

特色专题

AI 智能体辅助金融抗量子密码迁移: 迁移点发现与适配建议

李一¹, 丁津泰^{2,3}, 刘锐³

摘要 量子计算的快速发展对金融基础设施的公钥密码体系构成了根本性、系统性威胁。尽管各国监管机构在陆续出台抗量子密码迁移的指引框架,但金融机构的实际进展普遍缓慢。其核心瓶颈在于存量系统规模庞大、密码调用点数量众多且较为分散,而具备抗量子计算密码工程能力的专家极为稀缺。据此,提出 AI(artificial intelligence)智能体辅助金融抗量子密码迁移的思路,将 AI 智能体在迁移中的职责界定为面向不具备抗量子密码背景的存量系统责任人的专家知识投送:发现需要迁移的密码调用点,判定其业务语义属性与应适用的合规要求,生成可追溯至标准条文的适配建议与改造方案,并将决定权保留于人。围绕这一定位,提出识别、判定、生成、验证 4 类能力及相应的自主性分级,分析了迁移点漏检、属性误判、建议不合规、规则库滞后与过度信任 5 类特有失效模式;并以金融领域应用的开源分布式数据库 TiDB 为例,分析了迁移点的实际规模与排查难度,验证了将智能体定位于专家知识投送的必要性与可行性。

关键词 AI 智能体; 抗量子密码迁移; 密码资产测绘; 密码敏捷性; 金融业

量子计算的快速发展正在对当前公钥密码体系形成重要威胁,其本质是传统公钥密码学依赖的大整数分解、椭圆曲线离散对数等数学难题,在 Shor 量子算法^[1]面前会从“计算不可行”变为“多项式时间可解”^[2]。对于金融业,这种冲击将直接影响目前金融市场基础设施(financial market infrastructure, FMI)的网络安全基础。目前,从跨境支付系统、实时全额结算系统,到证券结算、中央存管机构的交易确权机制,再

到客户身份认证、敏感数据加密等基础环节,几乎所有金融业务流程都建立在可被量子计算破解的密码算法之上。传统密码体系依赖的计算复杂性假设,在经典计算下具备足够安全冗余,但在量子计算面前已不再成立^[3]。

全球风险研究所(Global Risk Institute)调研显示,受访密码学专家对具备密码破解能力的量子计算机(cryptographically relevant quantum computer, CRQC)将在 10~15 年内出现给出了较高概率估计,其中 15 年内出现的概率估计多在 5 成以上^[4]。面对这一威胁,国际清算银行(Bank for International Settlements, BIS)、欧盟委员会(European Commission, EC)^[5]、七国集团(G7)^[6]等已明确要求金融体系提前启动抗量子计算密码(post-quantum cryptography, PQC)迁

1. 中国人民银行金融研究所, 北京 100800

2. 西交利物浦大学数学与物理学院, 苏州 215123

3. 西交利物浦大学后量子迁移交叉实验室(PQC-X), 苏州 215123

收稿日期: 2026-05-13; 修回日期: 2026-08-24

基金项目: 国家重点研发计划项目(2023YFC3305500)

作者简介: 李一, 高级工程师, 研究方向为金融科技与量子安全, 电子信箱: yil@pbc.gov.cn

引用格式: 李一, 丁津泰, 刘锐. AI 智能体辅助金融抗量子密码迁移: 迁移点发现与适配建议[J]. 科技导报, 2026, 44(16): 123-131;

doi:10.3981/j.issn.1000-7857.2026.05.00058

移,避免在 CRQC 可用时陷入系统性安全瘫痪。这是因为金融业的 PQC 迁移是一个系统性复杂工程,迁移周期较长,不能等到 CRQC 问世再开始。2025 年 7 月 BIS 在《金融系统量子就绪路线图》中提出警示^[7]:“即使 CRQC 的问世时间存在不确定性,金融机构从现在起就需启动抗量子密码迁移——任何延迟都将导致当前数据暴露于未来的量子攻击之下”。这是因为恶意主体可通过网络窃听、数据泄露等方式,先获取当前金融交易的加密报文或敏感数据并长期存储,待 CRQC 问世后再解密这些信息。这对金融数据的长期保密性构成了不可逆挑战,甚至可能违反《通用数据保护条例》(General Data Protection Regulation, GDPR)等全球数据保护法规中“数据需在生命周期内持续受保护”的核心要求。

