

# Pamięć wirtualna w AS/400

Jan Posiadała

19 listopada 2002

# Spis treści

<b>1</b>	<b>Wprowadzenie - co to takiego AS/400</b>	<b>3</b>
<b>2</b>	<b>Organizacja pamięci</b>	<b>4</b>
2.1	Koncepcja wymiany . . . . .	4
2.2	Koncepcja adresowania . . . . .	4
2.3	Jedna tablica stron . . . . .	4
2.4	Własności pamięci . . . . .	4
<b>3</b>	<b>Tłumaczenie adresu</b>	<b>5</b>
3.1	Tłumaczenie adresu czyli odwrócona tablica stron . . . . .	7
3.2	Wpis w tablicy stron . . . . .	7
3.3	Tryby dostępu do pamięci . . . . .	9
3.4	Strategia wymiany czyli przybliżanie LRU . . . . .	10
3.5	TLB translation lookaside buffer . . . . .	10

# 1 Wprowadzenie - co to takiego AS/400

AS/400 jest systemem komputerowym firmy IBM. AS pochodzi od Application System i Advanced System. Najnowsza wersja systemu (V4R4) działa na PowerPC - opracowanym przez Apple Computer, Motorolę i IBM 64-bitowym procesorze typu RISC.

**(TI)MI** - Technology-Independent Machine Interface to interfejs na potrzeby aplikacji i (częściowo) oprogramowania systemowego. Całkowicie zasłania sprzęt przed oprogramowaniem położonym wyżej (a więc uniezależnia je od zmian sprzętowych). Ułatwia integrację różnych części systemu.

**LIC** - Licensed Internal Code to część systemu operacyjnego leżąca poniżej MI.

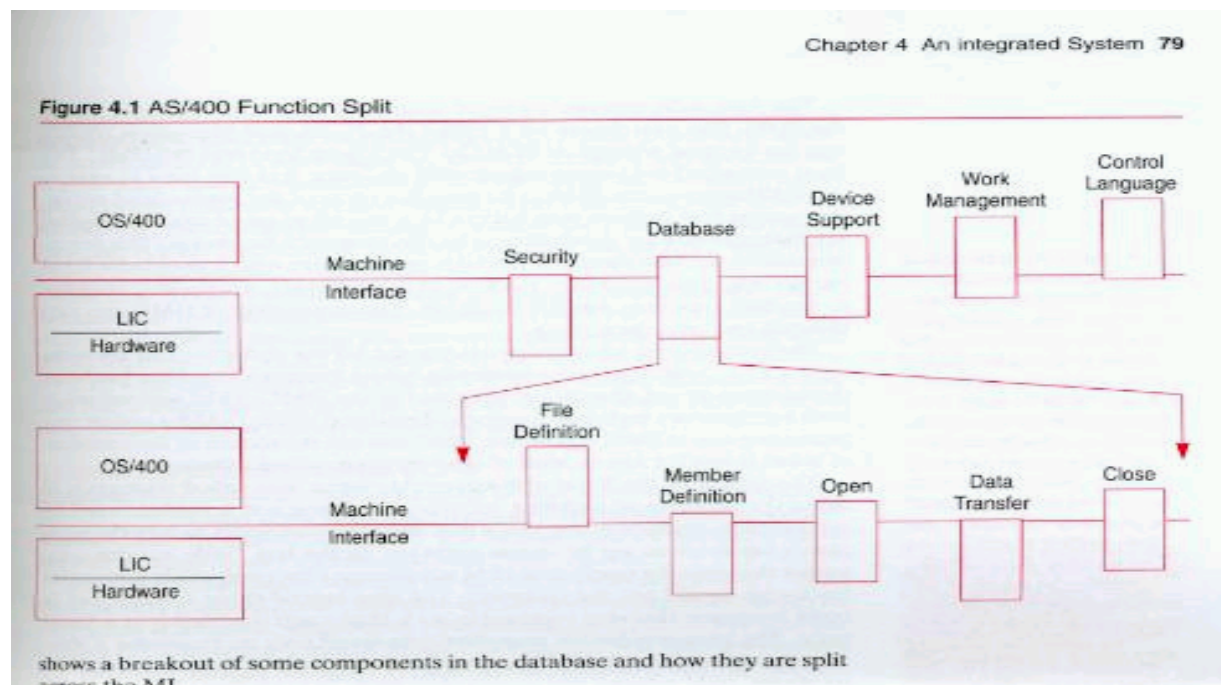
**OS/400** - obiekty i programy leżące powyżej MI.

