

C++內存管理 101 – 從最基礎到最繁複

課程簡介

內存 (memory, 臺灣術語稱為“記憶體”) 是電腦中的“腦”嗎？CPU 才是腦，CPU 才是計算機的三魂六魄。但若沒有內存，一切只存在於虛無縹渺間，等同於不存在。

內存曾經是最寶貴也最昂貴的週邊資源，現代程序員無法想像 DOS 時代對內存的錙銖必較。

俱往矣，且看今朝。我們(似乎)有用不完的便宜內存。但表象之下是操作系統和標準庫做了大量工作。而如果你開發內存高耗軟件，或處於內存受限環境下(例如嵌入式系統)，就有必要深刻了解操作系統和標準庫為你所做的內存管理，甚至需要自行管理內存。

本課程分為六講：

第一講：Primitives

C++語言中與內存相關的所有基礎構件 (constructs)，包括 `malloc/free`, `new/delete`, `operator new/operator delete`, `placement new/placement delete`，我將探討它們的意義、運用方式和重載方式。並以此開發一個極小型內存池 (memory pool)。

第二講：`std::allocator`

標準庫的興起，意味我們可以擺脫內存管理的繁複瑣屑，直接使用容器。但是容器背後的分配器(allocator)攸關容器的速度效能和空間效能。我將比較 Visual C++, Borland C++, GNU C++標準庫中的 allocator，並深入探索其中最精巧的 GNU C++ allocator 的設計。

第三講：`malloc/free`

`malloc/free` 是所有內存管理手段的最後一哩；通過它才和操作系統搭上線。當然你也可以直接調用 system API，但不建議。因此理解 `malloc/free` 的內部管理至為重要。我將以 Visual C++ 的 CRT (C RunTime Library)所帶的 `malloc/free` 源代碼為基礎，深度探索這最基礎最關鍵的內存分配與釋放函數。

第四講：`loki::allocator`

即使知名如 GNU C++ pool allocator，也有其小缺陷。Loki (一套作風前沿的程序庫) 的 allocator 設計精簡功能完整幾無缺點，很值得我們深究。

第五講：其他主題

除了 std::allocator，GNU C++ 還帶不少 allocators，它們不是標準庫的一部分，可視為標準庫的擴充。我將探討這些擴充的 allocator，特別是 bitmap allocator。

我們談的不只是應用，還深入設計原理與實現手法。在理解了這麼多底層 (Windows Heap, CRT malloc/free, C++ new/delete, C++ allocators) 之後，也許你終於恍然大悟，再不需要自行管理內存了；或也許你終於有能力想像，該在何處以何種方式加強內存管理。

你將獲得整個 video 課程的完整講義（也就是 video 呈現的每一張投影片畫面），和一個完整程序包括代碼文件.cpp 和可執行文件.exe。你可以在視聽過程中隨時停格仔細閱讀講義，細細咀嚼我所繪製的各種示意圖和源代碼之間的流動路線——這的確很需要時間和腦力，卻能令你腦洞大開。

侯捷簡介：程序員，軟件工程師，臺灣工研院副研究員，教授。曾任教於中壢元智大學、上海同濟大學、南京大學。著有《無責任書評》三卷、《深入淺出 MFC》、《STL 源碼剖析》…，譯有《Effective C++》《More Effective C++》《C++ Primer》《The C++ Standard Library》…。

以下這份不算太細緻的主題劃分，協助您認識整個課程內容，以及在視聽過程中正確翻找講義。內中的編號就是講義的頁碼。

內存管理 101 – 從最基礎到最複雜

主題劃分

第一講：C++ 內存構件

Overview 1-10

內存分配的每一層面 11

四個基本層面的用法 12-14

基本構件之一 new/delete expressions 15-17

基本構件之二 array new/delete 18-21

基本構件之三 placement new/delete 22

基本構件之分配流程 23-24

基本構件之重載 25-34

Per-class allocator 第一版本 35-36

Per-class allocator 第二版本 37-38

Common static allocator (第三版本) 39-41
Macro allocator (第四版本) 42-43
GNU C++ allocator (第五版本) 樣貌 44
雜項討論 45-48

第二講：std::allocator

內存塊佈局 52
VC6 allocator 53
BC5 allocator 54
GNU allocator 55-60
GNU allocator 行為剖析 61-76
GNU allocator 源碼剖析 77-87
GNU allocator 檢討 88-89
GNU allocator 監視 90-91
GNU allocator 移植到 C 語言 92

第三講：malloc/free

VC6 和 VC10 的 malloc 比較 96-97
Small Block Heap (SBH) 初始化 98-99
SBH 行為分析 – 區塊大小之計算 100-104
SBH 行為分析 – 數據結構 105-107
SBH 行為分析 – 分配之詳細圖解 108
SBH 行為分析 – 分配+釋放之連續動作圖解 109-115
SBH 檢討 116-122

第四講：Loki::allocator

上中下三個 classes 分析 127
Loki::allocator 行為圖解 128-134
class Chunk 分析 135-137
class FixedAllocator 分析 138-140
Loki::allocator 檢討 141

第五講：Other Issues

GNU C++ 對 allocators 的描述 144-149
VS2013 標準分配器與 new_allocator 150
G4.9 標準分配器與 new_allocator 151
G4.9 malloc_allocator 152
G4.9 array_allocator 153-155

G4.9 debug_allocator 156
G4.9 _pool_alloc 157-159
G4.9 bitmap_allocator 160-170
G4.9 使用 G4.9 分配器 171-172

-- the end



內存管理

從平地到萬丈高樓

Memory Management 101

第一講 primitives

第二講 std::allocator

第三講 malloc/free

第四講 loki::allocator

第五講 other issues



侯捷

萬丈高樓平地起

源碼之前
了無秘密



你應具備的基礎

- 曾經動態分配並使用 **memory**
- 曾經使用過 C++ 標準庫的容器 (**containers**)



我們的目標

從平地到萬丈高樓，

從最基礎的 C++ 語言構件
到高知名度的內存管理器，

徹底了解高高低低的方方面面。



網絡有許多資源 …

Doug Lea's Home Page

Surface mail: Doug Lea, [Computer Science Department](#), [State University of New York at Oswego](#), [Oswego](#), NY 13126 | Voice: 315-312-2688 | Fax: 315-312-5424 | E-Mail: dl@cs.oswego.edu | [vita](#) | [shortbio](#)

Documents

Books

- [Online supplement](#) to the book [Concurrent Programming in Java: Design Principles and Patterns](#), (second edition) published November 1, 1999 by [Addison-Wesley](#)
- HTML edition of the book [Object Oriented System Development](#), by [Dennis Lea](#), and Penelope Faure, originally published in 1993 by [Addison-Wesley](#).
- [Companion site](#) for [Java Concurrency in Practice](#) by Brian Goetz, with Tim Peierls, Bowbeer, David Holmes, Doug Lea. Addison-Wesley, 2006.

Software

- [JSR166 \(JDK 1.5 java.util.concurrent and jsr166x\) specs, etc.](#)
- [Concurrency utilities for Java EE](#)
- Version 1.3.4 of [util.concurrent](#), a [Java](#) package containing locks, queues, threads, and other primitives, a precursor to `java.util.concurrent`.
- [collections](#), a [Java](#) package that was a precursor to some JDK collections.
- [Microscope](#): a cute Java applet.
- [Version 2.8.6 of malloc.c](#), and [malloc.h](#). Also a short article on [the design of this](#) some [plots](#). Also a [malloc tracer](#) written by [Wolfram Gloger](#) and a [sample trace](#) of [malloc](#).

A Memory Allocator

by [Doug Lea](#)

[A German adaptation and translation of this article appears in [unix/mail](#) December, 1996. This article is now out of date, and doesn't reflect details of current version of malloc.]

Introduction

Memory allocators form interesting case studies in the engineering of infrastructure software. I started writing one in 1987, and have maintained and evolved it (with the help of many volunteer contributors) ever since. This allocator provides implementations of the standard C routines `malloc()`, `free()`, and `realloc()`, as well as a few auxiliary utility routines. The allocator has never been given a specific name. Most people just call it *Doug Lea's Malloc*, or *dlmalloc* for short.

Doug Lea 自1986年起潛心研究 malloc 算法，其作品被稱為 DL Malloc。目前 Linux 的 glibc 中的 malloc 算法就是直接來自 Doug Lea，其他平台的 malloc 實現也或多或少受到 DL 的影響。DL Malloc 源碼可下載自 Doug Lea 個人主頁：
<http://gee.cs.oswego.edu/dl/>



天寶當年, DOS 640K, extented memory.

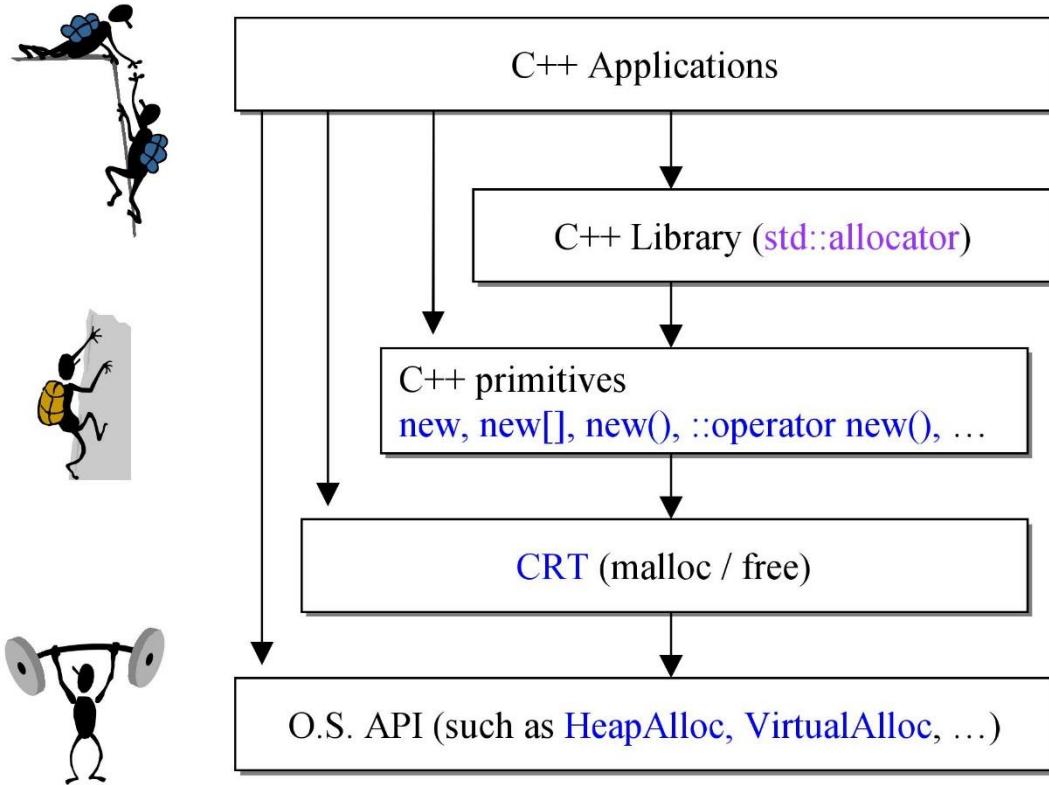


錙銖必較

俱往矣 且看今朝



C++應用程序, 使用memory的途徑





C++ memory primitives

分配	釋放	類屬	可否重載
<code>malloc()</code>	<code>free()</code>	C 函數	不可
<code>new</code>	<code>delete</code>	C++表達式 (expressions)	不可
<code>::operator new()</code>	<code>::operator delete()</code>	C++函數	可
<code>allocator<T>::allocate()</code>	<code>allocator<T>::deallocate()</code>	C++標準庫	可自由設計並以之搭配任何容器



C++ memory primitives

```
#01 void* p1 = malloc(512); //512 bytes
#02 free(p1);
#03
#04 complex<int>* p2 = new complex<int>; //one object
#05 delete p2;
#06
#07 void* p3 = ::operator new(512); //512 bytes
#08 ::operator delete(p3);
#09
#10 //以下使用 C++ 標準庫提供的 allocators。
#11 //其接口雖有標準規格，但實現廠商並未完全遵守；下面三者形式略異。
#12 #ifdef _MSC_VER
#13     //以下兩函數都是 non-static，定要通過 object 調用。以下分配 3 個 ints.
#14     int* p4 = allocator<int>().allocate(3, (int*)0);
#15     allocator<int>().deallocate(p4, 3);           ↗ 無用
#16 #endif
#17 #ifdef _BORLANDC_
#18     //以下兩函數都是 non-static，定要通過 object 調用。以下分配 5 個 ints.
#19     int* p4 = allocator<int>().allocate(5);
#20     allocator<int>().deallocate(p4, 5);
#21 #endif
#22 #ifdef _GNUC_
#23     //以下兩函數都是 static，可通過全名調用之。以下分配 512 bytes.
#24     void* p4 = alloc::allocate(512);
#25     alloc::deallocate(p4, 512);
#26 #endif
```



C++ memory primitives

```
#01 void* p1 = malloc(512); //512 bytes
#02 free(p1);
#03
#04 complex<int>* p2 = new complex<int>; //one object
#05 delete p2;
#06
#07 void* p3 = ::operator new(512); //512 bytes
#08 ::operator delete(p3);
#09
#10 #ifdef __GNUC__
#11     //以下兩函數都是 non-static，定要通過 object 調用。以下分配 7 個 ints
#12     void* p4 = allocator<int>().allocate(7);
#13     allocator<int>().deallocate((int*)p4, 7);
#14
#15     //以下兩函數都是 non-static，定要通過 object 調用。以下分配 9 個 ints.
#16     void* p5 = __gnu_cxx::__pool_alloc<int>().allocate(9);
#17     __gnu_cxx::__pool_alloc<int>().deallocate((int*)p5, 9);
#18 #endif
```



new expression

```
Complex* pc = new Complex(1, 2);
```

編譯器
轉為

```
Complex *pc;
try {
    1 void* mem = operator new( sizeof(Complex) ); //allocate
    2 pc = static_cast<Complex*>(mem);           //cast
    3 pc->Complex::Complex(1,2);                 //construct
    ... //注意：只有編譯器才可以像上面那樣直接呼叫 ctor
}
catch( std::bad_alloc ) {
    //若allocation失敗就不執行constructor
}
```

→ App. 欲直接調用 ctor, 可運用 placement new :
`new(p) Complex(1, 2);`

void *operator new(size_t size, const std::nothrow_t&)

... \vc98\crt\src\newop2.cpp

THROW0()

{ // try to allocate size bytes

void *p;

while ((p = malloc(size)) == 0)

{ // buy more memory or return null pointer

_TRY_BEGIN

if (_callnewh(size) == 0) break;

_CATCH(std::bad_alloc) return (0);

_CATCH_END

}

return (p);

}



The struct is used as a function parameter to `operator new` to indicate that the function should return a null pointer to report an allocation failure, rather than throw an exception.

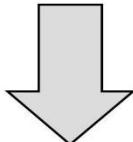
```
struct std::nothrow_t {};
```

—侯捷—



delete expression

```
Complex* pc = new Complex(1, 2);  
...  
delete pc;
```



編譯器轉為...

```
pc->~Complex();           //先析構  
operator delete(pc);    //然後釋放內存
```

```
void __cdecl operator delete(void *p) _THROW0()  
{ // free an allocated object  
    free(p);  
}
```

...\\vc98\\crt\\src\\delop.cpp



Ctor & Dtor 直接調用



```
96     string* pstr = new string;
97     cout << "str= " << *pstr << endl;
98
99 //! pstr->string::string("jjhou");
100 //! [Error] 'class std::basic_string<char>' has no member named 'string'
101 //! pstr->~string(); //crash
102     cout << "str= " << *pstr << endl;
```

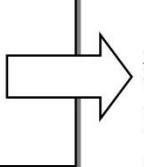
```
84 class A
85 {
86 public:
87     int id;
88     A(int i) : id(i) { cout << "ctor. this=" << this << " id=" << id << endl; }
89     ~A()           { cout << "dtor. this=" << this << endl; }
90 };
```

```
107     A* pA = new A(1);          //ctor. this=000307A8 id=1
108     cout << pA->id << endl;    //!
109 //! pA->A::A(3);            //in VC6 : ctor. this=000307A8 id=3
110 //! [Error] cannot call constructor 'jj02::A::A' directly
111
112 //! A::A(5);                //in VC6 : ctor. this=0013FF60 id=5
113 //!                         //dtor. this=0013FF60
114 //! [Error] cannot call constructor 'jj02::A::A' directly
115 //!                         // [Note] for a function-style cast, remove the redundant '::A'
116
117     cout << pA->id << endl;    //in VC6 : 3
118 //!                         //in GCC : 1
119
120     delete pA;               //dtor. this=000307A8
```



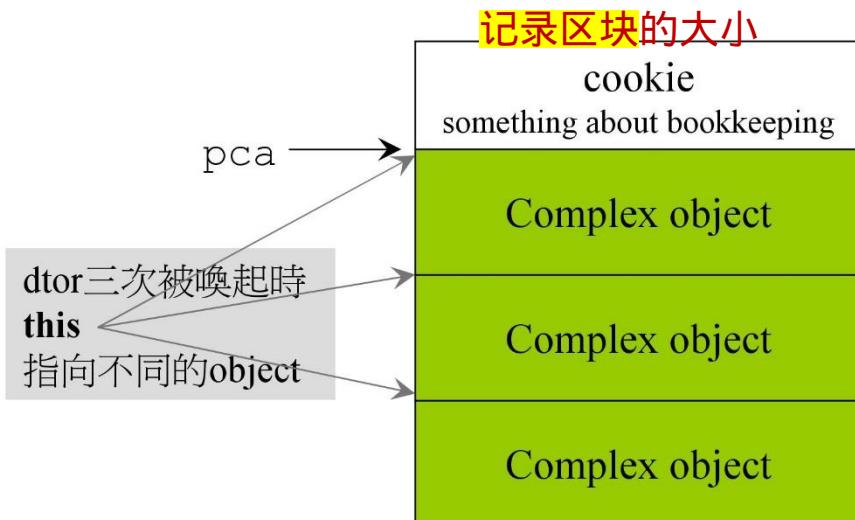
array new, array delete

```
Complex* pca = new Complex[3];  
//喚起三次ctor。  
//無法藉由參數給予初值  
...  
delete[] pca; //喚起3次dtor
```

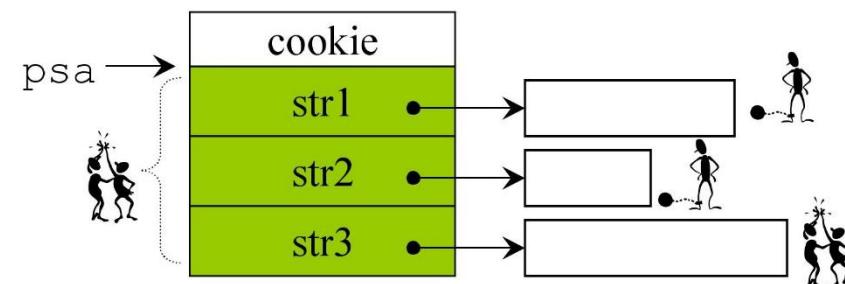


沒對每個 object 調用 dtor，有什麼影響？

- 對 class without ptr member 可能沒影響。
- 對 class with pointer member 通常有影響。



```
string* psa = new string[3];  
...  
delete psa; //喚起1次dtor
```





array new, array delete

```
class A
{
public:
    int id;

    A() : id(0) { cout << "default ctor. this=" << this << " id=" << id << endl; }
    A(int i) : id(i) { cout << "ctor. this=" << this << " id=" << id << endl; }
    ~A() { cout << "dtor. this=" << this << " id=" << id << endl; }
};
```

```
default ctor. this=0x3e398c id=0
default ctor. this=0x3e3990 id=0
default ctor. this=0x3e3994 id=0
buf=0x3e398c tmp=0x3e398c
ctor. this=0x3e398c id=0
ctor. this=0x3e3990 id=1
ctor. this=0x3e3994 id=2
buf=0x3e398c tmp=0x3e3998
dtor. this=0x3e3994 id=2
dtor. this=0x3e3990 id=1
dtor. this=0x3e398c id=0
```

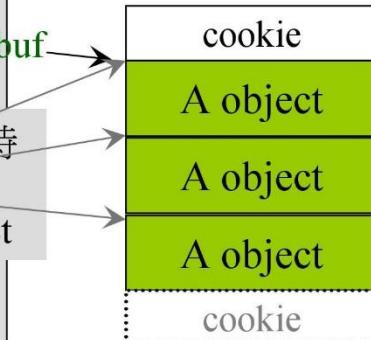
```
A* buf = new A[size]; //default ctor 3 次. [0]先於[1]先於[2])
//A必須有 default ctor,
//否則 [Error] no matching function for call to 'A::A()'
A* tmp = buf;

cout << "buf=" << buf << " tmp=" << tmp << endl;

for(int i = 0; i < size; ++i)
    new (tmp++) A(i); //ctor 3次
    placement new 賦值
cout << "buf=" << buf << " tmp=" << tmp << endl;

delete [] buf; //dtor 3 次 (次序逆反, [2]先於[1]先於[0])
```

dtor 三次被喚起時
this 指向不同的object

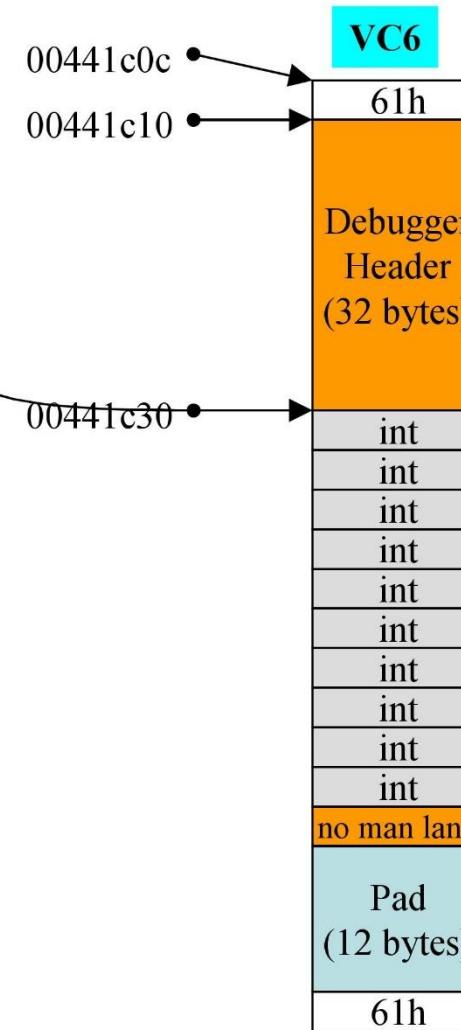




array size, in memory block

```
int* pi = new int[10]; //sizeof(pi) : 4  
delete pi;
```

```
int ia[10]; //from stack but not heap  
cout << sizeof(ia); //40
```

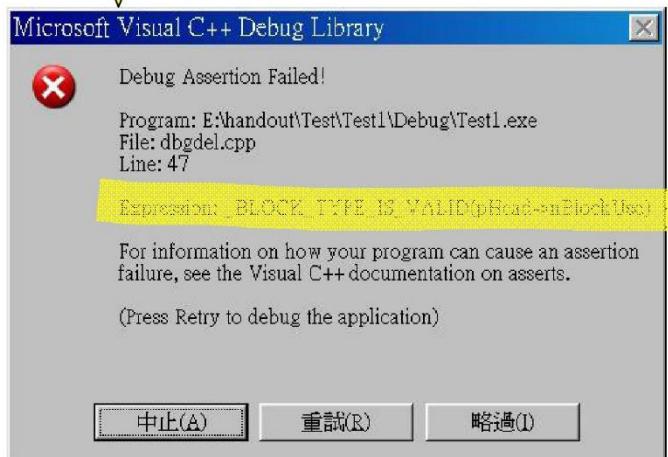




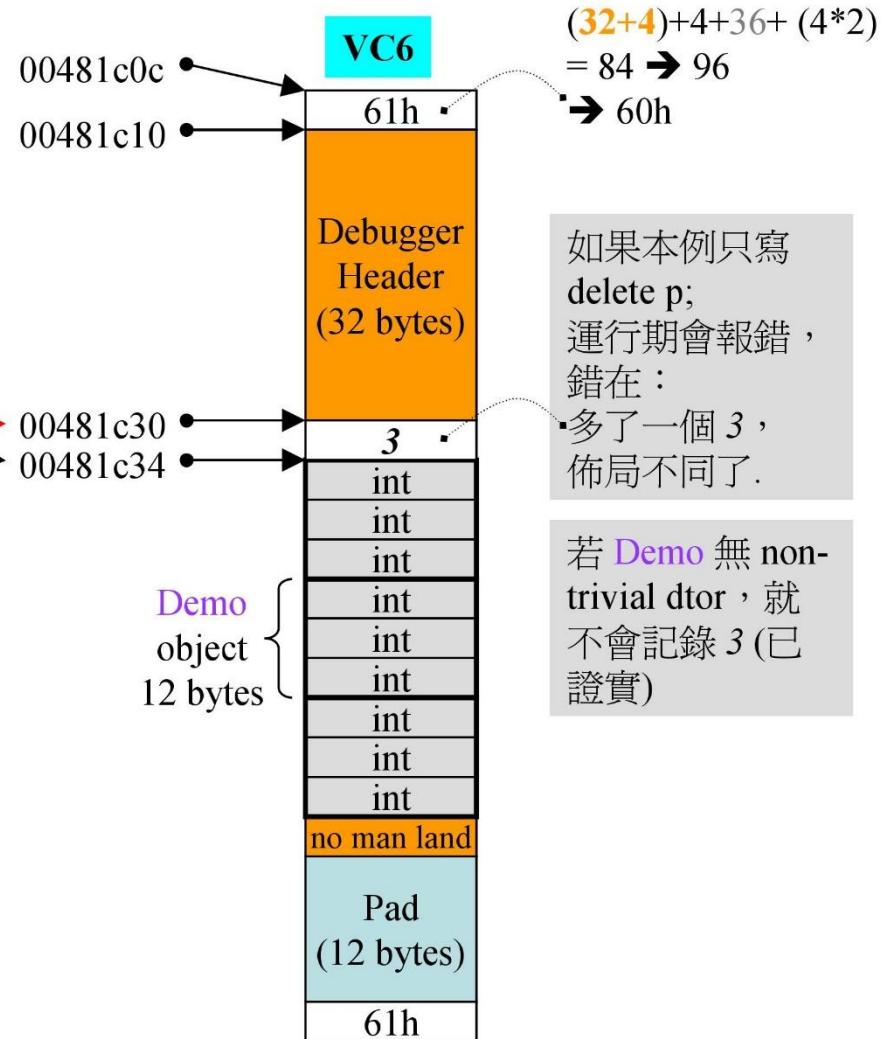
array size, in memory block

```
Demo* p = new Demo[3];
delete [] p;
```

若未寫 [],
debug mode 下出現
Assertion Failed.



```
Demo d[3]; //from stack but not heap
cout << sizeof(d); //36
```



如果本例只寫
delete p;
運行期會報錯，
錯在：
多了一個 3，
佈局不同了。

若 Demo 無 non-trivial dtor，就
不會記錄 3 (已證實)



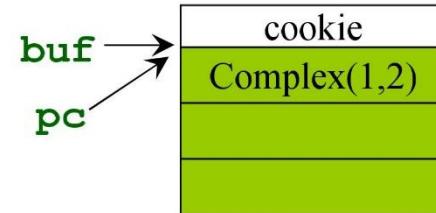
placement new

等同调用构造函数

- placement new 允許我們將 object 建構於 allocated memory 中。
- 沒有所謂 placement delete，因為 placement new 根本沒分配 memory.
亦或稱呼與 placement new 對應的 operator delete 為 placement delete.

注意，關於“placement new”，
或指 `new(p)`，
或指 `::operator new(size_t, void*)`

```
#include <new>
char* buf = new char[sizeof(Complex)*3];
Complex* pc = new(buf) Complex(1,2);
...
delete [] buf; //
```



編譯器轉為

```
Complex *pc;
try {
    1 void* mem = operator new(sizeof(Complex), buf); //allocate
    2 pc = static_cast<Complex*>(mem); //cast
    3 pc->Complex::Complex(1,2); //construct
}
catch( std::bad_alloc ) {
    //若allocation失敗就不執行constructor
}
```

```
void* operator new(size_t, void* loc)
{ return loc; }
```



C++應用程序, 分配內存的途徑

應用程序

```
Foo* p = new Foo(x);  
...  
delete p;
```

expression
(不可改變)
(不可重載)

```
Foo* p = (Foo*)operator new(sizeof(Foo));  
new (p) Foo(x);  
...  
p->~Foo();  
operator delete (p);
```

so, 我也可以模仿 new expression :

```
Foo* p = (Foo*)malloc(sizeof(Foo));  
new (p) Foo(x); //invoke ctor  
...  
p->~Foo(); //invoke dtor  
free (p);
```

member functions

Foo::operator new(size_t);
Foo::operator delete(void*);

1

可重載



我們有上↑下↓
兩個機會(兩個
地點)可以改變
內存分配機制

global functions

::operator new(size_t);
::operator delete(void*);

2

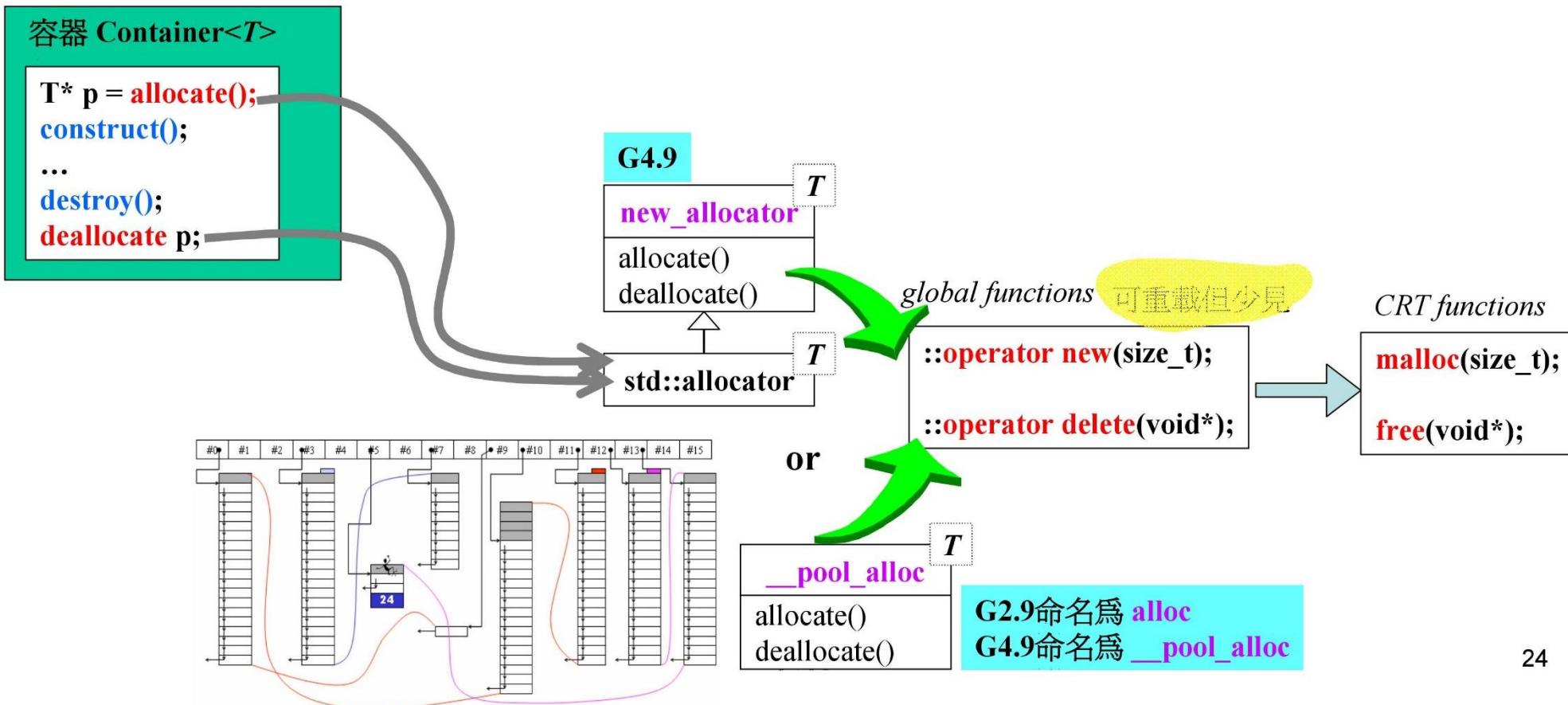
CRT functions

malloc(size_t);
free(void*);

可重載但少見



C++容器, 分配內存的途徑





重載 ::operator new / ::operator delete

小心，這影響無遠弗屆

```
void* myAlloc(size_t size)
{ return malloc(size); }
```

```
void myFree(void* ptr)
{ return free(ptr); }
```

///它們不可以被聲明於一個 namespace 內

```
inline void* operator new(size_t size)
```

```
{ cout << "jjhou global new() \n"; return myAlloc( size ); }
```

```
inline void* operator new[](size_t size)
```

```
{ cout << "jjhou global new[]() \n"; return myAlloc( size ); }
```

```
inline void operator delete(void* ptr)
```

```
{ cout << "jjhou global delete() \n"; myFree( ptr ); }
```

```
inline void operator delete[](void* ptr)
```

```
{ cout << "jjhou global delete[]() \n"; myFree( ptr ); }
```



—侯捷—

...\\vc98\\crt\\src\\newop2.cpp

```
void* operator new(size_t size, const std::nothrow_t&)
    _THROW0()
```

```
{ // try to allocate size bytes
```

```
void *p;
```

```
while ((p = malloc(size)) == 0)
```

```
{ // buy more memory or return null pointer
```

```
_TRY_BEGIN
```

```
if (_callnewh(size) == 0) break;
```

```
_CATCH(std::bad_alloc) return (0);
```

```
_CATCH_END
```

```
}
```

```
return (p);
```

```
}
```

...\\vc98\\crt\\src\\delop.cpp

```
void __cdecl operator delete(void *p) _THROW0()
```

```
{ // free an allocated object
```

```
free(p);
```

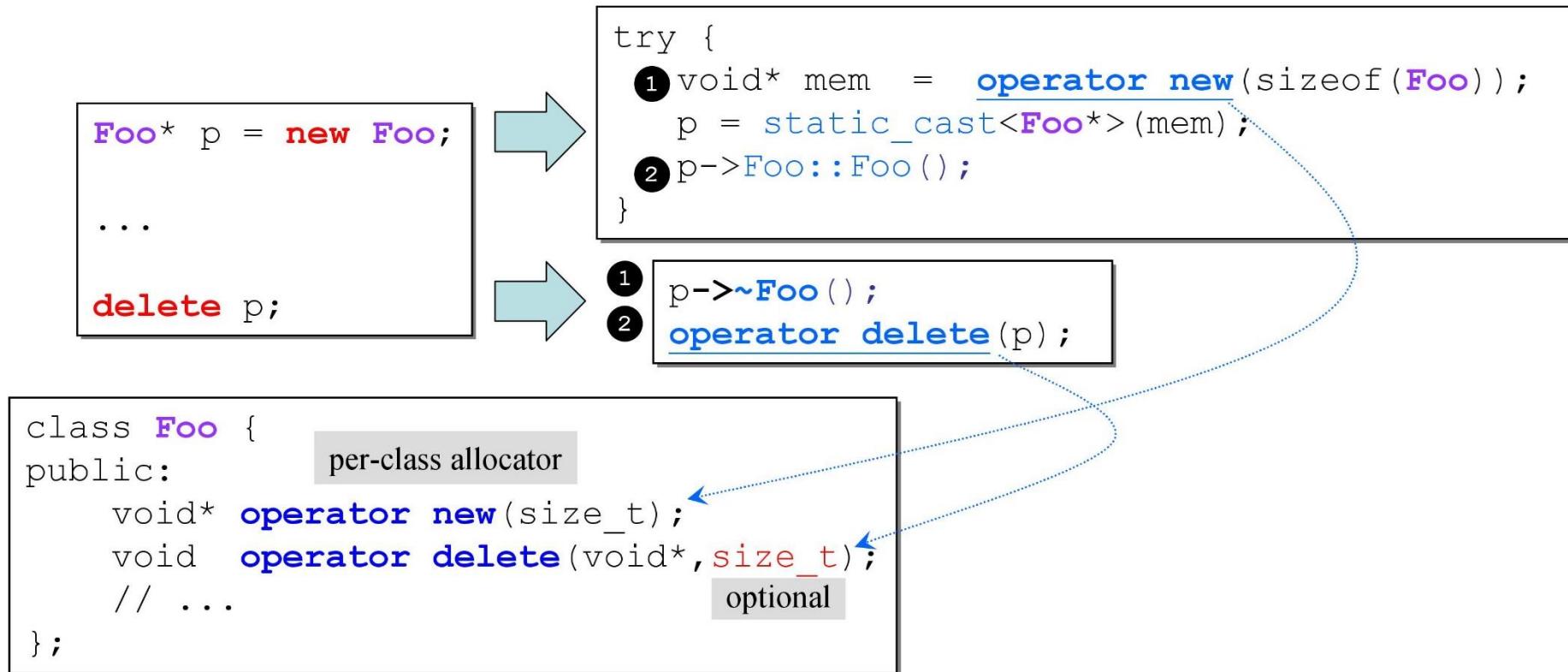
```
}
```





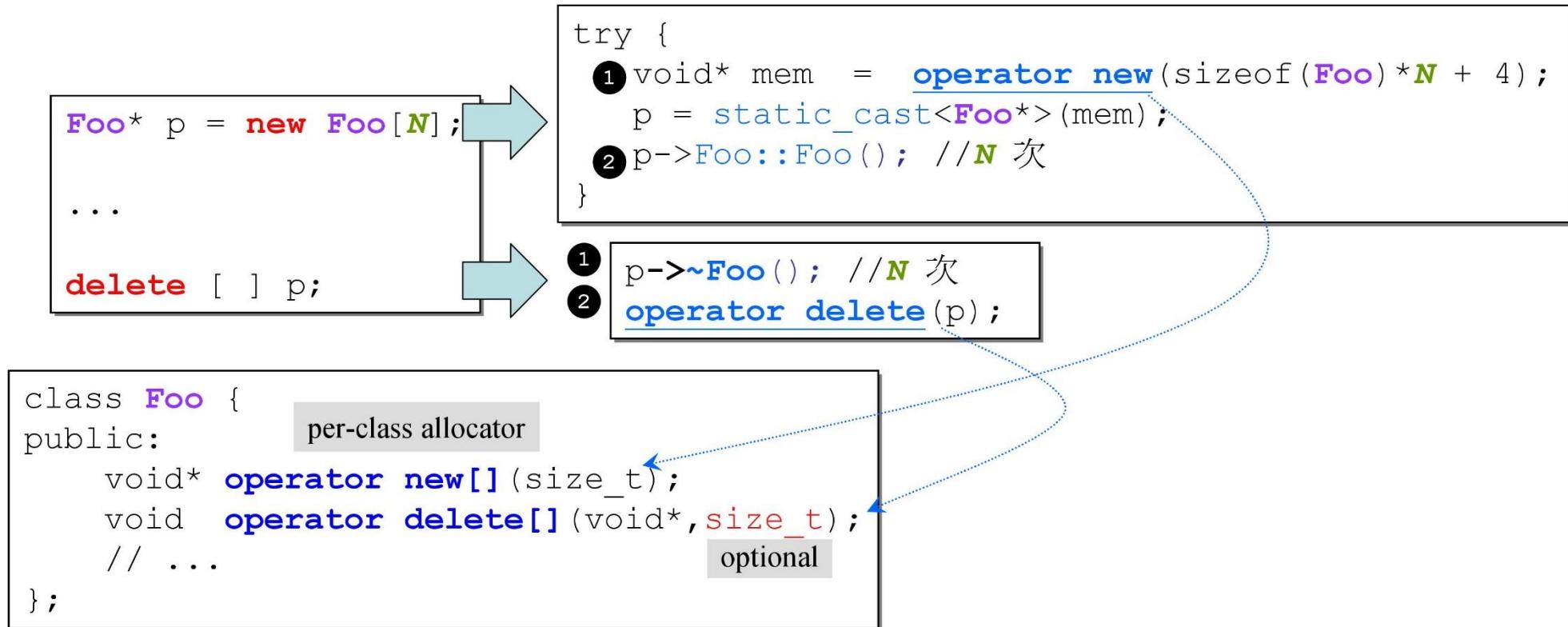
重载 operator new/operator delete

类内重载



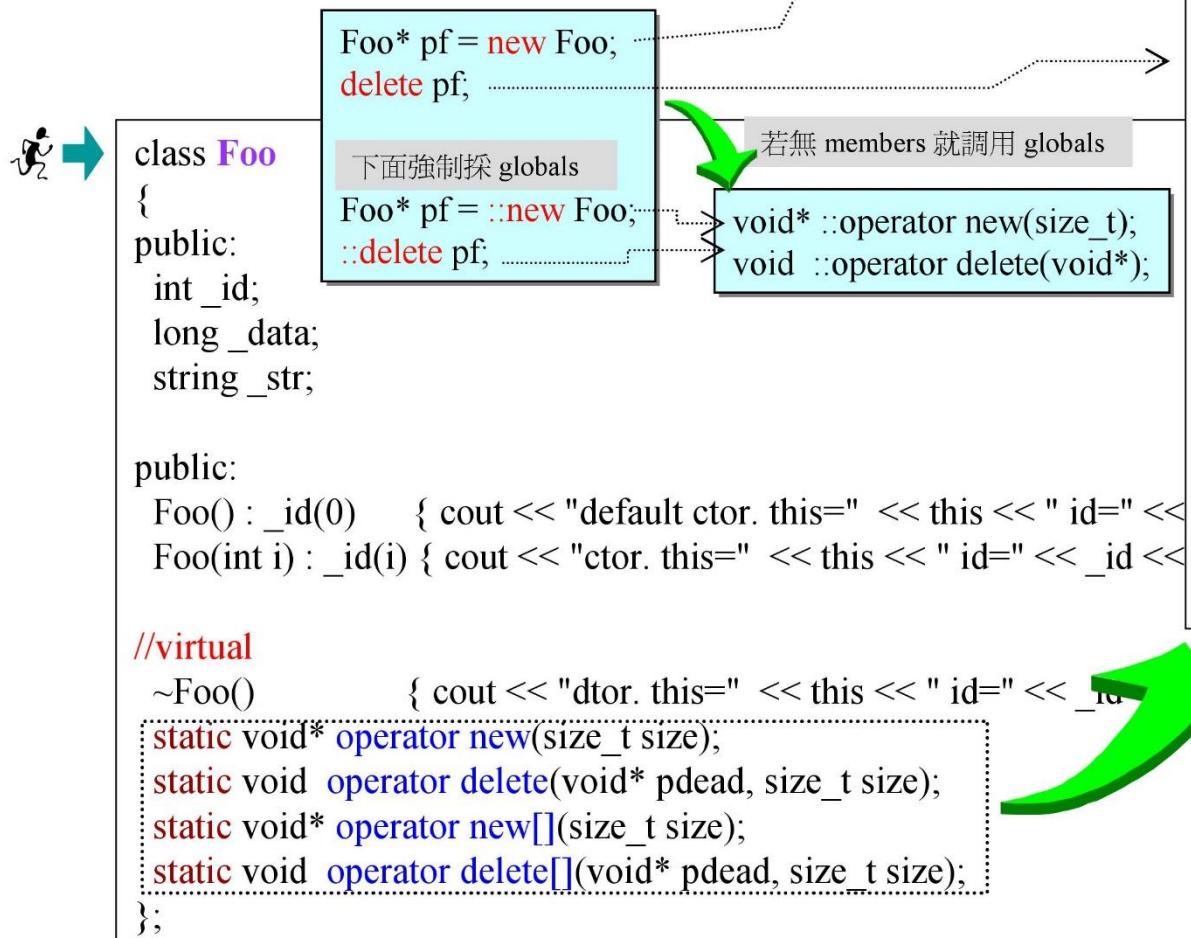


重載 operator new[]/ operator delete[]





示例, 接口



```
void* Foo::operator new(size_t size) {
    Foo* p = (Foo*)malloc(size);
    cout << .....
    return p;
}
void Foo::operator delete(void* pdead, size_t size) {
    cout << .....
    free(pdead);
}
void* Foo::operator new[](size_t size) {
    Foo* p = (Foo*)malloc(size);
    cout << .....
    return p;
}
void Foo::operator delete[](void* pdead, size_t size) {
    cout << .....
    free(pdead);
}
```



示例

Foo without
virtual dtor



```
cout << "sizeof(Foo)= " << sizeof(Foo);
```

```
1 Foo* p = new Foo(7);  
2 delete p;
```

```
3 Foo* pArray = new Foo[5];  
4 delete [] pArray;
```

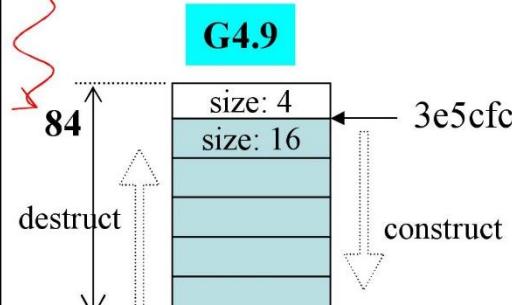
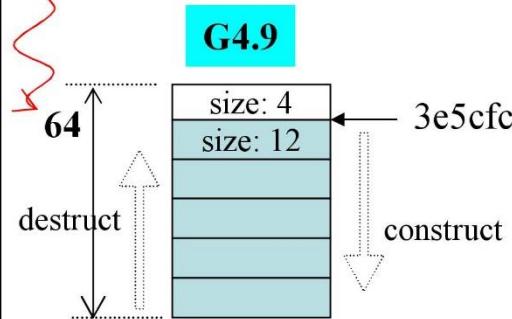
Foo with
virtual dtor



```
...._overload_array_new.....  
sizeof(Foo)= 12  
Foo::operator new(), size=12      return: 0x3e3988  
ctor. this=0x3e3988 id=7  
dtor. this=0x3e3988 id=7  
Foo::operator delete(), pdead= 0x3e3988 size= 12  
Foo::operator new[], size=64     return: 0x3e5cf8  
default ctor. this=0x3e5cf8 id=0  
default ctor. this=0x3e5d08 id=0  
default ctor. this=0x3e5d14 id=0  
default ctor. this=0x3e5d20 id=0  
default ctor. this=0x3e5d2c id=0  
dtor. this=0x3e5d2c id=0  
dtor. this=0x3e5d20 id=0  
dtor. this=0x3e5d14 id=0  
dtor. this=0x3e5d08 id=0  
dtor. this=0x3e5cf8 id=0  
Foo::operator delete[], pdead= 0x3e5cf8 size= 64
```



```
...._overload_array_new.....  
sizeof(Foo)= 16  
Foo::operator new(), size=16      return: 0x3e3988  
ctor. this=0x3e3988 id=7  
dtor. this=0x3e3988 id=7  
Foo::operator delete(), pdead= 0x3e3988 size= 16  
Foo::operator new[], size=84     return: 0x3e5cf8  
default ctor. this=0x3e5cf8 id=0  
default ctor. this=0x3e5d0c id=0  
default ctor. this=0x3e5d1c id=0  
default ctor. this=0x3e5d2c id=0  
default ctor. this=0x3e5d3c id=0  
dtor. this=0x3e5d3c id=0  
dtor. this=0x3e5d2c id=0  
dtor. this=0x3e5d1c id=0  
dtor. this=0x3e5d0c id=0  
dtor. this=0x3e5cf8 id=0  
Foo::operator delete[], pdead= 0x3e5cf8 size= 84
```





示例



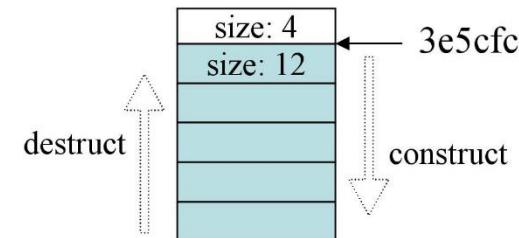
- ① Foo* p = ::new Foo(7);
- ② ::delete p;
- ③ Foo* pArray = ::new Foo[5];
- ④ ::delete [] pArray;

這樣調用 (也就是寫上
global scope operator ::) ,
會繞過前述所有
overloaded functions,
強迫使用 global version.

