

## 第四章 彌補預測誤差的資源配置方法與最佳化模型

本章針對彌補預測誤差的資源配置方法中的中央保留資源法與超額分配法做深入研究，並以圖形範例與最佳化模型來分析兩個不同的資源配置方法。

### 4.1 中央保留資源法

圖 4.1 以單一 Ingress Router 提出對某一鏈路(link)頻寬之需求為例， $\theta$  為資源需求量， $\tau$  為資源配置量， $C$  為容量(Capacity)，Ingress Router 對 BB 提出某一鏈路資源配置需求( $\theta$ )，BB 保留了部份資源於 Central Pool 中來彌補預測誤差，僅配置部分 Ingress Router 需求的資源( $\tau$ )。如圖所示，BB 對 Ingress Router 的需求並不充分滿足( $\tau < \theta$ )，餘下的資源( $C - \tau$ )則留至 Central Pool 中即時配置給有需要的 Ingress Router 提供服務。

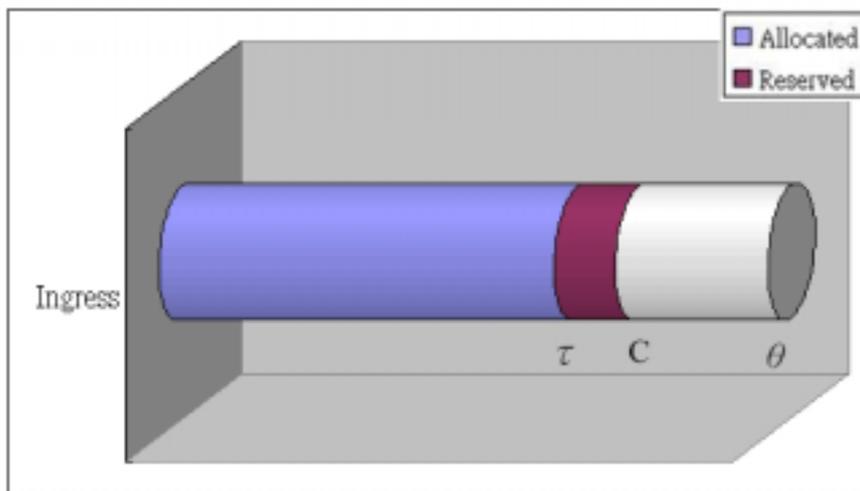


圖 4.1：單一 Ingress Router 資源保留與配置範例

再以圖 4.2 說明，圖中左右兩長條圖表示兩個 Ingress Router 的資源需求、配置與實際進入的訊務。比較實際進入兩個不同 Ingress Router 的訊務與其資源

批購量位置( $\theta$ )可知，兩個 Ingress Router 的預測都有誤差，欲進入 Ingress Router 1 (I1)的訊務比批購的資源還多，但進入 Ingress Router 2 (I2)的卻比批購的還少。假設 BB 剛好可以對每個 Ingress Router 十足配置其所批購的資源，因為進入 I1 的訊務超過預期( $\theta$ )，超過的部份會被 Ingress Router 拒絕，而 I2 卻有許多資源是浪費沒有使用。

中央保留資源法可以彌補這種缺陷，BB 保留部份的資源統籌分配，以即時配置的方式提供資源給有需要的 Ingress Router，必要時可以不完全滿足所有 Ingress Router 的需求。比較圖 4.3 利用 Central Pool 保留資源和圖 4.2 沒有 Central Pool 保留資源，相同的情況，I1 實際發生欲進入網域的訊務遠超過預測，而 I2 卻少於預測。圖 4.3 中 BB 對每個 Ingress Router 不足額配置資源，將 I1 與 I2 沒有配置出去的資源( $\theta - \tau$ )集中於 Central Pool 中，BB 會在執行時期將這部份保留的資源提供給 I1 使用(I1 斜線部分，on-demand acquired bandwidth)，如此一來，整體資源分配出去而沒有使用的部分(wasted bandwidth)會大大地降低，也減低因為預測誤差所造成資源的錯置。

