

Hacia una aproximación estructurada al diseño de redes de sensores inalámbricos (1ª Parte)

Por Ilya Bagrak, MeshNetics

Artículo cedido por Next For S.A., distribuidor de MeshNetics en España



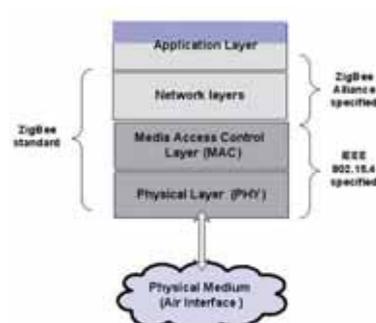
Figura 2. Topología de red en malla de ZigBee.

He aquí algunos enfoques prácticos, consejos y trucos para el diseño de redes de sensores inalámbricos que pueden ahorrarle muchos días del ciclo completo de diseño.

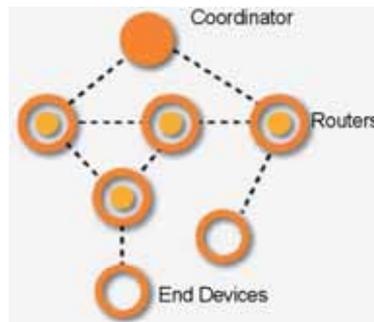
Introducción

Las redes de sensores inalámbricos, o WSNs (Wireless Sensor Networks), hacen referencia a una clase de sistemas distribuidos caracterizados por la operación autónoma y totalmente embebida de nodos de bajo consumo y reducido footprint que realizan medidas de sensores, adquisición de datos y control de un entorno objetivo, comunicándose inalámbricamente entre ellos. La adquisición y el control se llevan a cabo por medio de sensores conectados y actuadores que pueden ser administrados remotamente o bien a través de una aplicación embebida en el nodo. Los nodos que componen la red pueden ir desde un puñado a miles. Estando compuesto un sistema típico por cientos de nodos distribuidos a través de un edificio o de un espacio abierto. Muchas redes de sensores inalámbricos están basadas en estándares de redes inalámbricas propietarios, pero la tendencia reciente crece cada vez más hacia la estandarización de la comunicación inalámbrica de bajo consumo. ZigBee ofrece un estándar para la medición y control inalámbricos que está basado en unas especificaciones bien conocidas como es la 802.15.4. Si bien el documento 802.15.4 únicamente describe las capas PHY y MAC de la pila del protocolo, ZigBee se construye sobre 802.15.4 para ofrecer unas especificaciones de las capas de red y aplicación.

Figura 1. ZigBee se construye sobre 802.15.4 para ofrecer unas especificaciones de las capas de red y aplicación.



Entre los numerosos beneficios de ZigBee están el protocolo mesh para enrutado multi-salto y la entrega de datos, una especificación de seguridad y un conjunto de provisiones para asegurar la interoperabilidad a nivel de aplicación. En términos generales, ZigBee proporciona a los desarrolladores de aplicaciones embebidas una abstracción de mayor nivel para la gestión de redes y la interconexión con otros nodos.



A pesar de hacer referencia a ZigBee a lo largo de todo el artículo, muchas de las ideas y conclusiones del artículo serían igualmente válidas en el contexto de otros estándares que utilicen las capas MAC y PHY del 802.15.4. Con el fin de evitar cualquier confusión, se asumirá durante el resto del artículo que nuestro diseño objetivo implica un diseño de red multi-salto utilizando un protocolo de enrutamiento en malla (mesh), un esquema de modulación compatible con 802.15.4 y un protocolo de acceso al medio. En el artículo también se asume una familiaridad básica con las especificaciones ZigBee y 802.15.4.

Tamaño y organización de la Red

La organización y el tamaño de la Red es tal vez la elección de diseño más crítica, y que muy a menudo da forma y guía el resto del proceso de diseño. También puede tener un serio impacto pues las redes grandes son típicamente más difíciles de diseñar y mantener. Afortunadamente, existen enfoques para la implementación y mantenimiento de redes muy grandes sin por ello tener que

pagar las consecuencias de hacerlo.

