Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: August 20, 2016

A. Pelov, Ed.
Acklio
L. Toutain, Ed.
Institut MINES-TELECOM ; TELECOM Bretagne
Y. Delibie, Ed.
Kerlink

February 17, 2016

Constrained Signaling Over LP-WAN
draft-pelov-core-cosol-01

Abstract

This document presents a new type of long-range, low-rate radio technologies and an extensible mechanism to operate these networks based on CoAP. The emerging Low-Power Wide-Area Networks (LP-WAN) present a particular set of constraints, which places them at the intersection of infrastructure networks, ultra-dense networks, delay-tolerant networks and low-power and lossy networks. The main objectives of LP-WAN signaling is to minimize the number of exchanged messages, minimize the size of each message in a secure and extensible manner, all with keeping the fundamental principle of technology-independence (L2-independence). This document describes the use of the Constrained Application Protocol (CoAP) as the main signaling protocol for LP-WAN, over which minimal messages are exchanged allowing the full operation of the network, such as authentication, authorization, and management. The use of CoAP signaling provides a generic mechanism that can be applied to different LP-WAN technologies.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 20, 2016.

Table of Contents

1. Introduction

The goal of this document is to provide the necessary mechanisms to operate a Low-Power Wide-Area Network (LP-WAN) by using IETF CoAP [RFC7252] as a core signaling protocol.

Long-range, low-rate radio technologies have emerged in the past several years, and are the base for building LP-WANs. LP-WANs generally have the following characteristics:

- Work in narrow, license-free (ISM) bands with good propagation properties (< 1GHz)
- Low- to very-low throughput (1-200 kbps)
- Low-power operation (25 mW in Europe)
- Long-range communication capabilities (up to 30 km with line-of-sight, several km in urban environment)
- Strong channel access restrictions (1% to 10% duty cycling)
- Infrastructure-based
- Star topology

LP-WANs are built on radio communication technologies, which use advanced signal processing techniques and combination of appropriate modulation and coding approaches to provide the aforementioned radio characteristics.

The absence of license fees and the far-reaching connectivity allow for an extremely competitive pricing of LP-WANs compared to other networking technologies, e.g. cellular or mesh. LP-WANs are sometimes referred to as LPWA or LR-WAN (Low-Rate WAN). Even though LP-WANs are extremely limited in terms of network performance, they are enough for a wide class of applications, among which [LTN001]:

- Metering (water, gas, electricity)
- Infrastructure networks (water, gas, electricity, roads, pipelines, drains)
- Environment/Smart City (waste management, air pollution monitoring and alerting, acoustic noise monitoring, public lighting management, parking management, self service bike rental, digital board monitoring, water pipe leakage monitoring)
- Environment/Country side (soil quality, livestock surveillance, cattle and pet monitoring, climate, irrigation)
- Remote monitoring (house, building)
- Industrial (water tank, asset tracking)

- Automotive (vehicle tracking, impact detection, pay as you drive, assistance request, ...)
- Logistics (goods tracking, conservation monitoring)
- Healthcare (patient monitoring, home medical equipment usage)
- House appliances (pet tracking, white goods, personal asset)
- Truck (tyre monitoring)
- Identification (authentication)

The IEEE is studying LP-WANs, but limited to the case of low-energy critical infrastructure monitoring (LECIM), under the group IEEE 802.15.4k [IEEE.802-15.4k].

The combination of the above characteristics and the envisioned applications define a new class of networks with the following unique constraints:

- Potentially extremely high density (expected of up to 10k-100k+ end-devices managed by a single radio antena)
- Coexistence of delay-tolerant and critical applications (metering and alarms)
- Low-power, low-throughput, lossy connectivity (use of ISM bands)
- Limited payload (100 bytes max, typically less than 50 bytes, 12 bytes for UNB)

CoAP is a client-server protocol specialized for constrained networks and devices. CoAP is highly optimized, extensible, standard protocol, which in conjunction with the Concise Binary Object Representation (CBOR) is the ideal candidate for the signaling protocol of the control plane of an LP-WAN.

It can be used during all stages of the lifecycle of the network, e.g. discovery, authentication, operation. Furthermore, this can be achieved by following RESTful management paradigm, by using a particular resource tree definition or adopting COOL [I-D.veillette-core-cool].

