Perspectives

CARAF: Crypto Agility Risk Assessment Framework

Chujiao Ma, Luis Colon, Joe Dera, Bahman Rashidi, and Vaibhav Garg

Comcast CyberSecurity, Comcast Cable, 1800 Arch Street, Philadelphia, PA 19103, USA

*Correspondence address. Comcast CyberSecurity, Comcast Cable, 1800 Arch Street, Philadelphia, PA 19103, USA. Email: vaibhav_garg@comcast.com

Received 2 September 2020; revised 19 February 2021; accepted 30 April 2021

Abstract

Crypto agility refers to the ability of an entity to replace existing crypto primitives, algorithms, or protocols with a new alternative quickly, inexpensively, with no or acceptable risk exposure. These changes may be driven by regulatory action, advances in computing, or newly discovered vulnerabilities. Yet everyday operational needs may put crypto agility considerations on the back burner when deploying technology, designing processes, or developing products/services. Consequently, changes are often performed in an ad hoc manner. Transition from one crypto solution to another can then take a long time and expose organizations to unnecessary security risk. This paper presents a framework to analyze and evaluate the risk that results from the lack of crypto agility. The proposed framework can be used by organizations to determine an appropriate mitigation strategy commensurate with their risk tolerance. We demonstrate the application of this framework with a case study of quantum computing and related threats to cryptography in the context of TLS for Internet of Things.

Key words: agility; cryptography; risk assessment; quantum; IoT

Introduction

Enigma is one of the most well-known encryption systems in the world. At the time of its creation, it was considered as the strongest encryption system around [1]. It was only through the ingenuity of scientists from Poland, France, Britain, and others that Enigma was broken. Although the Allied forces took great care to keep this a secret, some historians have argued that Nazis may have known that Enigma was not secure [2, 3]. If so, why did the Nazis then not switch to a different encryption system? One potential answer lies in crypto agility or lack thereof. For starters, the Nazis would have needed a better alternative to Enigma. Furthermore, there were operational constraints. The staff, including those deployed on the front, would have to be trained in the use of a new algorithm and related hardware. They would also have to be issued new code books. Given that the Nazi forces were deployed all the way from Russia to France, this would have been extremely difficult. Thus, though in theory the Nazis could have switched to a different encryption solution, in practice this would have been quite difficult.

The ability to replace crypto primitives, algorithms, or protocols with limited impact on operations and with low overhead, such as costs, is referred to as crypto agility. Although most modern organizations rely on a swath of cryptography from RSA to AES, few have considered the risk of not accounting for crypto agility. Yet advances in crypto-analysis often result in the discovery of vulnerabilities in older cryptography [4, 5]. In addition, legal or regulatory mandates may require use of specific cryptography. Advances in computing, such as quantum computers, may necessitate switching to entirely different suites of algorithms with fundamentally different mathematical foundations [6].

Thus, crypto agility must be considered a business risk like any other, e.g. compliance and supply chain. How can organizations evaluate and mitigate this risk, i.e. the risk from crypto agility or lack thereof? This article aims to address this question. Specifically, we propose a 5D Crypto Agility Risk Assessment Framework (CARAF):

- First, organizations must determine the specific threat vector that is driving the crypto agility risk assessment.

The Author(s) 2021. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
• Second, they should identify the assets impacted by that threat vector.
• Third, they should evaluate the expected value of impacted assets being compromised.
• Fourth, they should identify the appropriate mitigation strategy based on the expected value of the compromised asset.
• Lastly, they should develop a roadmap that outlines how to implement the distinct mitigation strategies for the different classes of assets differentiated by risk.

The contribution of this work is a framework to approach security and risk management in a proactive manner instead of reactive. This targets future risks where an empirical approach to decision making will be constrained by limited or absent data. We begin by presenting background and related work in the “Background and Related Work” section. In the “Crypto Agility Risk Assessment Framework” section, we describe our CARAF. The “Case Study: Quantum Computing” section discusses the application of the framework with a case study of quantum computing in the context of TLS for Internet of Things (IoT). Finally, the “Conclusion” section concludes the article.

Background and Related Work

Crypto agility: a historical perspective

The need for crypto agility is well established. Vulnerabilities in older popular crypto systems have often created the need to switch to newer more secure alternatives. One example is the migration from SHA-1 to SHA-2. NIST banned all US federal agencies from using SHA-1 in 2010, and digital certificate authorities have not been allowed to issue SHA-1 certificates since 2016 [7]. Although most browsers will display an error message when encountering a SHA-1 certificate on the website, some let you bypass the error until much later.

Crypto agility can be difficult to implement without creating additional security exposure. TLS or SSL is an example of a protocol with some agility built in. TLS establishes an encrypted connection between a server and client using certificates with asymmetric and symmetric keys. It has built-in support for a number of ciphers which can be used optionally or interchangeably. This agility can, however, be used for other classes of cyberattacks. In BEAST [8] and CRIME [9], e.g. attackers were able to take advantage of the protocol’s built-in agility to switch to an insecure cipher. This also means that the TLS protocol is only as secure as the cipher that the client chooses to use. It is necessary to keep track of insecure ciphers but removing them proves to be difficult due to the need for backward compatibility. This results in fallback attacks, a recent example being POODLE [10].

Yet making the algorithm support more rigid to a handful of specified algorithms and key sizes can be difficult or result in operational challenges. Implementation problems pop up when the changes vary too greatly between versions. TLS v1.3, first published in 2014, has gone through dozens of revisions. However, different revisions cannot communicate with each other and it faces relatively high-failure rate for middleboxes. Middleboxes are network appliances that monitor and sometimes intercept traffic, and they are blocking traffic they do not understand, such as TLS v1.3. When presented with TLS 1.3, a large number of servers would disconnect instead of negotiating and reply with v1.2 [11]. Much effort was made to make initial communication for v1.3 look like v1.2 before TLS v1.3 was finally finalized on 21 March 2018 [12].

