AN1141: Thread Mesh Network Performance

This application note details methods for testing Thread mesh network performance. With an increasing number of mesh networks available in today's wireless market, it is important for designers to understand the different use cases among these networks and their expected performances. When selecting a network or device, designers need to know the network's performance and behavior characteristics such as battery life, network throughput and latency, and the impact of network size on scalability and reliability.

This application note demonstrates how the Thread mesh network differs in performance and behavior from other mesh networks. Tests were conducted using Silicon Labs' Thread Mesh software stacks and the Wireless Gecko SoC platform capable of running Thread Mesh and Proprietary protocols. The test environment was a commercial office building with active Wi-Fi and Zigbee networks in range. Wireless test clusters were deployed in hallways, meeting rooms, offices, and open areas. The methodology for performing the benchmarking tests is defined so that others may run the same tests. These results are intended to provide guidance on design practices and principles as well as expected field performance results.

Additional performance benchmarking information for other technologies is available at http://www.silabs.com/mesh-performance.

KEY POINTS

- Wireless test network in Silicon Labs Research and Development (R&D) is described.
- Wireless conditions and environments are evaluated.
- Mesh network performance including throughput, latency, and large network scalability is presented.

The information in this document is based on the SLThread implementation of Thread. SLThread reached 'end of service' in December 2019. Silicon Labs is replacing SLThread with an implementation of the more popular OpenThread. We anticipate that the results from OpenThread will be very close to the SLThread results.

1. Introduction and Background

Silicon Labs has provided performance testing results from embedded mesh networks as part of developer conferences and industry white papers. The basic performance data of throughput, latency, and impact of security can be used by system designers to define expected behavior. Testing is done across our different mesh network technologies – Zigbee, Thread, and Bluetooth and each are presented separately. This application note presents the Thread network performance.

1.1 Underlying Physical Layer and Packet Structure

Network performance is based on payload sizing since the application usage does not account for the packet overhead.

Thread uses IEEE 802.15.4 2006 with 127-byte packets and an underlying data rate of 250 kbps. The Thread packet format is shown in the figure below and results in a 63-byte payload. For payloads above 63 bytes, the Thread stack fragments using 6lowpan. Our data on performance is based on payload size as this is the design parameter of concern when building an application.

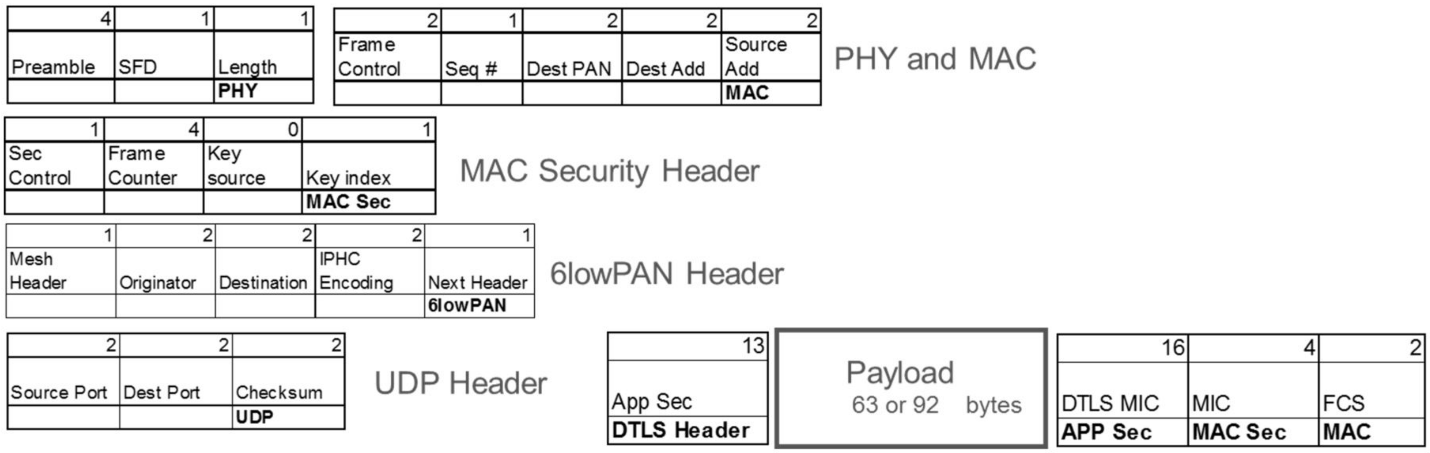


Figure 1.1. Thread Packet Format

1.2 Network Routing Differences

Thread supports both next hop routing as well as flooding. Thread networks maintain next hop routes to all routers as part of normal network maintenance instead of a device performing a flood based route discovery. Thread also minimizes the number of active routers to address scalability to large networks. Multicast flooding previously has been seen as a limitation for embedded 802.15.4 networks because the flooding in the presence of a large number of routers limited the frequency and reliability of multicast traffic. Note that the Thread network manages the number and spacing of active routers, so user intervention or management is not required.

Networks fragment larger messages into smaller ones to fit within their particular PHY limitations. For Thread, the fragmentation is done at the 6lowpan layer and is done source to destination (not at the individual hops).

For unicast forwarding within these networks, the message is forwarded as soon as the device is ready to send. For multicast forwarding, there are generally networking requirements for how messages are forwarded. For Thread devices, RFC 7731 MPL forwarding is used. For this method, the trickle timer is set to 64 milliseconds, so the devices back off a random amount up to this time before retransmitting.

Note: This performance data is for the Silicon Labs implementation of these mesh networking stacks. As is shown in the test network and infrastructure provided for this testing, no tests were performed with other stacks or systems.

2. Goals and Methods

This application note defines a set of tests performed to evaluate mesh network performance, scalability, and reliability. The test conditions and infrastructure are described, in addition to the message latency and reliability. This testing is conducted over actual wireless devices in test networks and not in simulation.

This testing is done to provide a relative comparison between the different mesh technologies to better understand and recommend their usage. Different network and system designs have different requirements for devices and networks. As such, no one network fulfills all possible network requirements. However, the three mesh networking technologies we are comparing are all aimed at low power and battery-operated mesh networks used for control and monitoring in the home and commercial buildings.

