
Introduction to the
Z-Wave Security
Ecosystem
Author: ABR

TABLE OF CONTENTS

1 ABBREVIATIONS 3
2 INTRODUCTION 4
3 SECURITY CHALLENGES IN CONNECTED HOME CONTROL 4
4 Z-Wave S2 SECURITY – UNDER THE HOOD 7
4.1 Security classes and network keys 7
4.2 Device Authentication 7
4.3 Key Exchange 8
4.4 Key integrity 9
4.5 Deterring lost devices 9
4.6 Secure multicast with no addressing overhead 9
4.7 Application status polling not needed 10
4.8 Expedited delivery 11
4.9 Energy efficiency 12
4.10 Active protection against attacks via Nonces 13
4.11 Secure gateway communication 13
REFERENCES 14

© Sigma Designs, Inc. 2016

SIGMA DESIGNS

1 ABBREVIATIONS

ABBREVIATION	EXPLANATION
DH	Diffie–Hellman. A method for encryption key exchange
DoS	Denial of Service. A popular Internet attack method
DSK	Device-Specific Key
DTLS	Datagram Transport Layer Security
ECDH	Elliptic Curve Diffie–Hellman. Advanced variant of Diffie-Hellman
IoT	Internet of Things
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
LAN	Local Area Network. Typically used to represent a home or office network
MAC	Media Access Control. Basic transmission control layer according to the OSI reference. This layer defines the Singlecast, Broadcast and Multicast frame formats
MPAN	Multicast Pre-Agreed Nonce. A Nonce used by members of a multicast group
NodeID	Node Identity. Unique identifier for a node in a Z-Wave network
Nonce	Number used Once
OOB	Out-of-band. Alternative to transmission over a network, e.g. keypad entry
OTA	Over The Air. Method for firmware update in contrast to "over the wire"
PAN	Personal Area Network. Short range and low power radio technologies are generally described as PANs. Therefore Z-Wave is a categorized as a PAN even though the outdoor range is up to 100 meters. Up to four repeaters may further extend the range.
PSK	Pre-Shared Key. OOB authentication method used to obtain trust between devices
QR	Quick Response code. Binary readable code often used for linking web pages
S0	Z-Wave Security 0. Relatively secure, yet as easy as non-secure Z-Wave
S2	Z-Wave Security 2. Strong encryption and authentication
SPAN	Singlecast Pre-Agreed Nonce. A Nonce for singlecast between two Z-Wave nodes
TLS	Transport Layer Security
UDP	User Datagram Protocol. A connectionless transport mechanism in IP networks
UI	User Interface
VPN	Virtual Private Network. technology for secure interconnection of LANs
WAN	Wide Area Network. Typically used to represent the Internet
Z/IP	Z-Wave for IP
Z/IP Gateway	IP-based gateway that presents Z-Wave nodes as IP hosts in the LAN

© Sigma Designs, Inc. 2016

2 INTRODUCTION

This whitepaper provides an overview of the Z-Wave Security ecosystem.

Whether one is providing remote access to, or securing communication between, Z-Wave (PAN) nodes from the Internet (WAN) via a home network (LAN), there are a number of challenges to consider. These include security attack threats, available cryptographic computation power, available network bandwidth, available code space, firewall policies, efficient battery operation and more.

S2 Security introduces best-in-class security in the PAN while maintaining the user friendliness and power efficiency, that Z-Wave is so well-known for. Consumer product manufacturers will appreciate that the S2 Security solution only requires a small code footprint in embedded devices while installers benefit from the simple installation procedure. S2 complements similar optimized mechanisms for IP domains that allow Z-Wave services to operate securely in an end-to-end fashion.

S2 Security may be considered as the first true smart home security solution. It enables secure communication for sensor devices that run for years on a single battery. At the same time it enables secure multicast addressing of lights, window coverings and similar devices.

Z-Wave nodes are added to the Z-Wave network (PAN) with Out-of-Band (OOB) authentication to ensure that they can be trusted.
A strong temporary key is used to assign keys for one or more security classes. This allows for segmentation of safety critical devices in the "S2 Access Control" class and sensors in the "S2 Authenticated" class, while the most constrained devices without authentication support are only allowed access to the "S2 Unauthenticated" class.
The Z/IP Gateway controls access to the Z-Wave network by only forwarding commands from trusted LAN clients or from a trusted Internet host such as a service provider portal. It must be expected that a typical home network (LAN) is compromised by malware or bots. DTLS is used to secure communication between LAN hosts and Z-Wave nodes. LAN hosts and Z-Wave nodes communicate via a Z/IP Gateway which terminates the DTLS encryption and strips Z/IP and IP headers before forwarding Z-Wave commands securely in the Z-Wave network.

