Constrained Radio Networks, Small Ciphertexts, Signatures, and Non-Interactive Key Exchange

John Preuß Mattsson, Göran Selander, Ben Smeets, Erik Thormarker
Ericsson

Abstract: Low Power Wide Area Networks (LPWANs) is a huge and very quickly growing market expected to reach over 1000 billion USD globally by 2027. Due to their relatively large ciphertext and signature sizes and the lack of Non-Interactive Key Exchange (NIKE) the currently selected PQC algorithms are unusable in many IoT systems using constrained radio networks. NIST and academia must work together to, if possible, standardize algorithms with smaller ciphertexts, smaller signatures, and with additional features such as NIKE.

1. Introduction

The NIST PQC Call for Proposals [1] included several IoT device aspects as requirements and evaluation criteria such as low power, constrained memory, or limited real-estate. The Call for Proposals also included some communication aspects such as bandwidth-constrained applications or Internet protocols with limited packet size.

The NIST report on the third round of the NIST PQC Standardization Process [2] focuses mostly on constrained device aspects such as clock cycles, memory, and storage, but has little discussion on constrained communication aspects except comparison of public key, ciphertext, and signature sizes.

There are many types of constrained radio networks. Low-Power Personal Area Networks (LPPANs) such as Bluetooth Low Energy has a reach of a few centimeters to a few meters. According to Data Bridge Market Research [3] the market size for Bluetooth Low Energy is expected to reach 27 billion USD by 2028, an annual growth rate of 19.6% throughout the forecast period. Low Power Wide Area Networks (LPWANs) with a reach of several kilometers is a huge and very quickly growing market. According to Maximize Market Research [4] the market for LPWANs is expected to reach over 1000 billion USD by 2027, an annual growth rate of 89.2% throughout the forecast period.

A PQC candidate with relatively small ciphertexts (SIKE with 346 bytes), and the candidates with the smallest signatures (GeMSS with 33 bytes and Rainbow with 66 bytes) were all eliminated from the NIST Post-Quantum Cryptography Standardization Process because of serious attacks that completely undermined the security. The selected PQC algorithms have much larger ciphertexts (Kyber with 768 bytes) and signatures (Falcon with 666 bytes and Dilithium with 2420 bytes), which work well for Web application but are very problematic or even unusable in many IoT systems using constrained radio networks.

In this position paper we give an overview of various constrains of LPWANs and explain why the currently selected algorithms are unusable in many of these constrained network technologies. The conclusions are that elliptic curve cryptography must be allowed to be used until the risk of Cryptographically Relevant Quantum Computers (CRQCs) is imminent and that academia and NIST must work together to, if possible, standardize algorithms with smaller ciphertexts, smaller signatures, and with additional features such as Non-Interactive Key Exchange (NIKE).

2. Constrained Radio Networks

Constrained radio networks are not only characterized by very small frame sizes on the order of tens of bytes transmitted a few times per day at ultra-low speeds, but also high latency, and severe duty cycles constraints. Devices are often low power so that battery-powered devices can be deployed with inexpensive batteries for over 10 years. Some constrained radio networks are also multi-hop where the already small frame sizes are additionally reduced for each additional hop. Too large payload sizes can easily lead to completely unacceptable completion times due to fragmentation into a large number of frames and long waiting times between frames can be sent (or resent in the case of transmission errors). This is especially true during network formation where many devices must send larger than normal messages during a limited time.

The number of different constrained radio network technologies is large and growing. Some examples of constrained network technologies are LoRaWAN, NB-IoT, Sigfox, Wi-SUN FAN, Bluetooth Low Energy, and IEEE 802.15.4. IEEE 802.15.4 is used in Zigbee, ISA100.11a, WirelessHART, MiWi, 6LoWPAN, 6TiSCH, Thread and SNAP. A few examples of LPWANs:

- LoRaWAN is often used with 51 bytes frames in Europe and 11 bytes frames in the US. The payload sizes for application data are even smaller as some of the bytes are needed for protocol headers such as IP. LoRaWAN is often used with duty cycles less than 1%. According to [5], Section 5.4.1, the duty cycle is by default relative to an observation time window (Tobs) of 1 hour, so a 1% duty cycle means that a device has a 36 second period where it can transmit information and then the device must pause transmission for an hour.