Normally, when analyzing data on network performance, we then consider what improvements can be made in the network to improve performance. Because of the limited data available publicly on mesh network performance in large networks today, it is difficult to have industry discussions on possible improvements or changes. For example, in commercial buildings there is concern over:

- Other network traffic, since there may be many subnets that interfere with each other.
- Wi-Fi interference from the normal building Wi-Fi infrastructure as these technologies are generally operating in the 2.4 GHz ISM band.
- Network throughput and latency as well as large network multicast latency and reliability, since multicasts are commonly used for lighting controls in dense office environments and users of the system expect responsiveness in lighting controls.

Note: The test results here are limited to comparisons of system performance under normal operating conditions, or under stress as noted in particular tests. This application note does not specifically address system interference or other such effects that have been addressed in other published results. However, testing is done in our Silicon Labs R&D facility where there are more than 100 Wi-Fi access points within RF range. The facility also has a 300-node Zigbee lighting network that is not part of this testing but is used for normal lighting control.

2.1 Review of Other Performance Testing

There are no specific, defined methods for evaluating and reporting large network reliability, scalability, or latency. Silicon Labs has published papers in the past comparing network performances based on network testing. This testing focused on device behavior and impact on battery life, and network throughput and latency. Large scale multicast testing also requires capturing accurate timing and reliability information from large distributed networks. All testing was conducted using Silicon Labs' Wireless Gecko SoC platform capable of running Zigbee, Thread, Bluetooth Mesh, and Proprietary RF protocols to eliminate the device itself as a difference in the testing. Previously published results showed differences between transceivers, network co-processors, and System-on-Chip designs. These tests all use System-on-Chip designs.

Other papers on performance include a master's thesis paper on IP network performance, "Experimental Study of Thread Mesh Network for Wireless Building Automation Systems", by Dapeng Lan, which was published by the Department of Software and Computer Systems at the Royal Institute of Technology in Sweden. This paper tested unicast and multicast performance per the following table:

Test	Packet	Added Load	Results
Unicast burst off load 0	Border router sends CoAP request 500 milliseconds after previous test. CoAP get with 10 bytes of payload and CoAP ack with 20 bytes payload.	No	Average round trip 1 hop 28.98 milliseconds 2 hop 49.42 3 hop 74.64 4 hop 94.22 5 hop 120.88 6 hop 150.37
Unicast burst load 0.2	same	Every device sends CoAP request to random device every 5 seconds	Average round trip 1 hop 29.06 milliseconds 2 hop 54.87 3 hop 88.25 4 hop 95.69 5 hop 117.78 6 hop 139.27
Unicast burst on load 0	Border router sends CoAP request immediately after previous test	No	Average round trip 1 hop 34.29 milliseconds 2 hop 62.38 3 hop 81.61 4 hop 112.99 5 hop 127.87 6 hop 142.65
Unicast burst load 0.2	same	Every device sends CoAP request to random device every 5 seconds	Average round trip 1 hop 32.89 milliseconds 2 hop 58.75 3 hop 88.50 4 hop 108.00 5 hop 119.52 6 hop 139.81

Test	Packet	Added Load	Results
Multicast reliability	Send multicast every 2 seconds from border router and check how many are received by nodes in the network. CoAP get with 10 bytes of payload.		Sent 1358 packets 1341 received 98.75% reliable
Multicast latency	Send multicast every 2 seconds from border router and count the time for complete coverage of network		Mean time to complete 92.62 milliseconds Max time 184 milliseconds

Unicast testing was done with a 23 node network, one device acting as a border router. The multicast testing was done on a 38 node network with one device acting as a border router. For Thread, 18 were routers and the remainder were end devices.

This data is provided because any such network testing is of interest to allow comparison. Note that different implementations of hardware and software will have different performances; there are no industry standards for this type of performance.

3. Test Network and Conditions

To minimize variability, device testing can also be performed in fixed topologies where the RF paths are wired together through splitters and attenuators to ensure the topology does not change over time and testing. This method is used for 7-hop testing to ensure network topology. MAC filtering can also be used to achieve the network topology.

A typical wired test configuration is shown below:

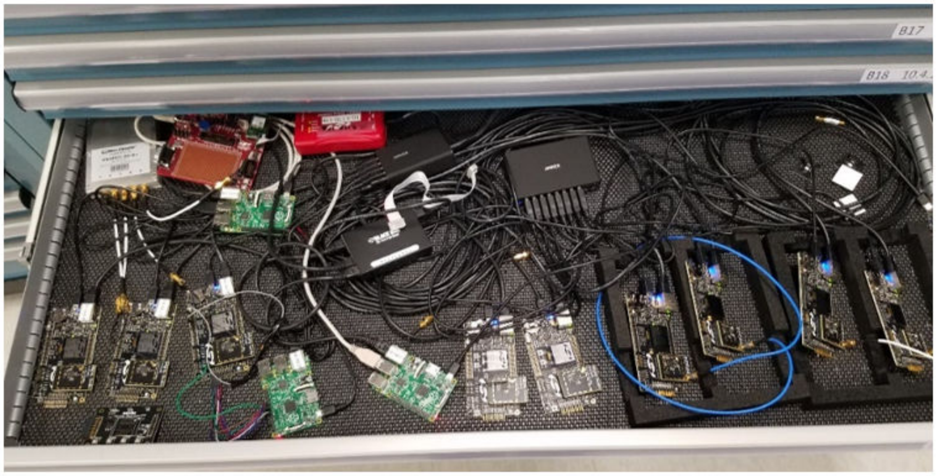


Figure 3.1. Wired RF devices in drawer with splitter and coax cable connectivity

Large network testing is best conducted in an open-air environment where device behavior is based on the existing and varying RF conditions. The Silicon Labs R&D facility is used for this open-air testing.

3.1 Facility and Test Network Conditions

The Silicon Labs R&D facility consists of a central core with an elevator shaft, other services with an open floor plan on the west end of the building, and offices and conference rooms on the east end. The overall facility measures approximately 120 feet by 200 feet. The image below shows the facility layout. The darker lines represent hard walls and everything else is split up with cube partitions.

