密码服务资源按需高效调度方案

寇文龙[1,2]，张宇阳[1]，李凤华[1,2,3]，曹晓刚[2,3]，李佳旻[1]，王竹[2,3]，耿魁[2]

（1. 西安电子科技大学网络与信息安全学院，陕西 西安 710071；2. 中国科学院信息工程研究所，北京 100093；
3. 中国科学院大学网络空间安全学院，北京 100049）

摘 要：随着“互联网+”时代数据安全传输和细粒度认证需求的日益增长，各类应用对密码服务的使用愈发频繁，如何处理随机交叉且峰值差异大的密码服务请求逐渐成为制约互联网服务安全应用的瓶颈问题。基于此，提出了一种高效的密码服务资源按需调度方案，实现了密码服务资源的高效差异化动态按需调度。首先，提出基于优化熵值法的密码设备归一化评价模型，实现对密码服务能力的描述和动态监测；然后，提出一种适用于不同密码服务需求的密码作业调度策略，并结合密码资源重构机制，实现对密码资源的差异化配置与调度；最后，通过理论分析和在实际生产环境中部署测试的方法，对所提方案进行验证。理论分析和实验测试结果表明，所提方案能较好地保证密码服务调度的高效性和可靠性，加解密吞吐率可达 56 Gbit/s。

关键词：密码资源；按需调度；高吞吐量；评价模型

中图分类号：TN92

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2022092

On-demand and efficient scheduling scheme for cryptographic service resource

KOU Wenlong[1,2], ZHANG Yuyang[1], LI Fenghua[1,2,3], CAO Xiaogang[2,3], LI Jiamin[1], WANG Zhu[2,3], GENG Kui[2]

1. School of Cyber Engineering, Xidian University, Xi’an 710071, China
2. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
3. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: With the increasing demand for data security transmission and fine-grained authentication in the “Internet+”era, the use of cryptographic services in various applications is becoming more frequent. How to deal with random cross and large peak difference cryptographic service requests has gradually become a bottleneck problem restricting the security application of various Internet services. Based on this, an efficient on-demand scheduling scheme for cryptographic service resources was proposed and the efficient differentiated dynamic on-demand scheduling of cryptographic service resources was realized. Firstly, the normalized evaluation model of cryptographic devices based on optimized entropy method was proposed to realize the description and dynamic monitoring of cryptographic service capabilities. Furthermore, a cryptographic job scheduling strategy suitable for different cryptographic service requirements was proposed, and the differential configuration and scheduling of cryptographic resources were realized by combining the cryptographic resource reconstruction mechanism. Finally, the proposed scheme was verified by theoretical analysis and deployment test in actual production environment. Theoretical analysis and experimental results show that the proposed scheme can better ensure the efficiency and reliability of cryptographic service scheduling, and the encryption and decryption throughput rate can reach 56 Gbit/s.

Keywords: cryptographic resource, demand-based resource scheduling, high throughput, evaluation model

收稿日期：2022−01−31；**修回日期：**2022−03−12

通信作者：耿魁，gengkui@iie.ac.cn

基金项目：国家重点研发计划基金资助项目（No.2018YFB0803903）；陕西省重点研发计划基金资助项目（No.2019ZDLGY12-09）

Foundation Items: The National Key Research and Development Program of China (No.2018YFB0803903), The Key Research and Development Program of Shannxi Province (No.2019ZDLGY12-09)

0　引言

随着“互联网+”战略的不断推进，互联网经济与各行各业不断融合，各种新业态和新型服务模式不断涌现，尤其是云服务、电子商务、电子支付、共享经济、大数据中心、社交网络的迅猛发展，直接导致了用户数量和业务种类均大幅增长。

随着企业上云和数字化转型升级的深化，数据正在成为企业的核心资产之一，在生产过程中发挥越来越大的价值。数据安全也成为广大企业和云服务商共同关注的话题之一。数据显示，2020—2021 年数据泄露的平均总成本增加了 10%，业务损失占数据泄露总成本的 38%。采用人工智能、安全分析和加密是降低数据泄露总成本的三大因素，与没有大量使用这些工具的公司相比，它们为公司节省了 125 万～149 万美元[1]。Facebook 被曝涉及数亿用户手机号码数据泄露，事件原因是服务器没有设置保护措施[2]。根据 2018 年全国性国密普查结果显示，所选取的等保三级以上系统中使用密码服务的仅占 24.77%[3]。因此，业务系统的安全性直接影响着业务的发展，密码服务是保障业务安全的核心基础。

互联网服务面临着业务类型多样、资源需求个性化、服务多轮交互、在线链接高并发、请求随机交叉、峰值差异大等巨大挑战。因此亟须实现对各类服务资源的高效差异化管理与动态利用，形成服务资源按需供给能力。在业务系统实现按需服务的同时，密码服务也应根据业务系统的特点和需求进行动态配置、管理和调度，实现提供密码按需服务的能力，从而满足互联网服务峰值差异大、高并发、需求个性化的业务特点和需求。然而，现有的密码系统、密码设备、各类密码计算资源尚不能根据需求进行动态配置、管理和调度，不具备满足差异化动态按需密码服务的能力。

针对上述问题，本文提出了一种高效的密码作业按需调度方案，为接入服务体系的用户和设备提供动态可扩展的密码服务资源，为密码服务设备提供设备运行监测、负载均衡、作业高效流转等服务，优化密码设备的服务供给能力，保障密码服务的高效性和可靠性。本文的主要贡献如下。

1) 提出了基于优化熵值法的密码设备归一化评价模型。该模型通过分析密码设备的运行状态和服务供给能力，可客观公平地反映密码设备之间的差异，实现对密码服务能力的描述和动态监测，为保障密码服务质量的可靠性和实现密码服务能力的可描述性提供支撑。

2) 提出了适用于不同密码服务需求的密码作业调度策略。该策略可适应密码服务在服务质量、服务效率等方面的差异化需求，并结合密码资源重构策略，实现对密码资源的差异化配置与调度，满足典型密码服务场景的服务需求。

3) 实现了高性能的密码按需服务系统，并在实际生产环境中部署和测试。测试结果表明，所提方案在满足差异化密码服务需求的同时，能较好地保证密码服务调度的高效性和可靠性，加解密吞吐率可达 56 Gbit/s。

