

2022 年腾讯基础平台技术犀牛鸟专项研究计划

研究命题

命题方向	命题题目
计算基础架构	命题一：基于云图形渲染的混合视频编码理论模型探索
	命题二：基于 AVS3 的实时通信编解码技术研究
	命题三：腾讯典型业务场景中 CPU benchmark 提取
	命题四：RISC-V 指令集优化研究
	命题五：面向触觉级 XR 的立体视频压缩传输联合优化技术研究
	命题六：负载感知型融合调度算法研究
	命题七：云原生容器热迁移技术研究
	命题八：下一代智能代码服务工蜂 2.0 关键技术研究
网络	命题九：面向超复杂异构网络的自适应传输控制关键技术研究
	命题十：面向动态异构网络的多路径调度关键技术研究
数据中心	命题十一：超大型数据中心设备健康管理方案研究
	命题十二：超大型数据中心制冷系统智能化节能方案研究

命题一：基于云图形渲染的混合视频编码理论模型探索

研究概要描述：

目前阻碍云图形商业化的两大因素是网络传输延迟和带宽，如果能在保证画质的情况下将视频流比特率大幅降低，那么可以有效的缓解网络传输延迟和带宽问题，使得基于云图形的各种商业场景成为可能。不同于传统静态图片流的视频编解码，基于云图形渲染的混合视频编码有着额外的图形信息输入，使得进一步降低带宽成为可能。图形信息输入包括：

- 1) 编码所要求的各类信息都可以由图形提供，包括感兴趣区域，深度场，运动估计等；
- 2) 图形本身提供了上采样的能力；
- 3) 图形渲染可以控制帧率，也可以预知第 N+1 的运动估计；
- 4) 图形输出的像素可以重排列，获得更大的编码压缩空间，甚至图形可以将输出分解成几部分，让每部分更加容易被压缩。

技术目标：

- 1) 提出一种新的服务于云图形的视频编解码的理论模型，其解码端可以适用于移动设备平台；
- 2) 视频编解码效率理论提升 20%。

实验资源：

- 1) 腾讯提供腾讯云计算资源和测试资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题二：基于 AVS3 的实时通信编解码技术研究

研究概要描述：

AVS3 视频编码标准是我国具备自主知识产权的信源编码标准，目前主要应用于广电领域，在实时通信场景还没有落地应用的情况。

目前影响 AVS3 在实时通信场景落地的主要问题有：

- 1) AVS3 标准的编码复杂度相比 H.265 高 20 倍左右，优化加速难度大；
- 2) 开源的 AVS3 编码器缺少可用于实时场景的优化算法；
- 3) AVS3 标准缺少一些如变分辨率参考等对实时通信场景友好的技术等。

技术目标：

- 1) 研究并实现适用于实时场景的 AVS3 优化算法，实现编码过程加速，可用于自研的 AVS3 实时编码器；
- 2) 基于 AVS3 标准研究变分辨率参考等适用于实时场景的技术，可用于下一代 AVS 标准提案。

实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供测试视频资源；
- 3) 腾讯提供专家指引和现场落地环境。

命题三：腾讯典型业务场景中 CPU benchmark 提取

研究概要描述：

大量研究显示数据中心的平均资源利用率从 10%到 50%不等，为了针对性的设计高性能 CPU、进一步提高资源利用率，需要了解如下问题：

- 1) 对数据中心业务进行详细分析，了解不同业务的具体性能特点并对其进行特征提取和性能分析；
- 2) 针对不同的业务生成相应的小型 benchmark，用于 CPU 设计时性能仿真和优化。

技术目标：

- 1) 分析腾讯典型业务场景，对各类业务进行特征分类；
- 2) 详细分析不同类型业务的特点，基于科学方法获取各类业务的特点并编写对应的 benchmark；
- 3) 对上述 benchmark 进行校准，确保其反应的特点与真实业务场景表现一致。

实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题四：RISC-V 指令集优化研究

研究概要描述：

随着 RISC-V 生态的发展壮大，基于 RISC-V 指令集的服务器级别高性能 CPU 也将会出现。为了设计实现基于 RISC-V 指令集的高性能 CPU，需要对现有指令集进行详细分析。目前正处于快速发展过程中的 RISC-V 指令集存在以下问题需要进行探索：

- 1) 现有指令集是否能满足高性能 CPU 对指令集的需求？是否需要新增针对高性能 CPU 的指令集？
- 2) 针对当前现有 RISC-V 指令集，如何在微结构层面进行优化设计，提高 CPU 的 IPC？

