Systèmes Avancés
M2 2015

Gestion Mémoire

Exercice N°3 1

- ! "e# est #$interva##e %$a%resses &"$"n 'rocess"s 'e"t tenter %$attein%re(
- ! "e##e est/sont #es interva##es %$a%resses)va#i%es* &"$"n 'rocess"s 'e"t attein%re(
- ! "e se 'asse+t+i# s$i# essaye %$accé%er , %$a"tres a%resses(
- -n 'rocess"s 'e"t+i# e..ect"er n$im'orte &"e##e o'ération s"r "ne a%resse va#i%e(

Exercice N°3 1)s"ite*

- ! "e#s sont #es /esoins %$"n 'rocess"s en terme %e 0estion)a##ocation / %é+a##ocation* mémoire(
- ! "e#s sont #es /esoins %" noya" en terme %e 0estion mémoire(
- A+t+on to"1o"rs /esoin %e services %e 0estion mémoire(
- ! "e##e tai##e %e mémoire '2ysi&"e 'e"t+on con.i0"rer s"r "ne mac2ine(

Es'ace A%ressa0e Process"s

- A##ocation / 3i/ération %e 4ones mémoires 'o"r #e 'rocess"s
- 5i..érence entre
  - A##ocation / #i/ération %$"n interva##e %$a%resses
  - A##ocation / #i/ération %es 'a0es '2ysi&"es so"s+ 1acentes6

Exercice N° 3 2

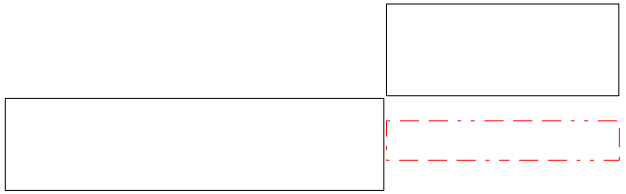
- ! "e#s sont #es a' 'e#s systèmes7 services &"i mo%i.ient #$es'ace %$a%ressa0e %$"n 'rocess"s(
- ! "e#s a' 'e#s7 accès7 instr"ctions in.#"ent s"r #$a##ocation7 #a 'résence7 #a %é+a##ocation %es 'a0es '2ysi&"es so"s+1acentes(

A''e#s et ré0ions

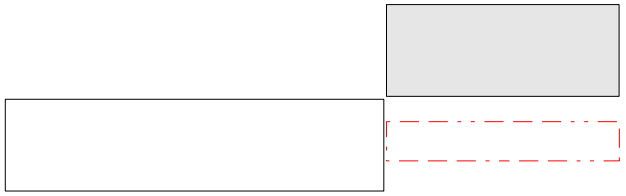
- 8é0ion9
  - :nterva##e %$a%resses conti0" avec #es m;mes %roits %$accès)#ect"re7 #ect"re/écrit"re7 exéc"tion *
- A''e#s9

  brk
  execve
  _exit
  fork
  mmap, munmap
  shmat, shmdt...