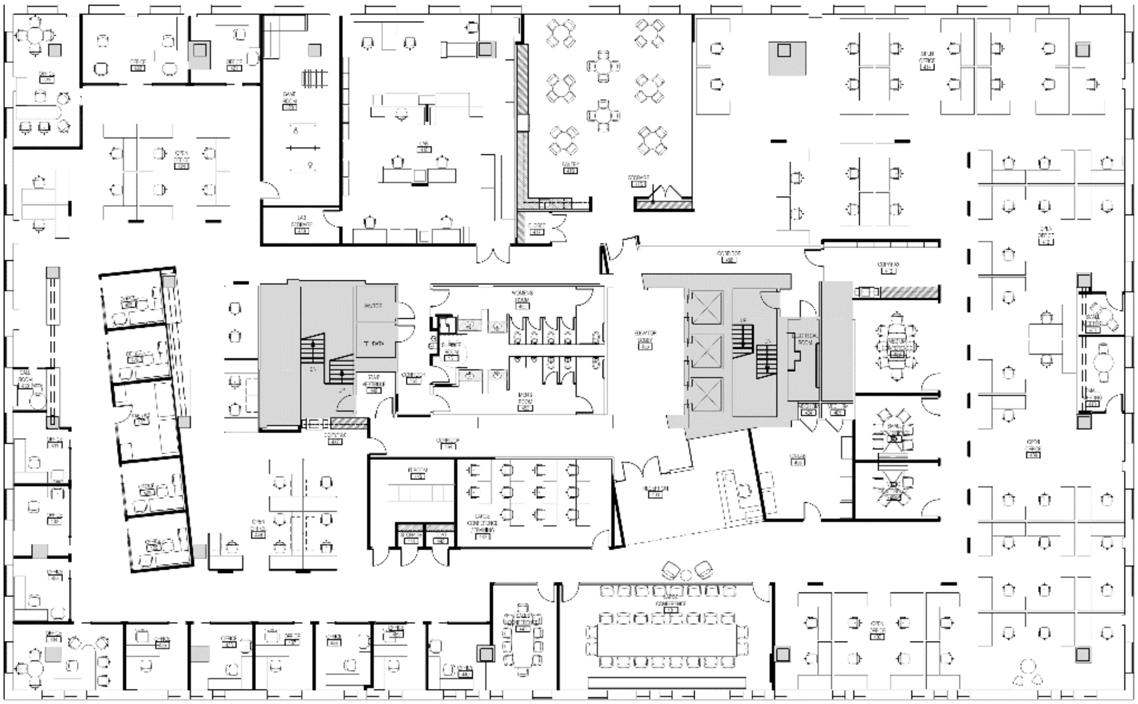
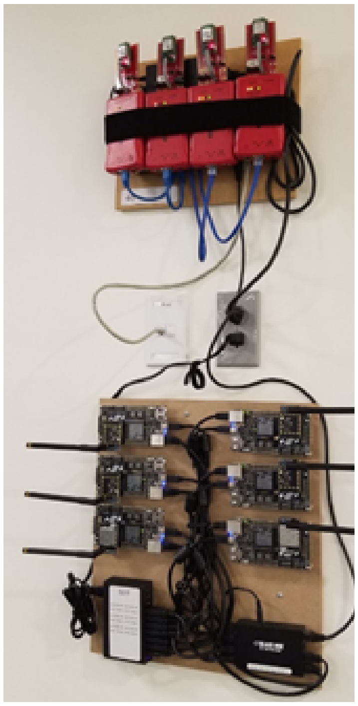


Figure 3.2. Silicon Labs facility layout used for wireless testing

The testing devices are installed at various locations around the facility. These devices all have Ethernet backchannel connectivity to allow:

- Firmware updates
- Command line interface
- Scripting
- Timing analysis
- Packet capture
- Energy measurements



Four EM35x Devices using PoE

Six EFR32MG (Mighty Gecko) Devices

Multi-band support to allow testing both
2.4 GHz (PCB antenna) and proprietary sub-GHz
protocols (external antenna)

USB power and Ethernet connectivity

Figure 3.3. Typical Testing Cluster

The testing clusters are spread throughout the facility in both high and low locations, open areas, and enclosed meeting rooms and offices.

Figure 3.4. Testing Clusters in the Silicon Labs R&D Facility

This test network has devices added or removed from it on a regular basis, but at the time of this testing it consisted of the following devices:

- EM35xx devices
- EFR32 TM Mighty Gecko devices

This network represented devices that were used for open-air testing by the networking and software quality assurance teams. All devices are controlled from a central test server and infrastructure, which allows scripted regression testing or manual testing by engineers.

3.2 Wireless Conditions in the Facility

The Silicon Labs R&D facility has a full Zigbee lighting control system including motion and lighting sensors and switches. This is not part of the test network and is used as a normal building control system independent of any testing being run.

The Silicon Labs R&D facility is also downtown and, in addition to our existing Wi-Fi infrastructure, there are over 100 Wi-Fi access points within RF distance of the facility. The following charts were taken as a snapshot of a normal work day Wi-Fi scan. This is considered the normal Wi-Fi background traffic.