1　相关工作

本节主要从设备服务能力评价体系和资源调度算法两方面入手，对相关研究进行论述。

1.1　设备服务能力评价体系

设备服务能力评价是指通过获取设备部分或者全部所需的指标数据，并根据算法对指标数据进行处理，得到一个评价结果。

焦扬等[4]提出了基于马尔可夫链的六维服务质量（QoS, quality of service）评价体系，有效满足了云环境中 QoS 可靠性评估需求。但是该方案没有考虑不同服务之间的关联关系，无法对多种服务组合的场景进行服务能力评价。Wang 等[5]提出了一种基于信任和隐私感知的云服务评估模型，引入客户满意度等参数来动态更新信任值，确保实际的服务质量。然而，该方案仅依靠信任值来度量服务质量，缺乏其他如用户满意度等必要的参考信息，影响服务质量评价的准确性。Jiang 等[6]提出了一种基于云模型的综合服务质量定性评价方法，利用高斯云变换将属于不同指标的概念组合在一起来评价服务质量。但是由于对指标概念划分粒度较粗，可能会导致指标概念被不正确划分，从而影响服务质量评价的准确性。

林闯等[7]认为 QoS 的评价指标仅反映了技术层面的性能，而忽略了用户的主观因素，提出用户体验质量（QoE, quality of experience）的概念，综合考虑服务、用户和环境层面的影响因素，直接反映了用户对服务的认可程度。阳小兰等[8]提出了一种基于博弈优化调度的筛选加权评价方法，通过用户 QoE 评价等多个指标的博弈，能够准确地评价资源

调度的有效性和准确性。但是该方案缺乏对服务质量的详细描述，导致资源匹配精度不高。Li 等[9]采用主观检验方法归纳出一个多 QoS 度量的多元函数，来评估视频流业务的整体 QoS。Song 等[10]提出了一种以用户为中心的客观 QoE 评价模型，综合考虑了技术、内容、上下文、用户等影响因素。文献[9-10]所提方案过于依赖用户的主观体验，会降低服务质量评估的准确性。

现有的设备服务能力评价方法都是根据部分或全部 QoS、QoE 或者其他参数指标，针对一种特定的服务进行评价，缺少对不同设备提供若干种差异化服务组合能力的评价，尤其是密码服务，通常包括签名、验签、哈希、加密、解密等多种不同的密码算法运算功能。

1.2 资源调度算法

资源调度算法一直是学术界的研究热点，陆续形成了几类典型的调度方案和架构。

Prassanna 等[11]针对突发性工作负载的问题，提出了基于阈值的多目标模因优化轮询调度算法。但该方案需要对其平衡性和负载的分散性进行优化以提高算法的效率。Patel 等[12]提出了一种增强型负载均衡 Min-Min 算法，将完成时间最长的任务分配给适当的资源。Grandl 等[13]提出了 Altruistic 调度方法，优先考虑将部分完成时间长的任务抢占完成时间短的任务的资源。文献[12-13]所提方案结合了 Min-Min 算法和 Max-Min 算法的优点，但是动态调整能力不足，无法根据任务和负载的实时变化来调整调度方案。苏命峰等[14]基于多重贪婪算法设计了 MQoS 云资源调度算法，综合量化云任务执行时间等 5 个指标，取得云计算系统最大效用。但该方案中指标权重是一个预设值，不能根据服务需求进行动态调整。马小晋等[15]提出一种基于改进模拟退火算法的虚拟机调度优化方法，在资源利用率、执行成本、负载均衡 3 个方面达到平衡。但该方案主要支持的是云中科学应用的实现，限制了算法的应用范围。Jana 等[16]提出了一种改进的粒子群优化算法，专注于调度算法中平均调度长度和执行成功率 2 个关键目标，有效地提高了云计算的业务性能。但该方案成本较高，需要耗费大量的内存，并且在某些情况下会关闭部分关键的虚拟机。Jiang 等[17]提出了一种基于动态一致性哈希的集群负载均衡算法，有效地降低了服务响应时间，提高了系统的吞吐量。然而，该方案中性能权重、负载参数等指标需要根据实验数据进行设定，不利于算法在其他应用场景下进行扩展。

李莉等[18]设计了一种同时支持非关联任务和关联任务的负载均衡作业调度算法，能实现高速的密码处理吞吐率。然而，该方案从密码算法层面对作业调度进行优化，缺少对用户密码服务需求的考虑。Li 等[19]提出了一种基于虚拟化技术的密码资源管理框架，给出了一种加密服务虚拟机的动态迁移方法，可以实现密码服务虚拟机的调度和迁移。但是该方案是对静态密码服务资源的调度，没有考虑密码服务资源的动态变化，不能很好地满足用户差异化的密码服务需求。

现有的资源调度算法对密码资源指标的影响考虑较少，而密码服务调度需要以密码资源指标为主，综合考虑网络带宽、系统负载、内存使用情况等多种影响因素。此外，还需要根据密码服务需求的变化动态调整调度策略，保证密码服务的高效性和可靠性。

2 密码服务调度系统模型

为了提高整个密码设备的服务性能和服务质量，在业务系统提供服务时，密码资源应根据业务系统的特点和需求进行动态配置、管理和调度。如图 1 所示，典型的密码服务模型包括密码服务应用、密码服务调度和密码设施。其中，密码服务应用主要包括各种使用密码服务的应用和系统；密码服务调度主要包括运行状态管理与监控、密码服务需求分析与调度和密码设备配置管理；密码设施包括密钥管理、密码设备集群和密码设备重构管理。密码服务以密码设施为基础，能提供基本的密码算法运算功能，密码服务调度在密码设施的基础上，将面向应用场景的密码功能进行整合，对密码服务应用提供易部署、易使用的密码服务，同时结合运行状态管理与监控，合理地调度密码设备集群中的密码设备，提高密码服务的服务质量和效率。

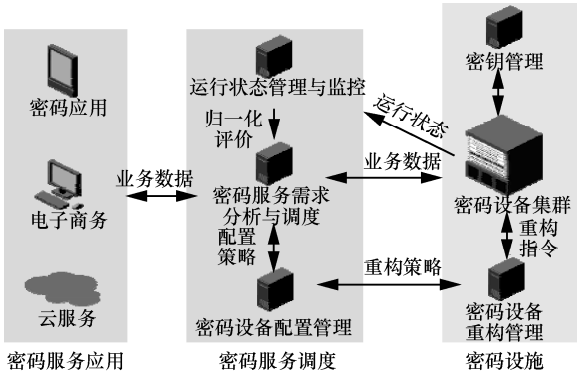


图 1 密码服务模型

本文着重对密码服务系统中调度系统展开研究，密码服务调度系统模型如图 2 所示。其中，密码服务需求分析对密码服务需求进行解析，结合密码设备的属性、运行状态以及资源使用情况等信息，对密码设备进行归一化评价，在评价结果的基础上通过计算密码设备的负载距离来生成密码作业调度策略，如果需要重构，则生成密码作业迁移策略和密码设备配置策略，并反馈密码服务供给能力；密码作业管理根据密码作业迁移策略和密码作业调度策略对密码作业进行管理。

