

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2025 年 3 月 6 日 (06.03.2025)



(10) 国际公布号
WO 2025/044675 A1

(51) 国际专利分类号:
G06F 9/50 (2006.01)

(21) 国际申请号: PCT/CN2024/109422

(22) 国际申请日: 2024 年 8 月 2 日 (02.08.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202311125060.1 2023年9月1日 (01.09.2023) CN

(71) 申请人: 阿里云计算有限公司 (ALIBABA CLOUD COMPUTING LTD.) [CN/CN]; 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号云谷园区 1-2-A06 室, Zhejiang 310030 (CN)。

(72) 发明人: 彭成 (PENG, Cheng); 中国北京市朝阳区望京东园四区 9 号楼-阿里中心. 望京 A 座, Beijing 100102 (CN)。 何旻 (HE, Min); 中国上海市浦东新区盛荣路 388 弄百佳通 1 号楼 2-5 层, Shanghai 200120 (CN)。 郑晓 (ZHENG, Xiao); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号云谷园区, Zhejiang 310030 (CN)。

(74) 代理人: 北京同钧律师事务所 (BEIJING TONGJUN LAW FIRM); 中国北京市海淀区西直门北大街 32 号院 1 号楼 7 层 808 室, Beijing 100082 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ,

(54) Title: VIDEO MEMORY ALLOCATION METHOD, VIDEO MEMORY REQUEST METHOD, AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 一种显存的分配方法、请求方法、设备及存储介质

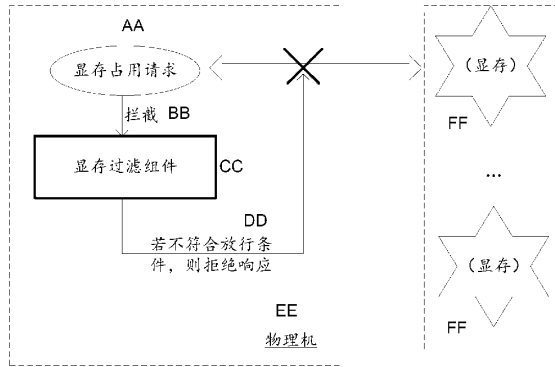


图 2

AA Video memory occupation request
BB Interception
CC Video memory filter assembly
DD If a clearance condition is not met, refuse to respond
EE Physical machine
FF Video memory

(57) Abstract: Provided in the embodiments of the present disclosure are a video memory allocation method, a video memory request method, and a device and a storage medium. In the embodiments of the present disclosure, it is proposed that a video memory occupation request initiated by an application in a physical machine is intercepted, and a video memory occupation upper limit value is configured for each application in the physical machine; on this basis, when the video memory occupation request is intercepted, whether the sum of a video memory amount to be occupied that is requested by the application this time and a video memory amount which has

[见续页]

IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

- 关于申请人有权要求在先申请的优先权(细则4.17(iii))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

been occupied by the application has exceeded the video memory occupation upper limit value corresponding to the application can be determined, and if the sum thereof has exceeded the video memory occupation upper limit value, a refusal to respond to the video memory occupation request is given. Therefore, a video memory occupation request generated in a physical machine can be filtered on the basis of video memory occupation upper limit values configured for different applications, so as to effectively isolate video memories occupied by the different applications, and thus the problem of the video memories being excessively occupied by some of the applications is avoided, such that the normal performance of the applications that share the video memories is ensured.

(57) 摘要: 本公开实施例提供一种显存的分配方法、请求方法、设备及存储介质。在本公开实施例中, 提出对物理机中应用程序发起的显存占用请求进行拦截, 并为物理机中的各个应用程序分别配置显存占用上限值; 基于此, 在拦截到显存占用请求的情况下, 可判断应用程序本次所请求的待占用显存量和应用程序已占用的显存量之和是否已经超过了应用程序对应的显存占用上限值, 若已经超过, 则拒绝响应显存占用请求。这样, 可基于为不同应用程序配置的显存占用上限值, 对物理机中发生的显存占用请求进行过滤, 以实现针对不同应用程序所占用的显存进行有效隔离, 从而避免显存被部分应用程序过量占用的问题, 进而保证共用显存的各个应用程序的正常性能。