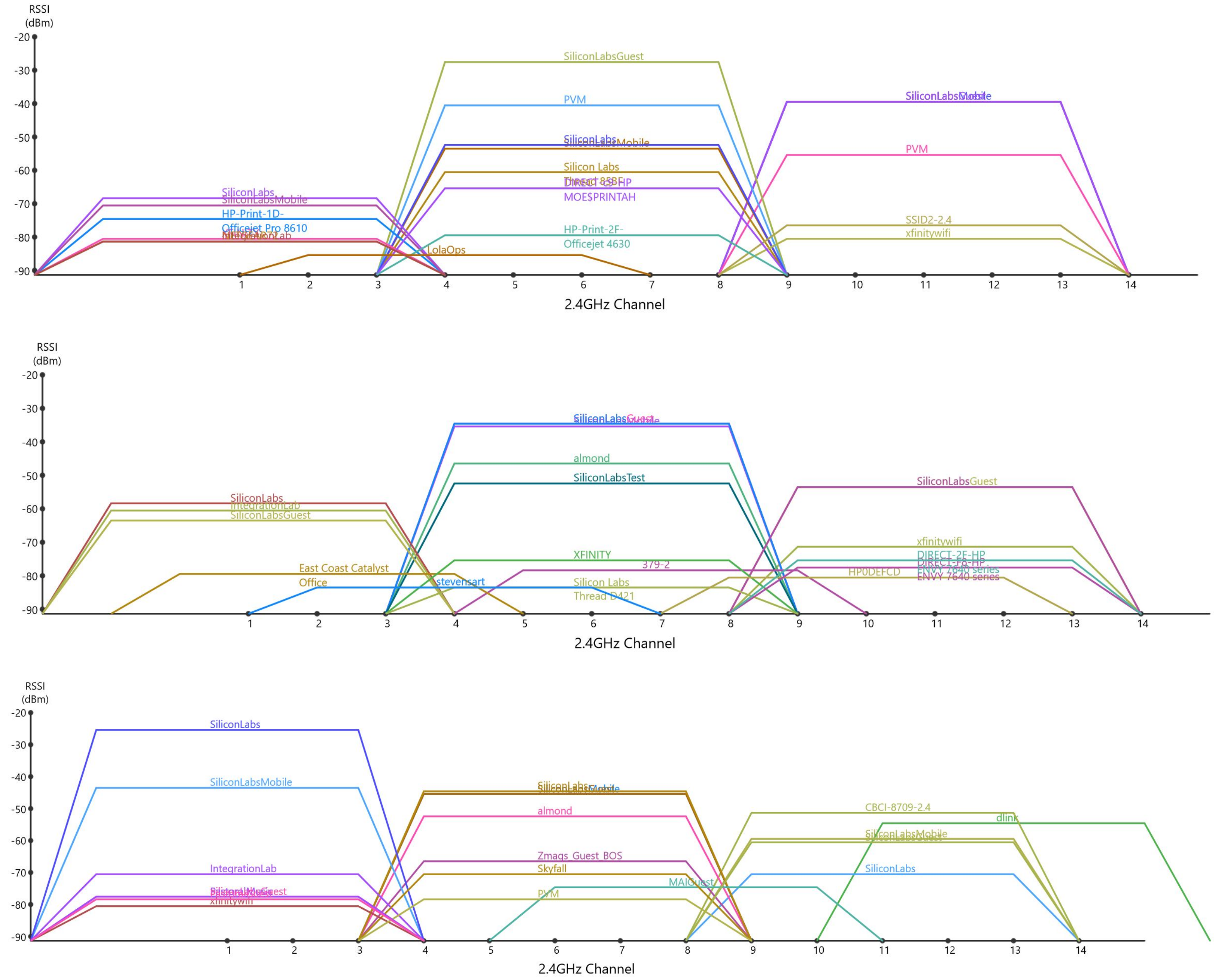


Figure 3.5. Wi-Fi Scans on a Normal Day

These Wi-Fi scans were taken at the southeast corner office, the west side, and the north side in the main conference room, respectively. These locations showed 62, 104, and 83 Wi-Fi access points within RF range.

3.3 Typical Test Network

Within the test network, a given test can be selected and used for a given set of devices. The network is established and devices joined using the Ethernet backchannel to send commands to devices. A typical network during testing is shown below. The black and grey lines show the router connectivity and strength.

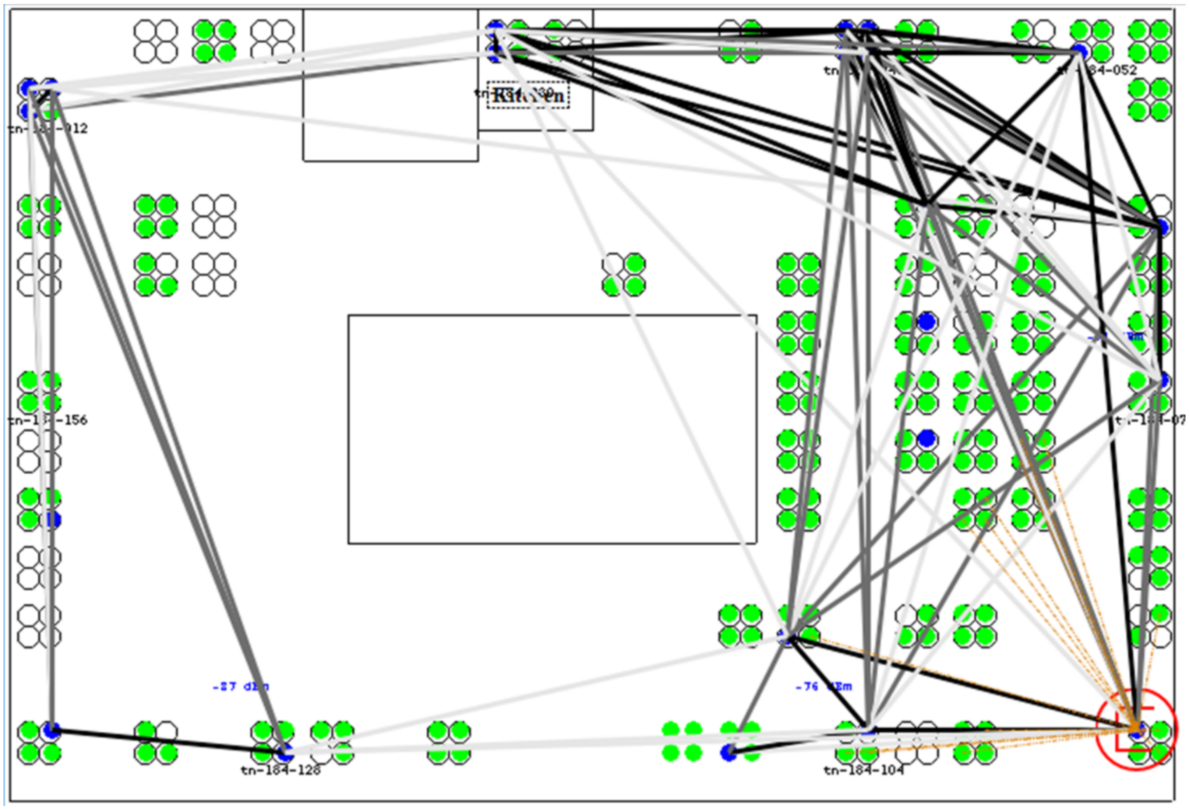


Figure 3.6. A Typical Network During Testing

4. Testing and Results

4.1 Throughput and Latency

The throughput and latency is tested in a controlled network (wired configuration) to test each hop against different packet payloads.