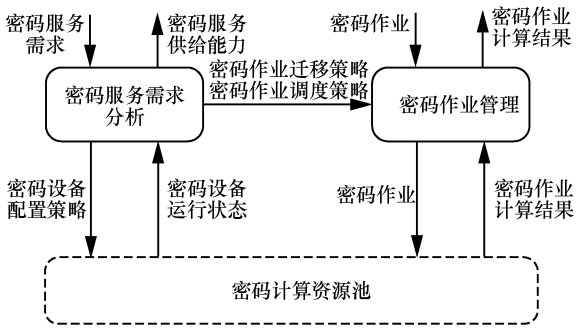


图 2　密码服务调度系统模型

3　基于优化熵值法的密码设备评价方法

为了客观公平地反映密码设备之间的差异，实现对密码服务能力的描述，为密码服务调度提供支撑，本节提出一种基于优化熵值法的密码设备评价方法。

首先，定义 m 个密码设备的集合为 $C=\{c_1,c_2,\cdots,c_m\}$，在 n 项资源评价指标下的原始采集数据矩阵为

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

其中，x_{ij} 表示第 i 个密码设备中第 j 项资源的资源使用率，$\boldsymbol{X}_i=(x_{1i},x_{2i},\cdots,x_{mi})^{\mathrm{T}}$，$i=1,\cdots,n$ 表示第 i 项资源在 m 个密码设备中的资源使用率。

随后，对 $\boldsymbol{X}_i$ 进行线性归一化处理

$$y_{ij}=\frac{x_{ij}-\min\{x_j\}}{\max\{x_j\}-\min\{x_j\}},i=1,\cdots,m$$

得到标准化数据矩阵为

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix},y_{ij}\in[0,1]$$

其中，在第 j 项资源使用率指标下，第 i 个密码设备的特征比重为 $b_{ij}=\dfrac{y_{ij}}{\sum_{i=1}^{m}y_{ij}},0\leqslant b_{ij}<1$，其相应的熵值为 $e_j=-K\sum_{i=1}^{m}b_{ij}\ln b_{ij}$。其中，$K=\dfrac{1}{\ln m}$ 为调节系数。

考虑到在计算密码设备的特征比重时会出现 $b_{ij}=0$ 的情况，进而导致计算熵值时出现 $\ln 0$，影响后续计算结果，因此，在计算密码设备的特征比重时添加一个平滑项，令第 j 项资源使用率指标下，第 i 个密码设备的特征比重为 $b_{ij}=\dfrac{y_{ij}+\lambda}{\sum_{i=1}^{m}y_{ij}+2m\lambda},0\leqslant b_{ij}<1$。其中，$\lambda\geqslant 0$。

接着，计算信息熵冗余度 $d_j=1-e_j$，进而得到各项资源指标的权重为 $w_j=\dfrac{d_j}{d_1+\cdots+d_n},j=1,\cdots,n$，且 $\sum_{j=1}^{n}w_j=1$。

最后，计算当前密码设备的服务能力评价系数 $S_i=\sum_{j=1}^{n}w_jb_{ij},\ i=1,\cdots,m$。

根据密码设备的自身特点，本文选择密码设备的密码运算资源使用率和网络带宽资源利用率作为密码服务资源的评价指标。其中，密码运算资源使用率以密码设备的 I/O 吞吐率为衡量指标，包括各种密码算法的运算速度；网络带宽资源利用率则以密码设备整体的 I/O 吞吐率和最大网络带宽的比值作为衡量指标。

4　高效密码作业调度策略

针对在线业务高效调度、资源动态配置的需求，在基于优化熵值法的归一化评价方法的基础上，综合生成密码作业调度策略。首先通过密码设备评价方法得出服务能力评价系数，在此基础上根据密码服务请求期望和当前密码服务资源使用情况，生成密码作业调度策略，同时配合密码设备动态重构，扩展了密码服务能力，解决了密码资源的动态扩展和单一设备利用率不高的问题。

4.1 按需服务的密码作业调度策略

密码按需服务调度策略首先对密码服务需求进行解析，将密码服务需求转化为密码服务应用对完成密码作业所需密码资源的期望，结合当前密码设备运行状态，对能否满足密码服务需求进行评估。若满足需求则生成密码作业调度策略，计算密码设备的期望负载距离，将密码作业调度到具体的密码设备；若不满足需求则生成密码设备配置策略，对指定的密码设备进行重构，待重构完成后，更新密码设备资源属性，生成新的密码作业调度策略。

密码服务需求由密码服务需求标识符、密码服务类型、密码服务有效期、密码服务最大速率、密码服务最小速率、密码算法标识符、密码算法参数、工作模式等参数组成。各项参数描述如下。

1) 密码服务需求标识符。密码服务请求的唯一标识符。

2) 密码服务类型。密码服务请求的具体服务类型，为加密、解密、签名、验签、哈希等一种或几种服务的组合。

3) 密码服务有效期。密码服务请求的密码服务时间，可以表示为起始时间和终止时间，或者起始时间和持续时间。

4) 密码服务最大速率。密码服务请求的最大需求，包括最大加密速度（bit/s 或 byte/s）、最大解密速度（bit/s 或 byte/s）、最大签名速度（次/秒）、最大验签速度（次/秒）、最大哈希速度（bit/s 或 byte/s）。

5) 密码服务最小速率。密码服务请求的最小需求，包括最小加密速度（bit/s 或 byte/s）、最小解密速度（bit/s 或 byte/s）、最小签名速度（次/秒）、最小验签速度（次/秒）、最小哈希速度（bit/s 或 byte/s）。

6) 密码算法标识符。密码服务请求所使用的算法类型。

7) 密码算法参数。密码算法的参数信息包括签名算法曲线参数和密钥长度、哈希算法分组长度、分组密码算法的密钥长度和分组长度等。

8) 工作模式。密码算法所用的模式，例如商密 SM4 算法包括电子密码模式（ECB, electronic codebook mode）、密码分组链接模式（CBC, cipher block chaining mode）、密文反馈模式（CFB, cipher feedback mode）、输出反馈模式（OFB, output feedback mode）和计数器模式（CTR, counter mode）5 种模式。

