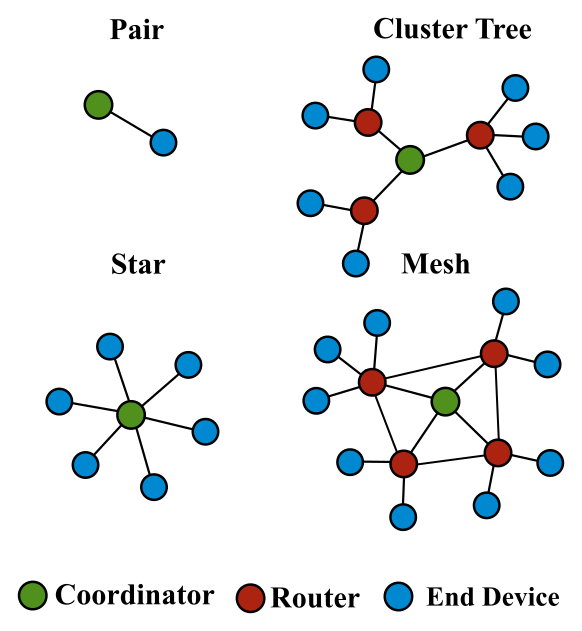
**Objetivo de la monitorización**

Identificar los factores ambientales presentes en un cultivo de cacao en zona rural del municipio de San Cayetano, N.S, Colombia, y determinar las acciones necesarias para mejorar la producción y controlar la proliferación de plagas.

**Red inalámbrica de Sensores**

La red inalámbrica de sensores se implementó siguiendo la topología *ZigBee Cluster Tree*, con 7 nodos: un nodo coordinador controlado con una *Raspberry Pi 3* y tres pares de nodos enrutadores - nodos finales, controlados con microcontroladores ATMega 328. En la figura 1 se puede observar la configuración escogida y la distribución de los nodos.

****

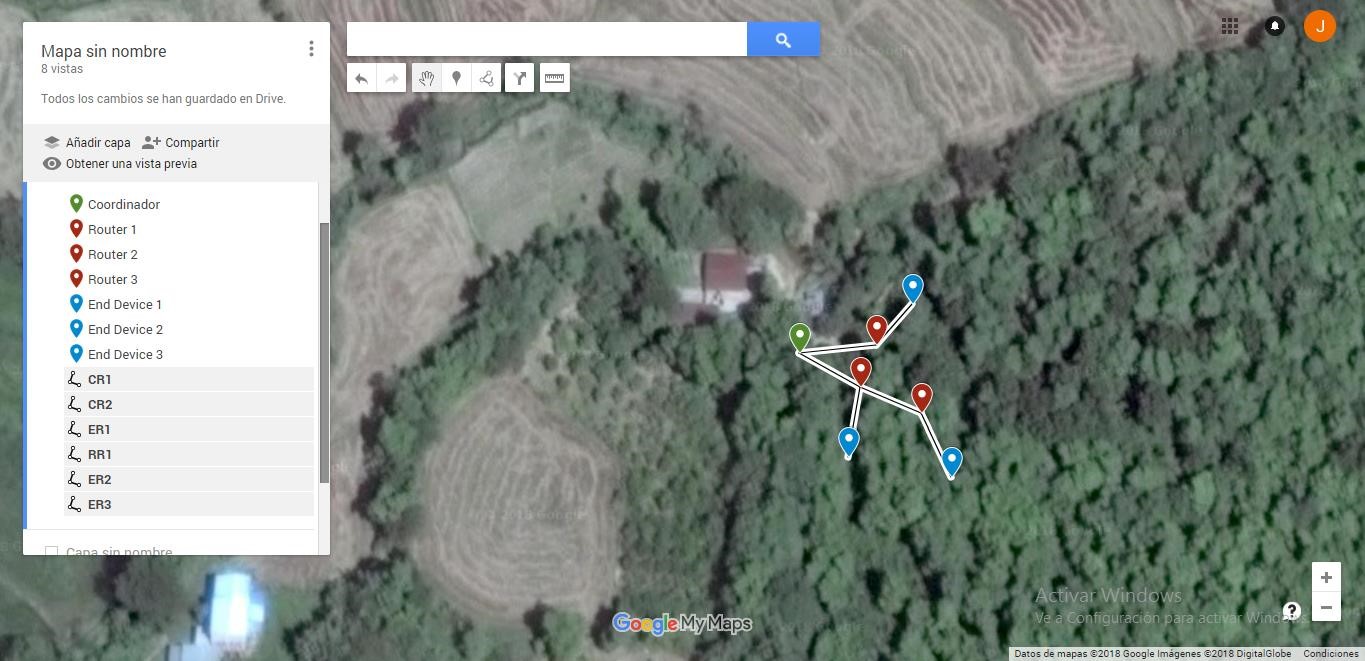
****

Fig. 1. Topologías ZigBee, vista satelital de la red implementada.