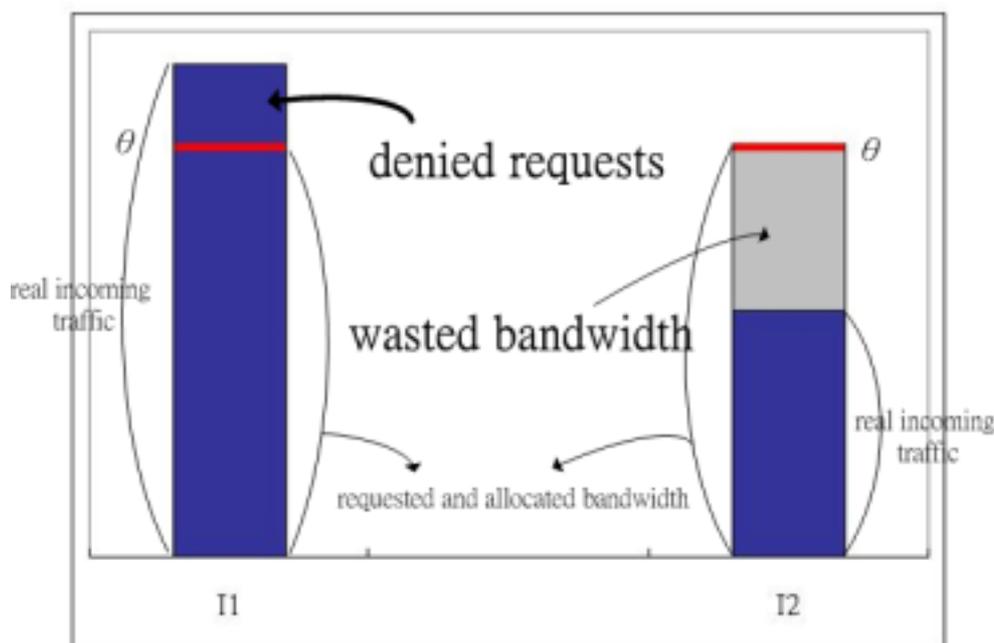


圖 4.2：頻寬需求、頻寬配置與實際訊務示意圖

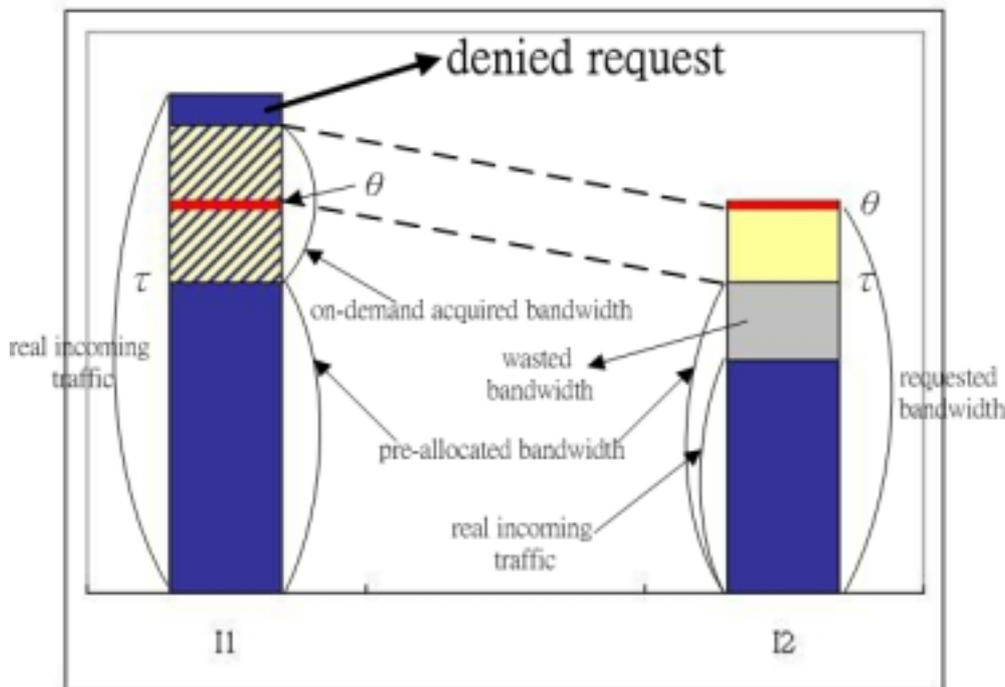


圖 4.3：中央保留資源法的資源保留與配置

#### 4.1.1 資源配置量與使用分析

中央保留資源法是保留部份的網路資源，當 Ingress Router 批購的資源不足時，Ingress Router 可向 BB 即時要求配置資源以服務新進的訊務，BB 檢視 Central Pool 中所保留的資源存量，提供給 Ingress Router 使用。然而這一部分的工作勢必造成網路負荷的增加，如同先前所提，當 BB 保留過多的資源在 Central Pool 中時，可以使網路資源的使用率相對提高，然而保留太多資源會使中央資源保留法退化成即時配置，失去事先規劃的優點，增加系統負荷；反之，如果 BB 將大部分的資源批發出去，當預測越不準確，Central Pool 中可以提供即時服務的資源又不足，此法彌補預測誤差的能力就大為降低。