The normal configuration is to test to 7 hops. Testing is done with one source node and a series of destination nodes to allow the number of hops to be varied.

This testing is done using the following configuration:

1. CoAP confirmable type messaging with an Acknowledgement
2. Packet payload from 10 bytes to up to 300 bytes in 10-byte increments for the latency test
3. From 1 to 7 hops using the leader as the source
4. Using 1 packet in flight
5. Sending as fast as possible given ack timing
6. Measuring round trip latency (source to destination to source) in milliseconds

For each of these different mesh networks, packet fragmentation occurs differently as we increase payload size as noted above. The use of larger packet sizes is up to the application layer, but we provide comparison data here to indicate relative performance as fragmentation occurs.

4.1.1 Thread Multi-hop latency

Thread timing has been measured as round trip time for a CoAP message of a given payload size.

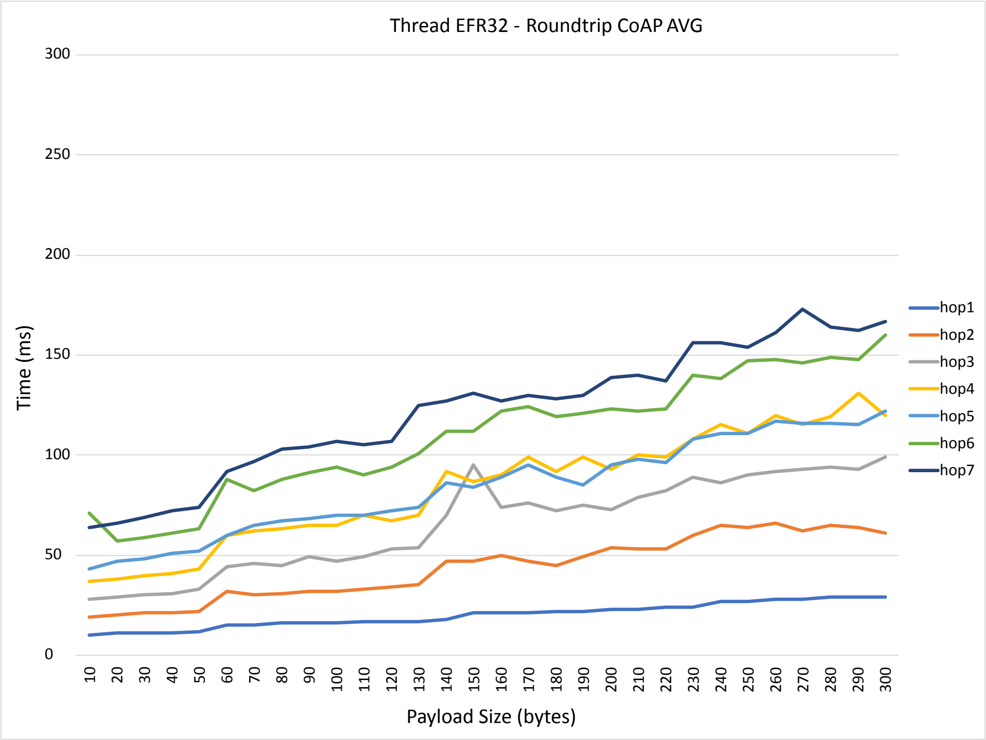


Figure 4.1. Thread EFR32 – Roundtrip CoAp AVG

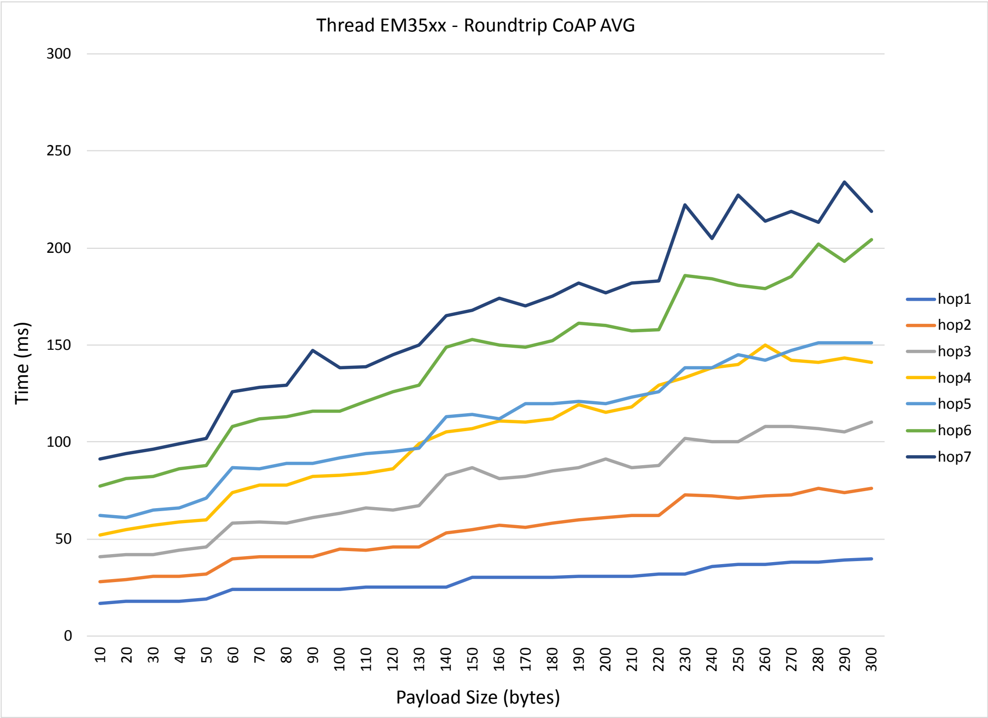


Figure 4.2. Thread EM35xx – Roundtrip CoAp AVG

Several items are worth noting in this multi-hop latency test.

- We do see better performance on the EFR32 platform than the EM35x platform.
- For a 1 hop payload out to 300 bytes, Thread maintains under a 50 millisecond round trip latency, which is excellent.