美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)于 2024 年已正式公布首批 PQC 标准,包括 ML-KEM(module-lattice-based key-encapsulation mechanism, CRYSTALS-Kyber)、ML-DSA(module-lattice-based digital signature algorithm, CRYSTALS-Dilithium)、SLH-DSA(stateless Hash-based digital signature algorithm, SPHINCS+)^[8],并于 2025 年增补 HQC(Hamming Quasi-Cyclic)作为 ML-KEM 的补充方案^[9]。美联储研究人员^[10]、欧洲中央银行(European Central Bank, ECB)、新加坡金融管理局(Monetary Authority of Singapore, MAS)^[11]、香港金融管理局(Hong Kong Monetary Authority, HKMA)^[12]等亦相继开展金融量子安全研究并明确迁移的紧迫性(其中新加坡金融管理局的试点聚焦量子密钥分发,与本文讨论的 PQC 算法迁移属不同技术路线),但全球金融机构的实际进展却面临多重难题^[13]。造成这一困境的核心原因在于,传统迁移方案高度依赖人工介入。从梳理分散在数百个系统中的密码资产,到针对不同业务场景测试抗量子算法的兼容性,再到解决版本冲突、硬件适配等技术问题,每一个环节都需要密码学专家与业务人员的深度协作,不仅周期漫长,还容易因人为疏漏留下安全隐患。

造成迁移进展缓慢的原因,通常被归结为技术复杂与周期漫长,但更根本的约束在于人。PQC 迁移所需的判断,几乎每一项都要求同时理解密码学、协议实现与具体业务语义,而同时具备这 3 方面能力的工

程师在全球范围内稀缺。一家大型金融机构往往拥有数百至数千个业务系统,加之供应商交付与外包代码,其密码调用点数以万计;若要求每一个存量单位都配备一批具备 PQC 知识的工程师,在可预见的迁移窗口内并不现实。因此,迁移的真正瓶颈不是“选哪一种算法”,而是如何把稀缺的密码专家知识规模化地投送到每一个存量代码点上。因为通过标准化的算法与参数集数量有限,且被标准与监管政策事实锁定。

具备自主感知、推理与工具调用能力的 AI 智能体(artificial intelligence agent),为这一知识投送问题提供了可行的技术载体。与传统的规则驱动型扫描工具不同,智能体能够在开放、异构的存量环境中理解代码与配置的语义,并结合外部规则库给出带依据的判断。智能体在抗量子计算密码迁移中辅助不具备 PQC 背景的存量系统责任人发现需要迁移的密码调用点,判定其属性与应适用的合规要求,生成可追溯至标准条文的适配建议与改造方案。本文首先界定金融 PQC 迁移的问题结构,说明算法与参数由标准和政策唯一确定、不构成决策空间;进而给出智能体的角色定位、能力边界与自主性分级,阐述面向存量系统的迁移点发现、属性判定与适配建议生成,讨论迁移后的密码敏捷性运维与相应风险,并以开源系统为例给出不依赖生产环境的验证设计。

1 金融抗量子密码迁移的问题结构

要界定 AI 智能体在迁移中的合理职责,首先需要区分迁移工作中哪些环节的答案由标准与政策唯一确定,哪些环节属于开放域问题。这一区分决定了自动化的边界:前者可判定、可校验,适合规模化自动执行;后者存在固有不确定性,必须保留人工复核。

在算法与参数层面,可选空间远比通常想象的狭小。NIST 在其 PQC 标准化评估框架中,以对高级加密标准(advanced encryption standard, AES)、安全散列算法(secure Hash algorithm, SHA)等对称密码的攻击代价为基准,定义了 1~5 共 5 个安全类别(security category),并在 FIPS 203、FIPS 204、FIPS 205 中为每个参数集指定了对应类别^[14]。但实际可用的档位只有 3 个:ML-KEM 的 512、768、1024 分别对应类别

1、3、5, SLH-DSA 与 HQC 同样只覆盖类别 1、3、5, ML-DSA 的 44、65、87 对应类别 2、3、5, 其中仅 ML-DSA-44 落在类别 2。更重要的是, 这 3 个档位在实践中进一步被政策与协议锁定: 美国 CNSA 2.0 对国家安全系统直接指定 ML-KEM-1024 与 ML-DSA-87^[15]; 而互联网与商用基础设施的事实标准是混合组合 X25519MLKEM768, 对应类别 3。这种收敛并非偶然, 而是互联互通的必要条件——协议套件、证书配置文件与协议参数注册共同将参数集固定下来, 单个系统一般不具备、也不应具备自行选择的权力。

在中国, 这一约束更为刚性。《中华人民共和国密码法》^[16]与《商用密码管理条例》^[17]要求关键信息基础设施使用符合国家标准、经检测认证的商用密码产品; 截至本文撰写时, PQC 的商用密码国家/行业标准仍在推进制定之中^[18]。这意味着中国金融机构现阶段并不存在“自主选择抗量子算法”的空间, 其可行的准备工作是摸清密码资产底数、推进密码敏捷性改造、形成迁移就绪度评估, 使系统具备在标准落地后快速切换的能力。在管理层面, 行业规范已先行一步: 2026 年 4 月发布的团体标准《银行业抗量子计算密码迁移管理指南》^[19], 就迁移的风险评估、迁移规划、系统改造、测试验证与过渡期运维提出了管理要求; 但其定位是管理指引, 并不指定具体算法, 这与前文的判断一致。