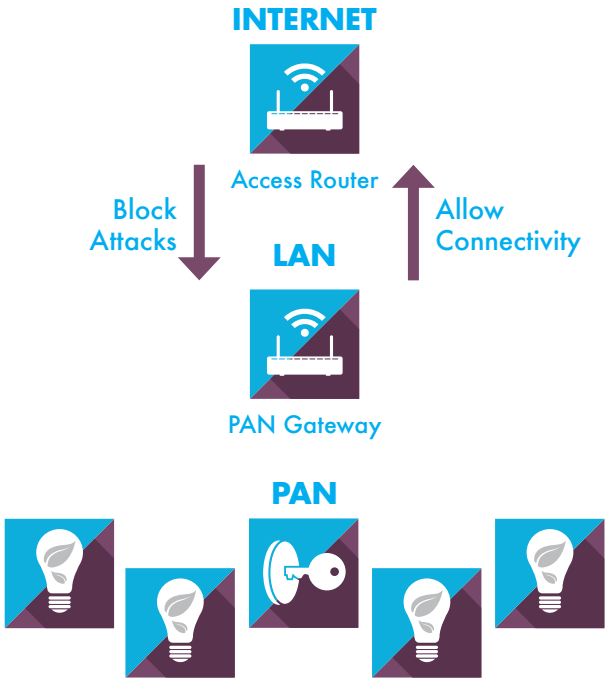
Users do not want to mess with firewalls. Therefore, the Z/IP Gateway punches a hole from the LAN side of the firewall by creating a secure TLS-based tunnel to a service provider portal in the Internet (WAN). The portal only accepts trusted gateways and the gateway only accepts trusted internet hosts.

3 SECURITY CHALLENGES IN CONNECTED HOME CONTROL

Z-Wave services span three networking domains (PAN, LAN and WAN). Starting from the WAN side, the Internet has evolved to be a very hostile place. An unprotected host will be attacked only minutes after connecting to the Internet. Therefore, the first line of defense for the home is an updated firewall that blocks all connection attempts made from the outside. In many cases, the firewall resides in the Internet Service Provider's access router. In theory, the router could be configured to accept incoming communication from trusted home control clients. However, the home control service is often provided by another service provider which has no control over the Internet Service Provider's equipment. Just to make things more complicated, additional firewalls and NAT translators may follow after the access router.

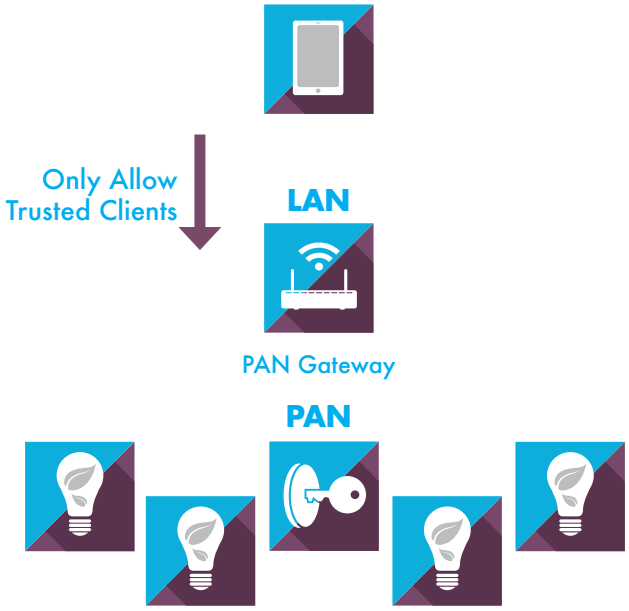
© Sigma Designs, Inc. 2016

It is generally not recommended to open ports in firewalls and due to the above issues it might not provide the intended connectivity to a home control gateway anyway. The solution is to have the home control gateway to establish a secure connection via its LAN interface to an already trusted portal server in the Internet that is managed by the home control service provider.
Having a "default deny" policy for incoming traffic protects not only the home control gateway but the entire LAN.