技术目标：

- 1) 对比当前 RISC-V 指令集和 ARM、X86 架构指令集在高性能 CPU 领域的优缺点，具体到增删某个具体指令对不同性能指标（IPC、可执行文件大小等）的影响；
- 2) 设计 micro-benchmark，逆向分析当前 ARM、X86 架构的高性能 CPU 对指令进行 crack、fuse 的方法；
- 3) 参考 ARM、X86 架构 CPU 对指令进行 crack、fuse 的方法，输出针对当前 RISC-V 指令集的 crack、fuse 推荐方法，并对其进行详细性能分析。

实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题五：面向触觉级 XR 的立体视频压缩传输联合优化技术研究

研究概要描述：

触觉级 XR (tactile X-Reality) 是全真互联的关键性技术之一，要求在 6DoF (degree of freedom)、8K 单眼视场 (FoV, field of view) 约束下，用户的运动-感知延迟 (MOP latency, motion-to-photon latency) 不超过 20ms。经典 VR (virtual reality) 方案通常采用成熟模块直接组合，如 MPEG-MIV^[1] 压缩+DASH SRD^[2] 方案，视频渲染、压缩、传输和头显之间隔离度较高，全链路未针对 6DoF 视频进行优化，用户的 MOP latency 受限于用户侧网络容量限制，从而极易诱发用户晕屏 (cybersickness)，导致各类 XR 应用商业化进程受阻。为此，学术界和工业界均在探索面向 6DoF 视频的全链路优化技术，提升传输质量，如基于超分辨率的 3D-Mesh 模型压缩重构^[3]，基于 FoV 角度预测的立体视频帧间压缩传输^[4]等。

为探索现有网络环境下 6DoF 单眼 8K FoV 视频的低延迟传输方案，使能触觉级 XR (tactile X-Reality)，拟开展面向触觉级 XR 的立体视频压缩传输联合优化技术研究。具体而言，针对腾讯提取的典型用户网络模态，研究高体验低延迟的 6DoF FoV 视频传输优化技术。特别是，寻找低成本高精度 FoV 预测方法，以及适应不同网络模态的 6DoF FoV 码率决策策略；同时，在 FoV 运动轨迹约束下，寻求 6DoF 视频帧的帧间最大相关一致性，改进 3D 视频渲染、压缩、传输、头显全链路流程，降低全链路 6DoF 视频传输带宽消耗；并在 5G SA 或 Wi-Fi 6+ 环境下，完成触觉级 XR (tactile X-Reality) 的原型系统可行性测试。

期望通过本项目得到一种低带宽低延迟的 6DoF FoV 视频传输方案，并探索 6DoF 8K 视频压缩传输联合优化的可行性与性能极限，同时在 5G SA 或 Wi-Fi 6+ 环境下初步实现触觉级 XR (tactile X-Reality) 的原型系统验证。

技术目标：

- 1) 提供一种低计算开销 FoV，结合自适应码率技术，相比非 FoV 方案，降低 6DoF 视频传输带宽消耗 85% 以上；
- 2) 提供一种新型的 6 DoF 视频压缩-传输联合优化方案，实现 3D 模型压缩-传输-超分重建，相比非超分重建方案，带宽消耗降低 20% 以上（实验室环境）；
- 3) 提供一套仿真测试系统，能在 5G SA 或 Wi-Fi 6+ 环境下，完成触觉级 XR (tactile X-Reality) 的原型验证，运动-感知延迟低于 20ms（实验室环境）。

实验资源：

- 1) 腾讯提供包含 Oculus 头显、Pico Neo 3 头显等设备的 VR 测试系统；
- 2) 腾讯提供包含典型网络模态的通讯网络 trace 数据集；
- 3) 腾讯提供包含软/硬件网损的超低延迟视频仿真验证系统。

参考文献：

- [1] J. M. Boyce *et al.*, "MPEG Immersive Video Coding Standard," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1521-1536, Sept. 2021.
- [2] <https://github.com/gpac/gpac/wiki/MPEG-DASH-SRD-and-HEVC-tiling-for-VR-videos>
- [3] Ed. Smith *et al.*, "Multi-view silhouette and depth decomposition for high resolution 3d object representation." in *Proceedings of NeurIPS*, Dec. 2018.