- 6TiSCH is a multi-hop mesh network using the time-slotted channel hop (TSCH) mode of IEEE 802.15.4. The frames are at most 127 bytes but drops for each hop. The actual payload size depends on configuration but in a typical 5-hop network the actual payload size can be about 45 bytes [7]. A possible performance metric is the network formation time for initial installation, taking account dependency on intermediate nodes being installed at the same time.

- NB-IoT is a LPWAN operating in licensed spectrum that is not characterized by fixed sized frames. The effects of larger messages are not as extreme as in LoRaWAN and 6TiSCH but as in all LPWAN, the byte count has a significant impact on time and energy consumption.

For an overview of Low-Power Wide Area Networks (LPWANs) and their limitations, see [6] and [7]. The reason why the LPWAN networks are so constrained and will continue to be constrained are that available spectrum and battery capacity is limited, and the relatively high energy costs associated with radio transmission.

In constrained radio networks, the processing energy costs are typically almost negligible compared to the energy costs for radio and the energy costs for sensor measurement, see e.g., [8]. In active radio networks, transmission, receiving, and listening typically requires similar amounts of power. In passive radio networks, the energy to transmit at one endpoint can be almost eliminated as the device antenna works as a mirror. The single most important factor to reduce power consumption in devices connected via radio is to keep the number of bytes or frames as low as possible. Keeping the number of bytes or frames low is also essential for low latency and time to completion as well as efficient use of spectrum to support a large number of devices. Due to the ultra-low speeds and severe duty cycle constraints, the number of bytes is often a much more important factor for latency and time to completion than the number of clock cycles.

To reduce overhead and processing in constrained radio networks, IETF has created several working groups and technologies for constrained networks, e.g., (here technologies in parenthesis when the name is different from the working group): 6lo, 6LoWPAN, 6TiSCH, ACE, CBOR, CoRE (CoAP, OSCORE), COSE, LAKE (EDHOC) ROLL (RPL), and LPWAN (SCHC).

3. Cryptography in Constrained Radio Networks

The power consumption for symmetric cryptography like AES and SHAKE is completely negligible in devices using radio, as the power consumption is several orders of magnitudes lower than the power for transmission, receiving, and listening over the radio [9]. The power consumption for traditional asymmetric cryptography such as ECDHE or EdDSA is typically within the same order of magnitude as actively transmitting or receiving a message over radio [9]. NIST Lightweight Cryptography (LWC) [10] might be useful to lower hardware chipset area, memory usage, or storage needed for software implementations of symmetric algorithms but will not have any effect on energy consumption in constrained radio networks (unless the symmetric algorithm is the major energy drain in an asymmetric algorithm). Saarinen has calculated the computational energy costs for many of the algorithms in NISTs PQC standardization project [11]. As can be seen from [9] and [11] the energy costs for encapsulation/decapsulation with Kyber or signing/verification with Dilithium or Falcon will be almost negligible compared to transmitting/receiving the encapsulations or signatures over radio. The absolutely most important factor for power consumption and battery lifetime in IoT systems using radio is the size of the messages required for key exchange or signatures. As stated in [12]:

> *"In wireless and sensor networks, conserving battery power is a prime concern, and so the energy cost of communication must be minimised. Thus using key establishment methods that minimise the number of bits that need to be transmitted is of fundamental importance."*