定义 E_i 为对第 i 种密码服务资源的需求期望，表示密码服务最大速率和密码服务最小速率的均值，采用此期望可以减少密码设备出现满载或过载的可能性，提高密码服务的可靠性。

定义 SAT_{ij} 为第 i 个密码设备中第 j 种密码服务资源的密码作业饱和度，x_{ij} 为当前第 i 个密码设备中第 j 种密码服务资源的速率，x_{ij}^{best} 为当前第 i 个密码设备中第 j 种密码服务资源的理论最大速率，则 SAT_{ij} 可表示为

$$\mathrm{SAT}_{ij}=\frac{E_j+x_{ij}}{x_{ij}^{\mathrm{best}}},\ i=1,\cdots,m,j=1,\cdots,n$$

定义 a_j 为隶属函数，表示为

$$a_j=\begin{cases}1, & \text{任务需求含有该资源}\\ 0, & \text{任务需求不含该资源}\end{cases}$$

定义 dist_i 为密码设备的期望负载距离，表示为 $\mathrm{dist}_i=\sum_{j=1}^{n}a_j(1-\mathrm{SAT}_{ij})$。若 $\mathrm{dist}_i\geqslant 0$，则说明第 i 个密码设备满足密码服务需求，可将密码作业调度至该密码设备进行密码算法运算；若 $\mathrm{dist}_i<0$，则说明第 i 个密码设备无法满足密码服务需求。

按需服务的密码作业调度策略步骤如下。

Step1 解析密码服务请求，计算密码服务请求的需求期望，转到 Step2。

Step2 获取每个密码设备的密码运算资源使用率和网络带宽资源利用率，计算每个密码设备的服务能力评价系数并排序，将对应的密码设备索引值依次放入密码设备候选队列，转到 Step3。

Step3 判断密码设备候选队列是否还有可选择的密码设备，若没有则转到 Step4，否则计算当前选择的密码设备的期望负载距离，若期望负载距离大于或等于 0，则将该密码服务请求中的密码作业添加到当前密码设备的密码作业队列中，转到 Step5；若期望负载距离值小于 0，则当前密码设备无法满足密码服务需求，将该设备从密码设备候选队列中移除，转到 Step3。

Step4 生成密码设备配置策略，对指定的密码设备进行密码资源重构，待重构完成后，更新密码设备的密码资源配置，转到 Step2。

Step5 生成密码服务调度策略，转到 Step1。

4.2 算法复杂度分析

通过分析可知，密码服务资源按需调度方案的

算法复杂度主要由基于优化熵值法的密码设备归一化评价算法的复杂度和密码作业调度算法的复杂度两部分组成。调度方案整体的复杂度与密码设备的数量、密码资源评价指标项的数量和单位时间内密码服务请求的数量有关。

基于优化熵值法的密码设备归一化评价模型的算法复杂度主要为计算服务能力评价系数的复杂度，也就是模型中数据矩阵的量级，为 $O(m+n)$，其中 m 为密码设备的数量，n 为资源评价指标的数量。密码作业调度算法需要对用户的期望和当前的密码计算资源进行匹配，对单个密码服务请求进行分配的算法复杂度为 $O(\log m)$，则密码作业调度算法的复杂度为 $O((k+m)\log m)$，其中 k 为单位时间内密码服务请求的数量。

在实际应用中，密码设备的数量总是远小于单位时间内密码服务请求的数量，密码资源评价指标的数量也有限，因此本文所提调度方案的算法复杂度在最优情况下为 $O(k)$，最差情况下为 $O(k\log m)$。

4.3　安全性分析

密码服务调度系统的安全性以密码计算资源池的安全性为基础。首先，密码服务需求数据和密码作业数据中不包含任何诸如密钥之类的敏感信息，保证了业务数据通信的安全性。其次，密码服务调度系统在和密码服务应用进行通信时，基于身份认证机制保证密码服务应用的合法性，防止非法或者恶意应用对密码服务系统的访问。最后，密码服务调度系统的需求解析和密码作业调度可通过不同的数据通路与密码服务应用通信，实现业务数据和请求数据的分离，数据安全性更高。

5　系统实现

5.1　系统设计

为验证本文所提密码服务资源按需调度方案的有效性，设计实现了密码按需服务原型系统，整体架构如图 3 所示。该系统对外提供签名、验签、摘要、加解密、密钥交换、证书操作等通用密码服务，并根据不同密码服务需求动态调度、管理密码设备。

具体来说，密码服务需求分析模块根据密码服务需求动态生成密码设备配置需求、密码作业调度策略，为密码计算单元柔性重构模块和密码作业管理模块提供支撑；密码作业管理模块根据密码作业、密码计算单元中密码资源的属性和使用情况，动态调度密码设备中的密码计算单元进行密码计算，实现密码作业虚拟化、按需调度与管理；密码设备运行状态管理模块根据密码作业调度信息和密码设备返回的密码设备运行状态，生成新的密码资源使用情况；密码设备配置管理模块对密码设备以及密码计算单元的各类属性和使用情况进行精细化、细粒度管理；密钥管理模块对整个密码服务系统的对称密钥、非对称密钥和证书进行管理；密码计算单元柔性重构模块根据密码设备配置管理模块下发的重构和配置指令对密码计算单元中的密码芯片、芯片中的块、知识产权（IP, intellectual property）核的数据传输带宽、数据缓冲区大小、密码算法类型、密码算法速率、密码算法参数、工作模式等相关属性进行细粒度的重构与配置。