[4] A. Zhang *et al.*, “YuZu: Neural-Enhanced Volumetric Video Streaming,” in *Proceedings of the USENIX NSDI*, May 2022.

命题六：负载感知型融合调度算法研究

研究概要描述：

随着云计算轻量高密化和业务场景的复杂化，大规模集群资源效率和运营成本问题的严重程度和解决难度与日俱增。操作系统作为资源核心管理者，所提供的传统调度算法和策略受到基础性冲击。传统操作系统提供公平调度策略，分组/进程量级控制，和有限的优先级分类配置接口；云操作系统（云资源管理系统）依照用户定义需求进行资源配置，通常提供一定弹性的调度策略来提升资源池的效率。

整体上，无论是传统操作系统和云操作系统都缺乏理解业务负载的能力，无法建立业务负载对资源的实际消费模型，无法判断业务负载对资源满足的实际敏感度，进而造成资源共享度不够，可重用资源的低周转率，最终使得系统的整体吞吐受限，资源利用效率低。同时，对于多样化的资源类型，现有系统都是提供分类调度管理，没有结合业务对多样资源的需求及其关联性进行融合调度。比如，CPU 调度系统与内存调度系统的关联度就较低，可能出现两类资源互相等待的场景。

近年，业界出现打破公平性的调度算法，此类算法通常要求用户进行业务分类定义，算法在不同类别业务之间实现非公平性的调度，比如允许绝对抢占，长期饥饿等。这类算法的局限性在于实际业务场景中难以实现简单的分类，或者过多的分类之间难以实现准确合理的优先级区分。

技术目标：

- 1) 研究一种具备业务负载感知能力，并且能跨资源类调度的算法；
- 2) 在 linux 系统上实现一个针对 CPU/内存/IO/网络等资源的算法原型系统。

实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题七：云原生容器热迁移技术研究

研究概要描述：

随着云计算技术的不断发展，越来越多的用户使用云原生容器，而业务需求的容器资源规格不同，导致会存在资源碎片，造成平台成本的浪费；同时为了最大化资源的利用效能，资源会被共享，共享资源间存在竞争冲突问题。对此，我们需要对集群中的业务执行重调度从而实现集群维度的全局最优编排。

但实际场景中业务需求重调度不中止计算，因此我们需要借助于容器热迁移技术来实现任务中止前状态的保存与任务重调度后状态的恢复。本命题期望设计并实现一种云原生容器热迁移技术，实现业务无感知的容器配置、容器运行状态和容器数据卷的迁移。

技术目标：

- 1) 实现业务无感知的 CPU 容器热迁移；
- 2) 实现对 k8s 无侵入的融合方案，能够以云原生方式落地。

可提供实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题八：下一代智能代码服务工蜂 2.0 关键技术研究

研究概要描述：

由于大型企业涉及的业务种类繁多(如游戏、社交、视频等)、业务逻辑复杂，软件研发的复杂性急剧增加。面对复杂多样的软件研发，特别是研发过程中的资产(代码库和资料库等)管理问题，如何做到规范、高效、安全、智能的管理是当前提高研发效能的热点问题。具体包括能否快速检索企业的海量代码并有效复用，能否基于用户需求自动生成代码和描述文档，智能高效地评审代码，从而缩短研发周期，规范研发模式，为多样化用户及时提供按需用户体验。另外，由于全球研发资产的多源异构性对数据的安全传输、快速同步提出了新的挑战。以上这些问题都是当前企业面临的研发管理痛点。而当前的代码管理平台，称为代码管理 1.0，包括 GitHub、GitLab 以及工蜂都是采用基于 Git 的代码协作开发模式，不能很好的解决以上问题，亟需基于人工智能和大数据技术，研究并突破下一代代码智能服务关键技术，构建工蜂 2.0，夯实腾讯在业界的影响力。具体需要解决的问题描述如下：

- 3) 如何从海量代码库中学习代码行为模式，构建智能的研发环境，包括代码补全、代码生成、代码描述文档生成等，从而提高开发效率，规范开发过程，提升软件质量；
- 4) 如何根据历史的 CR 信息提取高质量的代码评审模式，自动生成代码评审，从而提高 Committer 评审代码的效率；
- 5) 如何在全球多中心、跨地域的协同开发模式下保障代码库、资料库的快速、可靠、安全的同步和传输。