Lo mejor para las redes basadas en ZigBee son redes entre 300 y 500 nodos. Esto puede no parecer mucho, pero si uno tiene en cuenta que todos estos nodos funcionan en el mismo canal físico, se envían datos unos a otros al mismo tiempo, enrutan datos unos en nombre de otros y tratan al mismo tiempo de preservar la integridad general de la red, mediante el envío periódico de mensajes de control, parecerá una red muy ruidosa y congestionada. Téngase también en cuenta que la 802.15.4, sobre la que se basa el estándar ZigBee, utiliza un protocolo CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance - Acceso Múltiple por Detección de Portadora con Evasión de Colisiones), es decir, no hay dos nodos dentro del radio de alcance uno del otro que puedan transmitir simultáneamente. Si lo hacen, ambos fracasan y ambos lo reintentarán en un momento posterior. Si la red ya está congestionada, los reintentos producen un fallo de transmisión en cascada por la que más y más nodos intentan acceder al medio, el aire, sólo para encontrarlo ocupado.

De hecho, uno de los principales retos al aumentar el tamaño de las redes más allá de unos pocos cientos de nodos, es la gestión eficaz de la congestión de la red. La otra es la optimización de los recursos del sistema utilizados para almacenar el estado de la pila interna en tiempo de ejecución. Las siguientes secciones esbozan tres estrategias alternativas para abordar el problema de la congestión.

Densidad de la Red

Debe ser inmediatamente evidente por qué "una red de 300 nodos" en realidad nos dice muy poco acerca de la organización de la red. Debido al escenario de colisiones descrito anteriormente un criterio importante de la salud de la red es también su densidad, es decir, el número de nodos que se encuentran dentro del radio de alcance unos de otros; o para decirlo en otros términos - ¿cuántos nodos puede de media un nodo escuchar? Nuestra recomendación es que este número sea menos de 5, ya que

este número permite tanto redundancia como un medio de comunicación bastante despejado. Cualquier cosa más allá de 7 es probable que cargue la red creando segmentos muy saturados.

Una pregunta pertinente es entonces ¿cómo yo, como diseñador de un sistema determino el número de nodos que están al alcance? Una estrategia clara es la de personalizar la aplicación embebida en los nodos. La información sobre los nodos vecinos es en realidad una parte fundamental del funcionamiento del protocolo en una red ZigBee. De hecho, los nodos difunden activamente, haciendo broadcast, información sobre sí mismos, y esta información es recibida por cualquier otro nodo al alcance. La tabla de vecinos puede ser consultada por la aplicación residente en el nodo y el número de entradas únicas calculado. La aplicación puede entonces disponer la transmisión de este diagnóstico a un nodo designado. Evidentemente, esto sólo tiene sentido en la instalación de la red, cuando todavía son posibles cambios en la densidad de la red. Una vez que la red está en marcha y funcionando, la información de la densidad tendrá únicamente una función consultiva en la resolución de problemas.

Tenga también en cuenta que si el tamaño de la tabla de vecinos es inferior al número de nodos que rodean a un nodo, la pila ZigBee obligará se verá obligada a retirar periódicamente entradas de la tabla. Esta retirada también puede repercutir negativamente en el rendimiento global de la red, forzando al redescubrimiento de las rutas, aunque ninguno de los nodos en el camino se haya ido offline. Así pues, además de limitar la densidad de la red para evitar la congestión, la densidad de la red también debe considerarse en el contexto de los recursos del sistema, tales como el tamaño de la tabla de vecinos.