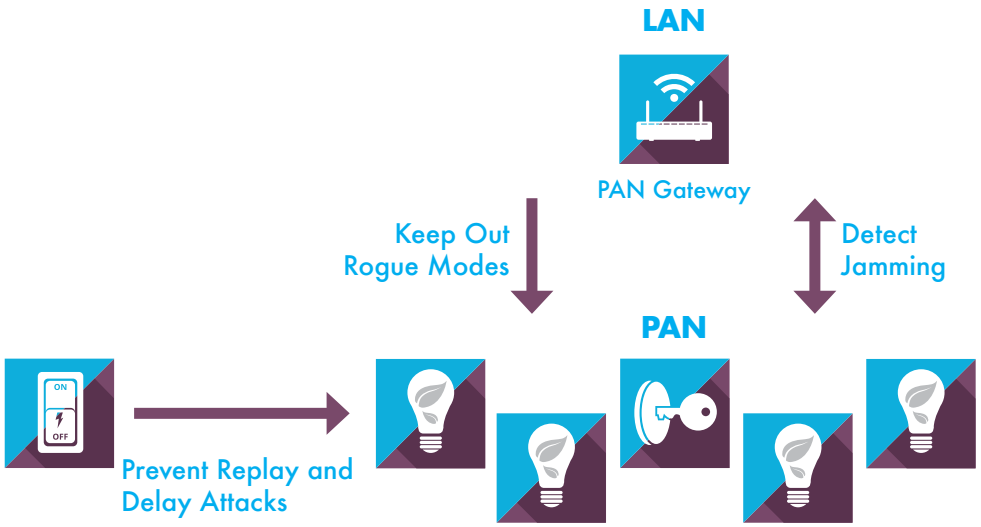


It is attractive to connect directly to the home control gateway via the LAN when the user is at home as the latency is lower and the system works even when the Internet connection is down.
Just as the WAN, the LAN has also evolved to be a very hostile place in most homes. Kids click on links on compromised web pages or download games with viruses, trojans and other malware. Adults are tricked to open attachments in apparently trustworthy phishing mails.
Thus, LAN hosts cannot rely on the LAN being safe today.
In home control terms, this means that LAN clients such as tablets and smart phones must use secure and trusted communication when accessing home control resources via the home control gateway. It is also required that the administration web interface of the home control gateway complies with best design practices for non-default user passwords and passwords as well as rejection of brute-force login attempts via exponential delay ramp-up in case of wrong credentials. Additional guidelines may be found in [HC SECURITY].

© Sigma Designs, Inc. 2016



PAN networks like Z-Wave may also be attacked. The simplest attack is radio jamming. Wi-Fi networks can also be jammed. Jamming may be used to obstruct PAN-based alarm systems. It is therefore critical that the Z-Wave gateway monitors periodical heartbeat signals from all nodes in the Z-Wave network so that jamming can be detected within minutes after the attack is initiated. Another threat is the unintended inclusion of a rogue node. Neighbors may inadvertently add nodes to each other's (wrong) networks or, more critically, an attacker may succeed in including a debugging node and subsequently extract the network keys from that node. Out-of-Band (OOB) authentication must be required for joining nodes in order to prevent the inclusion of rogue nodes.
An attacker may use a wireless sniffer to capture transmissions over the air and replay such transmissions at later time. Encrypted commands may also be replayed. This type of attack is prevented by making each transmission unique and ignoring earlier transmissions in the receiver.



© Sigma Designs, Inc. 2016

4 Z-WAVE S2 SECURITY - UNDER THE HOOD

Many Z-Wave building blocks serve multiple purposes. For instance, an acknowledgement mechanism may at the same time ensure secure delivery of a command, maintain secure multicast state information, assist in saving power and minimize the use of the radio spectrum.

This chapter provides a detailed walk-through of the mechanisms provided by Z-Wave S2 nodes.

4.1 Security classes and network keys

Like WiFi, S2 security [S2] also operates with the concept of a network key. All nodes may use this key to communicate to each other. S2, however, divides the logical Z-Wave network into three dedicated security classes, which each have a unique network key. A given S2 security class not only identifies the network key to use but also dictates the rules applying to authentication of a new node during inclusion. The "S2 Access Control" class is the most trusted class, intended for access control devices like door locks and garage doors. The "S2 Authenticated" class is used for all normal household devices such as sensors and light dimmers. The "S2 Unauthenticated" class is the least trusted class and is only intended for the most constrained controllers that, due to a limited user interface, are not capable of authenticating a node joining the network. An example is a key fob used to control a few lamps in a country cabin. Finally, a dedicated S0 key may be handed out to S2 enabled nodes for interoperability with S0-only nodes. S2 nodes transfer the S0 key just as the "S2 Unauthenticated" class key; using the temporary ECDH key for the key exchange, rather than using the S0 key exchange which has a known vulnerability.