与之相对, 迁移工作中真正的不确定性集中在存量侧, 可分为 3 类。其一是发现问题: 密码用法可能以硬编码的算法名、自研加密封装库、配置文件项、第三方软件开发工具包调用乃至仅存二进制的遗留模块等形态存在, 跨越多种语言与运行环境, 形态长尾且持续变化, 基于关键字与正则的静态规则难以覆盖。其二是属性判定问题: 同一处 RSA-2048 签名调用, 究竟承担的是内部日志的完整性校验, 还是对外报文的不可否认性保障, 决定了它是否必须迁移、应对齐哪一档要求, 以及优先级如何排序; 这一判断依赖数据生命周期、对外暴露面、证书链层级与业务语义等上下文, 无法由算法名本身得出。其三是改造问题: 从“应当用什么”到“这段代码怎么改、依赖库是否支持、硬件安全模块是否提供相应机制、改完如何验证”, 仍需大量工程判断。

由此可以得到本文的基本判断: 迁移中“选什么”的部分是确定性的映射问题, 其可选项少、答案唯一、依据可追溯, 恰恰因此适合由机器批量执行并逐条校验; 而“有哪些点需要迁、每个点是什么性质、如何改”的部分才是开放域问题, 也正是稀缺的密码专家知识被消耗的地方。AI 智能体的价值应当落在后者; 其对前者的作用是准确、可追溯地应用规则, 而非行使裁量权。

2 AI 智能体在迁移中的角色定位与能力边界

在人工智能的技术谱系中, AI 智能体属于自适应智能系统, 具备自主感知、决策与执行能力。传统 AI 系统(如规则引擎等)的本质是“工具”: 它们只能在预设的规则范围内执行单一任务, 无法应对未被规则覆盖的异常场景^[20]。而智能体则更接近“协作主体”: 它们能够通过内置的传感器感知周围环境的动态变化, 基于预设的目标自主规划任务流程, 甚至在遇到障碍时通过强化学习优化执行策略。例如, 当 PQC 迁移过程中遇到软件版本冲突导致抗量子算法无法部署时, 传统工具会直接报错并停止执行, 而智能体会自动检索版本兼容库, 尝试替代部署方案, 直到问题的合理解决。

本文从迁移任务本身反推智能体所需能力。迁移的一次完整作业可分解为“发现—判定—生成—验证”4 个必要环节, 任一环节缺失, 作业链条即告中断。据此, 面向金融 PQC 迁移的 AI 智能体需具备下列 4 类能力。

1) 识别能力: 在异构存量环境中发现密码调用点。其对象不限于运行时可观测的行为, 还包括源代码中的算法调用与常量、构建与部署配置、依赖组件与第三方软件开发工具包, 以及缺失源码的二进制模块。与基于关键字匹配的扫描工具相比, 智能体的增量在于处理长尾形态——自研封装库对底层算法的间接调用、跨语言调用链, 以及以配置项而非代码形式表达的算法选择, 均需结合上下文语义方能识别。

2) 判定能力: 为每一个已发现的调用点确定其业务语义属性与应适用的合规要求。属性包括所保护数据的保密期限、是否跨机构或对外暴露、在证书链中的层级、是否承担不可否认性等; 合规要求则来自

国家标准、检测认证目录与监管指引所构成的规则库。判定的输出不是一个算法名,而是一条带依据的结论:该点是否必须迁移、应对齐何种要求、优先级如何,以及支撑该结论的标准条文与置信度。

3) 生成能力:将判定结论转化为可执行的改造方案,包括目标配置、依赖组件升级路径、接口与报文长度变更的影响范围、硬件安全模块与中间件的适配方式,必要时直接生成代码补丁与配置变更。对于当前尚无法直接迁移的系统,生成的应当是密码敏捷性改造方案——将分散的算法调用收敛至统一抽象层,使其在标准落地后具备快速切换的能力。

4) 验证能力:对自身输出进行检验并给出可信度。包括改造后的功能回归与互操作性测试、建议与规则库的一致性自检,以及对低置信度判定的显式上报。验证能力是前3类能力可被信任的前提,也是人工复核得以聚焦的依据——智能体的价值不在于宣称自己正确,而在于准确标出自己何处不确定。

上述4类能力并不意味着智能体应当独立完成迁移。考虑到金融系统的合规约束与变更风险,有必要明确其自主性边界。本文将智能体在迁移中的自主性划分为4级:L1仅输出迁移点清单,以密码物料清单(cryptography bill of materials, CBOM)等标准化形式呈现;L2在清单基础上补充属性标注与风险定级;L3进一步给出推荐配置、依据条文与置信度;L4生成改造补丁并在测试环境中完成自验证。4级之上,生产环境的任何变更一律经人工审批后实施;置信度低于阈值的判定必须上报,而非按默认值处理。金融场景下的合理上限是L4加人工审批。

在这一界定下,智能体既不是可以取代密码专家的自主决策者,也不只是提高扫描速度的自动化工具,而是稀缺专家知识的投送载体:它把少数专家对标准的理解,以可校验、可追溯的形式,规模化地施加到每一个存量代码点上,从而使不具备PQC背景的系统责任人也能够按正确的路径完成迁移。