En los casos en que la colocación física de los nodos es fija debido a los requisitos de la aplicación, la densidad de la red puede ser fácilmente controlada mediante la reducción de la potencia de salida de los transceptores en la zona congestionada. La reducción de la potencia de salida tiene el mismo efecto conceptual que alejar los nodos, haciendo menos probable que uno oiga a otro. Los fabricantes tienden a establecer la potencia de salida al máximo para garantizar el máximo alcance de operación y la mejor calidad de enlace. En nuestra experiencia, la potencia de salida puede

ser fácilmente reducida en aplicaciones para interiores, donde el rendimiento del alcance es menos crítico. Como regla empírica, la reducción de 3 dBm de potencia de salida reduce el alcance efectivo por un factor de 1,5.

Una última consideración en la densidad es el límite teórico en el que la tasa de error crece lo suficiente como para desencadenar el efecto en cascada mencionado anteriormente. Naturalmente, este parámetro depende de cuánta información es transmitida por la propia aplicación. En nuestra experiencia, el límite es de aproximadamente 25 nodos dentro del rango de alcance unos de otros si cada nodo transmite un paquete de tamaño máximo por segundo. La densidad límite parece mantenerse independientemente de la implementación de la pila, lo que sugiere que está relacionado con una operación CSMA más fundamental en el nivel MAC. Uno puede extrapolar los límites de densidad de los nodos enviando una vez cada n segundos, por ejemplo, multiplicando la densidad máxima por un factor de $1.2n$. Esta aproximación no debe usarse nunca como una guía precisa para la densidad de la red, dado que el valor real dependerá de la proporción de routers y dispositivos finales en la red.

Particionamiento/ Segmentación de la Red por Canal

Cuando ni la eliminación de nodos ni la reducción de la potencia de salida son una opción, los diseñadores de sistemas pueden elegir combatir la congestión de la red particionando una red que se encuentra funcionando en un solo canal en varias redes que operan en diferentes canales. La 802.15.4 prescribe 16 canales en el rango de frecuencia de 2,4-GHz, 10 canales más en el rango de frecuencia de 900-MHz y 1 canal en la frecuencia de 800-MHz. Poner redes en diferentes canales tiene el efecto de la disociación total de las redes unas de otras, pero también impone la complejidad adicional de tener un coordinador para cada una de las redes individuales y la conexión de estos coordinadores entre sí, si fuera necesario para el sistema.

El enfoque de la partición por canal tiene más sentido todavía cuando las redes particionadas ya se tratan como lógicamente separadas. Por ejemplo, en una aplicación de automatización de un edificio, a menudo es conveniente disponer de una red en cada uno de los pisos del edificio que opere en su propio canal. Si hay un coordinador de red previsto para cada planta, entonces no tiene mucho sentido para los nodos comunicarse de unas plantas a otras, especialmente si ello puede afectar al rendimiento de la red (como puede ser el caso con redes de alta densidad).

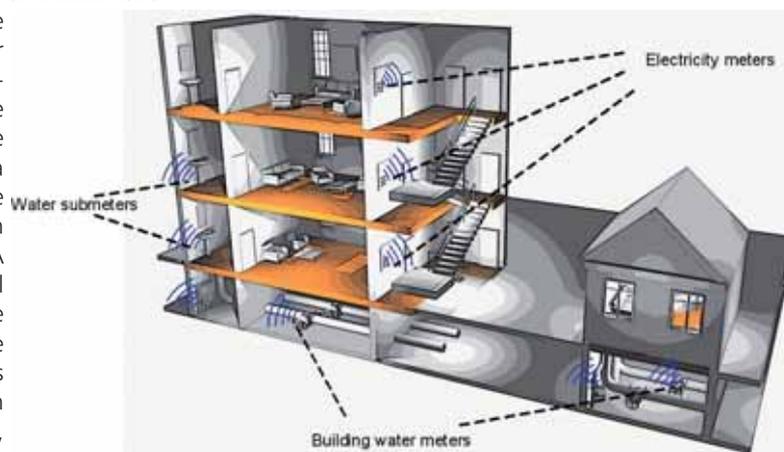


Figura 3. Medidores inalámbricos en un edificio residencial.

