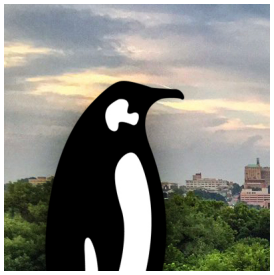

Contribution ID: **262** Type: **not specified**

Introducing PAGEMAP_SCAN IOCTL for Windows syscalls translation and CRIU

Monday, November 13, 2023 2:35 PM (25 minutes)

Windows APIs GetWriteWatch() and ResetWriteWatch() are used to get and clear the write-tracking state atomically of any number of pages in memory. Only the kernel can keep track of this state efficiently through the memory management component. Linux Kernel lacked this support.

Soft-dirty PTE flag was used initially. But it had to be left alone because of its short-comings and no way to fix it after years. UFFD_FEATURE_WP_ASYNC and UFFD_FEATURE_WP_UNPOPULATED are the new features which have been added in Userfaultfd to keep track of write-tracking state asynchronously and correctly. PAGEMAP_SCAN IOCTL has been added to filter the information about PTE bits and only get desired data. It is used to perform get and clear the write-tracking state atomically.

By using it, CRIU doesn't need to freeze processes to pre-dump their memory and it'll have the accurate information about pages to the moment of dumping them.

We'll discuss its evolution, current implementation, use cases, and benchmarks. The IOCTL is at v33. This discussion aims to advance patches and attract more users to this interface.

Primary author: Mr ANJUM, Muhammad Usama (Collabora)

Co-author: VAGIN, Andrei

Presenters: VAGIN, Andrei; Mr ANJUM, Muhammad Usama (Collabora)

Session Classification: Containers and checkpoint/restore MC

Track Classification: LPC Microconference: Containers and checkpoint/restore MC