Transition in Internet infrastructure is particularly difficult as digital signatures on certificates maybe expected to last decades. Once a particular signature algorithm is used to issue a long-lived certificate, it will be used by many relying parties and none of them can stop supporting it without invalidating all of the subordinate certificates [13]. In addition, due to the inability of legacy systems as well as resource constrained devices to support new algorithms, it has proven difficult to remove or disable old weakened algorithms. Thus, despite knowing that all algorithms will eventually become obsolete, migration is often a long and difficult process.

Drivers for crypto agility

New technology—quantum

Disruption from new technology, such as quantum computing, can compromise the security of cryptographic algorithms. For example, Shor’s algorithm can be used to violate the assumptions underlying most widely deployed public key crypto systems [14]. Simultaneously, Grover’s search algorithm provides a quadratic speedup on unstructured search problems, which affect the computational security of the symmetric key crypto systems and hash functions [15]. Table 1 provides a summary of the impact of large-scale quantum computers on common crypto algorithms [6]. Thus, a large enough quantum computer may require a transition that will be informed by the crypto agility of the impacted assets. Public key crypto systems will have to be replaced with quantum safe alternatives, while symmetric key solutions may require doubling key sizes to provide the same level of security. The former may be more challenging as it will require changes in both the impacted assets and the backend support infrastructure.

Algorithmic and operational vulnerabilities

In October 2017, the ROCA vulnerability was discovered in a software library implementing RSA, which impacted billions of security devices and smartcards [16]. In 2017, the improper issuance of SSL certificates from Symantec allowed malicious actors to setup corporate shells and phishing sites. In 2018, all existing Symantec SSL certificates were blocked by Google Chrome, and Symantec had to re-issue all its certificates [17]. In July 2018, NIST proposed a 5-year timeline to disallow use of the 3DES algorithms [18].

These examples illustrate the potential for discovering vulnerabilities in existing crypto systems and their operational deployments. Such discoveries require organizations to respond with appropriate mitigation actions. Lack of crypto agility may impede an organization’s ability to respond with adequate velocity, especially in the absence of responsible disclosure, such as by malicious actors.

Legal, regulatory, and ethics

Legal and regulatory mandates are one of the less visible drivers of crypto agility. Ideally, any public policy efforts in cybersecurity would be technology neutral. In practice, this is not always the case. For example, some governments may feel the need to manage what crypto systems are used within their jurisdiction. The Chinese government mandates SM(x) class of crypto algorithms for their vendors. In USA, FCC is working with telecommunications companies to implement STIR/SHAKEN protocols to address the issue of robocalling [19].

In addition to specific public entities preferring certain types of cryptography, there might also be requirements for creating lawful access mechanisms [20]. Arguably the most popular example of this is the Clipper chip, which leveraged the concept of key escrow [21].
New technologies such as quantum computing may require rethinking the management of the current crypto systems [29]. The security challenges of these systems are beyond the scope of this article and are available elsewhere [23–25]. However, we must note that implementing systems like Clear will require re-engineering the backend systems and therefore impinge on crypto agility.

Crypto agility solutions

Crypto agility can be facilitated with the adoption of a service software layer, or a gateway application, between applications and hardware security modules. Senetas CN series hardware encryptors focus on hardware agility by providing a flexible Field Programmable Gate Arrays architecture that enables in-field upgrades [26]. However, this requires dedicated and proprietary hardware and only works for network encryption. Cryptomathic Crypto Service Gateway 3.10 is a module that can be implemented on existing products [27]. It provides a cryptographic control center that acts as a Hardware Security Module service and a crypto policy management interface. However, this still requires the final end-user applications or endpoints to be able to support the appropriate keys and relevant algorithms. InfoSec Global AgileSec is a multicrypto platform security system. It consists of a cryptographic toolkit at endpoints and a management server infrastructure, which remotely deploys policy and sets it for cryptography across a diverse set of remote software and devices [28].

New technologies such as quantum computing may require switching to a completely new set of algorithms instead of better management of the current crypto systems [29]. NIST, e.g. is still in the process of identifying quantum-safe replacements for current public key algorithms [6, 30]. In the absence of established quantum safe alternatives to classical public key systems, the solution may be to deploy hybrid solutions. These will in theory remain secure if at least one of the underlying cryptographic schemes remains unbroken. However, they can be slower, have a larger footprint for key storage, and be less efficient [31]. Concurrently, there is also a collaborative effort on digital certificates compatible with both classical and quantum-safe cryptographic algorithms from ISARA, Cisco, CableLabs, and DigiCert. Users will be able to download a hybrid root certificate and request a hybrid end entity certificate, then connect to the TLS server using the hybrid certificate such that either classical or quantum-safe cipher suites are used for digital signature.

In an enterprise setting, consideration must be given to the cryptography as well as key management, policy enforcement, monitoring, usability, and updates. In order for a system to be crypto agile, all subcomponents of the system will need to be crypto agile as well. Businesses not only have to consider the technical aspect but also the implementation as well. For example, updating the encryption algorithm for a database requires not only a change in algorithms but also re-encryption to prevent data loss. Whereas if a database is hashed, it may not be possible to simply upgrade to a new hashing algorithm, but instead use both the old and the new algorithm in serial, as the existing data cannot be rehashed. Some technical solutions may not be practical to implement due to scalability concerns. Alternatively, some devices may not be capable of updates. Thus, any effort to increase crypto agility must begin with a risk assessment that takes a holistic view that includes consider controls, operational feasibility and capability, third-party vendors and incident response plans for mitigation.