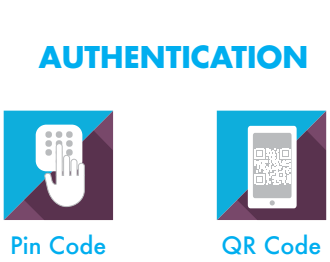
S0 and all three S2 classes use AES-128 [AES] encryption. The US Government considers AES-128 safe enough for classified information up to the SECRET level [NSA_AES]. Combined with S2 authentication and Nonce scrambling, there is no known way to break this protection – even with the aid of a super computer.
A node may request and be granted access to several security classes during inclusion but it only accepts incoming commands in the most trusted of the granted classes. This means that a light bulb being a member of both the "S2 Authenticated" and "S2 Unauthenticated" classes only accepts commands encrypted with the "S2 Authenticated" class key.

A controlling node that is member of all security classes may control a door lock via the "S2 Access Control" class key and a lamp via the "S2 Authenticated" class key.
A node may be granted a subset of the requested classes. For instance, a light bulb may request membership of both the "S2 Authenticated", "S2 Unauthenticated" and "S0" classes but be granted only the "S0" class by a constrained S2 controller.

4.2 Device Authentication

In a wireless environment, there is a real risk that a foreign node is included accidentally or due to evil intent. The S2 authentication process allows an including controller to verify that a joining node is indeed the physical device that it claims to be. Depending on the UI, an including controller may allow the installer to enter a Device-Specific Key (DSK) string of decimal digits that can be read visually or scanned as a QR code. The DSK is the first 16 bytes of the 32-byte long ECDH public key of the joining node.

© Sigma Designs, Inc. 2016



If a joining node is granted membership of the "S2 Access Control" or "S2 Authenticated" class by an including controller, the joining node advertises a garbled ECDH public key where the first 16 bits have been set to zero. This may sound like a vulnerability. An including controller could quickly scan through the 65536 possibilities to reverse engineer the garbled bits of the ECDH public key and then include the joining node without waiting for the installer. However, the ECDH public key is not a secret. The purpose is solely to force the installer to confirm that this is indeed the node that should be included.

If a joining node is only granted membership of the "S2 Unauthenticated" class, the node advertises the complete ECDH public key so that the authentication step can be skipped.

4.3 Key Exchange

Distributing a network key is a classic chicken-and-egg problem. Transferring a key to a new node over an insecure media requires a secure channel, but one cannot secure the channel without proper keys. Many Z-Wave devices do not have an interface that allows one to enter a key locally and it would be cumbersome if they had. Diffie-Hellman key exchange solves this problem.

A popularly illustration of Diffie-Hellman key exchange [DIFFIE-HELLMAN] is the mixing of colors and finding it hard to separate them again. This is known as a one-way function. A similar mathematical one-way function is raising large prime numbers to the power of large numbers. Given sufficiently large numbers, even today's supercomputers will have a hard time reversing this operation. By adding an initial step to the process, two nodes may establish a shared secret key for communication over an insecure network only by using the one-way function.
The size of exchanged keys is reduced significantly by using Elliptic Curve cryptography [ECDH]. Due to concerns that earlier curves had been compromised, the security industry moved to Curve25519 [CURVE 25519] in 2014. This at the same time enabled constrained IoT devices such as S2 nodes to use ECDH. Comparatively, Curve25519 requires less computation power than earlier curves.

S2 nodes use an ECDH shared key to derive a temporary link key. The link key allows an including controller to transfer one or more network keys securely to a joining node. Keeping link keys between all nodes in a network would require much non-volatile memory and it would make secure multicast impractical. S0 and S2 uses AES-128 based network keys which are symmetric. This means that all nodes in a given S2 Security Class can encrypt and decrypt commands using the same network key.

© Sigma Designs, Inc. 2016

4.4 Key integrity

S2 protects its keys in a number of ways. Asymmetric ECDH keys used for the temporary secure channel are only used to send a few frames before nodes switch to the assigned S2 Network keys. There is no known attack that allows an attacker to break ECDH with Curve25519. There is no known risk of exposing S2 Network keys during authenticated inclusion.

AES encrypted payload may exhibit recognizable patterns if used to encrypt substantial amounts of data. This may allow an attacker to make qualified guesses and determine the key via brute force methods. All variants of AES use a block size of 128 bits. The key is considered safe until half the entropy has been used, so it can be calculated that nodes may send 2^{64} ~ 300 million terabytes before AES payload does not look entirely random to an attacker.
If an S2 node sends permanently at 100kbit/s, sending this amount of data will take 750 million years.