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. LP-WAN Technologies

2.1. Radio technologies

There are two classes of LP-WAN radio technologies, using different radio modulation approaches:

- Ultra Narrow Band (UNB)
- Spread-spectrum (SS)

An example of UNB is the technology developed and promoted by SigFox [SigFox]. Semtech LoRa [LoRa] uses a direct-sequence spread-spectrum with orthogonal codes (OSSS).

Both approaches have their advantages and will coexist in the future, as there are currently several operators, which deploy the two types in the same areas.

2.2. Physical Layer Characteristics

At the physical layer, the important part is the possibility to reconstruct the signal at long distances. The used ISM bands are defined around the world (e.g. 868 MHz in Europe and 900 MHz in USA) and require a 1% (or 10%) duty cycling, or alternatively - advanced detection and channel reallocation techniques. In reality, all deployed networks use the duty cycling limitation, with the following distinction. There is one 100kHz band in which 10% duty cycling is allowed, with a slightly more emission power. The rest of the bands are limited at 1% duty cycling and very restricted power of emission (e.g. 25 mW in Europe).

UNB LP-WANs make the distinction between Uplink and Downlink, first depending on the modulation, and second with the 10% duty-cycling channel been used for the Downlink. OSSS LP-WANs make no such distinction, although for the operation of a network, an operator can chose to use the same Uplink/Downlink channel separation.

Note that the 1% or 10% duty-cycle limitation counts for all trafic originating from an electronic equipment, e.g. an antena managing 100k objects must obey the same limitation as an end-device, with all

frames emitted from the antena (data, acknowledgements) counting towards its quota.

2.2.1. Ultra Narrowband LP-WAN radios

Ultra Narrowband (UNB) technologies generally possess the following physical layer characteristics [LTN003]:

- Uplink:
  - channelization mask 100kHz (600 kHz USA)
  - baud rate 100 bauds (600 bauds USA)
  - modulation BPSK
- Downlink:
  - channelization mask: dynamic selection
  - down link baud rate: 600 baud
  - modulation scheme: GFSK
  - downlink transmission power: 500 mW, 10% duty cycle

2.2.2. Spread-spectrum LP-WAN radios

OSSS technologies possess the following physical layer characteristics [LTN003]:

- channelization mask: from 8 kHz to 500 kHz (depending on spreading factor)
- chip rate: 8 kcps up to 500 kcps
- data rate: 30-50 000 bps
- modulation scheme: equivalent to DSSS with orthogonal signaling

No particular distinction is made between the Uplink and the Downlink.

2.3. MAC Layer Characteristics

Several proprietary MAC frame formats exist for UNB and OSSS. However, they are designed to operate the network in a centralized, highly-vertically-integrated fashion. The only standard MAC frame

format is the IEEE 802.15.4k, which is based on the well-known IEEE 802.15.4 with the addition of a fragmentation sub-layer.

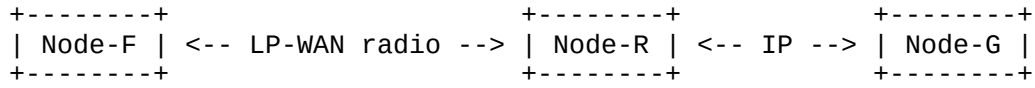
The channel access method is based on ALOHA, although it is up to the network operator to chose if an appropriate end-node polling should be implemented.

3. CoSOL Architecture

3.1. General LP-WAN architecture

We can identify three types of entities in a typical LP-WAN. These are:

- Node-F: far-reachable node, e.g. the end-point, object, device.
- Node-R: radio relay, bridging the LP-WAN radio technology to a different medium (often a LAN or cellular WAN).
- Node-G: gateway node, interconnection between the radio-relay node and the Internet.