Risk assessment frameworks

Regardless of the availability of technical or cryptographic solutions to crypto agility, any transition will be inherently expensive, with additional overhead, and thus requires careful planning. One solution is to use risk assessment to determine the optimum allocation of resources to ensure minimal exposure to risk. At its very core, the expected value of a risk is a product of probability of the risk materializing and the cost of impact. Different risk assessment frameworks vary in how they operationalize this concept. Risk assessments can be qualitative or quantitative. Quantitative assessments use equations to measure the risk in terms of definite numbers, such as estimated cost of assets, percentage of assets compromised, and cost of mitigation. Qualitative assessments use surveys or interviews to engage relevant parties for insights. The appropriateness of different risk assessment methodologies may be driven by the nature and size of the business, regulatory landscape, best practices, etc.

NIST SP 800-30 is a seminal and popular framework for assessing IT risk including security risk [32]. It is clearly structured when it comes to planning and implementation. However, the NIST framework focuses on assessing the risk of technology and there is no asset identification or consideration for controls needed for organizational risk assessment. In contrast to NIST SP 800-30, which is US focused, ISO/IEC 27005 is an international information security standard published by the International Organization for Standardization [33]. It places security in the context of the overall management and processes of a company. Although ISO is influenced by NIST, it allows different computational methods to calculate risk and covers technology, people and process, thus providing a more holistic picture. Distinct from NIST and ISO is OCTAVE, which is self-directed and customizable. It approaches security risks from an operational and organizational view and addresses technology in a business context [34]. ISO 27005 and OCTAVE are focused on information security while NIST 800-30 is broader and can be applied to systems, applications, or information. OCTAVE also does not produce a quantitative measure of the risks. Thus, even within the same risk domain there can be very different approaches to conducting risk assessments.

A well-known risk assessment framework for cybersecurity is NIST Cybersecurity Framework or CSF [35]. It is a 5D framework: (i) identify, (ii) protect, (iii) detect, (iv) respond, and (v) recover. Risk assessment should begin with an inventory of assets. Once assets have been identified, the organization must deploy security controls to protect them. Next, organizations must deploy tools to detect any attacks on identified assets. In case of an attack, the

### Table 1: Impact of quantum computing on common cryptographic algorithms

| Algorithm       | Type               | Purpose                  | Impact                  |
|-----------------|--------------------|--------------------------|-------------------------|
| AES             | Symmetric key      | Encryption               | Larger key sizes        |
| SHA-2, SHA-3    | Hash               | Hash functions           | Larger output           |
| RSA             | Public key         | Signatures, key establisement | No longer secure |
| ECC             | Public key         | Signatures, key exchange | No longer secure        |
| DSA finite field| Public key         | Signatures, key exchange | No longer secure        |

More recently, Ray Ozzie proposed a four-step process called Clear [22]. The security challenges of these systems are beyond the scope of this article and are available elsewhere [23–25]. However, we must note that implementing systems like Clear will require re-engineering the backend systems and therefore impinge on crypto agility.
enterprise must respond with appropriate mitigation actions. If the response is not adequate and the asset is compromised, the final step is to have a recovery mechanism in place to bring the asset back online. NIST’s CSF, as with most risk assessment methodologies, is meant to address known threats while crypto agility is more forward looking and meant to prepare organizations for eventual change.

In contrast, Mosca’s XYZ quantum risk model determines when it is time to prepare for quantum threats [36]. X refers to the duration that information should be kept secure. Y refers to the time needed to migrate to quantum-safe solution. Z is the estimate on when identified threat actors will have access to quantum technology. If X + Y > Z, then there is a lack or deficiency of security. The Mosca’s risk assessment is separated into six phases. Phase 1 is to identify information assets and their current cryptographic protection. Phase 2 is to research the state and estimate the timelines for availability of quantum computers and quantum-safe cryptography. Because the technology is still new and constantly changing, what is the most risky today may be different tomorrow, and continuous monitoring is needed. Phase 3 is to identify the threat actors, then estimate their time to access quantum technology and likelihood of exploits. Phase 4 is to identify the organization’s quantum vulnerability using the lifetime of the assets and time required for updates or migration. Phase 5 is to determine the quantum risk by calculating whether the business assets will become vulnerable before the organization can move to protect them. Phase 6 is to identify and prioritize the activities required to maintain awareness with roadmap or plan for migration.

Mosca’s model, though future looking, focuses on quantum and does not explicitly address crypto agility or provide guidelines on how to assess or address risk aside from a general timeline. For example, how should organizations prepare to respond to mass certificate and key replacement events? Simultaneously, how should they demonstrate continued policy compliance for all certificates (or document exceptions)? The lack of a formal framework makes it difficult for practitioners to have a common taxonomy, across different organizations with distinct business models, to reason about crypto agility risks. Whereas a formal framework will facilitate informed decision making to accept, mitigate, or reject the risk from lack of crypto agility as well as plan to address the risk when appropriate. In the next section we introduce a framework that satisfies these needs.

Crypto Agility Risk Assessment Framework
The “Background and Related Work” section discussed the difficulties in implementing crypto agility as well as the challenges that emerge from the lack thereof. These difficulties are exacerbated if crypto agility is approached by enterprises in an ad hoc manner without regard to the underlying technology, compensating controls, and lifecycle management. Decisions regarding crypto agility should then consider it a business risk and address exposure based on a comprehensive risk assessment. In this section, we present a 5D (or phase) framework to support this assessment, referred to as the CARAF.

Phase 1: identify threats
We begin by identifying the threat that a CARAF-based assessment aims to address. This differentiates CARAF from other risk frameworks and allows assessors to discount the assets that will not be impacted by the threat in question. For example, if assets are likely to be phased out before the need for crypto transition they can be considered out of scope. Similarly, if the threat only impacts software assets, hardware assets can be considered out of scope. Assessors can then explicitly address assets that are impacted by the threat. This enables a more optimized and realistic assessment framework, especially as most organizations have a wide variety of assets and exhaustive inventories are rare.