It is unfortunately not uncommon with constrained IoT systems that still operate without any security, and when they do use security, they often rely on symmetric Pre-Shared Keys (PSK) for authentication and key exchange. PSK authentication has the major weakness that it typically uses permanent key identifiers sent in cleartext, which enables attackers to identify and track endpoints. PSK key exchange has the major weakness that it does not offer forward secrecy implying that a breach in the PSK supply chain leads to that all derived keys, and all protected messages are compromised. Pairwise PSK authentication and key exchange are also not usable in groups as the number of keys that need to be distributed scales as $O(n^2)$. Symmetric group keys which are also commonly used in constrained IoT are even worse as misbehaving group members can passively read messages between other members of the group or impersonate and actively inject messages between other members of the group. Endpoint identity protection should be a requirement for all future systems where it is technically possible. Always assuming breach such as key compromise and minimizing the impact of breach are essential zero trust principles. Using key exchange with forward secrecy such as ephemeral Diffie-Hellman should therefore be a requirement for all future systems where it is technically possible. In fact, running Diffie-Hellman a single time is not enough, as an attacker that compromises a single session key (static key exfiltration [13]) can passively eavesdrop on all (future) messages and actively inject messages into a connection (e.g., a TLS 1.3 connection) between the real endpoints. To better follow zero trust principles, Diffie-Hellman should be rerun frequently which limits the impact of key compromise to a limited time and data and forces the attacker to dynamic key exfiltration which has a much higher risk of detection [13]. ANSSI requires ephemeral Diffie-Hellman key exchange every hour or 100 GB for IPsec [14], but the requirement to frequently run ephemeral Diffie-Hellman is a general principle that should be followed by all systems aligning with zero trust principles. The Signal Protocol [15] is designed from the start to enable very frequently rerunning of Ephemeral Diffie-Hellman and EDHOC [16] is designed to be as compact as possible so that rerunning Ephemeral Diffie-Hellman does not come with too high latency and energy costs. To offer acceptable security and privacy, the use of PSKs should be phased out or complemented with asymmetric cryptography for client identity protection and forward secrecy whenever technically possible. To offer acceptable best practice privacy 3GPP 5G has e.g., introduced public server keys for client identity protection with ECIES [17] and to align with zero trust principles and supply chain security, there is work on complementing the PSK key exchange with ECDHE [18]. Hopefully both of these additions can be made mandatory to use in the future.

To offer acceptable privacy and align with zero trust principles many constrained IoT networks have or are planning to migrate to asymmetric cryptography. Not only because of better privacy and alignment with zero trust, but also because public key cryptography offers much simpler and scalable key distribution. Elliptic Curve Diffie Hellman is a perfect fit as it is efficient, has small key sizes, and can be used in a large number of different and very useful ways. Using the e, s, ee, es, se, ss notation from the Noise protocol framework [19], a simple ECDH key exchange can e.g., be done in the following ways:

1. An Ephemeral-Ephemeral key exchange. This is the type of key exchange required to get forward secrecy. This type of key exchange is used in e.g., IKEv2, TLS 1.3, WireGuard, Signal, and EDHOC.

```
  -> e
  <- e, ee                                                  (1)
```

2. A Static-Ephemeral key exchange. This is useful for systems with pre-configured static keys as information can be encrypted in the “e, se” message. It is used in e.g., ECIES/HPKE, Noise IK, and WireGuard.

```
  <- s
  ...                                                          (2)
  -> e, se
```

3. An Ephemeral-Static key exchange. This is useful as the “s, es” message gives implicit authentication and that the whole AKE can be built with only ECDH and no signatures. It is used in e.g., Noise XX, Noise IK, WireGuard, and EDHOC.

```
  -> e
  <- s, es                                                     (3)
```

4. A Static-Static key exchange also known as a Non-Interactive Key Exchange (NIKE) [12]. This is useful as no messages are needed for the key exchange. Note that the static public keys would typically be distributed by a centralized node. Another commonly used NIKE is Blom’s scheme [20]. NIKE is used in in e.g., CMS, COSE, Group OSCORE, and HDCP, and is often suggested as a future security solution for authentication and/or key exchange in many IoT systems using very constrained radio where even Ephemeral-Ephemeral ECDH is problematic.

```
  -> s
  <- s
  ...                                                          (4)
  -> ss
```

Note that the above minimalistic message flows are typically not used as is, but instead used as parts of a larger authenticated key exchange. All of the above types of Diffie-Hellman Key exchange are used significantly in deployed and planned systems. The current NIST PQC algorithms are all KEMs or Signature algorithms. KEMs can be used directly as a drop-in replacement to (1) and (2) but not to (3), (4), (5), or (6), where (5) is the combination of (1) and (2), and (6) is the repeated interleaved use of (1).

