente, en el que se destaca que las potencias promedio y las modas son las que se tiene una mayor frecuencia.



Gráfica 47 Cuartil de potencia

En la gráfica anterior observamos los cuartiles de las potencias mínimas, máximas, promedio y las más frecuentes, en la cual observamos los cuartiles se encuentra en un valor escalar en cada uno de las muestras siendo el 25 % de las potencias con valores bajos, el 50% valores medios a comparación cuarto cuartil en el que el 75 % de las potencias se encuentran los valores más altos. En el 100% de las muestras las potencias máximas son mayores a -20 dBm, una potencia mínima tanto para el 25 %,50% como para el 75% de los datos entre -60 y- 80 dBm. Entre la potencia promedio y la moda se encuentra una similitud con una potencia mayor a -40 y menor a los -20 dBm.

Correlación tasa de retransmisión y potencia de varias redes

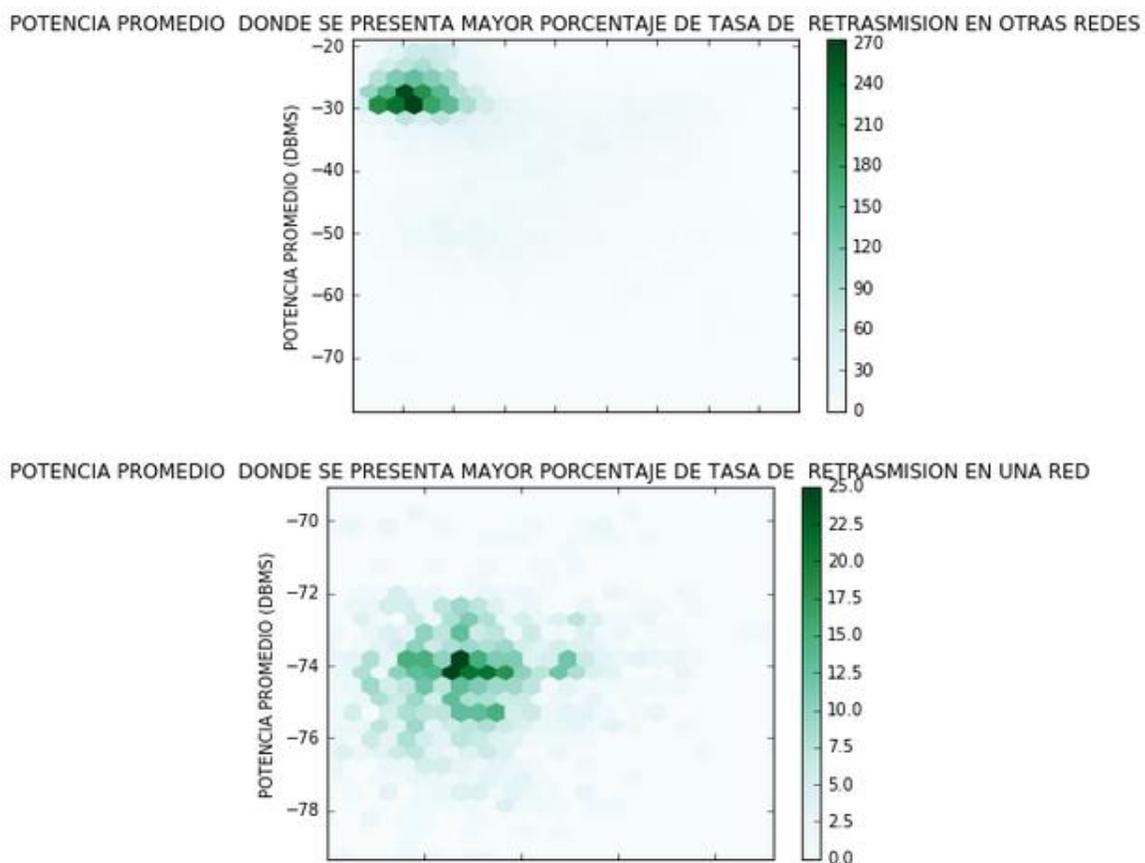


Gráfica 48 Correlación tasa de retransmisión y potencia de varias redes

En el análisis realizado a la tasa de retransmisión y la potencia que se presenta en una red determinada con respecto a las demás redes, las diferencias son

notables. En primera instancia en una red la tasa de retransmisión es de 0% a 20% con una potencia que oscila entre los -70 y -78 dBm, a comparación con las demás redes en que la retransmisión sobre pasa el 60%, donde presenta un promedio de transmisión del 20 % con una potencia entre -20 y -40 dBm. Estas gráficas se obtienen a partir de las capturas realizadas a distintas redes eficientes y deficiente

Estas gráficas no muestran con exactitud la concurrencia de la variable de ciertos valores, pero si da una clara observación de donde existe mayor concentración de la potencia en la que se presentaron las retransmisiones. Y se obtienen como resultado del análisis realizado, la potencia con la que se presentan el mayor porcentaje de retransmisión.



Gráfica 49 Hexagonal-binning potencia promedio donde se presenta mayor porcentaje de tasa de retransmisión en una red

La gráfica es la representación de potencia promedio la cual indica la suma de los valores y los divide para dar un valor medio de lo que se tiene la variable potencia, donde se presenta mayor porcentaje de tasa de retransmisión en una red, podemos observar el punto exacto de concentración los cuales son en -74 dBm con una concentración de 22.5 retransmisiones, y la gráfica de otras redes

donde observamos el punto exacto de concentración los cuales son en -30 dBm con una concentración de 270 retrasmisiones.

7.2.3 Correlación de la variable estaciones

Tabla 18 Correlación de la variable estaciones con respecto a retry rate, conectividad eficiente

	Retry	Numero_estaciones
rangos		
(0, 100]	0	4
(100, 200]	0	3
(200, 300]	0	8
(300, 400]	0	4
(400, 500]	0	3
(500, 600]	0	5
(600, 700]	0	3
(700, 800]	0	5
(800, 900]	0	3
(900, 1000]	0	3
(1000, 1100]	0	2
(1100, 1200]	0	4
(1200, 1300]	0	2
(1300, 1400]	0	3
(1400, 1500]	0	5
(1500, 1600]	0	4
(1600, 1700]	0	2
(1700, 1800]	0	5
(1800, 1900]	0	4
(1900, 2000]	0	4
(2000, 2100]	0	4
(2100, 2200]	0	3
(2200, 2300]	0	4
(2300, 2400]	0	4

	Retry	Numero_estaciones
rangos		
(2400, 2500]	0	8
(2500, 2600]	0	3
(2600, 2700]	3	7
(2700, 2800]	0	6
(2800, 2900]	0	5
(2900, 3000]	0	6
...