圖 4.4 舉例說明了資源配置量與整體資源的使用情況。 $\theta$  為 Ingress Router 對 BB 所提出的預測資源需求量，實體部份表示實際發生欲進入網域的資料量， $\tau$  為 BB 配置給 Ingress Router 使用的資源量，在此分別討論四個不同配發

量( $\tau_1$   $\tau_4$ )發生的情況。

$\tau_1$  配置超過進入訊務的資源，由圖中可以看出，批發出去的資源並沒有被 Ingress Router 使用完，最上方空白部分明顯的浪費了。

$\tau_2$  配置量接近進入網域的資料量，不會有過多即時資源配置的需求增加系統與網路的負荷，也不會因配置過多的資源造成浪費。

$\tau_3$  配置量不敷使用，有部分的資料傳送需要即時要求配置資源。

$\tau_4$  表示沒有預先配置的資料量，換句話說，Ingress Router 傳送所有訊務所需要的資源都是以即時配置的方式取得，在大型網路或網域資料量大時，系統可能會沒有辦法負荷，網路的效能也可能會因負荷過高而大大降低。

由上面的分析可以知道，BB 所批發的資源量儘可能接近進入網域的資料量，避免不必要的浪費，且不會造成系統與網路過大的負擔，加長系統的反應時間。

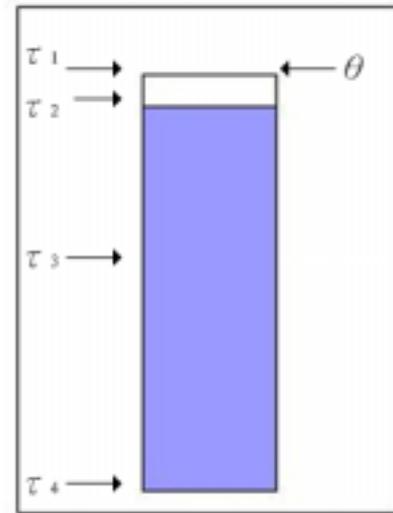


圖 4.4：資源需求與配置量

#### 4.1.2 最佳化模型

使用中央保留資源法，Ingress Router 需求的資源可以分成批發預售( $\tau$ )和即時需求( $> \tau$ )兩部份。而即時需求又可分為合乎預期( $\theta$ )與超出預測( $> \theta$ )兩部份。

如果 Ingress Router 需要使用超過  $\theta$  的資源，表示 Ingress Router 的預測過低，導致 Ingress Router 需要以較高價格臨時批購資源來服務使用者。若 BB 並未十足分配 Ingress Router 所提出預購的資源，導致 Ingress Router 必須即時要求資源配置，反而增加網路額外的負擔，因此必須防止 BB 為了增加即時批購

的獲利而故意不足額分配資源，使得網路負荷增加。其價格設計如下，如果 Ingress Router 向 BB 預定資源但未獲得足夠的配置，當 Ingress Router 需要使用超過 BB 所配置的資源( $\tau$ )，且還是少於預先要求配置的資源量( $\theta$ )時，這部份的資源雖然 Ingress Router 需要即時向 BB 要求資源配置，但 BB 此時須以低於預購的價格提供資源。

由上述說明可知，中央保留資源法的目標是以保留部份資源的方式，增加些許 BB 與網路的負荷來降低資料流被拒絕服務的比例、提升網路資源的使用率與系統業者的獲利。假設某一段欲預測的時段(CTP)的流量機率分佈與其參考時段(RTP)的流量機率分佈相同，我們可以利用 RTP 的歷史統計分佈來預測 CTP 的流量。

以單一 Ingress Router 為例，圖 4.5 為某 Ingress Router 對某一鏈路(link)的資源需求機率分佈圖，資源配置量與資源價格的關係圖。Ingress Router 預測可能進入網路的資料量，BOA 經過計算後向 BB 批購資源( $\theta$ )，BB 經過整體考量後配置資源給 Ingress Router ( $\tau$ )。

表 4.1 為中央保留資源法中所使用的符號，參考圖 4.5， $C_1$  為預購每單位價格， $C_2$  為臨時批購單位價格，較特別的  $C_1'$  表示使用已向 BB 預訂資源但未獲得配置的部分。其中  $C_2$ 、 $C_1'$  兩部分的資源需即時向 BB 要求資源配置來提供服務。根據前述說明， $C_1$ 、 $C_1'$ 、 $C_2$  應符合  $C_2 < C_1 < C_1'$  關係方為合理。