System operacyjny dla AS/400 jest połączeniem OS/400 i LIC.

Z biegiem czasu coraz więcej funkcji przesuwano z OS/400 do LIC, głównie w celu zwiększenia wydajności. Na poziomie LIC część funkcji przesunięto do mikro kodu.

**SLIC** - System Licensed Internal Code: LIC dla systemów opartych na procesorach RISC (jądro systemu operacyjnego AS/400).

SLIC nie jest mikrojądrem, ma jednak wiele własności mikrojądra.



Kod obecnej wersji systemu jest napisany w C++.

## 2 Organizacja pamięci

### 2.1 Koncepcja wymiany

Standardowa pamięć wirtualna obejmuje dwa poziomy pamięci: pamięć główną i dyskową. Części plików są przechowywane na dysku w obszarze wymiany (ponadto oczywiście w pamięci). W AS/400 nie ma obszaru wymiany - cały system plików staje się częścią pamięci wirtualnej.

Pamięć operacyjna staje się rodzajem pamięci podręcznej dla całego dysku, a nie jedynie zarezerwowanej części dysku. Konieczny staje się długi adres, który pozwoli zaadresować całą przestrzeń dyskową w systemie. Adres jest 64-bitowy.

### 2.2 Koncepcja adresowania

Adres wirtualny jest tłumaczony do adresu rzeczywistego zanim nastąpi odwołanie do pamięci. Architektura Power-PC ma jeszcze jeden poziom adresowania, zwany adresami efektywnymi generowanymi przez programy. Adresy efektywne muszą być najpierw przetłumaczone do adresów wirtualnych które z kolei, są tłumaczone do adresów fizycznych.

Pamięć składa się z ramek (na komputerze PowerPC mają one rozmiar 4 KB). Obiekty są podzielone na strony. Duży obiekt może zajmować wiele stron, pojedyncza strona nigdy nie zawiera więcej niż jednego obiektu. Strona sprowadzona do pamięci przez jeden proces jest dostępna dla każdego innego, np. dowolna liczba zadań może współdzielić instrukcje programu (takie nieograniczone współdzielenie ogranicza konieczność transmisji dyskowych).

### 2.3 Jedna tablica stron

W AS/400 jest tylko jedna tablica stron dla wszystkich użytkowników. Nie ma więc potrzeby czyszczenia TLB podczas przełączania kontekstu. Nie trzeba też czyścić pamięci podręcznych (tych między pamięcią główną a procesorem). Dlatego przełączenie kontekstu jest szybkie. Można więc wykonywać je często (eksperymentalnie zbadano, że średnio wykonuje się je raz na 1200 instrukcji). Dzięki temu AS/400 zapewnia dobrą wydajność w środowisku interakcyjnym, zorientowanym na aplikacje działające w trybie transakcji. Można więc dołączać wiele terminali. Pojedynczy duży AS/400 może obsługiwać ponad 2000 działających współbieżnie użytkowników.

### 2.4 Własności pamięci

Właśnie przedstawiliśmy niektóre własności organizacji pamięci w AS/400. Przejrzemy listę tych własności, razem nowymi, o których wcześniej nie mówiliśmy (omówimy je później):

- Wielkość strony - 4096 B
- Zakres adresu efektywnego  $2^{64}$  B
  - Liczba segmentów efektywnych  $2^{40}$
  - Rozmiar segmentu  $2^{24}$  B

- Dwa tryby tłumaczenia adresu
  - Tłumaczenie poprzez tablicę segmentów - opcja bezpieczeństwa C2
  - Tłumaczenie bezpośrednio
- Dwa specjalne typy adresów efektywnych, identyfikowane 12 najbardziej znaczącymi bitami adresu efektywnego. Specjalne postępowanie dla tych adresów.
  - hex 800 (efektywny = rzeczywisty) adresy te odwzorowują całą pamięć. Można zaadresować  $2^{28}$  segmentów.
  - hex 801 (efektywny = rzeczywisty dyskowy) adresy rzeczywiste na urządzeniach wejścia-wyjścia. Można zaadresować  $2^{28}$  segmentów.
- Zakres adresu wirtualnego
  - Liczba segmentów wirtualnych  $2^{40} - 2^{29}$  ( $2^{40} - 2 * 2^{28}$ )
  - Wielekość segmentu wirtualnego  $2^{24}$  B
- Zakres adresu rzeczywistego  $2^{52}$  B

Powyższe własności dotyczą 64-bitowego procesora AS/400.