CARAF aims to address a probable future security threat, allowing the organization to be proactive instead of reactive. As the threat is in the future, there may not be enough information on possible risk vectors to accurately identify impact, likelihood, or exposure. For example, consider the threat of quantum computing. NIST posits that a quantum computer capable of breaking 2000-bit RSA in a matter of hours could be built by 2030 for a budget of about a billion dollars [6]. Others hypothesize that there is a 15% chance that RSA and ECC will be broken by 2026, with a 50% chance by 2031 [37]. The likelihood and timeline for practical quantum computers may change due to new research and requires continuous evaluation. There are physical engineering challenges that need to be worked out as well, such as the limited number of qubits available [38]. Thus, it is unclear when a large enough quantum computer with the ability to factor RSA will materialize [6].

Furthermore, depending on the nature of the threat not all assets may be similarly impacted. For example, a large quantum computer will impact public key crypto algorithms more severely than symmetric key algorithms. It may be adequate to just double the key size for symmetric key algorithms, but public key algorithms will need to be replaced with quantum-safe alternatives, which will necessitate a greater change management effort.

Finally, depending on the category of the threat, the impact will be different:

- Regulatory requirements, both voluntary or otherwise, from governments, are inevitable. These are usually accompanied with timelines for transition as well as guidance for impacted parties. Thus, it may be easier for organizations to plan a response for this threat vector.
- Newly discovered vulnerabilities are, by their nature, unexpected and may impact mission critical applications. However, responsible disclosure can help organizations plan appropriately. In addition, they can learn from existing case studies of prior transitions, e.g. SHA1 to SHA2. If new exploits are discovered for known vulnerabilities, mitigation may already exist, e.g. a patch for the vulnerable subcomponent or alternatively a compensating control.
- Disruption from new technology, such as quantum computing, is the most difficult to address, due to the lack of a concrete timeline for threat manifestation as well as prior instances of transition. This is also where a crypto agility risk assessment may be the most informative and is the focus of the case study in the “Case Study: Quantum Computing” section. Aside from quantum computing, new lightweight cryptography approaches may replace extant resource intensive alternatives to improve performance.

Thus, CARAF starts by identifying the threats or future risks. The next step is to inventory the impacted assets.

Phase 2: inventory of assets
The security risk exposure of distinct assets will differ based on the nature of the threat. Consider, e.g. PCI-DSS v3.1 that deprecated the use of all versions of SSL as well as TLS 1.0 and required a move to TLS 1.1 and beyond [39]. The corresponding organizational response should have focused on assets that process PCI data and use...
TLS 1.0 or SSL. Thus, once an organization determines the threat vector driving crypto agility the next step is to inventory a list of impacted assets. These assets refer to systems with independent crypto components that support either confidentiality, integrity, or availability. The specific scope will be determined by the organization and their use case. Thus, an IoT ecosystem with two smart light bulbs and a central hub can be simultaneously described as three distinct assets or one single asset.

For large enterprises, the number of impacted assets may be too large to address simultaneously. Assets can then be categorized and prioritized according to the nature of the assets and the expected security risk exposure. Specifically, organizations may consider and document the following factors when taking inventory:

- **Scope:** Any inventory must begin with identifying the appropriate scope, which will be determined by the nature of the threat recognized in Phase 1. For example, in the case of PCI driven threats any non-PCI systems can be considered out of scope. However, in some cases the scope is not clear cut due to the inter-dependencies of services and devices. Thus, any dependencies of the assets in question must also be considered when applicable.
- **Sensitivity:** Organizations must prioritize in-scope assets with higher expected risk. Thus, it is important to understand where and how the asset is used, what is the impact of the asset being compromised, i.e. loss of confidentiality or integrity, or lost, i.e. become unavailable.
- **Cryptography:** Organizations must determine the cryptographic solutions that are being used to secure the in-scope assets with adequate sensitivity. It is important to assess the security of the algorithms in use along with the appropriateness of related properties, such as key lengths.
- **Secrets management:** Inventories must also include information about the management of secrets related to individual crypto-solutions. These may include but are not limited to keys, passwords, API tokens, and certificates as well as the frequency of use and updates.
- **Implementation:** Inventories must also note how the crypto-solutions are implemented. A hardcoded solution or one based in hardware, such as a hardware root of trust, will be more difficult to address than a software-based alternative. Even for a software-based solution, ease of increasing security will be contingent on whether there is a system for automated management of server/trust/key stores as well as if it is possible to remotely update software with appropriate mutual authentication.
- **Ownership:** No inventory can be considered complete without information about asset ownership. For large companies or non-technological industries, the crypto assets may come from third-party vendors instead of (or in addition to) internal product teams. Cryptography as a Service is an emerging trend in cloud computing. Even traditional computing widely uses open source crypto-libraries, such as OpenSSL. Documenting these upstream and downstream dependencies is critical as one risk of third-party products used by an organization is the lack of adequate and timely updates; some components may even be end-of-life. Responsible owners would then have appropriate response plans.
- **Location:** The location of an asset will impact how crypto agile it is. For example, on premise assets may require a different update process compared to those in the cloud. Jurisdictional constraints may also impinge agility. China, e.g. regulates all internal use of cryptography whereas USA does not [40].

These factors will help organizations to determine the assets that need to be prioritized for risk mitigation from the threat determined in phase 1. In addition, it will help highlight the extent of knowledge gaps so the organizations can plan accordingly. For example, it is not uncommon for central asset ownership repositories to have missing or dated information. Even when ownership is known, some asset owners may be unclear on what cryptography is used and how keys are stored while others may have a detailed change management plan. Most importantly, a survey based on the factors detailed above will help organizations to assess how crypto agile their assets are and understand the challenges to mitigation. This will help organizations estimate the risk exposure (Phase 3), appropriate risk mitigation strategy (Phase 4), and finally, develop a roadmap to implement that strategy (Phase 5).