表 4.1：中央保留資源法符號表

$\theta$	Required total bandwidth
$\tau$	Committed total bandwidth
$C_1$	Per unit cost of committed bandwidth
$C_2$	Per unit cost of on-demand allocated bandwidth without pre-order
$C_1'$	Per unit cost of on-demand allocated bandwidth with pre-order
P	Expected profit

$p_i$	Potential probability of incoming traffic $i$
$N$	Maximum possible incoming traffic

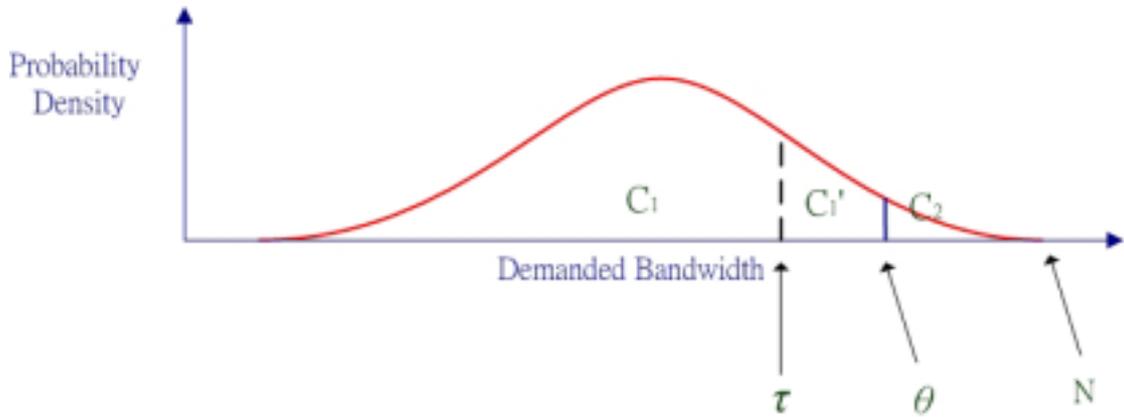


圖 4.5：資源需求機率分佈、資源配置量與資源價格關係圖

保留於 Central Pool 中使用的資源收入期望值為( $\tau$  至  $\theta$  每單位為  $C_1'$ 、 $\theta$  至  $N$  每單位為  $C_2$ )：

$$\sum_{i=\tau}^{\theta} p_i \times C_1' \times (i - \tau) + \sum_{i=\theta}^N p_i \times [C_2 \times (i - \theta) + C_1' \times (\theta - \tau)] \quad (4-1)$$

所以總收入期望值  $P$  為式子 4-1 加上預先分配資源所獲收入  $C_1 \times \tau$ ：

$$P = C_1 \times \tau + \sum_{i=\tau}^{\theta} p_i \times C_1' \times (i - \tau) + \sum_{i=\theta}^N p_i \times [C_2 \times (i - \theta) + C_1' \times (\theta - \tau)] \quad (4-2)$$

BB 綜合所有 Ingress Router 所提出的預訂需求做事先規劃，保留部份資源以即時配置的方式來彌補資源需求的預測誤差，即時配置的資源，需減少以  $C_1'$  價格賣出資源的可能，提高獲。由上可知，影響中央保留資源法獲利結果在於兩項因素：

- 保留統籌資源較為有利之條件
- 最佳資源保留量

等式 4-2 為一個  $\tau$  的函數，求最佳解時  $\sum_{i=\tau}^{\theta} p_i \times C_1' \times (i - \tau)$  部份的複雜度為  $O(\tau^2)$ ， $C_1 \times \tau$  與  $\sum_{i=\theta}^N p_i \times [C_2 \times (i - \theta) + C_1' \times (\theta - \tau)]$  的複雜度為  $O(\tau)$ ，不同情況下之最佳解可以簡單的數學方法求出。

## 4.2 超額分配法

圖 4.6 至 4.8 可以幫助瞭解超額分配法的概念。圖 4.6 上面兩個長條圖中， $\theta$  與  $\tau$  分別表示兩個 Ingress Router ( $I_1$ 、 $I_2$ ) 的需求與 BB 的配置，下方的長條圖表示鏈路的資源使用情況，由圖可以看出，BB 分配出去的資源超過鏈路的容量(Capacity,  $C$ )，超額分配的資源為 $(\tau_1 + \tau_2 - C)$ 。