La tabla anterior con los rangos de tasa de retransmisión con respecto al variable número de estaciones con conectividad eficiente, todo esto se puede observar en la tabla con rangos de tramas de (0,100] así hasta llegar al rango en este caso hasta (2900,3000] el cual fue escogido para mostrar una pequeña fracción de los rangos que se obtuvieron los cuales son mayores, al observar la tabla se puede dar cuenta que el porcentaje de retransmisión es de 0, el mayor número de estaciones que es 8 se ve reflejada en el rango de (2400,2500], el menor número de estaciones es 2 y se repiten en dos rangos los cuales son (1000,1100] y (1200, 1300]

Tabla 19 Correlación de la variable estaciones con respecto a retry rate, conectividad deficiente

	Retry	Numero_estaciones
rangos		
(0, 100]	14	3
(100, 200]	14	3
(200, 300]	5	3
(300, 400]	10	4
(400, 500]	20	4
(500, 600]	21	5
(600, 700]	31	2
(700, 800]	47	3
(800, 900]	16	7
(900, 1000]	21	3
(1000, 1100]	27	2

	Retry	Numero_estaciones
rangos		
(1100, 1200]	43	3
(1200, 1300]	31	6
(1300, 1400]	66	4
(1400, 1500]	35	4
(1500, 1600]	52	4
(1600, 1700]	54	4
(1700, 1800]	34	5
(1800, 1900]	25	7
(1900, 2000]	27	4
(2000, 2100]	34	3
(2100, 2200]	54	1
(2200, 2300]	35	5
(2300, 2400]	20	4
(2400, 2500]	45	4
(2500, 2600]	46	2
(2600, 2700]	57	1
(2700, 2800]	52	1
(2800, 2900]	61	3
(2900, 3000]	20	6
...

Tabla anterior con los rangos de tasa de retransmisión con respecto al variable número de estaciones con conectividad deficiente, todo esto se puede observar en la tabla con rangos de tramas de (0,100] así hasta llegar al (2900,3000] los cuales se escogieron para mostrar una parte de los datos que se tiene. Por otra parte, se observa el porcentaje de retransmisiones el cual el mayor es de 66 por ciento en el rango de (1300,1400] con 4 número de estaciones, el menor porcentaje de retransmisiones es de 5 en el rango de (200,300] con 3 número de estaciones. El mayor número de estaciones es 7 el cual se presentan en un rango de (800,900] 16 por ciento de retransmisiones.

Tabla 20 Correlación Spearman variable estaciones respecto a retry rate con conectividad eficiente

	RetryNormal	RetrySpearman
Numero_estaciones	-0.115963	-0.122728

Teniendo en cuenta la tabla anterior el método Spearman arroja un resultado de -0.122 y el método normal el resultado fue -0.115) al tener estos resultados se puede decir que en la variable número de estaciones, existe una correlación baja con una relación inversa, dado que los resultados tienden a 0 por la izquierda.

Tabla 21 Correlación Spearman variable estaciones respecto a retry rate con conectividad deficiente

	RetryNormal	RetrySpearman
Numero_estaciones	0.197688	0.174323

Teniendo en cuenta la tabla anterior a la cual se realizó los diferentes métodos como son el método Spearman con un resultado de (0.174) y el método normal con un resultado de (0.197) para una conectividad deficiente. Existe una relación directa, pero con una correlación parcial entre estas dos variables.

Tabla 22 Correlación Spearman con p-valúes variable estaciones respecto a Retry rate con conectividad eficiente

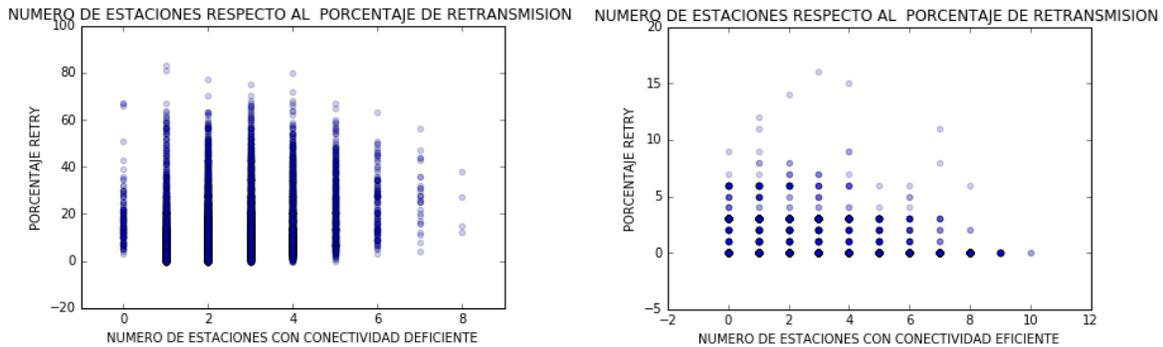
	Corr Spearman Rank(co-corr,p-value)
Retry-Numero_estaciones	(-0.122727617572, 4.2933489147e-98)

En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de número de estaciones con retry en una conectividad eficiente. Retry con respecto al número de estaciones (-0.122, 4.293e-98) estos valores fueron los que arrojaron los cálculos. Dado que el coeficiente de correlación es negativo se puede decir que entre el variable número de estaciones y la tasa de retransmisión existe una correlación inversa, partiendo de lo anterior si la retry aumenta el número de estaciones tiende a disminuir. El p-value nos indica que no se acepta la hipótesis.

Tabla 23 Correlación Spearman con p-values variable estaciones respecto a retry rate con conectividad deficiente

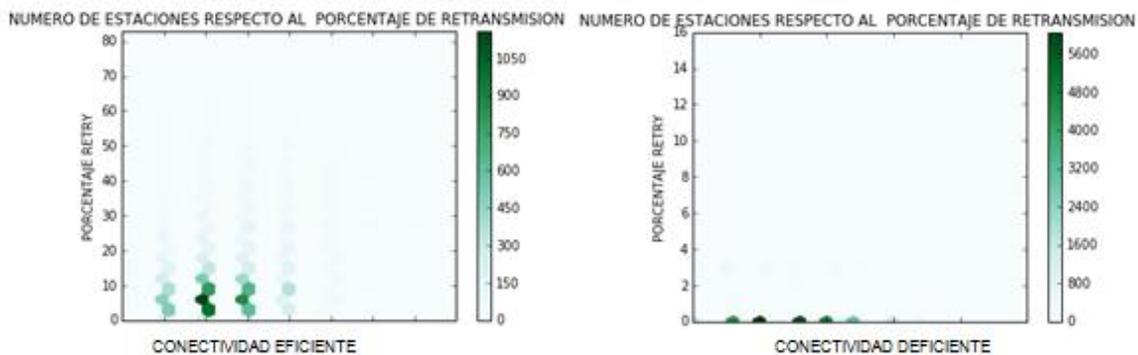
	Corr Spearman Rank(co-corr,p-value)
Retry-Numero_estaciones	(0.174323131274, 1.0862161385e-91)

En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de número de estaciones con retry en una conectividad deficiente. Retry con respecto al número estaciones (0.174, 1.086e-91). Dado que el coeficiente de correlación es positivo se puede decir que entre el variable número de estaciones y la tasa de retransmisión existe una correlación directa, el p-value (1.086e-91) este valor nos indica que se acepta la hipótesis.