- Even out to 7 hops with 300 bytes of payload, the round trip latency is under 200 milliseconds for EFR32 and slightly over for EM35x.
- For most applications where the payload is within one packet (50-60 bytes), Thread will maintain less than a 100 millisecond round trip latency out to 7 hops.

4.2 Network Tests versus Network Size

Open air large network testing is required to validate stack performance under less controlled conditions. These networks are configured within normal Silicon Labs office space with normal Wi-Fi interference, other network operations, and building control systems. No attempt is made to isolate these network RF conditions.

The networks to be tested for each stack include:

- Small network: 24 devices
- Medium network: 1 – 48 devices
- Medium network: 2 – 96 devices
- Large network: 1 – 144 devices
- Large network: 2 – 192 devices

Note: For any of these tests, the specific number of devices is acceptable within +/- 10% of these test network targets for a given set of testing. Testing in this large network is done in SoC mode for devices.

These networks are all configured as powered devices unless there is specific testing for sleeping end.

For each of these networks the testing will validate reliability and latency for a set of traffic conditions. Testing is intended to be done over 100 messages but longer runs with 10000 messages for reliability are also done. The same devices were used across the tests to keep the topology and density of the different test runs similar. The actual over the air conditions will vary and that cannot be controlled in these tests.

4.2.1 Thread Large Network Testing Results

Thread testing has been done using the latest version of the Silicon Labs Thread stack.

The graphs below represent the Thread network multicast behavior. The Thread network limits the number of routers to 32 or less, and those devices that are active routers can change over time as the network grows or conditions change. Those devices that are not active routers are referred to as Router Eligible End Devices (REEDs) and they act as Always On children. The initial device sends the broadcast, which is heard by all routers within RF range, as well as any REED devices having that initial device as a parent. REEDs are required by the Thread specification to synchronize with a single primary parent, but also synchronize with at least three additional parents to improve multicast reliability.

Thread devices use a multicast backoff of 32-64 milliseconds before a device relays the multicast.