8é0ions

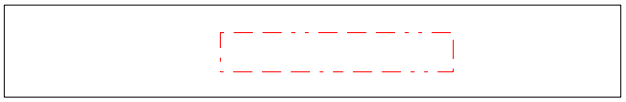


Ajout d'une région contiguë



Ajout d'une région contiguë
avec des droits différents

Réduction d'une région

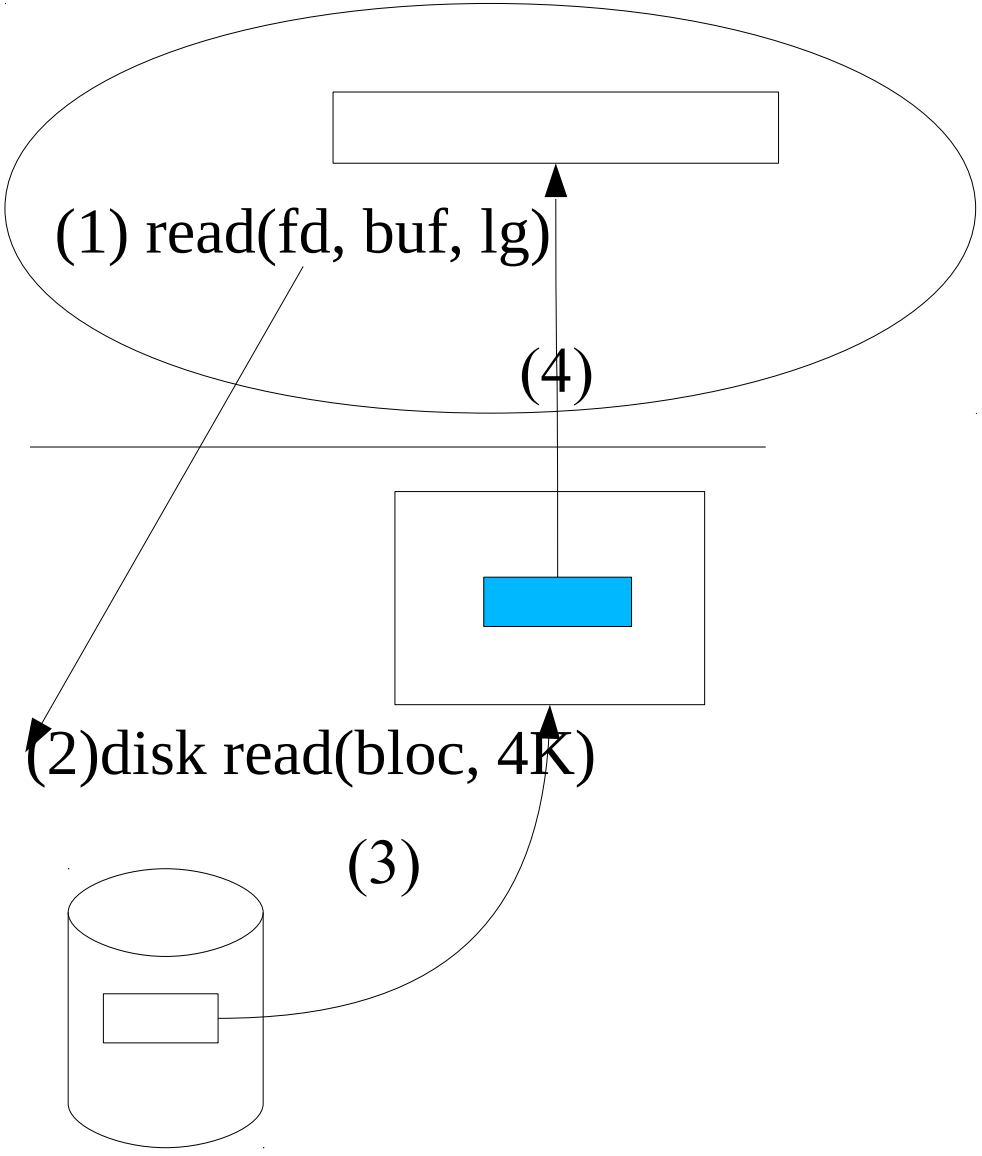


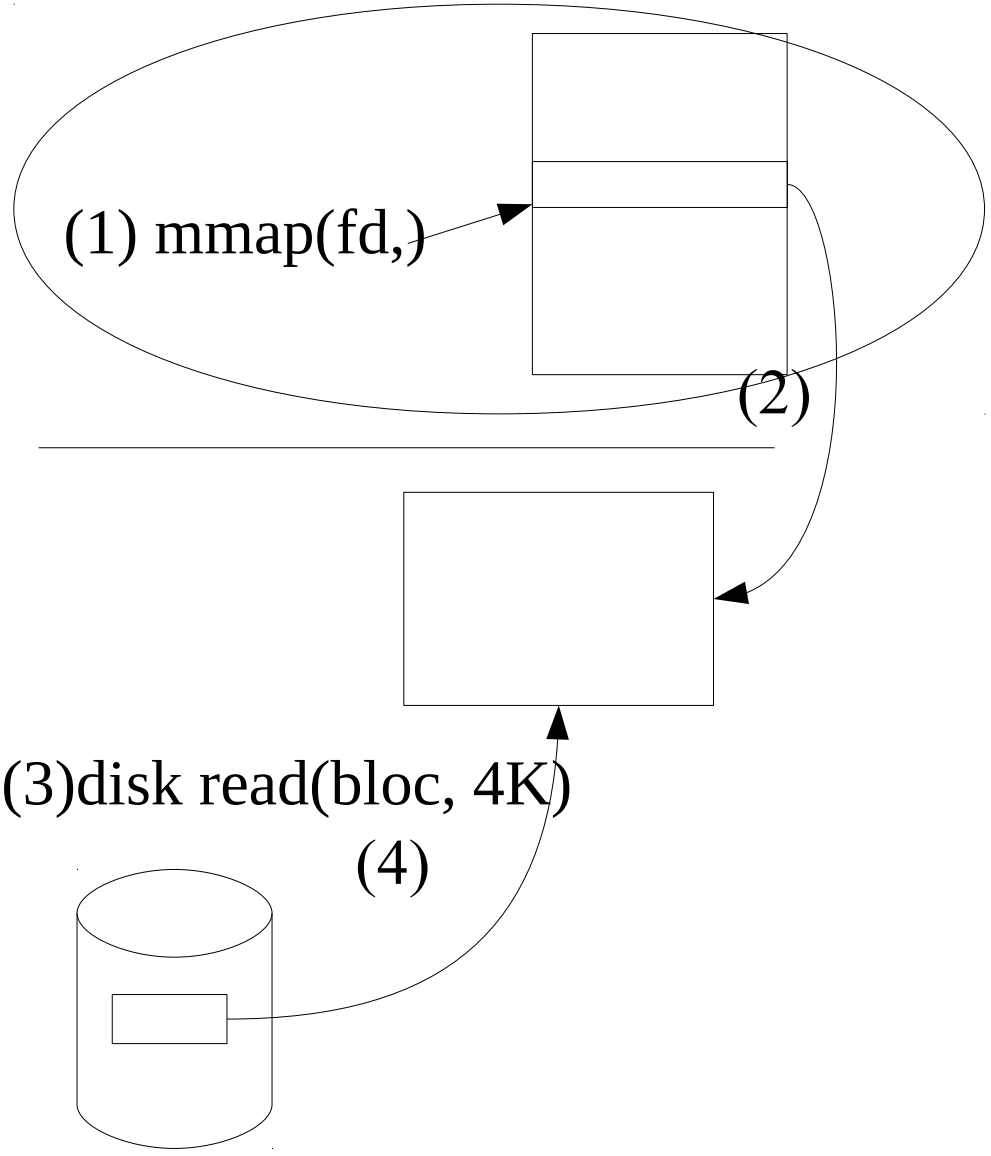
Une région => 2 régions

Memory ma'

- void *mmap(void *addr, size_t length,
  int prot, int flags ,
  int fd, off_t offset);
  - <rée "ne ré0ion mémoire %e tai##e length , #$a%resse addr addr[! 'ermet %$accé%er a" caractère %" .ic2ier fd &"i se tro"ve ;tre , #a 'osition offset 'ar ra''ort a" %é/"t %" .ic2ier
  - Si MAP=AN>N?M>-S @AB ma##oc
  - Protection©9 8EA5/D8:EE/EFE<
  - G#a0s©9 SHA8E5/P8:IAEE

8ea% vs Memory ma'





Memory #ocJ

- int mmap(void *addr, size_t length);
  - lerro"i##e en mémoire #es 'a0es corres'on%antes
    - Ne .orce 'as #es 'a0es '2ysi&"es , ;tre a##o"ées
    - Em';c2e #es 'a0es a##o"ées)maintenant o" '#"s tar%* %$;tre KCsLa''ées o"tCM
- m#ocJa##
  - M<3=<-88ENE
  - M<3=G-E-8E
- '#ocJ

ma##oc

- En réa#ité7 '#"s com'#exe &"e #$"ti#isation "s"e##e
  - I aria/#e %$"n système , #$a"treC6
    - <. 2tt'9//'"/#i/ /o"#%er i/m com/in.ocenter/aix/vNr1/in%ex 1s'(to'icA O2Gcom i/m aix 0en'ro0cO2G%ocO2G0en'ro0cO2Gsys=mem=a##oc 2tm
  - -ti#ise mma' 'o"r #es 0ran%s /#ocs
  - 8é%"it)o" 'as* #e tas si 'ossi/#eC6
  - KC<ac2eCM
  - >'tions %e KC%e/"OCM
- Mémoire 0arantie o" 'asC(