S2 Network keys are physically stored in S2 nodes. It is theoretically possible for a skilled person to disassemble an alarm sensor and physically extract network keys by connecting directly into discrete memory circuits of the sensor. State-of-the-art system-on-chip designs, such as the Sigma Designs Z-Wave 500-series chip, therefore feature on-chip non-volatile memory which can only be read back by firmware running in the same chip and which is cleared if new firmware is programmed into the chip via the external programming interface.

4.5 Deterring lost devices

It may be impossible to physically reset the network key of a lost device, e.g. if a key fob is stolen from a car. The situation may be noticed by the user or the central gateway may detect that a device has gone missing because the periodic keep-alive notifications stop.

One response would be to start distributing a new network key to all other nodes in the network. This would, however, require that all nodes could run with both the old and new keys – and corresponding old and new singlecast and multicast Nonces – for a longer period before the new key was completely phased in as many sensors only wake up rarely to report their presence to the gateway. The central gateway would even have to remember all previous keys since the user may have bundled (pre-configured) products lying around for months or years before they are unpacked and installed. Updating a key via an older old key makes no sense if there is no trust in the old key. Having an online enabled method for re-enabling the secure ECDH key exchange would open for a new possible attack vector.

The lost key fob of the above example may be blacklisted in the gateway, thereby preventing the key fob from triggering gateway rules.
If the key fob is later found under the driver's seat in the car, the rules can quickly be re-enabled by removing the key fob from the gateway blacklist again.

If a lost device directly controls other Z-Wave nodes, it is necessary to mark these nodes as failing and re-include them with new NodeIDs.

4.6 Secure multicast with no addressing overhead

Native Z-Wave multicast uses a bit-indexed address field to identify receivers of a particular transmission. This has the attractive property that receivers do not have to keep any state on memberships of multicast groups.

© Sigma Designs, Inc. 2016

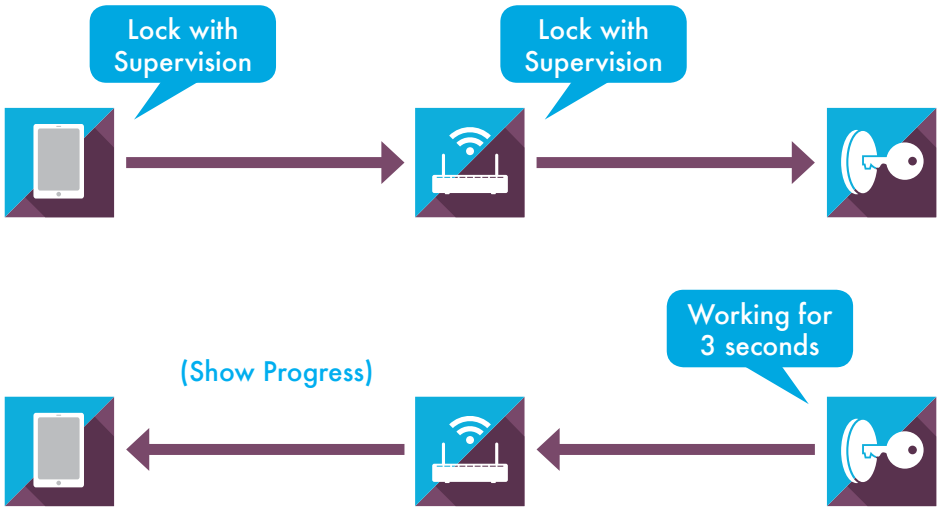
S2 receivers need to be aware of their S2 Multicast memberships. S2 makes a virtue of necessity and uses plain Z-Wave Broadcast frames for S2 Multicast transmissions, thereby avoiding the 30-byte overhead associated with Z-Wave Multicast transmissions. The length of an S2 Multicast frame is therefore comparable to the S2 Singlecast follow-up frame. S2 Multicast receivers recognize S2 Multicast transmissions from the S2 Multicast group owner's NodeID and a one-byte S2 Multicast group ID carried in the S2 Multicast frame header.

S2 uses an auto-updating 13 byte multicast Nonce (MPAN) that all members of an S2 Multicast group must know. Acknowledged S2 Multicast is a two-stage process consisting of one unacknowledged S2 Multicast frame which reaches all group members within range followed by a S2 Singlecast follow up transmission to each group member. Out-of-sync MPANs are automatically updated during the follow-up singlecast session between the transmitter and a multicast group member.