### 3 Tłumaczenie adresu

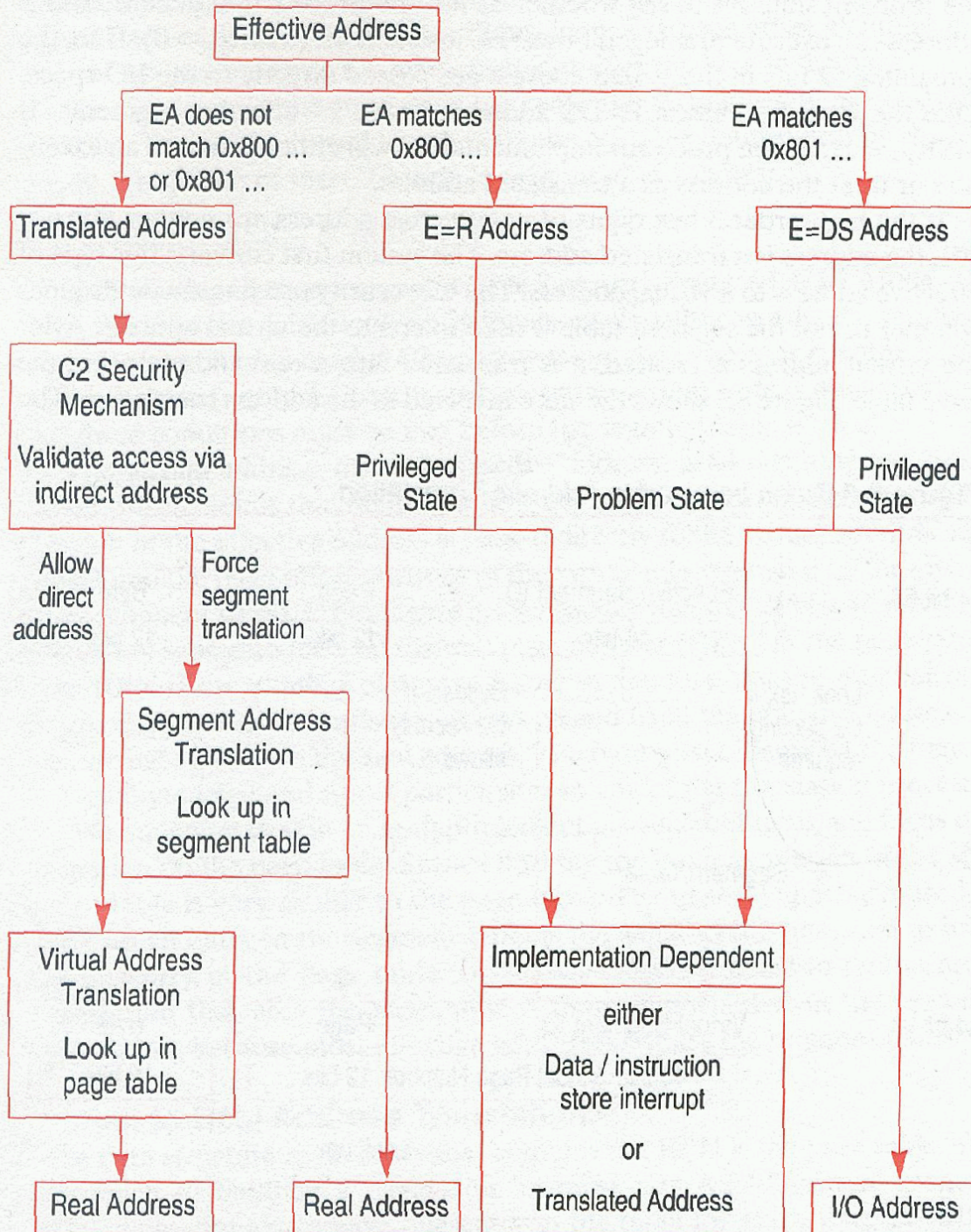
Pierwszą czynnością jest sprzętowe sprawdzenie 12-bitowego prefiksu adresu efektywnego .