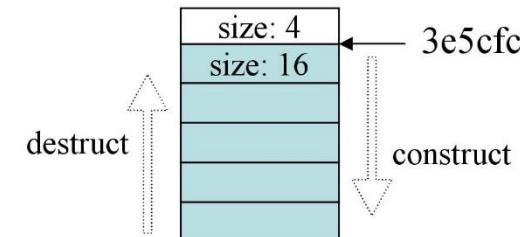
```
ctor. this=0x3e5ce0 id=7
dtor. this=0x3e5ce0 id=7
default ctor. this=0x3e5cf0 id=0
default ctor. this=0x3e5d08 id=0
default ctor. this=0x3e5d14 id=0
default ctor. this=0x3e5d20 id=0
default ctor. this=0x3e5d2c id=0
dtor. this=0x3e5d2c id=0
dtor. this=0x3e5d20 id=0
dtor. this=0x3e5d14 id=0
dtor. this=0x3e5d08 id=0
dtor. this=0x3e5cf0 id=0
```

↑ 都沒有進入我重載的
operator new(), operator delete(),
operator new[](), operator delete[]().

G4.9



G4.9





重載 new() / delete()

在已分配的特定內存创建对象
是C++实现的operator new版本

我們可以重載 class member `operator new()`，寫出多個版本，前提是每一版本的聲明都必須有獨特的參數列，其中第一參數必須是 `size_t`，其餘參數以 `new` 所指定的 `placement arguments` 為初值。出現於 `new (.....)` 小括號內的便是所謂 `placement arguments`。

`Foo* pf = new (300, 'c') Foo;`

或稱此為
placement operator delete.

我們也可以重載 class member `operator delete()`，寫出多個版本。但它們絕不會被 `delete` 調用。只有當 `new` 所調用的 ctor 拋出 exception，才會調用這些重載版的 `operator delete()`。它只可能這樣被調用，主要用來歸還未能完全創建成功的 object 所佔用的 memory。

示例

```
class Foo {  
public:  
    Foo() { cout << "Foo::Foo()" << endl; }  
    Foo(int) { cout << "Foo::Foo(int)" << endl; throw Bad(); }  
  
    // (1) 這個就是一般的 operator new() 的重載  
    void* operator new(size_t size) {  
        return malloc(size);  
    }  
    // (2) 這個就是標準庫已提供的 placement new() 的重載 (的形式)  
    // (所以我也模擬 standard placement new, 就只是傳回 pointer)  
    void* operator new(size_t size, void* start) {  
        return start;  
    }  
    // (3) 這個才是嶄新的 placement new  
    void* operator new(size_t size, long extra) {  
        return malloc(size+extra);  
    }  
    // (4) 這又是一個 placement new  
    void* operator new(size_t size, long extra, char init) {  
        return malloc(size+extra);  
    }  
.....(接下頁)
```

class Bad { };

故意在這兒拋出 exception ,
測試 placement operator delete.

//(5) 這又是一個 placement new, 但故意寫錯第一參數的 type
// (那必須是 size_t 以符合正常的 operator new)
//! void* operator new(long extra, char init) {
// [Error] 'operator new' takes type 'size_t' ('unsigned int')
// as first parameter [-fpermissive]
//! return malloc(extra);
//! }

示例 (續)

..... (續上頁)

//以下是搭配上述 placement new 的各個所謂 placement delete.

//當 ctor 發出異常，這兒對應的 operator (placement) delete 就會被調用.

//其用途是釋放對應之 placement new 分配所得
//(1) 這個就是一般的 operator delete() 的重載

```
void operator delete(void*,size_t)
{ cout << "operator delete(void*,size_t)" << endl;
```

//(2) 這是對應上頁的 (2)

```
void operator delete(void*,void*)  
{ cout << "operator delete(void*,vo
```

//(3) 這是對應上頁的 (3)

```
void operator delete(void*,long)
```

//(4) 這是對應上頁的 (4)

```
void operator delete(void*,long,char)
```

```
private:  
    int m_i;  
};
```

即使 operator delete(...) 未能一一對應於 operator new(...), 也不會出現任何報錯。你的意思是：放棄處理 ctor 發出的異常。

Foo start;

① Foo* p1 = new Foo;

② Foo* p2 = new (&start) Foo;

③ Foo* p3 = new (100) Foo;

④ Foo* p4 = new (100,'a') Foo;

⑤ Foo* p5 = new (100) Foo(1);

```
Foo* p6 = new (100,'a') Fo
```

Foo* p7 = new (&start)

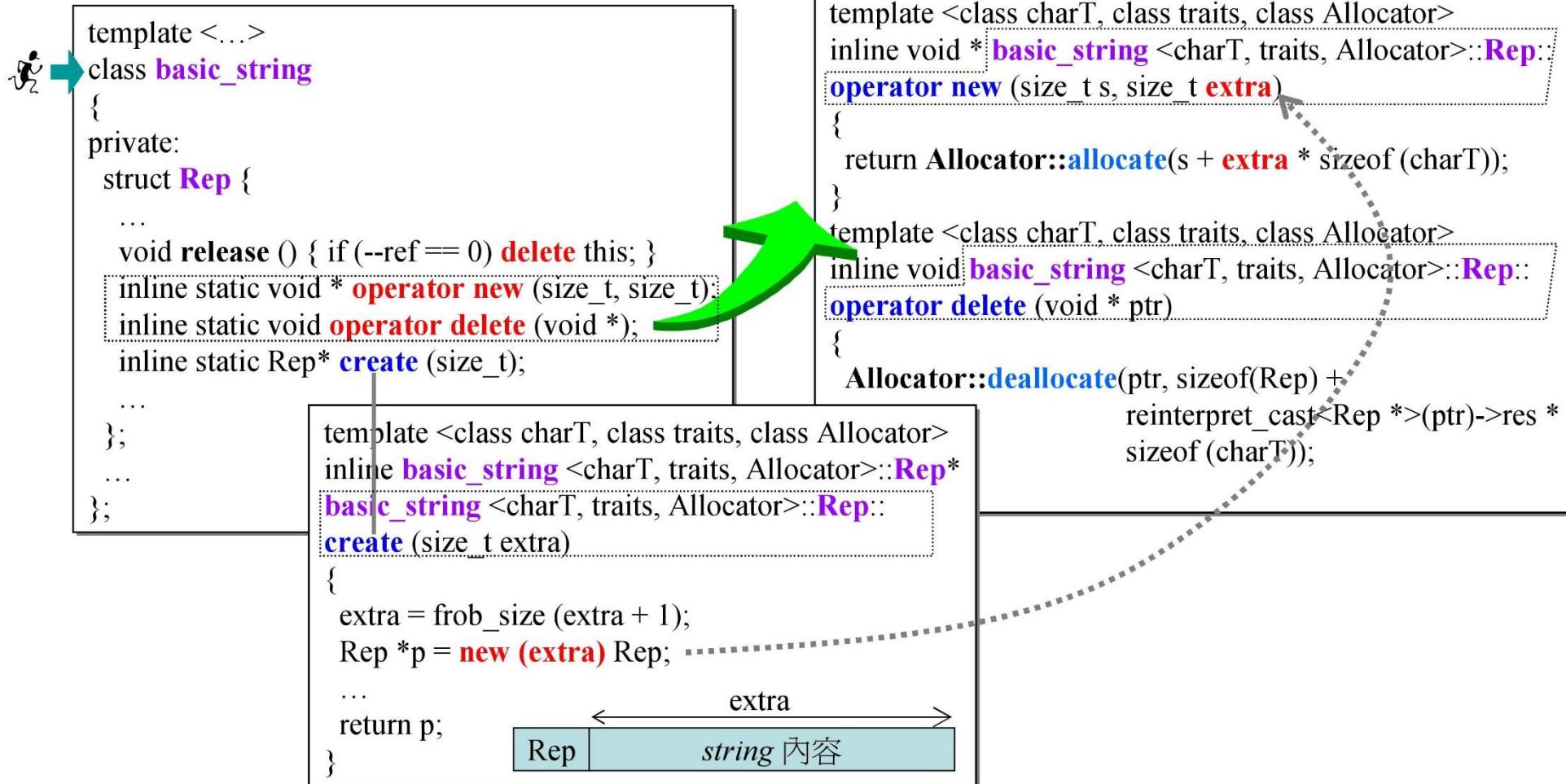
```
Foo::Foo()
① operator new(size_t size), size= 4
Foo::Foo()
② operator new(size_t size, void* start), size= 4  start= 0x22fe8c
Foo::Foo()
③ operator new(size_t size, long extra)  4 100
Foo::Foo()
④ operator new(size_t size, long extra, char init)  4 100 a
Foo::Foo()
⑤ operator new(size_t size, long extra)  4 100
Foo::Foo<int>
terminate called after throwing an instance of 'jj07::Bad'
```

ctor 抛出異常

► 奇怪, G4.9 沒調用 operator delete (void*,long), 但 G2.9 確實有.

VC6 warning C4291: 'void * __cdecl Foo::operator new(~)' no matching operator delete found; memory will not be freed if initialization throws an exception

basic_string 使用 new(extra) 擴充申請量



per-class allocator, 1

ref. C++Primer 3/e, p.765

减少malloc调用
先分配一大块内存再切

```
#include <cstddef>
#include <iostream>
using namespace std;

class Screen {
public:
    Screen(int x) : i(x) { };
    int get() { return i; }

    void* operator new(size_t);
    void operator delete(void*, size_t);
    // ...
    // 這種設計會引發多耗用一個
    // next 的疑慮。下頁設計更好
private:
    Screen* next;
    static Screen* freeStore;
    static const int screenChunk;
};

Screen* Screen::freeStore = 0;
const int Screen::screenChunk = 24;
```

但多了个指针

```
void* Screen::operator new(size_t size)
{
    Screen *p;
    if (!freeStore) {
        //linked list 是空的，所以申請一大塊
        size_t chunk = screenChunk * size;
        freeStore = p =
            reinterpret_cast<Screen*>(new char[chunk]);
        //將一大塊分割片片，當做 linked list 串接起來
        for (; p != &freeStore[screenChunk-1]; ++p)
            p->next = p+1;      链表
        p->next = 0;
    }
    p = freeStore;
    freeStore = freeStore->next;
    return p;
}
```

沒有對應的 delete，
但這不算 memory leak.

```
void Screen::operator delete(void *p, size_t)
{
    //將 deleted object 插回 free list 前端
    static_cast<Screen*>(p)->next = freeStore;
    freeStore = static_cast<Screen*>(p);
}
```



per-class allocator, 1

ref. C++Primer 3/e, p.765



```
cout << sizeof(Screen) << endl; //8  
  
size_t const N = 100;  
Screen* p[N];  
  
for (int i=0; i< N; ++i)  
    p[i] = new Screen(i);  
  
//輸出前 10 個 pointers, 比較其間隔  
for (int i=0; i< 10; ++i)  
    cout << p[i] << endl;  
  
for (int i=0; i< N; ++i)  
    delete p[i];
```

左邊是寫了 member operator new/delete 的結果，
右邊是沒寫因而使用 global operator new/delete 的結果

8
0x3e48f0
0x3e48f8
0x3e4900
0x3e4908
0x3e4910
0x3e4918
0x3e4920
0x3e4928
0x3e4930
0x3e4938

間隔8

8
0x3e48e0
0x3e48f0
0x3e4900
0x3e4910
0x3e4920
0x3e4930
0x3e4940
0x3e4950
0x3e4960
0x3e4970

間隔16

per-class allocator, 2

ref. Effective C++ 2e, item10

```
class Airplane {  
private:  
    struct AirplaneRep {  
        unsigned long miles;  
        char type;  
    };  
private:  
    union { AirplaneRep rep; //此欄針對使用中的objects  
            Airplane* next; //此欄針對 free list 上的object  
    };  
public:  
    unsigned long getMiles() { return rep.miles; }  
    char getType() { return rep.type; }  
    void set(unsigned long m, char t) {  
        rep.miles = m; rep.type = t;  
    }  
public:  
    static void* operator new(size_t size);  
    static void operator delete(void* deadObject, size_t size);  
private:  
    static const int BLOCK_SIZE;  
    static Airplane* headOfFreeList;  
};  
  
Airplane* Airplane::headOfFreeList;  
const int Airplane::BLOCK_SIZE = 512;
```

```
void* Airplane::operator new(size_t size)  
{  
    怎會有誤？當繼承發生時  
    //如果大小有誤，轉交給 ::operator new()  
    if (size != sizeof(Airplane))  
        return ::operator new(size);  
  
    Airplane* p = headOfFreeList;  
    if (p) //如果 p 有效，就把 list 頭部下移一個元素  
        headOfFreeList = p->next;  
    else {  
        //free list 已空，申請(分配)一大塊  
        Airplane* newBlock = static_cast<Airplane*>  
            (::operator new(BLOCK_SIZE * sizeof(Airplane)));  
  
        //將小塊串成一個 free list，  
        //但跳過 #0，因它將被傳回做為本次成果  
        for (int i = 1; i < BLOCK_SIZE-1; ++i)  
            newBlock[i].next = &newBlock[i+1];  
        newBlock[BLOCK_SIZE-1].next = 0; //結束 list  
        p = newBlock;  
        headOfFreeList = &newBlock[1];  
    }  
    return p;  
}
```



7



per-class allocator, 2

ref. Effective C++ 2e, item10

```
// operator delete 接獲一個內存塊  
// 如果大小正確，就把它加到 free list 前端  
void Airplane::operator delete(void* deadObject,  
                                size_t size)  
{  
    if (deadObject == 0) return;  
    if (size != sizeof(Airplane)) {  
        ::operator delete(deadObject);  
        return;  
    }  
  
    Airplane* carcass =  
        static_cast<Airplane*>(deadObject);  
  
    carcass->next = headOfFreeList;  
    headOfFreeList = carcass;  
}
```

未还内存 但未泄漏 还在手上

```
8  
0x3e4ce0 A 1000  
0x3e4d00 B 2000  
0x3e4d20 C 5000000  
0x3e4cd8  
0x3e4ce0  
0x3e4ce8  
0x3e4cf0  
0x3e4cf8 間隔8  
0x3e4d00  
0x3e4d08  
0x3e4d10  
0x3e4d18  
0x3e4d20
```



```
cout << sizeof(Airplane) << endl;  
size_t const N = 100;  
Airplane* p[N];  
  
for (int i=0; i< N; ++i)  
    p[i] = new Airplane;
```

//隨機測試 object 正常否

```
p[1]->set(1000,'A');  
p[5]->set(2000,'B');  
p[9]->set(500000,'C');  
...
```

//輸出前 10 個 pointers,

//用以比較其間隔
for (int i=0; i< 10; ++i)
 cout << p[i] << endl;

for (int i=0; i< N; ++i)
 delete p[i];

```
8  
0x3e4900 A 1000  
0x3e4930 B 2000  
0x3e4970 C 5000000  
0x3e4910  
0x3e4900  
0x3e48f0  
0x3e48e0  
0x3e4920 間隔16  
0x3e4930  
0x3e4940  
0x3e4950  
0x3e4960  
0x3e4970
```

左邊是寫了 member operator new/delete 的結果，
右邊是沒寫因而使用 global operator new/delete 的結果



static allocator

當你受困於必須為不同的 classes 重寫一遍幾乎相同的 member operator new 和 member operator delete 時，應該有方法將一個總是分配特定尺寸之區塊的 memory allocator 概念包裝起來，使它容易被重複使用。以下展示一種作法，每個 allocator object 都是個分配器，它體內維護一個 free-lists；不同的 allocator objects 維護不同的 free-lists。

```
class allocator
{
private:
    struct obj {
        struct obj* next; //embedded pointer
    };
public:
    void* allocate(size_t);
    void deallocate(void*, size_t);
private:
    obj* freeStore = nullptr;
    const int CHUNK = 5; //小一些以便觀察
};
```

```
void
allocator::deallocate(void* p, size_t)
{
    //將 *p 收回插入 free list 前端
    ((obj*)p)->next = freeStore;
    freeStore = (obj*)p;
}
```

```
void* allocator::allocate(size_t size)
{
    obj* p;
    if (!freeStore) {
        //linked list 為空，於是申請一大塊
        size_t chunk = CHUNK * size;
        freeStore = p = (obj*)malloc(chunk);
        //將分配得來的一大塊當做 linked list 般,
        //小塊小塊串接起來
        for (int i=0; i < (CHUNK-1); ++i) {
            p->next = (obj*)((char*)p + size);
            p = p->next;
        }
        p->next = nullptr; //last
    }
    p = freeStore;
    freeStore = freeStore->next;
    return p;
}
```



static allocator

专属于class

寫法
十分
制式

```
class Foo {  
public:  
    long L;  
    string str;          指针指着链表  
    static allocator myAlloc;  
public:  
    Foo(long l) : L(l) { }  
    static void* operator new(size_t size)  
    { return myAlloc.allocate(size); }  
    static void operator delete(void* pdead, size_t size)  
    { return myAlloc.deallocate(pdead, size); }  
};  
allocator Foo::myAlloc;
```

```
class Goo {  
public:  
    complex<double> c;  
    string str;  
    static allocator myAlloc;  
public:  
    Goo(const complex<double>& x) : c(x) { }  
    static void* operator new(size_t size)  
    { return myAlloc.allocate(size); }  
    static void operator delete(void* pdead, size_t size)  
    { return myAlloc.deallocate(pdead, size); }  
};  
allocator Goo::myAlloc;
```

這比先前的設計乾淨多了，application classes 不再與內存分配細節糾纏不清，所有相關細節都讓 allocator 去操心，我們的工作是讓 application classes 正確運作。



static allocator, 示例與結果

```
Foo* p[100];

cout << "sizeof(Foo)= " << sizeof(Foo) << endl;
for (int i=0; i<23; ++i) { //隨意看看結果
    p[i] = new Foo(i);
    cout << p[i] << ' ' << p[i]->L << endl;
}

for (int i=0; i<23; ++i) {
    delete p[i];
}
```

```
Goo* p[100];

cout << "sizeof(Goo)= " << sizeof(Goo) << endl;
for (int i=0; i<17; ++i) { //隨意看看結果
    p[i] = new Goo(complex<double>(i,i));
    cout << p[i] << ' ' << p[i]->c << endl;
}

for (int i=0; i<17; ++i) {
    delete p[i];
}
```

sizeof(Foo)= 8

0x3e5f90 0
0x3e5f98 1
0x3e5fa0 2
0x3e5fa8 3
0x3e5fb0 4
0x3e5d40 5
0x3e5d48 6
0x3e5d50 7
0x3e5d58 8
0x3e5d60 9
0x3e6080 10
0x3e6088 11
0x3e6090 12
0x3e6098 13
0x3e60a0 14
0x3e6128 15
0x3e6130 16
0x3e6138 17
0x3e6140 18
0x3e6148 19
0x3e61d0 20
0x3e61d8 21
0x3e61e0 22

每五个均相邻



macro for static allocator

```
class Foo {
public:
    long L;
    string str;
    static allocator myAlloc;
public:
    Foo(long l) : L(l) { }
    static void* operator new(size_t size)
    { return myAlloc.allocate(size); }
    static void operator delete(void* pdead, size_t size)
    { return myAlloc.deallocate(pdead, size); }
};
allocator Foo::myAlloc;
```

寫法
十分
制式

```
// DECLARE_POOL_ALLOC -- used in class definition
#define DECLARE_POOL_ALLOC()
public:
```

```
    void* operator new(size_t size) { return myAlloc.allocate(size); } \
    void operator delete(void* p) { myAlloc.deallocate(p, 0); } \
```

protected:

```
    static allocator myAlloc;
```

```
// IMPLEMENT_POOL_ALLOC -- used in class implementation file
#define IMPLEMENT_POOL_ALLOC(class_name)
allocator class_name::myAlloc;
```

```
class Foo {
    DECLARE_POOL_ALLOC()
public:
    long L;
    string str;
public:
    Foo(long l) : L(l) { }
};
IMPLEMENT_POOL_ALLOC(Foo)
```

```
class Goo {
    DECLARE_POOL_ALLOC()
public:
    complex<double> c;
    string str;
public:
    Goo(const complex<double>& x) : c(x) { }
};
IMPLEMENT_POOL_ALLOC(Goo)
```



macro for static allocator,示例與結果

```
Foo* pF[100];
Goo* pG[100];

cout << "sizeof(Foo)= " << sizeof(Foo) << endl;
cout << "sizeof(Goo)= " << sizeof(Goo) << endl;

for (int i=0; i<23; ++i) { //隨意看看結果
    pF[i] = new Foo(i);
    pG[i] = new Goo(complex<double>(i,i));

    cout << pF[i] << ' ' << pF[i]->L << '\t';
    cout << pG[i] << ' ' << pG[i]->c << '\n';
}

for (int i=0; i<23; ++i) {
    delete pF[i];
    delete pG[i];
}
```

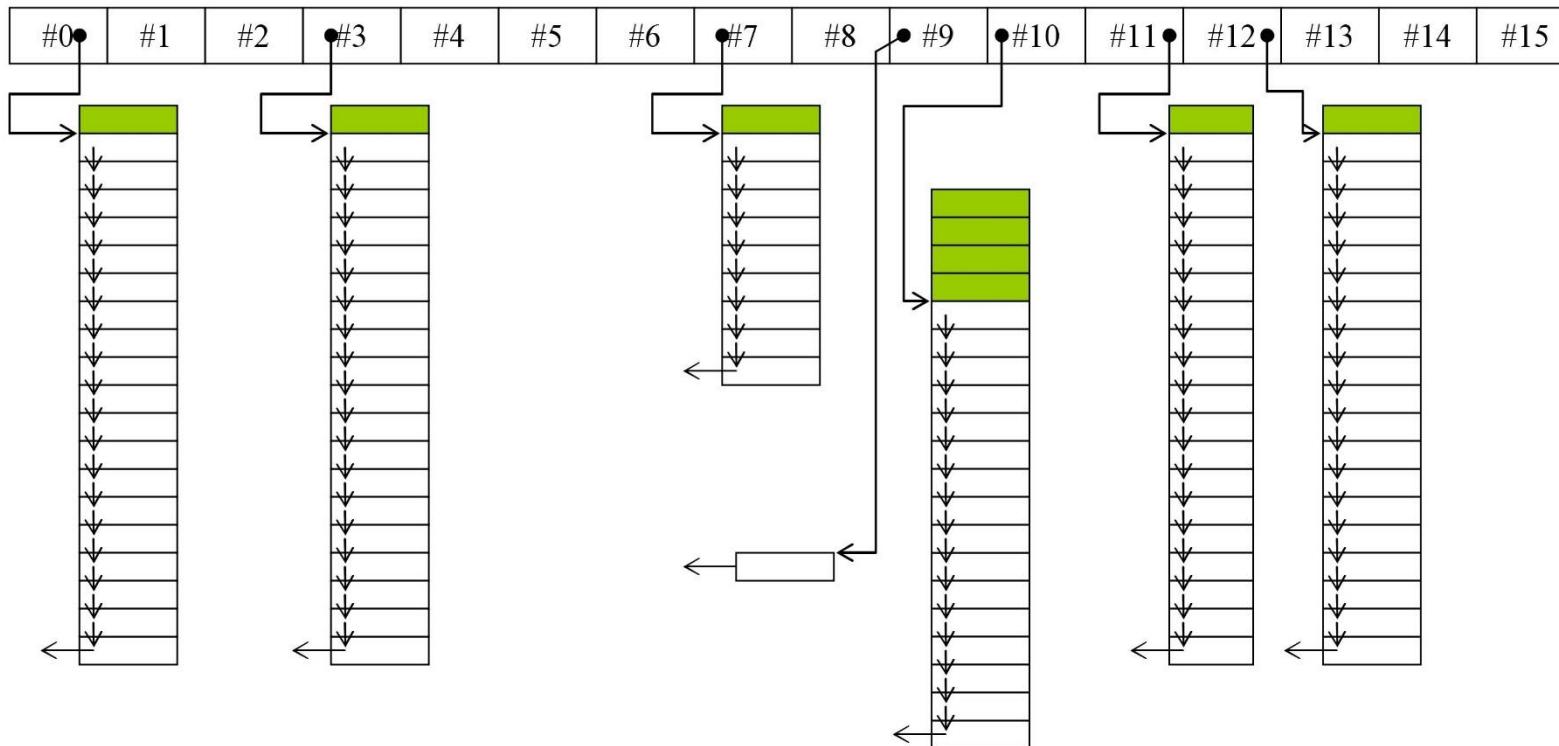


sizeof(Foo)= 8
sizeof(Goo)= 24
0x3e5fc0 0 0x3e6630 <0,0>
0x3e5fc8 1 0x3e6648 <1,1>
0x3e5fd0 2 0x3e6660 <2,2>
0x3e5fd8 3 0x3e6678 <3,3>
0x3e5fe0 4 0x3e6690 <4,4>
0x3e6158 5 0x3e66b0 <5,5>
0x3e6160 6 0x3e66c8 <6,6>
0x3e6168 7 0x3e66e0 <7,7>
0x3e6170 8 0x3e66f8 <8,8>
0x3e6178 9 0x3e6710 <9,9>
0x3e67d8 10 0x3e6808 <10,10>
0x3e67e0 11 0x3e6820 <11,11>
0x3e67e8 12 0x3e6838 <12,12>
0x3e67f0 13 0x3e6850 <13,13>
0x3e67f8 14 0x3e6868 <14,14>
0x3e6978 15 0x3e69a8 <15,15>
0x3e6980 16 0x3e69c0 <16,16>
0x3e6988 17 0x3e69d8 <17,17>
0x3e6990 18 0x3e69f0 <18,18>
0x3e6998 19 0x3e6a08 <19,19>
0x3e6b18 20 0x3e6b48 <20,20>
0x3e6b20 21 0x3e6b60 <21,21>
0x3e6b28 22 0x3e6b78 <22,22>



global allocator (with multiple free-lists)

將前述 allocator 進一步發展為具備 16 條 free-lists，並因此不再以 application classes 內的 static 呈現，而是一個 global allocator —— 這就是 G2.9 的 `std::alloc` 的雛形。



— 侯捷 —



new handler

當 operator new 沒能力為你分配出你所申請的 memory，會拋一個 std::bad_alloc exception。某些老舊編譯器則是返回 0—你仍然可以令編譯器那麼做：

```
new (nothrow) Foo;
```

此稱為 nothrow 形式。

拋出 exception 之前會先（不只一次）調用一個可由 client 指定的 handler，以下是 new handler 的形式和設定方法：

```
typedef void (*new_handler)();
new_handler set_new_handler(new_handler p) throw();
```

設計良好的 new handler 只有兩個選擇：

- 讓更多 memory 可用
- 調用 `abort()` 或 `exit()`

... \vc98\crt\src\newop2.cpp

```
void* operator new(size_t size, const std::nothrow_t&)
    _THROW0()
{
    // try to allocate size bytes
    void *p;
    while ((p = malloc(size)) == 0)
        { // buy more memory or return null pointer
            _TRY_BEGIN
                if (_callnewh(size) == 0) break;
            _CATCH(std::bad_alloc) return (0);
            _CATCH_END
        }
    return (p);
}
```



new handler

```
#include <new>
#include <iostream>
#include <cassert>
using namespace std;

void noMoreMemory()
{
    cerr << "out of memory";
    abort();
}

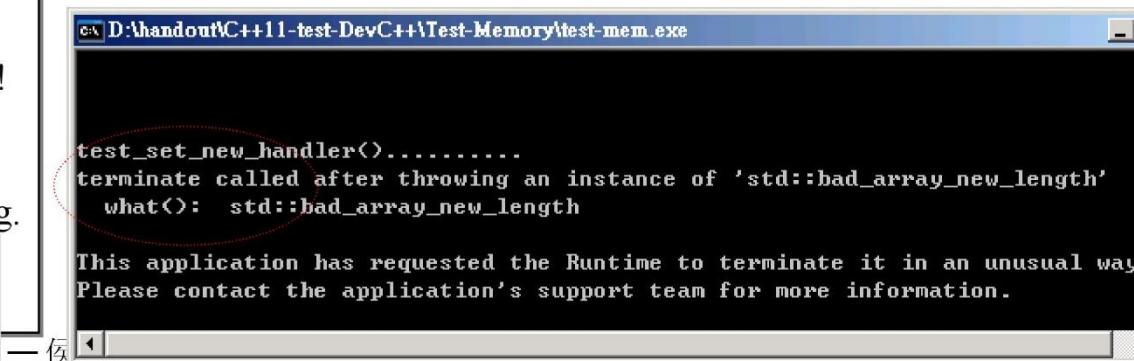
void main()
{
    set_new_handler(noMoreMemory);

    int* p = new int[1000000000000000]; //well, so BIG!
    assert(p);

    p = new int[1000000000000000]; //GCC warning.
    assert( // 執行結果：
           // BCB4 版會出現 out of memory.
           //           Abnormal program termination
           // 非常好，符合預期。);
}
```

msdev\vc98\crt\src\setnewh.cpp

```
new_handler
__cdecl set_new_handler(
    new_handler new_p)
{
    // cannot use stub to register a new handler
    assert(new_p == 0);
    // remove current handler
    _set_new_handler(0);
    return 0;
}
```



本例之 new handler 中若無調用 abort()，執行後 cerr 會不斷出現 “out of memory”，需強制中斷。這樣的表現是正確的，表示當 operator new 無法滿足申請量時，會不斷調用 new handler 直到獲得足夠 memory.



=default, =delete

```
class Foo {  
public:  
    Foo() = default;  
    Foo(const Foo&) = delete;  
    Foo& operator=(const Foo&) = delete;  
    ~Foo() = default;  
    ...  
};
```

it is not only for constructors and assignments,
but also applies to operator new/new[],operator
delete/delete[] and their overloads.

==== =default, =delete

```
class Foo {  
public:  
    long _x;  
public:  
    Foo(long x=0) : _x(x) { }  
};
```

✗ static void* operator new(size_t size) = **default**; [Error] cannot be defaulted
✗ static void operator delete(void* pdead, size_t size) = **default**;
 static void* operator new[](size_t size) = **delete**;
 static void operator delete[](void* pdead, size_t size) = **delete**;

```
class Goo {  
public:  
    long _x;  
public:  
    Goo(long x=0) : _x(x) { }  
  
    static void* operator new(size_t size) = delete;  
    static void operator delete(void* pdead, size_t size) = delete;  
};
```

[Error] use of deleted function ...

✗ Foo* p1 = new Foo(5);
 delete p1;
✗ Foo* pF = new Foo[10];
 delete [] pF;

✗ Goo* p2 = new Goo(7);
✗ delete p2;
Goo* pG = new Goo[10];
delete [] pG;

內存管理

從平地到萬丈高樓

Memory Management 101

第一講 primitives

第二講 std::allocator

第三講 malloc/free

第四講 loki::allocator

第五講 other issues

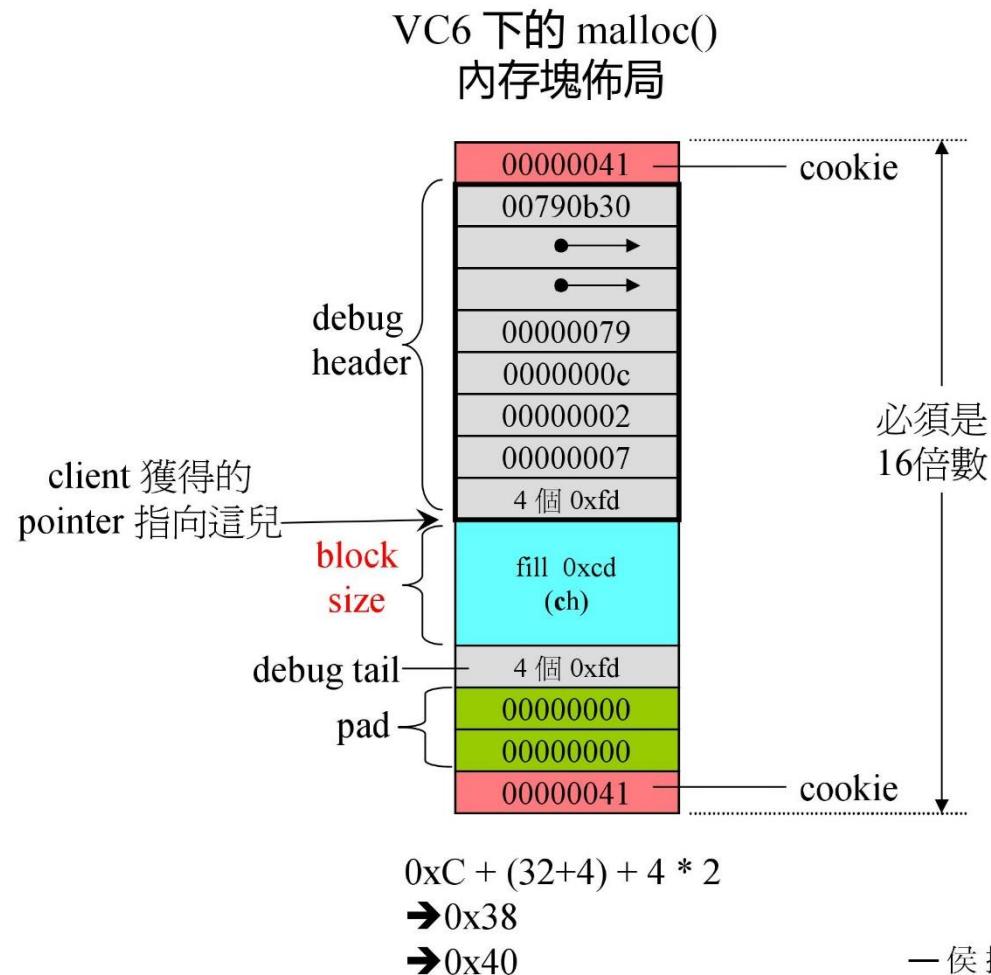


侯捷

西北有高樓
上與浮雲齊



VC6 malloc()





VC6 標準分配器之實現

VC6 所附的標準庫，其 std::allocator 實現如下 (<xmemory>)

```
template<class _Ty>
class allocator {
public:
    typedef _SIZT size_type;
    typedef _PDTI difference_type;
    typedef _Ty _FARQ *pointer;
    typedef _Ty value_type;
    pointer allocate(size_type _N, const void *)
    { return (_Allocate((difference_type)_N, (pointer)0)); }
    void deallocate(void _FARQ * _P, size_type)
    { operator delete(_P); }
};
```

#ifndef _FARQ
#define _FARQ
#define _PDTI ptrdiff_t
#define _SIZT size_t
#endif

VC6+ 的 allocator 只是以 ::operator new 和 ::operator delete 完成 allocate() 和 deallocate()，沒有任何特殊設計。

没有做内存管理
只是 malloc

```
template<class _Ty,
        class _A = allocator<_Ty>>
class vector
{ ...};
```

```
template<class _Ty,
        class _A = allocator<_Ty>>
class list
{ ...};
```

```
template<class _Ty,
        class _A = allocator<_Ty>>
class deque
{ ...};
```

```
template<class _K,
        class _Pr = less<_K>,
        class _A = allocator<_K>>
class set { ...};
```

其中用到的 _Allocate() 定義如下：

```
template<class _Ty> inline
_ty _FARQ *_Allocate(_PDTI _N, _Ty _FARQ *)
{ if (_N < 0) _N = 0;
  return ((_Ty _FARQ*) operator new((_SIZT) _N * sizeof(_Ty))); }
```

//分配 512 ints.
int* p = allocator<int>().allocate(512, (int*)0);
allocator<int>().deallocate(p, 512);



BC5 標準分配器之實現

BC5 所附的標準庫，其 std::allocator 實現如下 (<memory.stl>)

```
template <class T>
class allocator
{
public:
    typedef size_t           size_type;
    typedef ptrdiff_t        difference_type;
    typedef T*               pointer;
    typedef T                value_type;

    pointer allocate(size_type n,
                      allocator<void>::const_pointer = 0) {
        pointer tmp =
            _RWSTD_STATIC_CAST(pointer, (::operator new
                (_RWSTD_STATIC_CAST(size_t, (n*sizeof(value_type))))));
        _RWSTD_THROW_NO_MSG(tmp == 0, bad_alloc);
        return tmp;
    }
    void deallocate(pointer p, size_type) {
        ::operator delete(p);
    }
    ...
};
```

沒有做內存管理
只是 malloc

BC5 的 allocator 只是以 ::operator new 和 ::operator delete 完成 allocate() 和 deallocate()，沒有任何特殊設計。

```
template <class T,
          class Allocator=allocator<T>>
class vector
{ ...};
```

```
template <class T,
          class Allocator=allocator<T>>
class list
{ ...};
```

```
template <class T,
          class Allocator=allocator<T>>
class deque
{ ...};
```



//分配 512 ints.
int* p = allocator<int>().allocate(512);
allocator<int>().deallocate(p,512);

■■■ G2.9 標準分配器之實現

G2.9 所附的標準庫，其 std::allocator 實現如下 (<defalloc.h>)

G2.9 的 allocator 只是以 ::operator new 和 ::operator delete 完成 allocate() 和 deallocate()，沒有任何特殊設計。

```
template <class T>
class allocator {
public:
    typedef T      value_type;
    typedef T*     pointer;
    typedef size_t size_type;
    typedef ptrdiff_t difference_type;
    pointer allocate(size_type n) {
        return ::allocate((difference_type)n, (pointer)0);
    }
    void deallocate(pointer p) { ::deallocate(p); }
};
```

```
template <class T>
inline T* allocate(ptrdiff_t size, T*) {
    set_new_handler(0);
    T* tmp = (T*)
        (::operator new((size_t)(size*sizeof(T))));
    if (tmp == 0) {
        cerr << "out of memory" << endl;
        exit(1);
    }
    return tmp;
}
template <class T>
inline void deallocate(T* buffer)
    ::operator delete(buffer);
}
```

G++ <defalloc.h> 中有這樣的註釋：**DO NOT USE THIS FILE** unless you have an old container implementation that requires an allocator with the HP-style interface. **SGI STL uses a different allocator interface.** SGI-style allocators are not parametrized with respect to the object type; they traffic in void* pointers. **This file is not included by any other SGI STL header.**



G2.9 容器使用的分配器, 不是 std::allocator 而是 std::alloc



```
template <class T,  
         class Alloc = alloc>  
class vector {  
    ...  
};
```

```
template <class T,  
         class Alloc = alloc>  
class list {  
    ...  
};
```

```
template <class T,  
         class Alloc = alloc,  
         size_t BufSiz = 0>  
class deque {  
    ...  
};
```



```
//分配 512 bytes  
void* p = alloc::allocate(512);  
//也可以這樣alloc().allocate(512);  
alloc::deallocate(p,512);
```

```
template <class Key,  
         class T,  
         class Compare = less<Key>,  
         class Alloc = alloc>  
class map {  
    ...  
};
```

```
template <class Key,  
         class Compare = less<Key>,  
         class Alloc = alloc>  
class set {  
    ...  
};
```



而今安在哉？ std::alloc vs. __pool_alloc

```

class __pool_alloc_base
{
protected:
    enum { __S_align = 8 };
    enum { __S_max_bytes = 128 };
    enum { __S_free_list_size = (size_t) __S_max_bytes / (size_t) __S_align };

    union __Obj
    {
        union __Obj* __M_free_list_link;
        char __M_client_data[1]; // The client sees this.
    };

    static __Obj* volatile __S_free_list[__S_free_list_size];
    // Chunk allocation state.

    static char* __S_start_free;
    static char* __S_end_free;
    static size_t __S_heap_size;
    ...
}

template<typename _Tp>
class __pool_alloc : private __pool_alloc_base
{ ... };

```

— 侯捷 —

G2.9

用例：
`vector<string> vec;`

G4.9

G4.9 標準庫中有許多 extented allocators,
其中 __pool_alloc 就是 G2.9 alloc 的化身.