4.7 Application status polling not needed

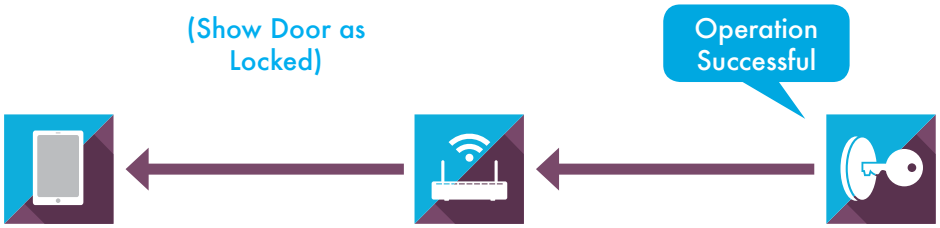
Z-Wave's bandwidth is limited. Yet, a number of control applications have a need for precise status information. One example includes a door lock control panel, which needs to know if the door was really locked after sending a "Lock" command. It may take several seconds for that operation to complete.

The classic approach would be to use polling but Z-Wave Plus puts restrictions on the use of polling as it affects the performance of the entire network and may further violate regional spectrum legislation. With S2, there is no need for polling. All S2 capable devices support the Supervision Command Class [Supervision CC] as it also serves to acknowledge the secure delivery of an S2 encrypted command.
The Supervision Report allows a destination node to advertise its current and future operational state. A receiving door lock may immediately advertise its current Supervision state as "Working" with an expected duration of three seconds, and later, its updated Supervision state "Success".



This makes it possible for a control application to update its UI in a more responsive way as it is notified as soon as the state changes rather than having to poll until a certain state is reached.