3in"x >>M

- <omment "ti#iser KCa" mie"xCM #a 8AMC(
  - Garantir &"e #a mémoire)virt"e##ement* a##o"ée sera e..ectivement %is'oni/#e &"an% on en a"ra /esoin
    - Mémoire KCréservéeCM m;me si 'as "ti#isée
    - AB '#ace #i/re non "ti#isée
  - Gaire %" KCs"r/ooJin0CM
    - Eant &"e #a mémoire)virt"e##ement* a##o"ée rée##ement "ti#isée est '#"s 'etite &"e #a mémoire '2ysi&"e)'#"s sLa' évent"e#* AB Pas %e 'ro/#ème
    - Sinon AB on 'e"t se tro"ver , co"rt %e mémoire s"r "n %é.a"t %e 'a0e AB se0mentation .a"#t P s"r "ne a%resse va#i%eC69+)

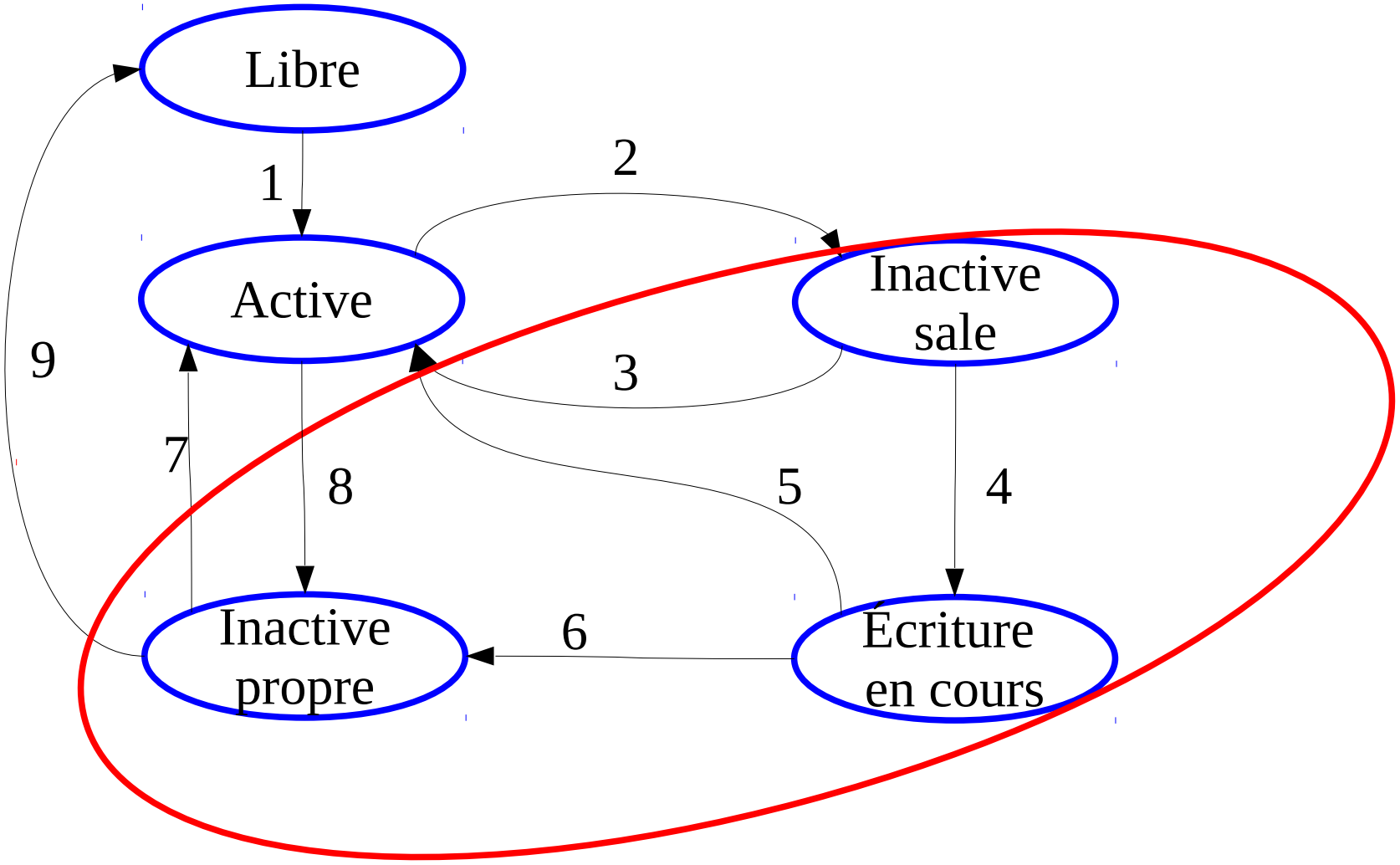
3in"x >>M

- /'roc/sys/vm/'anic=on=oom
  - Si , co"rt %e mémoire AB 'anic
- /'roc/sys/vm/oom=Ji##=a##ocatin0=tasJ
  - OC9 examine to"s #es 'rocess"s
  - 1C9 t"e #e 'rocess"s &"i %éc#enc2e #e mécanisme
- /'roc/sys/vm/overcommit=memory
  - OC9 véri.ication KC#axisteCM
  - 1C9 'as %e véri.ication
  - 2C9 es'ace virt"e# @ SLa' Q)8AM R ratio/100*

3in"x >>M

- /'roc/sys/vm/overcommit=ratio
  - 8atio %e #a .orm"#e 'réce%ente
- /'roc/S'i%T/oom=score
- /'roc/S'i%T/oom=score=a%1