3 面向存量系统的迁移点发现与适配建议

按照第1章给出的问题结构,本章讨论迁移作业中开放域部分的技术路径,即从存量系统出发,依次完成迁移点发现、属性判定与适配建议生成、改造方

案生成与灰度验证3个步骤。需要说明的是,这3个步骤对应迁移的准备与实施阶段;迁移完成后的运行阶段另在第4章讨论。

3.1 迁移点发现:从密码资产测绘到密码物料清单

密码资产测绘是PQC迁移的第一步,也是最基础的工作。只有准确把握“哪些系统使用了哪些密码算法”,才能制定科学的迁移策略。但传统的人工测绘方式存在明显的局限性,主要有金融机构的IT(information technology)系统通常由数百个异构模块组成;密码算法可能分散在操作系统、数据库、应用服务器等多个层级;人工梳理不仅耗时耗力,还容易遗漏边缘系统的密码应用;遗留系统(legacy systems)、定制化系统具有特殊的隐藏加密逻辑等。

迁移点发现需要在多个层面同时展开:网络侧观测实际协商的算法与证书,主机与中间件侧解析配置,代码侧进行静态分析与调用链追踪,对无源码模块则需依赖二进制特征与导入符号分析。这些手段各有盲区,任何单一手段都不足以形成完整底数^[21]。AI智能体在其中的作用,是把多源结果关联到统一的资产模型上,并对规则难以覆盖的长尾形态作出判断,例如识别出某个自研工具类实际封装的是SHA1 withRSA,或某个配置项的取值最终决定了传输层安全协议握手中所协商的签名算法。需要指出的是,密码资产发现在工程上难以做到完备,硬编码密钥、私有协议实现与纯二进制模块始终存在遗漏可能,因此发现环节的目标应当表述为显著提高覆盖率并明确标注不确定区域,而非宣称完全识别。以本文第5章分析的分布式数据库TiDB为例,其组件间双向认证与用户身份认证均以证书为基础,Go实现的组件建立在标准库crypto/x509之上,盘点需同时覆盖证书解析、加载、验证、对端校验与编码5类调用形态,相关调用点分散于4个核心组件,其中若干处因Go标准库不支持抗量子签名而构成硬阻塞。

发现结果应当以标准化、可比对的形式固化,而非仅呈现为一张资产图谱。密码物料清单为此提供了可行载体^[22]:它以结构化方式记录每一处密码用法的算法、参数、用途、所在组件及其依赖关系,可随软件版本一同维护与传递,也便于在监管报送与供应链协同中复用。在此基础上,智能体可按业务重要性、

数据保密期限与对外暴露面对迁移点排序,输出分批迁移的优先级建议,并对使用已明确不再推荐算法的调用点作出显著标记。

3.2 迁移点属性判定与适配建议生成

发现之后的关键步骤,是判定每一个迁移点的属性并据此给出适配建议。这一步骤在传统做法中由密码学专家逐个系统人工完成,其耗时与专家可得性直接决定了迁移的整体进度,也正是本文所指的瓶颈所在。需要强调的是,这一步骤中“应当对齐何种要求”的答案由标准与政策唯一确定,本身不构成决策空间;智能体的任务是准确地把规则应用到正确的对象上,而不是在多个候选之间行使裁量权。智能体的判定依托一个以合规文本为唯一数据源的规则库,并结合从存量系统中提取的上下文,输出带依据的结论。判定所依赖的信息可归为以下3类。

1) 合规约束(刚性)。由国家标准、检测认证目录与监管指引直接给出,是不可协商的底线。智能体在此的职责是核验该迁移点依据现行规定应当满足的要求,并标出现状与该要求之间的偏差,同时给出依据条文的出处。智能体不对算法安全性作独立判断,其规则库随标准发布与安全公告更新而更新。

2) 业务语义属性(需推导)。包括所保护数据的保密期限、调用是否跨越机构边界、在证书链中的层级、是否承担不可否认性,以及所属系统的业务重要性。这一类信息无法从算法名得出,需结合数据流向、接口定义、部署拓扑与业务文档推断,是判定环节不确定性的主要来源,也是必须保留人工复核的部分。例如,长保密期数据的加密调用,即使当前所用算法尚未被要求淘汰,也应因“先存储后解密”威胁而获得较高的迁移优先级。由此可进一步得到一条可操作的排序规则:密钥协商与密钥封装类调用应优先于数字签名类调用迁移。前者一旦被记录,未来即可解密历史流量;而签名仅用于实时认证,量子计算能力无法用于伪造已经发生的交互。这一规则可从当前开源协议栈实现的部署顺序中得到印证,详见第5章。

3) 工程约束(需测算)。包括密钥与签名长度增大对报文、存储与带宽的影响,握手与签名验签时延的变化,以及硬件安全模块、中间件、依赖库对目标机制的支持情况。需要明确的是,抗量子算法与现行

公钥算法在算法层面不可互换,所谓“兼容”指的是协议与系统层面能否并行承载2套算法——传输层安全协议的密钥共享扩展、双证书链与复合证书都属于此类工程手段。因此本环节评估的是适配路径与改造工作量,而非算法本身的兼容性。