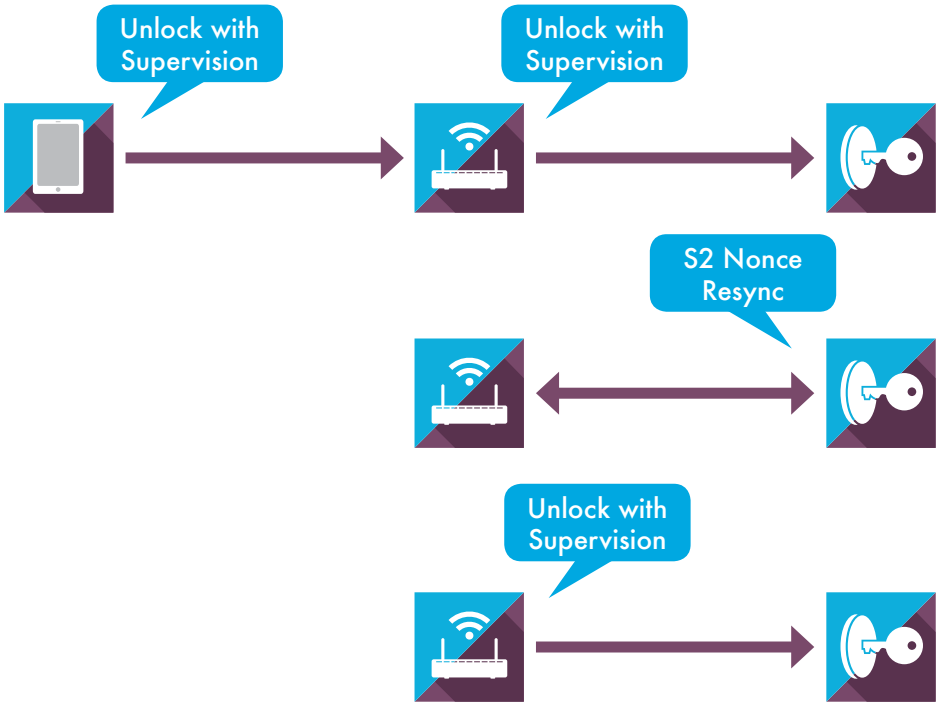
© Sigma Designs, Inc. 2016



4.8 Expedited delivery

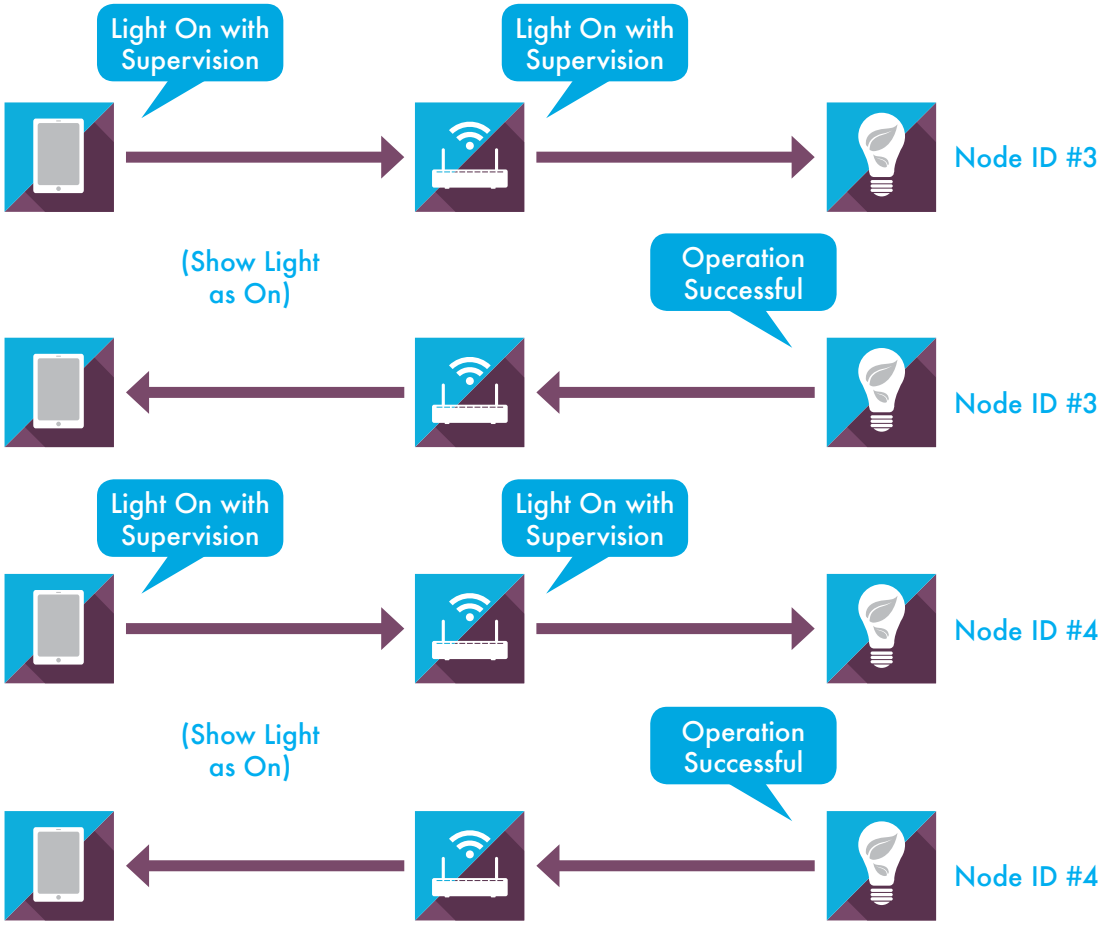
Many first-generation security systems rely on a mandatory challenge-response handshake to protect the network key and to prevent replay attacks. Auto-updated Nonces remove the need for this for S2 Singlecast as well as S2 Multicast transmissions.
If needed, a receiving node may return an error message indicating that singlecast and/or multicast Nonces are out of sync; forcing the sending node to resynchronize with the receiver.

The S2 transport layer integrates the S2 out-of-sync Nonce error messages with the Supervision application-layer acknowledgements. , which are mutually exclusive: If a frame can be decrypted, a Supervision Report is returned. If the frame cannot be decrypted, an S2 out-of-sync Nonce error message is returned. After Nonce resynchronization, the command may re-transmitted.



The Supervision Report allows the transmitter to go on sending commands to other nodes immediately without waiting for an S2 out-of-sync Nonce error message.

© Sigma Designs, Inc. 2016



4.9 Energy efficiency

Low-power radios often use almost the same power when listening as when actively transmitting. Further, the total power consumption of a device depends significantly on the actual application. A battery powered door lock may spend the majority of its battery on driving motors rather than on listening for commands over the radio.

On the other hand, a temperature sensor may be sensitive to the time spent sending or receiving. Consider a sensor based on the Sigma Designs Z-Wave 500 series chip that reports the temperature every five minutes and queries a mailbox for pending commands every 30 minutes.
Without security, such a sensor draws an average current of 2.2µA.
The S0 challenge-response overhead leads to an average current of 5.9µA.
A similar S2 sensor uses only 30% more energy than the non-secure sensor (2.9µA).