<yc#e %e vie %es 'a0es '2ysi&"es



http://www.redhat.com/magazine/001nov04/features/vm/#vm-page-state-machine

Eransition %es 'a0es

- 1C9 A##ocation %$"n 'a0e)exC9 %é.a"t %e 'a0e*
  - 3a 'a0e va rester active tant &"$e##e sera accé%ée ré0"#ièrement en #ect"re / écrit"re
- 2C9 Pa0e mo%i.iée mais non accé%ée %e'"is KC"n certain tem'sCM
- 3C9 Pa0e mo%i.iée &"e #$on accè%e , no"vea" avant &"$on ait entre'ris
- UC9 Pa0e mo%i.iée &"e #$on ve"t sa"ve0ar%er s"r %is&"e
  - SLa' o" système %e .ic2iers
- 5C9 Gina#ement cette 'a0e &"e #$on vo"#ait sa"ve0ar%er7 est , no"vea" accé%ée

Eransition %es 'a0es

- N 9 3es co'ies mémoires et %is&"es ont #e m;me conten"7 #a 'a0e n$a 'as été accé%ée récemment
- 7C9 #a 'a0e 'ro're vient %$;tre accé%ée %e no"vea"
- VC9 "ne 'a0e non mo%i.iée 'as accé%ée récemment
- WC9 !"an% e nom/re %e 'a0es #i/res est tro' /as7 on re'e"'#e #e stocJ en recyc#ant %es 'a0es inactives 'ro'res
  - 8emise , 4éro

- ti#isation %e %émons

- Po"r 0érer #a mémoire7 3in"x "ti#ise %es t2rea%s
  - I ariations s"ivant #es versions
  - Jscan% / /%.#"s2
  - JsLa'% / J/#ocJ%

Xesoins %" noya"

- Ero"ver %es 'a0es #i/res 'o"r #es attri/"er , "n 'rocess"s)co%e7 %onnées7 'i#e7 tas *
- Ero"ver %es 'a0es #i/res 'o"r y a##o"er %es str"ct"res %e %onnées %" noya")tasJ=str"ct7 .ic2iers7 'i#es systèmes7 *
- Accé%er a" conten" %es 'a0es)'a0e .rame*
  - 3ect"re / écrit"re7 insta##er %" co%e7 %es %onnées
  - <ac2er %es /#ocs %is&"es7
  - Permettre #$accès en 5MA 'ar #es 'éri'2éri&"es
- Et '#"s encore

5escri'te"r %e 'a0e

- <2a&"e 'a0e '2ysi&"e est re'résentée 'ar "n %escri'te"r %e 'a0e7 m;me si e##e n$est 'as "ti#isée %ans "n es'ace %$a%ressa0e

```
flags       (locked, error, dirt", slab,...)
_count       compteur de r#f#rence
_mapcount    nombre de $%& r#f#ren'ant
private      utilisation par le no"au
mapping      (tile pour cache de pages, anon.
index        (tilisations diverses
lru          )ha*nage least recentl" used
```

Accès a"x 'a0es '2ysi&"es

- 3e noya" %oit 'o"voir accé%er , to"tes #es 'a0es via son es'ace virt"e# %e 1 Gi0a6
- Si mémoire '2ysi&"e %e 'etite tai##e 'as %e ' /
- Si 0ran%e mémoire '2ysi&"e7 comment .aire(
- Pas /esoin %$avoir accès , to"tes #es 'a0es en m;me tem's >".
- Par ai##e"rs certaines 'a0es)mémoire /asse* sont #es se"#es 'o"vant .aire %" 5MA /"s :SA

Accès a"x 'a0es '2ysi&"es

- 3e noya" 'ar #"i+m;me a "ne tai##e Y#imitéeY7 0énéra#ement Z3 o" U MX
- A ra''orter , 1GX6
- 3e noya" "ti#ise "n Yma''in0Y %e #a mémoire '2ysi&"e #"i 'ermettant %$accé%er a"x 'a0es %ont i# a /esoin
- Ma''in0 varia/#e s"ivant #a tai##e %e #a mémoire '2ysi&"e
  - Mémoire @ VWNMX ma''in0 .ixe 1 'o"r 1
  - Mémoire BVWNMX B UG/ ma''in0 %ynami&"e
  - Mémoire B UG i%em7 mais 3 nivea"x %e ta/#es

Mémoire Non -ni.orme

- En Ystan%ar%Y7 3in"x 0ère #es mac2ines N-MA
- Notion %e Yn ["%sY)*no%e** mémoire
- <2a&"e *n ["%* %istin0"e trois ty'es %e 4ones
  - \>NE=5MA9 @ 1N MX 'o"r 5MA/:SA en 'riorité
  - \>NE=N>8MA3 @ VWN MX
  - \>NE=H:GH B VWNMX
- 3es a##ocations mémoire 'e"vent tenir com'te %e ces 4ones

Poo# 8éservé

- A##ocation mémoire9
  - Si 'a0es #i/res7 >]
  - Sinon7 #i/érer %es 'a0es7 atten%re #a .in %e #i/ération7 reto"rner "ne %es 'a0e)s* #i/érée)s*
    - Pe"t nécessiter "ne attente /#o&"ante
  - 5ans certains cas7 on ne 'e"t 'as /#o&"er #e c2emin %$exéc"tion en co"rs %ans #e noya")traitement %$interr"'tion*
  - A##ocation Y*atomi&"e*Y GGP=AE>M:< ne /#o&"e 1amais
    - -ti#ise %es 'a0es Yréservées Y)entre 12V]/ et NU M/*
    - Prises 'ro'ortionne##ement %ans \>NE=5MA7 \>NE=N>8MA3

A##ocations mémoire

- 3e noya" 3in"x a '#"sie"rs a##ocate"rs
  - A##ocation %e 'a0es 9 ***buddy system***
  - A##ocations %$o/1ets ty'és)ex9 *tasJ=str"ct** 9 ***slab***
  - A##ocations %e 4ones mémoire)ty'e ma##oc*

X"%%y system

- Permet %$a##o"er %es 4ones
  - 5e tai##e a#i0née s"r "ne ' "issance %e 2
    - 1 'a0e7 2 'a0es7 U 'a0es7 V 'a0es 512 et 102U 'a0es
  - 3a 'remière 'a0e reto"rnée a "ne a%resse '2ysi&"e a#i0née s"r #a m;me ' "issance %e 2
- Str"ct"re %e %onnées9
  - 11 t;tes %e #iste c2a^nées corres'on%ant a"x 11 tai##es Y%eman%a/#esY