**Phase 3: risk estimation**

For a medium to large organization, even a well-scoped inventory will need to be prioritized for risk mitigation based on exposure. A generic formula for risk estimation is “Risk = Probability × Impact.” “Probability” refers to the likelihood or frequency of exposure. In cybersecurity, this is informed by factors such as a threat actor’s motivation and experience, or if there are any mitigation or controls in place. “Impact” refers to the consequence of a risk materializing. For cybersecurity in an enterprise context, this is often the cost to company if an asset is compromised, i.e. loss of confidentiality, integrity, or availability. In traditional risk-based decision domains, actuarial information like statistics or records of previous events are used to calculate probability and impact, thereby modeling risk.

There are challenges in applying traditional risk models to cybersecurity. Cyber-insurance models, e.g. have struggled with the lack of information about past incidents [41]. Lack of information about incidents is particularly challenging in the context of crypto agility as the goal is to model the risk of an event that has not yet materialized. For example, let’s consider the threat of quantum computing to current cryptosystems. It is not meaningful to consider the frequency of exposure to quantum computing.

Instead, we provide a more specialized case of the general formula with a different abstraction of probability of exposure, i.e. the time to exposure. For example, the probability of exposure to quantum computers may be 15% by 2026 and 50% by 2031 [42]. Thus, the time to exposure is a probability distribution, that can be reduced to discrete values for a risk assessment. Impact in this case is measured as the cost of updating an asset to a secure state within the required timeline. Cost will be determined by some of the factors documented in the inventory (Phase 2). A more crypto agile asset will be less expensive to migrate and therefore pose overall lower risk. Therefore, crypto agility risk is a function of the time to migrate and the cost to migrate, i.e. “Risk = Timeline × Cost.”

**Timeline**

The timeline to exposure builds on Mosca’s Model [36] by including the information from Phases 1 and 2:

- **X (Shelf-life)** refers to the remaining lifespan of the device or data during which they must be protected. It can be ascertained from lifecycle management information, which should include lifespan, end-of-life information, etc. For example, information...
assets with legal hold will have jurisdictional mandates on retention.

- **Y (Mitigation or remediation)** refers to the number of years needed to replace or upgrade the asset, or time needed for deletion of data and recall of devices if those assets are to be phased out. In addition to time required for migration and mitigation for security measures, the time needed to fix implementation and reliability or performance issues to ensure smooth operation should also be included. This will be informed by the cryptography used, secrets management, implementation, availability of ownership information, as well as location. For example, on premise assets may be remediated faster as those are within the control of the enterprise. In contrast, cloud-based assets may take longer to address and the enterprise will depend on the cloud provider to make the necessary changes.

- **Z (Threat)** refers to the number of years before the threat vector results in a compromise. Although timeline-rating for X and Y can be deduced from the factors recorded as part of the inventory of assets in Phase 2, Z is independent of the inventory and comes from the threat assessment in Phase 1. For example, if the threat is from new technologies, then Z will have to be adjusted to account for any advances in research that shortens or lengthens the time horizon of the threat materializing.

For a quantitative risk assessment, we score the three components between 1 and 4, or low risk to critical, respectively (Table 2). The values 1–3 are from low to high based on how soon the future threat may be realized with the value of 4 means the threat is already here. For example, a rating of 4 would indicate that a regulation that deprecates certain algorithms has already passed or a quantum computer large enough to factor RSA has already been constructed.

The ratings for each component can be averaged across assets with similar sensitivity to produce a timeline-risk score, which should align with one of the values within the weighted impact rating model. For example, if an X + Y > Z then the organization’s risk level will likely be a High (3) or Critical (4), which means the enterprise infrastructure will succumb to quantum attacks in ‘Z’ years. On the opposite spectrum, if an X + Y < Z equation is true, then the risk level is likely Low (1) or Medium (2), and the organization likely has time to mitigate risks.

### Cost

The cost of mitigating the risk will vary depending on the type of assets and availability of resources for each organization. However, the mitigation will be more cost effective for more crypto agile assets and by corollary organizations. Crypto agility depends on four design considerations [43]:

- Implementation independence: Code is independent from the cryptographic implementation and managed separately. For example, there are no hard coded dependencies on a specific algorithm.
- Simplicity: Management is centralized through user-friendly interface to reduce risk of usage errors with clear and easily understood guidelines.
- Flexibility: Platform allows plug-and-play installation of the different cryptographic modules.
- Performance: Crypto tasks, such as key generation or decryption, have limited impact on operational overhead.

These design considerations will help estimate the extent to which an asset is crypto agile. The cost of risk mitigation may be computed by referring to the information collected during inventory of assets:

- **Cryptography**: The type of cryptography is important. For example, if the information assets are being encrypted with one algorithm, it may be necessary to decrypt and then re-encrypt when upgrading to a more secure alternative. This may be an expensive and difficult exercise if the enterprise databases are spread out. Systems may have to be brought offline, adding to costs. Hashed data are more challenging to migrate than the encrypted data. It is considered best practice to salt and hash passwords. If the enterprise moves to different hashing algorithms, they may have to issue password resets for all impacted accounts. In the best case, this creates operational overhead. In the worst case, this is a self-inflicted Denial of Service attack.
- **Secrets management**: Changes in algorithms may also impact secrets management or supporting infrastructure, such as key management systems and certificate issuing systems. Errors in the process can cause challenges. Let’s Encrypt had to revoke and reissue approximately 3 million certificates due to a bug in the system. These challenges can become extremely expensive if the tokens being reissued are hardware-based. It is estimated the cost RSA $66 billion to reissue SecureID tokens. Although Let’s Encrypt and RSA’s costs were not due to a crypto agility driven change, arguably these mistakes were made under Business As Usual conditions. If threat vectors drive changes in cryptography, it is not unreasonable that similar mistakes can be made.
- **Implementation**: Whether the implementation is hardware or software based will also add to costs. Specter and meltdown have demonstrated the difficulties of remediating hardware-based vulnerabilities [44]. Even software patching may require system downtime as well as potentially bricking the system if the patch was not adequately tested.
- **Ownership**: If the ownership involves third-party vendors, it may be more expensive as they may not be contractually required to remediate risk. Thus, their decision to remediate will depend on their internal risk estimation, which may not align with that of the enterprise. In some cases, it may make more fiscal sense for the vendor to lose the contract rather than remediate risk. In this case, the enterprise will have to either pay the vendor more to provide a patch, onboard a new system from a different vendor (assuming there is an alternative), or develop/find a compensating control.
- **Location**: An organization’s ability to mitigate risk may be further hampered by location. For example, jurisdictions may inform risk mitigation costs. In USA, internal use of cryptography is not regulated. However, approval must be granted by a central governing entity in China. Any risk mitigation will then be constrained by regulatory approvals and may increase costs [40].