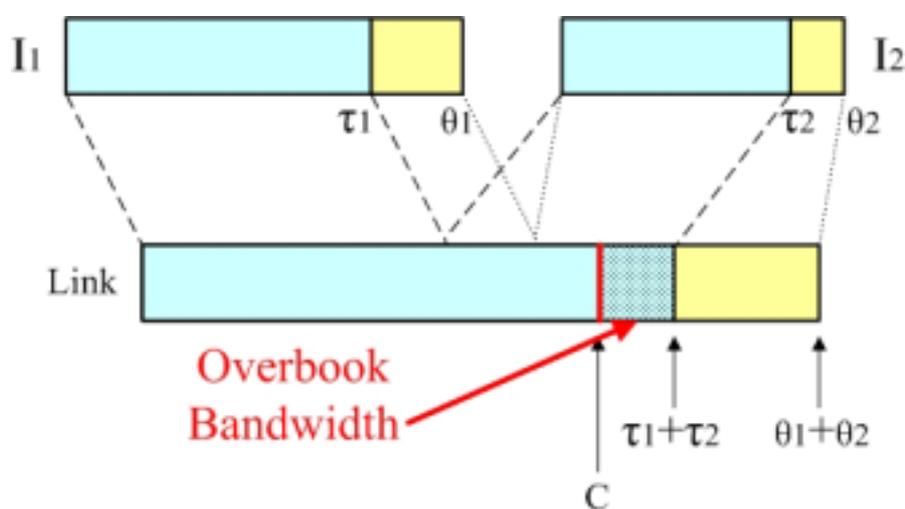


圖 4.6：超額配置範例

再以圖 4.7 和 4.8 的兩個例子說明。圖 4.7 是一個沒有超額配置的例子，Ingress Router 的資源需求預測總合 $(\tau_1 + \tau_2)$ 超過鏈路容量( $C$ )，所以 BB 僅能配置總合為  $C$  的資源給兩個 Ingress Router。兩個 Ingress Router 進入的訊務量為最上方的長條圖，由於預測上的誤差， $I_2$  浪費了許多配置的資源，但是  $I_1$  卻有許多的訊務被拒絕服務。相較於圖 4.8，BB 超額配置資源給兩個 Ingress Router，使得更多的訊務可以被允入，所以相同的情況下，因  $I_2$  預測誤差可能造成的浪費可以被  $I_1$  多允入的部份所彌補。所以只要允入的訊務不超過網路的負荷，造成網路的超載，資源的使用率是可以有效提升，系統業者也可因超賣而獲得更多的收入並維持網路的品質。