G2.9

```

enum { __ALIGN = 8 };
enum { __MAX_BYTES = 128 };
enum { __NFREELISTS = __MAX_BYTES/__ALIGN};

template <bool threads, int inst>
class __default_alloc_template
{
private:
    typedef __default_alloc_template<false,0> alloc;

    union obj {
        union obj* free_list_link;
    };

    static obj* volatile free_list[__NFREELISTS];
    ...
    // Chunk allocation state.

    static char* start_free;
    static char* end_free;
    static size_t heap_size;
    ...
}

```



而今安在哉？ std::alloc vs. __pool_alloc

```
// Allocate memory in large chunks in order to avoid fragmenting the
// heap too much. Assume that __n is properly aligned. We hold the
// allocation lock.
char*
__pool_alloc_base::__M_allocate_chunk(size_t __n, int& __nobjs)
{
    char* __result;
    size_t __total_bytes = __n * __nobjs;
    size_t __bytes_left = __S_end_free - __S_start_free;
    ...
    __try {
        __S_start_free = static_cast<char*>(__operator new(__bytes_to_get));
    }
    __catch(const std::bad_alloc&){
        // Try to make do with what we have. That can't hurt. We
        // do not try smaller requests, since that tends to result
        // in disaster on multi-process machines.
        size_t __i = __n;
        for(; __i <= (size_t) __S_max_bytes; __i += (size_t) __S_align) {
            ...
        }
        // What we have wasn't enough. Rethrow.
        __S_start_free = __S_end_free = 0; // We have no chunk.
        __throw_exception_again;
    }
    __S_heap_size += __bytes_to_get;
    __S_end_free = __S_start_free + __bytes_to_get;
    return __M_allocate_chunk(__n, __nobjs);
}
```

G4.9

G4.9 標準庫中有許多 extented allocators,
其中 `__pool_alloc` 就是 G2.9 的 `alloc` 的化身.

```
template <bool threads, int inst>
char* __default_alloc_template<threads, inst>::
chunk_alloc(size_t size, int& nobjs)
{
    char* result;
    size_t total_bytes = size * nobjs;
    size_t bytes_left = end_free - start_free;
    ...
    start_free = (char*)malloc(bytes_to_get);
    if(0 == start_free) {
        int i;
        obj* volatile *my_free_list, *p;

        //Try to make do with what we have. That can't
        //hurt. We do not try smaller requests, since that tends
        //to result in disaster on multi-process machines.
        for(i = size; i <= __MAX_BYTES; i += __ALIGN) {
            ...
        }
        end_free = 0; //In case of exception.
        start_free = (char*)malloc_alloc::allocate(bytes_to_get);
        //This should either throw an exception or
        //remedy the situation. Thus we assume it
        //succeeded.
    }
    heap_size += bytes_to_get;
    end_free = start_free + bytes_to_get;
    return(chunk_alloc(size, nobjs));
}
```

G2.9



G4.9 標準分配器之實現

```
template<typename _Tp>      <.../bits/new_allocator.h>
class new_allocator
{
...
pointer allocate(size_type __n, const void* = 0) {
    if (__n > this->max_size())
        std::__throw_bad_alloc();
    return static_cast<_Tp*>(
        ::operator new(__n * sizeof(_Tp)));
}

void deallocate(pointer __p, size_type)
{ ::operator delete(__p); }
...
};
```

define __allocator_base __gnu_cxx::new_allocator

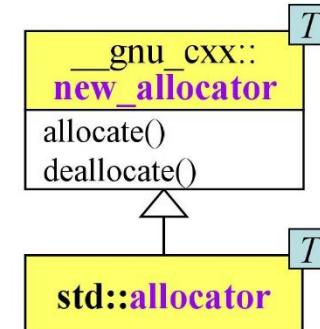
```
template<typename _Tp>      <.../bits/allocator.h>
class allocator: public __allocator_base<_Tp>
{
...
};
```

— 侯捷 —

```
template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class vector : protected _Vector_base<_Tp, _Alloc>
{ ... };

template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class list : protected _List_base<_Tp, _Alloc>
{ ... };

template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class deque : protected _Deque_base<_Tp, _Alloc>
{ ... };
```



■■■ G4.9 pool allocator 用例

```
//欲使用 std::allocator 以外的 allocator, 就得自行 #include <ext/...>
#include <ext/pool_allocator.h>
```

```
template<typename Alloc>
void cookie_test(Alloc alloc, size_t n)
{
    typename Alloc::value_type *p1, *p2, *p3;
    p1 = alloc.allocate(n);
    p2 = alloc.allocate(n);
    p3 = alloc.allocate(n);
```

```
cout << "p1= " << p1 << '\t' << "p2= " << p2 << '\t' << "p3= " << p3 << '\n';
```

```
alloc.deallocate(p1,sizeof(typename Alloc::value_type));
alloc.deallocate(p2,sizeof(typename Alloc::value_type));
alloc.deallocate(p3,sizeof(typename Alloc::value_type));
```

```
}
```



```
cout << sizeof(__gnu_cxx::__pool_alloc<int>) << endl; //1
vector<int, __gnu_cxx::__pool_alloc<int>> vecPool;
cookie_test(__gnu_cxx::__pool_alloc<double>(), 1);
//相距 08h (表示不帶 cookie)
```



```
p1= 0xae4138 p2= 0xae4140 p3= 0xae4148
```

```
p1= 0xae25e8 p2= 0xaddeb50 p3= 0xadd090
```

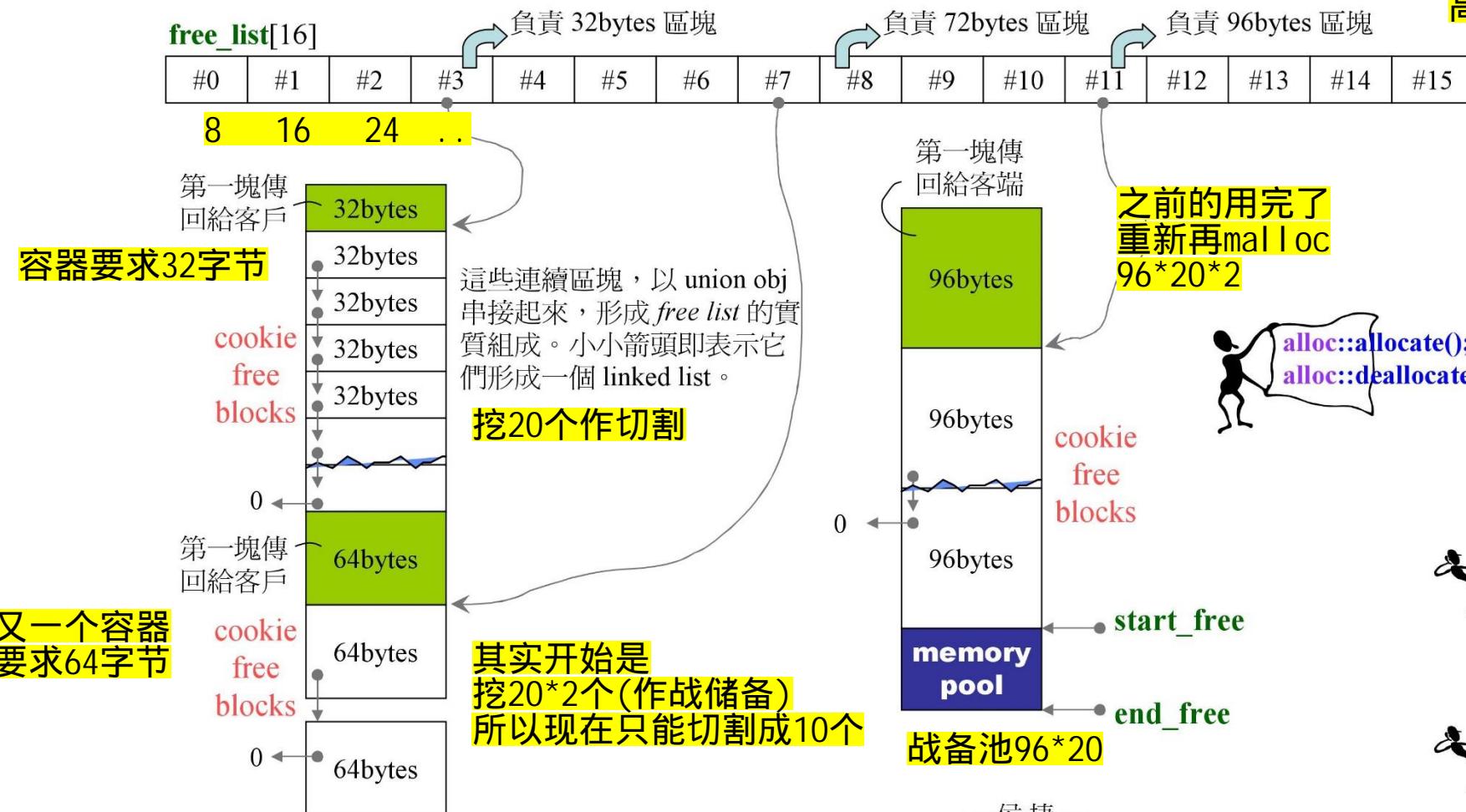


```
cout << sizeof(std::allocator<int>) << endl; //1
vector<int, std::allocator<int>> vec;
cookie_test(std::allocator<double>(), 1);
//相距 10h (表示帶 cookie)
```



```
p1= 0x3e4098 p2= 0x3e4088 p3= 0x3e4078
```

■■■ G2.9 std::alloc 運行模式



—侯捷—

高于上限128则直接调用malloc

```
template <class T,  
         class Alloc = alloc>  
class vector { ...};  
  
template <class T,  
         class Alloc = alloc>  
class list { ... };  
  
template <class T,  
         class Alloc = alloc,  
         size_t BufSiz = 0>  
class deque { ... };
```

```
template <class Key,  
         class T,  
         class Compare= less<Key>,  
         class Alloc = alloc>  
class map { ... };
```

```
template <class Key,  
         class Compare = less<Key>,  
         class Alloc = alloc>  
class set { ... };
```

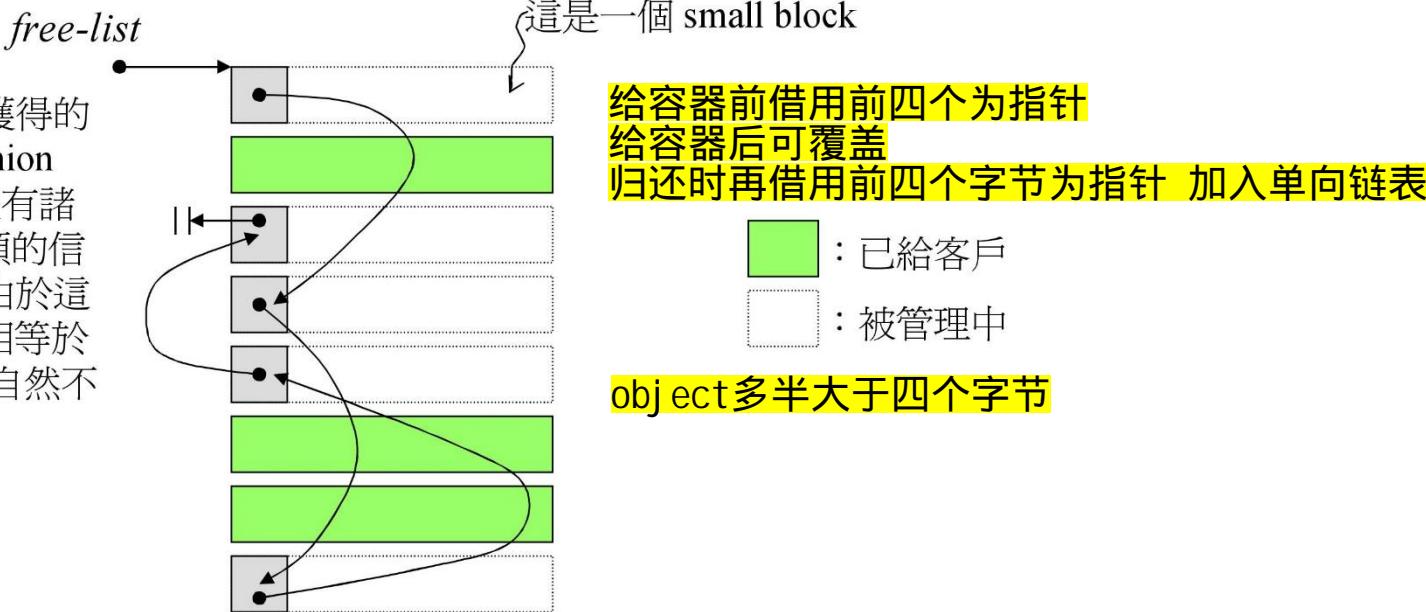
容器发出的需求会自动调到8的边界



embedded pointers

嵌入式指针

當客戶端獲得小區塊，獲得的即是 `char*`（指向某個 `union obj`）。此時雖然客戶端沒有諸如 `LString` 或 `ZString` 之類的信息可得知區塊大小，但由於這區塊是給 `object` 所用，相等於 `object` 大小，`object` ctor 自然不會逾份。



改用 struct
可

```
union obj {
    union obj* free_list_link;
    char client_data[1]; // client sees this
};
```



G2.9 std::alloc 運行一瞥.01

free_list[16]

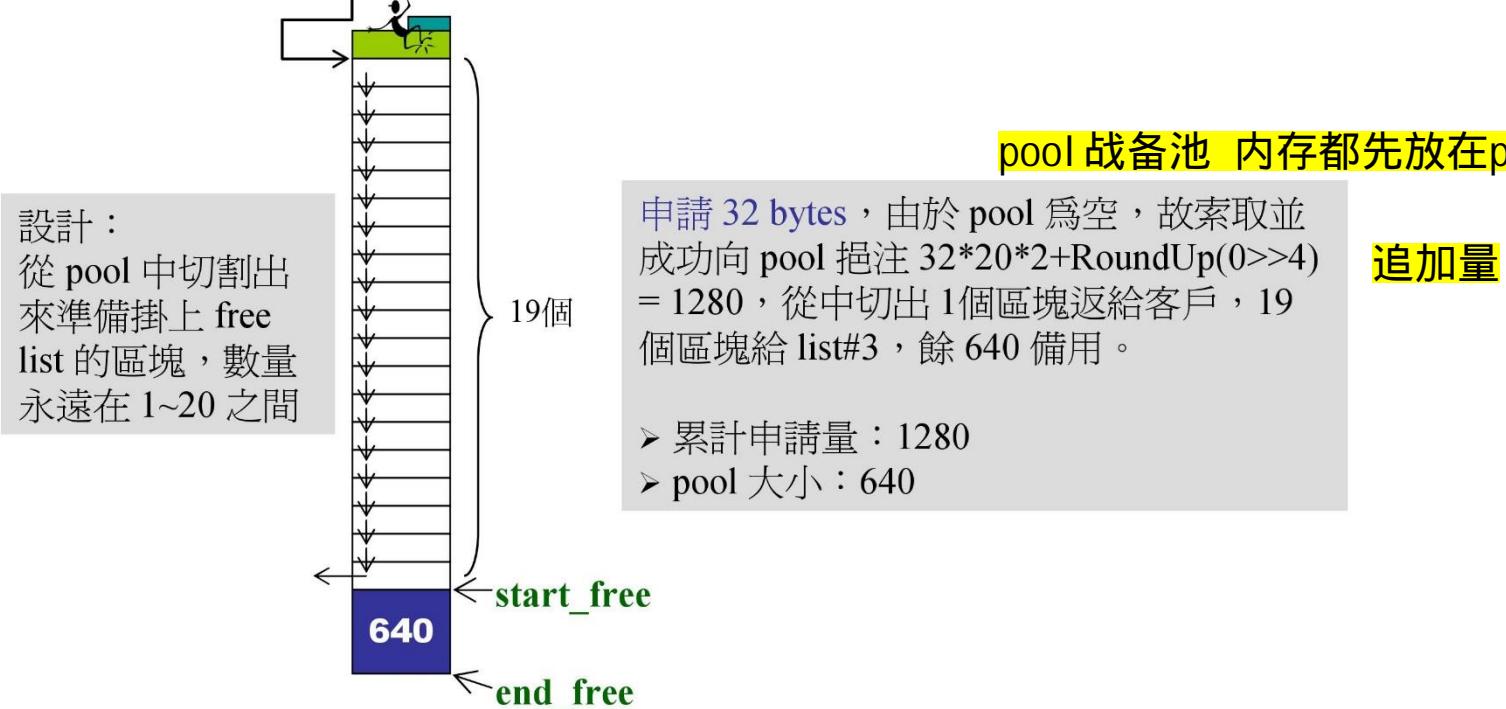
#0 #1 #2 #3 #4 #5 #6 #7 #8 #9 #10 #11 #12 #13 #14 #15



G2.9 std::alloc 運行一瞥.02

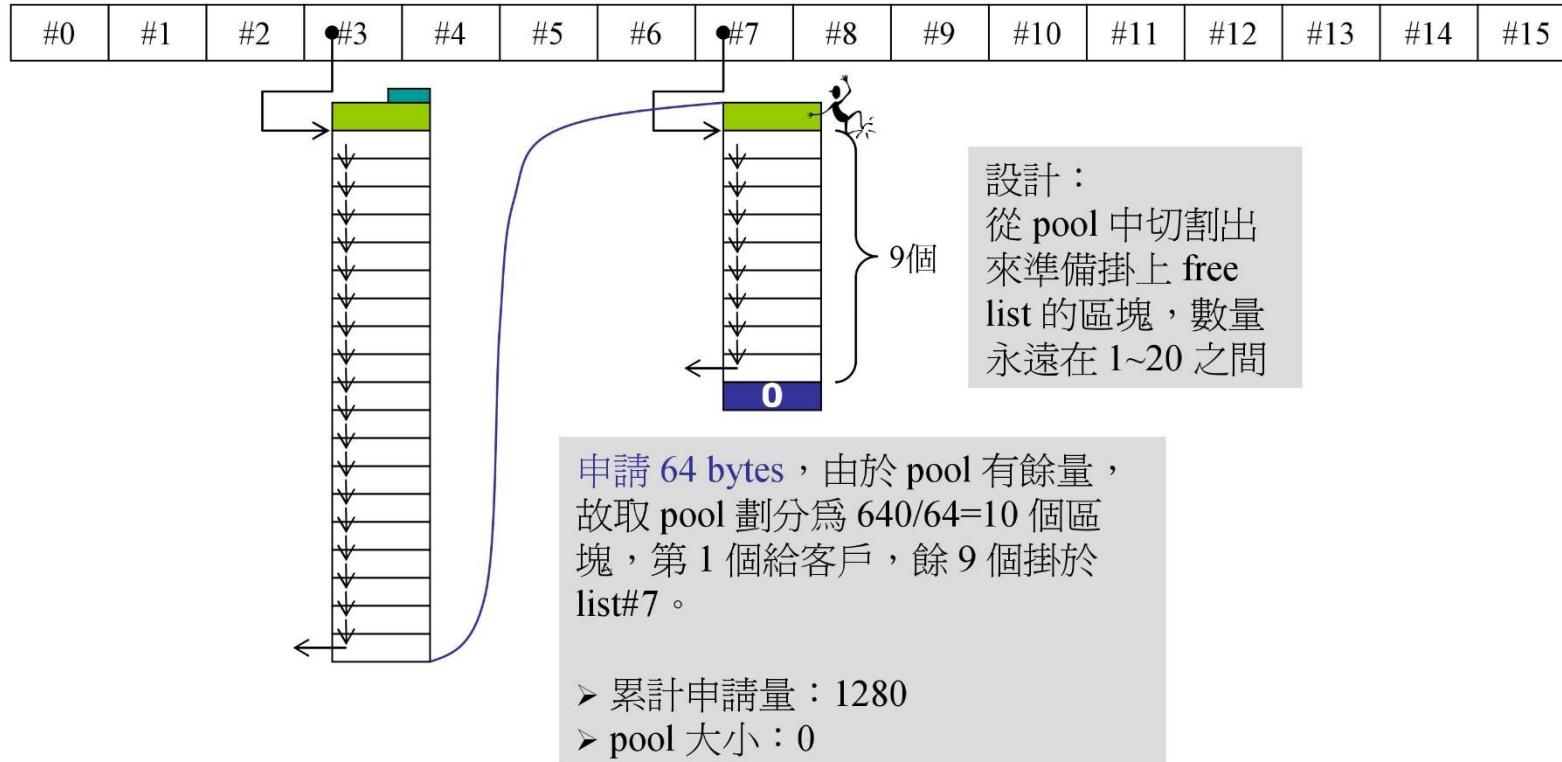
free_list[16]

#0	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----



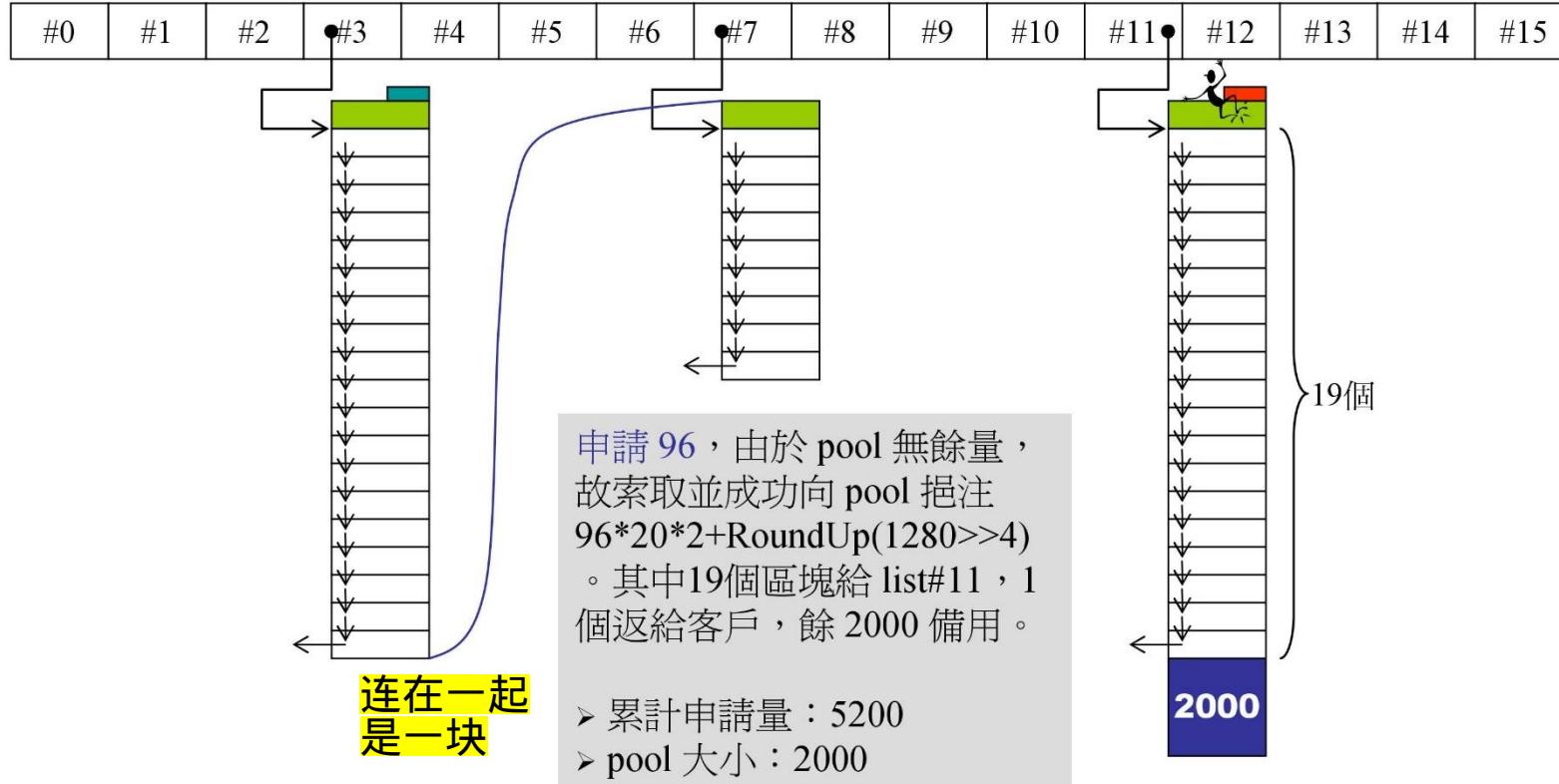


G2.9 std::alloc 運行一瞥.03



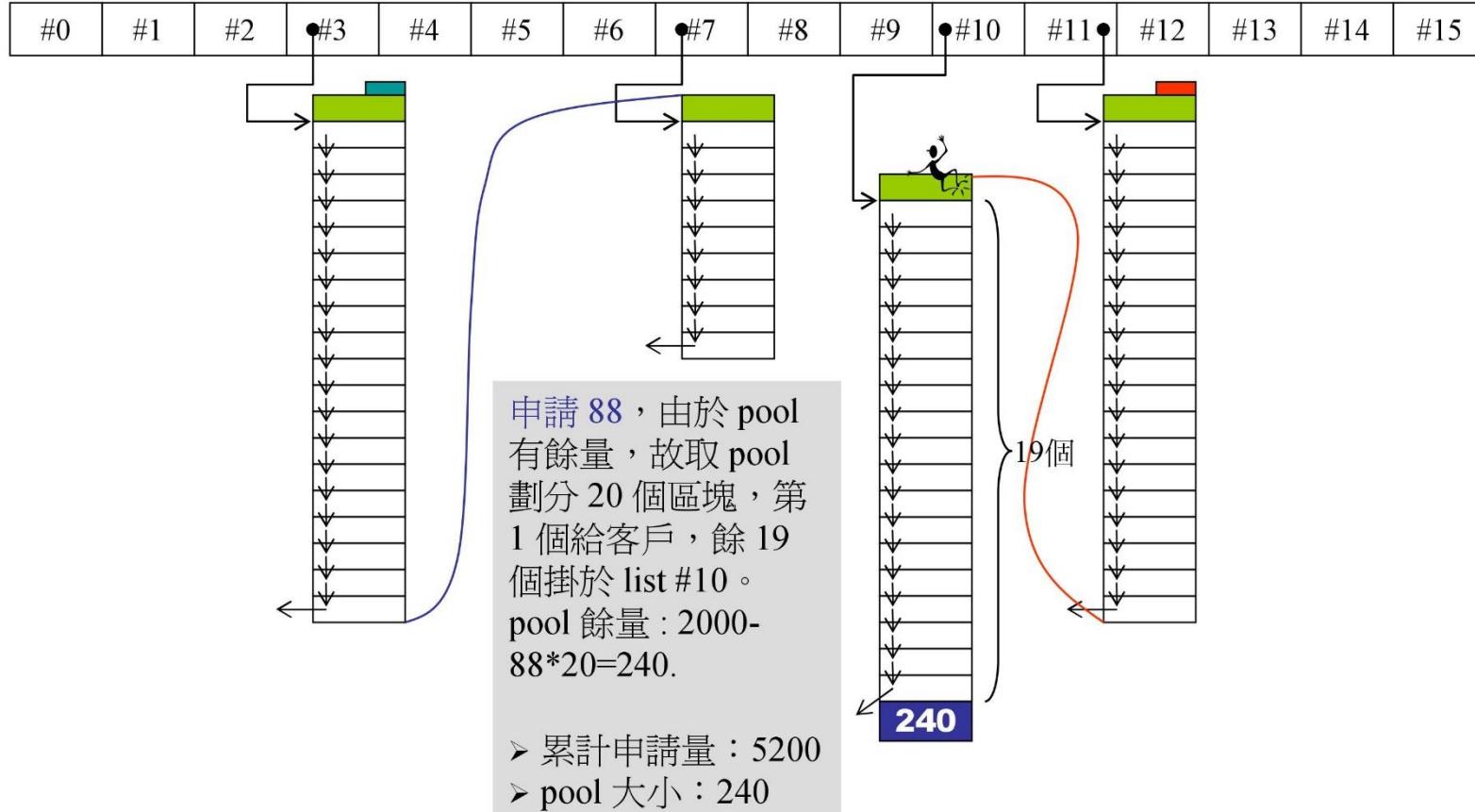


G2.9 std::alloc 運行一瞥.04



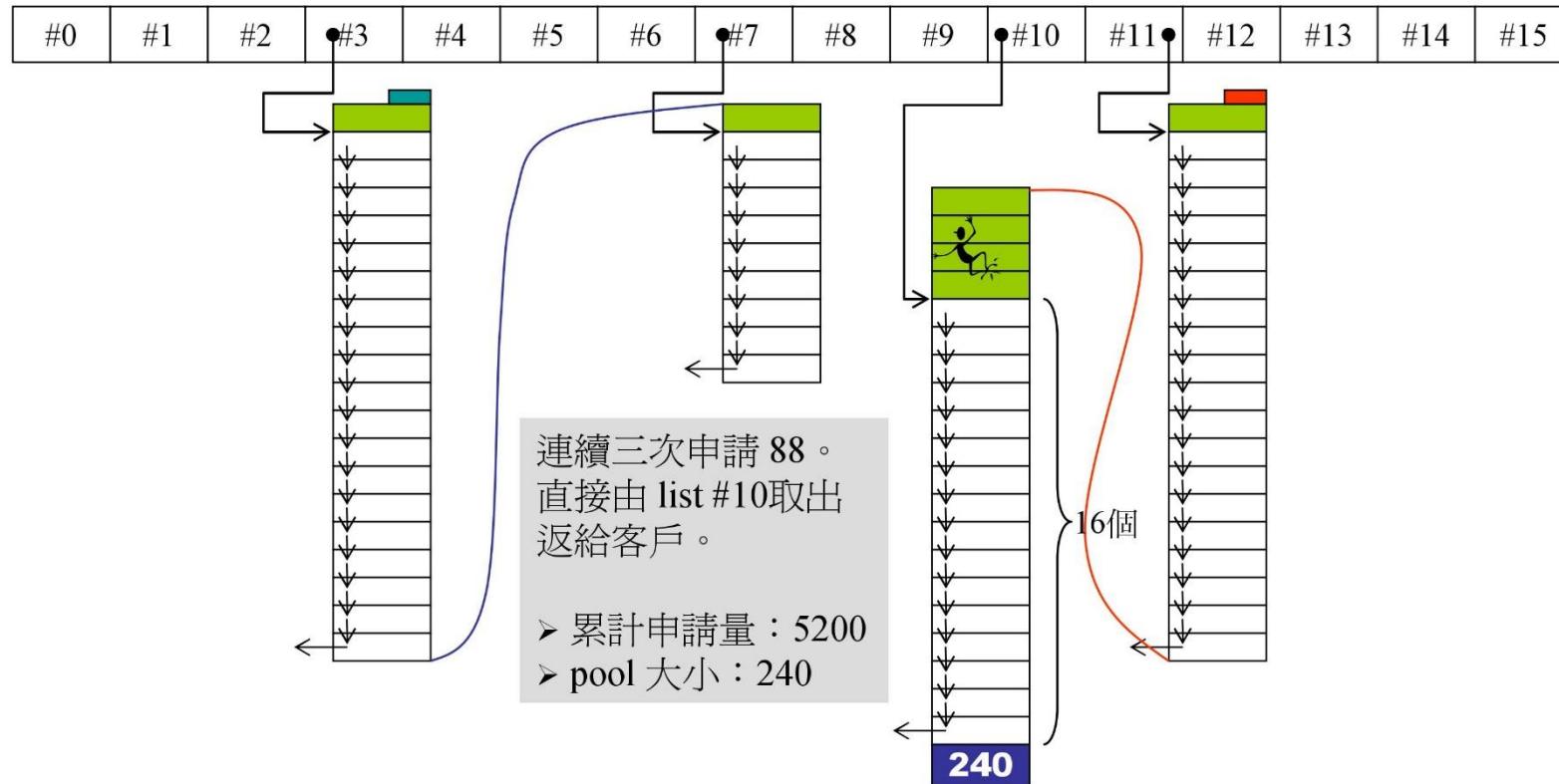


G2.9 std::alloc 運行一瞥.05



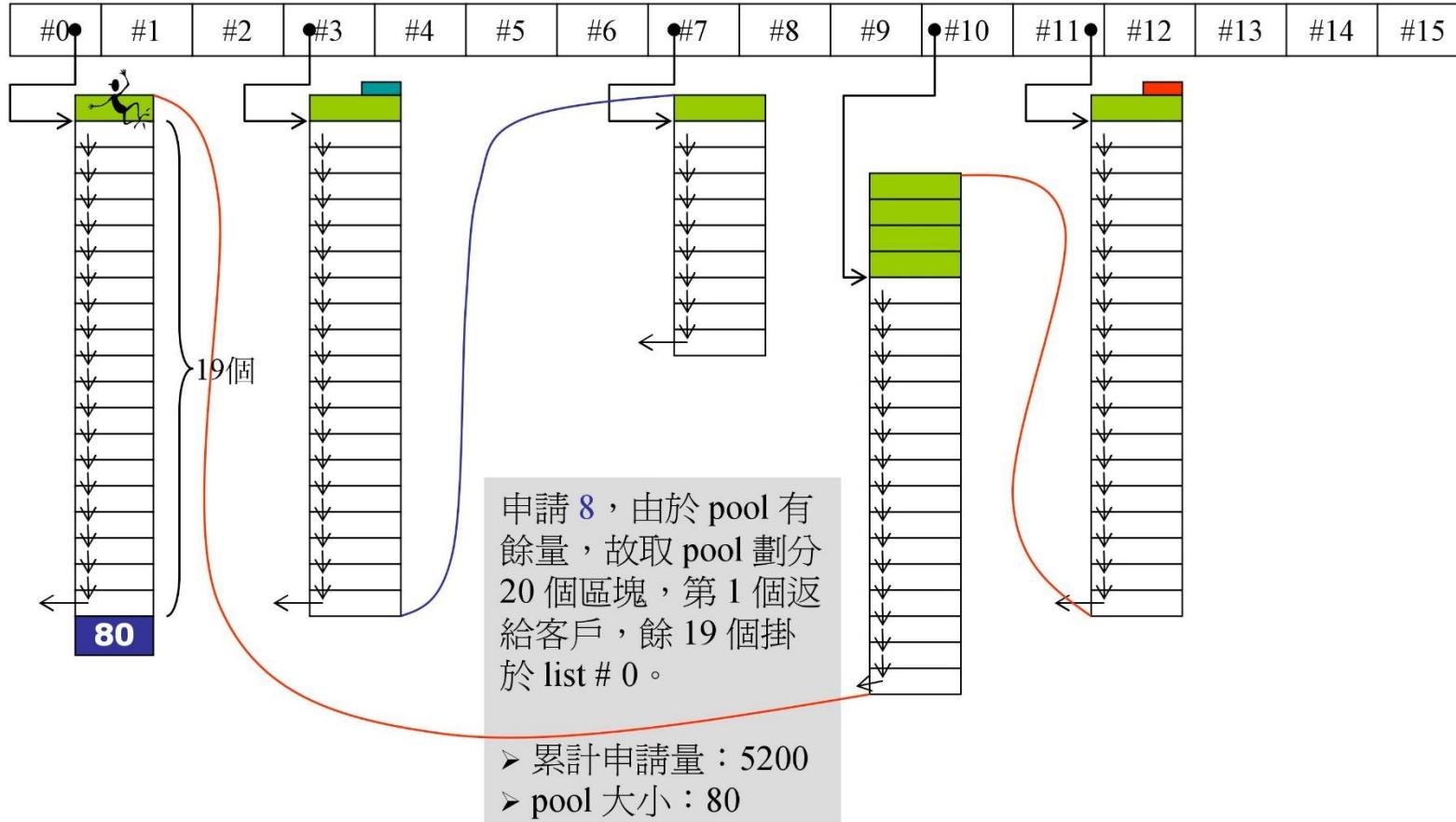


G2.9 std::alloc 運行一瞥.06



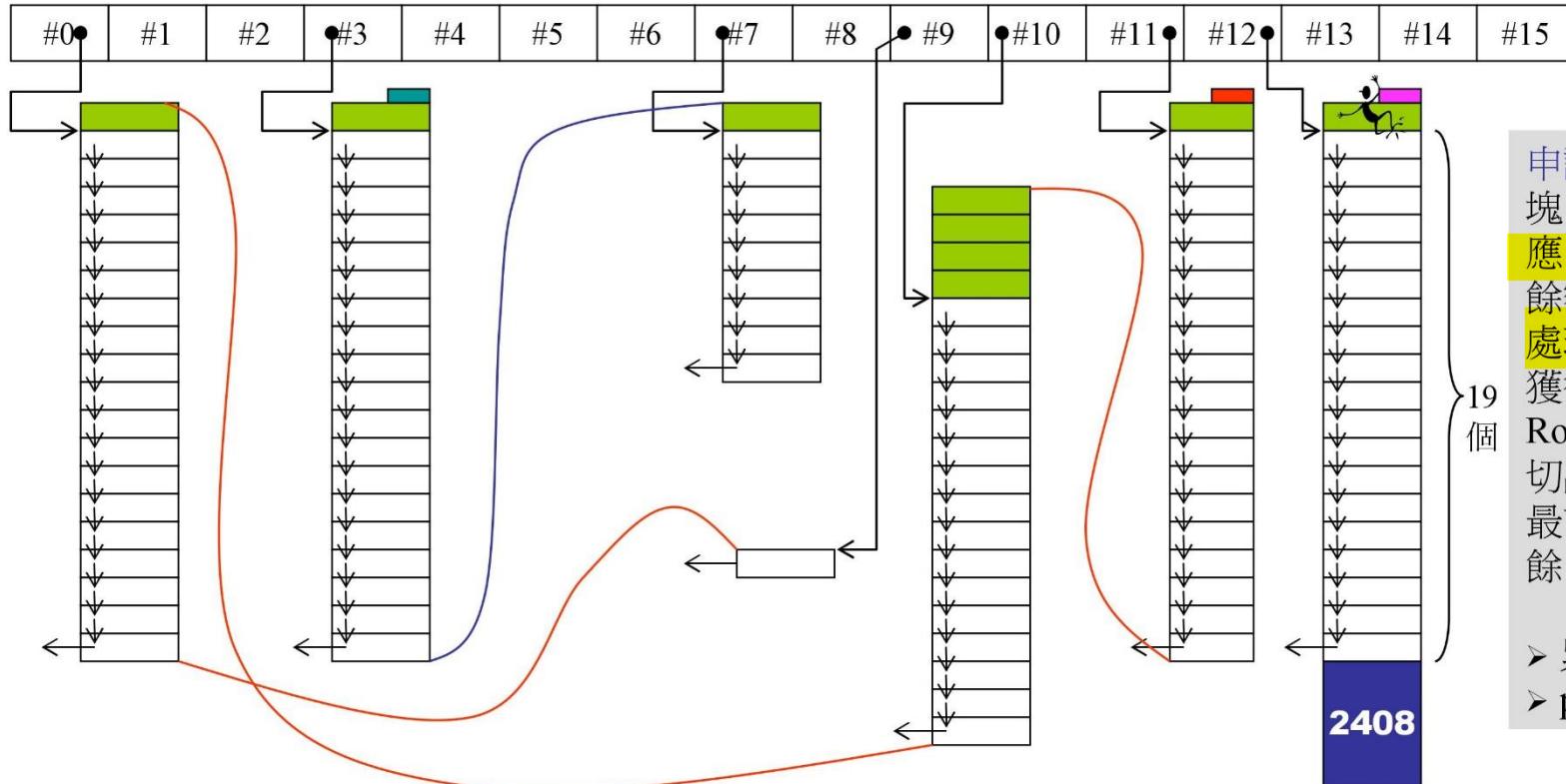


G2.9 std::alloc 運行一瞥.07





G2.9 std::alloc 運行一瞥.08

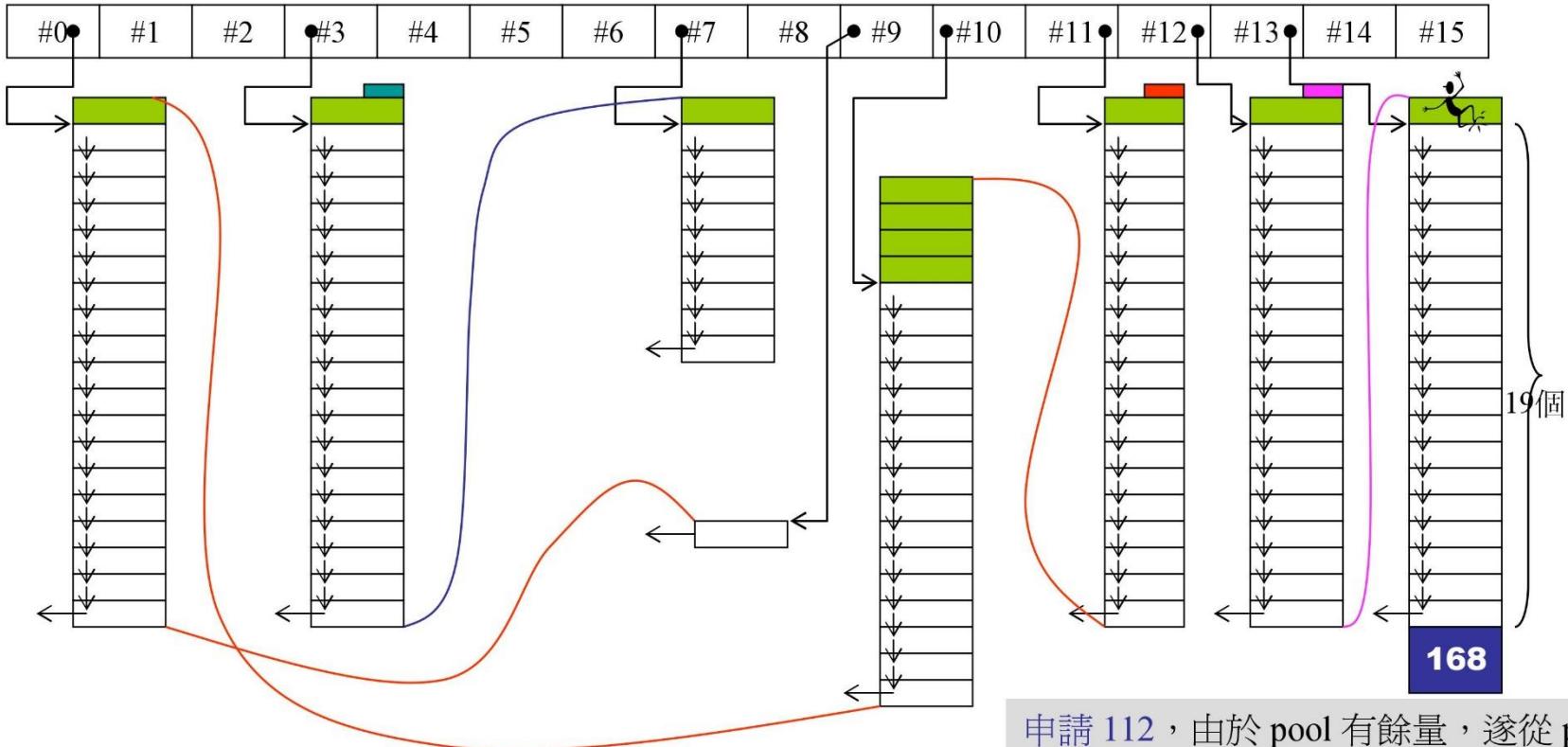


申請 104，list #12 無區塊，pool 餘量又不足供應 1 個，於是先將 pool 餘額撥給 list #9 (碎片處理)，然後索取並成功獲得挹注 $104 * 20 * 2 + \text{RoundUp}(5200 > 4)$ ，切出 19 個給 list #12，最前頭那個返給客戶，餘 2408 備用。

- 累計申請量：9688
- pool 大小：2408



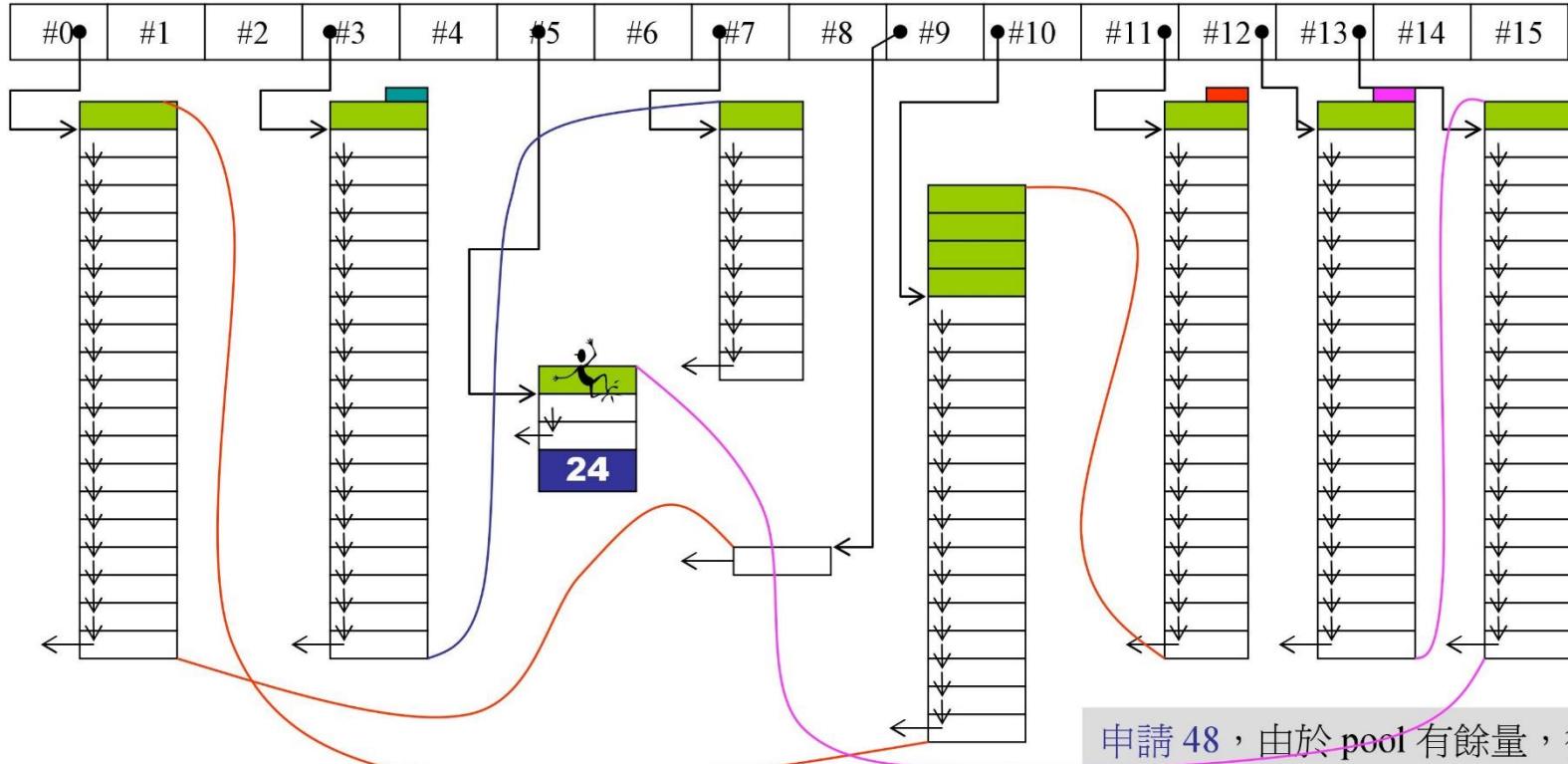
G2.9 std::alloc 運行一瞥.09



申請 112，由於 pool 有餘量，遂從 pool 取 20 個
區塊，撥 1 個返給客戶，留 19 個掛於 list #13.

- 累計申請量：9688
- pool 大小： $2408 - 112 \times 20 = 168$.