**Instrumentos de medición**

Los nodos *router* y *end device* integraron tres sensores del fabricante DFRobot para la medición de temperatura, humedad relativa, humedad del suelo e intensidad lumínica:

* **Temperatura y humedad relativa (Sensor DHT11)**
* Temperatura de (20°*C*−50°*C*)±5°*C*
* Humedad Relativa de (20%−90%)±2%
* **Humedad de la Tierra (basado en transistor NPN)**
* 0 – 100 Pw
* **Intensidad Lumínica (Sensor PT550)**
* 1 Lux a 6000 Lux

Los nodos se instalaron sobre las ramas de los árboles de cacao, ubicando los sensores DHT11 y PT550 a una altura de 1,5 m sobre el suelo, los sensores de intensidad lumínica se sembraron a una profundidad de más o menos 8 cm.

**Transmisión y almacenamiento**

Los valores registrados en los nodos se ordenan en una trama de datos en modo API y se transmiten hasta el nodo coordinador, posteriormente los datos se extraen de la trama de datos y se almacenan en una base de datos SQL alojada en un servidor Apache instalado en la Raspberry Pi 3.

**Información recopilada**

Se dispone de datos para tres periodos de medición distintos registrados con los nodos 2, 3 y 5:

* 09 de mayo 10:25 a.m. - 10 de mayo 02:38 p.m.
* 23 de mayo 10:08 a.m. - 26 de mayo 07:01 a.m.
* 05 de junio 10:21 a.m. - 08 de junio 06:40 p.m.

**El algoritmo**

Se aplica un filtro para los valores que pueden presentarse en los registros de cada variable.

1. Se selecciona un rango de valores permitidos para las variables, con base en límites ingresados por teclado y el rango de Tukey para la variable.

* El rango ingresado puede considerar el rango de medición de los sensores o ser producto de un análisis gráfico de las variables (histograma, gráfico de dispersión, entre otros)
* El rango de Tukey.
  + Diagrama de cajas: representación gráfica de la distribución de los datos, permite identificar el valor mínimo y máximo, la mediana de los datos, los caurtiles, el rango intercuartil, los límites o bisagras de Tukey y la presencia de valores atípicos extremos.
  + Cuartiles (**Q**): indicadores de los valores que corresponden al 25%, 50% y 75% de los valores de los datos.
  + Primer cuartil (**Q1**): valor mayor o igual al 25% de los datos.
  + Segundo cuartil (**Q2**): valor mayor o igual al 50% de los datos, corresponde a la media de los datos.
  + Tercer cuartil (**Q3**): valor mayor o igual al 75% de los datos.
  + Rango intercuartil (RIQ): distancia entre Q1 y Q3.
  + Rango de Tukey, establece límites de confianza de los datos, de modo que cualquier valor que se encuentra a una distancia de por debajo de Q1 o por encima de Q3 es considerados como atípico.
* Histograma: representación de la distribución de los datos según sus valores y cantidad de veces que estos se repiten.
* Correlación: indica la dirección y la fuerza de la relación entre dos variables.
* Diagrama de dispersión: representación gráfica de la relación entre dos variables.
* Se hace un recorrido de valores entre el límite inferior escogido (desde ahora se nombrará ) y el límite inferior de Tukey (). Según la diferencia entre se realiza un total de límites y iteraciones.
* Para cada iteración se buscan valores menores a . Cuando se presenta un valor () por debajo de éste se reemplaza por el equivalente al promedio de los valores registrados en días anteriores exactamente a la misma hora.
* Luego de haber reemplazado todos los valores atípicos para una iteración se procede se calcula la correlación () entre la variable analizada () y otra variable fuertemente relacionada con ésta () y se almacena en un diccionario con el respectivo valor de .
* ……

A partir de un procedimiento iterativo en el que se recorren los límites superiores e inferiores entre un rango escogido y el rango de Tukey, se determina el rango que garantiza la mejor correlación entre la variable filtrada y la variable que presenta la mayor relación con ésta.

Con la aplicación de este filtro se consigue la eliminación de valores atípicos extremos, que pueden corresponder a descalificación de los sensores o una falla en el funcionamiento del mismo por defectos de fabricación (Se sospecha que también pudieron ser causados por la presencia de insectos dentro de los nodos, dado que éstos se encontraban a la intemperie y en las visitas realizadas al escenario de medición se encontraron chinches y hormigas dentro de los nodos).

Posteriormente se aplican ventanas móviles y se determinan los registros que presentan mayores desviaciones estándar, éstos son reemplazados por el valor correspondiente de la mediana móvil obtenida. De manera que se suavizan los atípicos enmascarados dentro del rango de datos.

Sin embargo al algoritmo le falta reducir el impacto que pueden causar los atípicos en los valores siguientes su ocurrencia