5. A single public ephemeral key used in both Static-Ephemeral and Ephemeral-Ephemeral Diffie-Hellman. It is used in e.g., Noise XX, Noise IK, WireGuard, and EDHOC.

```
  <- s
  ...
  -> e, se                                                     (5)
  <- e, ee
```

6. A single public ephemeral key used in two different Ephemeral-Ephemeral Diffie-Hellman. The Signal Diffie-Hellman Ratchet consists of repeated interleaved use of this pattern.

```
-> e
<- e, ee                                                          (6)
-> e, ee
```

The current NIST PQC algorithms are an acceptable replacement for Ephemeral-Ephemeral ECDH (1) and signatures in non-constrained protocols built on SIGMA-I such as IKEv2 and TLS 1.3. The messages will be significantly larger but the performance in unconstrained use cases such as the Web will be acceptable.

For systems such as Noise IK, Noise XX, WireGuard, Signal, or EDHOC that uses Static-Ephemeral ECDH (2), Ephemeral-Static ECDH (3), or the combinations (5) and (6) the NIST PQC algorithms are not a direct replacement at all. Using the NIST KEMs or signatures leads to much larger messages and/or more flights. This might be acceptable for non-constrained use cases where WireGuard is typically used but leads to completely unacceptable performance degradation in many use cases where EDHOC is used. EDHOC [16] is a protocol designed by the IETF Lightweight Authenticated Key Exchange (LAKE) working group. Its main design goal it to standardize an AKE with as small messages as possible. The message sizes in EDHOC mode 3 (similar to Noise XX) are in a typical use case 37, 45, and 19 bytes long for a total of 101 bytes. Using Kyber and Falcon, the sizes would be 773, 1440, and 679 bytes long for a total of 2896 bytes, an increase with about 2800%! But as stated in [2], Falcon may be infeasible to implement on constrained devices. Using Kyber and Dilithium, the sizes would be 773, 3194, and 2433 bytes long for a total of 6400 bytes, an increase with about 6200%! Using only Kyber would increase the number of bytes more than using Kyber and Falcon, and also require more flights, which would further increase latency.

Group OSCORE [21] is an addition to CoAP enabling group requests with a signature or pairwise communication between any nodes in the group with Static-Static ECDH i.e., NIKE. A group request in OSCORE consists of a few bytes of header in addition to the payload and the signature. Changing 64 bytes ECDSA or EdDSA to 666 bytes Falcon increases the overhead with around 900%. Changing 64 bytes ECDSA or EdDSA to 2420 bytes Dilithium increases the overhead with around 3600%! A pairwise unicast message uses Static-Static ECDH and consists of a few bytes of header, the payload, and typically an 8 bytes AEAD tag. Changing from 0 bytes NIKE to 768 bytes Kyber increases the overhead for the first message with around 9000%.

Under optimal conditions, the energy usage for transmission and reception are proportional to the number of bytes, but because of decreased signal-to-noise ratios when many devices transmit at the same time, the power usage and single-fragment latency is often a superlinear function of the total number of bytes. In constrained radio networks such as LoRaWAN with severe duty cycles constraints, the time to completion is not proportional to the number of bytes as just a single extra byte or the need for retransmission might trigger an hour pause before continued transmission. For the most constrained networks it might take many hours or even days to transmit a single Dilithium signature.

Summary and Conclusions

Constrained radio networks such as LPWANs is a quickly growing market expected to reach over 1000 billion USD globally by 2027. Constrained radio networks are not only characterized by very small frame sizes on the order of tens of bytes transmitted a few times per day at ultra-low speeds, but also high latency, and severe duty cycles constraints. Due to their relatively large ciphertext and signature sizes and the lack of Non-Interactive Key Exchange (NIKE) the currently selected PQC algorithms have serious negative performance impact and are for that reason unusable in many IoT systems using constrained radio networks. Using the selected PQC algorithms instead of ECDH would in many cases increase the number of bytes with several thousand percent. Due to duty cycles the increase in time to completion is often even larger. The size of the messages sent over radio is the single most important factor for power consumption and battery lifetime in IoT systems using radio.