■ ■ ■ G2.9 std::alloc 運行一瞥.10

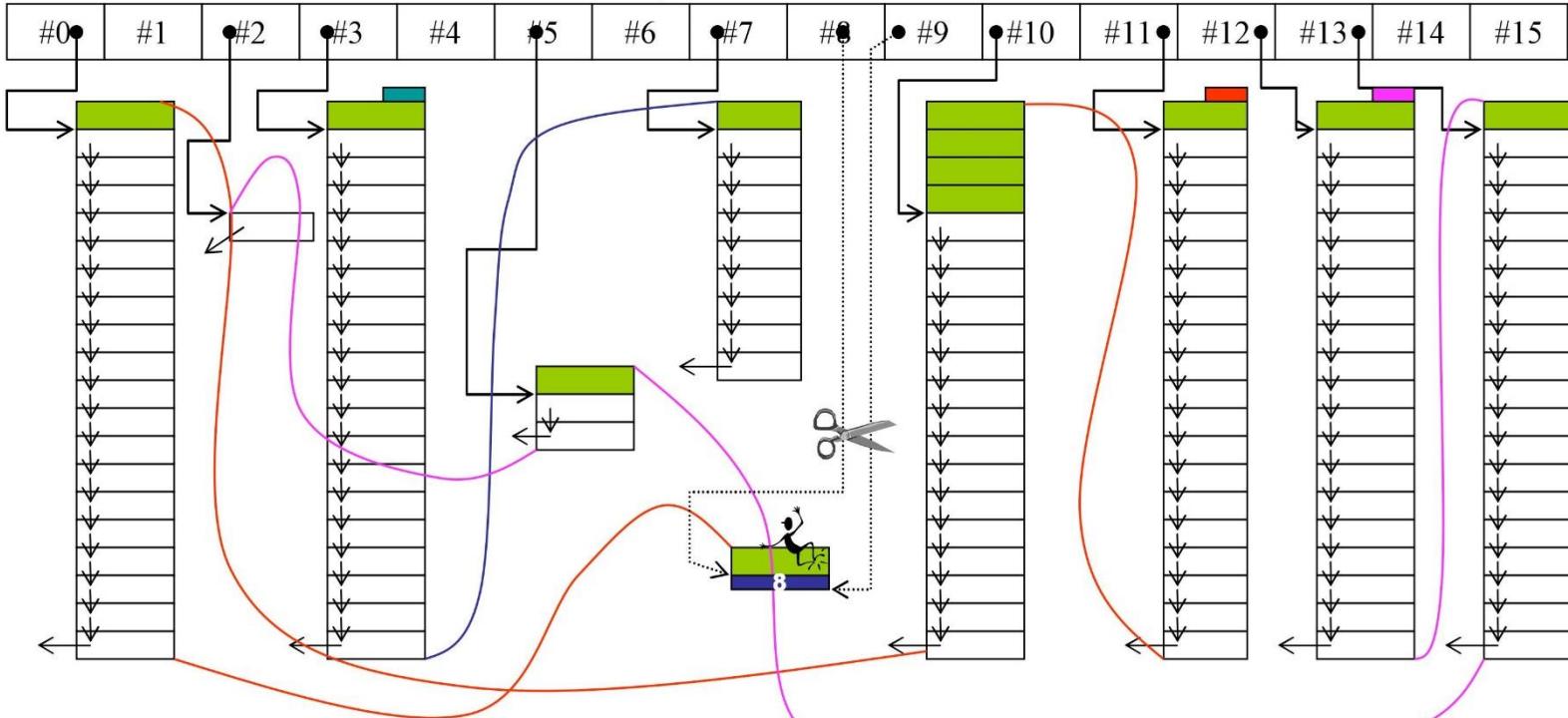


申請 48，由於 pool 有餘量，從 pool 取 3 個區塊，撥 1 個返給客戶，2 個掛於 list #5。

- 累計申請量：9688
 - pool 大小： $168 - 48 * 3 = 24$



G2.9 std::alloc 運行一瞥.11



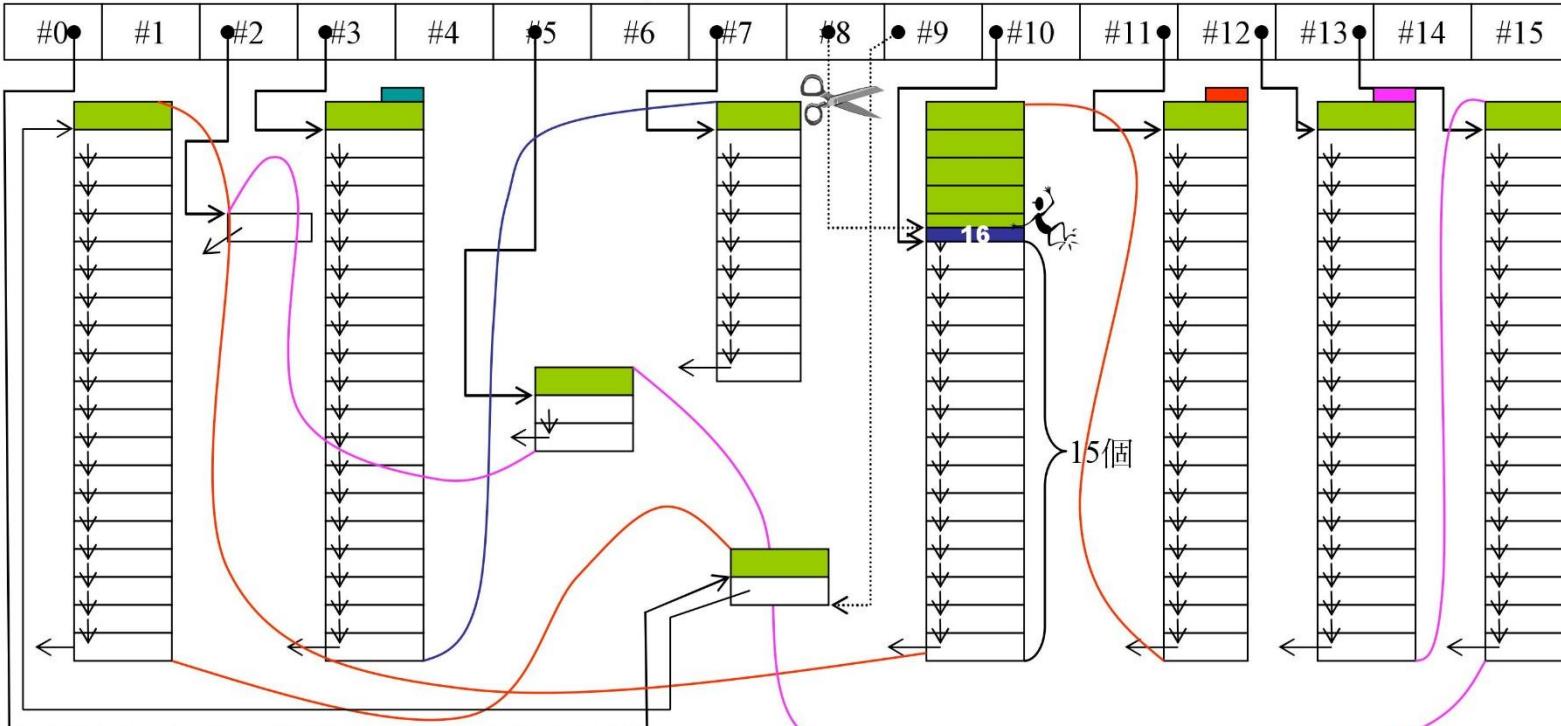
申請 72，list#8 無可用區塊，pool 餘量又不足供應 1 個，於是先將 pool 餘額撥給list#2，然後索取
 $72*20*2+\text{RoundUp}(9688>>4)$ ，但為觀察系統邊界，我將 system heap 大小設為 10000，目前已索取 9688，因此
無法滿足此次索取，於是 alloc 從手中資源 **取最接近之 80 (list#9)** 回填 pool，再從中切出 72 返回客戶，餘 8。

- 累計申請量：9688
- pool 大小：8

向右找 有空的则借来



G2.9 std::alloc 運行一瞥.12

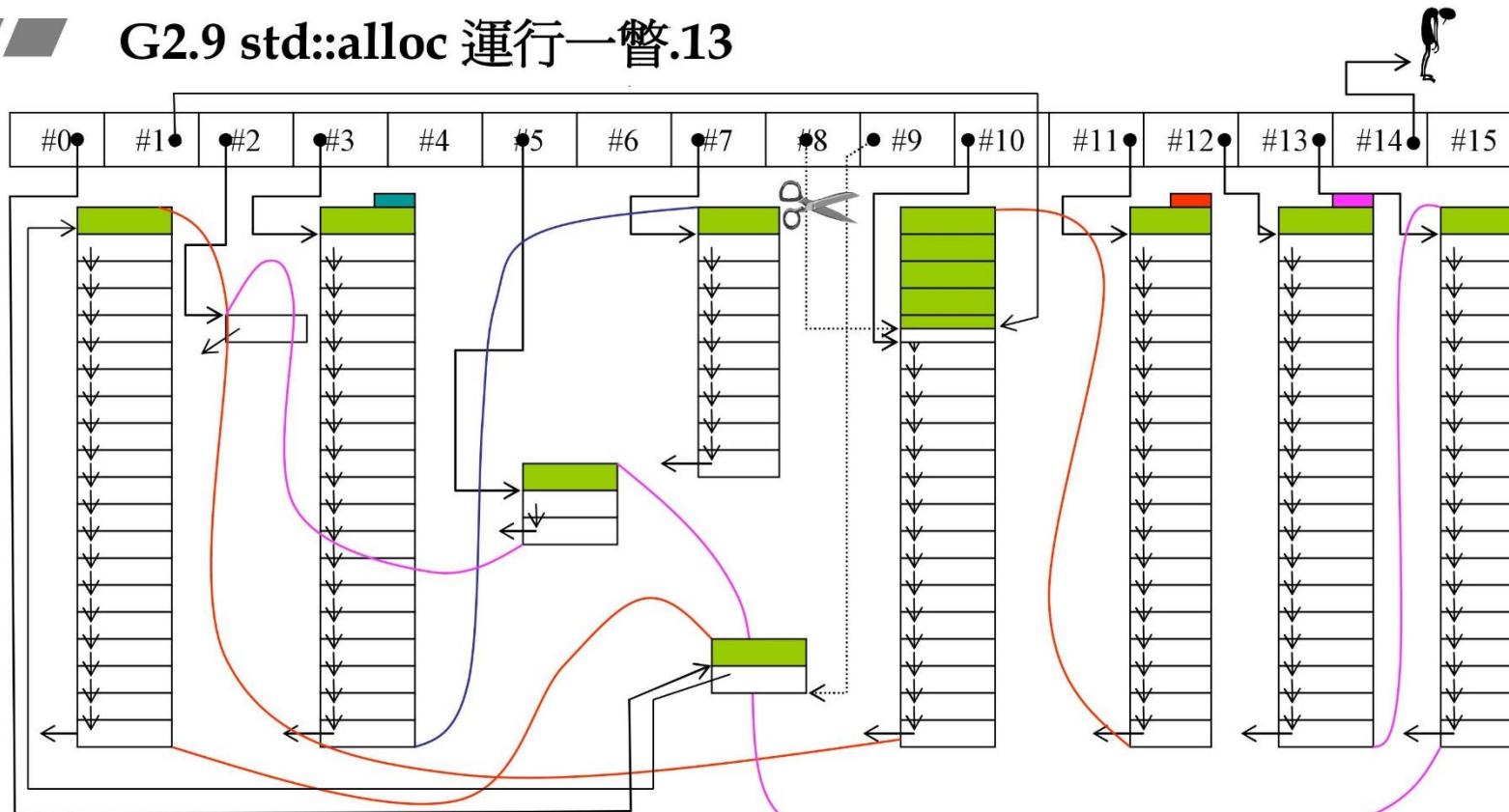


申請 72，list#8無可用區塊，pool 餘量又不足供應1個，於是先將 pool 餘額撥給 list#0，然後索取
 $72*20*2+RoundUp(9688>>4)$ ，但 system heap 大小已被我設為 10000，目前已索取 9688，因此無法滿足此次
索取，於是 alloc 從手中資源取最接近之 88 (list#10) 回填 pool，再從中切出 72 返給客戶，餘 16。

- 累計申請量：9688
- pool 大小：16



G2.9 std::alloc 運行一瞥.13

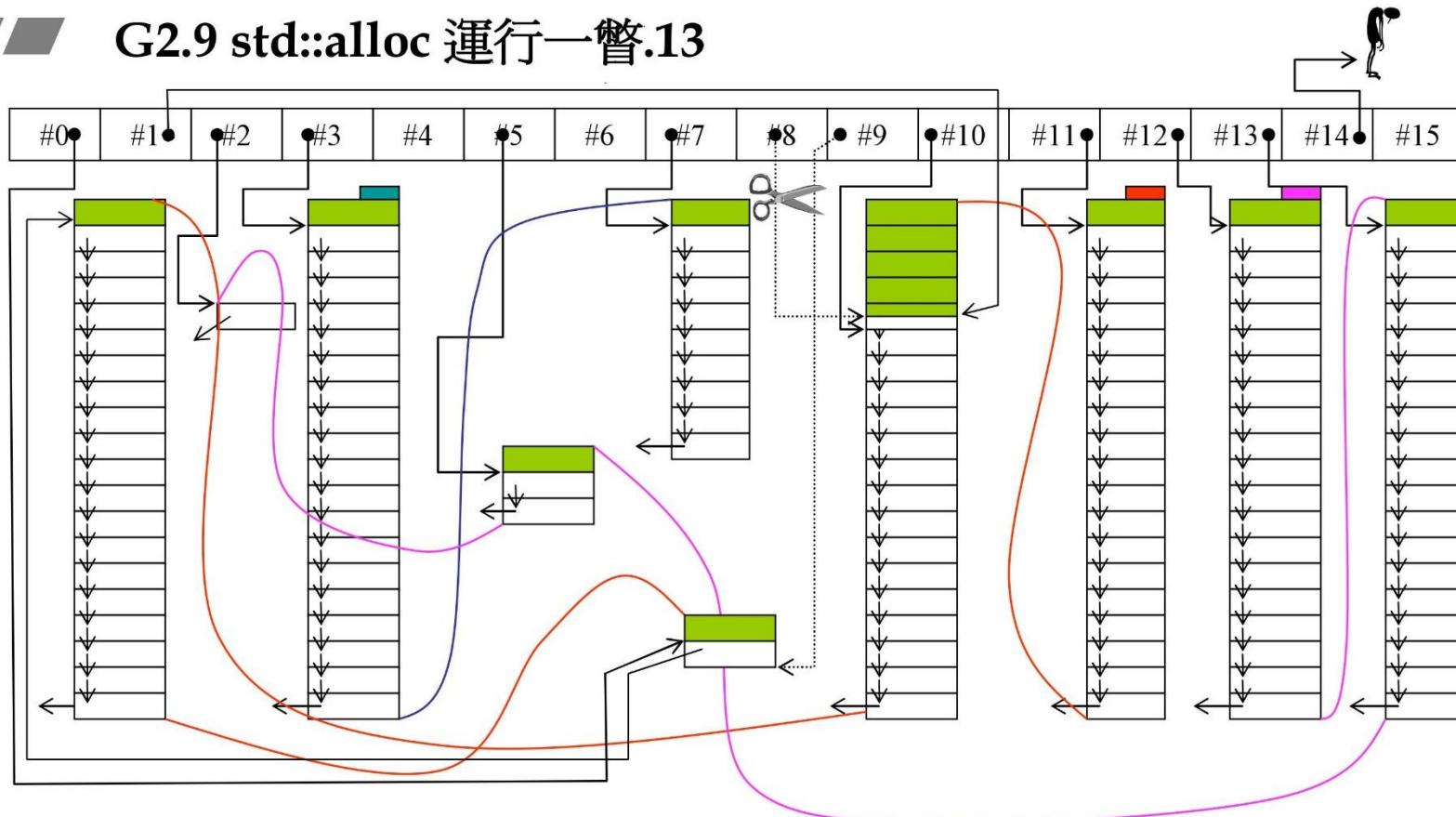


申請 120，list#14無可用區塊，pool 餘量又不足供應 1 個，於是先將 pool 餘額撥給 list#1，然後索取 $120*20*2+RoundUp(9688>>4)$ ，但 system heap 大小已被我設為 10000，目前已索取 9688，因此無法滿足此次索取，於是 alloc 從手中資源取最接近者回填 pool，但是找不到，終於窮途末路山窮水盡日薄西山。

- 累計申請量：9688
- pool 大小：0



G2.9 std::alloc 運行一瞥.13



累計量 : 9688

- 檢討**
- 檢討 1 : `alloc` 手中還有多少資源？以上白色皆是。可不可以將白色小區塊合成為較大區塊供應給客戶？唔，技術難度極高。
 - 檢討 2 : `system heap` 手中還剩多少資源？ $10000 - 9688 = 312$ 。可不可以將失敗的那次索取量折半...再折半...再折半...最終當索取量 ≤ 312 便能獲得滿足。

G2.9 std::alloc 源碼剖析, 1

```
#008 // 第一級分配器
#010 template <int inst>
#011 class __malloc_alloc_template {
#012 private:
#013     static void* oom_malloc(size_t);
#014     static void* oom_realloc(void*, size_t);
#015     static void (*__malloc_alloc_oom_handler)();
#016     //上面這個指針用來指向new-handler(if any)
#017 public:
#018     static void* allocate(size_t n)
#019     {
#020         void *result = malloc(n);    //直接使用malloc()
#021         if (0 == result) result = oom_malloc(n);
#022         return result;
#023     }
#024     static void deallocate(void *p, size_t /* n */)
#025     {
#026         free(p);                  //直接使用free()
#027     }
#028     static void* realloc(void *p, size_t /* old_sz */, size_t new_sz)
#029     {
#030         void * result = realloc(p, new_sz); //直接使用realloc()
#031         if (0 == result) result = oom_realloc(p, new_sz);
#032         return result;
#033     }
#034     static void (*set_malloc_handler(void (*f)()))()
#035     { //類似 C++ 的 set_new_handler().
#036         void (*old)() = __malloc_alloc_oom_handler; //記錄原new-handler
#037         __malloc_alloc_oom_handler = f; //把f 記起來以便爾後呼叫
#038         return(old); //把原先的handler傳回以便日後可恢復
#039     }
#040 };
```

```
#001 // #include <cstdlib>
#002 #include <cstddef>
#003 #include <new>
#004
#005 #define __THROW_BAD_ALLOC \
        cerr << "out of memory"; exit(1)
```

如果改用 operator new()，便可直接使用 C++ 的 set_new_handler()，不必寫出這般模擬。

相當於

```
typedef void (*H)();
H set_malloc_handler(H f);
```

my handler
void (f*)()



G2.9 std::alloc 源碼剖析, 2

```
#042 template <int inst>
#043 void (*__malloc_alloc_template<inst>::__malloc_alloc_oom_handler)() = 0;
#044
#045 template <int inst> //本例template參數'inst'完全沒派上用場
#046 void* __malloc_alloc_template<inst>::oom_malloc(size_t n)
#047 {
#048     void (*my_malloc_handler)();
#049     void* result;
#050
#051     for (;;) { //不斷嘗試釋放、分配、再釋放、再分配...
#052         my_malloc_handler = __malloc_alloc_oom_handler; //換個名稱
#053         if (0 == my_malloc_handler) { __THROW_BAD_ALLOC; }
#054         (*my_malloc_handler)(); //呼叫handler，企圖釋放memory
#055         result = malloc(n); //再次嘗試分配memory
#056         if (result) return(result);
#057     }
#058 }
```

有可能在#037被改掉

my handler

```
#060 template <int inst>
#061 void * __malloc_alloc_template<inst>::oom_realloc(void* p, size_t n)
#062 {
#063     void (*my_malloc_handler)();
#064     void* result;
#065
#066     for (;;) { //不斷嘗試釋放、分配、再釋放、再分配...
#067         my_malloc_handler = __malloc_alloc_oom_handler; //換個名稱
#068         if (0 == my_malloc_handler) { __THROW_BAD_ALLOC; }
#069         (*my_malloc_handler)(); //呼叫handler，企圖釋放memory
#070         result = realloc(p, n); //再次嘗試分配memory
#071         if (result) return(result);
#072     }
#073 }
```

是為了 multi-threading ?!



G2.9 std::alloc 源碼剖析, 3

```
#074 //-----
#075 typedef __malloc_alloc_template<0> malloc_alloc;
#076
#077 //一個換膚工程，將分配單位由 bytes 個數改為元素個數
#078 template<class T, class Alloc>
#079 class simple_alloc {
#080 public:
#081     static T* allocate(size_t n)           //一次分配n個T objects
#082     { return 0 == n? 0 : (T*)Alloc::allocate(n*sizeof(T)); }
#083     static T* allocate(void)             //一次分配1個T objects
#084     { return (T*)Alloc::allocate(sizeof(T)); }
#085     static void deallocate(T* p, size_t n) //一次歸還n個T objects
#086     { if (0 != n) Alloc::deallocate(p, n*sizeof(T)); }
#087     static void deallocate(T *p)          //一次歸還1個T objects
#088     { Alloc::deallocate(p, sizeof(T)); }
#089 };
```

用例：

```
template <class T, class Alloc = alloc>
class vector {
protected:
    // 專屬之分配器，每次分配一個元素大小
    typedef simple_alloc<T, Alloc>
        data_allocator;
    result = data_allocator::allocate(); //分配1個元素空間
    ...
    data_allocator::deallocate(result);
```

G2.9 std::alloc 源碼剖析, 4

```

#098 //本例兩個template參數完全沒派上用場
#099 template <bool threads, int inst>
#100 class __default_alloc_template {
#101 private:
#102     //實際上應使用 static const int x = N
#103     //取代 #094~#096 的 enum { x = N }, 但目前支援該性質的編譯器不多
#104
#105     static size_t ROUND_UP(size_t bytes) {
#106         return (((bytes) + __ALIGN-1) & ~(__ALIGN - 1)); → 若bytes為13，則(13+7) & ~7
#107     }
#108     union obj { //type definition
#109         union obj* free_list_link;
#110     }; //改用 struct 亦可
#111
#112 private:
#113     static obj* volatile free_list[__NFREELISTS];
#114
#115     static size_t FREELIST_INDEX(size_t bytes) {
#116         return (((bytes) + __ALIGN-1)/__ALIGN - 1); → 若bytes為8，則(8+7)/8-1=0
#117     }
#118
#119     // Returns an object of size n, and optionally adds to size n free list.
#120     static void *refill(size_t n);
#121
#122     // Allocates a chunk for nobjs of size "size". nobjs may be reduced
#123     // if it is inconvenient to allocate the requested number.
#124     static char* chunk_alloc(size_t size, int &nobjs);
#125
#126     // Chunk allocation state.
#127     static char* start_free; //指向'pool'的頭
#128     static char* end_free; //指向'pool'的尾
#129     static size_t heap_size; //分配累計量
#130

```

若bytes為13，則(13+7) & ~7
即 10100 & 11000，得 10000即 16

#090 //---
#091 //第二級分配器
#092 //---
#094 enum { __ALIGN = 8}; //小區塊的上調邊界
#095 enum { __MAX_BYTES = 128}; //小區塊的上限
#096 enum { __NFREELISTS = __MAX_BYTES / __ALIGN}; //free-list個數

若bytes為8，則(8+7)/8-1=0
若bytes為16，則(16+7)/8-1=1
若bytes為20，則(20+7)/8-1=2
若bytes為24，則(24+7)/8-1=2

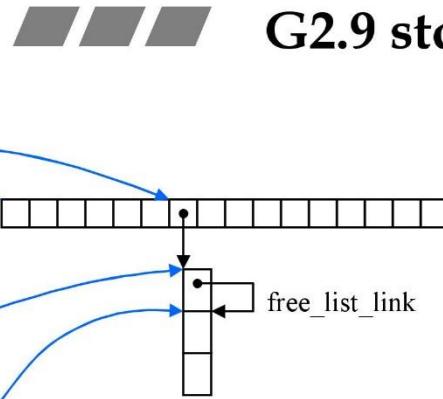
G2.9 std::alloc 源碼剖析, 5

```

#132 public:
#133
#134     static void* allocate(size_t n) //n must be > 0
#135     {
#136         obj* volatile *my_free_list; //obj**
#137         obj* result;
#138
#139         if (n > (size_t) __MAX_BYTES) { //改用第一級
#140             return (malloc_alloc::allocate(n));
#141         }
#142
#143         my_free_list = free_list + FREELIST_INDEX(n);
#144         result = *my_free_list;
#145         if (result == 0) { //list為空
#146             void* r = refill(ROUND_UP(n)); //挹注之
#147             return r;
#148         } //若往下進行表示list內已有可用區塊
#149         *my_free_list =
#150             result->free_list_link;
#151         return (result);
#152     }

```

refill() 會充填 free list 並
返回一個（其實就是第
一個）區塊的起始地址



如果這 p 並非當初從 **alloc** 取得，仍可併入 **alloc** 內 (這不好).
如果 p 所指大小不是 8 倍數，甚至會帶來災難

```

#154     static void deallocate(void* p, size_t n) //p 不為 0
#155     {
#156         obj* q = (obj*)p;
#157         obj* volatile *my_free_list; //obj**
#158
#159         if (n > (size_t) __MAX_BYTES) {
#160             malloc_alloc::deallocate(p, n); //改用第一級
#161             return;
#162         }
#163         my_free_list = free_list + FREELIST_INDEX(n);
#164         q->free_list_link = *my_free_list;
#165         *my_free_list = q;
#166     }
#167
#168     static void* reallocate(void* p, size_t old_sz,
#169                             size_t new_sz); //此處略列
#170 }

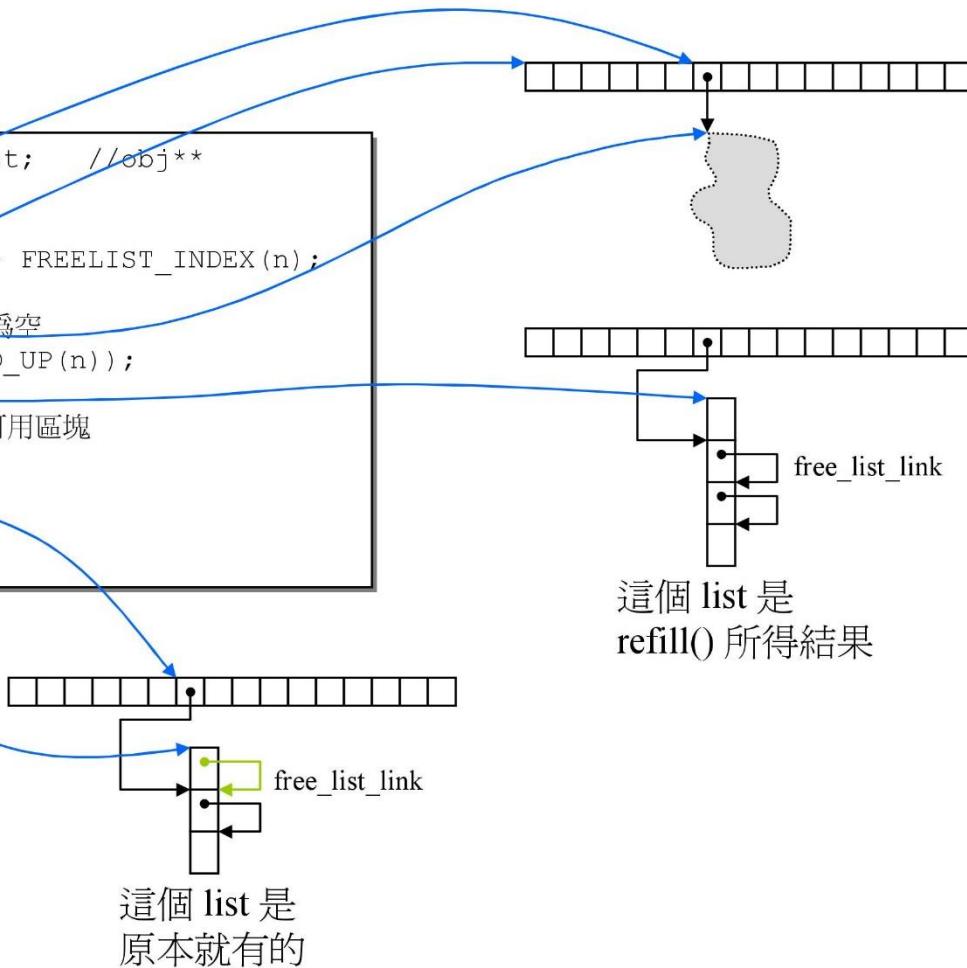
```

没有free



G2.9 std::alloc 源碼剖析, 5

```
#136     obj* volatile *my_free_list;    //obj**
#137     obj* result;
...
#143     my_free_list = free_list + FREELIST_INDEX(n);
#144     result = *my_free_list;
#145     if (result == 0) { //list為空
#146         void* r = refill(ROUND_UP(n));
#147         return r;
#148     } //若往下進行表示list內已有可用區塊
#149     *my_free_list =
#150         result->free_list_link;
#151     return (result);
#152 }
```

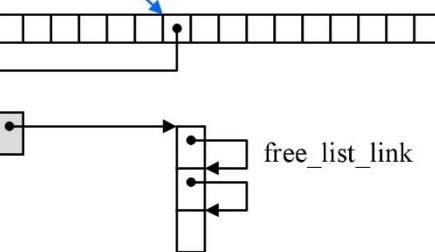
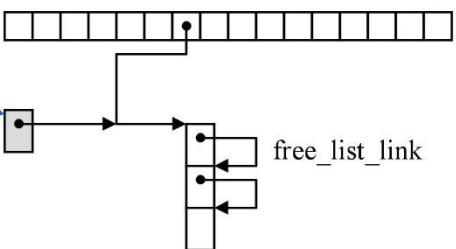
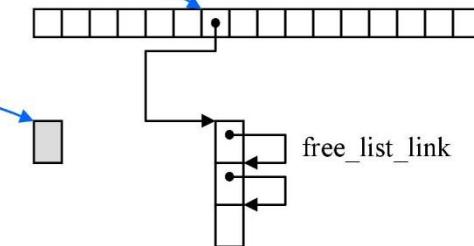


G2.9 std::alloc 源碼剖析, 5

```
#154     static void deallocate(void* p, size_t n) //p 不為 0
#155     {
#156         obj* q = (obj*)p;
#157         obj* volatile *my_free_list; //obj**
...
#163         my_free_list = free_list + FREELIST_INDEX(n);
```

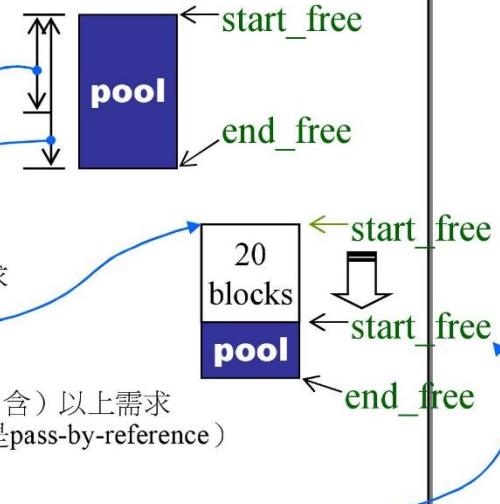
```
#164     q->free_list_link = *my_free_list;
```

```
#165     *my_free_list = q;
#166 }
```



G2.9 std::alloc 源碼剖析, 6

```
#172 // -----
#173 // We allocate memory in large chunks in order to avoid fragmenting the
#174 // malloc heap too much. We assume that size is properly aligned.
#175 // We hold the allocation lock.
#176 //-----
#177 template <bool threads, int inst>
#178 char*
#179 __default_alloc_template<threads, inst>::
#180 chunk_alloc(size_t size, int& nobjs)
#181 {
#182     char* result;
#183     size_t total_bytes = size * nobjs;
#184     size_t bytes_left = end_free - start_free;
#185
#186     if (bytes_left >= total_bytes) { //pool空間足以滿足20塊需求
#187         result = start_free;
#188         start_free += total_bytes; //調整(降低)pool水位
#189         return(result);
#190     } else if (bytes_left >= size) { //pool空間只足以滿足一塊(含)以上需求
#191         nobjs = bytes_left / size; //改變需求數(注意nobjs是pass-by-reference)
#192         total_bytes = size * nobjs; //改變需求總量(bytes)
#193         result = start_free;
#194         start_free += total_bytes; //調整(降低)pool水位
#195         return(result);
#196     } else { //pool空間不足以滿足一塊需求
#197         size_t bytes_to_get = 2 * total_bytes + ROUND_UP(heap_size >> 4); //現在打算從system free-store取這麼多來挹注
#198
#199         //首先嘗試將 pool 做充份運用
#200         if (bytes_left > 0) { //如果pool還有空間,
#201             obj* volatile *my_free_list = free_list + FREELIST_INDEX(bytes_left); //找出應移轉至第#號free-list(區塊儘可能大)
#202             free_list += FREELIST_INDEX(bytes_left);
#203             //將pool空間編入第#號free-list(肯定只成1區塊)
#204             ((obj*)start_free)->free_list_link = *my_free_list;
#205             *my_free_list = (obj*)start_free;
#206     }
```



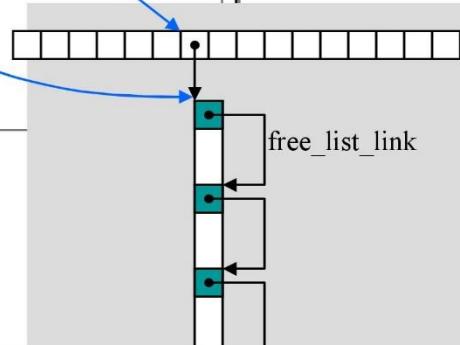
G2.9 std::alloc 源碼剖析, 7

```

#207     start_free = (char*)malloc(bytes_to_get); //從system free-store取這麼多,注入pool
#208     if (0 == start_free) {                      //失敗!以下試從free-list找區塊
#209         int i;
#210         obj* volatile *my_free_list, *p;
#211
#212         //Try to make do with what we have. That can't hurt.
#213         //We do not try smaller requests, since that tends
#214         //to result in disaster on multi-process machines.
#215         for (i = size; i <= _MAX_BYTES; i += _ALIGN) { //例88,96,104,112...
#216             my_free_list = free_list + FREELIST_INDEX(i);
#217             p = *my_free_list;
#218             if (0 != p) { //該free-list內有可用區塊,以下釋出一塊(only)給pool
#219                 *my_free_list = p -> free_list_link;
#220                 start_free = (char*)p;
#221                 end_free = start_free + i;
#222                 return(chunk_alloc(size, nobjs)); //遞歸再試一次
#223                 //此時的pool定夠供應至少一個區塊
#224                 //而任何殘餘零頭終將被編入適當free-list
#225             }
#226         }
#227         end_free = 0; //至此,表示 memory 已山窮水盡.
#228         //改用第一級,看看 oom-handler 能否盡點力
#229         start_free = (char*)malloc_alloc::allocate(bytes_to_get);
#230         //這會導致拋出異常,或導致memory不足的情況得到改善
#231     }
#232     //至此,表示已從system free-store成功取得很多 memory
#233     heap_size += bytes_to_get; //累計總分配量
#234     end_free = start_free + bytes_to_get; //挹注pool(調整尾端)
#235     return(chunk_alloc(size, nobjs)); //遞歸再試一次
#236 }
#237 }
```

→ p易讓人誤解,實為
obj** my_free_list;
obj* p;

i 從 size + _ALIGN
開始迭代更佳.



G2.9 std::alloc 源碼剖析, 8

```
#240 // Returns an object of size n, and optionally adds to size n free list.
#241 // We assume that n is properly aligned. We hold the allocation lock.
#242 //-----
#243 template <bool threads, int inst>
#244 void* __default_alloc_template<threads, inst>::
#245 refill(size_t n) //n 已調整至 8 的倍數
#246 {
#247     int nobjs = 20; //預設取 20 個區塊 (但不一定能夠)
#248     char* chunk = chunk_alloc(n,nobjs); //nobjs 是 pass-by-reference
#249     obj* volatile *my_free_list; //obj**
#250     obj* result;
#251     obj* current_obj;
#252     obj* next_obj;
#253     int i;
#254 }
```

```
#255     if (1 == nobjs) return(chunk); //實際得1，交給客戶
#256     //以下開始將所得區塊掛上 free-list
#257     my_free_list = free_list + FREELIST_INDEX(n);
#258     //在 chunk 內建立 free list
#259     result = (obj*)chunk;
#260     *my_free_list = next_obj = (obj*)((char*)chunk + n);
#261     for (i=1; ; ++i) {
#262         current_obj = next_obj;
#263         next_obj = (obj*)((char*)next_obj + n);
#264         if (nobjs-1 == i) { //最後一個
#265             current_obj->free_list_link = 0;
#266             break;
#267         } else {
#268             current_obj->free_list_link = next_obj;
#269         }
#270     }
#271     return(result);
#272 }
```

所謂切割就是把指針所指處轉型為obj，再取其next_obj指針繼續行事，一而再三...

G2.9 std::alloc 源碼剖析, 9

■■■ G2.9 std::alloc 觀念大整理

```
list<Foo> c; //假設 sizeof(Foo)+(4*2) <= 128  
c.push_back( Foo(1) );
```



~~c.push_back(new Foo(2)); //錯誤（元素型別不符）~~

Foo* p = new Foo(2);

c.push_back(*p);



delete p;

■■■ G2.9 std::alloc 批鬥大會

```
#136 obj* volatile *my_free_list;
```

```
#210 obj* volatile *my_free_list, *p; ➔
```

```
#208 if (0 == start_free) {
```

```
#218 if (0 != p) {
```

```
#255 if (1 == nobjs) return(chunk);
```

```
obj* *p1, *p2;
```

➔ ~~obj** p1, *p2;~~

➔ ~~obj **p1, *p2; ➔ obj** p1; obj* p2;~~

➔ ~~obj** p1, **p2;~~

➔ ~~obj** p1; obj** p2;~~

==不小心写成一个= 这样反过来写编译器就会报错 值得学习

```
#197 size_t bytes_to_get =
```

```
...  
#207 start_free = (char*)malloc(bytes_to_get);
```

```
typedef void (*new_handler)();  
new_handler set_new_handler(new_handler p) throw();
```

|||

```
typedef void (*H)();  
static H set_malloc_handler(H f);
```

```
#212 //Try to make do with what we have. That can't hurt.
```

```
#213 //We do not try smaller requests, since that tends
```

```
#214 //to result in disaster on multi-process machines.
```

deallocate() 完全沒有調用 free() or delete
(源於其設計上的先天缺陷)

没有变量记录给出去的指针

G4.9 pool allocator 運行觀察

//我要一個可累計總分配量和總釋放量的 operator new/delete.

//除非 user 直接使用 malloc/free,
//否則都避不開它們, 這就可以累計總量.
static long long countNew = 0;
static long long timesNew = 0;

//小心，這影響無遠弗屆

//它們不可被聲明於 namespace 內

```
inline void* operator new(size_t size) {  
    //cout << "jjhou global new(), \t" << size << "\t";  
    countNew += size;  
    ++timesNew;  
    return myAlloc( size );  
}
```

```
inline void* operator new[](size_t size) {  
    cout << "jjhou global new[](), \t" << size << "\t";  
    return myAlloc( size );  
}
```

... 繼右

```
void* myAlloc(size_t size)  
{ return malloc(size); }  
void myFree(void* ptr)  
{ return free(ptr); }
```

重載 ::operator new / ::operator delete

小心，這影響無遠弗屆

```
void* myAlloc(size_t size)  
{ return malloc(size); }  
void myFree(void* ptr)  
{ return free(ptr); }
```

... 接左

//天啊, 以下(1)(2)可並存並由(2)抓住流程 (但它對我這兒測試無用).

//若只存在 (1), 抓不住流程. 為什麼?

//在 class members 中必須二擇一 (任一均可), 否則編譯報錯.

```
inline void operator delete(void* ptr, size_t size) { //(1)  
    myFree( ptr );  
}
```

```
inline void operator delete(void* ptr) { //(2)  
    myFree( ptr );  
}
```

```
inline void operator delete[](void* ptr, size_t size) { //(1)  
    myFree( ptr );  
}
```

```
inline void operator delete[](void* ptr) { //(2)  
    myFree( ptr );  
}
```



G4.9 pool allocator 運行觀察

Q: 我能不能觀察到 malloc() 真正分配出去的總量 (含所有 overhead) ?

A: **不能**; 除非...



```
//C++/11 alias template
template <typename T>
using listPool = list<T, __gnu_cxx::__pool_alloc<T>>;
```

```
//reset countNew
countNew = 0;
timesNew = 0;

listPool<double> lst;
for (int i=0; i< 1000000; ++i)
    lst.push_back(i);
cout << "::countNew= " << ::countNew << endl;
        //16752832 (注意, node 都不帶 cookie)
cout << "::timesNew= " << ::timesNew << endl; //122
```



```
//reset countNew
countNew = 0;
timesNew = 0;

list<double> lst;
for (int i=0; i< 1000000; ++i)
    lst.push_back(i);
cout << "::countNew= " << ::countNew << endl;
        //16000000 (注意, node 都帶 cookie)
cout << "::timesNew= " << ::timesNew << endl; //1000000
```

```
list<double, __gnu_cxx::__pool_alloc<double>> lst;
```

■■■■ G2.9 std::alloc 移植至 C

std::alloc 整個由 static data members 和 static member functions 構成，整體只一個 class 而無分枝，因此移植至 C 很容易。需注意：

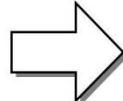
`chunk_alloc(size_t size, int& nobjs);`

pass by reference 移至 C 時或需改為 pass by pointer.

移植步驟

- 去除 templates
- 將 union obj 移至 _default_alloc_template class 外部獨立
- 將所有 static data 移為 global
- 將 __malloc_alloc 改名為 malloc_alloc, 將 __default_alloc 改名為 alloc
- 將 malloc_alloc 的所有 static functions 移為 global
- 將 alloc 的所有 static functions 移為 global
- 將 .cpp 改為 .c，將上述 pass by reference 改為 pass by pointer，再將：

```
union obj {
    union obj* free_list_link;
};
```



```
typedef union __obj {
    union __obj* free_list_link;
} obj;
```

或：

```
typedef struct __obj {
    struct __obj* free_list_link;
} obj;
```

內存管理

從平地到萬丈高樓

Memory Management 101

第一講 primitives

第二講 std::allocator

第三講 malloc/free

第四講 loki::allocator

第五講 other issues



侯捷

胸中自有丘壑

觸類旁通

VC6 內存分配

```
ExitProcess(code)
    _initterm(,,) //do terminators
        __endstdio(void)
    _initterm(,,) //do pre-terminators
doexit(code, 0, 0)
⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ __crtGetEnvironmentStringsA()
③ GetCommandLineA()
    __sbh_alloc_new_group(...)
    __sbh_alloc_new_region()
    sbh_alloc_block(...)
    heap_alloc_base(...)
    heap_alloc_dbg(...)
    nh_malloc_dbg(...)
    malloc_dbg(...)
② __ioinit() // initialize lowio
    __sbh_heap_init()
① heap_init(...)
mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()
```

SBH: Small Block Heap

```
if (size <= __sbh_threshold) { //3F8, i.e. 1016
    pvReturn = __sbh_alloc_block(size);
    if (pvReturn) return pvReturn;
}
if (size == 0) size = 1;
size = (size + ...) & ~(...);
return HeapAlloc(crtheap, 0, size);
```

VC10 內存分配

```

    graph TD
        A[_heap_init(...)] --> B[_initterm(..)]
        A --> C[mainCRTStartup()]
        C --> D[_initterm(..)]
        D --> E[_initstdio()]
        E --> F[_initterm(..)]
        F --> G[_doexit(code, 0, 0)]
        G --> H[_exit(code)]
        H --> I[main()]
        I --> J[_initterm(..)]
        J --> K[_initstdio()]
        K --> L[_initterm(..)]
        L --> M[_cinit()]
        M --> N[_setenvp()]
        N --> O[_setargv()]
        O --> P[_crtGetEnvironmentStringsA()]
        P --> Q[GetCommandLineA()]
        Q --> R[_heap_alloc_base(size_t size)]
        R --> S[_nh_malloc_dbg(...)]
        S --> T[_malloc_dbg(...)]
        T --> U[_iointi()]
        U --> V[_heap_alloc_base(...)]
        V --> W[_nh_malloc_dbg(...)]
        W --> X[_malloc_dbg(...)]
        X --> Y[_iointi()]
        Y --> Z[_heap_alloc_base(...)]
        Z --> AA[_forceinline void * __cdecl _heap_alloc (size_t size)]
        AA --> BB[_if (_crtheap == 0) { ... }]
        BB --> CC[_return HeapAlloc(_crtheap, 0, size ? size : 1);]
    
```

Call graph illustrating the flow of control from the CRT startup code through various initialization routines to the final memory allocation code in `_heap_alloc_base()`.

The graph shows the following sequence of calls:

- `_heap_init()` calls `_initterm()`, `mainCRTStartup()`, and `_iointi()`.
- `mainCRTStartup()` calls `_initterm()`, `_initstdio()`, `_iointi()`, and `_cinit()`.
- `_iointi()` calls `_heap_alloc_base()`.
- `_heap_alloc_base()` calls `_nh_malloc_dbg()`.
- `_nh_malloc_dbg()` calls `_malloc_dbg()`.
- `_malloc_dbg()` calls `_iointi()`.
- `_iointi()` calls `_heap_alloc_base()`.
- `_heap_alloc_base()` calls `_nh_malloc_dbg()`.
- `_nh_malloc_dbg()` calls `_malloc_dbg()`.
- `_malloc_dbg()` calls `_iointi()`.
- `_iointi()` calls `_heap_alloc_base()`.
- `_heap_alloc_base()` calls `_forceinline void * __cdecl _heap_alloc (size_t size)`.
- `_forceinline void * __cdecl _heap_alloc (size_t size)` checks if `_crtheap == 0`.
- If `_crtheap == 0`, it performs error handling: `_FF_MSGBANNER()`, `_NMSG_WRITE(_RT_CRT_NOTINIT)`, and `_crtExitProcess(255)`.
- Finally, it returns `HeapAlloc(_crtheap, 0, size ? size : 1)`.