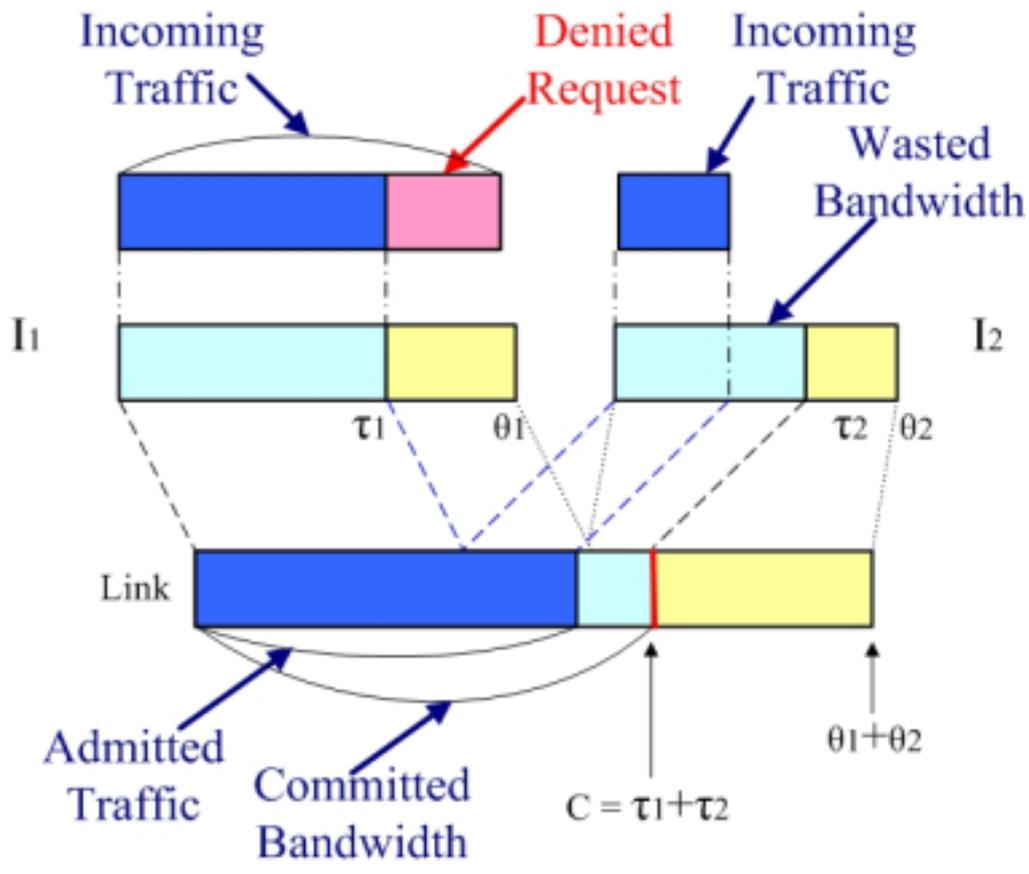


圖 4.7：沒有使用超額分配法的資源使用情況

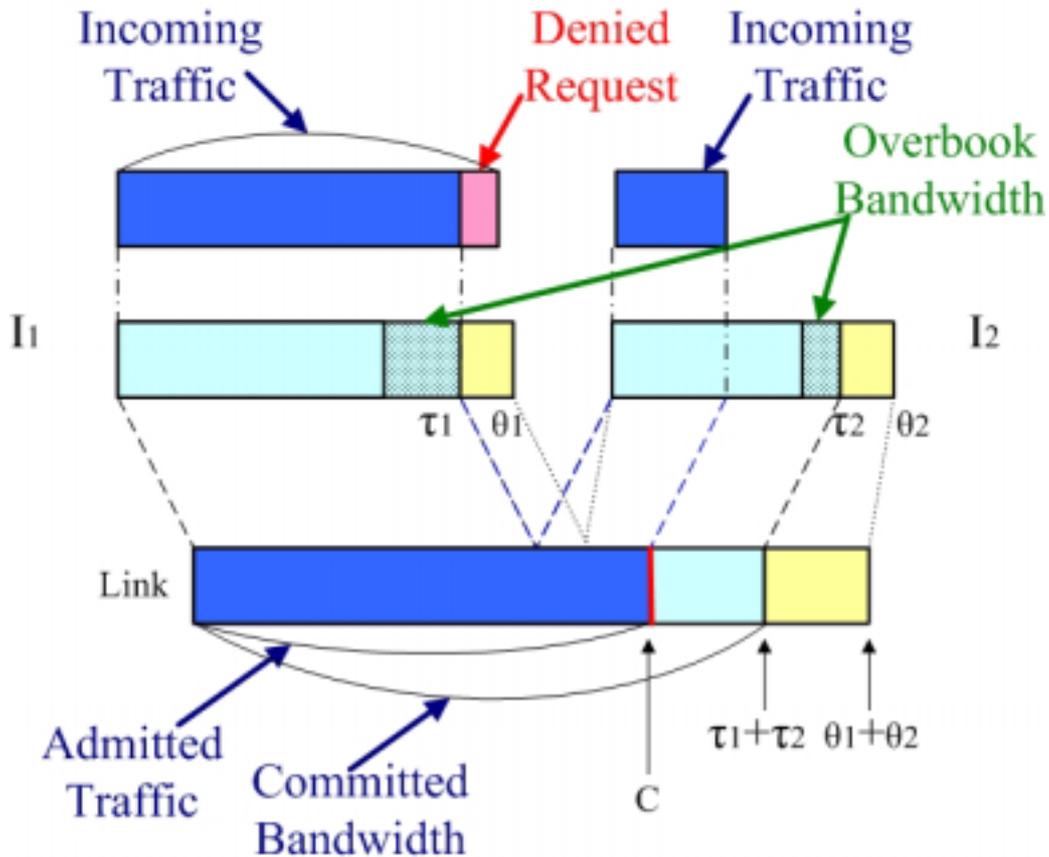


圖 4.8：使用超額分配法的資源使用情況

#### 4.2.1 資源配置與需求資源分佈分析

圖 4.9 為超額分配的比率和訊務遺失(packet loss)的比率間可能的關係，如圖所示，若超額分配過量使網路湧進過多資料而無法處理時，網路的擁塞情況會使得系統品質下降，嚴重時甚至會癱瘓整個網路。