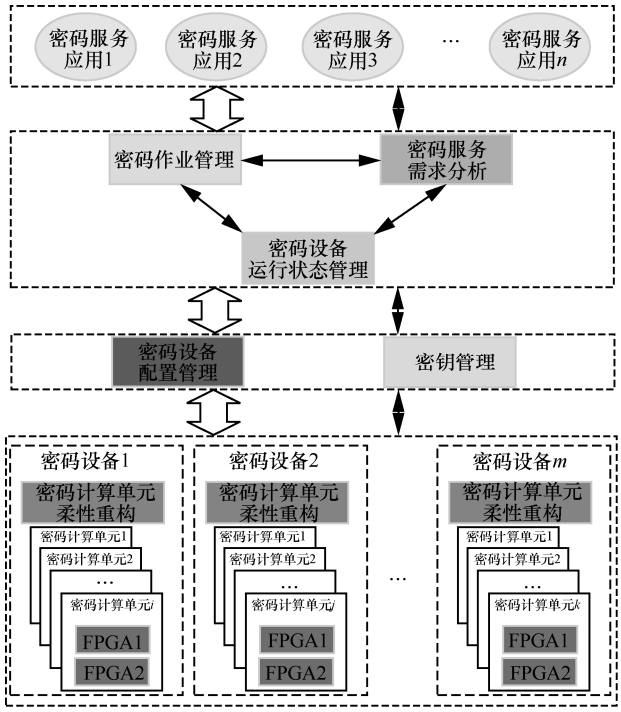


图 3　密码按需服务原型系统架构

5.2　密码服务资源按需调度算法实现

密码服务资源按需调度算法的主要流程如图 4 所示，结合密码服务需求分析、密码作业管理、密码设备配置管理和密码计算单元柔性重构等模块来生成密码服务调度策略。密码服务资源按需调度算法主要步骤说明如下。

Step1　密码服务需求分析模块接收来自密码应用提出的密码服务需求，转到 Step2。

Step2　密码设备运行状态管理模块获取各密码计算单元的运行状态、密码算法处理速率和网络吞

吐率等信息，根据本文提出的基于优化熵值法的归一化评价方法，计算密码计算单元的密码服务能力评价系数，转到 Step3。

Step3 对所有密码计算单元的服务能力评价系数排序，从服务能力评价系数最高的密码计算单元开始计算期望负载距离，转到 Step4。

Step4 根据密码计算单元的期望负载距离判断是否有密码计算单元满足需求，如果没有，则向密码应用返回无法满足需求，转到 Step1；否则向密码应用返回满足需求。然后继续判断是否需要重构，如果不需要重构，转到 Step6；如果需要重构，则首先生成密码作业迁移策略，并发送给密码作业管理单元，完成重构之前的密码作业迁移工作，转到 Step5。

Step5 根据新的密码设备配置生成密码资源配置策略，并发送给密码计算单元柔性重构模块，转到 Step6。

Step6 生成密码作业调度策略，转到 Step1。

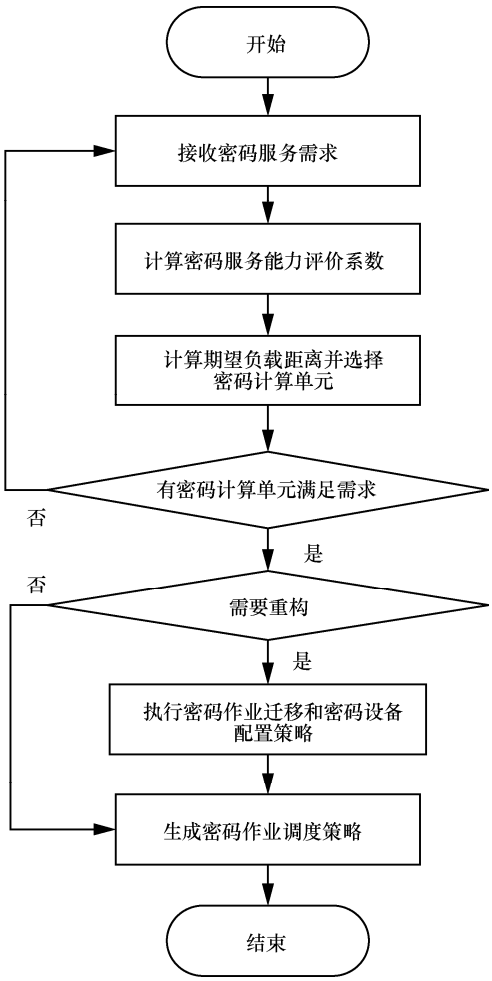


图 4 密码服务资源按需调度算法的主要流程

5.3 密码计算单元柔性重构实现

密码计算单元柔性重构流程如图 5 所示。当密码服务需求分析模块判断当前的密码资源配置无法满足密码服务需求时，会生成密码设备配置策略。其中，密码计算单元柔性重构算法如算法 1 所示，具体步骤说明如下。

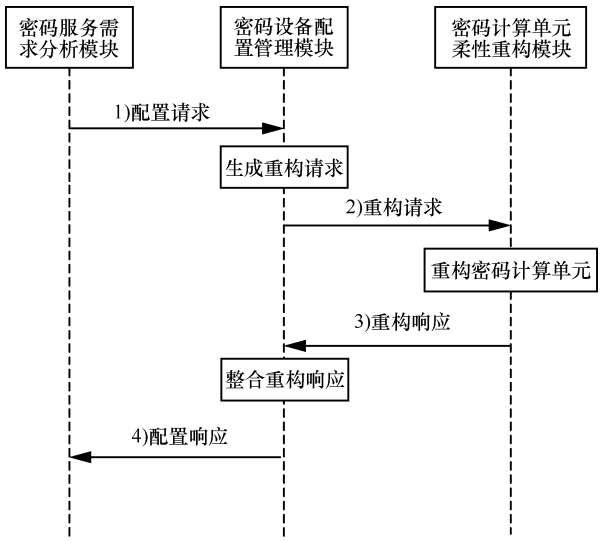


图 5 密码计算单元柔性重构流程

算法 1 密码计算单元柔性重构算法

输入 密码设备配置策略 CONF

输出 密码计算单元重构指令 REBUILD

1) 初始化密码计算单元重构指 REBUILD = ∅
2) for each c in CONF do;
3) 获取密码计算单元 c 对应的密码设备配置类型 cfgType
4) create rebuild
5) rebuild . $c = c$
6) rebuild .cfgType=cfgType
7) REBUILD ← REBUILD ⋃ rebuild
8) end for
9) return REBUILD

Step1 密码服务需求分析模块在解析密码应用的密码服务需求时，如果现有的密码资源配置不能满足需求，则根据密码服务需求的密码算法类型、密码算法参数、密码算法性能要求等指标生成密码设备配置策略，并发送给密码设备配置管理模块，转到 Step2。