3.3 改造方案生成与灰度验证

判定结果最终需要落为可执行的改造方案。在过渡阶段,工程上普遍采用传统算法与抗量子算法并行支持的做法:在完全迁移到抗量子算法之前,系统同时承载2套算法,以维持与新旧两侧的互操作。智能体在此可生成分批切换方案并跟踪执行:按系统与通道分批,而非按流量比例切换;先在非核心链路与测试环境完成互操作性与性能验证,确认时延、报文长度与吞吐的变化处于可接受范围后,再逐步扩大范围;每一批次均须具备明确的回滚路径。切换范围、时点与回滚判据由人工确认,智能体负责方案生成、影响分析、执行跟踪与异常上报。

4 密码敏捷性运维与迁移风险分析

4.1 迁移后的密码敏捷性运维

迁移完成并不意味着一劳永逸。抗量子算法的具体实现仍可能被发现缺陷,标准与推荐配置也会随之调整,因此系统需要具备持续的监测与快速切换能力。这一阶段的工作已超出本文所讨论的迁移点发现与适配建议的范围,此处仅作简要说明,以保持迁移生命周期叙述的完整性。

在密钥生命周期维度,智能体会监控密钥的生成、存储、使用、销毁全生命周期。例如,私钥的使用频率是否突然增加,是否在非工作时间被访问等。如果某私钥在特殊时点被频繁使用,智能体会识别为异常,可能是密钥泄露的迹象,这时智能体会立即触发密钥轮换。这种动态防御机制,能够在攻击者完成攻击前,就将风险控制在萌芽状态,为金融机构提供了持续的安全保障。因此,金融信息系统应采用模块化、可快速切换的密码敏捷性设计架构,兼容多类传统密码与PQC,以实现平滑迭代、低成本迁移。

PQC签名是PQC体系的核心应用场景,其用于保障金融交易的不可伪造性与不可否认性,是跨境支付、证券结算等核心业务的安全基石。但传统的密钥

轮换机制存在明显的缺陷: 轮换周期较为僵化, 缺少密码敏捷架构, 缺乏动态应急调整能力, 难以应对量子威胁下密钥泄露、算法漏洞等突发风险。AI 智能体的执行层模块, 可以实现抗量子签名密钥轮换的智能化。尤其是当智能体感知到密钥存在风险时, 会立即触发紧急轮换。例如, 如果硬件安全模块(hardware security module, HSM)设备的日志显示有异常访问记录, 或者抗量子算法被安全机构标记为“存在安全漏洞”, 智能体会在几分钟内完成新密钥的生成与部署, 同时将旧密钥作废。此外, 智能体可结合阈值签名技术进一步强化密钥安全防护。

智能体还可持续同步密码管理机构与金融监管机构发布的安全公告与标准更新, 将其映射到既有的密码物料清单上, 快速定位受影响的系统与调用点, 生成变更方案并提请审批。需要强调的是, 生产环境的算法切换不应由智能体自动触发执行; 其价值在于把“哪些系统受影响、影响范围多大、如何切换”这一原本需要专家逐个系统排查的工作, 在短时间内完成。

4.2 迁移点发现与判定的特有失效模式

在本文界定的定位下, AI 智能体的失效不再主要表现为“选错算法”, 而是表现为发现与判定环节的误差, 及其经由规模化应用而被放大的后果。本节讨论的失效模式主要包括迁移点漏检、属性误判、建议不合规、规则库滞后与过度信任 5 类。

1) 迁移点漏检风险。这是本定位下最严重的失效模式。未被发现的密码调用点不会产生任何告警, 却会在迁移被宣告完成之后继续以量子脆弱的形态运行; 由于“已完成迁移”的判断带来虚假的安全感, 其危害甚至超过迁移尚未开始的状态。漏检的来源包括无源码的遗留模块、私有协议实现、动态加载的组件与运行期才确定的配置。缓解手段是多手段交叉验证与覆盖率的显式度量: 智能体应当报告其扫描覆盖了哪些范围、哪些区域无法确证, 而不是仅输出一份看似完整的清单。第 5 章的案例即显示, 数量最多的证书解析类调用分散于 4 个核心组件, 人工逐点排查极易遗漏其下游的证书校验环节, 而这类遗漏通常不会产生编译或运行期错误, 因此难以被常规测试发现。

2) 属性误判风险。属性判定依赖对业务语义的推断, 是不确定性的主要来源。低估将导致本应迁移的调用点被排在低优先级甚至被判定为无需迁移; 高

估则造成不必要的改造成本与性能损失。误判在单个系统上表现为局部错误, 但当同一判定规则被批量应用于数千个调用点时, 系统性偏差会被成倍放大。因此判定结果必须携带置信度, 低置信度样本必须进入人工复核队列。

3) 建议不合规风险。智能体给出的配置若指向未纳入国家标准或未通过检测认证的算法与实现, 则该建议本身即构成合规问题。规避这一风险的前提, 是规则库以合规文本为唯一数据源, 且每条建议都可回溯至具体条文; 智能体不得基于自身对算法安全性的判断, 给出超出规则库范围的推荐。