NIST and academia must work together to, if possible, standardize algorithms with smaller ciphertexts, smaller signatures, and with additional features such as NIKE. While Rainbow is completely broken, the Oil and Vinegar scheme still offer very small signatures but has quite large public keys. Blom's scheme [20] is unconditionally secure (i.e., also against quantum attacks) as long as no more than k users are compromised. Unfortunately, the scheme has a master key, which is not acceptable in many use cases, and the security completely fails when $k + 1$ users are compromised, which happened in the case of HDCP. The isogeny-based CSIDH [22] seems like an extremely promising algorithm for constrained radio networks, not only are the 64 bytes public keys and encapsulations/ciphertexts extremely small, just twice the size of on ECDH public key, but CSIDH also supports NIKE and is a direct replacement for EDCH in all of the patterns (1)–(6). That CSIDH is a bit slower than currently selected PQC algorithms is not a showstopper for constrained radio networks as message size is typically the dominant factor for energy, efficient use of spectrum, and latency. Given that CSIDH requires significantly more computation than the selected PQC algorithms but results in significantly smaller messages it is uncertain how an optimized CSIDH implementation would compare on total energy consumption such as Figure 10 of [2]. With its resulting much smaller messages, CSIDH would however drastically lower latency and time to completion in many constrained radio networks compared to the currently selected PQC algorithms. The small message sizes would also mean efficient use of spectrum enabling support of a large number of devices. Note that in constrained sensors, the CPU is mostly only used to read and transmit sensor measurements. That the CPU is occupied is not blocking other tasks like it does on a Web server. More work on CSIDH would undoubtedly lead to more optimized code as well as new ideas for algorithmic optimizations. To work well in constrained radio networks, the message sizes need to align with the tens of bytes transmitted a few times per day that the networks are designed for. Infrequently sending a few hundred bytes is acceptable in many constrained networks but sending a thousand bytes is not feasible in more constrained networks. Note that static keys often do not need to be sent over constrained links, as they can be provisioned or accessed over non-constrained links. Moreover, signatures can in many cases be replaced by a symmetrical MAC from an Ephemeral-Static or Static-Static key exchange by changing the architecture and protocols, as long as the proving node is online. NIST should publicly acknowledge that post-quantum algorithms for constrained radio networks is a major unsolved problem and that NIST might have a fifth round in the PQC project or a separate Lightweight Post-Quantum Cryptography Project. This would encourage academia and funding institutions to prioritize this important area. The actual specification and publication of Post-Quantum Non-interactive key exchanges could also be done in IRTF CFRG which recently has worked as an excellent complement to NIST for globally recognized cryptographic algorithms.

As there will not be any NIST approved post-quantum algorithms usable in constrained radio networks for the foreseeable future, it is essential that elliptic curve cryptography is allowed to be used until the risk of Cryptographically Relevant Quantum Computers (CRQCs) is imminent. For national security systems encrypting TOP SECRET information that need to stay secret for many many decades, the migration to PQC is urgent. Other systems such as DNSSEC are rightfully adopting a wait-and-see strategy [23]:

> *"without massive and unexpected discoveries in both quantum physics and engineering for quantum computers, there is no chance that a cryptographically relevant quantum computer (CRQC) could be built in the next decade, and possibly not for many decades. Even after waiting a decade, there will clearly be years if not decades of warning before a CRQC will be built, and that amount of time will be more than sufficient for the DNSSEC community to adopt one or more appropriate signature algorithms based on post-quantum cryptography (PQC)."*

Unless NIST allows use of ECC until the risk of CRQCs is imminent, constrained radio networks will likely use (group) PSK authentication and (group) PSK key exchange with its bad privacy and failure to adhere to zero trust principles. It is still uncertain if a CRQC will ever be built.