Step2 密码服务配置管理模块解析密码设备配置策略，根据配置策略中的密码算法类型、密码算

法性能要求等指标生成重构指令，并将重构指令发给指定的密码计算单元柔性重构模块，转到 Step3。

Step3 密码计算单元柔性重构模块在收到重构指令之后，根据重构指令的要求对密码计算单元进行重构，并等待重构完成，转到 Step4。

Step4 密码计算单元柔性重构模块将重构结果返回给密码设备配置管理单元，转到 Step5。

Step5 密码设备配置管理单元将所有密码计算单元柔性重构模块返回的重构结果进行汇总并分析，生成新的密码设备配置信息，然后将新的密码设备配置信息返回给密码服务需求分析模块。

6　系统测试与分析

6.1　实验环境

实验环境拓扑如图 6 所示，包括应用服务器、调度服务器和密码设备。其中，应用服务器共 5 台，模拟不同密码服务应用发送的密码服务请求；调度服务器 1 台，部署本文所提调度方案；密码设备 1 台，共 8 个现场可编程门阵列（FPGA, field programmable gate array）密码计算单元，提供密码算法运算功能。应用服务器和调度服务器之间通过 100 Gbit/s 交换机连接，调度服务器和密码设备之间通过 100 Gbit/s 电缆线连接。

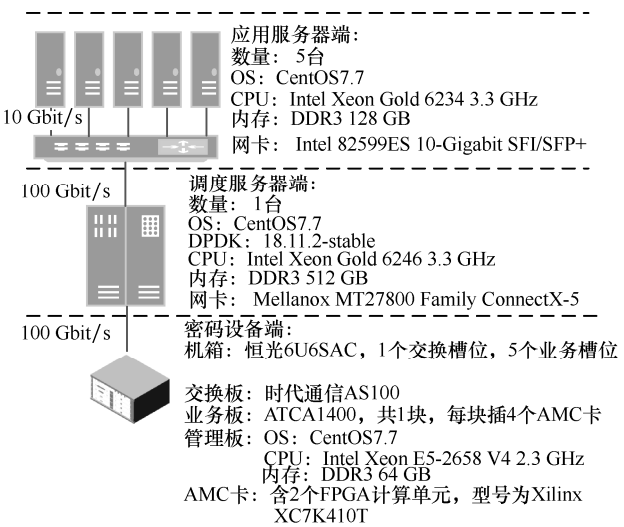


图 6　实验环境拓扑

在密码算法方面，本文采用商用密码算法 SM2、SM3、SM4，提供 SM2 算法签名、SM2 算法验签、SM3 算法哈希和 SM4 算法 ECB 模式加解密等密码运算服务。在 FPGA 上实现 3 种密码设备配置，每种配置的算法 IP 核个数如表 1 所示。

表 1　3 种密码设备配置的算法 IP 核个数

配置	SM2 IP 核/个	SM3 IP 核/个	SM4 IP 核/个
配置 A	4	3	1
配置 B	6	1	1
配置 C	0	8	1

本文所实现的 SM2、SM3、SM4 算法的单核性能以及 3 种密码设备配置的性能如表 2 所示。

表 2　SM2、SM3、SM4 算法的单核性能以及 3 种密码设备配置的性能

配置	SM2 签名性能/（次·秒$^{-1}$）	SM2 验签性能/（次·秒$^{-1}$）	SM3 哈希性能/(Gbit·s^{-1})	SM4ECB 性能/(Gbit·s^{-1})
单核性能	5 000	2 000	1.0	7.0
配置 A	20 000	8 000	3.0	7.0
配置 B	30 000	12 000	1.0	7.0
配置 C	0	0	7.0	7.0

6.2　测试及分析结果

应用服务器发送密码服务请求至调度服务器，经过调度算法的计算，将密码作业调度至密码设备中的 FPGA 密码计算单元，待密码作业完成后将结果返回给应用服务器。每个应用服务器发送的请求数从 20 增加到 100，调度服务器端分别采用文献[12]提出的增强型负载均衡 Min-Min 算法和文献[17]提出的基于动态一致性哈希的集群负载均衡算法与本文所提调度算法进行对比，分别测试 3 种调度算法的密码作业最大完成时间、单位时间可服务请求数量和 FPGA 密码计算单元平均负载。3 种调度算法进行对比实验时，应用服务器发送的密码服务请求数量和种类完全相同，并运行多次取平均值。

应用服务器发送的密码服务请求从以下 3 个维度出发进行设计。

1) 密码服务请求中的密码服务类型从实际应用场景出发，包含一种或多种密码服务，以适用不同场景的密码服务需求。例如，SM2 算法签名和 SM3 算法哈希适用于电子发票或电子签章等应用；SM4 算法 ECB 模式加解密，适用于视频加密等应用。

2) 密码服务请求中的密码服务最大速率和最小速率，根据测试环境中的密码设备配置的性能以及 FPGA 密码计算单元的数量等因素进行设计，范围覆盖 SM2 算法签名、SM2 算法验签、SM3 算法哈希和 SM4 算法 ECB 加解密等密码服务性能的最大值，以检验调度算法中密码作业迁移和密码设备重构算法的可靠性和效率。

3) 密码服务请求中密码服务有效时长从分钟到小时不等，以检验调度算法的稳定性。

随着密码服务请求数量的增加，单个密码作业的最大完成时间的变化情况如图 7 所示。从图 7 中可以看出，随着密码服务请求数量的增加，3 种调度算法的密码作业最大完成时间都在增加，当密码服务请求数量较少时，3 种调度算法的差异不太明显，但是随着密码服务请求数量的增加，FPGA 计算单元的负载逐渐增大，另外 2 种调度算法由于不考虑密码作业迁移和 FPGA 计算单元动态配置，密码作业排队时间增加显著，与本文调度算法的差距越来越大。

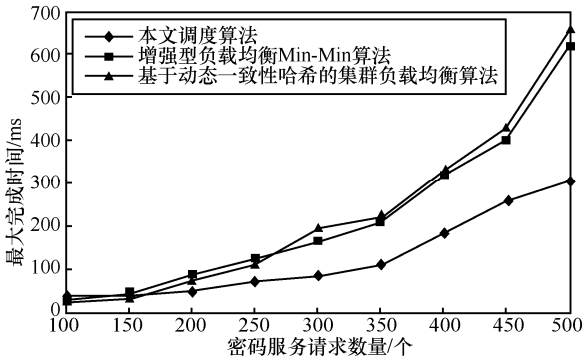


图 7 单个密码作业的最大完成时间的变化情况

随着密码服务请求数量的增加，单位时间可服务请求数量的变化情况如图 8 所示。从图 8 中可以看出，随着密码服务请求数量的增加，3 种调度算法的单位时间可服务请求数量都在增加，当密码服务请求数量较少时，3 种调度算法的差异不太明显，都能够满足大部分的密码服务请求，但是随着密码服务请求数量的增加，3 种调度算法的单位时间可服务请求数量均达到峰值，由于本文调度算法实现了密码作业迁移和 FPGA 计算单元动态配置，使单位时间可服务请求数量要高于另外 2 种调度算法。