Del mismo modo, las particiones pueden ser por habitaciones o por áreas de oficina. En primer lugar y principalmente la decisión de particionar por un canal físico debe ser impulsada por las necesidades de la aplicación y el número de canales disponibles. Esta forma de particionamiento también puede complicar aún más el diseño general si nodos de distintas particiones deben comunicarse unos con otros.

La máscara de canal, como la potencia de salida del transceptor, es un parámetro ajustable de la red. Se puede asignar una máscara de canal adecuada al coordinador de la red y a los dispositivos de la red que debe establecer la conexión con el coordinador. Como un procedimiento estándar un coordinador realiza una exploración de detección de energía para escoger el canal de la máscara de canal con menor "actividad".

Tenga en cuenta que aparte de la asignación de una máscara con un único canal en ella, no hay ninguna forma de forzar a un coordinador a operar en un determinado canal. Es muy aconsejable, sin embargo, mantener al menos dos canales en la máscara de canal pues la interferencia in-band de redes Wi-Fi,

auriculares Bluetooth, y otros dispositivos electrónicos pueden hacer que toda la red esté inoperativa por una ocupación temporal de un canal específico, de otra forma considerado libre. Tener múltiples canales en la máscara de canal permitirá al coordinador una elección de canales y la flexibilidad de conmutar a un canal menos congestionado en el futuro.

Particionamiento/ Segmentación de la Red por PANID

Una última estrategia que es ciertamente menos eficaz que la separación física por canales es la separación por PANID. El PANID es un identificador único que es conocido por todos los dispositivos en una determinada red. Los dispositivos que utilicen distintos PANIDs no serán capaces de comunicarse unos con otros, lo que permite que múltiples redes que se sitúen en el mismo espacio sin que los datos de una red aparezcan en la otra. De hecho, varias redes con diferentes PANIDs pueden utilizar todas el mismo canal físico para comunicarse unos con otros.

A pesar de que la congestión en el aire no se verá afectada por la división de la red por PANIDs, la sobrecarga de procesamiento en cada uno de los nodos puede verse reducido por el filtrado del tráfico de los dispositivos de las redes adyacentes (con otro PANID) en las capas más bajas de la pila. En la mayoría de los casos, el filtrado realmente se hará automáticamente por el hardware PHY, liberando la aplicación de recursos y tiempo. Para aplicaciones donde el procesamiento de la aplicación y la sobrecarga de enrutamiento predominen, el particionado por PANID puede ser muy eficaz en el alojamiento de más nodos en un área dada.

Uno de los beneficios de la segmentación basada en el PANID es el número relativamente elevado de posibles segmentos únicos, 216 para ser exactos. Si bien existen sólo 26 canales físicos, los PANIDs ofrecen un enfoque mucho más escalable para la implementación de la segmentación, aunque menos eficaz. La mayoría de los sistemas de acceso inalámbrico utilizarán una combinación de la estrategia de segmentación por el PANID y por el número de canal. Al final, la combinación de estrategias ofrece la mejor combinación de escalabilidad y eficacia en el tratamiento de la congestión de la red.

Throughput

Otro aspecto de diseño que requiere una cuidadosa consideración es el rendimiento (throughput), es decir la cantidad total de datos útiles que la aplicación desea transferir por unidad de tiempo. Muchos ingenieros de sistemas suponen erróneamente que hay más ancho de banda del que está realmente disponible (a veces errando por más de un orden de magnitud), lo que lleva a la implementación de una aplicación con un bajo rendimiento o directamente no operativa y a diseños fallidos. El culpable de este habitual problema es la capacidad del medio anunciada de 250 kbps para enlaces de red 802.15.4. En la práctica a lo que se refiere el número es al límite físico teórico, es decir, el ancho de banda efectivo en el nivel PHY. Se hace caso omiso de los retrasos de los protocolos introducidos por todas las capas de la pila por encima de la física, la sobrecarga de la manipulación y el análisis de cada paquete, los tiempos de acceso al medio, las confirmaciones (acknowledgements) y los errores de bits. En nuestra experiencia, un enlace punto a punto establecido entre dos nodos ZigBee a una distancia de un salto, no puede soportar una tasa de transmisión de más de 110 a 120-kbps. Con la introducción de confirmaciones, esta tasa se reduce aproximadamente a la mitad. En un entorno típico de red donde hay de tres a cinco nodos tratando de acceder al medio compartido en un momento dado, la tasa se reduce aún más, a decenas de kbps.

