



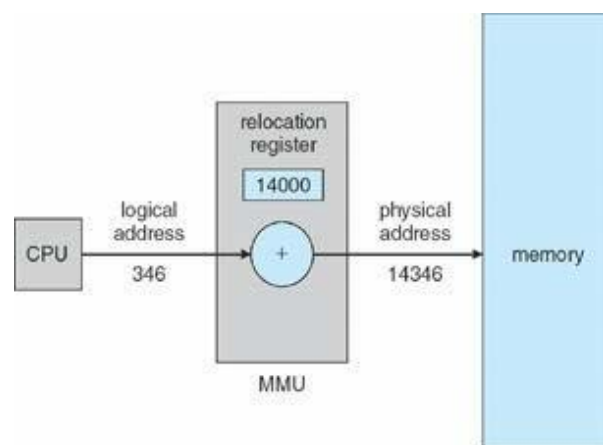
## >>> COURSE MATERIAL <<<

INTRODUCTION TO OPERATING SYSTEM / Course ID 502047

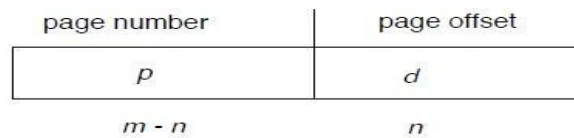
Thông tin dưới đây được dịch từ tài liệu StudyGuide tại trang OS-Book của sách giáo trình chính và phần Summary của sách. Mọi lỗi sai sót hay góp ý, xin gửi về cho tôi qua email [trantrungtin@tdtu.edu.vn](mailto:trantrungtin@tdtu.edu.vn)

### Study Guide for Lecture notes ch09

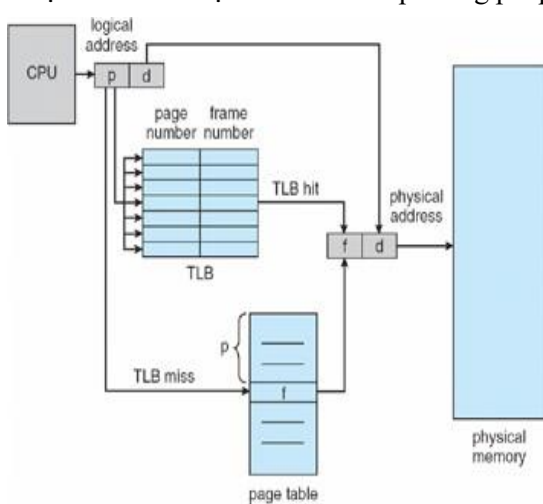
- Bộ nhớ cache (Cache) nằm ở giữa bộ nhớ chính và các thanh ghi CPU. Bộ nhớ cache còn chia làm nhiều mức như L1, L2, L3, ....
- Thanh ghi cơ sở (Base) và thanh ghi giới hạn (limit) định ra không gian địa chỉ luận lý cấp cho một tiến trình.
- Các địa chỉ trong mã chương trình biên dịch được liên kết (bind) đến các địa chỉ khả tái định vị. Công việc liên kết có thể xảy ra ở 3 thời điểm:
  - + Thời điểm biên dịch: [Compile time] nếu vị trí bộ nhớ được cấp đã được xác định trước, mã tuyệt đối [absolute code] sẽ được tạo ra.
  - + Thời điểm nạp: [Load time] phải sinh ra code khả tái định vị [relocatable code] nếu vị trí bộ nhớ chưa biết tại thời điểm biên dịch.
  - + Thời điểm thực thi: [Execution time] quá trình liên kết tạm hoãn đến thời điểm chạy nếu như tiến trình có thể bị di chuyển trong quá trình nó thực thi.
- Bộ quản lý bộ nhớ [Memory-Management Unit (MMU)] là thiết bị ánh xạ địa chỉ ảo sang địa chỉ vật lý.
- Kiểu sắp đặt đơn giản sử dụng các thanh ghi tái định vị [relocation register] để cộng vào giá trị cơ sở trong bộ nhớ.
- Hoán chuyển [Swapping] cho phép tổng không gian bộ nhớ vật lý của tất cả tiến trình vượt hơn kích thước của bộ nhớ vật lý. Để có thể thực hiện được như vậy, một vài tiến trình có thể bị hoán chuyển tạm thời ra kho lưu trữ (backing store) và sau đó sẽ được mang trở lại vào bộ nhớ chính để thực thi.
- Kho lưu trữ (Backing store): vùng đĩa cứng đủ lớn để chứa các bản sao của các ảnh bộ nhớ của các tiến trình bị tạm thời hoán chuyển ra.
- Mang ra, mang vào [Roll out, roll in]: hoán chuyển biến thể để định thời dựa trên mức độ ưu tiên của các tiến trình. Tiến trình có độ ưu tiên thấp sẽ bị dời ra [swap out] để tạo thêm vùng trống trong bộ nhớ cho tiến trình có độ ưu tiên cao có thể được nạp vào.
- Giải pháp cho bài toán cấp phát vùng nhớ động [Dynamic Storage-Allocation Problem] là
  - + Đầu tiên vừa [First-fit]: cấp phát lỗ trống [hole] đầu tiên tìm thấy và đủ lớn để nạp.
  - + Tối ưu [Best-fit]: cấp phát lỗ trống nhỏ nhất trong số tất cả lỗ trống đủ lớn để nạp (cần phải dò tìm tất cả bộ nhớ), lỗ trống mới sau khi cấp còn thừa ra cũng sẽ là nhỏ nhất.