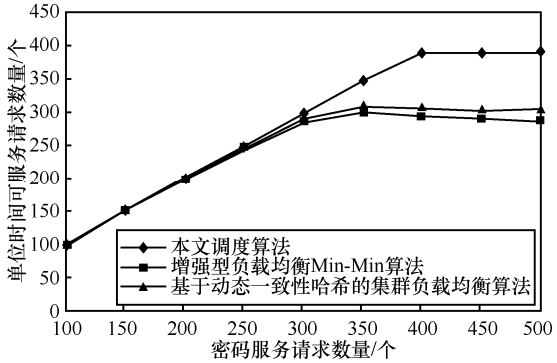


图 8 单位时间可服务请求数量的变化情况

本文调度算法、增强型负载均衡 Min-Min 算法和基于动态一致性哈希的集群负载均衡算法随着密码服务请求数量的增加 FPGA 负载率的变化情况如图 9～图 11 所示。从图 9 中可以看出，本文调度算法在尽量减少密码作业迁移和 FPGA 计算单元重构的前提下，将密码作业优先调度到同一个 FPGA 计算单元，因此在密码服务请求数量较少时只有一个 FPGA 计算单元有负载，并且随着密码服务请求数量的增加，同时工作的 FPGA 计算单元数量也随之增加。从图 10～图 11 中可以看出，其他 2 种算法的 FPGA 负载相对比较均衡，在密码服务请求数量较大的情况下，每个 FPGA 的负载均较高，当新的密码服务请求到来时，由于不考虑密码作业迁移和 FPGA 计算单元动态配置，FPGA 计算单元剩余计算能力不足以满足密码服务需求。

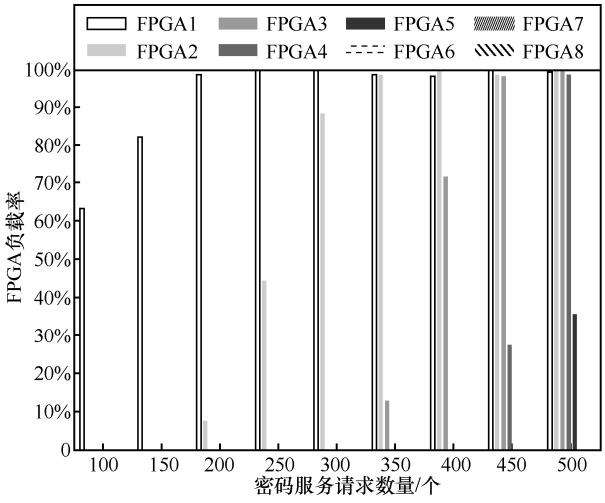


图 9 本文调度算法 FPGA 负载率

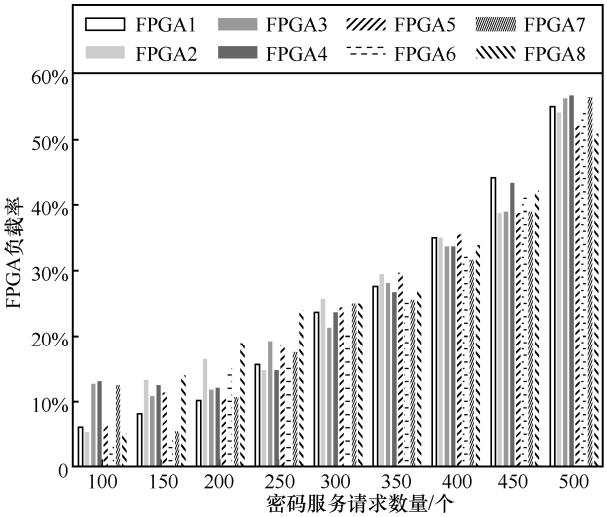


图 10 增强型负载均衡 Min-Min 算法 FPGA 负载率

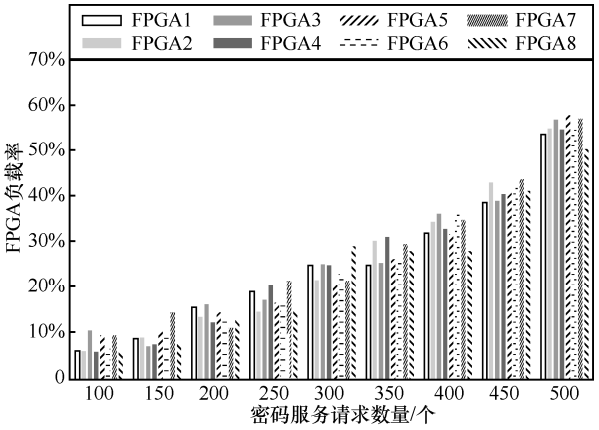


图 11　基于动态一致性哈希的集群负载均衡算法 FPGA 负载率

此外，在应用本文调度策略时，密码应用的密码服务请求在经过密码服务需求分析单元的计算后即生成密码作业调度策略，密码作业调度单元在接收到密码应用发送的密码作业时，根据密码作业调度策略，可直接将密码作业分配给对应的 FPGA 计算单元，时间和空间复杂度均为 $O(1)$，FPGA 计算单元 SM4 算法 ECB 模式加解密运算的吞吐率即密码作业调度的吞吐率。因此实验所用的 8 个 FPGA 计算单元均执行 SM4 算法 ECB 模式加解密运算，且负载率均接近于 100%时，每个 FPGA 计算单元的吞吐率均为 7 Gbit/s，密码服务系统整体吞吐率为 56 Gbit/s，实现了密码计算能力的线性增长。

从实验结果可以看出，本文所提密码按需服务方案在不增加新的硬件密码计算设备的基础上，通过调度服务系统与密码计算单元的柔性重构相结合的方式，能很好地满足不同场景下的密码服务需求。目前学术界和产业界推出的密码机，如 SJJ1524 服务器密码机只能通过增加硬件设备的方式来提高密码服务的性能，SJJ1601 云加密机虽然支持虚拟化，实现一台物理实体密码机提供多台虚拟密码机的服务，但是多台虚拟密码机的整体性能没有超过一台物理实体密码机的上限，不能满足不同场景的密码服务需求。

7　结束语

本文提出了一种高效的密码服务资源按需调度方案。通过使用基于优化熵值法的密码设备归一化评价模型实现对密码服务能力的描述和动态监测；同时，提出适用不同需求的密码作业调度策略，并结合密码资源重构策略，实现对密码资源的差异化配置与调度；实现将动态可扩展的密码服务资源提供给任何接入服务体系的用户和设备。在实际生产环境的部署和测试表明，所提方案在满足差异化的密码服务需求的同时，能较好地保证密码服务调度的高效性和可靠性，加解密吞吐率可达 56 Gbit/s。

参考文献：

[1] IBM. Cost of a data breach report 2021[R]. 2022.

