Post-Quantum Secure Over-the-Air Update of Automotive Systems

Joppe W. Bos\textsuperscript{1} [0000–0003–1010–8157], Alexander Dima\textsuperscript{2}, Alexander Kiening\textsuperscript{3} [0000–0003–2161–8662], and Joost Renes\textsuperscript{4} [0000–0003–1884–6330]

\textsuperscript{1} NXP Semiconductors, Leuven, Belgium
joppe.bos@nxp.com
\textsuperscript{2} PiNTeam GmbH, Germany
alexander.dima@pinteam.eu
\textsuperscript{3} DENSO AUTOMOTIVE Deutschland GmbH, Germany
a.kiening@eu.denso.com
\textsuperscript{4} NXP Semiconductors, Eindhoven, Netherlands
joost.renes@nxp.com

\textbf{Abstract.} With the announcement of the first winners of the NIST Post-Quantum Cryptography (PQC) competition in 2022, the industry has now a confirmed foundation to revisit established cryptographic algorithms applied in automotive use cases and replace them with quantum-safe alternatives. In this paper, we investigate the application of the NIST competition winner CRYSTALS-Dilithium to protect the integrity and authenticity of over-the-air update packages. We show how this post-quantum secure digital signature algorithm can be integrated in AUTOSAR Adaptive Platform Update and Configuration Management framework and evaluate our approach practically using the NXP S32G vehicle network processor. We discuss two implementation variants with respect to performance and resilience against relevant attacks, and conclude that PQC has little impact on the update process as a whole.

\textbf{Keywords:} Post-Quantum Cryptography · Over-the-Air Update · Migration.

\section{Introduction}

Digital signatures are one of the core cryptographic building blocks in modern digital security. The goal of this concept, invented by Diffie and Hellman \cite{DiffieHellman}, is to provide message authentication against the public key of a sender. Applications of digital signatures are countless and diverse. In this work we look at the application of digital signatures in automotive applications where secure boot and secure (over-the-air) update are two main applications to protect against modifications of the software or firmware with potentially malicious intentions.

In the last decades multiple digital signature designs have been standardized. These approaches are either based on Elliptic Curve Cryptography (ECC) \cite{ECC} or the Rivest–Shamir–Adleman (RSA) \cite{RSA} algorithm. However, assuming the
availability of a large-scale quantum computer the security of these “classical”
approaches is threatened by Shor’s quantum algorithm \[50\] which is able to re-
cover such ECC/RSA private keys in polynomial time. To prepare for this quan-
tum threat, alternative public-key algorithms are necessary. These are typically
referred to as post-quantum or quantum-safe algorithms. The new algorithms
are intended to run on classical hardware yet provide sufficient protection even
against adversaries that are in possession of a quantum computer. They are
not to be confused with quantum cryptography such as Quantum Key Distri-
bution (QKD), where the cryptographic algorithms also run on infrastructure
(partially) consisting of quantum computers.

Recognizing the threat of quantum computer, the US National Institute of
Standards and Technology (NIST) put out a call for proposals \[42\] in 2016
to submit candidates for a new standards consisting of post-quantum secure
algorithms. In July 2022, NIST has recommended one primary algorithm for
digital signatures: CRYSTALS-Dilithium \[14,35\] (denoted simply as Dilithium
in the remainder of this paper) In addition, two other signature schemes will
also be standardized: Falcon \[45\] and SPHINCS+ \[5,23\]. The final standard for
Dilithium is expected to be published in 2024, which will deviate slightly from
the original Dilithium proposal due to comments from industry and academia
that were received as part of the standardization process. The NIST standard-
ized algorithms are also receiving interest from European standardization and
norms bodies. For example, the German Federal Office for Information Security
(BSI) will consider including Dilithium into their Technical Guidelines for Cryp-
tographic Mechanisms TR-02102-1 \[15\] Section 4] after the final NIST standard
has been published.