技术目标：

基于以上问题，我们给出打造工蜂 2.0 需要突破的技术目标，主要涉及智能化(1、2、3、4、5)、网络(6)、安全(7)三个方向：

- 1) 研究代码补全方法，构建基于代码语义理解的代码补全规则集，以及基于 Transformer 的代码补全模型，通过规则+AI 模型的方式解决代码 Token (扩充 VSCode IntelliSense 未实现的规则)、代码行、代码块粒度的补全问题，对标业界先进工具(如 GitHub Copilot、TabNine、Kite 等)；
- 2) 研究并实现基于函数/类/文件等粒度的代码生成方法，解决特定领域(例如游戏、社交、视频等)的代码生成问题，对标业界先进工具(如 GitHub Copilot 等)；
- 3) 研究并实现基于函数/类/文件等粒度的代码注释(Code Comments)生成方法；
- 4) 挖掘代码变更行为模式，结合 AI 技术研究并实现代码提交日志(Commit Message)生成方法；
- 5) 研究 CR(Code Review)智能评审技术，挖掘高质量的代码评审模式，实现智能评审的辅助推荐功能；
- 6) 研究并实现多源异构场景下的海量代码文件、二进制文件的实时、快速、可靠的同步机制，特别是低带宽场景下的代码传输的局域网加速、边缘加速、地域 CDN 等技术；
- 7) 研究并实现海量代码存储的低损耗、高性能、快速加解密技术。

可选择 1 个或多个技术目标开展研究，并需在企业内（腾讯工蜂）实现工程落地，鼓励并支持发表相关领域的 CCF A 类学术论文。

实验资源：

- 1) 腾讯提供腾讯云计算资源；
- 2) 腾讯提供专家指引和现场落地环境。

命题九：面向超复杂异构网络的自适应传输控制关键技术研究

研究概要描述：

近几年，随着音视频业务呈现出爆炸式增长，流量传输优化成为学术界、工业界不断关注的焦点。大量事实说明，one-size-fits-all 的网络传输协议与算法不存在，而现有基于专家经验的固定规则式传输控制无法很好地适应网络状态的动态性和网络条件的多样性，如同一种拥塞控制参数（或算法）在不同的网络条件下会取得千差万别的效果。因此，研究并部署适配音视频业务的自适应传输控制技术成为各个 CDN 厂商构建核心竞争力的关键。

研究面向超复杂异构网络的自适应传输关键技术，改善用户侧 QoE，仍然面临以下难题：

- 1) 细粒度的网络 QoS 与粗粒度的用户 QoE 很难形成较为精确的匹配关系；
- 2) 网络状态的测量结果难以准确复用至拥塞控制参数的自动化配置；
- 3) 网络条件的差异化与应用服务的多样化难以精确指导拥塞控制算法的自适应选择；
- 4) 依托新型技术（如 RDMA、P4 等）的自适应传输控制解决方案还远未达到部署要求。

技术目标：

- 1) 基于网络 QoS 可实现用户 QoE 的准确推理（精度达到 80%以上），探索影响用户 QoE 的关键 QoS 指标；
- 2) 依托自适应拥塞控制，显著改善网络 QoS（如重传率降低 10%，有效吞吐提升 5%）和用户侧 QoE 指标（如卡顿频率与时长等）；
- 3) 研究成果具备现网部署能力；
- 4) 发表至少 1 篇 CCF B 类（或 SCI 二区）及以上的学术论文。

实验资源：

- 4) 腾讯提供腾讯云计算资源；
- 5) 腾讯提供可测量的真实网络数据；
- 6) 腾讯提供第三方性能测试平台。

命题十：面向动态异构网络的多路径调度关键技术研究

研究概要描述：

本项目主要研究如何利用 QUIC 多路径技术提升传输性能和可靠性。当前移动互联网用户多宿主接入方式 (WiFi/蜂窝) 日益普及, 业务服务大规模云化并提供多通 (电信/联通/移动等) 接入, 基于用户态的 QUIC 传输协议的标准化和生态成熟, 为多路径传输技术领域的发展和大规模应用带来重要机遇。但传统基于经验启发式的多路径调度策略, 存在难以很好适应真实移动用户网络的动态性和异构性的局限。

技术目标：

提出基于 DRL 的多路径调度算法, 相比现有多路径调度算法 (例如 Linux 内核支持的 MinRTT/ReMP/BLEST) 在真实海量用户的网络环境中：

- 1) 广告类业务成功率提升 (例如 20%) ;
- 2) 流媒体类业务卡顿率降低 (例如 20%) ;
- 3) 下载类业务明显带宽提升 (例如 20%) ;
- 4) 适合同时在服务端和移动端部署, 具有合理的资源消耗。