4) 规则库滞后风险。标准更新、参数集调整与安全公告发布之后, 若规则库未能及时同步, 智能体将持续输出过时建议, 且因其形式规范而更易被采信。规则库应具备明确的版本、生效范围与更新机制, 历史建议应可依据规则库版本追溯与重新评估。

5) 过度信任风险。本文所设定的服务对象, 恰恰是不具备 PQC 背景的存量系统责任人——这意味着他们通常不具备独立复核智能体输出的能力。这是该定位的固有脆弱性: 建议的形式越规范、表述越确定, 越容易被直接采纳。缓解手段一方面在于输出必须附带依据条文与置信度, 使复核有据可依; 另一方面在于流程设计上保留具备密码专业能力的审批环节, 尤其针对高优先级与低置信度的判定结果。

4.3 智能体自身的安全与治理风险

除上述与迁移作业直接相关的失效模式外, 智能体作为承载密码资产全量信息的系统组件, 其自身的安全与治理风险同样不可忽视, 主要包括被攻击风险、数据泄露风险、恶意操控风险与信任机制风险。

在 AI 系统被攻击风险方面, 智能体作为 PQC 迁移系统的核心组件, 本身也可能成为攻击者的目标。特别是在量子威胁环境下, 如果智能体的通信信道、存储介质或处理过程缺乏适当的保护, 可能被攻击者利用量子计算能力进行破解。

数据泄露风险是智能体面临的另一个重要安全威胁。智能体在 PQC 迁移过程中需要处理大量敏感信息, 包括密码资产清单、密钥材料、系统配置、用户隐私数据等。尤其是完整的密码物料清单本身即构成一份高价值的攻击地图, 它精确标明了机构内尚未完成迁移的脆弱位置, 其保护级别应不低于所描述的

系统本身。如果这些信息在存储、传输或处理过程中缺乏适当的保护,可能导致大规模的数据泄露事件。

在恶意操控风险方面,智能体可能被恶意行为者操控,导致其做出错误的决策或执行有害的操作。这种风险可能来自于训练数据的污染、模型参数的篡改、输入数据的伪造等多种途径。例如,如果攻击者能够向智能体提供虚假的安全事件信息,可能导致智能体触发错误的应急响应措施,影响系统的正常运行。

信任机制风险涉及智能体与其他系统和用户之间的信任关系。在 PQC 迁移环境中,智能体需要与多个系统交互,包括认证服务器、密钥管理系统、应用程序等。如果这些交互过程中的信任机制存在漏洞,可能导致身份伪造、权限滥用、数据篡改等安全问题。

5 案例分析与验证设计

本文界定的智能体职责,其可行性与难点可以在公开系统上得到具体说明,而不必依赖金融机构的生产环境。本章先以一个具备金融信息属性的开源系统为例,说明迁移点发现的实际规模与性质,再给出可执行的评测设计。

TiDB(一款开源分布式关系型数据库)面向高并发在线交易与实时分析场景,已被多家银行等金融机构用于重要业务系统,其客户端接入、组件间通信与用户身份认证均以 TLS(transport layer security)与 X.509 证书为安全基础,构成了与金融系统同构的密码依赖,且代码完全公开、可完整编译与测试,适合作为分析对象。其证书与公钥密码的使用贯穿 3 类通道:客户端与服务器之间的 TLS 接入、组件之间基于 X.509 证书的双向认证(并按证书通用名(common name, CN)限定可信调用方),以及基于 X.509 客户端证书的用户身份认证(可对 Subject、Issuer、主体备用名(subject alternative name, SAN)等字段提出要求)。这些调用为数众多,主要分布在 4 个核心组件(TiDB、PD、TiKV、TiFlash)之上,跨越 Go 与 Rust 技术栈,并可归为证书解析、加载与签发、链校验、对端 CN/SAN 校验与套件选择 5 类形态;其中以 Go 标准库 crypto/x509 实现的 TiDB、PD 等组件因标准库尚不支持抗量子签名而构成硬阻塞,而以 Rust 生态(OpenSSL/

rustls)实现的 TiKV 存储层则走不同的迁移路径——同一系统内 2 条技术栈的差异,正是异构存量环境迁移面复杂性的缩影。

这一案例说明了 2 点。其一,迁移点的数量与分散程度决定了人工逐点排查在工程上不可行,而其中多数调用点的改造动作却是可判定的。一旦确定某调用点属于何种形态、承担何种密码角色,应当路由到何处即由标准与封装设计唯一确定。其二,真正的困难不在于确定目标算法,而在于把调用点找全并判断其性质。同为证书解析调用,位于身份验证路径上的一处与位于日志工具中的一处,其迁移必要性与优先级完全不同,而这一区别无法从调用本身读出。这与第 1 章给出的问题结构一致。