Gráfica 50 Dispersión número de estaciones respecto al porcentaje de retransmisión

En la relacione que muestra el porcentaje de retransmisión con respecto al número de estaciones en las redes con conectividad deficiente y eficiente son diferentes tanto en la cantidad de estaciones como el porcentaje de retransmisión, dado que en la red eficiente no supera el 15 por ciento de retransmisión, pero tiene 10 estaciones mientras la red deficiente llega hasta el 80 por ciento de retransmisión tan solo con 8 estaciones.



Gráfica 51 Hexagonal Binning número de estaciones respecto al porcentaje de retransmisión

En la anterior gráfica, la red con conectividad deficiente la mayor parte del porcentaje de retransmisión se da con la entre 2 y 4 estaciones, mientras en la red con conectividad eficiente no se muestra estaciones con porcentajes de retransmisión, dado que son bajos,

7.2.4 Correlación de la variables access points

Tabla 24 Correlación de la variable access points con respecto a retry rate, conectividad eficiente

	Retry	Numero_AP
rangos		
(0, 100]	0	8
(100, 200]	0	8
(200, 300]	0	8
(300, 400]	0	6
(400, 500]	0	8
(500, 600]	0	7
(600, 700]	0	8
(700, 800]	0	8
(800, 900]	0	9
(900, 1000]	0	10
(1000, 1100]	0	10
(1100, 1200]	0	9
(1200, 1300]	0	8
(1300, 1400]	0	9
(1400, 1500]	0	7
(1500, 1600]	0	7
(1600, 1700]	0	8
(1700, 1800]	0	9
(1800, 1900]	0	9
(1900, 2000]	0	6
(2000, 2100]	0	7
(2100, 2200]	0	6
(2200, 2300]	0	6
(2300, 2400]	0	6
(2400, 2500]	0	8

	Retry	Numero_AP
rangos		
(2500, 2600]	0	7
(2600, 2700]	3	6
(2700, 2800]	0	6
(2800, 2900]	0	5
(2900, 3000]	0	6
...

La anterior tabla contiene rangos de tasa de retransmisión con respecto a la variable número de access points con conectividad eficiente, todo esto se puede observar en la tabla con rangos de tramas de (0,100] así hasta llegar al rango en este caso hasta (2900,3000] , se puede observar que hay 3 retransmisiones en el rango de (2600,2700] con 6 número de access points, el mayor número de access points se da en los rangos (900,1000], 1000,1100] con una tasa de retransmisión de 0 por ciento, se puede decir en este caso que la tasa de retransmisión es mínima.

Tabla 25 Correlación de la variables access points con respecto a retry rate, conectividad deficiente

	Retry	Numero_AP
rangos		
(0, 100]	14	2
(100, 200]	14	2
(200, 300]	5	2
(300, 400]	10	2
(400, 500]	20	2
(500, 600]	21	2
(600, 700]	31	3
(700, 800]	47	1
(800, 900]	16	2
(900, 1000]	21	2
(1000, 1100]	27	3
(1100, 1200]	43	2

	Retry	Numero_AP
rangos		
(1200, 1300]	31	1
(1300, 1400]	66	1
(1400, 1500]	35	1
(1500, 1600]	52	1
(1600, 1700]	54	1
(1700, 1800]	34	1
(1800, 1900]	25	1
(1900, 2000]	27	1
(2000, 2100]	34	1
(2100, 2200]	54	1
(2200, 2300]	35	1
(2300, 2400]	20	1
(2400, 2500]	45	1
(2500, 2600]	46	1
(2600, 2700]	57	2
(2700, 2800]	52	1
(2800, 2900]	61	1
(2900, 3000]	20	2

Tabla anterior con los rangos de rata de retransmisión con respecto a la variable número de estaciones con conectividad deficiente, todo esto se puede observar en la tabla con rangos de tramas de (0,100] así hasta llegar al rango en este caso hasta (2900,3000]. Se puede observar el mayor número de retransmisiones con 61 por ciento en un rango de (2800, 2900] con respecto al número de access points 1, 3 access points es el mayor número en los rangos de (600,700] con un número de retransmisiones de 31 por ciento.

Un gran número de access points no indica que el porcentaje de retransmisiones sea mayor y un bajo número de access points, no indica menos retransmisiones.

Tabla 26 Correlación Spearman de la variable access points con respecto a retry rate, conectividad eficiente

	RetryNormal	RetrySpearman
--	--------------------	----------------------

	RetryNormal	RetrySpearman
Numero_AP	0.010023	0.016347

Con el método Spearman el valor obtenido es (0.010) y el método normal el valor obtenido es (0.016) se puede decir que la variable número de access points no hay correlación con respecto a retry, dado que el coeficiente en ambos casos es muy mínimo.

Tabla 27 Correlación Spearman de la variable access points con respecto a retry rate, conectividad deficiente

	RetryNormal	RetrySpearman
Numero_AP	0.270598	0.353831

Los resultados dados por el método Spearman y el método normal si puede decir que el número de access points con respecto al porcentaje de retransmisión hay correlación pero tienden a ser una asociación positiva teniendo dado que los resultados tienden a 0 por la derecha.