[2] 万佳. Facebook 新漏洞：4.19 亿用户手机号码可公开访问，或遭遇重大安全风险[EB]. 2019.

[3] 云数据安全. 云上密码应用最佳实践[EB]. 2020.

[4] 焦扬, 陈喆, 梁员宁, 等. 基于马尔可夫过程的云服务组合 QoS 量化评估方法研究[J]. 计算机科学, 2015, 42(9): 127-133.
JIAO Y, CHEN Z, LIANG Y N, et al. Research on QoS quantitative evaluation method of cloud service composition based on Markov process[J]. Computer Science, 2015, 42(9): 127-133.

[5] WANG Y B, WEN J H, WANG X B, et al. Cloud service evaluation model based on trust and privacy-aware[J]. Optik, 2017, 134: 269-279.

[6] JIANG W X, GU C Z, WU J J. A quality-of-service evaluation method based on the cloud model for routing protocols in wireless sensor network[J]. International Journal of Distributed Sensor Networks, 2017: doi.org/10.1177/1550147717731247.

[7] 林闯, 胡杰, 孔祥震. 用户体验质量(QoE)的模型与评价方法综述[J]. 计算机学报, 2012, 35(1): 1-15.
LIN C, HU J, KONG X Z. Survey on models and evaluation of quality of experience[J]. Chinese Journal of Computers, 2012, 35(1): 1-15.

[8] 阳小兰, 钱程, 朱福喜. 基于云计算的大数据服务资源评价方法[J]. 计算机科学, 2018, 45(5): 295-299.
YANG X L, QIAN C, ZHU F X. Evaluation method of big data service resources based on cloud computing[J]. Computer Science, 2018, 45(5): 295-299.

[9] LI M F, LEE C Y. A cost-effective and real-time QoE evaluation method for multimedia streaming services[J]. Telecommunication Systems, 2015, 59(3): 317-327.

[10] SONG J R, YANG F Z, ZHOU Y C, et al. QoE evaluation of multimedia services based on audiovisual quality and user interest[J]. IEEE Transactions on Multimedia, 2016, 18(3): 444-457.

[11] PRASSANNA J, VENKATARAMAN N. Threshold based multi-objective memetic optimized round robin scheduling for resource efficient load balancing in cloud[J]. Mobile Networks and Applications, 2019, 24(4): 1214-1225.

[12] PATEL G, MEHTA R, BHOI U. Enhanced load balanced Min-Min algorithm for static meta task scheduling in cloud computing[J]. Procedia Computer Science, 2015, 57: 545-553.

[13] GRANDL R, CHOWDHURY M, AKELLA A, et al. Altruistic scheduling in multi-resource clusters[C]//Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Berkeley: USENIX Association, 2016: 65-80.

[14] 苏命峰, 王国军, 李仁发. 基于利益相关视角的多维 QoS 云资源调度方法[J]. 通信学报, 2019, 40(6): 102-115.
SU M F, WANG G J, LI R F. Multidimensional QoS cloud computing resource scheduling method based on stakeholder perspective[J]. Journal on Communications, 2019, 40(6): 102-115.

[15] 马小晋, 许华虎, 卞敏捷, 等. 基于改进模拟退火算法的虚拟机调度优化方法[J]. 通信学报, 2018, 39(S1): 278-287.
MA X J, XU H H, BIAN M J, et al. Virtual machine scheduling optimization method based on improved simulated annealing algorithm[J]. Journal on Communications, 2018, 39(S1): 278-287.

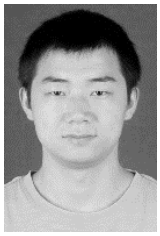
[16] JANA B, CHAKRABORTY M, MANDAL T. A task scheduling technique based on particle swarm optimization algorithm in cloud environment[C]//Soft Computing: Theories and Applications. Berlin: Springer, 2019: 525-536.

[17] JIANG X M, YANG H M, YANG Y, et al. Cluster load balancing algorithm based on dynamic consistent hash[J]. Journal of Intelligent & Fuzzy Systems, 2021, 41(3): 4461-4468.

[18] 李莉, 史国振, 耿魁, 等. 基于负载均衡的随机作业流密码服务调度算法[J]. 通信学报, 2018, 39(6): 11-19.
LI L, SHI G Z, GENG K, et al. Scheduling algorithm for stochastic job stream cipher service based on load balancing[J]. Journal on Communications, 2018, 39(6): 11-19.

[19] LI F L, JI H F, ZHOU H W, et al. A cryptographic resource management framework and dynamic migration method based on virtualization[C]//Proceedings of 2021 7th International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2021: 560-564.

[作者简介]



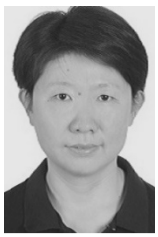
寇文龙（1990− ），男，河南许昌人，西安电子科技大学博士生，主要研究方向为信息安全。

张宇阳（1995− ），男，山东淄博人，西安电子科技大学硕士生，主要研究方向为电子与通信工程。

李凤华（1966− ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。

曹晓刚（1996− ），男，河北邢台人，中国科学院信息工程研究所博士生，主要研究方向为信息安全。

李佳旻（1993− ），男，山西吕梁人，西安电子科技大学博士生，主要研究方向为隐私计算、机器学习、联邦学习。



王竹（1972− ），女，山西太原人，博士，中国科学院信息工程研究所研究员，主要研究方向为密码理论与技术、安全协议。

耿魁（1989− ），男，湖北红安人，博士，中国科学院信息工程研究所高级工程师、硕士生导师，主要研究方向为网络安全、信息保护。