我們的計價模型應加入品質下降所導致之損失，如圖 4.10，當超額分配量過高，獲利極可能會快速下降。

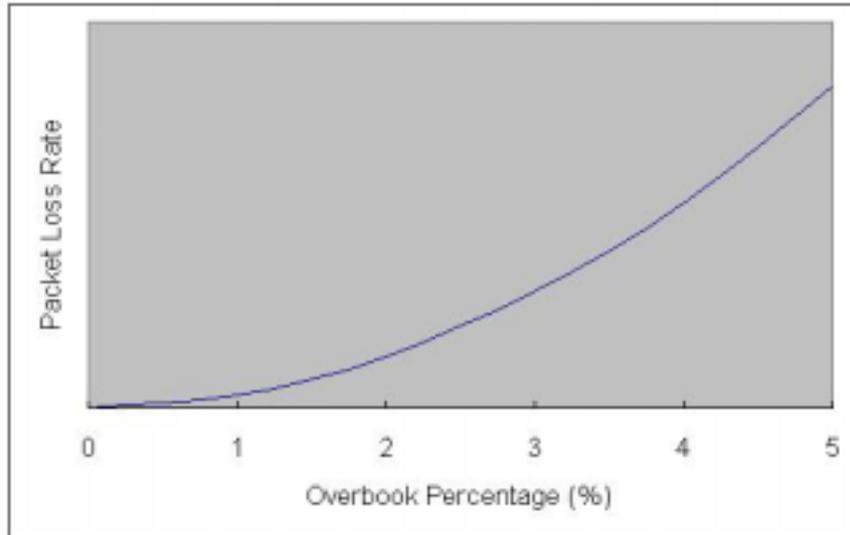


圖 4.9：超額分配與訊務遺失

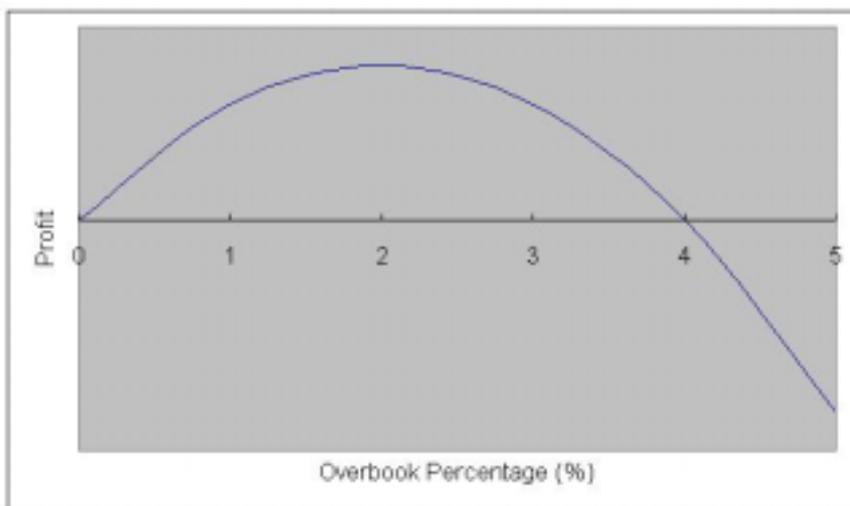


圖 4.10：超額分配與系統獲利

BB 接收所有 BOA 所提出的資源配置需求，統籌規劃整個網路的資源配置，並利用超額分配的方法來增加整體允入的資源量，提升網路資源的使用率。以整個網路而言，超額配置資源給各個不同的 BOA，在截長補短之後，從各個 Ingress Router 實際進入網路的資料流流量總和是可能小於預測的總和，所以在資源配置上利用超額分配的技術，本研究的研究目標在於尋求頻寬配置之最佳超額分配值。

最佳的超額分配必須避免太大膽的超額配置，降低網路品質，也要避免太保守的超額配置，使得網路資源使用率無法提升。網路供應商可以賺取超賣的頻寬費用，但也必須承擔因超賣所造成的品質降低所導致之損失。

以下以圖 4.11 至 4.14 為例說明核心網路中某一個鏈路上訊務需求及訊務允入之機率變化。圖 4.11 為核心網路上某條鏈路上單一 Ingress Router 頻寬需求的機率分佈圖。根據歷史統計資料，Ingress Router 會預測某一段欲預測的時段 (CTP)內可能進入網域的訊務量，經過計算後 Ingress Router 對 BB 提出資源配置的請求，包含網域上每個鏈路的資源需求，並將 Ingress Router 頻寬需求的機率分佈圖提供於 BB 參考，以便做出配置決定。