References

[1] NIST, "Submission Requirements and Evaluation Criteria for the Post-Quantum Cryptography Standardization Process"
https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/call-for-proposals-final-dec-2016.pdf

[2] NIST IR 8413, "Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process"
https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8413.pdf

[3] Data Bridge Market Research, "Global Bluetooth Low Energy Market – Industry Trends and Forecast to 2028"
https://www.databridgemarketresearch.com/reports/global-bluetooth-low-energy-market

[4] Maximize Market Research, "Low Power Wide Area Network Market – Global Industry Analysis and Forecast (2022-2027)"
https://www.maximizemarketresearch.com/market-report/global-low-power-wide-area-network-market/7252/

[5] ETSI 300 220-1, "Short Range Devices (SRD) operating in the frequency range 25 MHz to 1 000 MHz; Part 1: Technical characteristics and methods of measurement"
https://www.etsi.org/deliver/etsi_en/300200_300299/30022001/03.01.01_60/en_30022001v030101p.pdf

[6] IETF RFC 8376, "Low-Power Wide Area Network (LPWAN) Overview"
https://www.rfc-editor.org/rfc/rfc8376

[7] IETF LAKE WG, “Requirements for a Lightweight AKE for OSCORE”
https://datatracker.ietf.org/doc/html/draft-ietf-lake-reqs

[8] Taoufik, Bouguera, Diouris, Chaillout, Jaouadi, Andrieux, “Energy Consumption Model for Sensor Nodes Based on LoRa and LoRaWAN”
https://doi.org/10.3390/s18072104

[9] Meulenaer, Gosset, Standaert, Pereira, “On the Energy Cost of Communication and Cryptography in Wireless Sensor Networks”
https://perso.uclouvain.be/fstandae/PUBLIS/55b.pdf

[10] NIST Lightweight Cryptography
https://csrc.nist.gov/projects/lightweight-cryptography

[11] Saarinen, “Mobile Energy Requirements of the Upcoming NIST Post-Quantum Cryptography Standards”
https://arxiv.org/pdf/1912.00916.pdf

[12] Freire, Hofheinz, Kiltz, Paterson “Non-Interactive Key Exchange”
https://eprint.iacr.org/2012/732.pdf

[13] IETF RFC 7624, “Confidentiality in the Face of Pervasive Surveillance: A Threat Model and Problem Statement”
https://www.rfc-editor.org/info/rfc7624

[14] ANSSI DAT-NT-003, “Recommendations for securing networks with IPsec”
https://www.ssi.gouv.fr/uploads/2015/09/NT_IPsec_EN.pdf

[15] The Signal Protocol
https://signal.org/docs/

[16] Selander, Preuß Mattsson, Palombini, “Ephemeral Diffie-Hellman Over COSE (EDHOC)”
https://datatracker.ietf.org/doc/html/draft-ietf-lake-edhoc

[17] 3GPP TS 33.501, “Security architecture and procedures for 5G System”
https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169

[18] Arkko, Norrman, Torvinen, Preuß Mattsson, “Forward Secrecy for the Extensible Authentication Protocol Method for Authentication and Key Agreement (EAP-AKA' FS)”
https://datatracker.ietf.org/doc/html/draft-ietf-emu-aka-pfs

[19] Noise Protocol Framework
http://noiseprotocol.org

[20] Blom, “An optimal class of symmetric key generation systems”
https://www.cs.utexas.edu/users/lam/396m/papers/Blom.pdf

[21] Tiloca, Selander, Palombini, Preuß Mattsson, Park, “Group OSCORE - Secure Group Communication for CoAP”
https://datatracker.ietf.org/doc/html/draft-ietf-core-oscore-groupcomm

[22] Castryck, Lange, Martindale, Panny, Renes, "CSIDH: An Efficient Post-Quantum Commutative Group Action"
https://eprint.iacr.org/2018/383.pdf

[23] ICANN, "Quantum Computing and the DNS"
https://www.icann.org/en/system/files/files/octo-031-11feb22-en.pdf