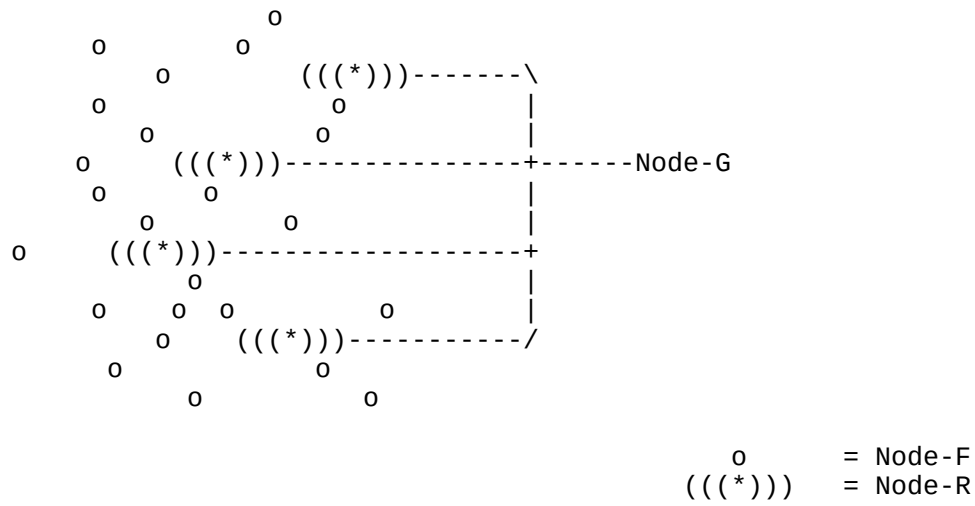
General architecture of an LP-WAN. LP-WAN radio technology is used only between the Node-F and the Node-R.

Figure 1

Of these, only Node-F and Node-R communicate through an LP-WAN radio technology. However, due to the extreme constraints of these technologies, they are always behind a gateway (Node-G). Note, that the Node-R and Node-G can be collocated, e.g. on a single hardware equipment.

The Node-G is connected to the Internet and is assumed to have sufficient computational resources to store a context for each of the Node-Fs. The strong limitation here is the radio link.

In an actual deployment, a (limited) set of Node-Rs cover a large area with a potentially very-high number of Node-Fs. A single Node-G is capable of controlling all Node-Rs.



An example coverage of an area with several Node-Rs. Note that a single Node-F may be covered by several Node-Rs.

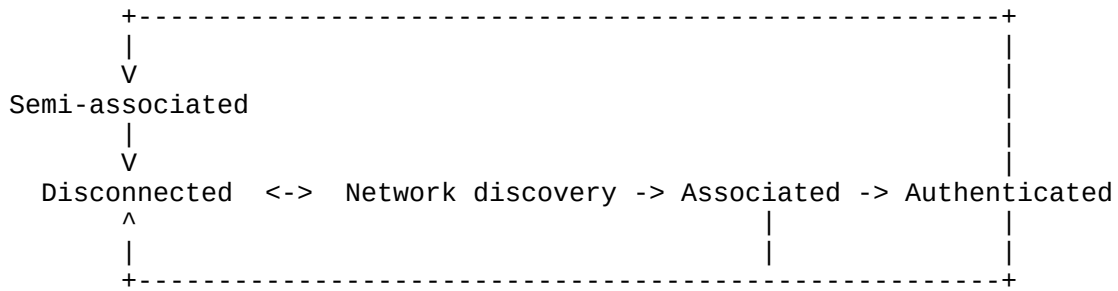
Figure 2

3.2. Node-F lifecycle

Similar to other wireless infrastructure-based technologies, a Node-F can go through several stages:

- Semi-Association
- Network Discovery
- Association
- Authentication
- Dissociation

The Node-F state machine is then the following:



Node-F connectivity state machine.

Figure 3

The Node-F can be in Semi-Associated mode. Upon start, and depending on the application, a Node-F can use a state of uni-directional communication, where it is considered semi-associated to the network. In that state, the Node-F broadcasts frames, handled by the Node-G, but the network cannot join the Node-F on a regular basis. This is a degraded LP-WAN operating mode and if caution is not used, can lead to significant scalability and evolvability issues.

The Network Discovery can be reactive or proactive. The former is based on detecting beacon frames sent periodically by the network (e.g. Node-G). The latter is implemented by the Node-F broadcasting probe request frames, to which all appropriate Node-Gs must respond.