**0x800** (efektywny = rzeczywisty) Adresami rzeczywistymi posługuje się wiele części SLIC. Pewne struktury SLIC (jak np. tablica stron) i procedury SLIC nie są stronicowane. Ich adres wirtualny jest równoważony adresowi rzeczywistemu. Część efektywnej przestrzeni adresowej dla tych adresów jest ustanowiona oddzielnie.

**0x801** (efektywny = rzeczywisty dyskowy) Zewnętrzna przestrzeń adresowa jest widziana przez procesor jako część pamięci. Adres jest używany do identyfikacji konkretnego urządzenia I/O dołączonego do systemu. Ten Typ architektury jest nazywany "wyjście-wejście odwzorowane jako pamięć" . Żaden specjalny zestaw instrukcji do komunikacji procesora z urządzeniami we/wy nie jest potrzebny.

Po wykryciu trybu 0x80X sprawdza się, czy proces jest upoważniony do wykonywania uprzywilejowanych instrukcji. Jeśli tak, dolne 52-bity są zwracane jako adres rzeczywisty. Jeśli nie, generowany jest wyjątek. Tak więc odwołania do pamięci w tym trybie wykonują się bez narzutu systemowego.

**Figure 8.2 Address Translation with Tags Active**



### 3.1 Tłumaczenie adresu czyli odwrócona tablica stron

Z 64-bitowego adresu wirtualnego jest wyodrębniany 52-bitowy prefiks VPN (*ang. virtual page number*). VPN jest używany do znalezienia odpowiedniej pozycji w odwróconej tablicy stron na której znajduje się (lub nie - błąd braku strony) rzeczywisty numer strony - RPN - czyli adres ramki w której znajduje się ta strona.

Pozostałe 12 bitów to przesunięcie na stronie - jest bezpośrednio doklejane do adresu rzeczywistego.

Każda pozycja odwróconej tablicy stron przechowuje co najwyżej 8 wpisów dla adresów wirtualnych stron, które aktualnie znajdują się w pamięci. Adresy wirtualne (VPN) są sklejane w grupy wpisów przy pomocy pewnej funkcji haszującej.

Wszystkiego czego można się dowiedzieć o tej funkcji to:

*"The hash function on an AS/400 takes some high-order bits out of the VPN and XORs them with some low-order bits from the VPN. This value is ANDed with mask bits from a special register that the size of the page table. Finally, this result is ORed with the real address of the page table. The result of this is a 52-bit real address into the page table."*

Do odnalezienia (potencjalnego) wpisu odpowiadającego danemu VPN używa się tej właśnie funkcji haszującej.

Grupy wpisów nazywane są PTEG (page table entry group). Po odnalezieniu odpowiedniej grupy znajdujące się tam wpisy zostaną przeszukane sekwencyjnie w celu odnalezienia wpisu (PTE) odpowiadającemu dokładnie adresowi, do którego nastąpiło odwołanie.

W większości implementacji AS/400, wielkość odwróconej tablicy stron (ilość grup PTEG) jest połową ilości stron mieszczących się w pamięci. Przy liczności grup równej 8 oznacza to, że tablica może pomieścić 4 razy więcej wpisów niż jest ramek w pamięci, a średnia liczba wpisów w grupie wynosi 2. Najważniejszą cechą funkcji haszującej jest właśnie równomierny rozrzut PTE po grupach.

W wyjątkowych sytuacjach - gdy wynik funkcji haszującej jest równy dla więcej niż 8 adresów wirtualnych stron znajdujących się w pamięci - odpowiednie wpisy muszą być trzymane w drugiej tablicy stron (a tak naprawdę używana jest druga funkcja haszująca).