X"%%y9 A##ocation

- Si #a #iste %e #a tai##e %eman%ée)2^n7 ran0 n* n$est 'as vi%e7 on renvoie 'remier é#ément %e cette #iste
- Si #a #iste %e ran0 n est vi%e
  - >n re0ar%e #a #iste %e ran0 s"'érie"r)nQ1*
  - Si 'as vi%e
    - >n 'ren% #e 'remier é#ément7 et on #e casse en 2 é#éments %e #a tai##e in.érie"re
    - >n ran0e #$é#ément restant #i/re %ans #a #iste %e ran0 in.érie"r7 et on renvoie #$a"tre
  - Si vi%e7 on 'asse a" ran0 nQ2
  - Si ran0 10 et vi%e7 erre"r

X"%%y9 3i/ération

- >n #i/ère "n é#ément 3 %e ran0 N
- Si #a #iste %e ran0 N contient "n é#ément E te# &"e E et 3 sont a%1acents)/"%%y*9
  - Gin %e 3 A 5é/"t %e E

    o"
  - 5é/"t %e E A Gin %e 3
  - >n com/ine ces 2 é#éments en "n é#ément %e ran0 NQ1
- Et on recommence

<ac2es %e Pa0es / <P-

- A''e#er #$a##ocate"r /"%%y 'o"r to"tes #es re&";tes serait tro' .ré&"ent AB ine..icace
- 3e Noya" maintient)'ar <P-* "n cac2e %e 'a0es7 en .ait 2
  - Pa0es Y2otY)i# y a "ne c2ance &"e #e cac2e mémoire contienne %es #i0nes %e cette 'a0e*
  - Pa0es Yco#%Y 'ar exem'#e 'o"r #es accès 5MA)ne 'asse 'as 'ar #e cac2e*
- Si nivea" %e cac2e tro' 2a"t
  - 8eto"r %es 'a0es en excès vers #e /"%%y
- Si nivea" %e cac2e tro' /as
  - 3e re0arnir %e'"is #e /"%%y

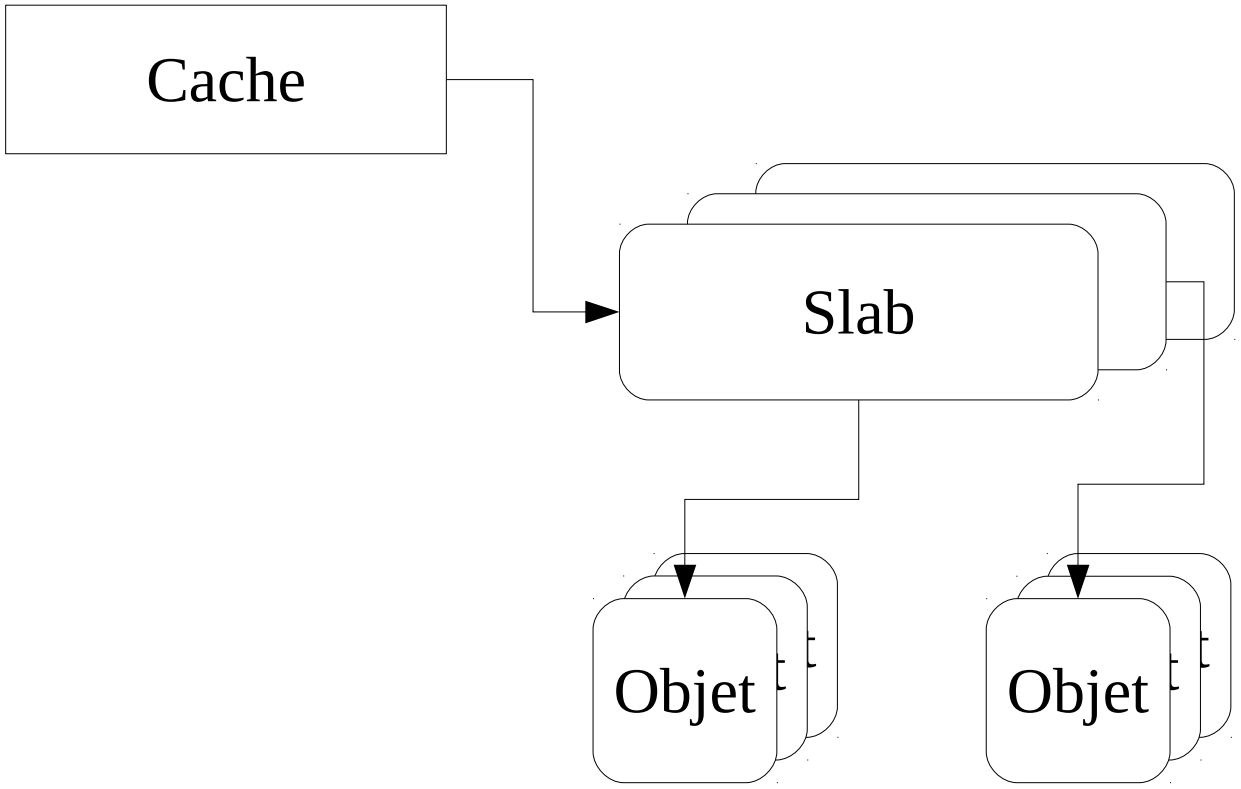
A##ocate"r YS#a/Y)'avé7 %a##e*

- A##ocation %e 'a0es n$est 'as 'rès e..icace 'o"r a##o"er %ynami&"ement %es o/1ets).ic2iers7 'rocess"s7 *
- 5i..érents ty'es %$o/1ets a##o"és/#i/érés %e manière ré'étée
- _viter #a .ra0mentation %e #a mémoire)c. ma##oc en es'ace "ti#isate"r*
- Ne 'as a#i0ner systémati&"ement #es o/1ets s"r %es .rontières %e 'a0e AB mei##e"re "ti#isation %es cac2es matérie#s

A##ocate"r YS#a/Y

- 5érivé %e So#aris 2 U7
- <2or"s avait "n mécanisme simi#aire
- :ntro%"it #a notion %e Ycac2e %$o/1etsY
  - \one o` to"s #es o/1ets sont %e m;me ty'e9 'rocess"s7 %escri'te"rs %e .ic2iers7
- En .ait "n cac2e est s"/%ivisé en Ys#a/sY
  - -n Ys#a/Y est constit"é %e 'a0es '2ysi&"es conti0"as7 to"s #es s#a/s ont #a m;me tai##e
  - <2a&"e Ys#a/Y contient %es o/1ets %" ty'e 'ré%é.ini