In this paper, we investigate the practical impact of Dilithium in one of
the key automotive use-cases: Over-The-Air (OTA) update. This is done from a
Tier-1 perspective where all the communication cost from host processor to the
secure processor is taken into account. For our experiment, we use the S32G
vehicle network processor and compare running the post-quantum secure sig-
nature verification in two settings. Firstly, the verification is performed on the
microprocessor which is based on the Arm Cortex-A53 and secondly, on the mi-
crocontroller which is part of the Hardware Security Engine (HSE) and based on
the Arm Cortex-M7. The A53 is larger and faster compared to the M7 but does
not offer security features to protect key-material or against advanced attacks
such as fault injection \[8,7\].

2 Background and Related Work

2.1 Post-Quantum Digital Signatures

In this work we follow the recommendation from NIST and focus on lattice-
based digital signatures. This area in cryptography has a rich history and a wide
variety of options relying on different hardness assumptions. One approach in
lattice-based cryptography is based on Regev’s work introducing the Learning
With Errors (LWE) problem \[47\], which relates to solving a “noisy” linear system
Table 1. Public key and signature sizes for each security level of Dilithium in bytes.

<table>
<thead>
<tr>
<th>Security level</th>
<th>Public key</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilithium-2</td>
<td>1312</td>
<td>2420</td>
</tr>
<tr>
<td>Dilithium-3</td>
<td>1952</td>
<td>3293</td>
</tr>
<tr>
<td>Dilithium-5</td>
<td>2592</td>
<td>4595</td>
</tr>
</tbody>
</table>

modulo a known integer. This problem can be used as the basis for a signature scheme, as shown by Lyubashevsky [34], by improving on his idea to apply Fiat-Shamir with aborts [33] to lattices. A more specialized version is based on the Ring Learning With Errors (R-LWE) problem [36,43], which works in a special ring (more specifically the ring of integers of a cyclotomic number field) that offers significant storage and efficiency improvements compared to LWE. Although R-LWE has additional algebraic structure and relies on the (worst-case) hardness of problems in ideal lattices, no significant concrete improvements in cryptanalysis are known. Finally, a combination of many of these ideas (plus various improvements) resulted in CRYSTALS–Dilithium [14] based on something which is known as Module-LWE. Table 1 gives an overview of the public-key and signature sizes for the three Dilithium parameter sets.

An alternative direction to realize post-quantum secure digital signatures is that of hash-based signatures. Two algorithms from this realm are known as extended Merkle Signature Scheme (XMSS) and Leighton-Micali Signatures [38] (LMS) and have been established as Requests For Comments (RFCs) by the Internet Engineering Task Force (IETF) [24,38]. Recently, they have been published as a NIST Special Publication 800-208 [12] as well. These approaches are based on well-established cryptographic primitives (i.e., hash functions) and have efficient signature verification times. The main downside is that both signature schemes are stateful on the end of the signer, which can seriously complicate key management. Their stateless counterpart SPHINCS+ [22] solves this issue, but at the cost of significantly increasing the signature size. Other alternatives explore digital signatures based on the hardness of multivariate quadratic equations, error-correcting codes and isogenies. Unfortunately either their keys or signatures are extremely large, or they are relatively inefficient, making them difficult to apply in embedded scenarios.

In June 2023 a new standardization process was started by NIST with the focus on small digital signature schemes. At the time of writing the submission deadline had passed, but the proposals had not been made public yet. This could lead to new, smaller signature schemes that could outperform Dilithium or hash-based signatures. Having that said, it will take a number of years for the proposals to be analyzed and finally published into a standard.

2.2 PQClean

The NIST standardization effort required inclusion of reference implementations in pure C, to demonstrate the correctness and efficiency of the proposed algo-
rithms. As there were no requirements with respect to software engineering standards, the code quality varied wildly and was often not directly fit for integration into higher level protocols. This observation led Kannwischer, Schwabe, Stebila and Wiggers [30] to develop PQClean, a collection of clean implementations of the NIST PQC proposals in pure C. For example, any implementation included in PQClean should check that code is valid in C99, passes functional tests, does not write outside provided buffers, etc. For a full list of requirements we refer to the Github repository.