Tabla 28 Correlación Spearman de con p-value la variable access points con respecto a retry rate, conectividad eficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-Numero_AP	(0.0163466011572, 0.00528132432335)

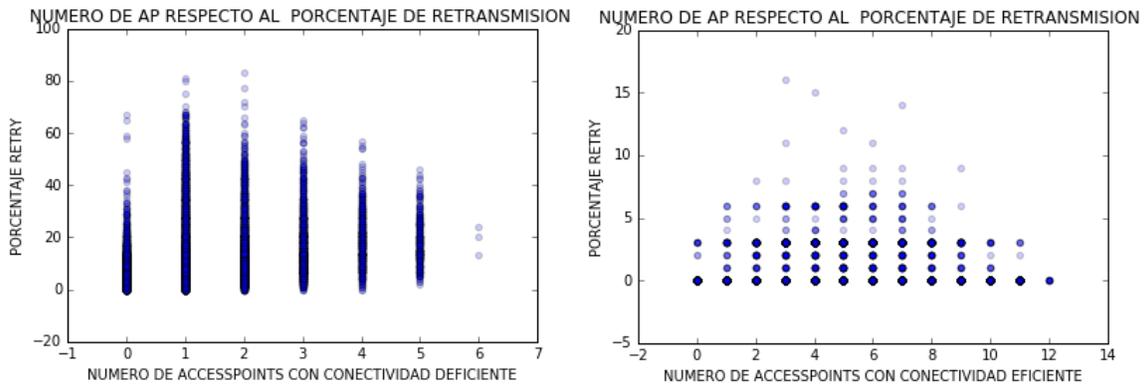
En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) de los números access points con una conectividad eficiente. Dado que el valor de retry con respecto al número de access points es de (0.016, 0.005) se puede decir que entre la variable número access points y la tasa de retransmisión existe una relación directa con una correlación lineal muy mínima.

Tabla 29 Correlación Spearman de con p-value la variable access points con respecto a retry rate, conectividad deficiente

	Corr Spearman Rank(co-corr,p-value)
Retry-Numero_AP	(0.353831422842, 0.0)

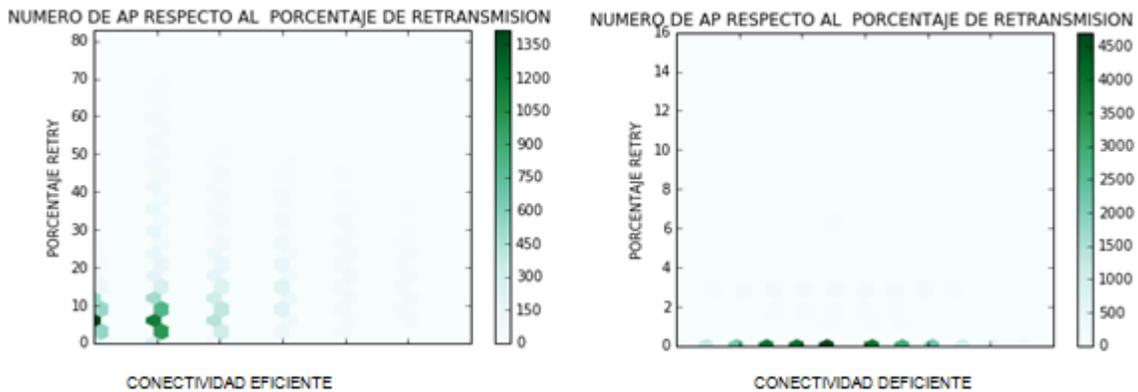
En esta tabla se observa los valores de coeficiente de correlación (co-corr), el valor p(p-value) del número de access points con retry con resultados (0.353, 0.0) en una conectividad deficiente. Dado que el coeficiente de correlación es positivo se puede decir que entre la variable access points y la tasa de retransmisión existe

una correlación directa. El p-value es de (0.0) con este resultado se puede decir que se rechaza la hipótesis.



Gráfica 52 Dispersión número de access points respecto al porcentaje de retransmisión

En la relación que muestra el porcentaje de retransmisión con respecto al número de access points en las redes con conectividad deficiente y eficiente son diferentes tanto el número de access points como el porcentaje de retransmisión, dado que en la red eficiente no supera el 15 por ciento con 12 access points, mientras la red deficiente llega hasta el 80 por ciento de retransmisión tan solo con 5 access points.



Gráfica 53 Hexagonal binning número de access points respecto al porcentaje de retransmisiones

En la anterior gráfica, la red con conectividad deficiente, la mayor parte del porcentaje de retransmisión se da con la presencia de pocos access points, mientras que en la red con conectividad eficiente se muestran access points con porcentajes de retransmisión bajos.

8 LISTA DE ACRÓNIMOS

AP	Accespoint
CTS	<i>clear to sent Formato de la trama de control</i>
RTS	<i>request to send Formato de trama de control</i>
ACK	<i>Tramas acknowledgement</i>
TBTTs	<i>Target Beacon Transmission Time</i>
TIM	<i>(solo presente en tramas generadas por ap)</i>
IBSS	<i>BSS independiente</i>
ID	<i>Identidad</i>
SSID	<i>The service set identifier (SSID) is a network name used to identify an 802.11 wireless network.</i>
RA	<i>the receiver address</i>
TA	<i>the transmitter address</i>
FCS	<i>frame check sequence, es la secuencia de verificación de trama</i>
DA	<i>Contiene la dirección Mac del nodo de destino final en la red.</i>
SA	<i>Contiene la dirección Mac del nodo que inició la trama.</i>
RA	<i>Contiene la dirección Mac que identifica al dispositivo inalámbrico que es el destinatario inmediato de la trama.</i>
TA	<i>Contiene la dirección Mac que identifica al dispositivo inalámbrico que transmitió la trama.</i>
FCS	<i>Contiene una comprobación de redundancia cíclica (crc) de 32 bits de la trama.</i>
AID	<i>Para la estación que transmitió la trama.</i>
DHCP	<i>dynamic host configuration protocol</i>
UDP	<i>user datagram protocol</i>
TCP	<i>transmission control protocol</i>
IP	<i>internet protocol</i>
ARP	<i>address resolution protocol</i>
FTP	<i>file tranfer protocol</i>
HTTP	<i>hypertext transfer protocol</i>
POP	<i>post office protocol</i>
SMTP	<i>simple mail transfer protocol</i>
TELNET	<i>telecommunication network</i>
ICMP	<i>internet control message protocol</i>
DNS	<i>domain name system</i>
SSID	<i>service set identifier</i>
BSS	<i>basic server set</i>
ESS	<i>extended service set</i>

CRC	<i>cyclic redundancy check</i>
RTT	<i>Round-Trip delay Time</i>
CSMA/CD	<i>Carrier Sense Multiple Access</i>
ARP	<i>address resolution protocol</i>
TCP/IP	<i>transmission control protocol</i>
RTS	<i>real time transport protocol</i>
HDLC	<i>High-Level Data Link Control</i>
LLC	<i>logical link control</i>
MAC	<i>Message Authenticity Check</i>
MAC	<i>Media Access Control</i>
STA	<i>Station</i>
RF	<i>Radio frequency</i>