所有內存管理
機制都繼續存
在，只不過，
深埋到 O.S. 層
面去了 (詳見
[Windows Heap
課程](#))

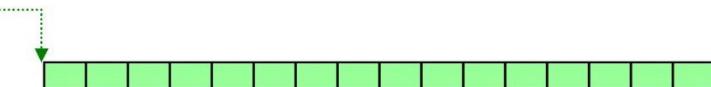
SBH 之始 – `_heap_init()` 和 `_sbh_heap_init()`

```
int __cdecl _heap_init (
    int mtflag
)
{
    // Initialize the "big-block" heap first.
    if ( __crtheap = HeapCreate( mtflag ? 0 : HEAP_NO_SERIALIZE,
                                BYTES_PER_PAGE, 0 ) == NULL )
        return 0; 4096

    // Initialize the small-block heap
    if ( __sbh_heap_init() == 0 )
    {
        ←
        HeapDestroy( __crtheap );
        return 0;
    }
    return 1;
}
不論 big-block heap or
small-block heap ,
只要失敗就 return 0 ,
別玩了。
```

heapinit.c

CRT 會先為自己建立一個 `_crtheap`，然後從中配置 SBH 所需的 headers, regions 做為管理之用。App. 動態配置時若 size > threshold 就以 `HeapAlloc()` 從 `_crtheap` 取。若 size <= threshold 就從 SBH 取 (實際區塊來自 `VirtualAlloc()`)



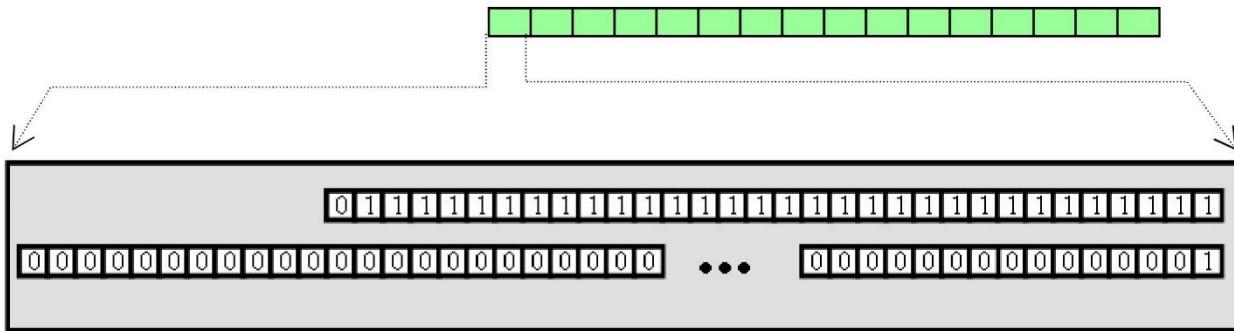
```
int __cdecl __sbh_heap_init (void)
{
    if ( !(__sbh_pHeaderList =
           HeapAlloc( __crtheap, 0, 16 * sizeof(HEADER) ) ) )
        return FALSE;

    __sbh_pHeaderScan = __sbh_pHeaderList;
    __sbh_pHeaderDefer = NULL;
    __sbh_cntHeaderList = 0;
__sbh_sizeHeaderList = 16;

    return TRUE;
}
```

sbheap.c

■■■ SBH 之始 – _heap_init() 和 _sbh_heap_init()



```
typedef unsigned int BITVEC;  
  
typedef struct tagHeader  
{  
    BITVEC      bitvEntryHi;  
    BITVEC      bitvEntryLo;  
    BITVEC      bitvCommit;  
    void *      pHeapData;  
    struct tagRegion * pRegion;  
}  
HEADER, *PHEADER;
```

VC6 內存分配

```

        ExitProcess(code)
        _initterm(,,) //do terminators
        __endstdio(void)
        _initterm(,,) //do pre-terminators
        doexit(code, 0, 0)
⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
    __initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ __crtGetEnvironmentStringsA()
③ GetCommandLineA()
    {
        __sbh_alloc_new_group(...)
        __sbh_alloc_new_region()
        sbh_alloc_block(...)

        __heap_alloc_base(...)
        __heap_alloc_dbg(...)
        __nh_malloc_dbg(...)
        malloc_dbg(...)

        ② __ioinit() // initialize lowio .....
        __sbh_heap_init()

        ① __heap_init(...)

        mainCRTStartup()
        KERNEL32! bff8b6e6()
        KERNEL32! bff8b598()
        KERNEL32! bff89f5b()
    }

```

```

/* Memory block identification */
#define _FREE_BLOCK      0
#define _NORMAL_BLOCK    1
#define _CRT_BLOCK        2
#define _IGNORE_BLOCK     3
#define _CLIENT_BLOCK     4
#define _MAX_BLOCKS       5

#ifndef _DEBUG
#define _malloc_crt  malloc
...
#else /* _DEBUG */
#define _THISFILE _FILE_
#define _malloc_crt(s) _malloc_dbg
(s, _CRT_BLOCK,
 _THISFILE, _LINE_)
...
#endif /* _DEBUG */

```

```

void __cdecl __ioinit (void)
{
    ...
    if ( (pio = _malloc_crt( IOINFO_ARRAY_ELTS * sizeof(ioinfo) ))
        == NULL )
    ...
}

```

```

typedef struct {
    long osfhnd;
    char osfile;
    char pipech;
} ioinfo;

```

32

8

100

VC6 內存分配

```

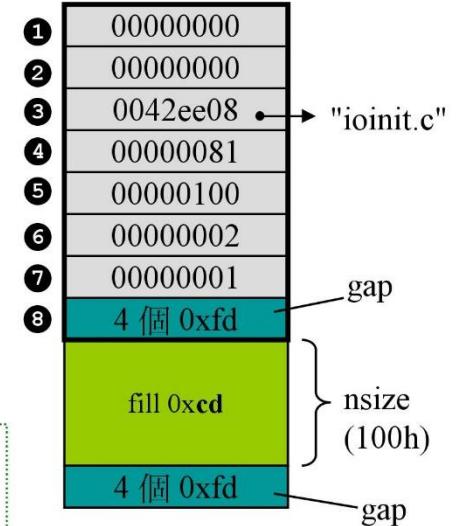
ExitProcess(code)
    _initterm(,,) //do terminators
        __endstdio(void)
    _initterm(,,) //do pre-terminators
doexit(code, 0, 0)
⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ __crtGetEnvironmentStringsA()
③ GetCommandLineA()
    [sbh_alloc_new_group()]
    [sbh_alloc_new_region()]
    sbh_alloc_block(...)
    _heap_alloc_base(...)
        heap_alloc_dbg(...)
            nh_malloc_dbg(...)
            malloc_dbg(...)
② _ioinit() // initialize lowio
    sbh_heap_init()
① _heap_init(...)
mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()

```

```

#define nNoMansLandSize 4
typedef struct _CrtMemBlockHeader
{
    struct _CrtMemBlockHeader * pBlockHeaderNext;
    struct _CrtMemBlockHeader * pBlockHeaderPrev;
    char * szFileName;
    int nLine;
    size_t nDataSize;
    int nBlockUse;
    long lRequest;
    unsigned char gap[nNoMansLandSize];
    /* followed by:
     * unsigned char data[nDataSize];
     * unsigned char anotherGap[nNoMansLandSize];
     */
} _CrtMemBlockHeader;

```



```

...
blockSize = sizeof(_CrtMemBlockHeader) + nSize + nNoMansLandSize;
...
pHead = (_CrtMemBlockHeader *) _heap_alloc_base(blockSize);
... (續下頁)

```

VC6 內存分配

```
ExitProcess(code)
_initterm(,,) //do terminators
    __endstdio(void)
    _initterm(,,) //do pre-terminators
doexit(code, 0, 0)

⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
        _initterm(,,) //do initializations
    ⑦ _cinit() // do C data initialize
    ⑥ _setenvp()
    ⑤ _setargv()
    ④ __crtGetEnvironmentStringsA()
    ③ GetCommandLineA()
        __sbh_alloc_new_group(...)
        __sbh_alloc_new_region()
        __sbh_alloc_block(...)

        _heap_alloc_base(...)
        ... heap_alloc_dbg(...)

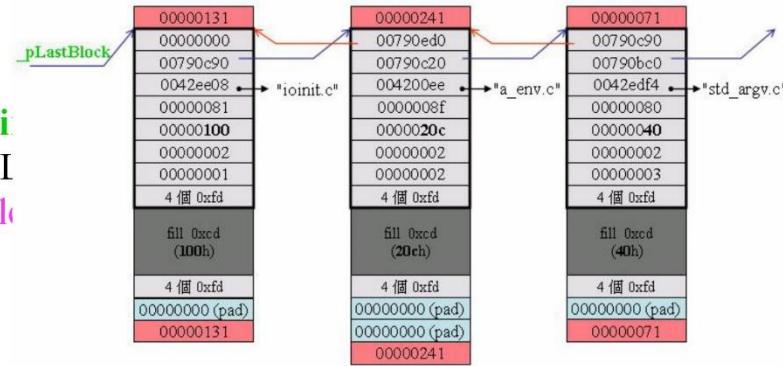
        __nh_malloc_dbg(...)
        _malloc_dbg(...)

    ② _ioinit() // initialize lowio
        __sbh_heap_init()

    ① _heap_init(...)

mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()
```

```
... (承上頁)  
if (_pFirstBlock)  
    _pFirstBlock->pBlockHeaderPrev = pHead;  
else  
    _pLastBlock = pHead;  
  
pHead->pBlockHeaderNext = _pFi  
pHead->pBlockHeaderPrev = NULI  
pHead->szFileName = (char *)szFil  
pHead->nLine = nLine;  
pHead->nDataSize = nSize;  
pHead->nBlockUse = nBlockUse;  
pHead->lRequest = lRequest;  
  
/* link blocks together */  
pFirstBlock = pHead;
```



```
static unsigned char _bNoMansLandFill = 0xFD;
static unsigned char _bDeadLandFill = 0xDD;
static unsigned char _bCleanLandFill = 0xCD;

static _CrtMemBlockHeader * _pFirstBlock;
static _CrtMemBlockHeader * _pLastBlock;
```

```

/* fill in gap before and after real block */
memset((void *)pHead->gap, _bNoMansLandFill, nNoMansLandSize);
memset((void *)(pbData(pHead) + nSize), _bNoMansLandFill, nNoMansLandSize);
/* fill data with silly value (but non-zero) */
memset((void *)pbData(pHead), _bCleanLandFill, nSize);
return (void *)pbData(pHead);

```

VC6 內存分配

```
ExitProcess(code)
    _initterm(,,) //do terminators
        __endstdio(void)
    _initterm(,,) //do pre-terminators
    doexit(code, 0, 0)
⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ __crtGetEnvironmentStringsA()
③ GetCommandLineA()
    __sbh_alloc_new_group(...)
    __sbh_alloc_new_region()
    sbh_alloc_block(...)
    __heap_alloc_base(...)

    __heap_alloc_dbg(...)
    __nh_malloc_dbg(...)
    malloc_dbg(...)

② __ioinit() // initialize lowio
    __sbh_heap_init()

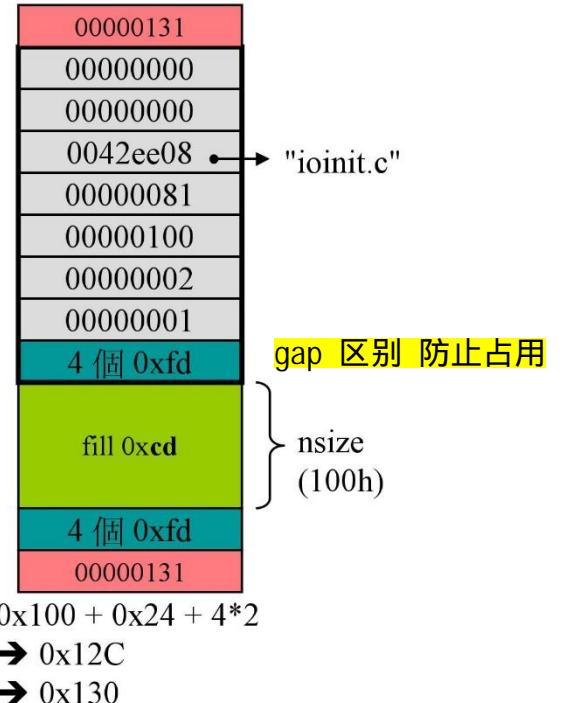
① __heap_init(...)

mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()
```

1024-8 8为cookie大小

```
if (size <= __sbh_threshold) { //3F8, i.e. 1016
    pvReturn = __sbh_alloc_block(size);
    if (pvReturn) return pvReturn;
}
if (size == 0) size = 1;
size = (size + ...) & ~(...);
return HeapAlloc(__crtheap, 0, size);
```

VC6 內存分配



```
ExitProcess(code)
    _initterm(,,) //do terminators
        _endstdio(void)
    _initterm(,,) //do pre-terminators
doexit(code, 0, 0)

⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        _initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ __crtGetEnvironmentStringsA()
③ GetCommandLineA()
    sbh_alloc_new_group(...)
    sbh_alloc_new_region()
    sbh_alloc_block(...)

    heap_alloc_base(...)
    heap_alloc_dbg(...)
    nh_malloc_dbg(...)
    malloc_dbg(...)

    计算要分
② _ioinit() // initialize lowio
    sbh_heap_init()

① heap_init(...)

mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()
```

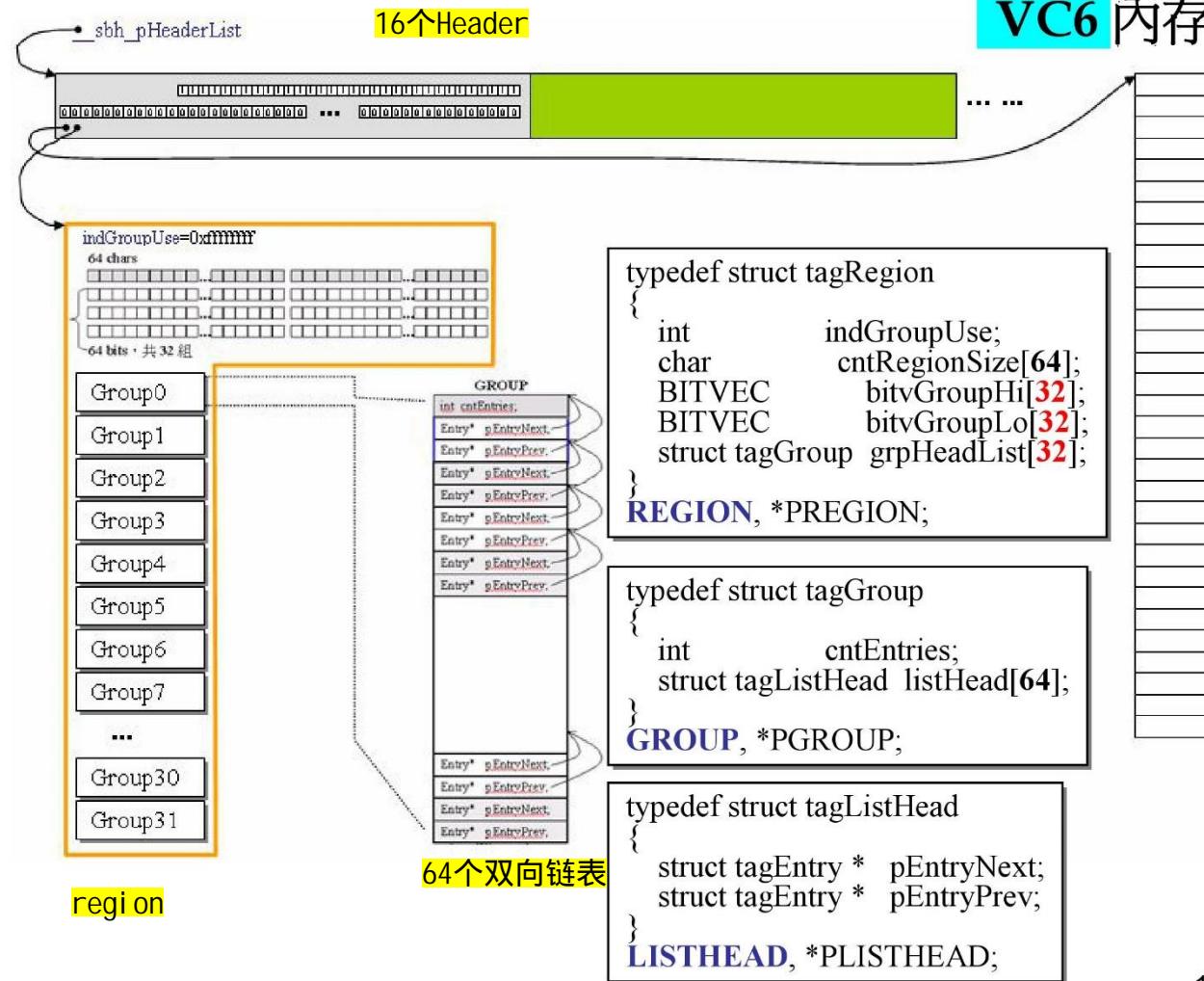
计算要分配的内存的大小

```
// add 8 bytes entry overhead and round up to next para size  
sizeEntry = (intSize + 2 * sizeof(int) + 16  
            + (BYTES_PER_PARA - 1))  
            & ~(BYTES_PER_PARA - 1);
```

调到16的倍数 类似之前的RoundUp

VC6 內存分配

虛擬地址空間



ExitProcess(code)
 _initterm(,,) //do terminators
 _endstdio(void)
 _initterm(,,) //do pre-terminators
 doexit(code, 0, 0)

⑨ exit(code)

⑧ main()
 _initterm(,,) //do C++ initializations
 _initstdio(void)
 _initterm(,,) //do initializations

⑦ _cinit() // do C data initialize

⑥ _setenvp()

⑤ _setargv()

④ _crtGetEnvironmentStringsA()

③ GetCommandLineA()
 sbh_alloc new_group(...)
 sbh_alloc new_region()
 sbh_alloc block(...)

内存管理

heap_alloc_base(...)
 heap_alloc_dbg(...)
 nh_malloc_dbg(...)
 malloc_dbg(...)

② _ioinit() // initialize lowio
 sbh_heap_init()

① heap_init(...)

mainCRTStartup()
 KERNEL32! bff8b6e6()
 KERNEL32! bff8b598()
 KERNEL32! bff89f5b()

VC6 內存分配

```

ExitProcess(code)
    _initterm(,,) //do terminators
        __endstdio(void)
    _initterm(,,) //do pre-terminators
doexit(code, 0, 0)
⑨ exit(code)
⑧ main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
    _initterm(,,) //do initializations
⑦ _cinit() // do C data initialize
⑥ _setenvp()
⑤ _setargv()
④ _crtGetEnvironmentStringsA()
③ GetCommandLineA()
    __sbh_alloc_new_group(...)
    __sbh_alloc_new_region()
    sbh_alloc_block(...)

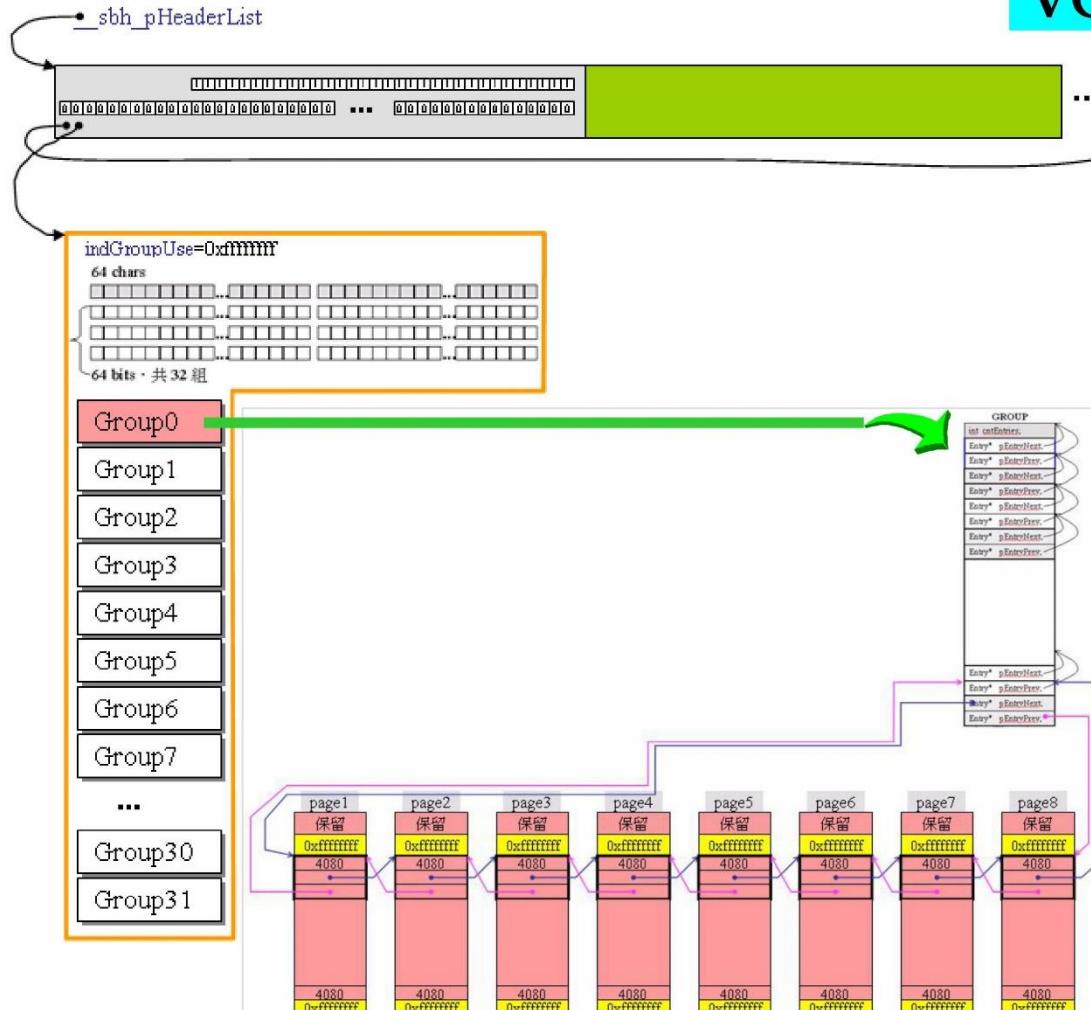
    _heap_alloc_base(...)
    _heap_alloc_dbg(...)
    _nh_malloc_dbg(...)
    malloc_dbg(...)

② _ioinit() // initialize lowio
    __sbh_heap_init()

① _heap_init(...)

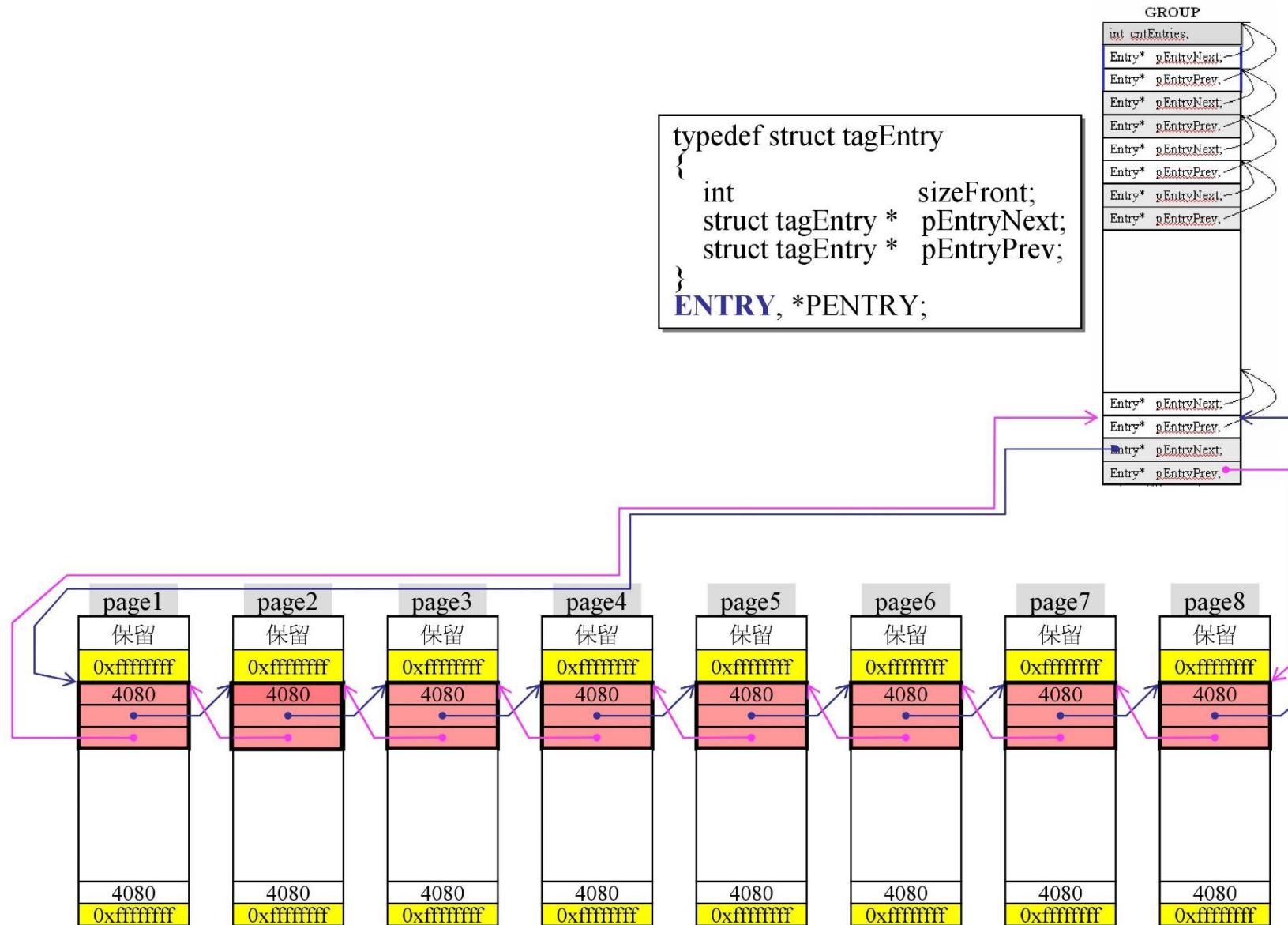
mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()

```

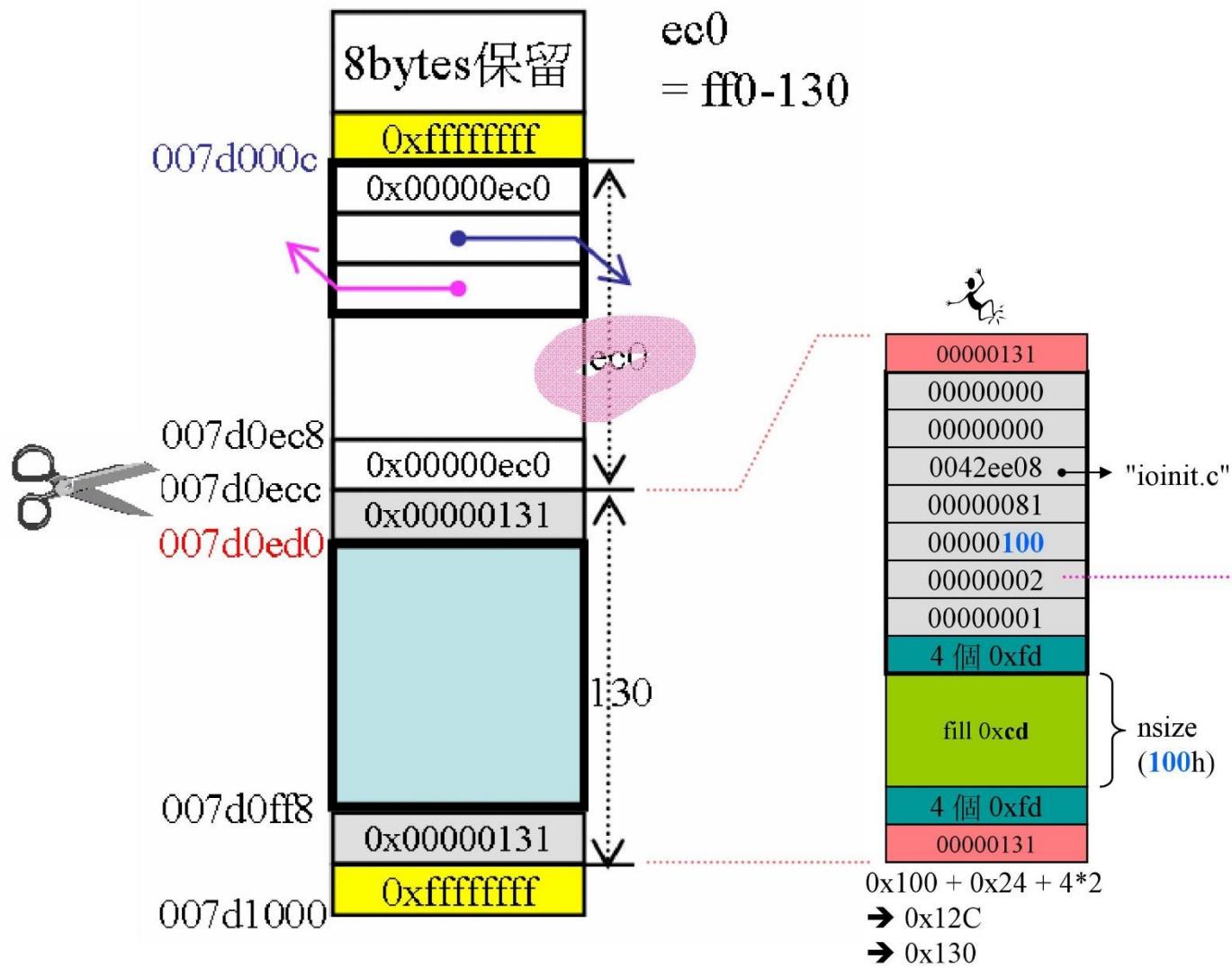


分成32块
由group管理
group就是双向链表

VC6 內存分配

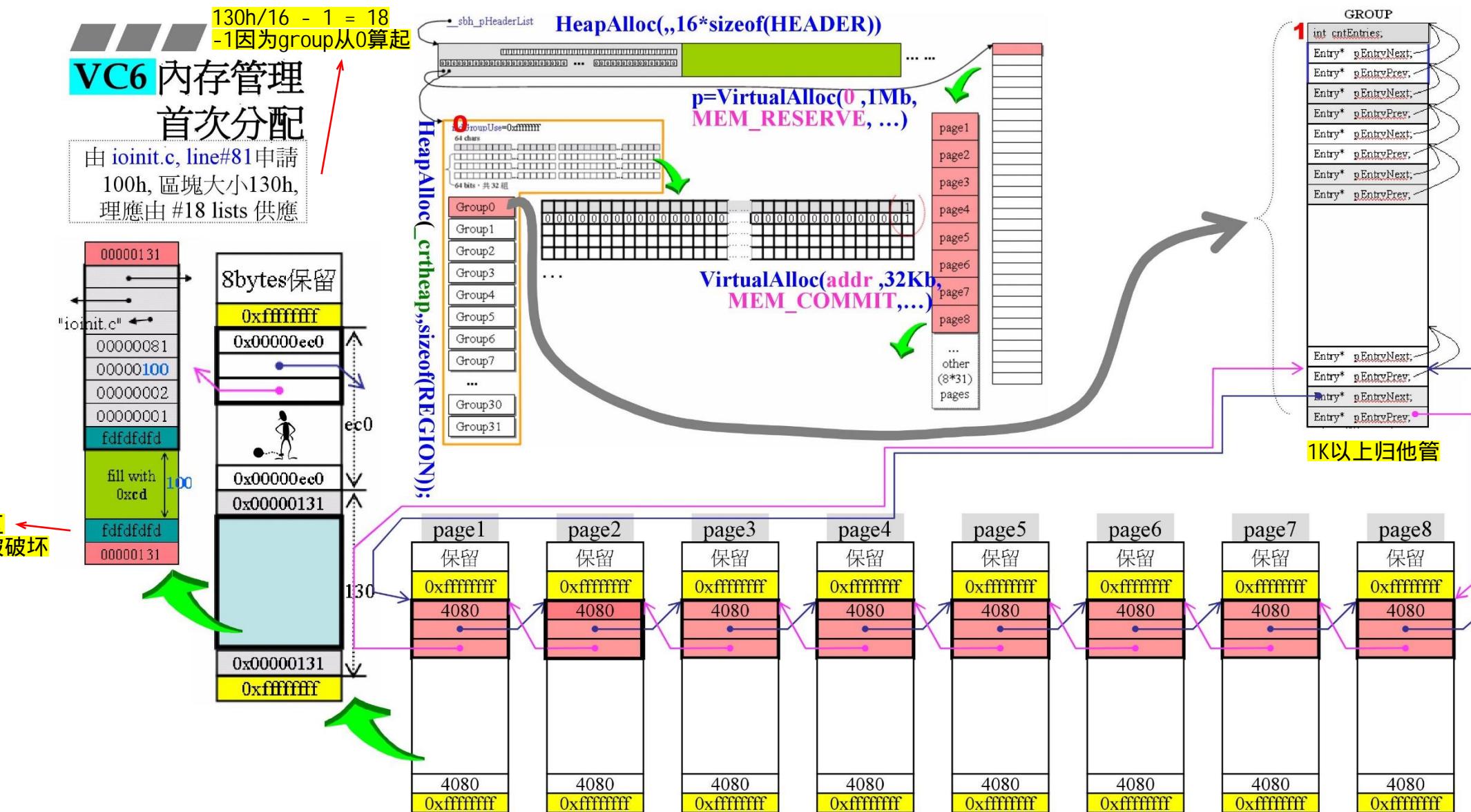


VC6 內存分配

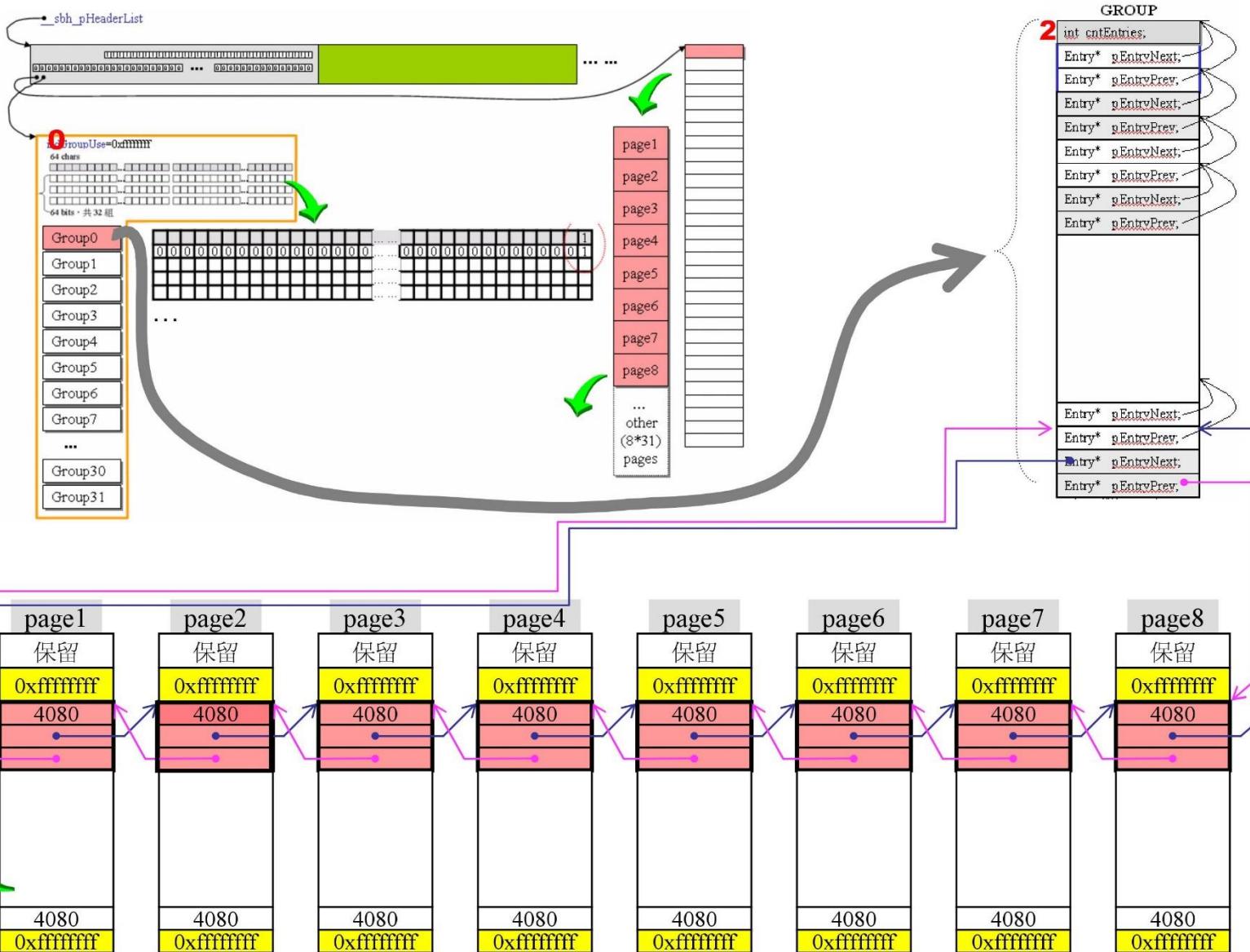
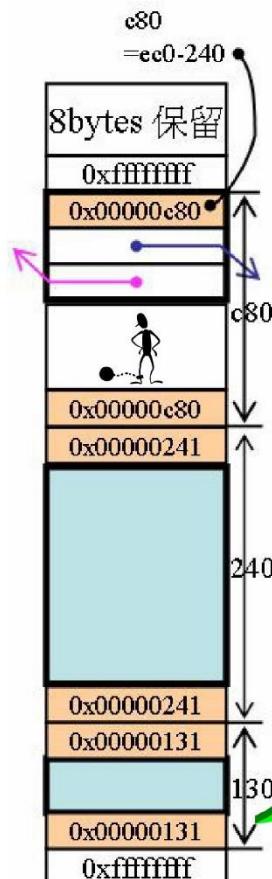


VC6 内存管理 首次分配

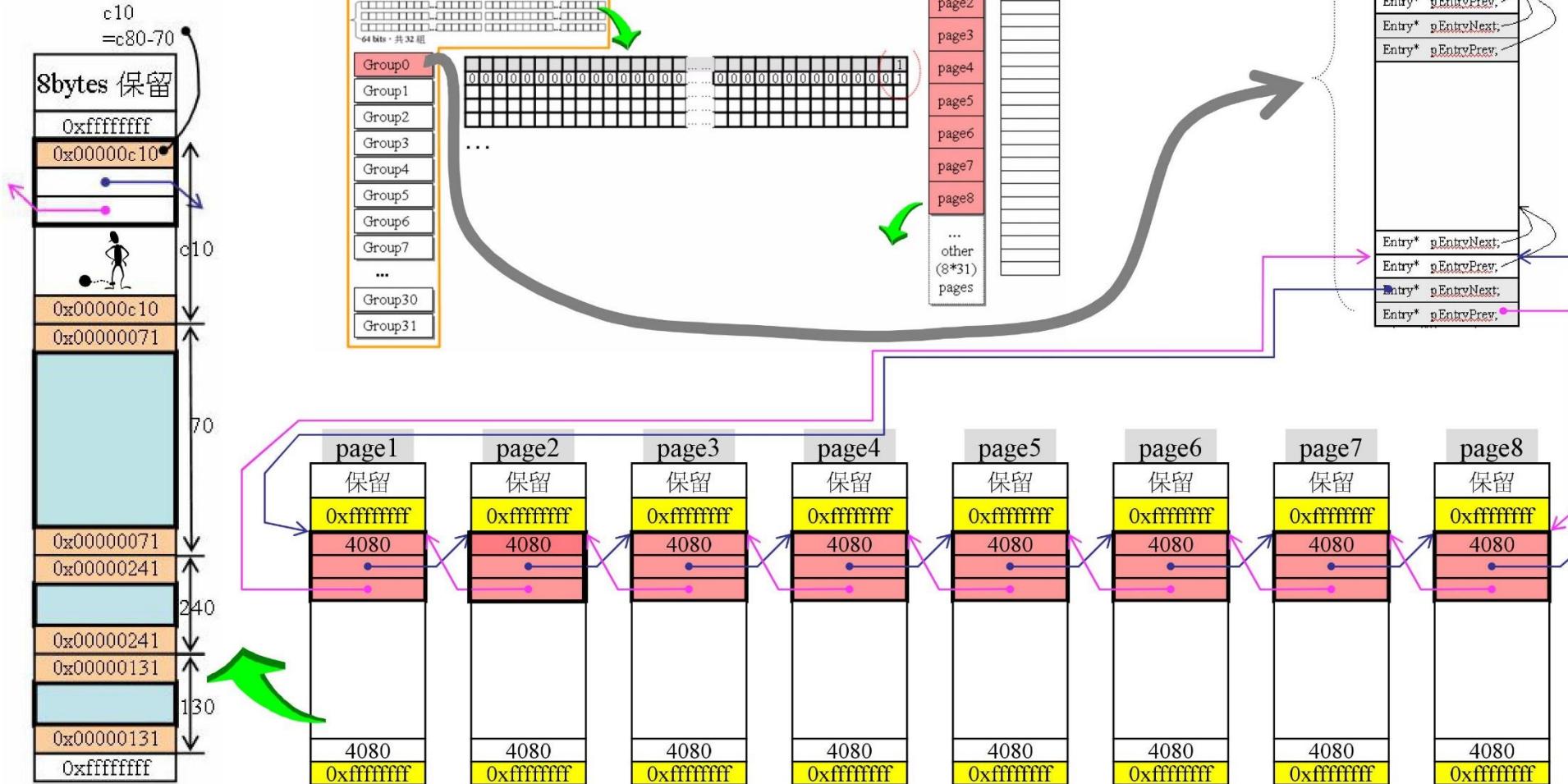
由 ioinit.c, line#81 申请
100h, 区块大小130h,
理应由 #18 lists 供應



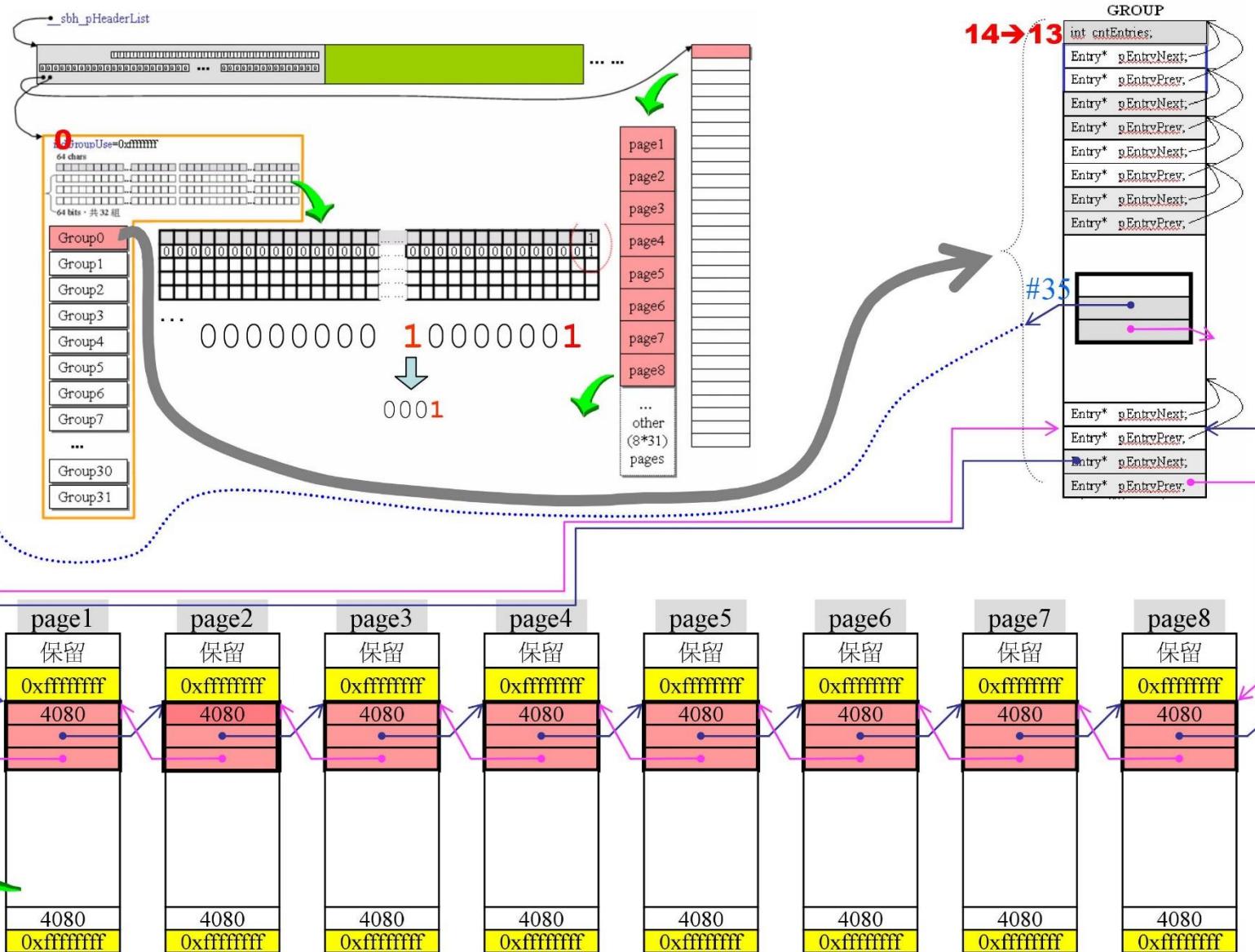
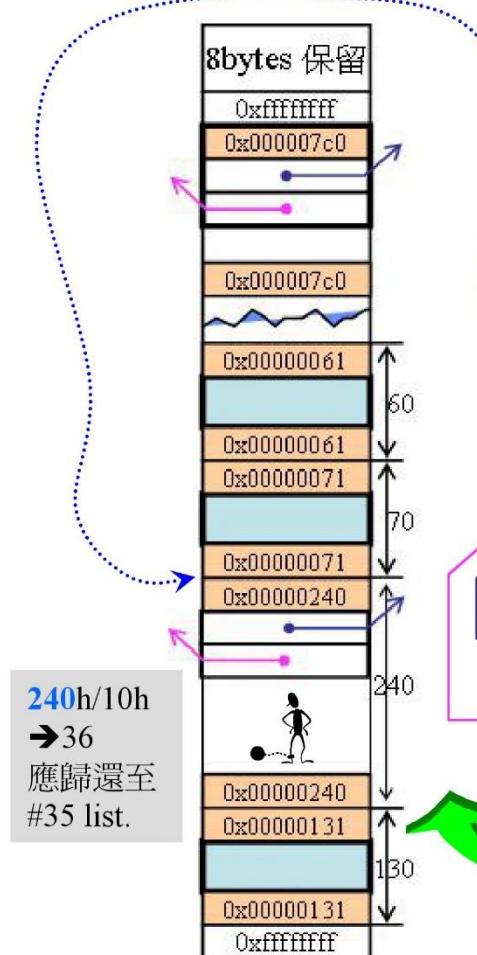
VC6 内存管理 第2次, 分配



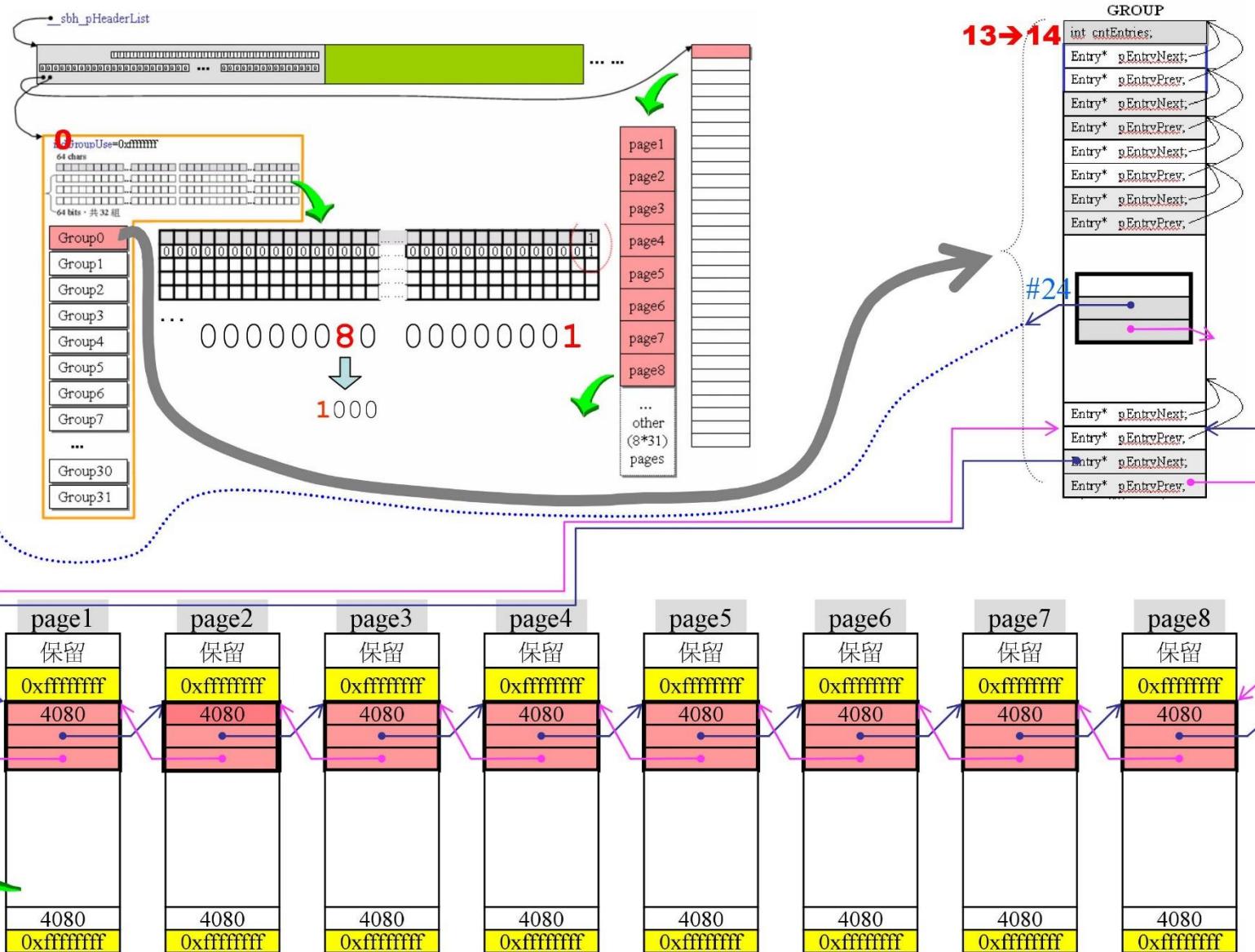
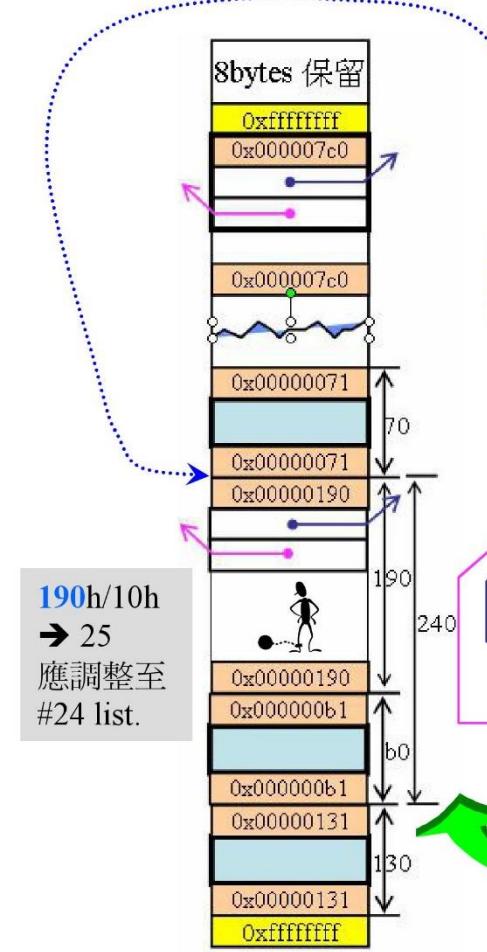
VC6 内存管理 第3次, 分配