Evidentemente, de 20 a 40 kbps es muy diferente de la máxima publicada de 250 kbps, y un ingeniero de sistemas sin experiencia no es consciente a menudo de la diferencia hasta que es demasiado tarde. El punto fundamental es que las redes de sensores inalámbricos nunca se pensaron para soportar aplicaciones con gran ancho de banda. En lugar de ello, el mercado objetivo son redes relativamente grandes con necesidad reducida de ancho de banda. Si uno de los sensores de su aplicación es una cámara de vídeo que genera un flujo de datos de 100 kbps, entonces una red de sensores inalámbricos y específicamente ZigBee será una mala elección. Existen tecnologías inalámbricas complementarias que harán un buen trabajo para esa función.

A pesar de que las discrepancias entre las expectativas del rendimiento y lo que la tecnología puede soportar en realidad son muy comunes, existen usos legítimos de la tecnología que requieren

de la optimización del ancho de banda para funcionar eficazmente en el entorno de destino. Consideremos, por ejemplo, un sensor de temperatura que muestrea 100 veces por segundo. Sería una mala elección de diseño enviar cada muestra en un paquete separado. De hecho, esto tendría consecuencias desastrosas para la salud de la red, teniendo en cuenta lo que hemos mencionado de la congestión de la red y el fallo de transmisión en cascada. Una estrategia mucho mejor es agregar múltiples muestras en un único paquete, ya que en la mayoría de los casos son preferibles los paquetes de tamaño máximo sobre los más pequeños. Cuando la agregación no es posible, puede utilizarse el procesamiento local para reducir los requisitos de ancho de banda. Considere ahora una aplicación típica de termostato con un umbral bajo y uno alto. No hay absolutamente ninguna necesidad de enviar las lecturas de temperatura intermedia, siempre y cuando las lecturas entren dentro del rango de temperatura aceptable. Cada nodo puede utilizar el procesamiento local a fin de determinar si las lecturas están fuera del rango prescrito, y transmitir los datos sólo cuando sea necesario para alertar a otro dispositivo remoto. Otras estrategias encaminadas a reducir la cantidad de datos que se envían pueden implicar algún tipo de resumen de datos o promediado de los mismos.

Debe tenerse en cuenta, sin embargo, que incluso las aplicaciones altamente optimizadas pueden encontrar congestiones cuando múltiples eventos físicos ocurren simultáneamente. Si la temperatura en toda una planta de un edificio sube simultáneamente, lo más probable es que muchos termostatos deseen transmitir sus lecturas de temperatura a la vez. Una práctica para luchar contra picos de ancho de banda como estos es el retraso aleatorio de las transmisiones. Obviamente, la retransmisión en la capa MAC es una forma de aleatoriedad, pero poco hace para prevenir las colisiones, sólo se preocupa una vez que ya se ha producido una colisión. Un retraso en la capa de aplicación, cuando está bien hecho, puede ser algo que realmente alivia las colisiones antes de que se produzcan. Cuando cientos de nodos necesitan transmitir un paquete una vez cada hora, el diseño óptimo tratará de repartir las transmisiones a lo largo de la hora, a fin de reducir al mínimo la posibilidad de colisiones. Es muy aconsejable que los ingenieros de sistemas empleen esta estrategia cuando las transmisiones en masa son probables. 