9 CONCLUSIONES

- En las redes a las cuales se le realizó el debido análisis, en las horas pico se observó que existe un gran porcentaje de retransmisión teniendo en cuenta que la potencia con la cual se reciben los paquetes son inferiores a los -30 dBm, esto sucede en la gran mayoría de las redes analizadas con conectividad deficiente.
- Se realizaron capturas a dos tipos de redes según su rendimiento en la cual se obtuvo:
 - En una red con conectividad eficiente, donde la percepción de funcionamiento es bueno, el porcentaje de retransmisiones no superó el 30%.
 - En una red con conectividad deficiente, donde la percepción de funcionamiento es bueno, el porcentaje de retransmisiones supera el 70%
- Existen redes que no presente retransmisiones, pero esto no quiere decir que sean redes con una conectividad eficiente o algunas donde su porcentaje de retransmisión sean altos el cual no indica que sea una red con conectividad deficiente, por lo tanto, el problema puede estar en la red cableada.
- Según los resultados en las tablas de frecuencias, mayor cantidad de estaciones no indica mayor porcentaje de retransmisión, y menor cantidad de estaciones no indica menor porcentaje de retransmisión. Esto nos permite decir que la conectividad de una red, si es eficiente o deficiente, no depende únicamente del número de estaciones
- En el análisis de los datos se encontró redes eficientes con gran número de access points y redes deficientes con pocos access points. Esto nos permite concluir que el mayor número de access points no infiere en el porcentaje de retransmisiones, o que sea proporcional, dicho de otra forma, que a mayor número de access points mayor tasa de retransmisión o a menor número de access points menor tasa de retransmisión.
- Al realizar el estudio de la potencia de una red seleccionada con respecto a las demás cercanas, se observa que la red por si sola maneja una tasa de retransmisión entre 0% y 66% con una potencia comprendida entre los -97.58 dBm a -92.18 dBm. Y en paralelo con las demás redes, presentan

una dispersión más grande en la que se llega tasas de retransmisiones mayores del 66 % con una potencia de -28 dBm según lo observamos en la gráfica de dispersión hexagonal. a partir de lo anterior una sola red tiene una tasa de retransmisión no muy alta con potencias bajas, resultado inverso a los que sucede con las demás redes, en la que mantiene porcentajes de retransmisión y potencias altas. por consiguiente, se puede deducir que se debe a interferencia entre las mismas redes.

10 ANEXOS

10.1 Captura de datos por medio de la herramienta Wireshark

Para el inicio de Wireshark, se realiza por consola con el comando *sudo* para ejecutarlo con privilegios de seguridad del usuario *root*, como lo muestra la *imagen 1*.

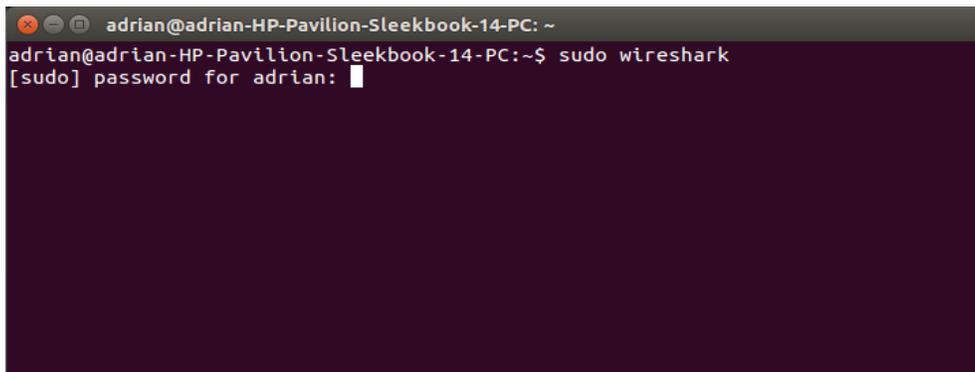


Ilustración 12 inicialización Wireshark

Al inicializar Wireshark la herramienta nos muestra una interfaz *imagen 2*, donde tenemos link de documentación, abrir archivos y las opciones de captura. Para nuestro caso seleccionamos wlan0 que con la que vamos a trabajar para la captura de redes Wi-Fi.

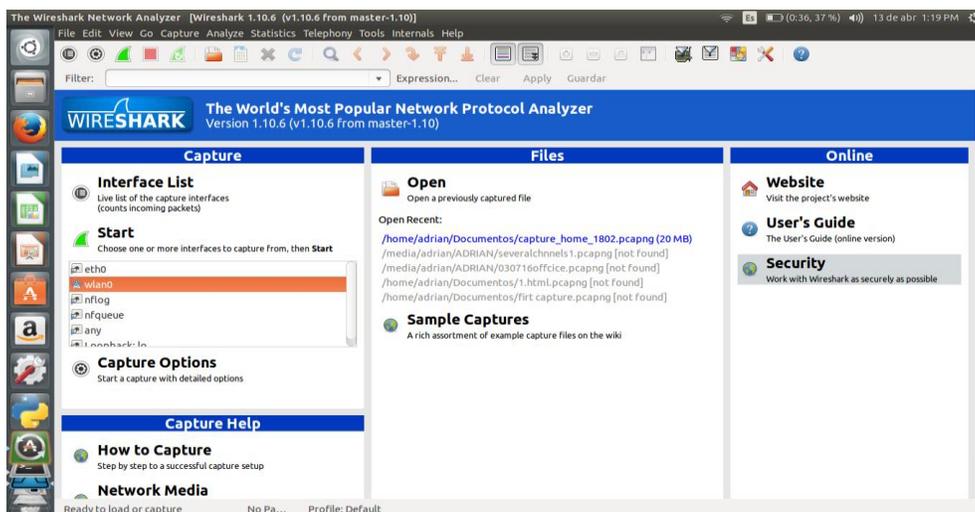


Ilustración 13 interfaz de la herramienta Wireshark

Para realizar la captura de la red Wi-Fi, configuramos la captura en modo monitor que es la que nos permite ver las tramas 802.11 y deshabilitar los filtros SSID y canales típicos de la wlan lo que nos permite capturar tráfico independientemente del SSID y el canal de tráfico.

En el medio inalámbrico al ser emisiones de radio éstas llegan a cualquier máquina dentro de su radio de alcance por lo que se asemejaría a una topología Hub en una red cableada. Esto lo realizamos dando doble clic sobre el interfaz de captura y seleccionado en modo monitor (*imagen 3*). O por la terminal dando el comando:

```
aircom-ng start wlan0
```

En modo *root*, este script se utiliza para activar en modo monitor las interfaces inalámbricas y volver en modo administrador.