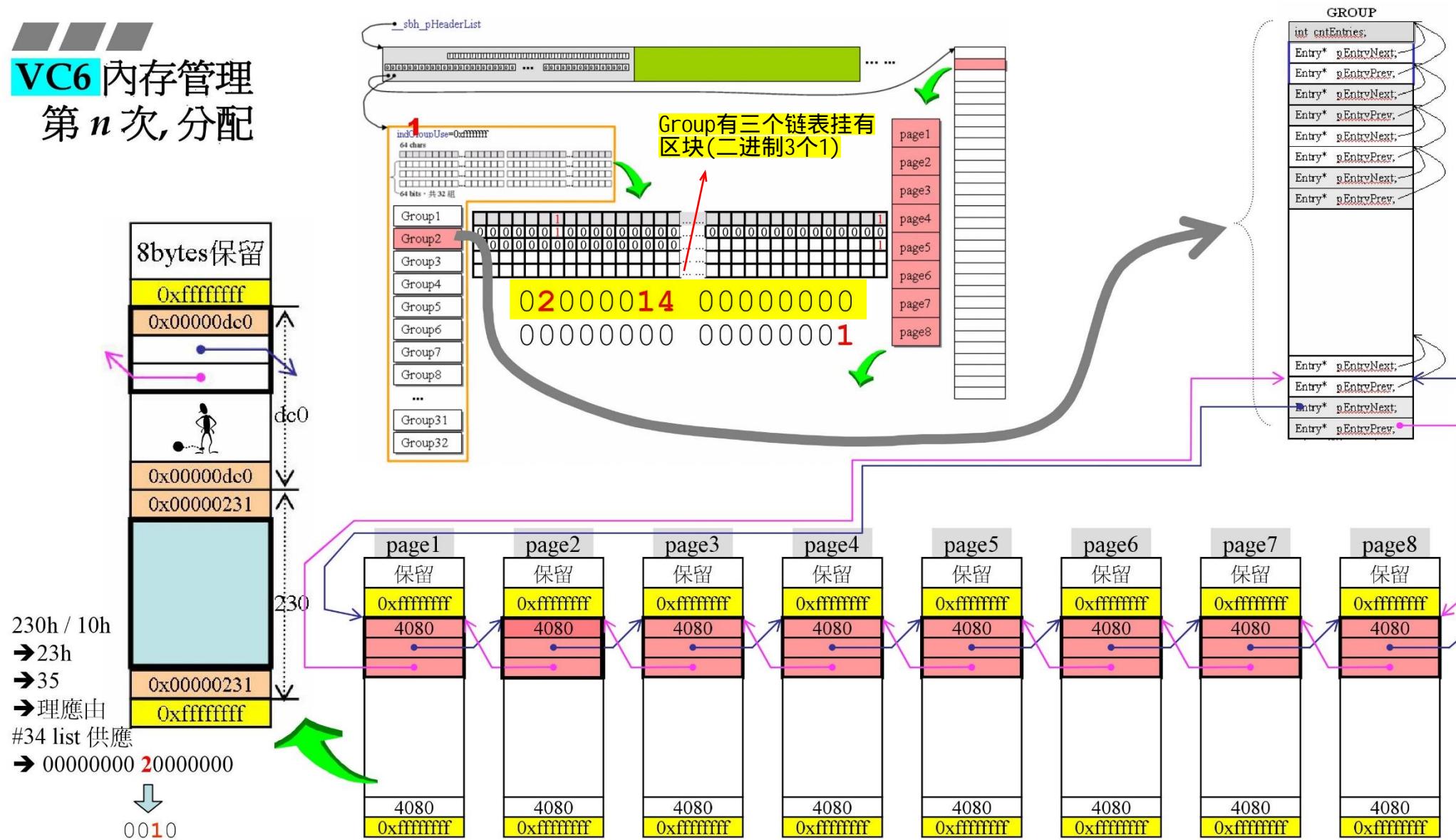
VC6 內存管理 第15次，釋還



VC6 內存管理 第16次, 分配



VC6 內存管理 第 n 次, 分配

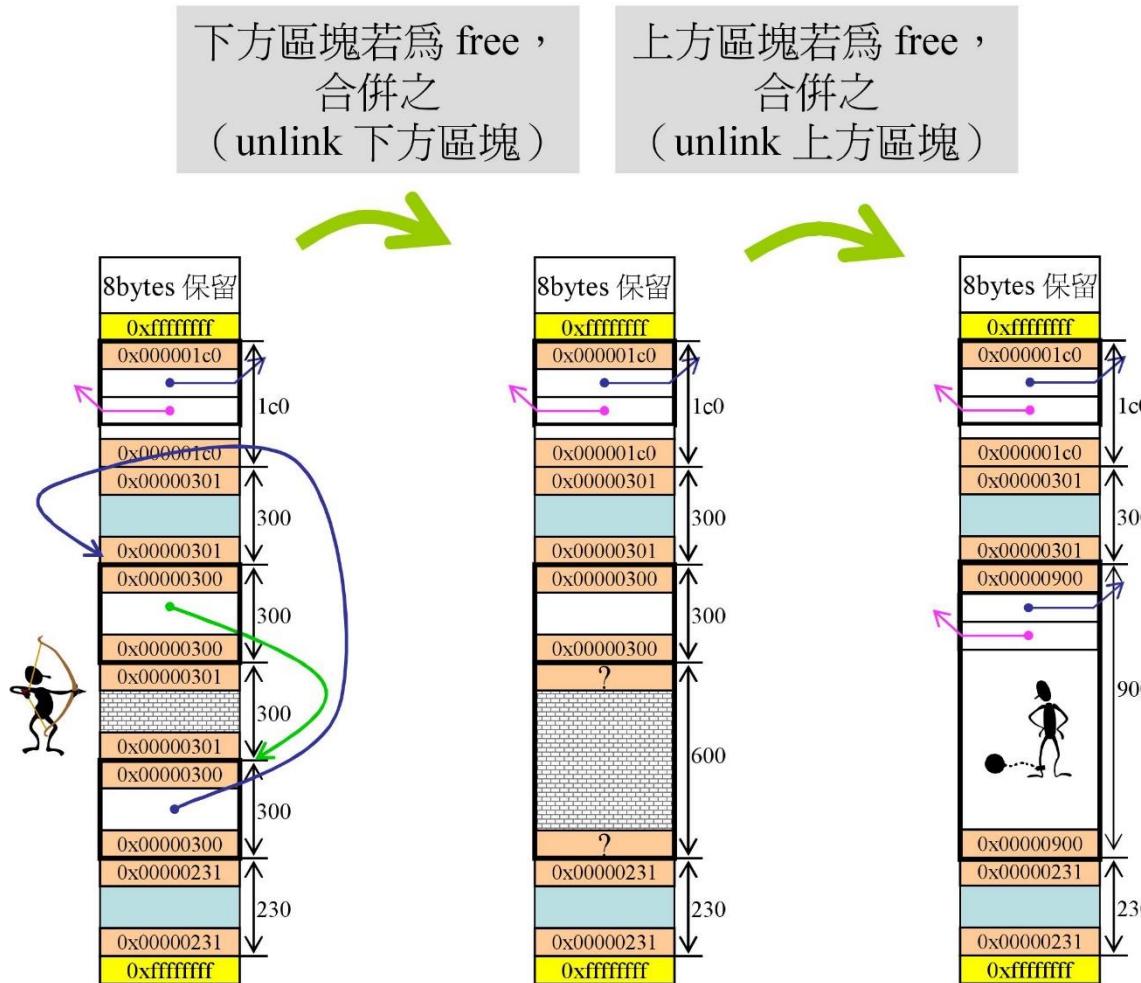




VC6 內存管理

區塊之合併

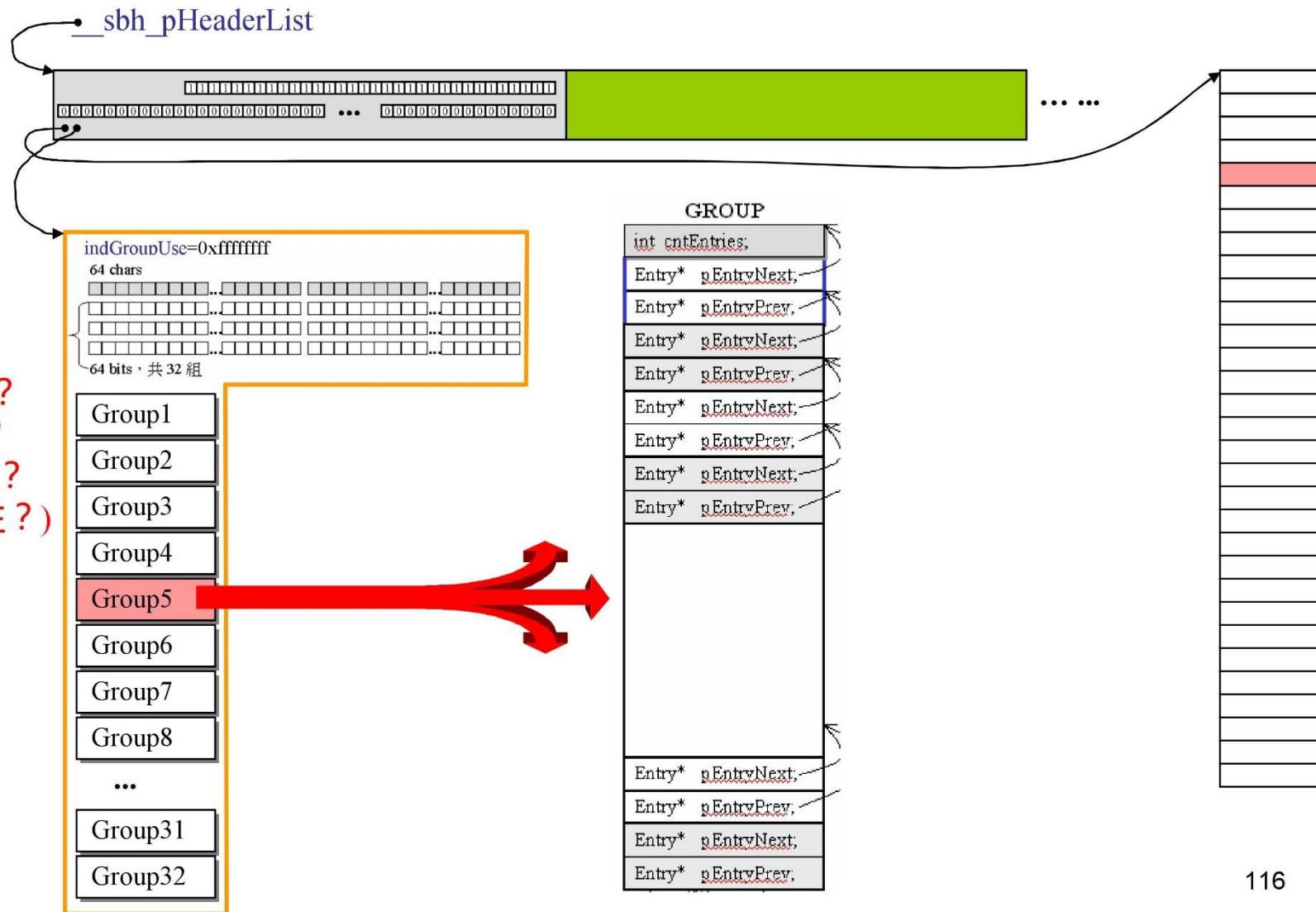
上 cookie
與
下 cookie
的作
用
往上合
并



VC6 內存管理 free(p)

p 花落誰家？

落在哪個 Header 內？
落在哪個 Group 內？
落在哪個 free-list 內？
(被哪個 free-list 鏈住？)

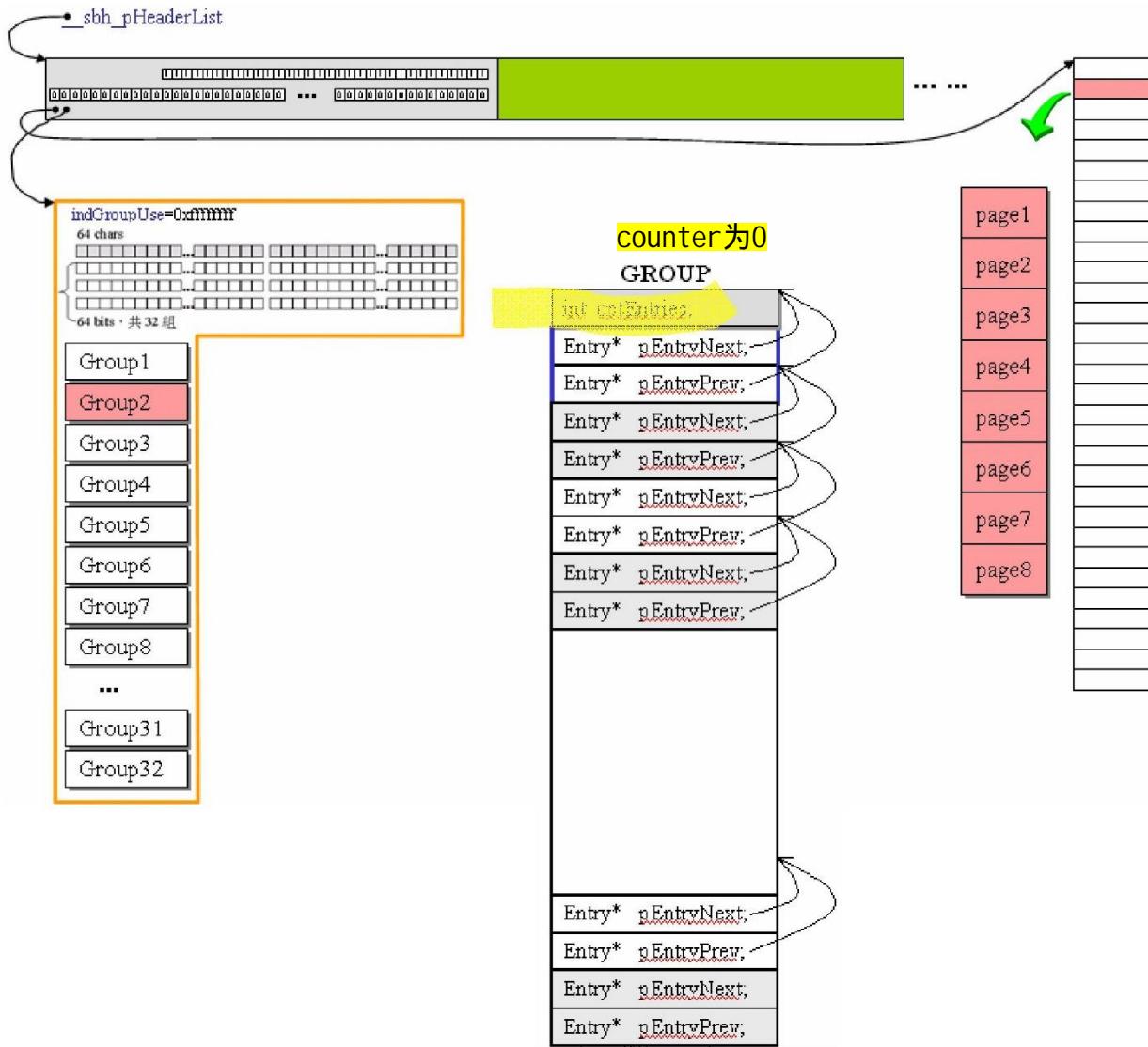




VC6 內存管理

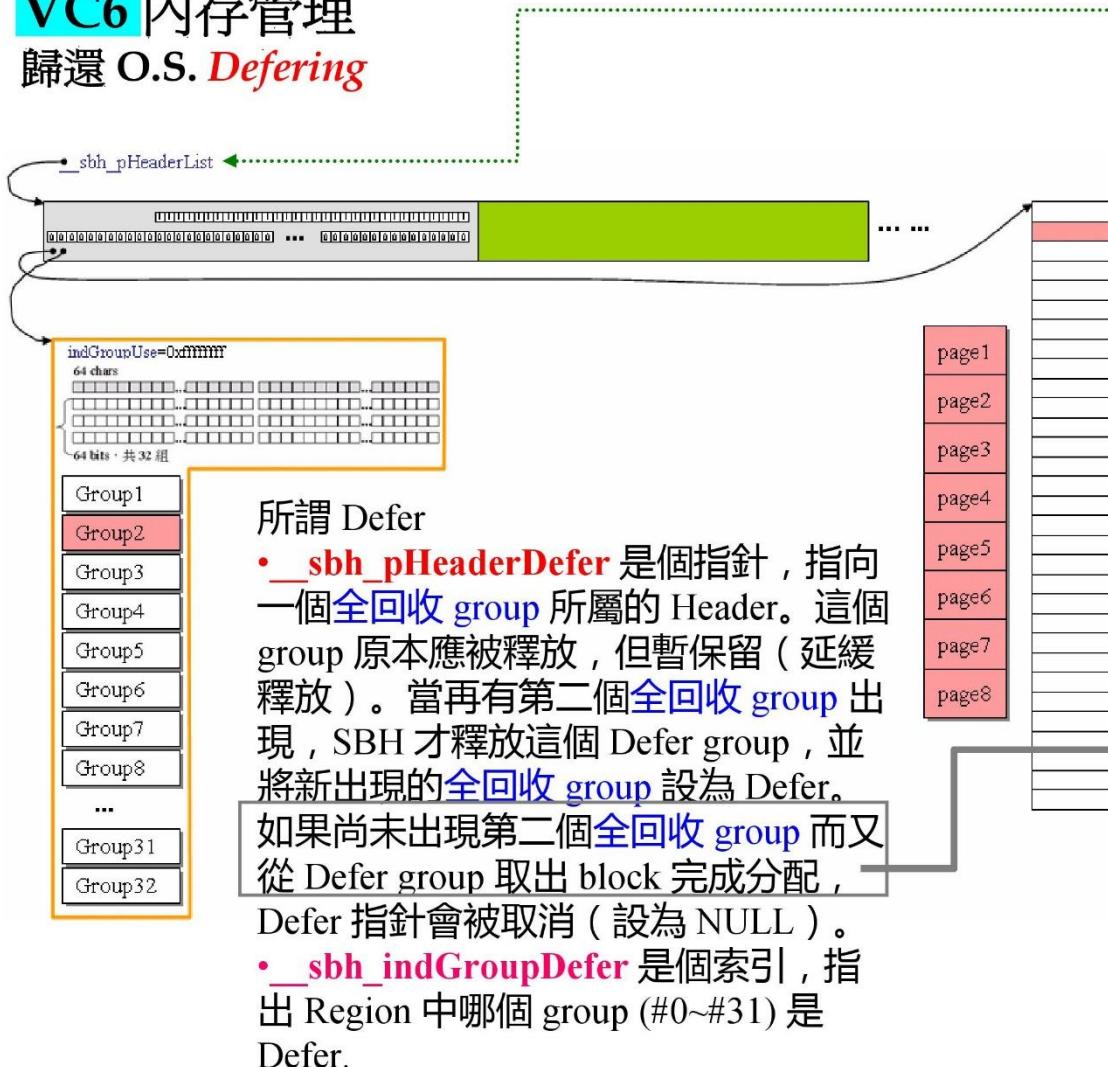
分段管理之妙
利於歸還O.S.

- 1, 如何判斷全回收？
- 2, 不要躁進！



VC6 內存管理

歸還 O.S. *Defering*



```

int __cdecl __sbh_heap_init (void)
{
    if (!(_sbh_pHeaderList =
        HeapAlloc(_crtheap, 0, 16 * sizeof(H HEADER)))
        return FALSE;

    _sbh_pHeaderScan = _sbh_pHeaderList;
    _sbh_pHeaderDefer = NULL;
    _sbh_cntHeaderList = 0;
    _sbh_sizeHeaderList = 16;

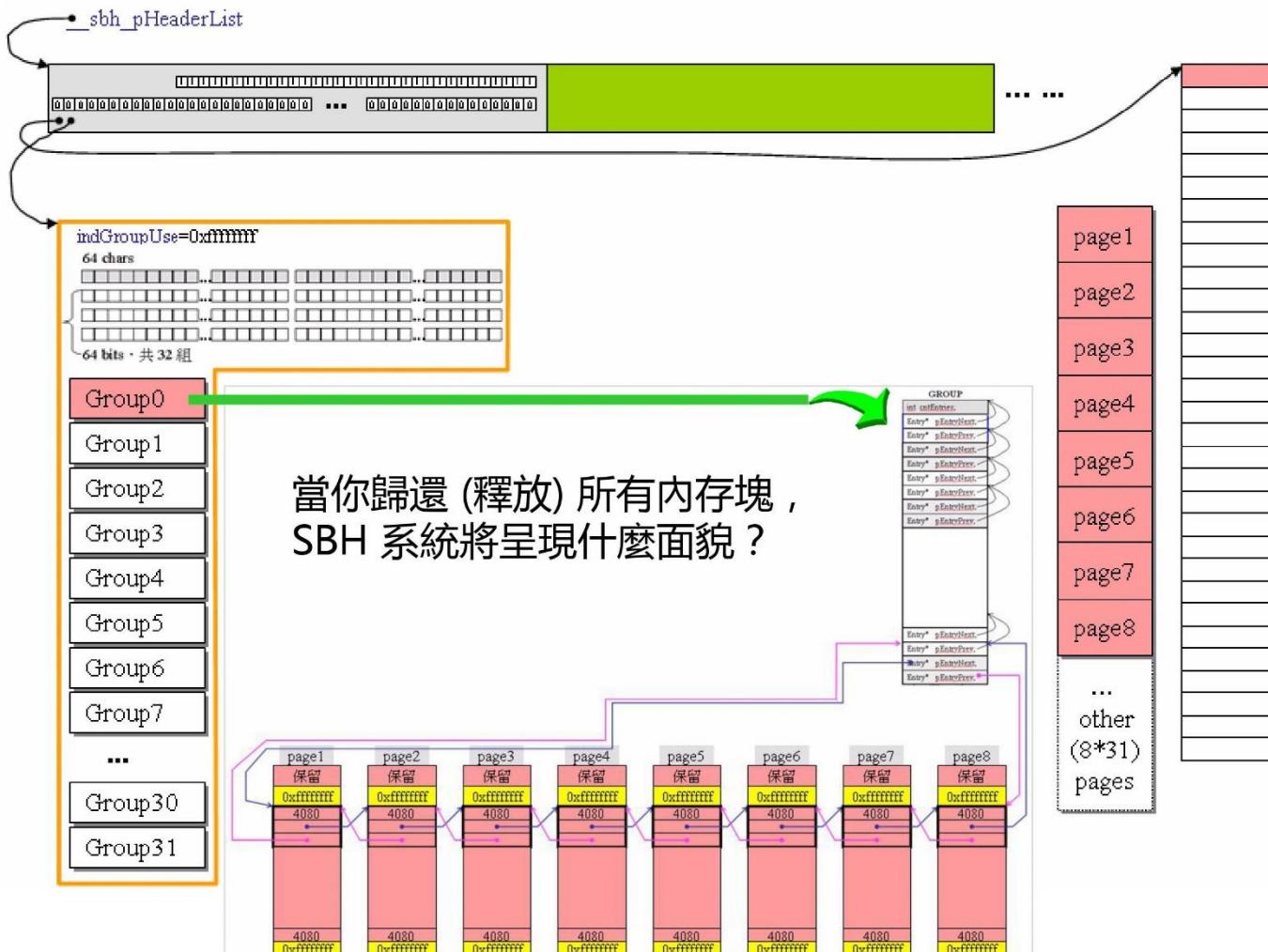
    return TRUE;
}

```

SBH 分配好 block 後，最終準備將 cntEntries 累加 1，但會先檢查它原本是否為 0；若成立，再看此次所在之 Header 是否 Defer 且此次所用之 Group 是否 Defer？都吻合就取消其 Defer 身份（令 **_sbh_pHeaderDefer = NULL**）。

此意味 Defer 必定是個全回收 Group。全新 Group 不會是 Defer，因為 Defer 指針不可能指向全新 Group。

VC6 內存管理



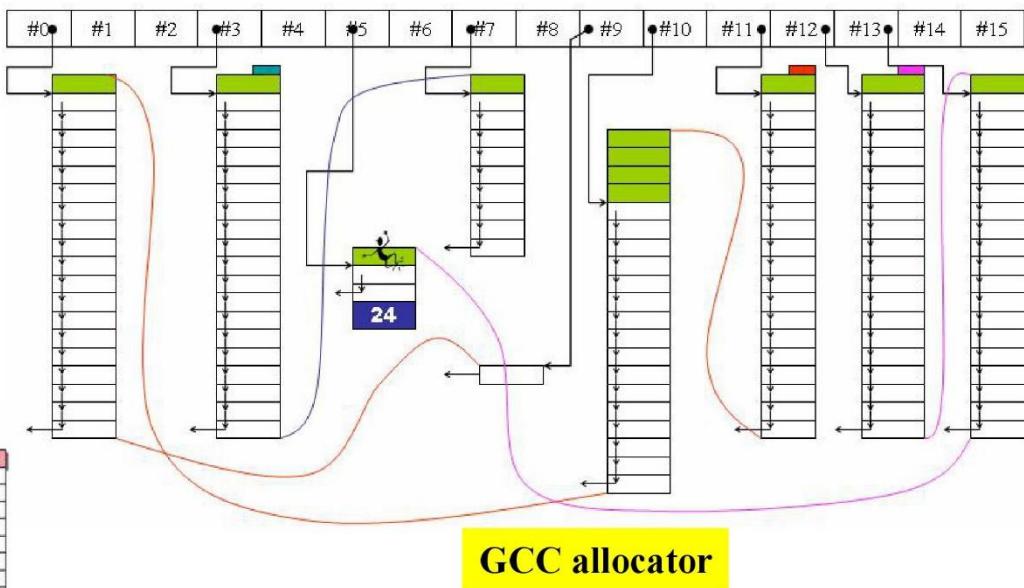
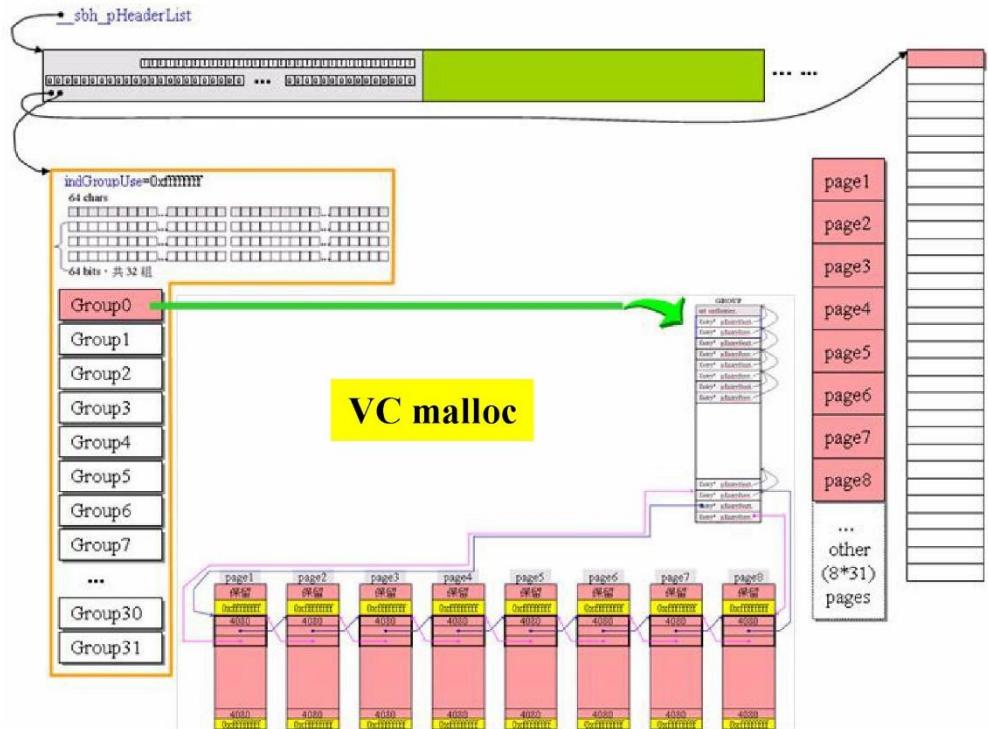
VC6, Heap State Reporting Functions

The following functions report the state and contents of the heap, and use the information to help detect memory leaks and other problems:

Function	Description
_CrtMemCheckpoint	Saves a snapshot of the heap in a _CrtMemState structure supplied by the application.
_CrtMemDifference	Compares two memory state structures, saves the difference between them in a third state structure, and returns TRUE if the two states are different.
_CrtMemDumpStatistics	Dumps a given _CrtMemState structure. The structure may contain a snapshot of the state of the debug heap at a given moment, or the difference between two snapshots. "Dumping" means reporting the data in a form that a person can understand.
_CrtMemDumpAllObjectsSince	Dumps information about all objects allocated since a given snapshot was taken of the heap, or from the start of execution. Every time it dumps a _CLIENT_BLOCK block, it calls a hook function supplied by the application, if one has been installed using _CrtSetDumpClient .
_CrtDumpMemoryLeaks	Determines whether any memory leaks occurred since the start of program execution, and if so, it dumps all allocated objects. Every time it dumps a _CLIENT_BLOCK block, it calls a hook function supplied by the application, if one has been installed using _CrtSetDumpClient .



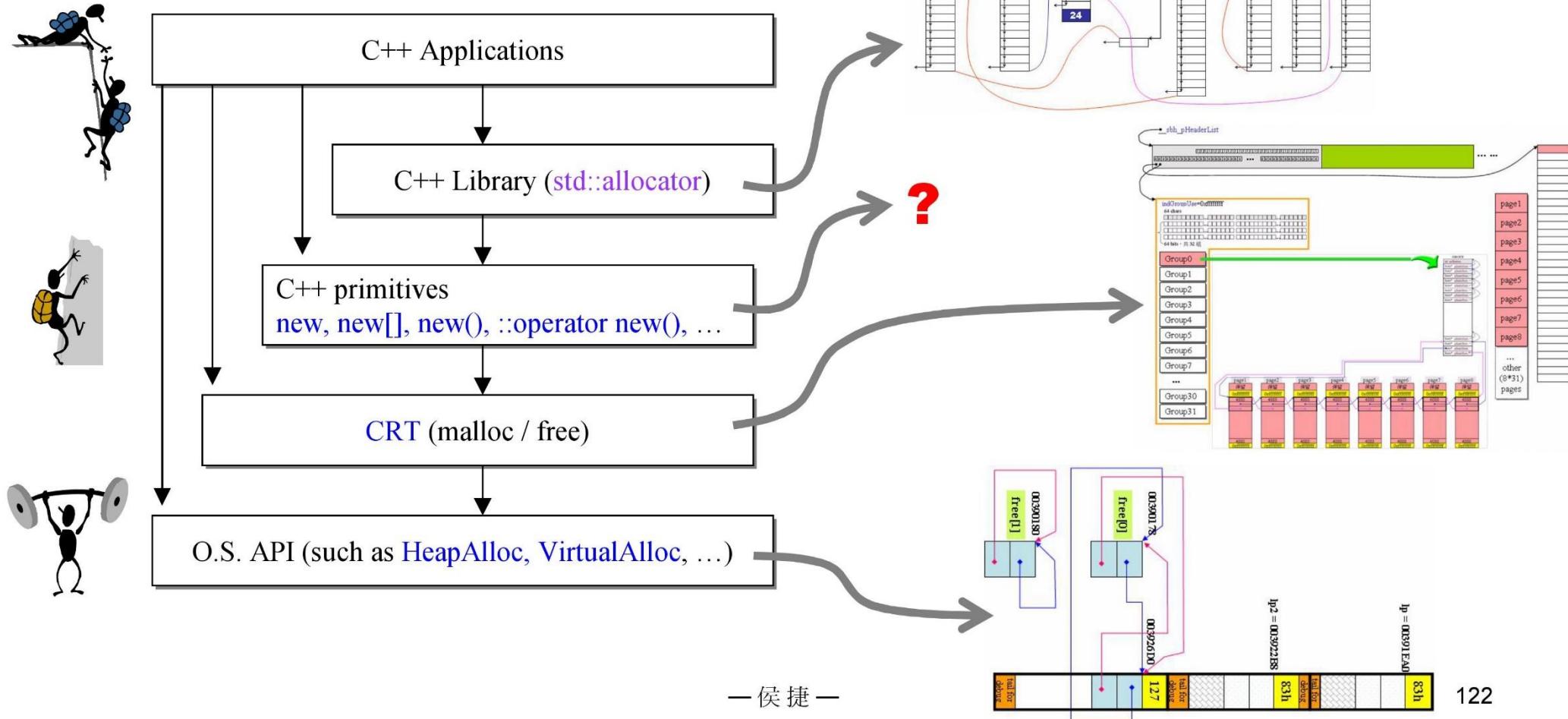
VC malloc + GCC allocator 亂點鴛鴦譜？



GCC allocator



疊床架屋，有必要嗎？



內存管理

從平地到萬丈高樓

Memory Management 101

第一講 primitives

第二講 std::allocator

第三講 malloc/free

第四講 loki::allocator

第五講 other issues



侯捷

成竹在胸



Loki Library

sourceforge.net/projects/loki-lib/

SOURCEFORGE Search Browse Enterprise Blog Jobs

SOLUTION CENTERS Go Parallel Resources Newsletters

Home / Browse / Development / Design / Loki

Loki Beta

Brought to you by: [aandrei](#), [rich_sposato](#), [syntheticpp](#)

Summary Files Reviews Support Wiki Mailing Lists Tickets News Discussion Code

★ 5.0 Stars (6)
↓ 187 Downloads (This Week)
Last Update: 2013-04-08

sf Download loki-0.1.7.exe

Browse All Files

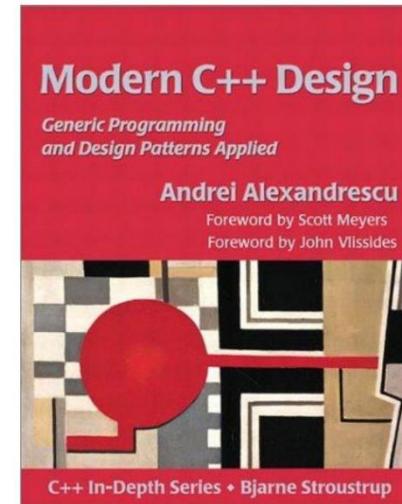
Description

A C++ library of designs, containing flexible implementations of common design patterns and idioms.

[Loki Web Site >](#)

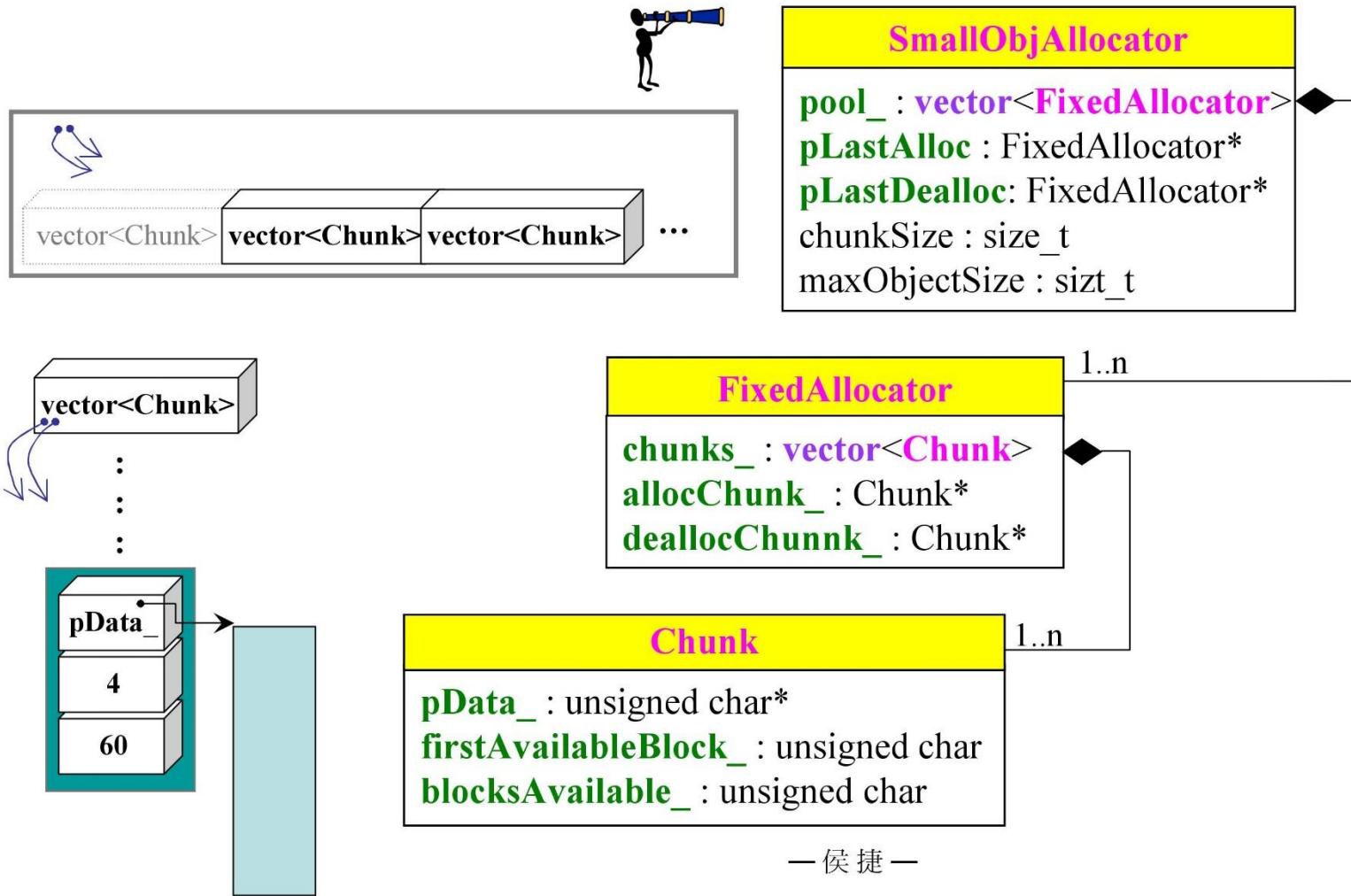
Categories License
Design MIT License

126



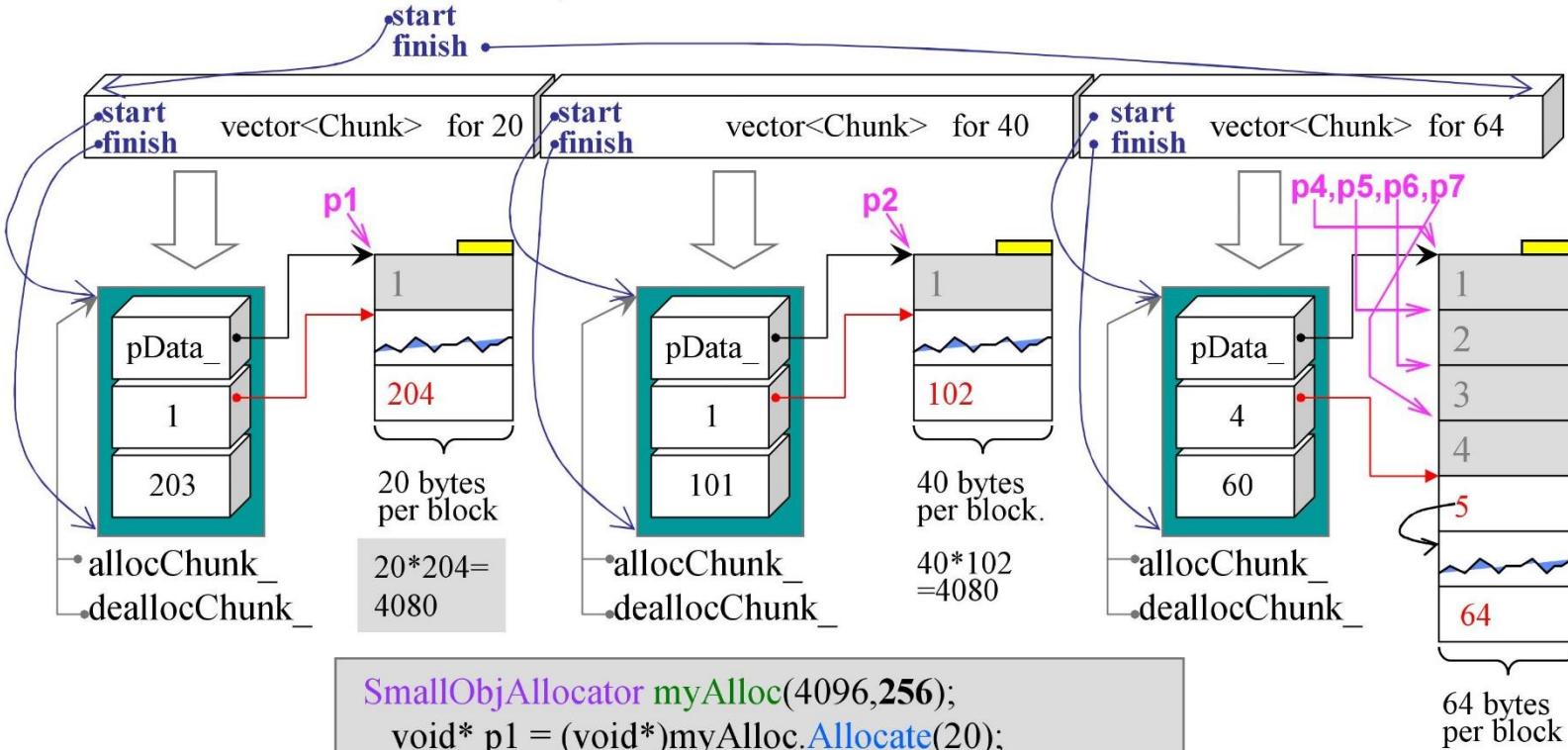


Loki allocator, 3 classes





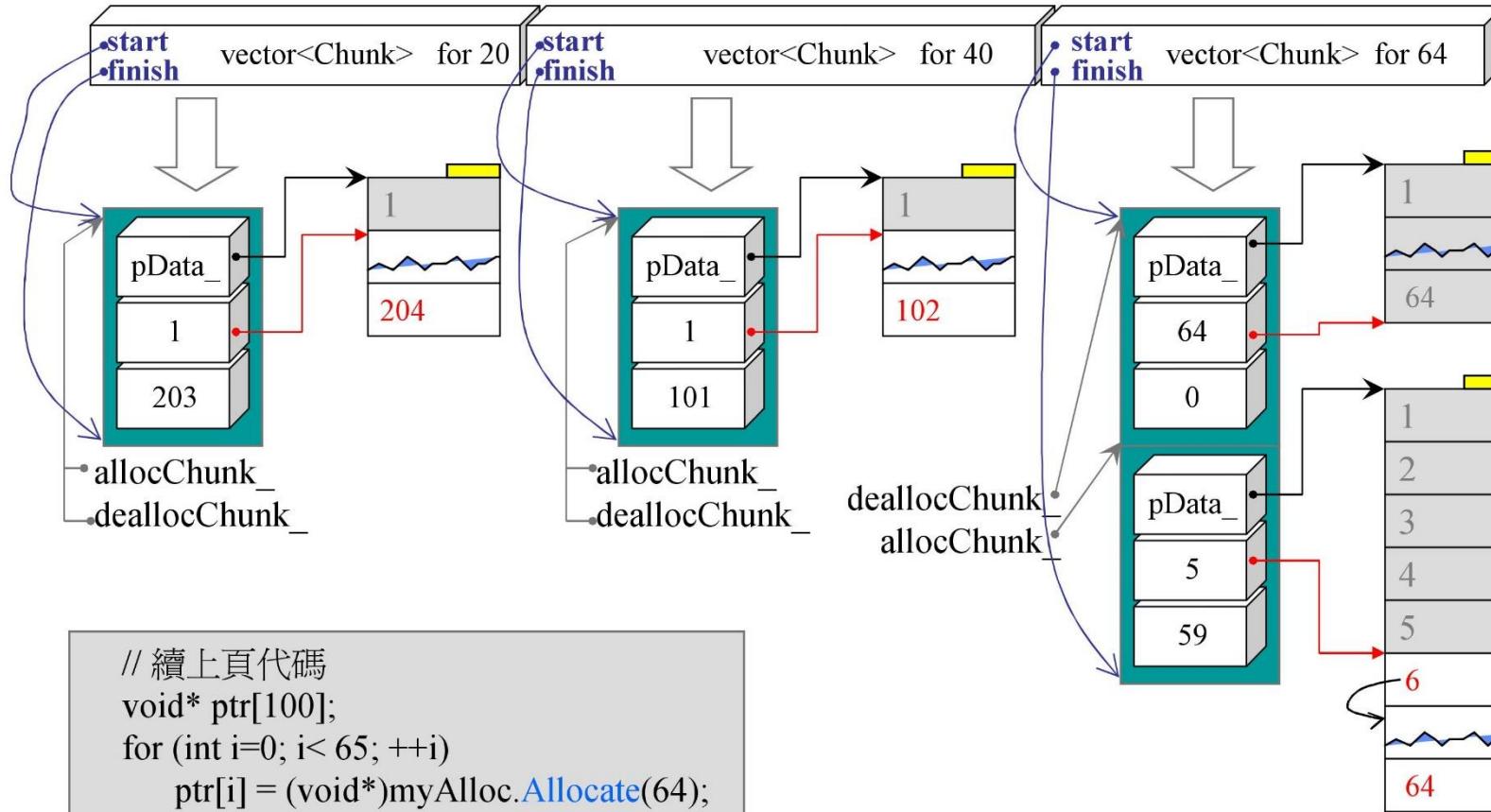
Loki allocator, 1



```
SmallObjAllocator myAlloc(4096,256);
void* p1 = (void*)myAlloc.Allocate(20);
void* p4 = (void*)myAlloc.Allocate(64);
void* p2 = (void*)myAlloc.Allocate(40);
void* p3 = (void*)myAlloc.Allocate(300); // > 256
void* p5 = (void*)myAlloc.Allocate(64);
void* p6 = (void*)myAlloc.Allocate(64);
void* p7 = (void*)myAlloc.Allocate(64);
```