Once a network has been discovered, the Node-F can associate to the network. The association creates the necessary (minimal) context on the Node-G, which initiates the authentication of the Node-F

The authentication is initiated by the Node-G, which should allow for the necessary AAA exchanges to take place. If the authentication is successful, the Node-F enters the Authenticated state. In this stage there is bi-directional communication between the Node-F and the Node-G. If the authentication is not successful, the Node-F enters Disconnected state. Once in Authenticated state, the Node-F can downgrade its connectivity to Semi-Associated mode.

The management of the node in Authenticated state is performed with COOL [I-D.veillette-core-cool]. As an example, managing the parameters of a Semtech LoRa device can be achieved through the use of the YANG module defined in [I-D.pelov-yang-lora]

Finally, the Node-F may decide to dissociate from the network by sending an explicit request. Upon dissociation the Node-G may release all contexts related to the Node-F and re-association

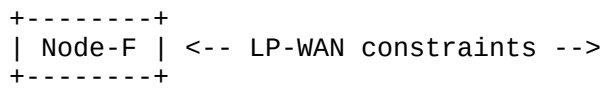
requires going through the authentication stage again. Node mobility is achieved by explicitly dissociating from the old Node-G and then authenticating to the new Node-G. Implicit dissociation is also possible upon the expiration of predefined timers, or in case of mobility optimization.

3.3. CoAP as Signaling Protocol for LP-WANs

Use as CoAP for signaling is implemented as follows. The MAC, network and/or transport layers MUST provide a mechanism to differentiate user data from signaling data frames (e.g. by using separate MAC addresses, IP addresses and/or UDP-ports). Both the Node-G and the Node-F are running CoAP servers for implementing the control plane. Frames exchanged over the LP-WAN radio interface and marked as "signaling data" are handled by the corresponding control plane CoAP servers.

The Node-G runs a (virtual) CoAP server for each Node-F. This server is identified with a DNS name, e.g. "node123.home.node-g.example.com", which can be used explicitly in the CoAP messages via the Proxy-Uri option if needed.

Note, that the Node-R acts only as a transceiver and as such is transparent from protocol point of view. As such, the following management scheme applies:



Node-F connectivity from protocol point of view.

Figure 4

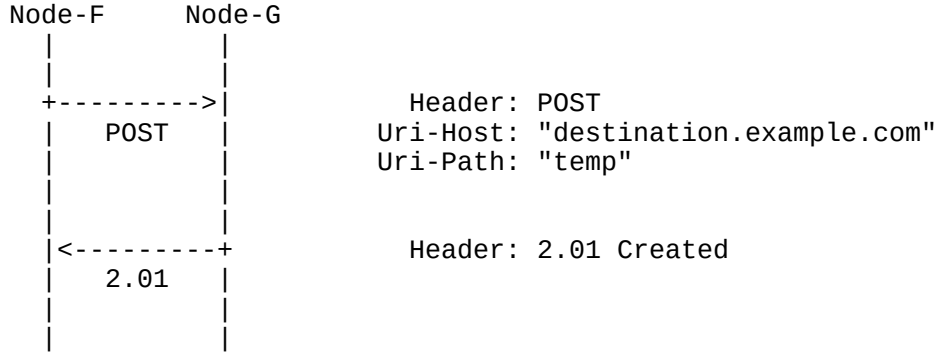
3.3.1. Semi-Association

When in a semi-associated state, a Node-F broadcasts its messages without performing network discovery, or association. If the Node-F is under the coverage of a Node-G, the Node-G will receive the broadcast, and forward the user data. The frames SHOULD be signed, so that they could be authenticated by the network. Layer 2 acknowledgements MUST be used, and in some cases piggybacking on them can provoke the Node-F to associate to the network.

The broadcast messages MUST include the necessary information to join the user data destination, and enough information for the Node-G to authenticate the message sender. This can be achieved through a Confirmable CoAP message, where the user data are POSTed to a well-

known resource defined on the Node-G. DTLS with integrity check can be used, with long-lived keys negotiated by the Node-F and the network. Alternatively, COSE objects may provide the necessary mechanisms.

Even though an application can be implemented by using only simplex association capabilities, there are non-negligible negative consequences related to scalability and evolvability in this case. For example, a Node-F which periodically broadcasts information will occupy the spectrum, even if there is no operator willing to accept its trafic. In addition, no channel access management can be applied.