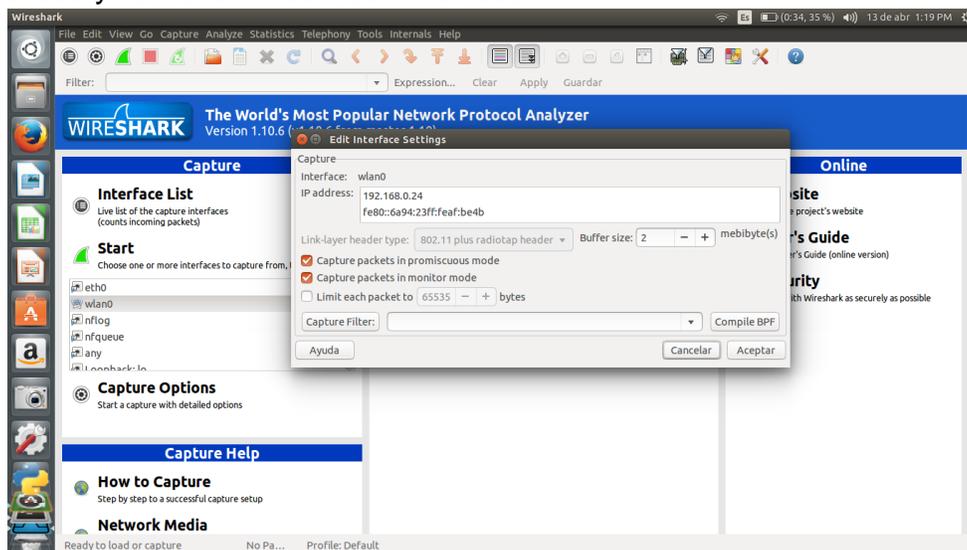


Ilustración 14 configuración de modo de captura inalámbrica

Se inicia la captura seleccionando “start”, (*imagen 3*) esto dará inicio a la captura y muestra de las tramas capturadas. Para finalizar la captura seleccionamos el recuadro rojo en la parte superior, derecha de la interfaz como lo muestra la siguiente imagen

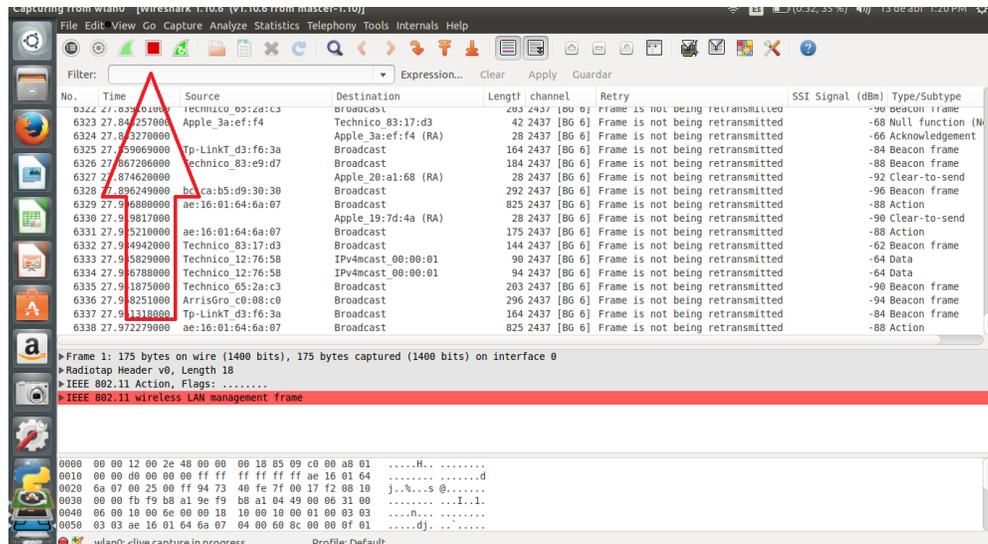


Ilustración 15 captura con la herramienta Wireshark

Al exportar la captura en formato “csv” se procede a guardar la captura y luego seleccionamos file → export packet dissections → as csv (figura 5), en la ventana emergente seleccionamos todos los paquetes “all packets” y damos un nombre a la exportación y damos aceptar, (figura 6).