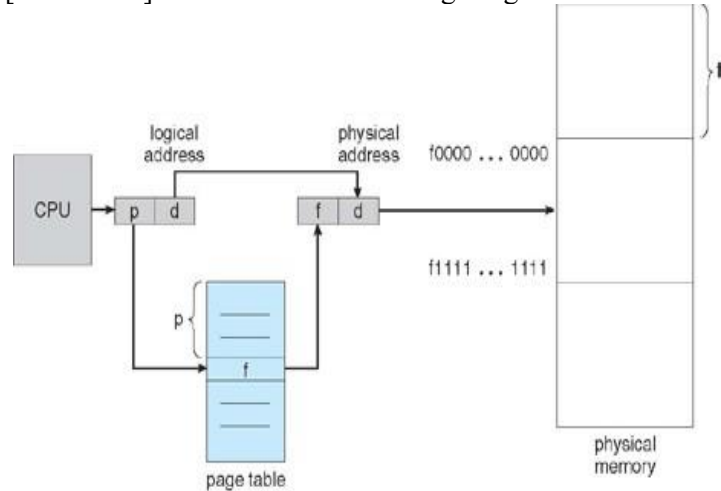
- + Tồi nhất [**Worst-fit**]: cấp lỗ trống có kích thước lớn nhất (và đủ lớn để nạp), lỗ trống mới sau khi cấp thừa ra cũng sẽ là to nhất.
- Phân mảnh ngoại [**External Fragmentation**]: tổng kích thước các lỗ trống trong không gian bộ nhớ có thể đáp ứng yêu cầu cấp phát, nhưng các lỗ trống không liên tục.
  - + Có thể giảm phân mảnh ngoại bằng cách phương pháp nén [**compaction**]: hoán chuyển các vùng nhớ tự do [free memory] thành một khối liên tục. Chỉ có thể thực hiện nếu quá trình tái định vị là động.
- Phân mảnh nội [**Internal Fragmentation**]: bộ nhớ cấp phát lớn hơn một ít so với yêu cầu cấp phát, lý do là kích thước cấp phát theo khối (ví dụ 1, 2, 4, 8, ... bytes) còn các yêu cầu thì nhiều kích cỡ (1, 2, 3, 4, 5, ... bytes).
- Trong kỹ thuật phân trang [Paging], bộ nhớ vật lý được chia làm các khung [frame] có kích thước cố định, kích thước là bội số của 2, nằm trong đoạn 512 byte và 16 MB.
- Bộ nhớ luận lý được chia làm các trang [page] có kích thước cố định và bằng kích thước của khung.
- Bảng phân trang [**Page table**] được sử dụng để dịch địa chỉ luận lý sang địa chỉ vật lý.
  - + Số trang [**Page number (p)**]: được sử dụng như là một chỉ số [index] trong bảng phân trang.
  - + Độ dịch trang [**Page offset (d)**]: kết hợp với địa chỉ cơ sở để định ra địa chỉ bộ nhớ vật lý.
  - + Danh sách các khung tự do [**Free-frame list**] được bảo trì để theo dõi các khung nào có thể dùng cho cấp phát.
- Với không gian địa chỉ luận lý có kích thước  $2^m$  bytes và kích cỡ mỗi trang  $2^n$  bytes, thì địa chỉ dùng cho vùng không gian đó có cấu trúc:



- Bộ đệm dịch tìm sang một bên [**Transition Look-aside Buffer (TLB)**] là một bộ nhớ cache của CPU để cho phân cứng quản lý bộ nhớ sử dụng nhằm nâng cao tốc độ dịch địa chỉ ảo.
  - + Thường là nhỏ - từ 64 đến 1024 mục.
  - + Khi bị trượt [miss], giá trị trượt sẽ được nạp vào TLB cho lần truy cập tiếp theo được nhanh hơn.
- Bộ nhớ TLB được tổ chức theo phương pháp kết hợp [associative] nên có thể tìm kiếm song song.

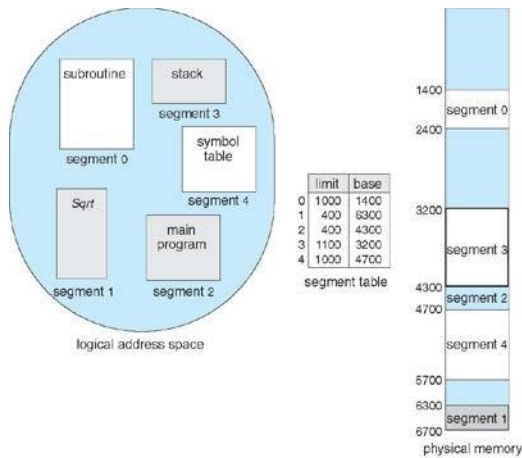


Phân trang với TLB

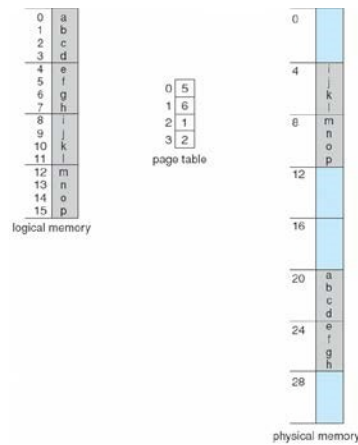


Phân trang không có TLB

- Thời gian truy cập hiệu quả [**Effective Access Time**]:  $EAT = (1 + \epsilon) \alpha + (2 + \epsilon)(1 - \alpha)$ 
  - $\epsilon$  = time unit,  $\alpha$  = tỉ lệ trúng
- Các bit hợp lệ / không hợp lệ [**Valid / invalid**] có thể được sử dụng để bảo vệ bộ nhớ.
  - + “Hợp lệ” nên trang được liên kết đang nằm trong không gian địa chỉ luận lý của tiến trình, vì vậy đó là một trang hợp pháp.
- Bảng phân trang có thể phân cấp [paged page tables].
- Bảng phân trang băm [**Hashed Page Tables**]: số trang ảo được băm [hashed] vào bảng phân trang.
- Bảng phân trang có chuỗi các phần tử băm đến cùng một vị trí
  - + Mỗi phần tử có (1) số trang ảo, (2) giá trị của khung trang được ánh xạ, (3) một con trỏ tới phần tử tiếp theo
  - + Tìm kiếm dọc theo chuỗi số trang ảo.
- Bảng phân đoạn [**Segment table**] – ánh xạ các địa chỉ vật lý 2-chiều.
- Các mục được bảo vệ với bit hiệu lực và quyền đọc / ghi / thực thi.



Ví dụ phân đoạn



Ví dụ phân trang

## Summary Chapter 09 of book “OS concepts 10<sup>th</sup> edition”

- Bộ nhớ là trung tâm cho hoạt động của một hệ thống máy tính hiện đại và nó bao gồm một mảng kích thước lớn của các byte, mỗi byte có địa chỉ riêng.
- Một cách để phân bổ một không gian địa chỉ cho mỗi tiến trình là thông qua việc sử dụng các thanh ghi cơ sở (base) và giới hạn (limit). Thanh ghi cơ sở giữ địa chỉ vật lý đầu tiên (nhỏ nhất) của vùng nhớ hợp pháp và thanh ghi giới hạn chỉ định kích thước của phạm vi bộ nhớ mà tiến trình đó được phép truy cập.