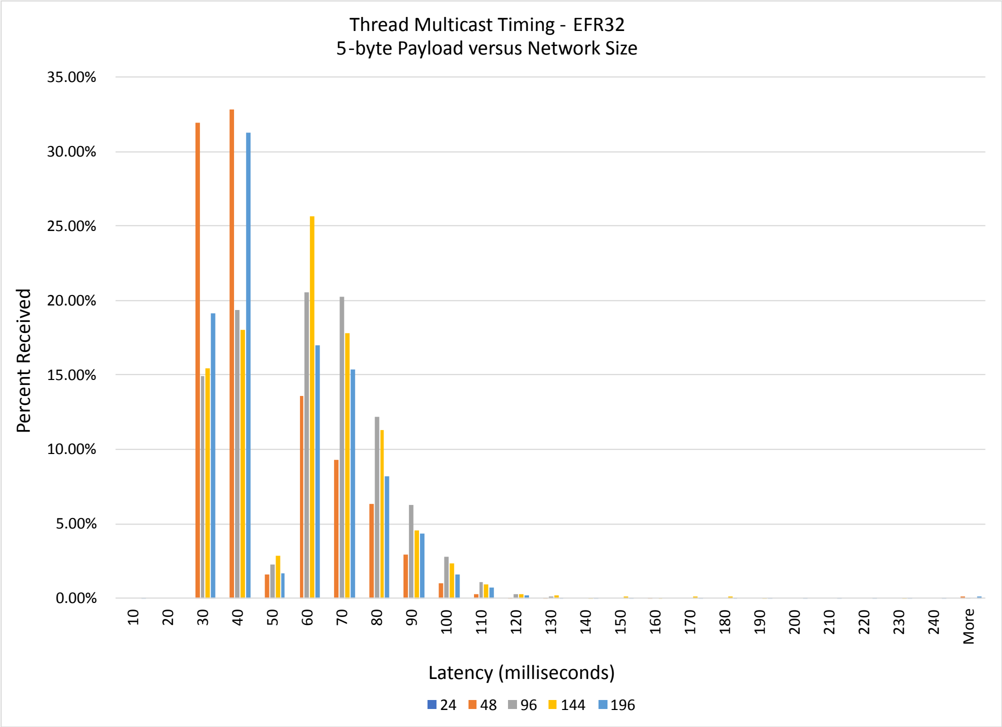


Figure 4.3. Thread Multicast Timing – EFR32 5-byte Payload vs Network Size

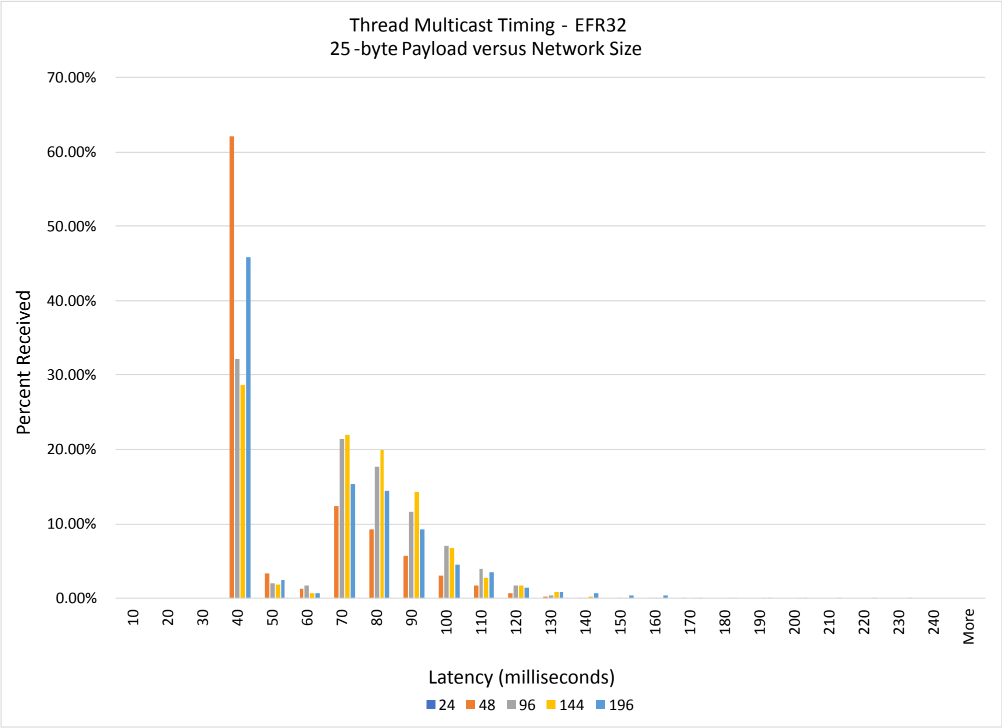


Figure 4.4. Thread Multicast Timing – EFR32 25-byte Payload vs Network Size

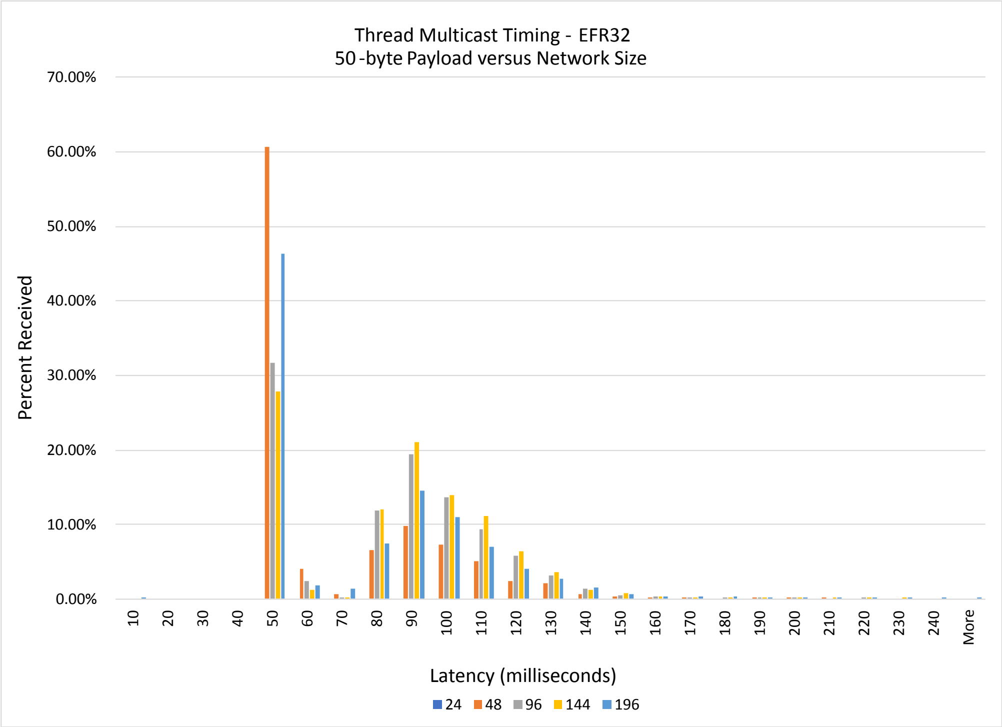


Figure 4.5. Thread Multicast Timing – EFR32 50-byte Payload vs Network Size

Of note in these results:

- We see the Thread network behavior is very consistent across network sizes, and the latency increases as the packet size increases.
- Thread latency does not spread out as much as network size increases. This is expected as there are fewer routers, so less congestion in the network.
- Thread is generally a little bit quicker than Zigbee on multicast performance, particularly as the network size increases.

- These tests all showed 100% reliability, but the larger scale testing below shows the network reliability.
- We tested at 3-second intervals to be consistent with the Zigbee results above. Shorter intervals on Thread do not show different performance so we ran the extended testing for reliability discussed below with 0.5 second intervals between broadcasts to show those results.

4.2.2 Large Network Extended Testing

Testing with the 192 node network was extended to run over 10000 multicasts to evaluate performance over a longer time period. These tests are also used to show reliability as some failures can be expected in longer tests.

The Thread duration testing was done with 0.5 second intervals between broadcasts because the network is more tolerant of shorter broadcast intervals. The histogram of these results is below.

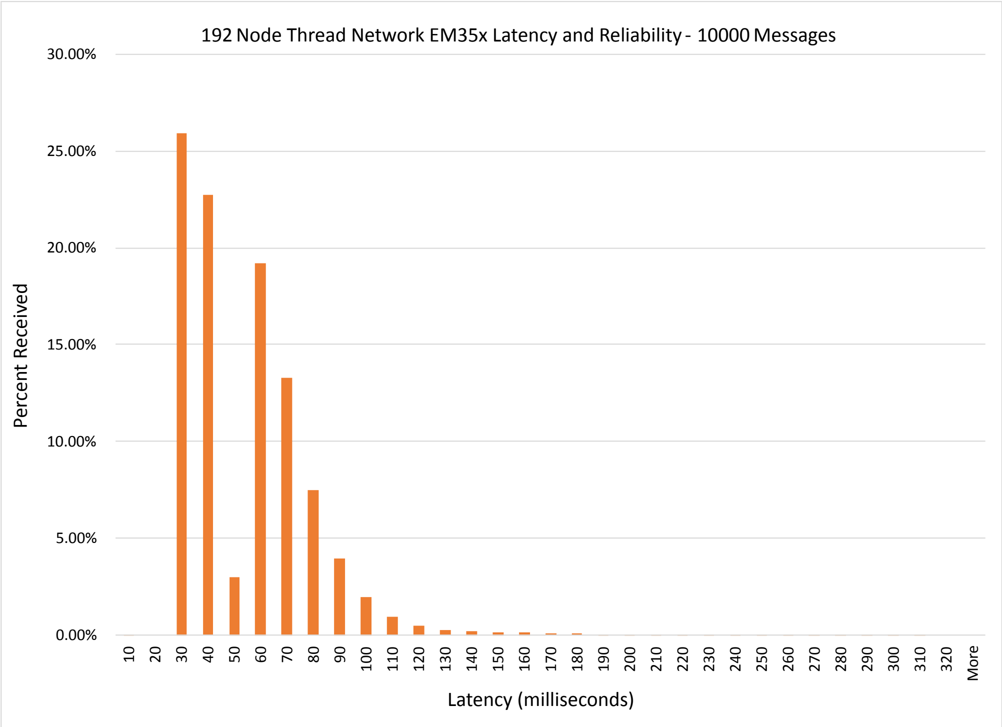


Figure 4.6. 192 Node Thread Network EM35x Latency and Reliability – 10000 Messages

The last packet received in this testing was at 303 milliseconds, which means that Thread does not spread the latency in larger networks. There is little difference in Thread network performance as the network scales to larger sizes. In this reliability testing, there were six lost messages resulting in a 99.9997% delivery. Using 200 milliseconds as an important latency for operation, 99.96% of the Thread packets were received within this time.