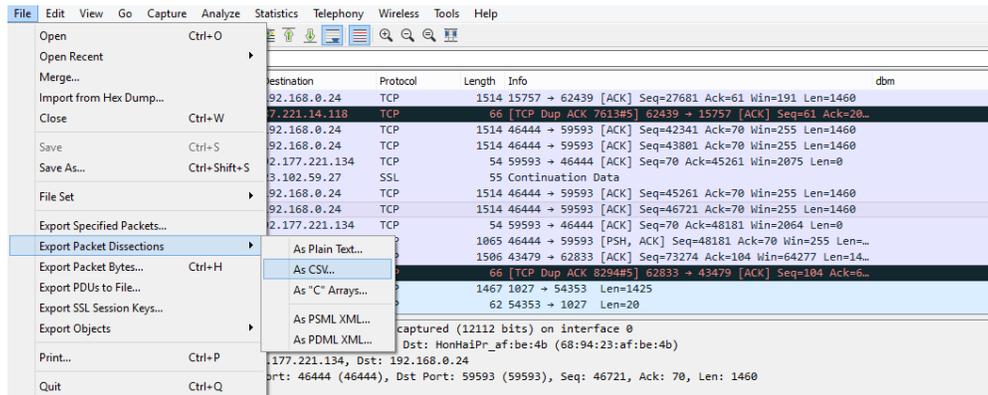


Ilustración 16 exportar un archivo a formato csv

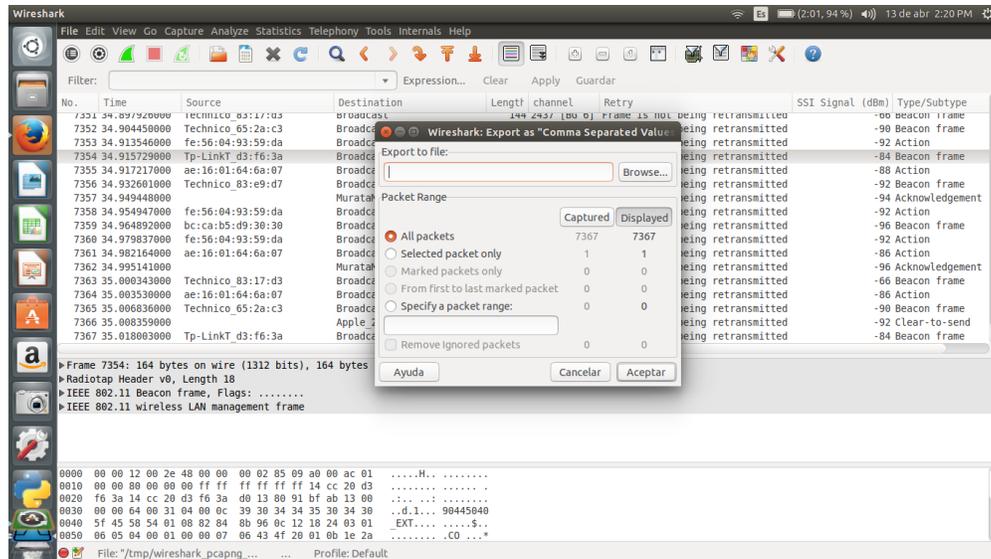


Ilustración 17 guardar archivo en formato CVS

10.2 Procesamiento de los datos capturados

Python es un lenguaje que ha tomado fuerza en los últimos años por ser un lenguaje multiparadigma y de código libre, en que se desarrollan módulos para el análisis de datos con una perspectiva estadística haciendo uso de algoritmos de Machine Learning. Por estas razones se elige para realizar el procesamiento de los datos capturados en Wireshark, las librerías pandas, numpy, matplotlib, scipy son las más frecuentes e implementadas en la investigación.

En el proceso la implementación de python se hace uso del administrador de paquetes anaconda para un mejor despliegue. Anaconda contiene por defecto ipython componente del paquete scipy, shell en la que se realiza todo el procesamiento y análisis de los datos capturados para la investigación, ya que esta integra varios módulos, añadiendo funcionalidades extras de un modo más interactivo por medio de la interfaz notebook más conocido como Jupyter.

Luego de tener instalado anaconda y configurado el ambiente de ipython, se inicia Jupyter notebook mediante el siguiente comando en la terminal de linux.

```
C:\Users\yearr>ipython notebook
```

Ilustración 18 iniciar notebook por terminal

Al ejecutar ipython, el panel de control de jupyter notebook se mostrará en el navegador con la siguiente dirección "<http://localhost:8888/tree>", de allí se podrá navegar por los diferentes directorios que contienen los cuadernos existentes.

 jupyter



Ilustración 19 interfaz jupyter notebook

Para el respectivo análisis de los datos, luego de tener identificadas las variables, se hace uso de los notebook, que permite crear un proyecto “cuaderno” por cada variable con su respectivo análisis, para ello damos clic en nuevo, para su creación y la ejecución del código en python.

La siguiente imagen muestra un notebook cuando es creado y no se ha realizados ejecuciones de código.

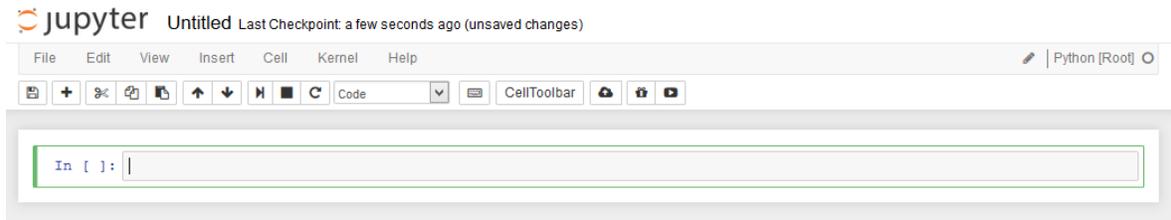


Ilustración 20 hoja de trabajo ipython

Para la ejecución del análisis de los datos se realiza la importación de librerías y el cargue de datos tanto en la caracterización como en la correlación. En la siguiente sección se describe las librerías necesarias e importadas y la carga de datos con su respectivo código python.

10.3 Importación de librerías

En cada uno de los notebook creados para el análisis de los datos se realiza la importación de la librería PANDAS que ofrece para Python unas estructuras de datos fáciles de usar y alto performance, por medio de funciones esenciales para el análisis de datos, la librería MATPLOTLIB para visualizaciones gráficas y NUMPY paquete fundamental para la manipulación de datos, permitiendo operaciones aritméticas sobres los elementos del data frame. Cabe resaltar que se

asignan alias en las importaciones, para realizar los llamados de las funciones que comprenden estas librerías en una forma corta.

```
import pandas as pd
import numpy as np
import matplotlib as p
% matplotlib inline
```

 (1)

La notación `% matplotlib inline` se ejecuta para la visualización de las gráficas en el notebook y no se despliegan o se muestran en otra ventana emergente.

11 BIBLIOGRAFÍA

- [1] 8. w. n. t. d. guide, 802.11 wireless networks the definitive guide, O'Reilly, 2005.
- [2] M. P. N. Gil, «Biblioteca universidad de servia,» [En línea]. Available: <http://bib.us.es/salud/sites/bib3.us.es.salud/files/revesppod4xxiii22012.pdf>. [Último acceso: 8 marzo 2017].
- [3] «statstutor,» [En línea]. Available: <http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf>. [Último acceso: 1 marzo 2017].
- [4] M. P. N. Gil, «biblioteca universidad de sevilla,» [En línea]. Available: <http://bib.us.es/salud/sites/bib3.us.es.salud/files/revesppod4xxiii22012.pdf>. [Último acceso: 8 marzo 2017].
- [5] «statstutor,» [En línea]. Available: <http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf>. [Último acceso: 1 marzo 2017].
- [6] M. S. Gast, 802.11 Wireless Networks, O'Reilly Media, 2005.
- [7] M. Gast, 802.11 wireless networks the Definitive Guide Second, 2005.
- [9] J. M. a. b. T. Seoane a, SEMERGEN - Medicina de Familia, Elsevier España: Copyright © 2007 Elsevier España S.L y Sociedad Española de Medicina Rural y Generalista (SEMERGEN). All rights reserved., 2007.
- [10] R. Shemonski, The Wireshark Field Guide, 2013.
- [11] v. manzano, «asignatura,» 2012. [En línea]. Available: <http://asignatura.us.es/dadpsico/apuntes/EstimacionEstadistica.pdf>. [Último acceso: 25 12 2016].
- [12] M. Gast, «802.11 wireless Networks the Definitive Guide,» de *802.11 wireless networks*, O'Reilly, 2005, p. 656.
- [13] B. d. I. Fuente, Correlación entre variables. Modelos de regresion simple (lineal,cuadratica,cubica), modelos de regresion multiple.
- [14] A. C. Estrada, SEGURIDAD POR NIVELES, MADRID, 2011.
- [15] A. B. Downey, «thinkstats2,» 2014. [En línea]. Available: <http://thinkstats2.com>. [Último acceso: 1 septiembre 2016].
- [16] D. A. W. David D. Coleman, Certified Wireless, Indiana: Copyright, 2006.
- [17] J. B. ,. K. ,. O. ,. P. ,. S. S. Cuervo Alder, «ScineceDirect,» 2007. [En línea]. Available: <http://bibliotecasenlinea.unillanos.edu.co:2051/science/article/pii/B9781597491709500205>. [Último acceso: 19 10 2016].