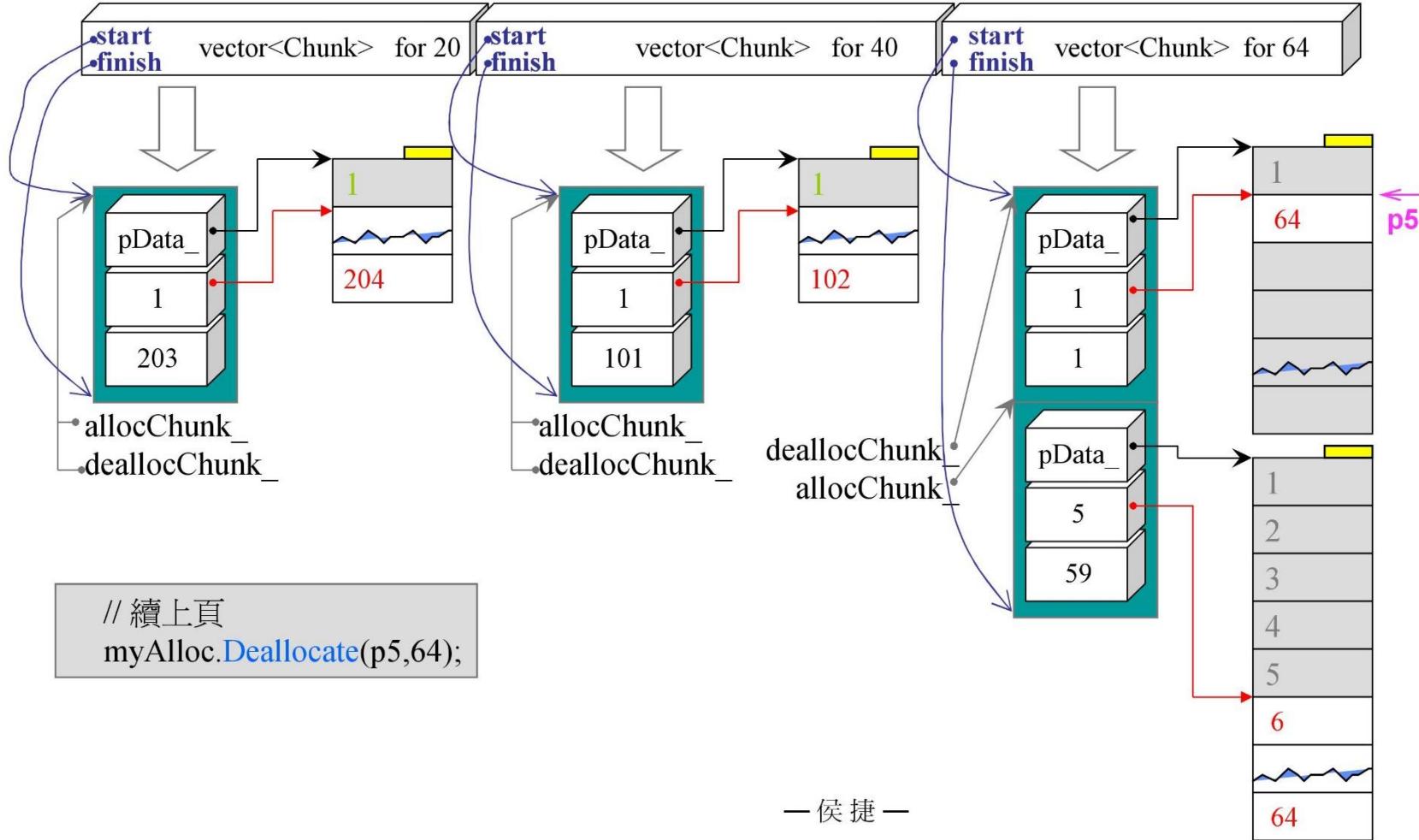


Loki allocator, 2





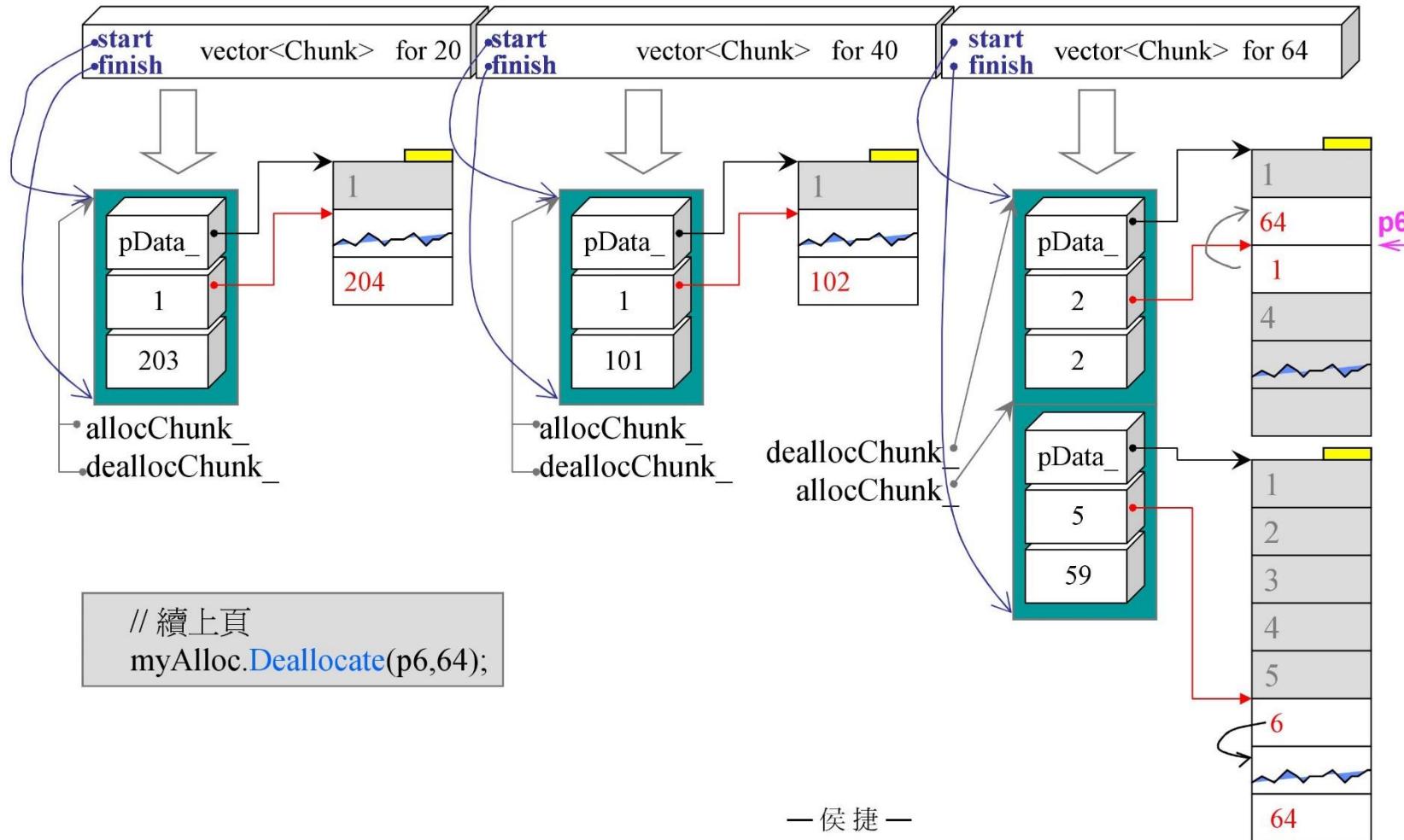
Loki allocator, 3



—侯捷—



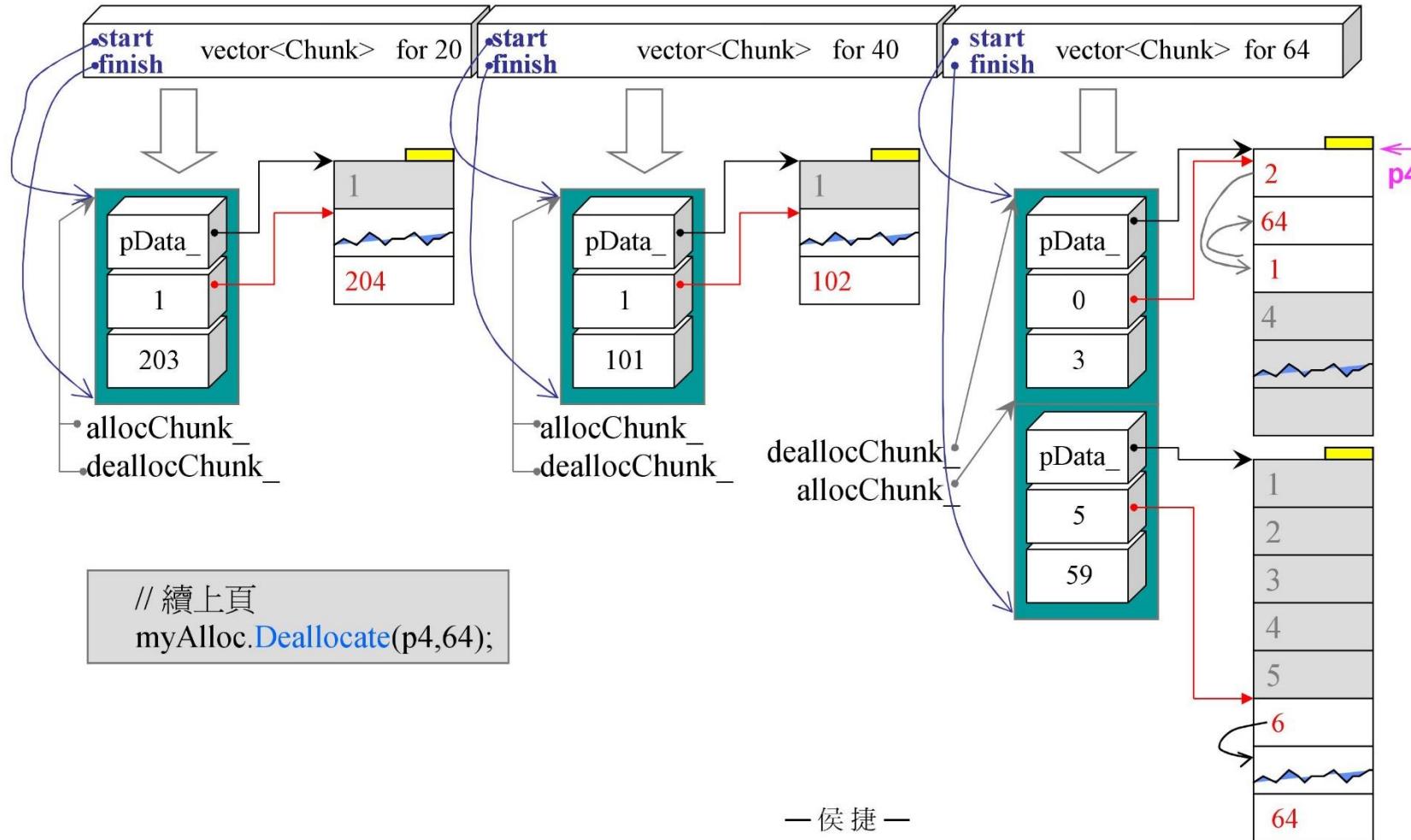
Loki allocator, 4



—侯捷—

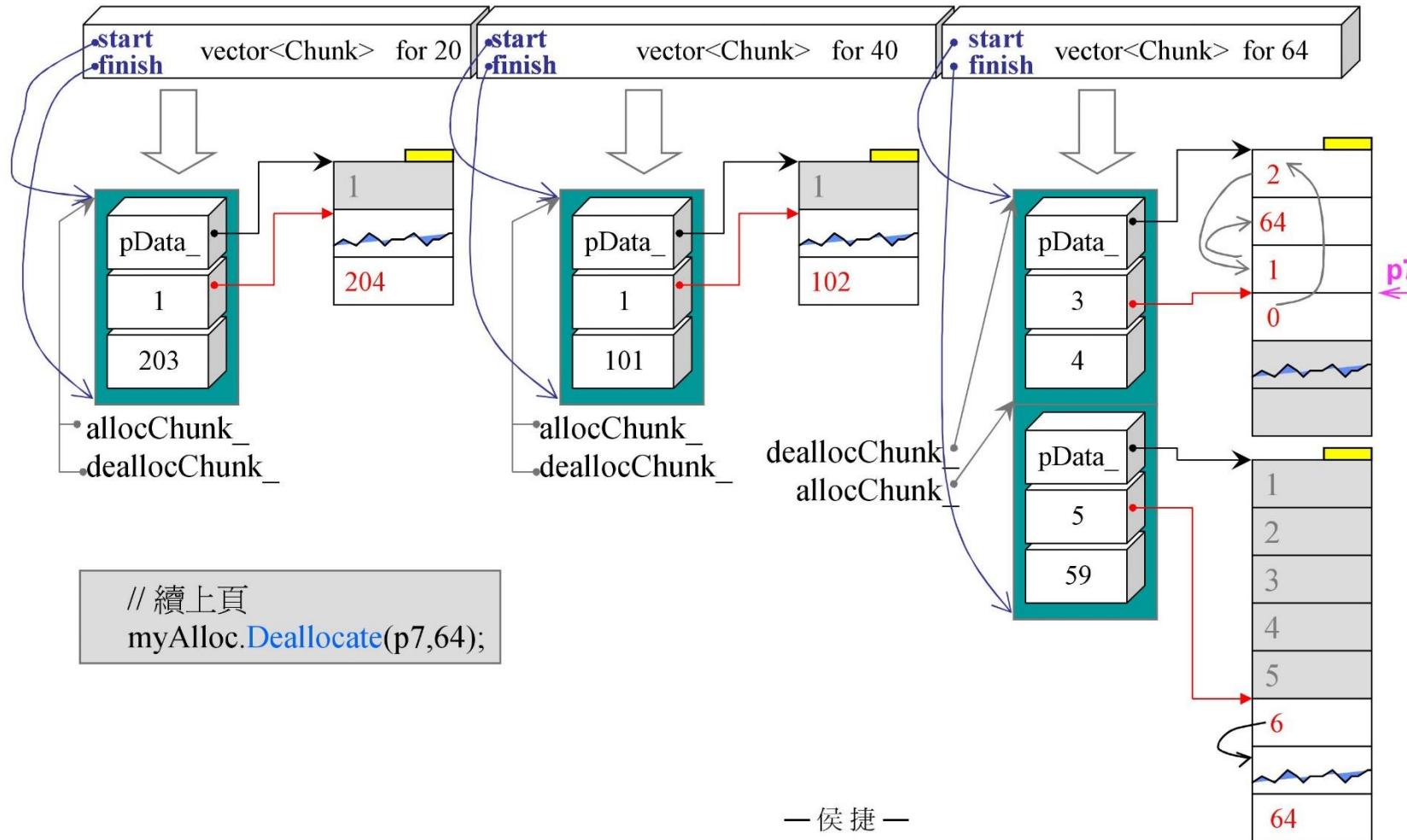


Loki allocator, 5





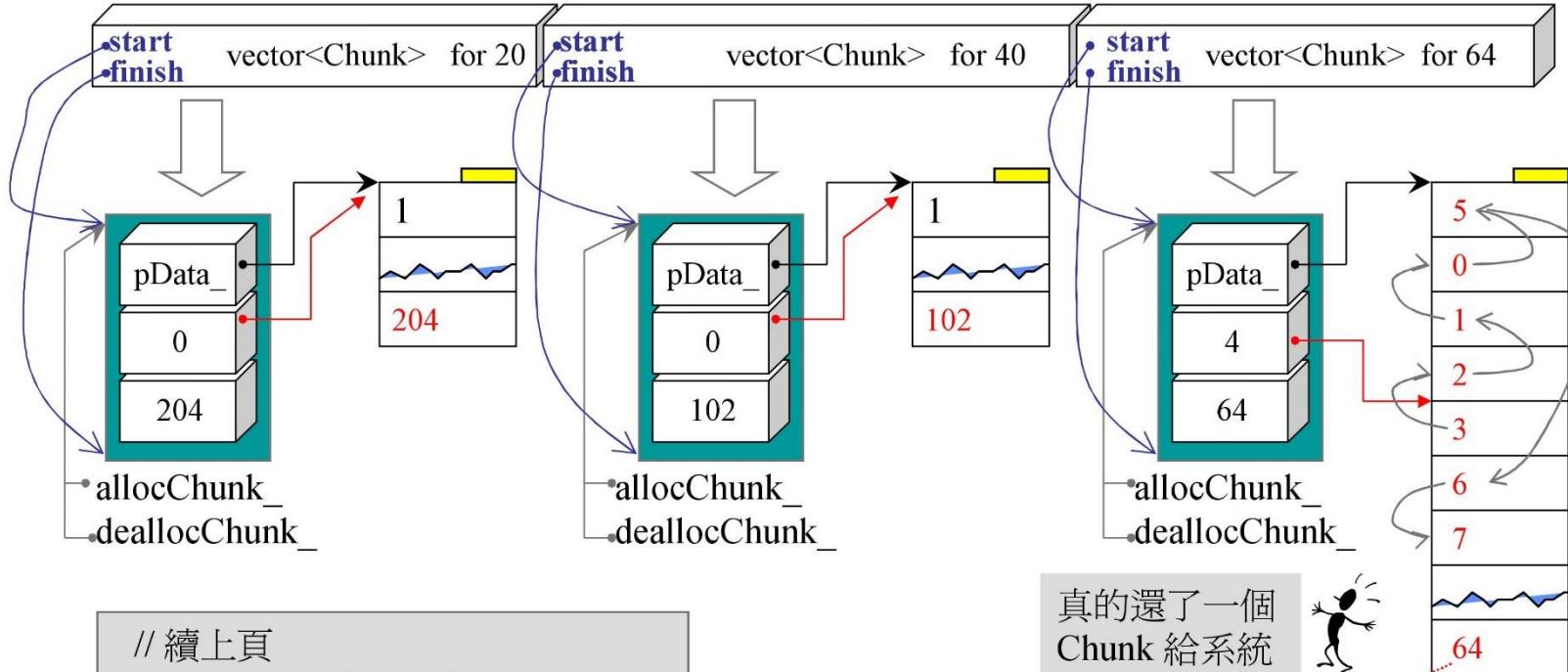
Loki allocator, 6



—侯捷—



Loki allocator, 7



```
// 續上頁
myAlloc.Deallocate(p1,20);
myAlloc.Deallocate(p2,40);
myAlloc.Deallocate(p3,300);
for (int i=0; i< 65; ++i)
    myAlloc.Deallocate(ptr[i], 64);
```

真的還了一個
Chunk 給系統

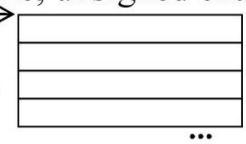


64 對本系統而言是錯誤數字，但它永遠不會被拿取（因為當它被拿取必是 blocksAvailable_ 為 0 的狀態，那就放棄拿取了）



Loki allocator, Chunk

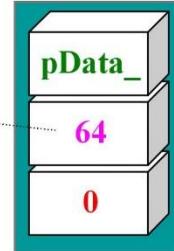
```
void FixedAllocator::Chunk::Init(std::size_t blockSize, unsigned char blocks)
{
    pData_ = new unsigned char[blockSize * blocks];
    Reset(blockSize, blocks);
}
```



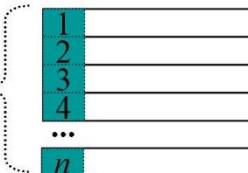
```
void FixedAllocator::Chunk::Reset(std::size_t blockSize,
                                  unsigned char blocks)
```

```
{
    firstAvailableBlock_ = 0;
    blocksAvailable_ = blocks;
```

```
    unsigned char i = 0;
    unsigned char* p = pData_;
    for (; i != blocks; p += blockSize) //流水號標示索引
        *p = ++i;
}
```

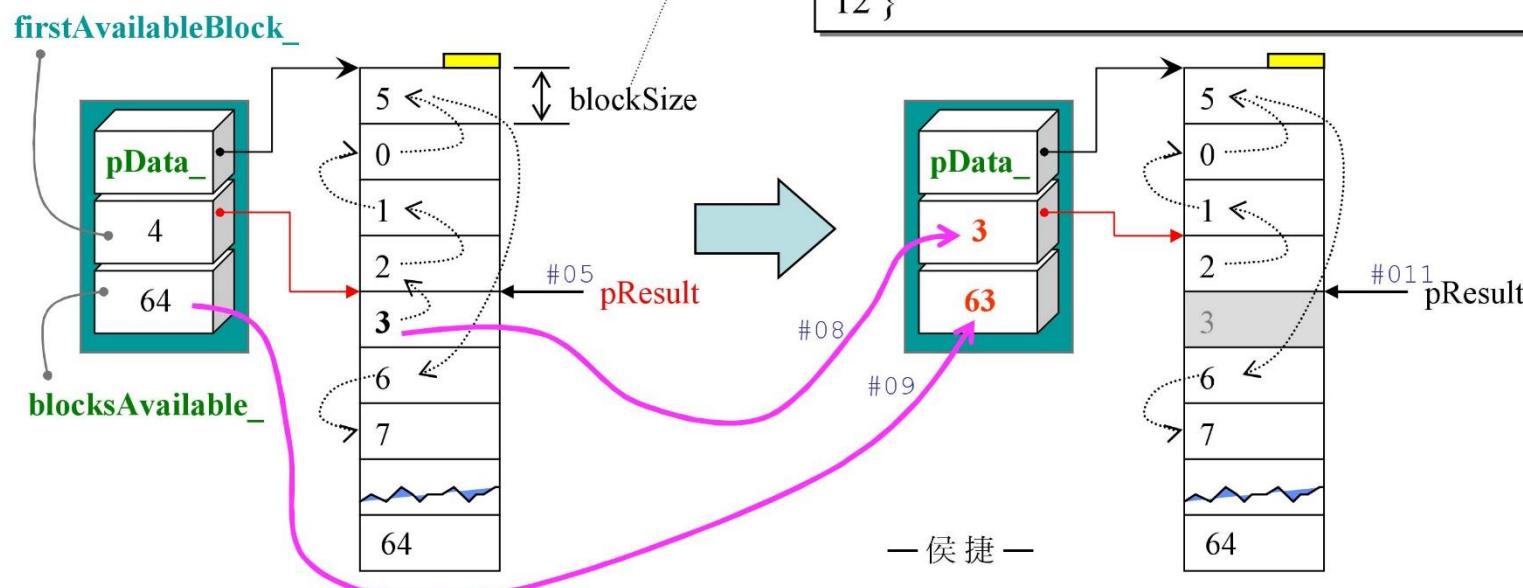


```
void FixedAllocator::Chunk::Release()
{
    delete[] pData_; //釋放自己
} //此函數被上一層調用
```





Chunk::Allocate()



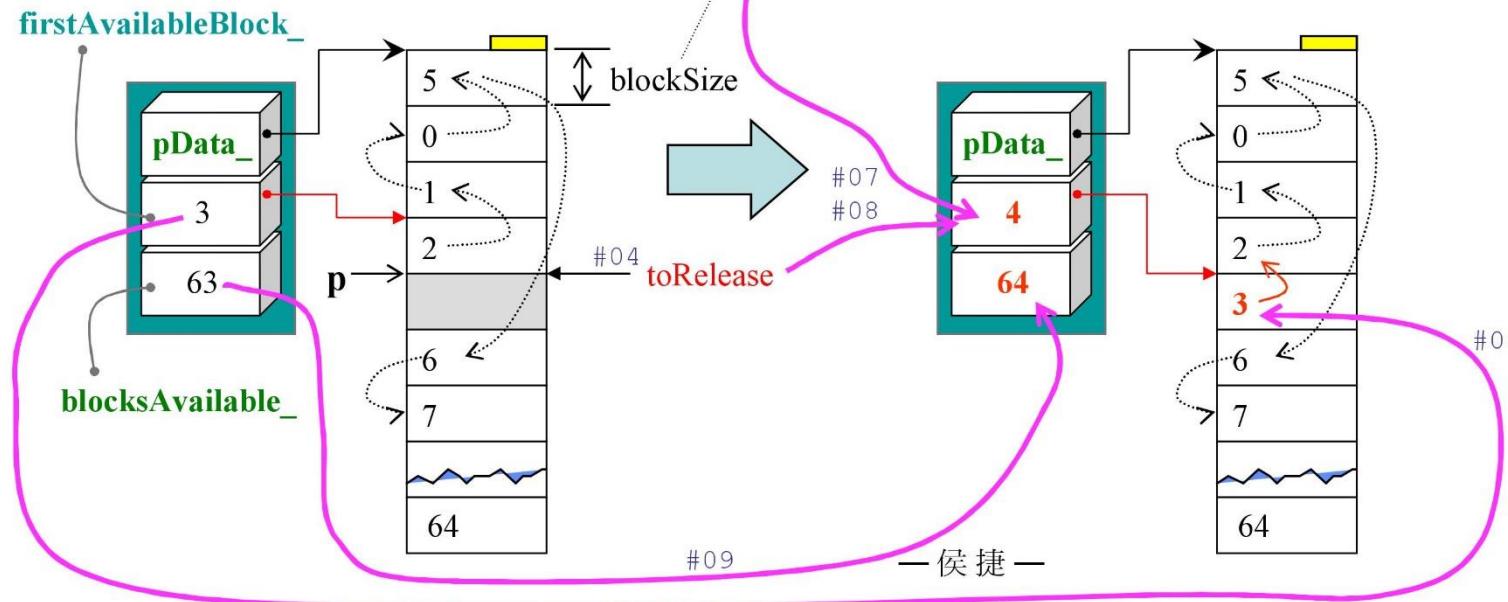
```
01 void* FixedAllocator::Chunk::Allocate(std::size_t blockSize)
02 {
03     if (!blocksAvailable_) return 0; //此地無銀
04
05     unsigned char* pResult = //指向第一個可用區塊
06     pData_ + (firstAvailableBlock_ * blockSize);
07
08     firstAvailableBlock_ = *pResult; //右側索引便是下一個可用區塊
09     --blocksAvailable_;
10
11     return pResult;
12 }
```

—侯捷—



Chunk::Deallocate()

```
01 void FixedAllocator::Chunk::Deallocate(void* p,
02                                         std::size_t blockSize)
03 {
04     unsigned char* toRelease = static_cast<unsigned char*>(p);
05
06     *toRelease = firstAvailableBlock_;
07     firstAvailableBlock_ = static_cast<unsigned char>(
08         (toRelease - pData_) / blockSize);
09     ++blocksAvailable_; //可用區塊數加 1
10 }
```



FixedAllocator::Allocate()

```
#01 void* FixedAllocator::Allocate()
#02 {
#03     if (allocChunk_ == 0 || allocChunk_->blocksAvailable_ == 0)
#04     { //目前沒有標定chunk或該chunk已無可用區塊
#05         Chunks::iterator i = chunks_.begin(); //打算從頭找起
#06         for (; ++i) //找遍每個chunk 直至找到擁有可用區塊者.
#07     {
#08         if (i == chunks_.end()) //到達尾端，都沒找著
#09         {
#10             // Initialize 創建 temp object 拷貝至容器然後結束生命
#11             chunks_.push_back(Chunk()); //產生 a new chunk 掛於末端
#12             Chunk& newChunk = chunks_.back(); //指向末端 chunk
#13             newChunk.Init(blockSize_, numBlocks_); //設好索引
#14             allocChunk_ = &newChunk; //標定,稍後將對此 chunk 取區塊
#15             deallocChunk_ = &chunks_.front(); //另一標定
#16             break;
#17         }
#18         if (i->blocksAvailable_ > 0)
#19         { //current chunk 有可用區塊
#20             allocChunk_ = &*i; //取其位址
#21             break; //不找了,退出 for-loop
#22         }
#23     }
#24 }
#25 return allocChunk_->Allocate(blockSize_); //向這個chunk 取區塊
#26 }
```

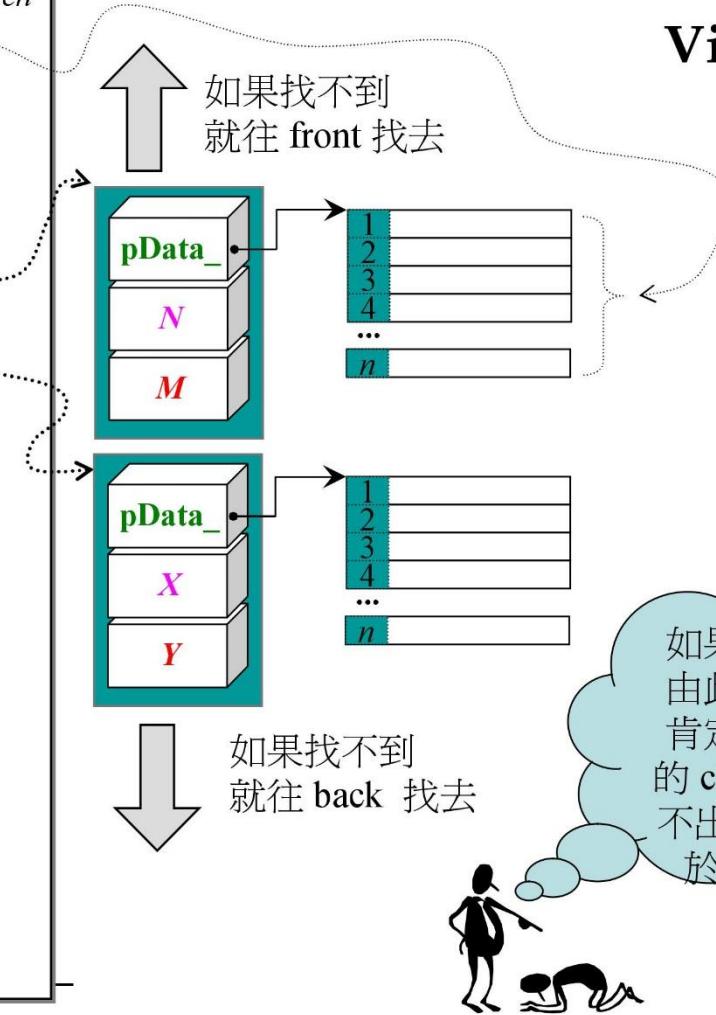
在此 chunk 找到可用區塊，
下次就優先由此找起

FixedAllocator::Deallocate()

```
void FixedAllocator::Deallocate(void* p)
{
    deallocChunk_ = VicinityFind(p);
    DoDeallocate(p);
}
```

FixedAllocator::VicinityFind()

```
#01 // FixedAllocator::VicinityFind (internal)
#02 // Finds the chunk corresponding to a pointer, using an efficient search
#03 FixedAllocator::Chunk* FixedAllocator::VicinityFind(void* p)
#04 {
#05     const std::size_t chunkLength = numBlocks_ * blockSize_;
#06
#07     Chunk* lo = deallocChunk_;
#08     Chunk* hi = deallocChunk_ + 1;
#09     Chunk* loBound = &chunks_.front();
#10    Chunk* hiBound = &chunks_.back() + 1;
#11
#12    for (;;) { //兵分兩路
#13        if (lo) { //上路未到盡頭
#14            if (p >= lo->pData_ && p < lo->pData_ + chunkLength)
#15                return lo; //p 落於 lo 區間內
#16            if (lo == loBound) lo = 0; //到頂了，讓 loop 結束
#17            else --lo; //否則往 lo 的 front 繼續找
#18        }
#19
#20        if (hi) { //下路未到盡頭
#21            if (p >= hi->pData_ && p < hi->pData_ + chunkLength)
#22                return hi; //p 落於 hi 區間內
#23            if (++hi == hiBound) hi = 0; //往 hi 的 back 繼續找
#24            //若到最尾則讓 loop 結束
#25        }
#26    }
#27 }
```



FixedAllocator:: DoDeallocate()

```

#0001 void FixedAllocator::DoDeallocate(void* p)
#0002 {
#0003     //
#0004     deallocChunk_->Deallocate(p, blockSize_);
#0005
#0006     if (deallocChunk_->blocksAvailable_ == numBlocks_) {
#0007         // 確認全回收, 那麼, 該 release 它嗎?
#0008         // 注意, 以下三種情況都重新設定了allocChunk_
#0009         Chunk& lastChunk = chunks_.back(); //標出最後一個
#0010         //如果“最後一個”就是“當前 chunk”
#0011         if (&lastChunk == deallocChunk_) {
#0012             ① // 檢查是否擁有兩個 last chunks empty
#0013                 // 這只需往 front 方向看前一個 chunk 便知.
#0014                 if (chunks_.size() > 1 &&
#0015                     deallocChunk_[-1].blocksAvailable_ == numBlocks_) {
#0016                     // 是的, 有 2 free chunks, 拋棄最後一個
#0017                     lastChunk.Release();
#0018                     chunks_.pop_back();
#0019                     allocChunk_ = deallocChunk_ = &chunks_.front();
#0020                 }
#0021             return;
#0022         }
#0023     }

```



```

#0024     if (lastChunk.blocksAvailable_ == numBlocks_) {
#0025         ② //兩個 free chunks, 拋棄最後一個
#0026         lastChunk.Release();
#0027         chunks_.pop_back();
#0028         allocChunk_ = deallocChunk_;
#0029     }
#0030     else {
#0031         ③ //將free chunk 與最後一個chunk 對調
#0032         std::swap(*deallocChunk_, lastChunk);
#0033         allocChunk_ = &chunks_.back();
#0034     }
#0035 }
#0036 }

```



Loki allocator 檢討

- 曾經有兩個 bugs, 新版已修正.
- 精簡強悍; 手段暴力 (關於 for-loop).
- 使用「以 array 取代 list, 以 index 取代 pointer」的特殊實現手法.
- 能夠以很簡單的方式判斷「chunk 全回收」進而將 memory 歸還給操作系統.
- 有 Deferring (暫緩歸還) 能力
- 這是個 allocator , 用來分配大量小塊不帶 cookie 的 memory blocks , 它的最佳客戶是容器 , 但它本身卻使用 vector , What happens ? 雞生蛋 蛋生雞 ?

內存管理

從平地到萬丈高樓

Memory Management 101

第一講 primitives

第二講 std::allocator

第三講 malloc/free

第四講 loki::allocator

第五講 the others

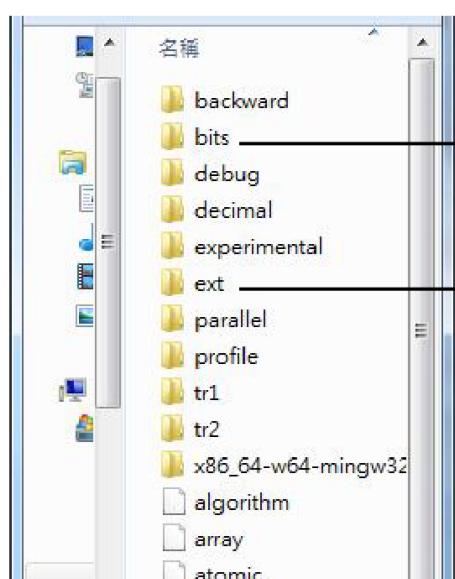


侯捷

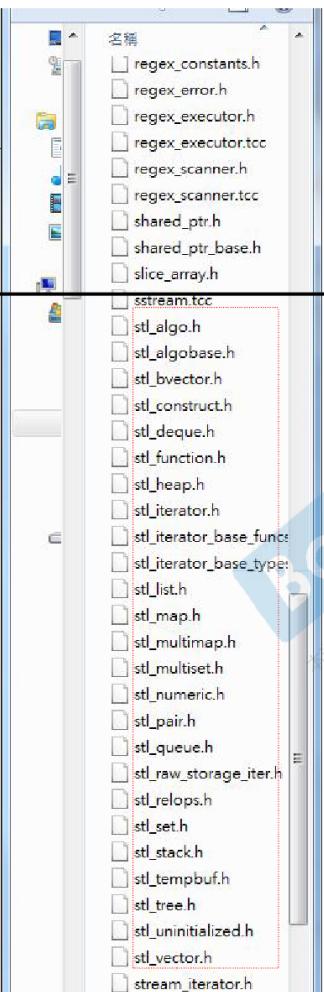


GNU C++

...\\4.9.2\\include\\c++



...\\include\\c++\\bits



...\\include\\c++\\ext

```
C:\命令提示字元
C:\Program Files\Dev-Cpp\MinGW64\lib\gcc\x86_64-w64-mingw32\4
xt 的目錄

2014/12/08 上午 06:32      5,177 array_allocator.h
2014/12/08 上午 06:32     31,324 bitmap_allocator.h
2014/12/08 上午 06:32     5,716 debug_allocator.h
2014/12/08 上午 06:32    6,189 extptr_allocator.h
2014/12/08 上午 06:32    4,466 malloc_allocator.h
2014/12/08 上午 06:32   22,973 mt_allocator.h
2014/12/08 上午 06:32    4,436 new_allocator.h
2014/12/08 上午 06:32   8,423 pool_allocator.h
2014/12/08 上午 06:32  24,537 throw_allocator.h
                           9 個檔案          113,241 位元組
                           0 個目錄  86,021,267,456 位元組可用
```

— 侯 捷 —



GNU C++ 對於 Allocator 的描述

當你將元素加入容器中，容器必須分配更多內存以保存這些元素，於是它們向其模板參數 **Allocator** 發出申請，該模板參數往往被另名為 (*aliased to*) `allocator_type`。甚至你將 `chars` 添加到 `string` class 也是如此，因為 `string` 也算是一個正規的 STL 容器。

每個元素類型為 `T` 的容器 (container-of-`T`) 的 `Allocator` 模板實參默認為 `allocator<T>`。其接口只有大約 20 個 public 聲明，包括嵌套的 (nested) `typedefs` 和成員函數。最重要的兩個成員函數是：

```
T* allocate (size_type n, const void* hint = 0);  
void deallocate (T* p, size_type n);
```

`n` 指的是客戶申請的元素個數，不是指空間總量。

這些空間是通過調用 `::operator new` 獲得，但 **何時調用** 以及 **多麼頻繁調用**，並無具體指定。

```
template <class T,  
         class Allocator=allocator<T>>  
class vector;
```

```
template <class T,  
         class Allocator=allocator<T>>  
class list;
```

```
template <class T,  
         class Allocator=allocator<T>>  
class deque;
```



GNU C++ 對於 Allocator 的描述

最容易滿足需求的作法就是每當容器需要內存就調用 `operator new`，每當容器釋放內存就調用 `operator delete`。這種作法比起分配大塊內存並緩存 (caching) 然後徐徐小塊使用當然較慢，優勢則是可以在極大範圍的硬件和操作系統上有效運作。

➤ `_gnu_cxx::new_allocator`

實現出簡樸的 `operator new` 和 `operator delete` 語意。

```
template<typename _Tp>
class new_allocator
{
    .....
    pointer allocate(size_type __n, const void* = 0) {
        .....
        return static_cast<_Tp*>
            (::operator new(__n * sizeof(_Tp)));
    }
    void deallocate(pointer __p, size_type)
    { ::operator delete(__p); }
};
```

➤ `_gnu_cxx::malloc_allocator`

其實現與上例唯一不同的是，
它使用 C 函數 `std::malloc` 和 `std::free`。

```
template<typename _Tp>
class malloc_allocator
{
    pointer allocate(size_type __n, const void* = 0) {
        .....
        pointer __ret = .....(std::malloc(__n * sizeof(_Tp)));
    }
    void deallocate(pointer __p, size_type) {
        std::free(.....(__p));
    }
};
```



GNU C++ 對於 Allocator 的描述

另一種作法就是使用**智能型 allocator**，將分配所得的內存加以緩存(*cache*)。這種額外機制可以數種形式呈現：

- 可以是個 **bitmap index**，用以索引至一個以 2 的指數倍成長的籃子(exponentially increasing power-of-two-sized buckets)
- 也可以是個相較之下比較簡易的 **fixed-size pooling cache**.

這裡所說的 **cache** 被程序內的所有容器共享，而 **operators new** 和 **operator delete** 不經常被調用，這可帶來速度上的優勢。使用這類技巧的 allocators 包括：

- **gnu_cxx::bitmap_allocator**
一個高效能 allocator, 使用 bit-map 追蹤被使用和未被使用(used and unused)的內存塊。
- **gnu_cxx::pool_allocator**
- **gnu_cxx::__mt_alloc**

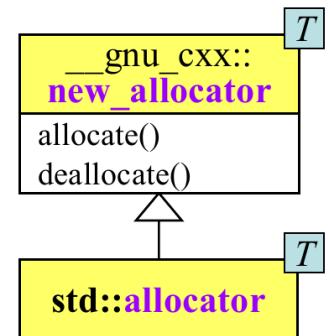
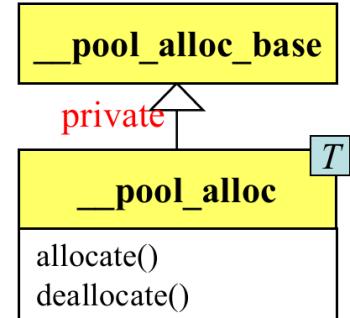
GNU C++ 對於 Allocator 的描述

Class allocator 只擁有 `typedef`, `constructor`, 和 `rebind` 等成員。它繼承自一個 `_high-speed extension allocators`。也因此，所有分配和歸還 (allocation and deallocation) 都取決於該 base class，而這個 base class 也許是終端用戶無法碰觸和操控的 (user-configurable).

很難挑選出某個分配策略說它能提供最大共同利益而不至於令某些行為過度劣勢。
事實上，就算要挑選何種典型動作以量測速度，都是一種困難。

GNU C++ 提供三項綜合測試 (three synthetic benchmarks) 用以完成 C++ allocators 之間的速度評比：

- **Insertion.** 經過多次 `iterations` 後各種 STL 容器將擁有某些極大量。分別測試循序式 (sequence) 和關聯式 (associative) 容器。
- **多線程環境中的 insertion and erasure.** 這個測試展示 allocator 歸還內存的能力，以及量測線程之間對內存的競爭。
- **A threaded producer/consumer model.** 分別測試循序式 (sequence) 和關聯式 (associative) 容器。





GNU C++ 對於 Allocator 的描述

另兩個智能型 allocator：

➤ _gnu_cxx::debug_allocator

這是一個外覆器 (wrapper)，可包覆於任何 allocator 之上。它把客戶的申請量添加一些，然後由 allocator 回應，並以那一小塊額外內存放置_size 信息。一旦 deallocate() 收到一個 pointer，就會檢查 size 並以 assert() 保證吻合。

➤ _gnu_cxx::array_allocator

允許分配一已知且固定大小 (known and fixed size) 的內存塊，內存來自 std::array objects。用上這個 allocator，大小固定的容器 (包括 std::string) 就無需再調用 ::operator new 和 ::operator delete。這就允許我們使用 STL abstractions 而無需在運行期添亂、增加開銷。甚至在 program startup 情況下也可使用。

— 侯 捷 —

```
ExitProcess(code)
    _initterm(,,) //do terminators
        __endstdio(void)
    _initterm(,,) //do pre-terminators
    doexit(code, 0, 0)
9 exit(code)
8 main()
    _initterm(,,) //do C++ initializations
        __initstdio(void)
    _initterm(,,) //do initializations
7 __cinit() // do C data initialize
6 __setenvp()
5 __setargv()
4 __crtGetEnvironmentStringsA()
3 GetCommandLineA()
    __sbh_alloc_new_group(...)
    __sbh_alloc_new_region()
    __sbh_alloc_block(...)
    _heap_alloc_base(...)
    _heap_alloc_dbg(...)
    _nh_malloc_dbg(...)
    _malloc_dbg(...)
2 __ioinit() // initialize lowio
    __sbh_heap_init()
1 __heap_init(...)

mainCRTStartup()
KERNEL32! bff8b6e6()
KERNEL32! bff8b598()
KERNEL32! bff89f5b()
```

VS2013 標準分配器 與 new_allocator

<.../list>

```
template<class _Ty> 《...\\xmemory0》 <-->
class allocator
: public _Allocator_base<_Ty>
{ // generic allocator for objects of class _Ty
public:
    typedef value_type *pointer;
    typedef size_t size_type;

    void deallocate(pointer _Ptr, size_type)
    { // deallocate object at _Ptr, ignore size
        ::operator delete(_Ptr);
    }

    pointer allocate(size_type _Count)
    { // allocate array of _Count elements
        return (_Allocate(_Count, (pointer)0));
    }

    pointer allocate(size_type _Count, const void *)
    { // allocate array of _Count elements, ignore hint
        return (allocate(_Count));
    }

    ...
};
```

《...\\memory》

```
...
#include <xmemory>
```

《...\\xmemory》

```
...
#include <xmemory0>
```

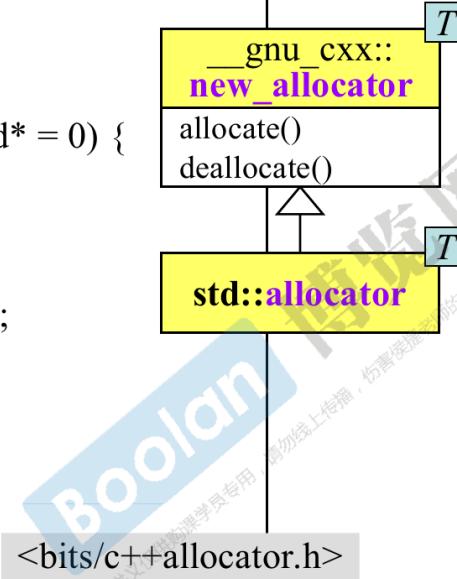
```
template<class _Ty,
         class _Alloc = allocator<_Ty>>
class list
: public _List_buy<_Ty, _Alloc>
{
    // bidirectional linked list
    ...
};
```

// TEMPLATE FUNCTION _Allocate

```
template<class _Ty> inline
_Ty *_Allocate(size_t _Count, _Ty *)
{
    // allocate storage for _Count elements of type _Ty
    void *_Ptr = 0;
    if (_Count == 0);
    else if ((( )(-1) / sizeof (_Ty)) < _Count)
        || (_Ptr = ::operator new(_Count * sizeof (_Ty))) == 0)
        _Xbad_alloc(); // report no memory
    return ((_Ty *)_Ptr);
}
```

G4.9 標準分配器 與 new_allocator

```
template<typename _Tp>           <.../ext/new_allocator.h>
class new_allocator
{
...
pointer allocate(size_type __n, const void* = 0) {
    if (__n > this->max_size())
        std::__throw_bad_alloc();
    return static_cast<_Tp*>(
        ::operator new(__n * sizeof(_Tp)));
}
void deallocate(pointer __p, size_type)
{ ::operator delete(__p); }
...
};
```



```
#include <ext\new_allocator.h>           <.../bits/allocator.h>
template<typename _Tp>
class allocator : public __allocator_base<_Tp>  std
{ ... };
```

—侯捷—

```
template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class vector : protected _Vector_base<_Tp, _Alloc>
{ ... };
```

```
template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class list : protected _List_base<_Tp, _Alloc>
{ ... };
```

```
template<typename _Tp,
         typename _Alloc = std::allocator<_Tp>>
class deque : protected _Deque_base<_Tp, _Alloc>
{ ... };
```

《...\\memory》

```
...
#include <bits\allocator.h>
```

G4.9 malloc_allocator

```
template<typename _Tp>
class malloc_allocator
{
    // NB: __n is permitted to be 0. The C++ standard says nothing
    // about what the return value is when __n == 0.
    pointer
    allocate(size_type __n, const void* = 0)
    {
        if (__n > this->max_size())
            std::__throw_bad_alloc();

        pointer __ret = static_cast<_Tp*>(std::malloc(__n * sizeof(_Tp)));
        if (!__ret)
            std::__throw_bad_alloc();
        return __ret;
    }
}
```

<.../ext/malloc_allocator.h>

T	<u>gnu_cxx::</u> malloc_allocator
	allocate()
	deallocate()

// __p is not permitted to be a null pointer.

```
void
deallocate(pointer __p, size_type)
{ std::free(static_cast<void*>(__p)); }
```

size_type

```
max_size() const _GLIBCXX_USE_NOEXCEPT
{ return size_t(-1) / sizeof(_Tp); }
```

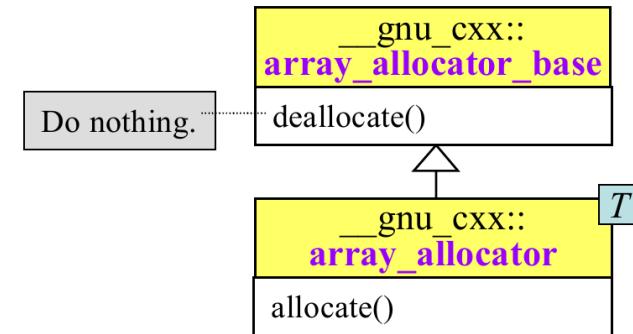
.....