实验资源：

- 1) 腾讯提供腾讯云计算资源;
- 2) 腾讯提供可测量的真实网络数据, 包括上 T 的带宽和上亿的用户数据;
- 3) 腾讯提供第三方性能测试平台;
- 4) 腾讯提供专家指引和现场落地环境。

命题十一：超大型数据中心设备健康管理方案研究

研究概要描述：

腾讯目前已经建立了遍布全球的数据中心，并构建了一体化集中监控平台，沉淀了多年，多地域，多维度的海量数据。在该基础上，数据中心运营进入智能化运营的阶段，即从传统的依赖于人工和经验的模式升级为依赖于数据和算法的智能模式。

在数据中心基础设施管理的各类业务中，数据中心基础设施设备健康管理是数据中心安全运营的生命线，如何准确地掌握设备健康状态，提前或者及时、准确地发现问题，如何更加合理科学地进行设备维保作业是当前行业面对的重大挑战。

腾讯数据中心已经发展到第四代—基于 T-block 架构的数据中心，具有模块化、标准化的特点，并设计了全新的弱电架构支持终端设备可监可控，在该基础上，需要研究一套具有创新性的数据中心基础设施设备健康管理解决方案，实现对于设备全生命周期的科学化管理。

技术目标：

以数据为驱动，综合各系统物理过程和业务流程，构建基于 T-block 架构的数据中心设备健康评估模型及预测模型，提出系统性的设备健康管理解决方案，以提升数据中心设备运维质量，提供准确的设备健康评估结果，提前发现设备问题，并为基于状态的设备维修和管理提供支撑。具体内容如下：

- 1) 健康状态分析和评估：根据数据中心设备运营特点，梳理影响设备性能的相关因素，建立设备健康评估指标体系，给出评估约束条件，综合设备劣化过程和风险等，建立综合的评估模型，实现对设备健康状态的分类和估计，要求设备恶化和故障状态的评估准确率达到 95%以上；
- 2) 故障预测：根据设备运行历史数据和故障记录，利用数据驱动方法建立设备状态及性能趋势变化的统计模型，以一定的置信水平给出设备故障概率。要求重要故障预测时间至少提前 5 天，要求重要故障预测准确率达到 90%以上；
- 3) 状态维护决策：维护决策优化建模，以提高数据中心设备维护保障效率为总体目标，综合考虑维护费用、故障风险及停机时间等约束条件，建立维护决策模型。自动给出维护建议，输出科学合理的维护方案和计划，要求决策准确率达到 95%以上。

实验资源：

- 1) 腾讯提供研发环境和计算资源；
- 2) 腾讯提供现场落地环境。

命题十二：超大型数据中心制冷系统智能化节能方案研究

研究概要描述：

腾讯拥有遍布全球的大型和超大型数据中心，在大量的数据中心运营管理过程中，面对了前所未有的挑战。其中最紧迫的问题是能源消耗，巨额的电费已经成为数据中心高速发展的瓶颈。随着“3060”双碳目标的发布，各地陆续出台限耗政策，严格地管控数据中心能耗水平，要求极低的运行 PUE。在这种大背景下，数据中心需要持续地探索智能化、创新性的技术来支撑绿色生态体系的建设。

腾讯数据中心已经发展到第四代一基 T-block 架构的数据中心，具有模块化、标准化的特点，并设计了全新的弱电架构支持终端设备可监可控，在该基础上，需要研究一套具有创新性的数据中心能效管理解决方案，实现能效的智能分析和调优。

技术目标：

针对能效管理，进行如下内容的研究：

- 1) 基于大数据和设备机理模型研究和设计 T-block 能耗评估算法，在腾讯数据中心现场实际落地，使平台具有能效量化评估能力。
- 2) 研究和设计 T-block 能耗优化算法和控制策略，在腾讯数据中心现场实际落地，使平台具有输出能效优化策略以及进行安全控制的能力。
- 3) 和腾讯研究团队一起完成 T-block 能效智能管理方案在腾讯智维平台的落地，要求在典型工况上的诊断准确率超过 95%，自动优化策略准确率超过 95%，制冷系统平均能耗降低 10%。

实验资源：

- 1) 腾讯提供研发环境和计算资源；
- 2) 腾讯提供现场落地环境。