对若干公开的密码工具与协议栈实现的观察,还提供了一条可操作的优先级规则。当前已具备抗量子能力的开源实现,普遍先在密钥协商环节部署经典算法与 ML-KEM 的混合方案,而证书签名仍沿用经典算法。这一顺序符合“先存储后解密”威胁的时间性:密钥协商一旦被记录,未来即可解密历史流量;而签名仅用于实时认证,量子计算能力无法用于伪造已经发生的交互。因此在属性判定中,密钥协商与密钥封装类调用应获得高于数字签名类调用的迁移优先级。

在案例分析的基础上,迁移点发现与适配建议的质量可以通过标注评测进一步量化,且同样不需要生产环境。评测语料可由具备真实密码用法的开源项目与一组包含已知用法形态的测试工程共同构成,由具备密码工程经验的人员标注密码调用点及其应有属性,形成基准集。指标分为 3 层:发现层面度量调用点的召回率与准确率,并按调用形态分层统计,以暴露长尾情形下的能力边界;判定层面度量属性判定的准确率与置信度的校准程度,即低置信度样本是否确实集中了错误;建议层面度量输出建议与规则库条文的符合率,以及不合规建议的发生率。对照组可设为基于关键字与正则的静态扫描,以及现有的密码资产发现工具,以说明智能体的增量究竟来自何处。该评测的实施是本文后续工作的重点。

6 结论

PQC 迁移是全球金融体系面临的一项长期而艰

巨的任务,它不仅涉及技术层面的算法替换,还涉及监管层面的规则协调,以及金融机构内部的流程重构。本文的基本判断是,迁移的真正约束不在“选哪一种算法”,而在于如何把稀缺的密码专家知识规模化地投送到存量系统数量庞大的密码调用点。据此,本文将 AI 智能体的职责界定为面向不具备 PQC 背景的存量系统责任人的专家知识投送:发现迁移点,判定其属性与应适用的合规要求,生成可追溯至标准条文的适配建议与改造方案,并将决定权保留于人。围绕这一定位,本文提出了识别、判定、生成与验证 4 类能力及相应的自主性分级,分析了迁移点漏检、属性误判、建议不合规、规则库滞后与过度信任 5 类特有失效模式,并以开源的金融领域使用的 TiDB 数据库为例分析了迁移点发现的实际规模与难点,给出了不依赖生产环境的评测设计。

本文的研究带来 3 点更具一般性的启示。其一(方法论方面),通过引入 AI 智能体辅助金融抗量子密码迁移,实现“专业知识投送与人机协同”,在高合规约束的金融场景中形成“能自动化的尽量自动化、需裁量的坚决留给人”的责任清晰路径;其 L1—L4 自主性分级亦可为其他高风险领域的负责任 AI 应用提供参照。其二(监管合规方面),智能体所基于的规则库与国家制度规定、标准、监管指引、检测认证目录等保持同步,使每条建议都可回溯至条文,契合中国抗量子密码标准正在制定、关键信息基础设施须使用合规密码产品的现实;在标准落地前,智能体的价值集中于密码资产测绘、密码敏捷性改造与迁移就绪度评估。其三(可推广性方面),本文不依赖生产环境的评测设计具有可迁移性,同样适用于能源、政务、电信等其他关键信息基础设施的抗量子迁移。

本文的局限在于第 5 章的量化评测尚未实施,能力划分与失效模式分析仍以案例与机制推演为主。后续工作将沿 3 条线索展开:一是落地第 5 章所述的标注评测,形成可复现基准并与关键字/正则扫描等工具对照;二是研究属性判定中不确定性的量化表达;三是探索跨机构的密码物料清单共享与供应链协同,将智能体应用从“迁移点发现与适配建议”延伸到系统全生命周期的密码敏捷性治理。

参考文献(References)

- [1] Shor P. Algorithms for quantum computation: Discrete logarithms and factoring[C]//Proceedings of 35th Annual Symposium on Foundations of Computer Science. Washington: IEEE Computer Society, 1994: 124–134.
- [2] Chen L, Jordan S, Liu Y K, et al. Report on post-quantum cryptography: NISTIR 8105[R]. Gaithersburg: NIST, 2016.
- [3] Bernstein D J, Lange T. Post-quantum cryptography[J]. *Nature*, 2017, 549(7671): 188–194.
- [4] Mosca M, Piani M. Quantum threat timeline report 2025[R]. Toronto: Global Risk Institute, 2026.
- [5] NIS Cooperation Group. A coordinated implementation roadmap for the transition to post-quantum cryptography[R]. Brussels: European Commission, 2025.
- [6] G7 Cyber Expert Group. G7 Cyber Expert Group statement on advancing a coordinated roadmap for the transition to post-quantum cryptography in the financial sector[R]. Washington, DC: U. S. Department of the Treasury, 2026.
- [7] Auer R, Dodson D, Dupont A, et al. Quantum-readiness for the financial system: a roadmap: BIS Papers No 158[R]. Basel: Bank for International Settlements, 2025.
- [8] National Institute of Standards and Technology. NIST releases first 3 finalized post-quantum encryption standards[EB/OL]. (2024–08–13) [2026–04–25]. <https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards>.
- [9] National Institute of Standards and Technology (NIST). NIST selects HQC as fifth algorithm for post-quantum encryption [EB/OL]. (2025–03–11). <https://www.nist.gov/news-events/news/2025/03/nist-selects-hqc-fifth-algorithm-post-quantum-encryption>.
- [10] Mascelli J, Rodden M. "Harvest now decrypt later": Examining post-quantum cryptography and the data privacy risks for distributed ledger networks: Finance and Economics Discussion Series 2025–093[R]. Washington, DC: Board of Governors of the Federal Reserve System, 2025.
- [11] Monetary Authority of Singapore, DBS Bank, HSBC, et al. Quantum key distribution proof-of-concept sandbox for the financial sector[R]. Singapore: Monetary Authority of Singapore, 2025.
- [12] Hong Kong Monetary Authority. The HKMA unveils "Fintech 2030" at the Hong Kong FinTech Week 2025 [EB/OL]. (2025–11–03) [2026–08–16]. <https://www.hkma.gov.hk/eng/news-and-media/press-releases/2025/11/20251103-3/>.
- [13] Bank for International Settlements. Project Leap phase 2: quantum-proofing payment systems[R]. Basel: BIS Innovation Hub, 2025(OTHP107).