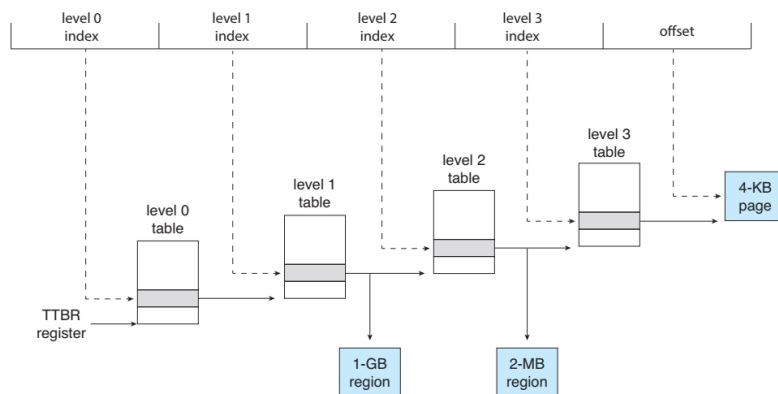


Figure 9.27 ARM four-level hierarchical paging.

- Địa chỉ tượng trưng tham chiếu liên kết đến địa chỉ vật lý thực tế có thể xảy ra trong quá trình (1) biên dịch, (2) nạp hoặc (3) thực thi.
- Một địa chỉ do CPU tạo ra được gọi là địa chỉ logic, nó được đơn vị quản lý bộ nhớ (MMU) dịch thành địa chỉ vật lý trong bộ nhớ chính.
- Một cách tiếp cận để phân bổ bộ nhớ là phân bổ các phân vùng liên tục của bộ nhớ có kích thước tùy biến. Các phân vùng này có thể được phân bổ dựa trên ba chiến lược: (1) lỗ trống đầu tiên, (2) lỗ trống vừa khít nhất và (3) lỗ trống to nhất.
- Hệ điều hành hiện đại sử dụng phân trang để quản lý bộ nhớ. Trong quá trình này, bộ nhớ vật lý được chia thành các khối có kích thước bằng nhau được gọi là khung trang và bộ nhớ luận lý được chia thành các khối có cùng kích thước (và cũng bằng kích thước một khung trang) được gọi là các trang.

- Khi phân trang được sử dụng, một địa chỉ logic được chia thành hai phần: số trang và độ dịch (trong nội bộ trang). Số trang đóng vai trò là chỉ mục trong bảng phân trang lên mỗi tiến trình có khung trang đang chứa trong bộ nhớ vật lý. Độ dịch là vị trí cụ thể trong khung trang được tham chiếu.
- Bộ đệm dịch tìm sang một bên (TLB / translation look-aside buffer) là bộ đệm phần cứng của bảng phân trang. Mỗi mục TLB chứa một số trang và khung trang tương ứng của nó.
- Sử dụng TLB trong dịch địa chỉ cho các hệ thống phân trang liên quan đến việc lấy số trang từ địa chỉ logic và kiểm tra xem khung trang liên kết với trang đó có nằm trong TLB không. Nếu có, khung trang được lấy từ TLB. Ngược lại, nó phải được lấy từ bảng trang.
- Phân trang phân cấp liên quan đến việc chia một địa chỉ logic thành nhiều phần, mỗi phần đề cập đến một cấp khác nhau của bảng trang. Khi địa chỉ mở rộng vượt quá 32 bit, số lượng mức phân cấp có thể trở nên lớn. Hai chiến lược giải quyết vấn đề này là các bảng trang băm (hashed page tables) và các bảng trang đảo ngược (inverted page tables).
- Hoán đổi (Swapping) cho phép hệ thống di chuyển các trang thuộc một tiến trình vào đĩa cứng (virtual memory) để tăng mức độ đa chương trình.
- Kiến trúc Intel 32 bit có hai cấp bảng trang và hỗ trợ kích thước trang 4 KB hoặc 4 MB. Kiến trúc này cũng hỗ trợ mở rộng địa chỉ trang, cho phép bộ xử lý 32 bit truy cập vào không gian địa chỉ vật lý lớn hơn 4 GB. Các kiến trúc x86-64 và ARMv9 là các kiến trúc 64 bit sử dụng phân trang phân cấp.