Besides the clean implementations, PQClean also includes optimized implementations for targeted architectures depending on the algorithm. It includes optimized Dilithium implementations for modern processors with support for AVX2, and for the 64-bit Arm architecture family AArch64. The latter is part of the ARMv8-A instruction set that is implemented in the Cortex-A53. We use this optimized implementation to run Dilithium on the S32G host processors.

2.3 AUTOSAR

The AUTomotive Open System ARchitecture (AUTOSAR [6]) is a consortium of automotive companies and other interested parties to specify a harmonized architecture and API for automotive middleware vendors to implement. These middlewares offer typical services to automotive applications such as communication, logging, and diagnostics. AUTOSAR specifies two architectures for different kind of Electronic Control Unit (ECUs): the Classic Platform to support applications with real-time requirements, and the Adaptive Platform for high-performance applications running on a POSIX-compatible operating system such as Linux or QNX [1]. Both Classic and Adaptive Platforms are designed to interact in an in-vehicle network.

For the Adaptive Platform, the AUTOSAR partners are developing a common Adaptive AUTOSAR Demonstrator (APD) to validate the specifications. This demonstrator can be used by partners for their own proof of concept developments. We used this APD as basis for our practical evaluation later.

2.4 S32G Vehicle Network Processors

We target an S32G vehicle network processor as the platform of choice for the impact assessment of integrating post-quantum cryptography in the over-the-air update protocol. This high-end automotive processor is developed by NXP Semiconductors and part of a larger S32 automotive platform which includes the S32K, S32R and S32V and is designed to meet the safety and security requirements in the automotive and industrial domains (i.e., compliance with IEC 61508 [25] and ASIL-D classification in ISO 26262 [26]). Typical uses include service-oriented gateways, domain controllers, vehicle computers and safety processors. The S32G consists of a combination of microcontrollers (MCUs) based
on the Arm Cortex-M7, and microprocessors (MPUs) based on Arm Cortex-A53. These application CPUs are combined with several types of memory (SRAM, DRAM, NOR/NAND Flash) and various hardware accelerators.

The precise configuration depends on the choice of model: we deploy the S32G274A which contains 3 Arm Cortex-M7 cores, 4 Arm Cortex-A53 cores, and 8 MB of system RAM. Each of the MCUs runs in a delayed lockstep configuration at a maximum frequency of 400 MHz and has 32 KB instruction and data caches. The MPUs are configured as 2 clusters of 2 cores each running at a maximum frequency of 1 GHz. Every core has access to 32 KB L1 instruction and data caches, while each cluster shares another 512 KB of L2 cache. Optionally, the A53 clusters can be configured to also run in a delayed lockstep setting, effectively removing one of the clusters from an application’s point of view but increasing the fault tolerance.

Most notably, the processor contains a Hardware Security Engine (HSE) which supports both symmetric and (classical) asymmetric cryptography accelerators, a random number generator, and dedicated secure memory. The HSE is responsible for the boot flow if secure boot is enabled as well as serving as the Root of Trust (RoT) for host applications: in our setting computing the post-quantum secure signature verification. The HSE Firmware (HSE-FW) has been extended by Bos et al. [9, Section 4.1] to include Dilithium into the signature verification service. This is a high-security low-memory software implementation which has protection against fault attacks, and supports all parameter sets of Dilithium v3.1.

2.5 Related work

A generic approach to reduce the memory consumption of Dilithium on constrained devices is presented by Bos, Renes, and Sprenkels in [10]. In the automotive domain there have been investigations on applying post-quantum cryptography to the setting of secure boot. An impact assessment of hash-based post-quantum secure schemes on secure boot is studied by Kampanakis, Panburana, Curcio and Shroff [27]. Hermelink, Pöppelmann, Stöttinger, Wang and Wan perform an investigation into Authenticated Key Exchange (AKE) combining different post-quantum cryptographic schemes [20]. In [10], Feritzmann, Vith, Florez and Sepúlveda analyze lattice-based Key Encapsulation Mechanisms (KEMs) for automotive systems. An investigation of the practical impact of migrating the secure boot flow on a Vehicle Network Processor using Dilithium is done by Bos, Carlson, Renes, Rotaru, Sprenkels, and Waters in [9].