A##ocate"r YS#a/Y



5escri'te"r %e cac2e

kmem_cache_t

+ arra" // cac2e #oca# 'ar <P- %$o/1ets #i/res
+ ob+size //tai##e %$o/1et
+ gfpflags // in%icate"rs a##ocation /"%%y system
+ lists // 0estion %es s#a/s)c. 'a0e s"ivante*
+ ctor, dtor // initia#isation7 %estr"ction %$o/1et
+ color // n/ %e co"#e"rs %es s#a/s
+ etc

Accès a"x s#a/s

```
kmem_list,
  struct list_head slabs_partial
    -- liste des slabs ni plein ni vide
  struct list_head slabs_full
    -- liste des slabs pleins
  struct list_head slabs_free
    -- liste des slabs totalement vides
  struct arra"_cache *shared
    -- cache partag# par tous les )$(
```

5escri'te"r S#a/

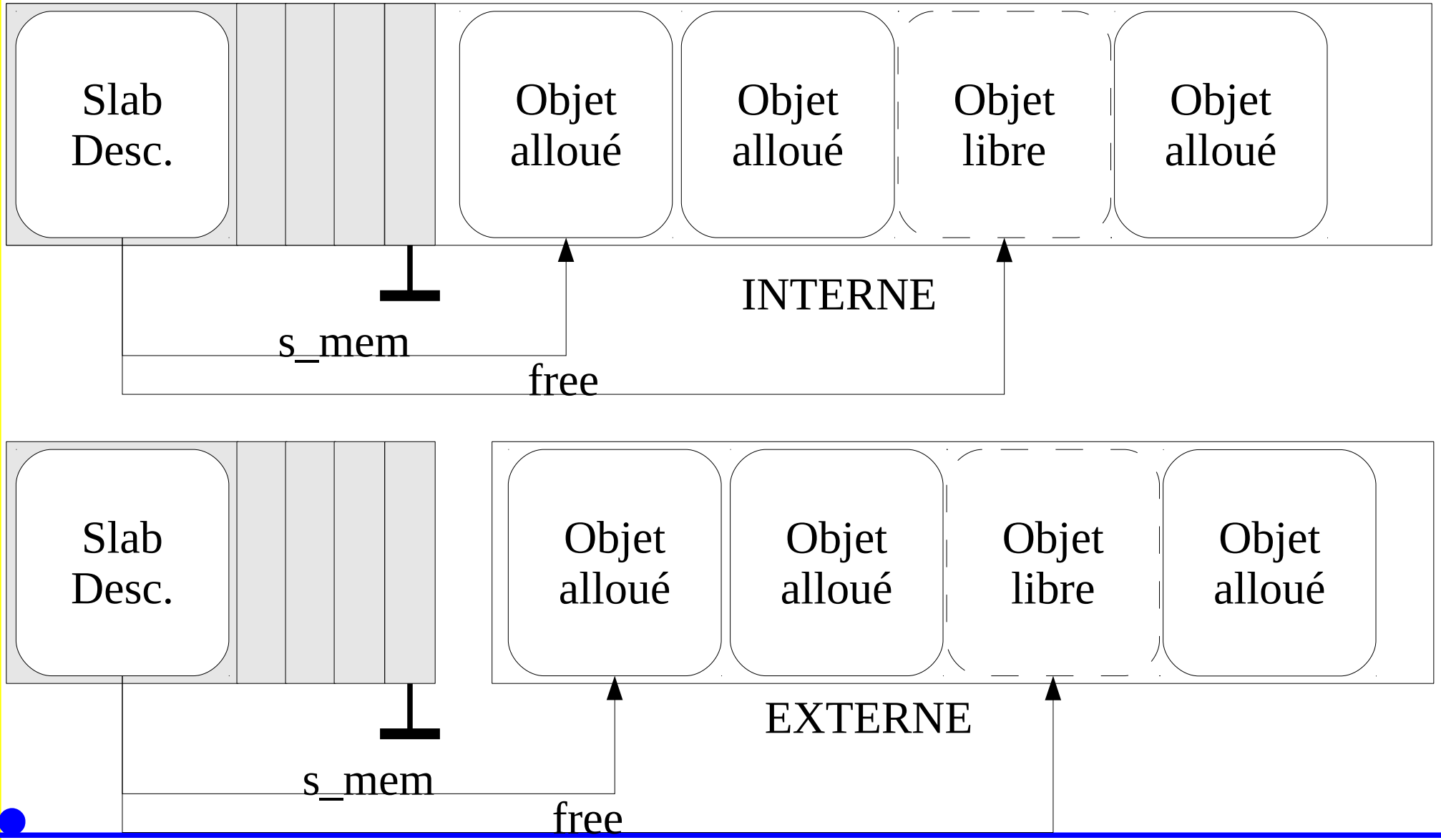
```
struct list_head list
-- cha*nage des slabs (partiel, vides, pleins)
unsigned long colouroff
-- offset premier ob+et, variation cache
void * s_mem
-- adresse premier ob+et
unsigned int inuse
unsigned int free
```

- 3es %escri'te"rs 'e"vent ;tre internes o" externes

S'écia#isation %es cac2es

- <ac2es 0énéra"x
  - cac2e=cac2e9 Po"r 0érer #es %escri'te"rs %es)a"tres* cac2es6
  - \ones mémoires non ty'ées7 mais %e #on0"e"rs 'ré%é.inies9 327 NU7 12V
    - S'écia#isées entre accessi/#es via 5MA o" non

S#a/ et >/1ets


Slab
Desc.
Objet
alloué
Objet
alloué
Objet
libre
Objet
alloué
INTERNE
s_mem
free
Slab
Desc.
Objet
alloué
Objet
alloué
Objet
libre
Objet
alloué
EXTERNE
s_mem
free

A##ocation / 3i/ération o/1ets

```
void * kmem_cache_alloc (
            kmem_cache_t *cachep,
            int flags)

void kmem_cache_free(
            kmem_cache_t *cachep,
            void  *ob+p)
```