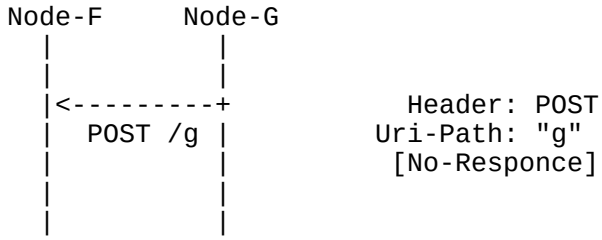
Sending a message in a semi-associated state.

Figure 5

3.3.2. Network Discovery

A network can be discovered by a Node-F reactively or proactively.

Reactive network discovery is based on the detection of periodic beacons emitted by the Node-G. The beacons are implemented with CoAP messages with the No-Response option [I-D.tcs-coap-no-response-option]. The Node-G POSTs its information to a well-known resource, e.g. "/network/node-G/" or a resource alias "/g". Alternatively, this could be achieved by POST-ting to a COOL container (e.g. POST /cool with data node ID = 1 for example).

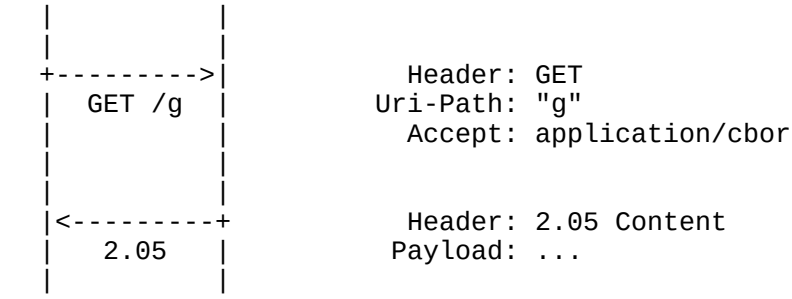


Reactive network discovery. The Node-G sends periodically beacon messages, containing information pertinent to this network.

Figure 6

The CoAP POST request is processed at the Node-F. A resource is created locally, with the representation, which provides the appropriate network parameters, e.g. network ID, Node-G ID, and other radio-related parameters, such as channel, beacon frequency and so forth. This information allows the Node-F to begin the authentication phase.

A Node-F may chose to proactively probe for the existence of network coverage. In that case, it sends a Confirmable CoAP GET request to obtain the information from a well-known resource, normally published by the beacon messages, e.g. "/network/node-G/" or a resource alias "/g" or COOL data node ID ("/cool" data node ID = 2 for example).



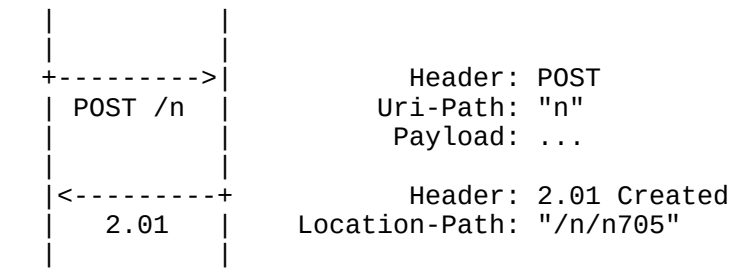
Proactive network discovery. The Node-F request the information of all surrounding Node-Gs.

Figure 7

Once the network is discovered, the Node-F has all necessary information to start the authentication phase.

3.3.3. Association

Before being able to communicate, the Node-F must associate to the network, and then eventually authenticate. The association phase signals to the Node-G that there is a new device willing to communicate with the network. This association SHOULD provide enough information to allow the Node-G to start the authentication process. For example, it may provide the AAA server, which could authenticate the Node-F, or its EAP-Identity. Note, that the Node-F may elect to mark the association message with the No-response option [I-D.tcs-coap-no-response-option], waiting for the subsequent authentication request from the Node-G.