W przypadku, gdy ani w podstawowej ani w drugiej tablicy nie odnaleziono wpisu dla żądanej strony, wywołany jest błąd braku strony. Część SLIC odpowiedzialna za zarządzanie pamięcią sprowadza stronę do pamięci (być może wyrzucając inną) i odnawia obraz pamięci (tablica stron).

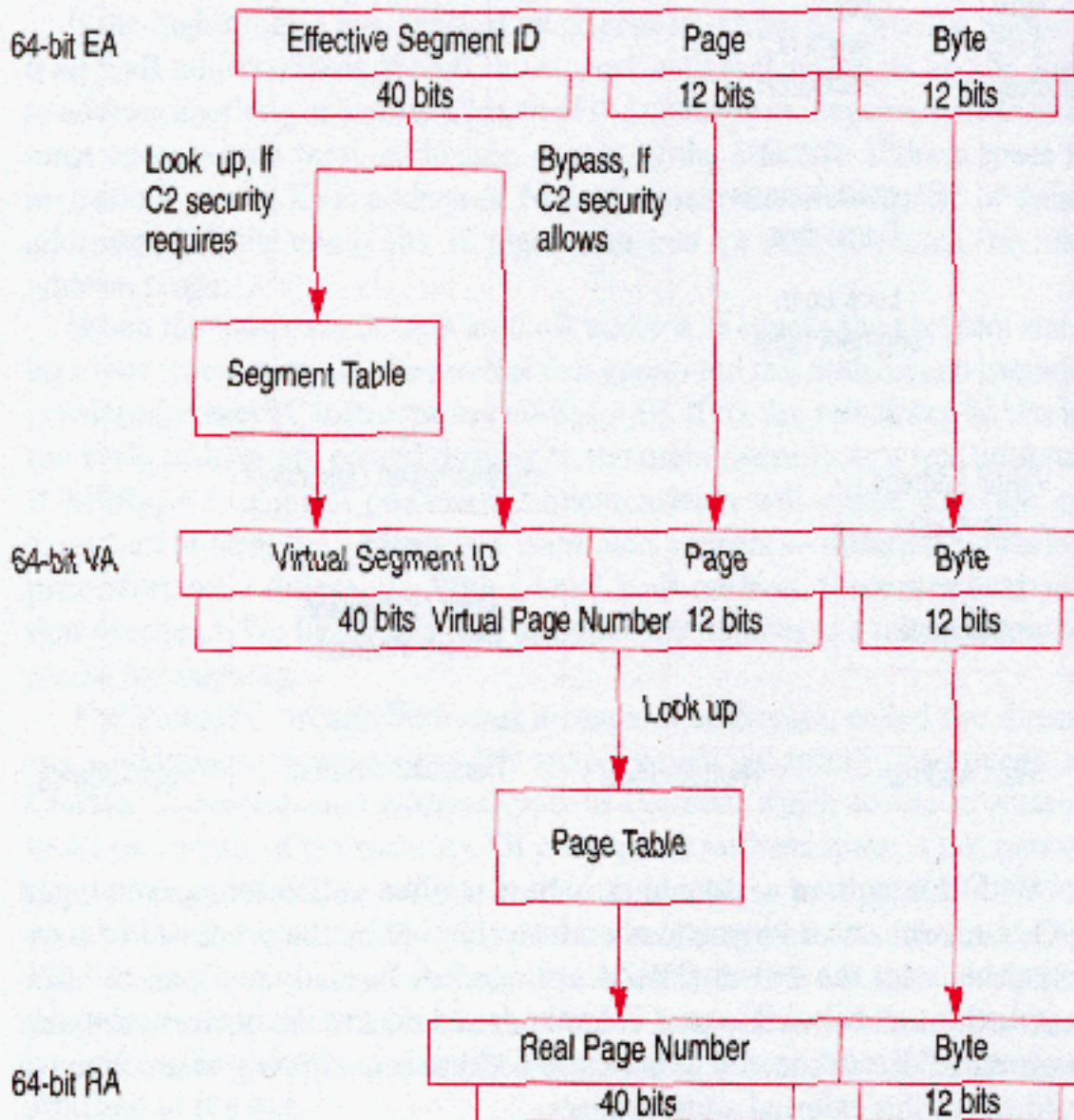
### 3.2 Wpis w tablicy stron

Wpis w tablicy stron ma 16 bajtów długości.