};

G4.9 array_allocator

<.../ext/array_allocator.h>

```
template<typename _Tp, typename _Array = std::tr1::array<_Tp, 1>>
class array_allocator : public array_allocator_base<_Tp>
{
public:
    typedef size_t      size_type;
    typedef _Tp         value_type;
    typedef _Array      array_type;
    .....
private:
    array_type* _M_array;
    size_type   _M_used;
    .....
public:
    array_allocator(array_type* __array = NULL) throw()
        : _M_array(__array), _M_used(size_type()) { }
    .....
```



Do nothing.



pointer

```
allocate(size_type __n, const void* = 0)
{
    if (_M_array == 0 || _M_used + __n > _M_array->size())
        std::__throw_bad_alloc();
    pointer __ret = _M_array->begin() + _M_used;
    _M_used += __n;
    return __ret;
};
```

G4.9 array_allocator

```
using namespace std;
using namespace std::tr1;
using namespace __gnu_cxx;
```

```
typedef ARRAY std::array<int, 65536>;
```

```
ARRAY* pa = new ARRAY;
```

```
array_allocator<int, ARRAY> myalloc(pa);
```

```
int* p1 = myalloc.allocate(1);
```

```
int* p2 = myalloc.allocate(3);
```

```
myalloc.deallocate(p1); //no-op
```

```
myalloc.deallocate(p2); //no-op
```

```
delete pa;
```

使用 `std::tr1::array` 或 `std::array` 均可。

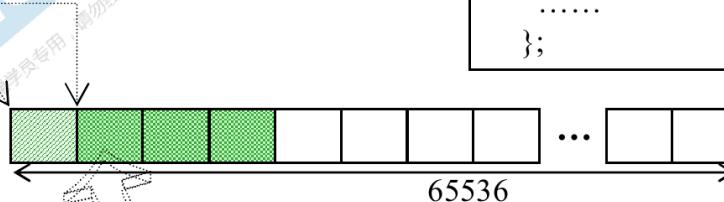
這種 array object
其內部將有 `int _M_instance[65536]`;

```
pointer allocate(size_type __n, const void* = 0)
{
    if (_M_array == 0 || _M_used + __n > _M_array->size())
        std::__throw_bad_alloc();
    pointer __ret = _M_array->begin() + _M_used;
    _M_used += __n;
    return __ret;
}
```

```
template<typename _Tp, std::size_t _Nm>
struct array
{.....
```

```
    value_type _M_instance[_Nm ? _Nm : 1];
```

```
    iterator begin()
    { return iterator(&_M_instance[0]); }
    size_type size() const { return _Nm; }
    .....
};
```



array_allocator 並不會
回收已給出的內存空間

G4.9 array_allocator

```
using namespace std;  
using namespace std::tr1;  
using namespace __gnu_cxx;
```

```
int my[65536];
```

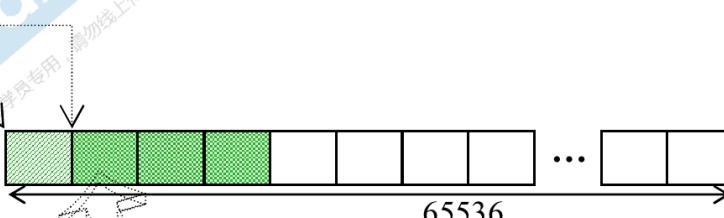
使用 std::tr1::array 或 std::array 均可.

```
array_allocator<int, array<int,65536>>  
myalloc(&my);
```

這種 array object
其內部將有 int _M_instance[65536];

```
int* p1 = myalloc.allocate(1);
```

```
int* p2 = myalloc.allocate(3);  
myalloc.deallocate(p1); //no-op  
myalloc.deallocate(p2); //no-op
```



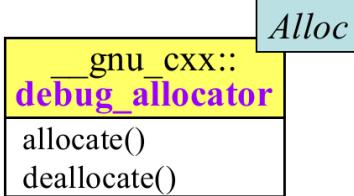
array_allocator 並不會
回收已給出的內存空間

G4.9 debug_allocator

```
template<typename _Alloc> <.../ext/debug_allocator.h>
class debug_allocator
{
    .....
private:
    size_type _M_extra; //通常是個 size_t;
    _Alloc _M_allocator;

    size_type _S_extra() {
        const size_t _obj_size = sizeof(value_type);
        return (sizeof(size_type) + _obj_size - 1) / _obj_size;
    } 計算 extra (bytes) 相當於幾個元素

public:
    debug_allocator(const _Alloc& __a)
        : _M_allocator(__a), _M_extra(_S_extra()) { }
    .....
```



```
pointer allocate(size_type __n) {
    pointer __res = _M_allocator.allocate(__n + _M_extra);
    size_type* __ps = reinterpret_cast<size_type*>(__res);
    *__ps = __n;
    return __res + _M_extra;
}

void deallocate(pointer __p, size_type __n) {
    using std::__throw_runtime_error;
    if (__p) {
        pointer __real_p = __p - _M_extra;
        if (*reinterpret_cast<size_type*>(__real_p) != __n)
            __throw_runtime_error
                ("debug_allocator::deallocate wrong size");
        _M_allocator.deallocate(__real_p, __n + _M_extra);
    } else
        __throw_runtime_error
            ("debug_allocator::deallocate null pointer");
}
.....
```

G2.9 容器使用的分配器不是 std::allocator 而是 std::alloc



```
template <class T,  
         class Alloc = alloc>  
class vector {  
    ...  
};
```

SGI style

→ //分配 512 bytes.
void* p = alloc::allocate(512);
//也可以這樣alloc().allocate(512);
alloc::deallocate(p,512);

```
template <class T,  
         class Alloc = alloc>  
class list {  
    ...  
};
```

```
template <class Key,  
         class T,  
         class Compare= less<Key>,  
         class Alloc = alloc>  
class map {  
    ...  
};
```

```
template <class T,  
         class Alloc = alloc,  
         size_t BufSiz = 0>  
class deque {  
    ...  
};
```

```
template <class Key,  
         class Compare = less<Key>,  
         class Alloc = alloc>  
class set {  
    ...  
};
```

而今安在哉 → __pool_alloc

```

class __pool_alloc_base
{
protected:
    enum { _S_align = 8 };
    enum { _S_max_bytes = 128 };
    enum { _S_free_list_size = (size_t)_S_max_bytes / (size_t)_S_align };

    union _Obj
    {
        union _Obj* _M_free_list_link;
        char      _M_client_data[1]; // The client sees this.
    };

    static _Obj* volatile _S_free_list[_S_free_list_size];
    // Chunk allocation state.

    static char* _S_start_free;
    static char* _S_end_free;
    static size_t _S_heap_size;
    ...
};

template<typename _Tp>
class __pool_alloc : private __pool_alloc_base
{ ... };

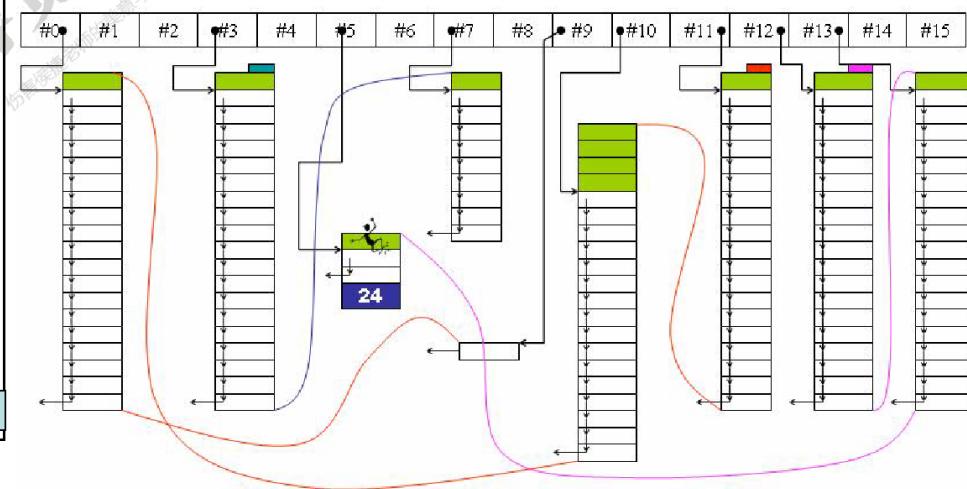
```

G4.9 <.../ext/pool_allocator.h>

G4.9 標準庫中有許多 extented allocators,
其中 `__pool_alloc` 就是 G2.9 `alloc` 的化身.

STD style

用例：
`vector<string,`
`__gnu_cxx::__pool_alloc<string>>`
`vec;`



— 侯捷 —

G4.9 __pool_alloc 用例

//欲使用 std::allocator 以外的 allocator, 必須自行 #include <ext/...>

```
#include <ext/pool_allocator.h>
```

```
template<typename Alloc>
```

```
void cookie_test(Alloc alloc, size_t n) <.....>
```

```
{
```

```
    typename Alloc::value_type *p1, *p2, *p3;
```

```
    p1 = alloc.allocate(n);
```

```
    p2 = alloc.allocate(n);
```

```
    p3 = alloc.allocate(n);
```

```
    cout << "p1= " << p1 << '\t' << "p2= " << p2
```

```
        << '\t' << "p3= " << p3 << '\n';
```

```
    alloc.deallocate(p1,sizeof(typename Alloc::value_type));
```

```
    alloc.deallocate(p2,sizeof(typename Alloc::value_type));
```

```
    alloc.deallocate(p3,sizeof(typename Alloc::value_type));
```

```
}
```



```
cout << sizeof(__gnu_cxx::__pool_alloc<int>) << endl; //1  
vector<int, __gnu_cxx::__pool_alloc<int>> vecPool;  
cookie_test(__gnu_cxx::__pool_alloc<double>(), 1);
```



p1= 0xae4138 p2= 0xae4140 p3= 0xae4148

相距 sizeof(double), 表示不帶 cookie

p1= 0xae25e8 p2= 0xaddeb50 p3= 0xadd090

相距 > sizeof(double), 並不表示就帶著 cookie.



```
cout << sizeof(std::allocator<int>) << endl; //1  
vector<int, std::allocator<int>> vec;  
cookie_test(std::allocator<double>(), 1);
```

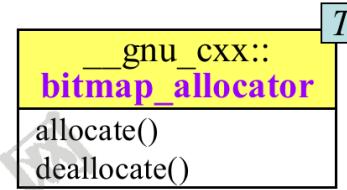


p1= 0x3e4098 p2= 0x3e4088 p3= 0x3e4078

相距(總是) == sizeof(double)=8, 可斷定帶著 cookie.

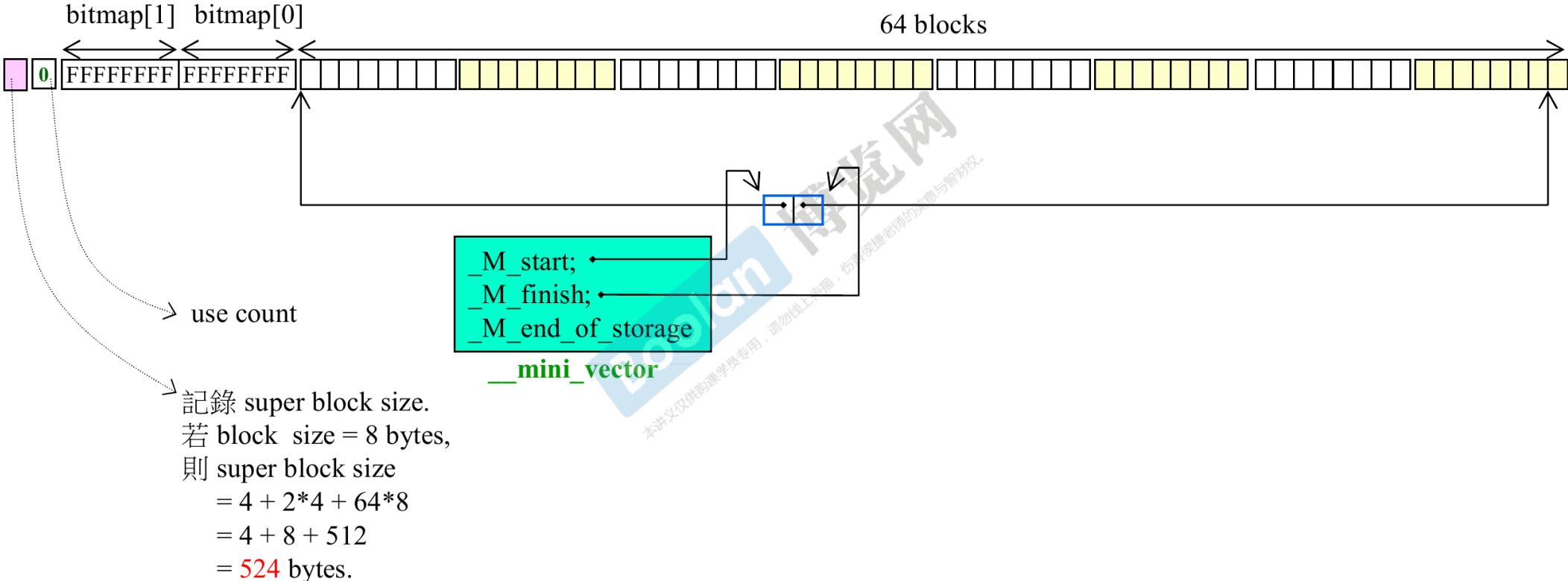
G4.9 bitmap_allocator

```
template<typename _Tp>           <.../ext(bitmap_allocator.h>
class bitmap_allocator : private free_list {
...
public:
    pointer allocate(size_type __n) {
        if (__n > this->max_size())
            std::__throw_bad_alloc();
        if (__builtin_expect(__n == 1, true))
            return this->_M_allocate_single_object();
        else {
            const size_type __b = __n * sizeof(value_type);
            return reinterpret_cast<pointer>(::operator new(__b));
        }
    }
    void deallocate(pointer __p, size_type __n) throw() {
        if (__builtin_expect(__p != 0, true)) {
            if (__builtin_expect(__n == 1, true))
                this->_M_deallocate_single_object(__p);
            else
                ::operator delete(__p);
        }
    }
}
```

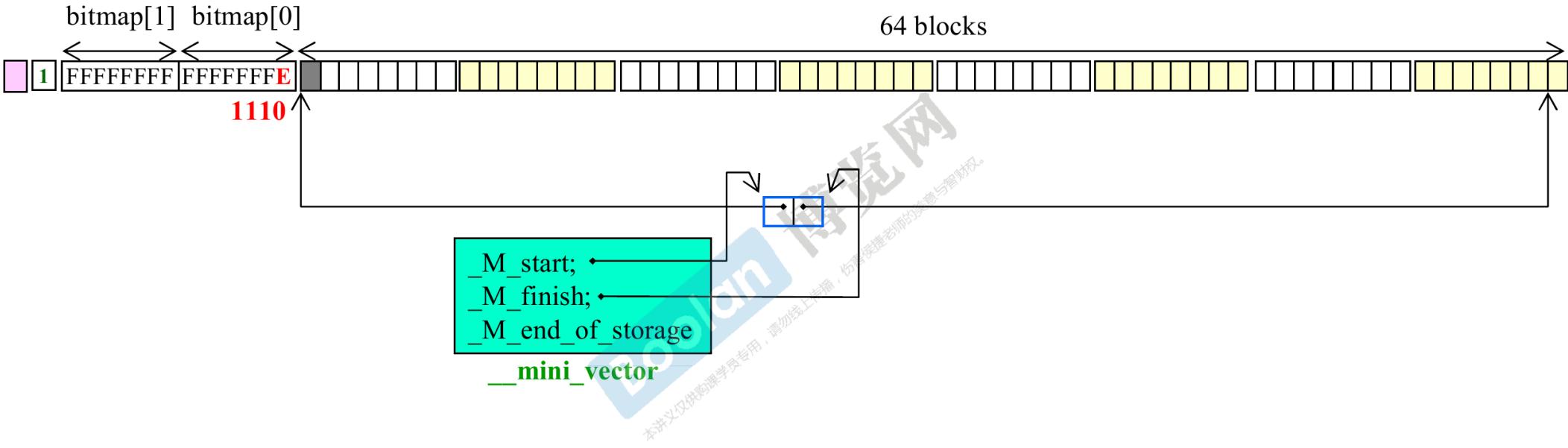


G4.9 bitmap_allocator,

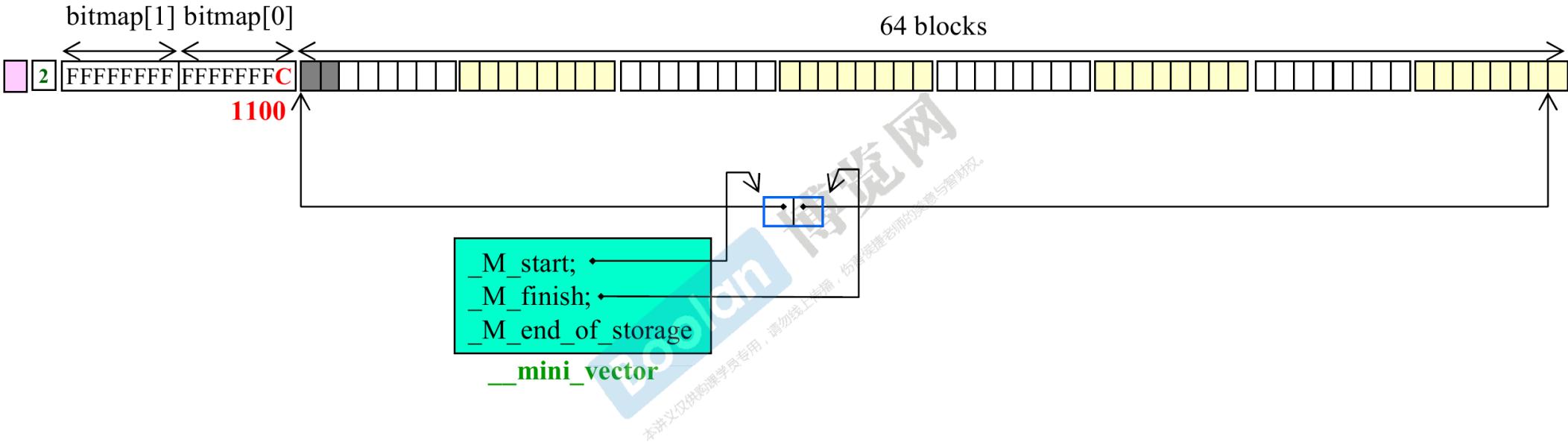
關於 blocks, super-blocks, bitmap, mini-vector



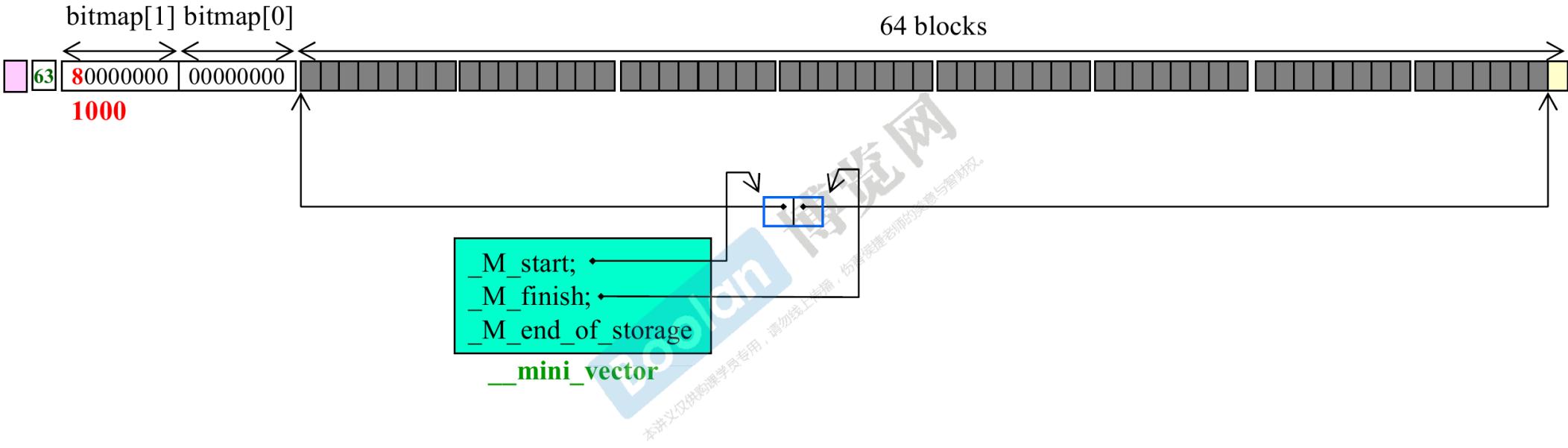
G4.9 **bitmap_allocator**,
關於 blocks, super-blocks, bitmap, mini-vector



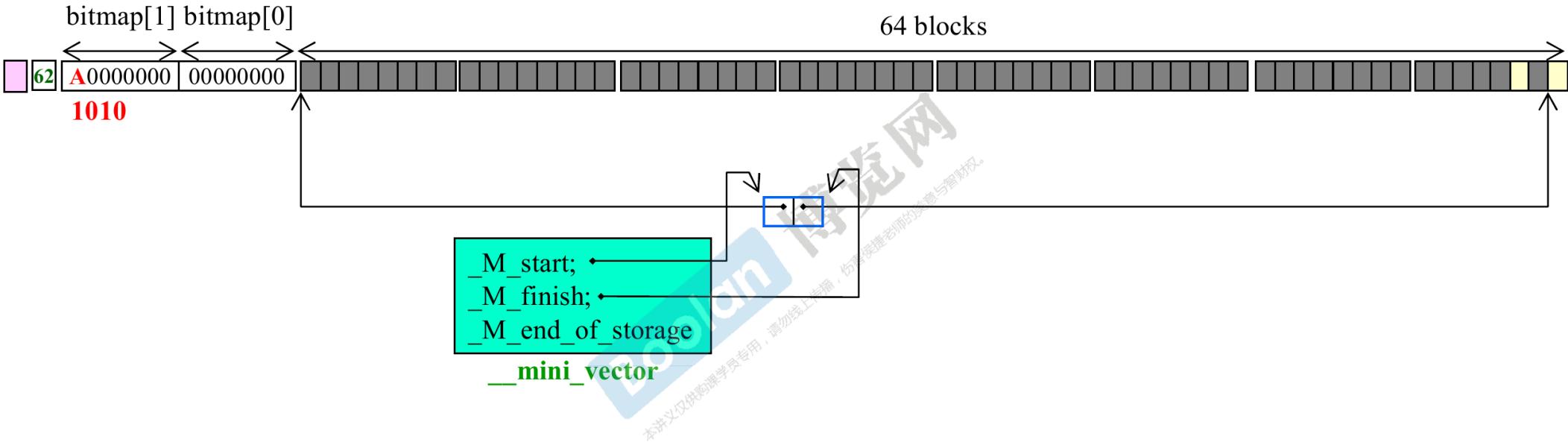
G4.9 bitmap_allocator, 關於 blocks, super-blocks, bitmap, mini-vector



G4.9 bitmap_allocator, 關於 blocks, super-blocks, bitmap, mini-vector



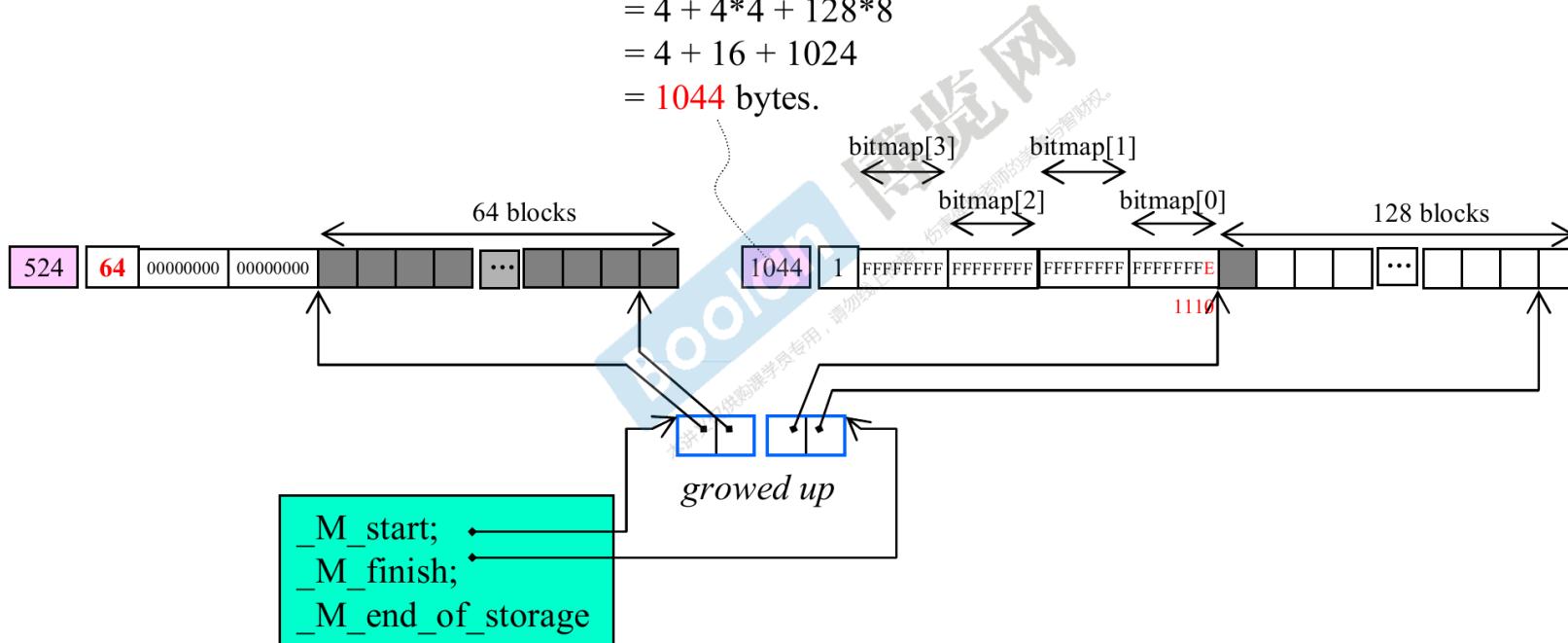
G4.9 bitmap_allocator, 關於 blocks, super-blocks, bitmap, mini-vector



G4.9 bitmap_allocator,

1st super-block 用罄, 啟動 2nd super-block

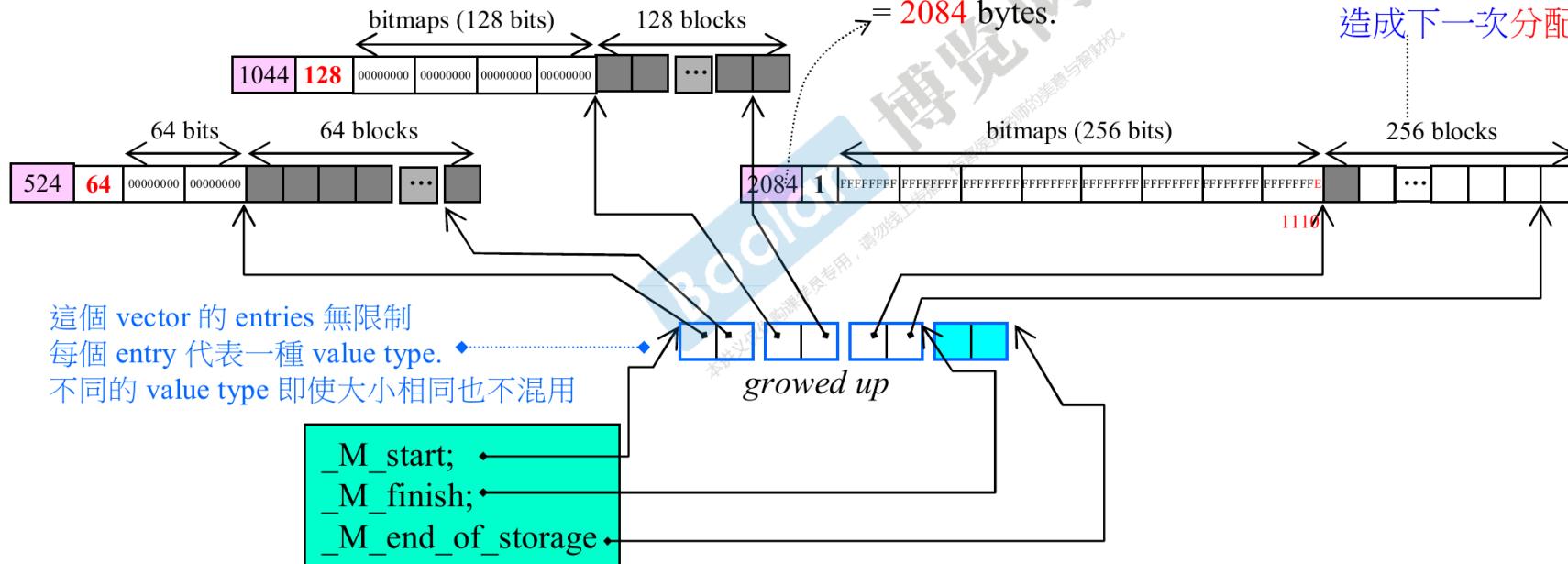
記錄 super block size.
若 block size = 8 bytes,
則 super block size
 $= 4 + 4*4 + 128*8$
 $= 4 + 16 + 1024$
 $= 1044$ bytes.



G4.9 bitmap_allocator, 2nd super-block 用罄, 啟動 3rd super-block

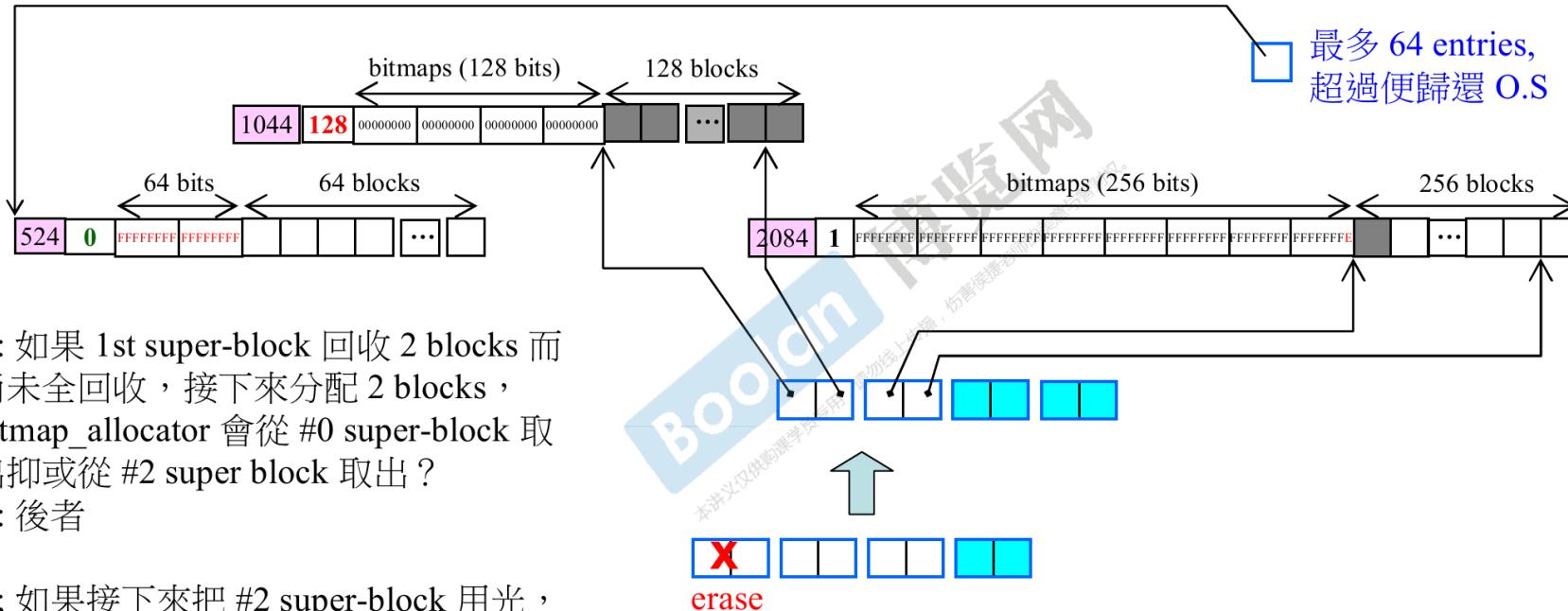
記錄 super block size.
 若 block size = 8 bytes,
 則 super block size
 $= 4 + 8*4 + 256*8$
 $= 4 + 32 + 2048$
 $= 2084$ bytes.

若不曾全回收, 則分配規模不斷倍增, 相當驚人. 每次全回收便造成下一次分配規模減半



G4.9 bitmap_allocator, 1st super-block 全回收

1st super block 全回收,
於是登錄到另一 vector
下次 分配規模 減半



Q: 如果 1st super-block 回收 2 blocks 而尚未全回收，接下來分配 2 blocks，bitmap_allocator 會從 #0 super-block 取出抑或從 #2 super block 取出？

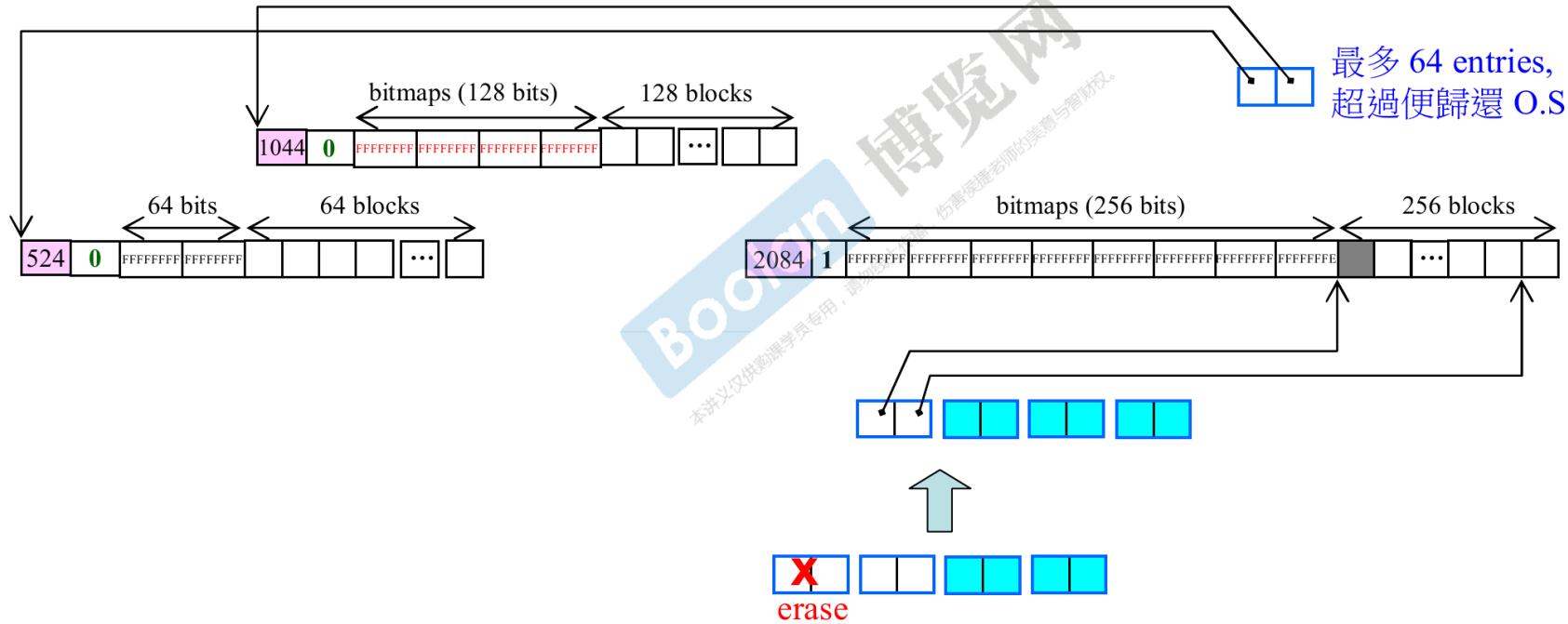
A: 後者

Q: 如果接下來把 #2 super-block 用光，然後分配 2 blocks，bitmap_allocator 會從 #0 super-block 取出抑或新建一個 #3 super block 並從中取出？

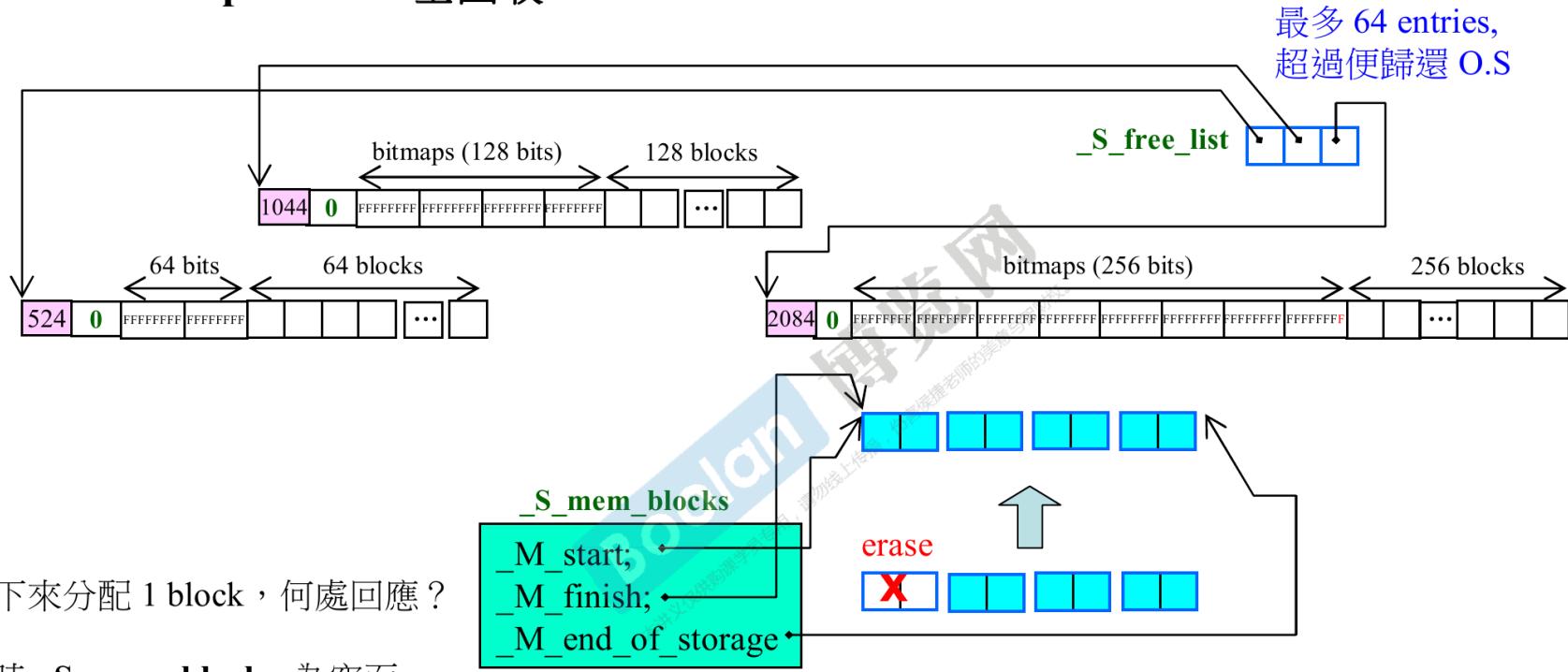
A: 前者

G4.9 bitmap_allocator, 2nd super-block 全回收

這個 vector 的元素排列將以 super block size 為依據. 因此當觸及 threshold (64), 新進者若大於最末者便直接 delete 新進者, 否則 delete 最末者然後再 insert 新進者. 若未觸及 64 則無條件 insert 至適當位置.



G4.9 bitmap_allocator, 3rd super-block 全回收

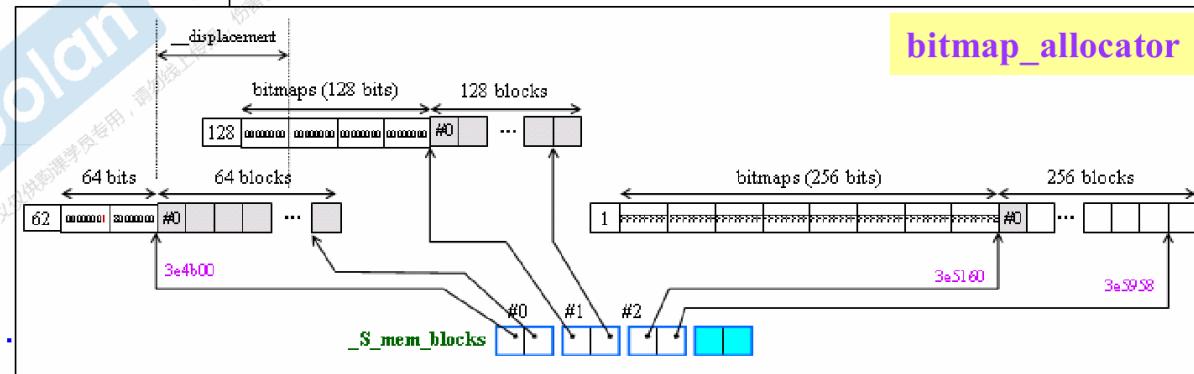
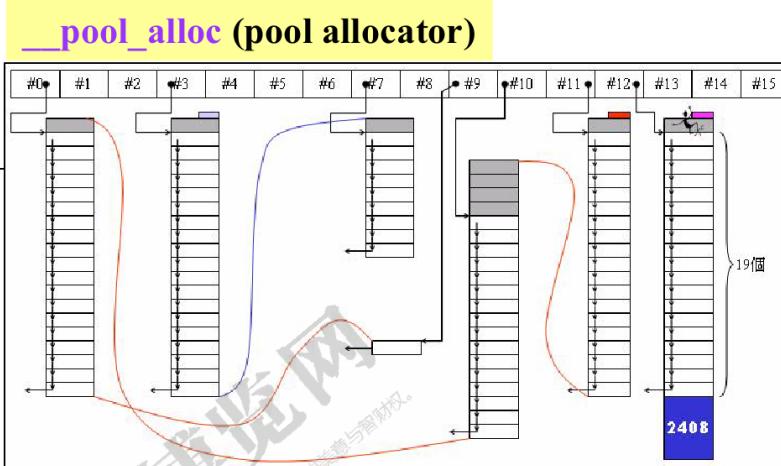


Q: 接下來分配 1 block，何處回應？

A: 此時 **_S_mem_blocks** 為空而 **_S_free_list** 有三個 super-blocks，於是取一個放進 **_S_mem_blocks**，然後遵循先前法則完成分配。

使用 G4.9 分配器

```
916 #include <list>
917 #include <stdexcept>
918 #include <string>
919 #include <cstdlib>      //abort()
920 #include <cstdio>        //snprintf()
921 #include <algorithm>    //find()
922 #include <iostream>
923 #include <ctime>
924
925 #include <cstddef>
926 #include <memory>        //內含 std::allocator
927 //欲使用 std::allocator 以外的 allocator, 得自行 #include <ext>...
928 #include <ext\array_allocator.h>
929 #include <ext\mt_allocator.h>
930 #include <ext\debug_allocator.h>
931 #include <ext\pool_allocator.h>
932 #include <ext\bitmap_allocator.h>
933 #include <ext\malloc_allocator.h>
934 #include <ext\new_allocator.h>
935 namespace jj20
936 {
937 void test_list_with_special_allocator()
938 {
939 cout << "\ntest_list_with_special_allocator()....."
940
941 list<string, allocator<string>> c1;
942 list<string, __gnu_cxx::malloc_allocator<string>> c2;
943 list<string, __gnu_cxx::new_allocator<string>> c3;
944 list<string, __gnu_cxx::__pool_allocator<string>> c4;
945 list<string, __gnu_cxx::__mt_alloc<string>> c5;
946 list<string, __gnu_cxx::bitmap_allocator<string>> c6;
```



使用 G4.9 分配器

```
948 int choice;
949 long value;
950
951     cout << "select: ";
952     cin >> choice;
953     if (choice != 0) {
954         cout << "how many elements: ";
955         cin >> value;
956     }
957
958 char buf[10];
959 clock_t timeStart = clock();
960 for(long i=0; i< value; ++i)
961 {
962     try {
963         sprintf(buf, 10, "%d", i);
964         switch (choice)
965         {
966             case 1 : c1.push_back(string(buf)); break;
967             case 2 : c2.push_back(string(buf)); break;
968             case 3 : c3.push_back(string(buf)); break;
969             case 4 : c4.push_back(string(buf)); break;
970             case 5 : c5.push_back(string(buf)); break;
971             case 6 : c6.push_back(string(buf)); break;
972             default: break;
973         }
974     }
975     catch(exception& p) {
976         cout << "i=" << i << " " << p.what() << endl;
977         abort();
978     }
979 }
980
981 cout << "a lot of push_back(), milli-seconds : "
982     << (clock()-timeStart) << endl;
```

```
989 //test all allocators' allocate() & deallocate();
990 int* p;
991 allocator<int> alloc1;
992 p = alloc1.allocate(1);
993 alloc1.deallocate(p,1);
994
995 __gnu_cxx::malloc_allocator<int> alloc2;
996 p = alloc2.allocate(1);
997 alloc2.deallocate(p,1);
998
999 __gnu_cxx::new_allocator<int> alloc3;
1000 p = alloc3.allocate(1);
1001 alloc3.deallocate(p,1);
1002
1003 __gnu_cxx::__pool_allocator<int> alloc4;
1004 p = alloc4.allocate(2);
1005 alloc4.deallocate(p,2);
1006
1007 __gnu_cxx::__mt_allocator<int> alloc5;
1008 p = alloc5.allocate(1);
1009 alloc5.deallocate(p,1);
1010
1011 __gnu_cxx::bitmap_allocator<int> alloc6;
1012 p = alloc6.allocate(3);
1013 alloc6.deallocate(p,3); }
```

The End