The other use-case is secure over-the-air software update. A survey on this topic in connected vehicles is performed by Halder, Ghosal, and Conti [18]. Some benchmark results using the post-quantum secure schemes Dilithium and Falcon in the setting of OTA update on an Arm Cortex-A53 are presented by Manna, Perazzo, Treccozzi, and Dini in [37]. In [3], Banegas, Zandberg, Baccelli, Herrmann, and Smith investigate low-power software update with the Software Updates for Internet of Things (SUIT) specification focusing on low-power IoT devices.
3 Over-the-Air Update

Over-the-air updates include services like software over-the-air (SOTA), firmware over-the-air (FOTA), and over-the-air provisioning (OTAP). To implement this, the AUTOSAR Adaptive Platform offers the specification of the functional cluster “Update and Configuration Management” (UCM) which is responsible for handling the vehicle side of the update process. This includes the following steps of the components shown in Figure 1:

1. Download the update package by the UCM client;
2. Check the package’s authentication tag (e.g., digital signature);
3. Evaluate the package’s manifest;
4. Interact with state management;
5. Apply the update.

In this section we describe the considerations we have taken for the implementation.

3.1 OTA Update Security

The verification of the authentication tag (i.e., step 2) is the most important step for our work. There are several ways in which a malicious entity can subvert the integrity of the package. Firstly, they can attempt to obtain a digital signature over a malicious package, which would pass the authentication check in step 2. This could be done by compromising the signing infrastructure and having direct access to the (secret) signing key, but for the purpose of this work such attacks are considered out of scope as they cannot be thwarted by countermeasures on the processor itself. Alternatively, the signing key can be retrieved by breaking the public-key cryptographic scheme that is deployed. This attack is a realistic scenario for a quantum adversary, as it allows them to efficiently obtain the signing key for classical systems such as RSA or ECDSA. Indeed, for this reason timely migration to post-quantum alternatives is critical. In this work we deal with this attack vector by relying on the quantum security of Dilithium.

Assuming that no signature can be created over a malicious package, a malicious entity could instead try to compromise the integrity of the public key. In that case the verification in step 2 would pass, but would verify authenticity with respect to an attacker controlled public key for which they can sign arbitrary packages. To trust the authenticity of update packages, strong protection on the public key is therefore required. For example, this can be achieved by storing the (hash of the) public key in Read-Only Memory (ROM) such as fuses.

Finally, the runtime execution of the verification of the authentication tag could be compromised. It is well known that such advanced security threats should be considered in the automotive domain [32]. One category of such a relevant security threat for signature verification are active attacks. An example of such an attack is fault analysis [8] and more specifically differential fault analysis (DFA) [7]. Such types of attacks work by utilizing under-powering and power
spikes, clock gitches, temperature attacks, optical attacks or electromagnetic injection to introduce faults during the execution. If some operations are skipped or performed incorrectly during the execution process, then this should lead to an incorrect calculation of the output which could be used to learn something about the secret key material used. When using signature verification no secret key material is used. However, a simple attack would be to fault the control if-statement if a signature is correct to always yield true. This would lead to unauthorized updates with invalid signatures to be installed on the target device. In the setting of Dilithium a survey of both passive and active attacks is given in [46].

We summarize an overview of the threats in Table 2. Although there are different ways to achieve security against these attack vectors, ultimately it boils down to a well-defined Root of Trust (RoT). The expected capabilities of RoTs are outlined for example by TCG [11, Part 1 §9.5.5] or GlobalPlatform [17]. Its implementation has to conform with the latest security standards and certifications (e.g., ISO 26262). Typically, the RoT comes in the form of a Hardware Security Module (HSM) or Hardware Security Engine (HSE).