- [14] National Institute of Standards and Technology. FIPS 203: Module–lattice–based key–encapsulation mechanism standard; FIPS 204: Module–lattice–based digital signature standard; FIPS 205: Stateless Hash–based digital signature standard[S]. Gaithersburg: NIST, 2024.
- [15] National Security Agency. The Commercial National Security Algorithm Suite 2.0 and quantum computing FAQ: version 2.1[R]. Fort Meade: NSA, 2024.
- [16] 全国人民代表大会常务委员会. 中华人民共和国密码法[Z]. 北京: 全国人民代表大会常务委员会, 2019.
- [17] 国务院. 商用密码管理条例[Z]. 北京: 中华人民共和国国务院, 2023.
- [18] 商用密码标准研究院. 关于开展新一代商用密码算法征集活动的公告[EB/OL]. (2025–02–05) [2026–08–16]. <https://www.niccs.org.cn/symmbzyjy/tzgg/pc/content/19374229>
- [19] 江苏省金融学会. 银行业抗量子计算密码迁移管理指南: T/JPSFFB 001–2026[S]. 南京: 江苏省金融学会, 2026.
- [20] Russell S J, Norvig P. Artificial intelligence: A modern approach[M]. 4th ed. Hoboken: Pearson, 2021.
- [21] National Institute of Standards and Technology, National Cybersecurity Center of Excellence. Migration to post–quantum cryptography: Preparation for considering the implementation and adoption of quantum safe cryptography: NIST SP 1800–38 (initial preliminary draft)[R]. Gaithersburg: NIST, 2023.
- [22] Ecma International. CycloneDX bill of materials specification: ECMA–424[S]. 2nd ed. Geneva: Ecma International, 2025.

AI–Agent–Assisted post–quantum cryptography migration in finance: Migration point discovery and adaptation recommendation

LI Yi¹, DING Jintai^{2,3}, LIU Rui³

1. Research Institute of the People's Bank of China, Beijing 100800, China

2. School of Mathematics and Physics, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China

3. Post–Quantum Migration Interdisciplinary Laboratory (PQC–X), Xi'an Jiaotong–Liverpool University, Suzhou 215123, China

Abstract The rapid advancement of quantum computing poses fundamental and systemic threats to the public–key cryptography underpinning financial market infrastructures. Although regulatory authorities across jurisdictions have issued policy frameworks for post–quantum cryptography (PQC) migration, practical progress remains slow. This paper argues that the bottleneck is not algorithm choice: the standardized PQC algorithms and their parameter sets are few in number and are effectively fixed by standards and regulatory policy, leaving little decision space. The real bottleneck is the asymmetry between the scale of legacy systems and the supply of cryptographic expertise. A large financial institution may contain tens of thousands of cryptographic call sites scattered across hundreds of heterogeneous legacy systems and vendor–delivered code, while engineers competent in PQC remain scarce, making it unrealistic to staff every legacy unit with such experts. Accordingly, this paper redefines the role of AI agents in migration as knowledge delivery to legacy system owners who lack a PQC background: discovering cryptographic call sites in code, configuration and third–party components; determining their business semantics and applicable compliance requirements; generating adaptation recommendations and remediation plans traceable to normative provisions; and leaving decision authority with human reviewers. Four capabilities are proposed – discovery, determination, generation and verification – together with an autonomy grading scheme. Five failure modes specific to this positioning are analyzed: missed migration points, attribute misjudgment, non–compliant recommendations, stale rule bases and over–trust. Using an open–source system with financial characteristics as a case, the paper further examines the actual scale and difficulty of migration point discovery, and presents an evaluation design that requires no production environment.

Keywords AI agent; post–quantum cryptography migration; cryptographic asset discovery; cryptographic agility; financial industry ●



(责任编辑 王志敏)