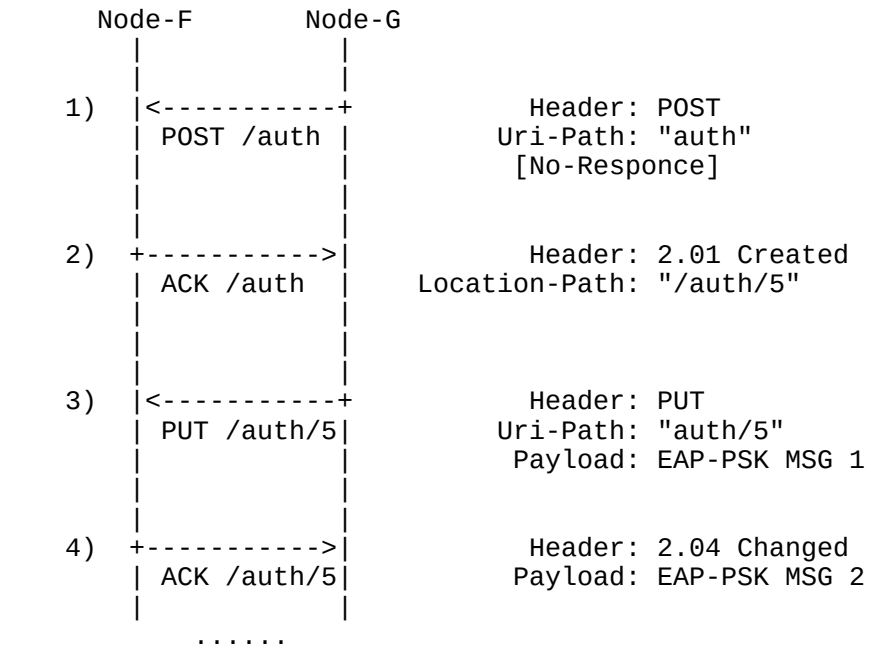
Node-F associates to a network, by creating a corresponding resource element on the Node-G.

Figure 8

3.3.4. Authentication

The EAP-over-CoAP [I-D.marin-ace-wg-coap-eap] specifies an approach to encapsulating EAP messages over CoAP. This allows to authenticate a Node-F, which wishes to join an LP-WAN, and negotiate the L2 encryption keys, and DTLS keying material.

As the Node-F has already associated to the Node-G, it is the Node-G that initiates the authentification request, by going directly to Step 1) of the EAP-over-CoAP specification.



Node-F and Node-G perform mutual authentication following EAP-over-CoAP.

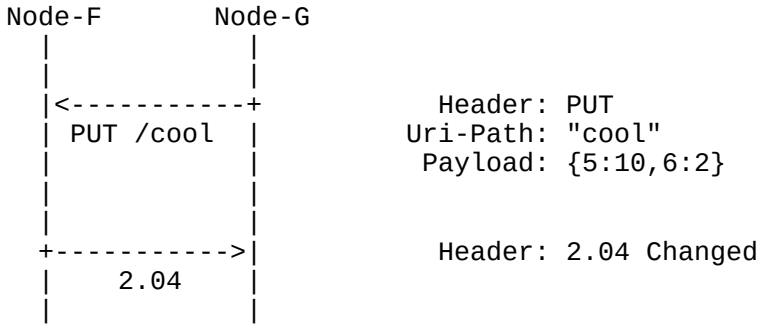
Figure 9

Upon the end of the authentication phase, a Master Shared Key (MSK) is known by the Node-F and the Node-G, and is used to generate DTLS encryption or integrity keys. Further communications should be encrypted/signed with the freshly derived keys.

3.3.5. Operation

Once the Node-F is authenticated to the network, it can send user data via the Node-G to any other end-point on the Internet.

During the operation of the Node-F, the network may need to change one or more parameters concerning the LP-WAN radio parameters of the Node-F. These changes may even concern parameters related to the Node-F itself (such as sleep cycles), its network parameters (e.g. IP addresses), and so forth. This is achieved through the use of COOL [I-D.veillette-core-cool]. The appropriate YANG modules must be present on the Node-F (e.g. a Semtech LoRa Node-F should implement the [I-D.pelov-yang-lora] module).