- [18] J. A. J. A. C. F. Carballar, de *Wi-Fi : lo que se necesita conocer*, 2010, p. 224.
- [19] «Wireshark,» [En línea]. Available: <https://www.Wireshark.org/>. [Último acceso: 10 08 2016].
- [20] «WIKIPEDIA,» 18 MARZO 2016. [En línea]. Available: https://es.wikipedia.org/wiki/Prueba_anecd%C3%B3tica. [Último acceso: 25 12 2016].
- [21] «significados,» [En línea]. Available: <http://www.significados.com/Wi-Fi/>. [Último acceso: 15 10 2016].
- [22] «minitab 17,» [En línea]. Available: <http://support.minitab.com/es-mx/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/what-is-a-hypothesis-test/>. [Último acceso: 25 12 2016].
- [23] «meccanismocomplesso,» [En línea]. Available: <http://www.meccanismocomplesso.org/en/hexagonal-binning/>. [Último acceso: 26 01 2017].
- [24] «IoT -A,» 1 JUNIO 2009. [En línea]. Available: <http://iot-spain.com/?p=3547>. [Último acceso: 24 NOVIEMBRE 2015].
- [25] «http://www.bdat.net/seguridad_en_redes_inalambricas/x187.html,» [En línea]. Available: http://www.bdat.net/seguridad_en_redes_inalambricas/x187.html. [Último acceso: 10 08 2016].
- [26] «<http://wiki.inf.utfsm.cl/images/thumb/b/b5/Trama80211.png/450px-Trama80211.png>,» [En línea]. Available: <http://wiki.inf.utfsm.cl/images/thumb/b/b5/Trama80211.png/450px-Trama80211.png>. [Último acceso: 10 08 2016].
- [27] «<http://bibing.us.es/proyectos/abreproy/11306/fichero/TEORIA%252F07+-+Capitulo+2.pdf>,» [En línea]. Available: <http://bibing.us.es/proyectos/abreproy/11306/fichero/TEORIA%252F07+-+Capitulo+2.pdf>. [Último acceso: 14 10 2016].
- [28] «Google book,» 20 ENERO 2004. [En línea]. Available: <https://books.google.com.co/books?id=O0PFAAAACAAJ&dq=herramientas+utilizadas+en+informatica+forense&hl=es&sa=X&ved=0ahUKEwjD0aOzn67JAhVM7yYKHSuwBslQ6AEIQTAE>. [Último acceso: 24 NOVIEMBRE 2015].
- [29] «gobierno digital,» 12 febrero 2015. [En línea]. Available: <http://colombiadigital.net>. [Último acceso: 24 noviembre 2015].
- [30] «Definicion ABC,» [En línea]. Available: <http://www.definicionabc.com/tecnologia/histograma.php>. [Último acceso: 25 12 2016].
- [31] «aula virtual universidad nacional de cordoba,» [En línea]. Available: <http://www.fca.proed.unc.edu.ar/mod/book/view.php?id=3270>. [Último acceso:

28 12 2016].

- [32] «Biblioteca de ingeniería Universidad de Sevilla,» [En línea]. Available: <http://bibing.us.es/proyectos/abreproy/11306/fichero/TEORIA%252F07+-+Capitulo+2.pdf>. [Último acceso: 8 marzo 2017].
- [33] A. C. Estrada, «Departamento de Tecnologías de la Información e las Comunicaciones,» [En línea]. Available: <http://www.tic.udc.es/~nino/blog/lasi/documentos/seguridad-por-niveles.pdf>. [Último acceso: 1 septiembre 2016].
- [34] A. C. Estrada, «Departamento de Tecnologías de la Información e las Comunicaciones,» [En línea]. Available: <http://www.tic.udc.es/~nino/blog/lasi/documentos/seguridad-por-niveles.pdf>. [Último acceso: 1 septiembre 2016].
- [35] A. C. Estrada, «DarFe,» [En línea]. Available: <http://www.darfe.es/joomla/>. [Último acceso: 1 septiembre 2016].
- [36] «wikipedia,» [En línea]. Available: [http://wiki.inf.utfsm.cl/index.php?title=EEE_802.11_wireless_LANs_\(%E2%80%9Cwi-%EF%AC%81%E2%80%9D\)](http://wiki.inf.utfsm.cl/index.php?title=EEE_802.11_wireless_LANs_(%E2%80%9Cwi-%EF%AC%81%E2%80%9D)). [Último acceso: 1 septiembre 2016].
- [37] «mrn-cciew,» [En línea]. Available: <https://mrncciew.com/2014/10/08/802-11-mgmt-beacon-frame/>. [Último acceso: 1 septiembre 2016].
- [38] «flylib,» [En línea]. Available: <http://flylib.com/books/en/2.519.1.35/1/>. [Último acceso: 1 SEPTIEMBRE 2016].
- [39] «flylib,» [En línea]. Available: <http://flylib.com/books/en/2.519.1.35/1/>. [Último acceso: 1 SEPTIEMBRE 2016].
- [40] «flylib,» [En línea]. Available: <http://flylib.com/books/en/2.519.1.35/1/>. [Último acceso: 1 SEPTIEMBRE 2016].
- [41] «flylib,» [En línea]. Available: <http://flylib.com/books/en/2.519.1.35/1/>. [Último acceso: 1 SEPTIEMBRE 2016].
- [42] G. Benitez, 13 12 2016. [En línea]. Available: <http://www.ingenieriasystems.com/2016/12/Trama-inalambrica-80211-y-Resumen-CCNA1-V5-CISCO-C4.html>. [Último acceso: 20 DICIEMBRE 2016].
- [43] [En línea]. Available: <https://statistics.laerd.com/statistical-guides/spearman-rank-order-correlation-statistical-guide.php>. [Último acceso: 21 marzo 2017].