### Phase 4: secure assets through risk mitigation

There are typically three options for risk mitigation:

- Secure the asset by spending resources. This may be rational when the value of asset is greater than the cost to secure it. It can be achieved by upgrading the asset with a new crypto solution
that mitigates the risk. An alternative is to implement compensating controls to reduce the risk exposure, which may be the best option for legacy assets that cannot be upgraded to a secure state.

- Accept the risk and maintain status quo. This is reasonable when the expected value of the risk is lower than organization’s risk tolerance.
- Phase out impacted assets. This option may apply if the value of the asset is lower than the expected risk, especially if the cost to secure is high.

The appropriate risk mitigation strategy will thus depend on the organizational risk tolerance and the expected value of risk determined in Phase 3 as the function of Timeline and Cost. A simplistic risk mitigation strategy is documented in Table 3.

**Phase 5: organizational roadmap**

On the basis of risk mitigation strategy, organizations will need to develop a tactical roadmap to address crypto agility (or the risks from a lack of). The success of any roadmap will, however, depend on having some foundational constructs in place. The enterprise must have coherent crypto policy that supports and guides different teams in making decisions about their cryptography choices. This policy must be enforced with an appropriate Responsible, Accountable, Consulted and Informed matrix. The enterprise crypto policy must tie in to associated organizational processes. Some examples are:

- Incident Response Plan: These should include plans to address vulnerabilities in internally approved crypto-solutions.
- Third Party Risk Assessment: These should consider their responsiveness to vulnerability disclosure.
- Security Architecture Reviews: These should assess the crypto agility of the proposed architecture.
- Product Development: Development teams should be trained to select solutions with greater crypto agility. For example, when picking an open-source component, they should prioritize components with greater flexibility, better support, and faster patching history.
- Change Management Plan: This should include defined process for updates to cryptography and supporting systems, e.g. key management systems, hardware security modules, etc.

The processes and policies must be complemented with appropriate technology to allow for greater agility. For example, Nmap scans may be mapped against asset inventories to identify assets with missing entries. Mechanisms for automated secure software updates must be leveraged whenever possible. Validating, replacing, and revoking certificates, keys, and algorithms should be similarly automated.

A crypto agility remediation roadmap builds on this foundation of enterprise crypto policy, associated process, and appropriate technology. First, the crypto policy should be updated to remove deprecated algorithms and incorporate any replacements. Second, associated processes should be leveraged to push those requirements.

For example, new crypto requirements should be pushed into third party contracts. Similarly, change management should be used to both update assets that are being secured and expedite timelines for assets that need to be phased out. Appropriate communications channels should be used to make developers aware of new requirements. A comprehensive list of actions is beyond the scope of this paper and will necessarily depend on the nature of the enterprise. Finally, enterprise should review the existing tooling to determine whether additional technical solutions are needed to implement the remediation plan. For example, if some assets can neither be phased out nor upgraded, e.g. due to resource constraints, it may be necessary to implement compensating technology, such as access control.

**Case Study: Quantum Computing**

Using CARAF may help organizations with the transition by prioritizing mitigation based on expected risk. Thus, in this section, we present a case study on how to operationalize CARAF for the use case of quantum computing. As early as 2015, NSA recommended that organizations prepare for the upcoming quantum resistant algorithm transition [45]. One possible solution using conventional hardware is post-quantum cryptography (PQC), or conventional ciphers based on mathematical problems other than factoring and discrete logarithm. NIST is currently reviewing PQC and quantum safe standards are expected to be out by 2024 [30]. Thus, US government may have an expectation that organizations (in particular vendors that provide critical services) transition to quantum safe alternatives in the not-too-distant future. This can be expected regardless of whether quantum computers that impact current crypto systems become practical by 2024 or not.

**Identify threat**

Quantum computing impacts the security of encryption schemes, hashing algorithms, as well as digital signatures [6], as noted in Table 1. Symmetric key and hashing algorithms will need larger key sizes and larger outputs to maintain their current security posture. Whereas for public key cryptography, systems will have to migrated from existing algorithms to quantum safe alternatives.

In a standard agility assessment, modularity, and abstractions in software and network implementations allow easy switching of cryptographic algorithms. However, quantum safe algorithms are based on fundamentally different underlying mathematical assumptions compared to existing solutions such as RSA or ECC. These assumptions add four additional constraints to crypto agility: (i) larger key sizes, (ii) larger outputs, (iii) greater time to encrypt/decrypt (or sign), and (iv) longer time to establish a secure channel or validate authentication. Correspondingly, the threat from quantum computing will include, in addition to standard crypto agility concerns, limited storage as well as constraints on operational overhead. These new crypto requirements can be difficult to implement in assets with limited space, high speed requirements, or hardcoded implementations of crypto, in which case they become crypto agility concerns.