### 3.2 OTA Update Implementation

As shown in Figure 1, cryptographic algorithms can be implemented in ECUs in two ways: in an HSM or in a software library executed directly on a host application core. While the HSM approach generally offers better protection mechanisms especially against physical attacks (see Section 3.1), it is also typically less flexible than executing the software on the host processor. Which approach to choose depends on performance requirements and the risk treatment decisions based on the Threat Analysis and Risk Assessment (TARA).

To implement support for digital signature verification, UCM can use the services offered by the CryptoAPI, which abstracts interaction with concrete cryptographic implementations by means of Crypto Providers. Crypto Providers are secure wrappers of cryptographic software libraries or secure element host drivers. They serve two purposes: standardizing user interaction and securing
crypto objects such as secret key material, secret seed material or special public keys. Ideally, to comply with the AUTOSAR architecture, we would add a new Crypto Provider for the HSE Linux host driver. Unfortunately, the Adaptive Platform Demonstrator does not feature a sufficiently up-to-date reference implementation of the CryptoAPI. Therefore, we decided to integrate the PQC-enhanced HSE Linux host driver with UCM and use low-level driver interfaces directly to perform the signature verification.

3.3 OTA Update Performance Requirements

It is clear that performance is critical for Over-the-Air updates. However, the performance issues do not come from executing the cryptographic operations but arise when scheduling OTA software updates at a massive scale over a cellular network (like 4G LTE to 5G). This becomes clear from the standardization survey for over-the-air updating in vehicles [19] where no standards mention specific latency requirements for the signature verification but the challenges of deploying large scale updates over the cellular network is mentioned over-and-over again. Security, on the other hand, is highlighted at many places. This is of course not a surprise since safety and security are key requirements in an update protocol: this reduces, for example, the risk to remotely install malware in the vehicle. A survey on security attacks and defense techniques for connected and autonomous vehicles can be found in [44].

In short, a (small) performance loss when migrating towards post-quantum cryptography is not expected to have an impact on the end user experience as it will be negligible in the overall update time. As we shall see in Section 4, this holds true for both the implementation in the HSM/HSE as well as on the host processor. From a security point of view, it therefore makes the HSE the preferred way of executing the signature verification. The host processor would be the preferred option only if the additional flexibility is more important than the loss in security guarantees.
4 Benchmark results and Discussion

In this section we elaborate on the practical details of the PQC secure OTA demonstrator, and discuss the pros and cons of the different options.

4.1 Development Setup

For development we used the S32G Vehicle Networking Evaluation Board (VNP-EVB) and Vehicle Networking Reference Design (VNP-RDB2). Both function virtually identically for our purposes, the latter being the more recent and recommended development board (though now already superseded by the RDB3 board). These boards come equipped with a Linux Board Support Package (BSP) for the Cortex-A53 MPUs that we used as a basis for integrating Dilithium into our software. The BSP provides Linux User IO (UIO) drivers and a messaging module (libhse) that provides direct low-level access to the HSE. See Figure 2 for an overview of the S32G2 software ecosystem. The use of Dilithium signature verification is almost completely transparent for a user: as is the case for ECC or RSA, a message, public key and signature are provided to a signature service in which the signing scheme can be specified. As part of the integration of Dilithium into the firmware, the signing scheme can simply be set to Dilithium instead of the classical choices of ECC or RSA. We refer to Figure 3 for the control flow when executing Dilithium signature verification on the HSE.

4.2 Executing the Cryptographic Algorithms

As explained in Section 2.4 our platform contains a microprocessor based on the Arm Cortex-A53 running at 1000 MHz which we will refer to as the host. The
Fig. 3. Overview of delivering data to the HSE in order to execute its functionality.

Hardware Security Engine utilizes a microcontroller based on the Arm Cortex-M7 running at 400 MHz. As the Cortex-A53 has a much more extensive architecture than the Cortex-M7, it is no surprise that most software can be executed significantly faster. This is especially the case as it runs at a much higher frequency. The difference in performance can typically be offset by additional cryptographic hardware support in the HSE, but this is not yet available for the S32G2. For future platforms (e.g., S32G3 with SHA3 acceleration), the difference between the host and the HSE will favor the HSE much more.