Individual frame delivery confirmation may not always be needed. One such example is an over the air (OTA) firmware update which employs its own retransmission state machine. S2 encryption can be used in one-frame-mode. This mode only has a small overhead compared to the non-secure variant. A node can receive an S2 protected OTA firmware update at virtually the same time and battery cost as a non-secure node. A node running S0 would need 200% more bandwidth and energy for the same task.

© Sigma Designs, Inc. 2016

4.10 Active protection against attacks via Nonces

S2 network keys are secret among trusted nodes and should not be revealed to others via predictable patterns in the payload. Therefore, any frame is scrambled by a long Nonce before encryption. The 13 byte singlecast Nonce (SPAN) is auto-updated before each new transmission. This prevents an attacker from recording a command and replaying it later.

An advanced variant of the replay attack is the delay attack: An attacker may 1) at the same time, record a command at the transmitter and jam reception at receiver end, and 2) when the command is retransmitted, record the retransmitted command at the transmitter, jam reception at receiver end and then replay the first recorded command. Most cars are vulnerable to this attack.
A sending S2 node uses the Supervision Command Class to request a delivery acknowledgement for the command. The attacker cannot construct a valid response for the second command; only return an old response or none at all. The sending S2 node can therefore warn the user of suspicious network activities.
Thus, this attack is not only very difficult to accomplish; it can also be detected by the sending S2 node.

4.11 Secure gateway communication

The Z/IP Gateway controls access to the Z-Wave network by only forwarding commands from trusted LAN clients or from a trusted Internet host such as a service provider portal.
An AES-based pre-shared key [TLS PSK] is used by all LAN hosts communicating to Z-Wave nodes via the Z/IP gateway. Two Z/IP Gateways may also interconnect two PANs securely using the open standard DTLS [DTLS].
UDP is used for low latency and to provide support for communication to battery devices in the PAN. UDP is unfortunately problematic in the Internet as it may be used for aggressive DoS attacks. Therefore, firewalls often block UDP.

Therefore, The Z/IP Gateway creates a dedicated tunnel from the inside (the LAN side) of the firewall to a service provider portal. This is known as firewall penetration. The Z/IP Tunnel uses the open standard TLS [TLS]. TLS is a secure overlay to TCP. RSA-1024 standard certificates [RSA] are incorporated in both ends to ensure two-way authentication so that a portal only accepts trusted gateways and a Z/IP Gateway only accepts communication from trusted internet hosts.

© Sigma Designs, Inc. 2016

REFERENCES

[TLS] - IETF RFC5246, Transport Layer Security (TLS) Protocol Version v1.1

[RSA] - US Pat 4,405,829, Rivest, Shamir, Adleman, Cryptographic communications system and method[DTLS] - IETF RFC4347, Datagram Transport Layer Security v1.0

[TLS PSK] - IETF RFC4279, Pre-Shared Key Ciphersuites for Transport Layer Security (TLS)

[ECDH] - Elliptic Curve Diffie-Hellman cryptography, http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-56Ar2.pdf

[CURVE 25519] - D. J. Bernstein, Curve25519: new Diffie-Hellman speed records. Proceedings of PKC 2006. URL: http://cr.yp.to/papers.html#curve25519

[AES] - FIPS197: Advanced Encryption Standard (AES), November 26, 2001, URL: http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf

[NSA_AES] - "National Policy on the Use of the Advanced Encryption Standard (AES) to Protect National Security Systems and National Security Information", June 2003, http://csrc.nist.gov/groups/ST/toolkit/documents/aes/CNSS15FS.pdf

[DIFFIE-HELLMAN] - Diffie, W.; Hellman, M. (1976). "New directions in cryptography". IEEE Transactions on Information Theory 22 (6): 644–654.

Introduction to DH:
https://www.khanacademy.org/computing/computer-science/cryptography/modern-crypt/v/diffie-hellman-key-exchange-part-1

https://www.khanacademy.org/computing/computer-science/cryptography/modern-crypt/v/diffie-hellman-key-exchange-part-2

[HC SECURITY] - Sigma Designs, SDS13349, Security considerations in Home Control installations

[S2] - Sigma Designs, SDS11274, Security 2 Command Class, version 1

[Supervision CC] - Sigma Designs, SDS13321, Supervision Command Class, version 1

© Sigma Designs, Inc. 2016