**Inventory of assets**

The NIST competition for replacements of current public key algorithm has reached the final round [30]. Provably secure PQC and standards are expected to be out by 2024, so any cryptographic assets to be phased out before then can be eliminated from this risk assessment. Given that symmetric key crypto systems and hashing algorithms require an easier fix, i.e. increase in key size or hash

| Table 3: risk mitigation assessment |
|-----------------------------------|
| Mitigation methods | Low cost | High cost |
| X + Y < Z | Phase out | Accept risk |
| X + Y > Z | Secure asset | Phase out |
output, we do not consider those to be within the scope of this assessment.

Instead we focus on public key crypto systems, which require migration to a different class of algorithms. Public key crypto systems are critical to establish authentication through the use of digital signatures as well as for encrypting data in transit by establishing session keys (which then use symmetric key cryptography).

- **Lifecycle Management**: We assume that there is no distinction between enterprise owned and third-party devices, i.e. the latter have to comply with the same lifecycle management expectations, which are enforced through third party contracts. This assumption implies that if an enterprise owned device is given an upgrade to address cryptographic risk or likewise rendered end of life, a similar device from a third-party vendor will also be either upgraded or discarded. If the third-party device has a longer timespan, then the risk estimates in the following section will be different.

### Risk estimation
As noted before, the expected value of risk is a function of the timeline of the risk and the cost of migration. The timeline for mitigation and shelf-life can vary widely depending on the implementation of the assets and the type of assets. The cost can also vary widely depending on the amount of assets and the organization. The ranges provided are based on industry estimates for frame of reference.

### Timeline
Timeline is based on three distinct factors.

- **Z (Threat)**: NIST posits that a quantum computer capable of breaking 2000-bit RSA in a matter of hours could be built by 2030 for a budget of about a billion dollars [6]. In addition, NIST is reviewing potential solutions for quantum-safe algorithms and is supposed to publish its recommendations between by 2024 [30]. Based on this we assume that if the threat is realized in 20+ years, i.e. twice of NIST’s estimated timeline, the risk may be low. Correspondingly, if the threat is realized in 10–20 years risk is medium, in 5–10 years is high, and 0–5 years is critical.

- **Y (Mitigation or remediation)**: If the asset uses a TLS implementation with existing support for quantum-safe algorithms, the time needed for mitigation will be less if the TLS implementation does not currently have support. Even when support is available, migration can take a long time. The SHA-1 to SHA-2 migration took approximately 10 years. Blackberry took 5 years to move from 3DES to AES while in control of all devices and servers [47]. For the purpose of this risk estimation, we assume that:
  - Enterprise owned assets with quantum support will take 5–10 years to migrate.
  - Enterprise owned assets without quantum support will take 11–20 years to migrate.
  - Third party owned assets with quantum support will take 11–20 years to migrate.
  - Third party owned assets without quantum support will take more than 20 years to migrate (if ever).

- **X (Shelf-Life)**: A consumer grade IoT device for a generic enterprise setting can have a lifetime from 2 to 20 years. For example, it is not uncommon for phones to be upgraded every 2 years. In contrast, printers and cameras may often last over 10 years. Thus, for the purpose of this risk estimation we can divide the assets to align with the quantum threat timeline as documented in Table 4.

### Cost
The next step is to estimate the cost of migrating to a quantum safe solution for each class of assets according to the timeline. The exact value of the migration will differ based on the organization, the type of IoT asset, etc. However, a few trends will likely apply across the board. First, the cost to migrate will decrease over time, as new tools similar to Open Quantum Safe (OQS) are developed that make integration of quantum safe algorithms easier. Even existing tools will undergo greater testing. As more entities try to use these tools in practice, they will publish improvements. In contrast, the IoT systems that use TLS implementations that already provide the option
to either use a post-quantum or a hybrid solution will be less expensive to migrate compared to those that do not. Based on that, Table 5 provides a qualitative cost estimate for quantum safe TLS migration for IoT assets.

Secure assets

We can use the information from Tables 4 and 5 to approximate the appropriate security mitigation, shown in Table 6 with explanation as follows according to color:

- (Gray) Low-risk IoT devices are those that are already scheduled to be phased out in the next 5 years. For these devices, regardless of ownership or implementation, the cost to migrate to a quantum safe solution is nontrivial. Thus, it is reasonable to accept the risk (assuming that the current timeline for phase out is ensured).
- (Blue) For any enterprise owned IoT devices with support for PQC where risk is medium or higher, the cost to migrate will be on the lower end. For these assets, the organization may want to upgrade to a quantum-safe or hybrid solution. The organization will have adequate time to test and the experience gleaned will help them prepare for a post-quantum world in other domains as well.
- (Green) For enterprise owned IoT devices with no support for PQC, the question is more challenging. In the near term, the cost of migration is likely to be high. They will have three mitigation options (in increasing order of difficulty):
  - Move to a different implementation of TLS.
  - Write a custom fork of the current implementation.
  - Implement a compensating control, such as a quantum safe wrapper for the TLS protocol, e.g. Golioqs wrapper for Go applications [48].
- Although the first may pose operational challenges, the remaining two may introduce additional security concerns. Thus, the appropriate mitigation may be to simply accept the risk, which in that time frame is medium.
- In contrast, the lifetime of IoT devices goes beyond 10 years, the risk of not using a quantum-safe version of TLS becomes higher. At the same time, the cost to migrate to a quantum-safe version of TLS goes down. With the additional time, the enterprise will have adequate time to re-architect their IoT device to use another implementation of TLS as well as to test it. Thus, it might be more rational to secure the asset. Alternatively, if upgrading is too difficult, the solution may be to phase out the insecure IoT device and replace with another device that supports a quantum-safe implementation of TLS.
- (Yellow) For third-party IoT devices that support PQC, securing the assets requires consideration of multiple factors. The availability of quantum safe alternatives within the TLS implementation reduces the cost of migration. However, as the device needs to be updated by the third party that owns it, enterprise needs to enforce mitigation through contracts. Some vendors may charge extra for the cost of development and integration. If the vendor has no previous experience with post-quantum algorithms and they are pressed for time, they may not perform adequate testing or inadvertently add bugs to the code.
- Thus, it may be reasonable to accept the risk in the short term, i.e. while the risk is medium. As the risk becomes high (or critical) and the corresponding cost to secure the asset goes down, the appropriate mitigation may be to either ask the vendor to provide an upgrade or switch to a different vendor with pre-existing defense against quantum threats.
- (Red) For third party IoT devices with no support for PQC. If these devices have a shelf-life of greater than 10 years, they should be phased out prior to the risk becoming high or critical. However, if the shelf-life is between 6 and 10 years, the risk is medium and the best option may be to accept the risk. The timeline will impose a high cost if the organization wants to mitigate the risk by either implementing a compensating control or switching to another vendor.