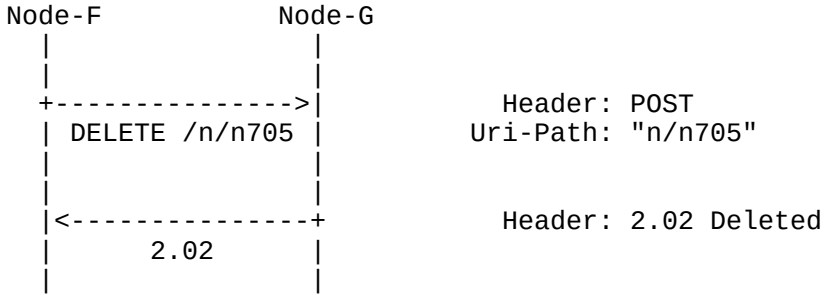


Example, in which the Node-G changes the Spreading Factor (e.g. COOL data node ID = 5) to 10, and the Channel (e.g. COOL data node ID = 6) to 2. Note, that the payload is encoded in CBOR.

Figure 10

3.3.6. Dissociation

If the Node-F wishes to deregister from the network, it could do so by deleting the context created upon association:



Node-F dissociates from the network by deleting its associated resources.

Figure 11

4. Acknowledgements

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

All drafts are required to have a security considerations section. See RFC 3552 [RFC3552] for a guide.

7. References

7.1. Normative References

[I-D.ietf-core-observe]
Hartke, K., "Observing Resources in CoAP", draft-ietf-core-observe-16 (work in progress), December 2014.

[I-D.marin-ace-wg-coap-eap]
Garcia, D., "EAP-based Authentication Service for CoAP", draft-marin-ace-wg-coap-eap-01 (work in progress), October 2014.

[I-D.pelov-yang-lora]
Pelov, A., Toutain, L., Delibie, Y., and A. Minaburo, "YANG module for LoRa Networks", draft-pelov-yang-lora-00 (work in progress), December 2015.

[I-D.tcs-coap-no-response-option]
Bhattacharyya, A., Bandyopadhyay, S., Pal, A., and T. Bose, "CoAP option for no server-response", draft-tcs-coap-no-response-option-13 (work in progress), November 2015.

[I-D.veillette-core-cool]
Veillette, M. and A. Pelov, "Constrained Objects Language", draft-veillette-core-cool-00 (work in progress), November 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC7252] Shelby, Z., Hartke, K., and C. Bormann, "The Constrained Application Protocol (CoAP)", RFC 7252, DOI 10.17487/RFC7252, June 2014, <http://www.rfc-editor.org/info/rfc7252>.

7.2. Informative References

[IEEE.802-15.4k]
Institute of Electrical and Electronics Engineers, "Low-Rate Wireless Personal Area Networks (LR-WPANs) - Amendment 5: Physical Layer Specifications for Low Energy, Critical Infrastructure Monitoring Networks., IEEE 802.15.4k", IEEE Standard 802.15.4, 2013.

[LoRa] Semtech, "https://web.archive.org/web/20150510011904/https://www.semtech.com/wireless-rf/lora.html", May 2015.

[LTN001] European Telecommunications Standards Institute, "Low Throughput Networks (LTN); Use Cases for Low Throughput Networks, ETSI GS LTN 001", IEEE ETSI GS LTN 001, 2014.

[LTN003] European Telecommunications Standards Institute, "Low Throughput Networks (LTN); Protocols and Interfaces, ETSI GS LTN 003", IEEE ETSI GS LTN 003, 2014.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, DOI 10.17487/RFC3552, July 2003, <http://www.rfc-editor.org/info/rfc3552>.

[SigFox] SigFox, "https://web.archive.org/web/20150628225901/http://www.sigfox.com/en/#!/technology", June 2015.

Authors' Addresses

Alexander Pelov (editor)
Acklio
2bis rue de la Chataigneraie
Cesson-Sevigne, Bretagne 35510
FR

Email: a@ackl.io

Laurent Toutain (editor)
Institut MINES-TELECOM ; TELECOM Bretagne
2 rue de la Chataigneraie
Cesson-Sevigne, Bretagne 35510
FR

Email: laurent.toutain@telecom-bretagne.eu

Yannick Delibie (editor)
Kerlink
1 rue Jacqueline Auriol
Thorigne-Fouillard, Bretagne 35235
FR

Email: yannick.delibie@kerlink.fr