5escri'te"r Mémoire

- <2a&"e 'rocess"s)"ti#isate"r* a "n %escri'te"r mémoire)mm_struct* accessi/#e %e'"is son %escri'te"r %e 'rocess"s)task_struct*7 'arta0é 'ar #es %i..érentes t2rea%s
- <ontient9
  - '0%9 'ointe"r 0#o/a# %irectory)ta/#e %e 'a0es*
  - 5i..érentes a%resses 'rocess"s9 co%e7 %onnées7 tas7 'i#e7 ar0s7 env7
  - 3iste %es %escri'te"rs %e ré0ions %" 'rocess"s
  - Statisti&"es7

5escri'te"r Mémoire

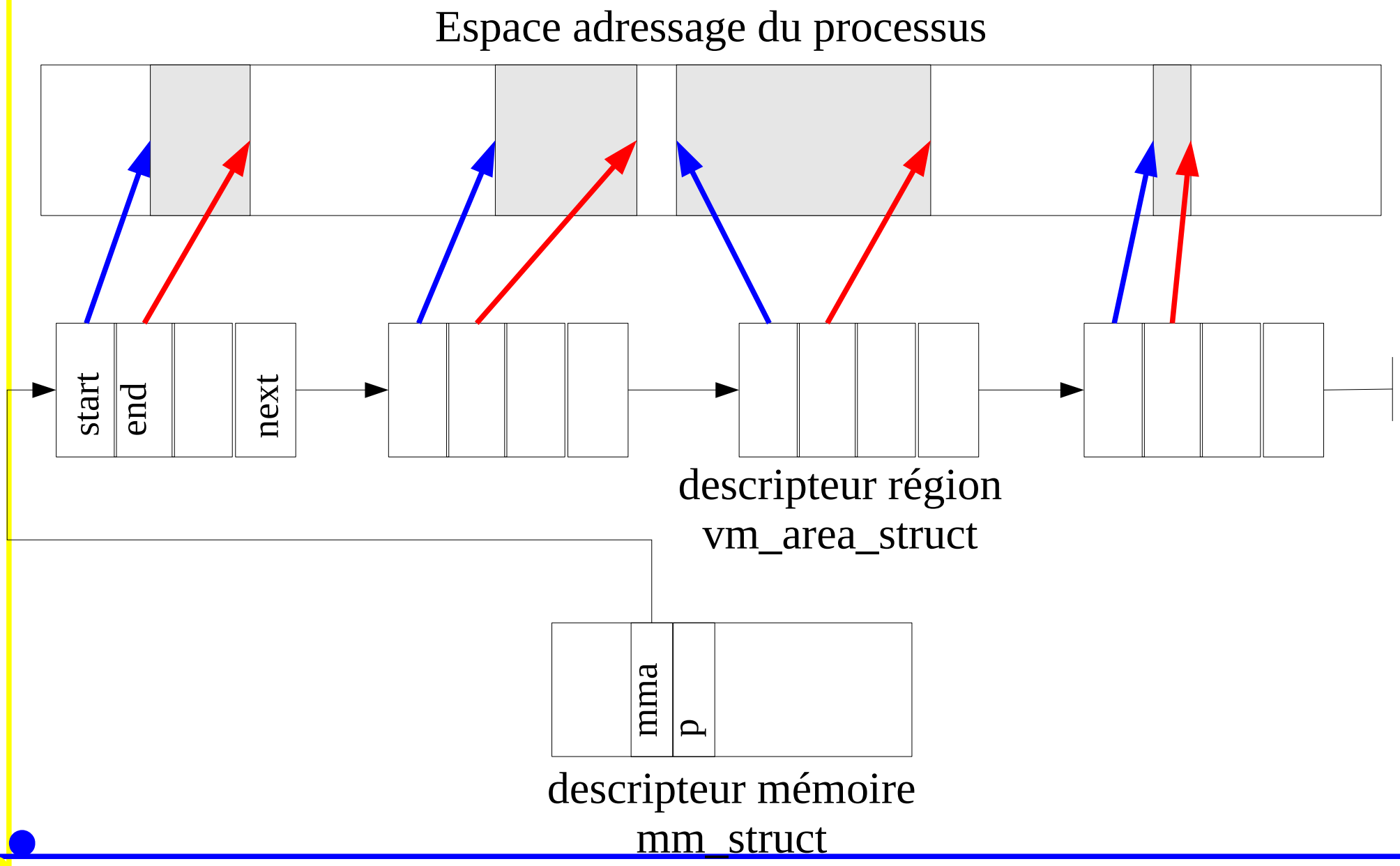
- <2a&"e 'rocess"s)"ti#isate"r* a "n %escri'te"r mémoire)mm_struct* accessi/#e %e'"is son %escri'te"r %e 'rocess"s)task_struct*7 'arta0é 'ar #es %i..érentes t2rea%s
- <ontient9
  - '0%9 'ointe"r 0#o/a# %irectory)ta/#e %e 'a0es*
  - 5i..érentes a%resses 'rocess"s9 co%e7 %onnées7 tas7 'i#e7 ar0s7 env7
  - 3iste %es %escri'te"rs %e ré0ions %" 'rocess"s
  - Statisti&"es7

5escri'te"r %e ré0ion

```
struct vm_area_struct .
   vm_start -- adresse de d#but
   vm_end   -- adresse de fin / 0
   vm_next  -- prochaine r#gion
   vm_page_prot -- protection
   vm_rb    -- gestion arbre rouge1noir
...
2
```

- Gérées %ans "ne #iste sim'#ement c2a^née)next* et a"ssi %ans "n ar/re ro"0e+noir