Organizational roadmap

Based on the security mitigation strategy identified in Table 6, the enterprise must determine a tactical roadmap next. For the low-risk scenario, the solution is to continue enforcing the organization’s existing technology change management plans. For medium risk where the organization accepts the risk, e.g. enterprise devices with no post-quantum support, the roadmap will include an exception process for the assets in question.

Third-party IoT devices with no post-quantum support pose the potential for high or critical risk, and the solution is to phase out. Here the enterprise will have to start the process of reviewing alternatives. This will include working with the procurement team to identify other vendors and including requirements around post-quantum security in the procurement guidelines.

The roadmap for mitigating the risk by securing the asset will require upgrading to a quantum safe alternative. It will be necessary to understand the trade-offs between different options before moving forward. There is a significant body of existing literature, e.g. [49–51], with detailed benchmarks. Although organizations can learn from prior work, the associated algorithms are being continuously updated. Furthermore, different implementations will result in distinct performance outcomes. Thus, organizations will need to invest in custom benchmarks that capture the constraints of their assets as well as the respective operational environment.

For the purpose of demonstration, we consider a simulation of TLS communication in a generic system using a Linux virtual machine. We use an x86-64 Ubuntu 19.10 Virtual Machine, running Linux Kernel 5.3.0-24-generic, and GNU C Compiler (gcc) 9.2.1. Currently, there are three libraries that integrate quantum safe alternatives into TLS connection: (i) ISARA Radiate, (ii) libpqcrypto, and (iii) OQS. Here we explore solutions using OQS, which has better community support compared to libpqcrypto and is not proprietary as is the case with ISARA Radiate.

OQS is a consortium of partners and contributors, led by the University of Waterloo, that have written and released open-source implementations of many PQC algorithms in C libraries on GitHub. The PQC for OpenSSL are still in development and does not include all versions of all candidate, so we chose two with different mathematical foundations for benchmarking to illustrate the trade-offs [30], shown in Table 7. (Note: Picnic-L3, and Picnic-L5 were not supported in OQS at time of benchmarking and thus do not have the corresponding key/certificate generation speeds.)

The two digital signature algorithms we benchmarked are Dilithium, which is lattice based algorithm and a finalist of round 3
of the NIST competition, and Picnic, which is hash based and an alternate of round 3 of the NIST competition. The difference in the mathematical foundation results in different overheads, as reflected in the benchmark table. Lattice based algorithms have: (i) larger key sizes, (ii) smaller signature size, and (iii) faster key/certificate generation speed. Similar trade-offs have been seen in benchmarks performed by Amazon for hybrid public key cryptography [31].

Depending on the resources of the IoT asset as well as the operational constraints, the organization will choose the best possible alternative. Many IoT assets may have limited storage. However, the key sizes for Dilithium are not significantly different from RSA (or Diffie–Hellman), which is usually at least 1024 bit. Simultaneously, the signature for RSA is the same order of magnitude as the key. The signatures for Dilithium are also similar in size to their keys. Picnic, while having much lower key sizes, has a signature size that is at least one order of magnitude greater than RSA. Thus, if the asset currently uses RSA, Dilithium may be an appropriate replacement. Organizational roadmap will need to ensure that the asset can switch algorithms, e.g. through a software update. However, the organizational roadmap will not have to plan for hardware upgrades as the signatures sizes as well as the key sizes are similar to that used for RSA currently.

### Conclusion

Lack of crypto agility is a risk that can hamper the ability of organizations to respond to changing regulatory and technology landscapes. At the same time, it is a difficult risk to address due to the lack of prior incidents and data on exposure. This manifests in real life examples, which show that the transition can be both expensive and take a long time. For example, the SHA-1 to SHA-2 transition took over 10 years. Thus, a systematic review of threats that impinge on cryptography is needed to ensure that adequate agility is being built in as well as appropriate mitigation options are available for legacy and third-party systems.

CARAF proposed here is the first framework that allows enterprises to undertake such reviews and create an associated playbook. Applying this framework to the emerging threat of quantum computing to a generic IoT system provides clear actionable guidance for a risk mitigation strategy, in particular, identifying areas that need to be prioritized for protection and where it may be reasonable to accept the risk. Furthermore, converting this strategy into a tactical roadmap provides a better understanding of the solution space as well as the inherent challenges.

### Conflict of interest statement

The authors of this papers are full time employees of Comcast Cable. The views expressed in this paper are those of the individual authors and do not reflect the official position of Comcast Cable.

### References

1. Ratcliff RA. *Delusions of Intelligence: Enigma, Ultra, and the End of Secure Ciphers*. Cambridge: Cambridge University Press, 2006.
2. Kahn D. *Seizing the Enigma: The Race to Break the German U-Boat Codes, 1939–1943*. Annapolis, Maryland, USA: Naval Institute Press, 2001.
3. Wesolowski S. The invention of enigma and how the polish broke it before the start of WWII. In: *IEEE Conference on the History of Telecommunications (Canada)*. University of Waterloo, 2001.
4. Fluhrer S, Mantin I, Shamir A. Weaknesses in the key scheduling algorithm of RC4. In: *International Workshop on Selected Areas in Cryptography*. Springer, 2001, 1–24.