4.3 Results

It is clear from Section 3.3 that the execution time of the signature verification algorithm typically is not the bottleneck in the setting of OTA update for automotive applications. However, to assess the impact of the ongoing migration towards post-quantum cryptography we quantify the impact of switching to the usage of Dilithium-2 signature verification. This impact is measured in the form of the latency of the verification algorithm.

Remark 1. The latency is not solely determined by the choice of signature scheme itself, but also by the length of the software update. Typically, all digital signature verification schemes pre-hash the variable-length input to a fixed-size digest, possibly including miscellaneous other data such as padding, a public key, a commitment, etc. Afterwards the digest is processed to create or check the final digital signature. This step is independent of the signature scheme chosen but the exact the choice of the hash function does differ in practice. For example, the ECDSA signature scheme starts by applying a FIPS 180 compliant hash function (e.g., SHA-256) to the input message. Dilithium signs (and verifies) arbitrary-length messages by hashing them together with the public key, using Keccak (the permutation underlying SHA-3). The choice of hash
function has no impact on the security of the public-key signature scheme, it can, however, have significant impact on the performance in practice. For example, the S32G274A used offers hardware support for SHA-2 and not for (variants of) SHA-3. Therefore we focus on the setting where a SHA-512 hash over the software is signed instead and subsequently verified.

We adjusted the OTA update protocol such that it uses the post-quantum secure Dilithium algorithm. For our experiments we utilized the Dilithium-2 parameter set: this means a public-key of 1 312 bytes and verifying a signature of 2 420 bytes (see Table 1). The 64-byte pre-hashed software update together with the signature is validated using the Dilithium signature verification algorithm. As explained in Section 3.2 two approaches are considered: running the verification on the host processor itself and executing the Dilithium verification on the HSE. Our benchmark figures include all overhead required to run the signature verification: i.e., the timer starts and stops at the host processor before and after the complete signature verification. All together, the Dilithium-2 verification time is 0.5ms on the host processor and 11.1ms on the HSE.

4.4 Discussion

As seen from Section 4.3, the latency on the host processor is lower than on the HSE. This difference in performance is explained by the fact that the host processor and the HSE differ significantly in terms of the offered performance: e.g., the instruction set as well as the clock speed of both platforms differ significantly. For example, if we compare the Dilithium-2 verification results on a Cortex-M7 by [21] to the ones from [4] on a Cortex-A72 then one already sees a difference of 5–6 times. Hence, if we also take the difference in clock-speed into account (another factor of $2.5 \times$) one expect that the host processor is around 14 times faster. However, we stress again that the host processor is unprotected against advanced attacks relevant in the automotive domain (see Section 3.1). The HSE, on the other hand, offers much more security features compared to the host processor. This includes protection against side-channel and fault attacks. The unprotected implementation in [21] requires 1 439 kilo cycles, measured on STM32F767ZI NUCLEO-144 development board. At 200 MHz this corresponds to about 7.1ms. This is faster than our implementation, but can be explained by the fact that the HSE includes fault protection and applies more aggressive time-memory trade-offs that impact performance further. Whereas the runtime RAM memory required in [21] is about 36 kB for Dilithium-2, the HSE requires less than 3 kB.

Looking ahead, much of the 11.1ms in the HSE signature verification is spent in the software implementation of the SHAKE-256 algorithm used in Dilithium verification. This does not come as a surprise as the embedded benchmarking platform PQM4 [28,29] (running on a ARM Cortex-M4) reports that over 80 percent of the verification time is spent in SHA3. Future generations (e.g., S32G3) will have dedicated SHA3 hardware available, pushing the performance to be
very close to the host processor. Nevertheless, in the context of a full over-the-air software update, a difference of 10ms in verification latency is negligible and will have no impact on the experience of the vehicle owner. Therefore the security features offered by the HSE significantly outweigh the small difference in performance.

References