Gestion Mémoire

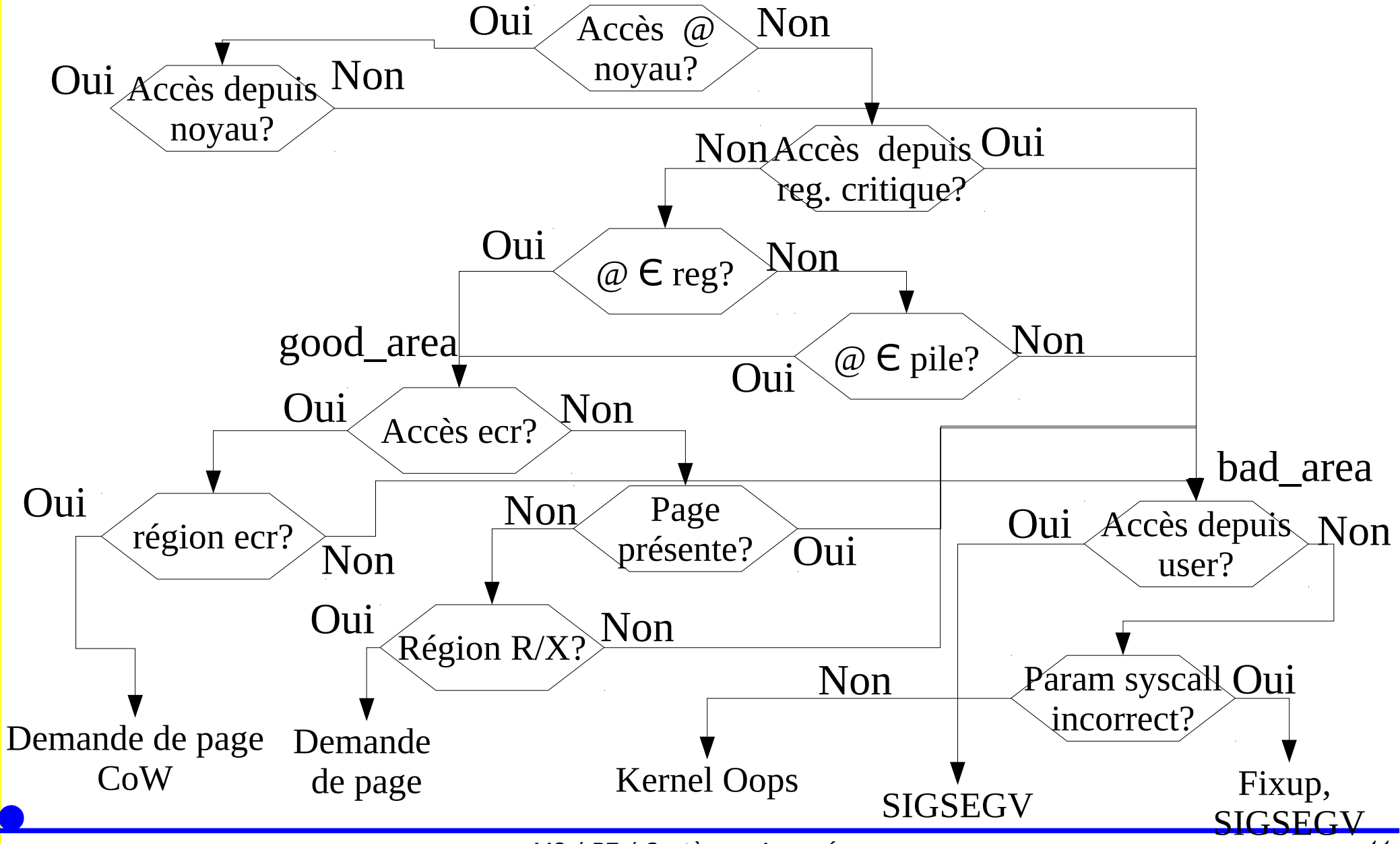


Eraitement %es exce'tions

3$a%resse .a"tive est+e##e %ans #$es'ace %$a%ressa0e %" 'rocess"s(

- >"i
- Est+ce "n 'ro/#ème %e 'rotection(
  - >"i AB Envoyer "n si0na#)S:GSEGI*
  - Non AB Accès va#i%e7 a##o"er "ne 'a0e '2ysi&"e
- Non
- Accès %e'"is mo%e "ti#isate"r(
  - >"i AB Envoyer "n si0na#)S:GSEGI*
  - Non AB X"0 noya" 9+)

Eraitement %es exce'tions



IMLare et #a mémoire

- Po"r 0érer e..icacement #a mémoire7 ImLare "ti#ise '#"sie"rs tec2ni&"es©9
  - Erans'arent Pa0e S2arin0
  - Xa##oonin0
  - <om'ression
  - SLa'

IMLare EPS

- 5ans c2a&"e IM to"rne "n >S
  - So"vent i%enti&"e %$"ne IM , #$a"tre
  - Pas %e so"rce comm"ne comme #es ino%es 'o"r #es 'rocess"s
  - Nom/re"ses 'a0es i%enti&"es)'#eines %e 4éros*
- :%enti.ation©9
  - <2ecJs"m '"is com'araison /ytes//ytes
  - -ti#isation %e <>D
  - :n2i/é s"r #es 3ar0e Pa0es)2MX*

IMLare Xa##oonin0

- 5ans c2a&"e IM o` #es o"ti#s ImLare sont insta##ésC9
  - >n insta##e "n %river KC/a##onCM
  - ! "i ca"se avec #$2y'ervise"r , #$ins" %e #$>S
  - Gon.#eC9 5eman%e %e #a mémoire , #$>S
    - Po"r #a re%onner so"s #e mantea" , #$2y'ervise"r
  - 5é0on.#eC9 réc"'ère ce &"$i# a 'r;té et #e ren% , #$>S
- Permet %e trans.érer tem'orairement #a mémoire %$"ne IM 'e" active , "ne IM %ans #e /esoin

IMLare <om'ression

- P#"tbt &"e KCsLa''erCM
  - <om'resser "ne 'a0e
  - Si tai##e com'ressée @ c 'a0e AB >]
  - P#"s ra'i%e &"$"ne entrée/sortie %is&"e
- Eec2ni&"e 'as s'éci.i&"e , ImLare
  - A:F AME 'ar exem'#e

IMLare SLa'

- 5ernier nivea" 'o"r .aire .ace , #a %eman%e mémoire©9
  - -ti#iser #e sLa'/o"t 7 sLa'/in
  - 5e #a mémoire K©'2ysi&"e©M %es IM
- 3$>S %ans #a IM 'e"t #"i m;me "ti#iser %" sLa'
- A éviter en 0énéra#7 sinon #es 'er.ormances se %é0ra%ent très série"sement