**Pierwsze 64 bity:**

AVPN	SW	//	H	V
0	57	60	62	63

**Figure 8.3** Steps Involved in Address Translation





BITY	NAZWA POLA	ZNACZENIE
0-56	AVPN	skrócony adres wirtualny strony
57-60	SW	zarezerwowany dla oprogramowania systemowego
62	H	identyfikator funkcji haszującej
63	V	ważność wpisu

Pierwsze 57-bitowe pole to tzw "skrócony" VPN. Dlaczego skrócony VPN jest dłuższy niż "normalny" VPN? Architektura PowerPC ma wspierać 80-bitowe adresy. Adres jest skrócony o 11 dolnych bitów które, były wykorzystane w funkcji haszującej. Zatem adresy 64-bitowe mają adresy skrócone długości 41, a 16 górnych bitów jest wyzerowanych.

### Drugie 64 bity:

//	TS	0..0	RPN	//	AC	R	C	WIMG	/	PP
0	1	2	12	51	54	55	56	57	60	62

BITY	NAZWA POLA	ZNACZENIE
1	TS	bit ustawienia "taga"
2-11		wyzerowane
12-51	RPN	fizyczny adres strony (numer ramki)
54	AC	porównanie adresu
55	R	bit odwołania
56	C	bit zapisu
57-60	WMIG	kontrola dostępu
62-63	PP	ochrona strony

### 3.3 Tryb dostępu do pamięci

Tryb dostępu procesora do danych na stronie jest definiowany przez cztery bity:

- W - natychmiastowe pisanie (*write through*)
- I - bez kopii w pamięci podręcznej (*caching inhibited*)
- M - spójność pamięci (*memory coherence*)
- G - chroniona sekwencyjność (*guarded storage*)

Dla stron o tłumaczonych adresach wartość tych flag pobierana jest z PTE. Dla trybu 0x800 pola mają wartości 0,0,1,1.

**natychmiastowe pisanie** wszelkie zmiany w pamięci podręcznej są eksportowane do pamięci głównej. Bit jest ustawiony np. w przypadku dzielenia pamięci przez wiele procesorów. Odczyt jest lokalny (z pamięci podręcznej), zapis globalny (do pamięci podręcznej i głównej).

**bez kopii w pamięci podręcznej** dostęp odbywa się bez poziomu pośredniego. Tryb ten jest przydatny w przypadku dużych bloków danych przetwarzanych sekwencyjnie.

**spójność pamięci i chroniona sekwencyjność** wymuszają spójność danych, gdy procesor (procesory) - dla zwiększenia efektywności - nie zachowuje "logicznej" sekwencyjności wykonania instrukcji.

### 3.4 Strategia wymiany czyli przybliżanie LRU

Mechanizm wymiany wybiera do wyrzucenia z pamięci te strony, które bity odwołania i zapisu mają ustawione na 0. Po każdorazowym wywołaniu algorytm wymiany ustawia wszystkim stronom bity odwołania na 0. Zatem strony są podzielone na dwie grupy. Te do których:

- odwołano się od ostatniej wymiany (1)
- nie odwołano się od ostatniej wymiany (0).

Utrzymywana jest także lista zmienionych stron, ale takich do których nie było odwołania ( $R=0, C=1$ ). Przy dostatecznej ilości takich stron strony te są zapisywane na dysku, ale pozostają w pamięci z wyzerowanym bitem zapisu. Zapobiega to gromadzeniu się wielu czasochłonnych czynności na jeden moment.

### 3.5 TLB translation lookaside buffer

W rejestrach utrzymywana jest tablica przechowująca PTE, które były ostatnio używane. W systemach AS/400 wielkość TLB pozwala na skuteczność odwołań rzędu 95%. Oznacza to że najwyżej co 20 tłumaczenie adresu wymaga więcej niż jednego fizycznego odwołania do pamięci.

## **Źródła**

1. Frank G. Soltis *Inside the AS/400* Duke Press, 1996
2. Janina Mincer-Daszkiewicz <http://rainbow.mimuw.edu.pl/SO-MSUI/>