5. Summary

Thread shows excellent reliability and latency below the 200-millisecond timing typically required for human interaction with devices. Even in multicast, large network conditions, every 0.5 seconds the Thread network handles the traffic and maintains latency and reliability. Thread shows little change in network behavior as the network scales up in size due to the flexible router configuration done at the network layer.

Thread does show better multi-hop latency performance running on the EFR32 than the EM35x platform. This is somewhat expected since the device is a newer architecture, runs at a higher clock speed, and has more RAM for packet handling.

As packet payload increases, the latency across the network also increases, but this is a smaller impact when testing 5, 25, and 50 bytes of payload.

5.1 Follow-up Testing Considerations

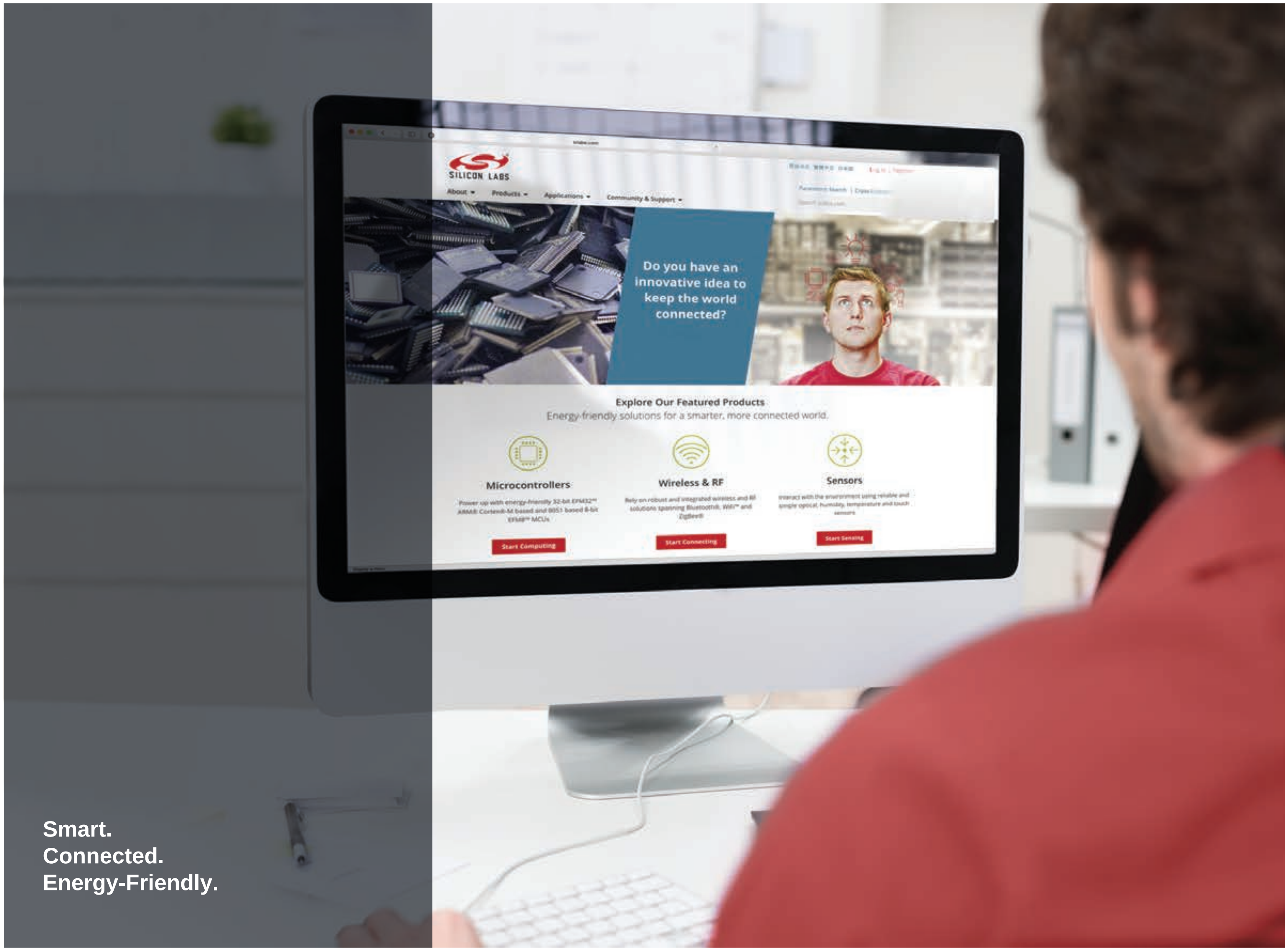
The testing described in this application note requires follow-up tests to further define the device behavior and network operations. The following specific items are noted for follow-up testing:

1. Failure testing can also be added by dropping nodes out of this network during these tests to evaluate recovery time and impact on reliability.
2. Testing should be performed with different device types running in System-on-Chip and Network Co-Processor (NCP) modes. Previous testing has revealed some differences between these modes of operation, so this should be further characterized.

5.2 Related Literature

This application note has provided information on Thread mesh networking. For information on Bluetooth and Zigbee mesh networking, and a comparison of all three technologies, refer to the following application notes:

- AN1137: Bluetooth Network Performance
- AN1138: Zigbee Mesh Network Performance
- AN1142: Mesh Network Performance Comparison



Products
www.silabs.com/products

Quality
www.silabs.com/quality

Support and Community
community.silabs.com

SILICON LABS

Silicon Laboratories Inc.
400 West Cesar Chavez
Austin, TX 78701
USA

http://www.silabs.com