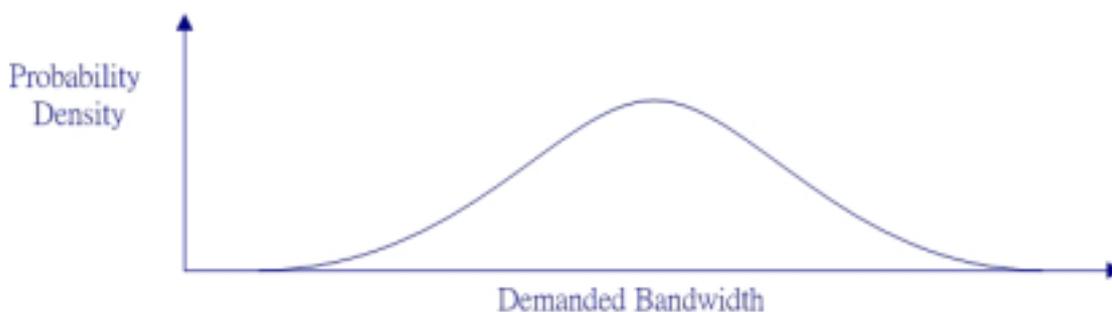


圖 4.11：單一 Ingress Router 需求資源機率分佈圖

當 BB 決定 Ingress Router 的資源配置量後，實際可以進入網路的資料量會受到資源配置量的限制，資源需求的機率分佈情況也會隨著變動。如圖 4.12 所示，超過資源配置量的資料皆不被允許進入網路，所以其發生的機率會累積於最大允許進入的位置，這一部份的機率是由後方超過資源配置量的機率累積所得，即圖中於最大配置量的虛線長度為後方面積所得，機率分佈圖由圖 4.11 變成圖 4.12。

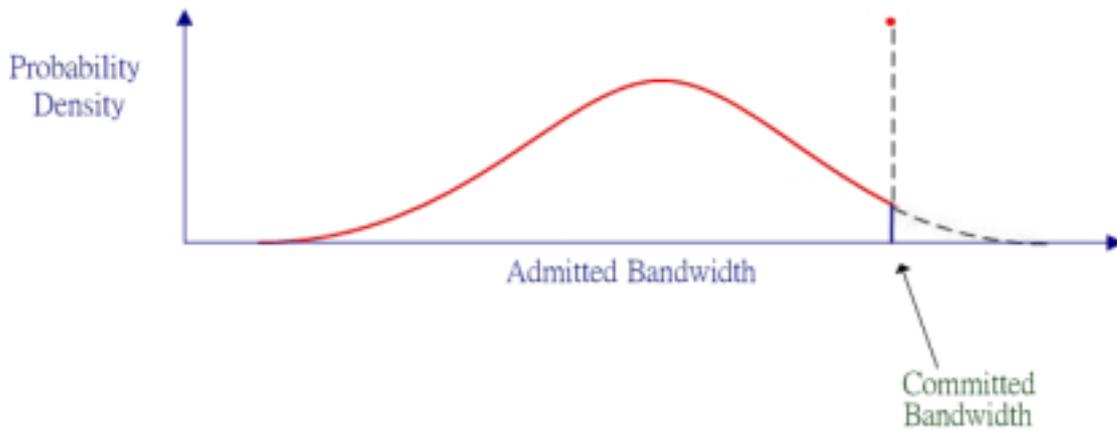


圖 4.12：單一 Ingress Router 受到資源配置量的需求資源機率分佈圖

同一條鏈路上各個 Ingress Router 需求總合的機率分佈可能會如圖 4.13 所示一般。



圖 4.13：需求資源總合機率分佈圖

當 BB 決定所有 Ingress Router 的資源配置量後，由於每個 Ingress Router 被允許使用的資源有限，所以資源需求的分佈圖形也會因而改變，在尾端的部分會有高量區的產生，如圖 4.14。



圖 4. 14：受到資源配置總量限制之資源需求總合機率分佈圖

#### 4.2.2 最佳化模型

表 4.2 為超額分配法最佳化模型所使用的符號說明表，模型中以單一 Ingress Router 舉例並配合圖形說明，需求分佈分為連續函數與不連續函數，模型介紹以連續函數為例。

$\tau$  為 BB 配置給某一 Ingress Router 的資源量(Booking Level)，也就是此 Ingress Router 可以允許進入的最大訊務量(committed bandwidth)，當此 Ingress Router 有超過  $\tau$  的資料想要通過此條鏈路時，由於最大訊務量的限制，Ingress Router 最多只能允許  $\tau$  的資料量通過，即  $\tau$  為此 Ingress Router 在此鏈路上可使用資源的上限。

表 4. 2：超額分配法符號表

$\theta$	Required total bandwidth
$\tau$	Committed total bandwidth
$m_1$	Per unit overbooked bandwidth
$m_2$	Per unit packet loss penalty
R	Net profit
$p(b)$	Potential probability of incoming traffic b

N	Maximum possible incoming traffic
$h(b)$	Packet loss function of committed bandwidth $b$
C	Link capacity

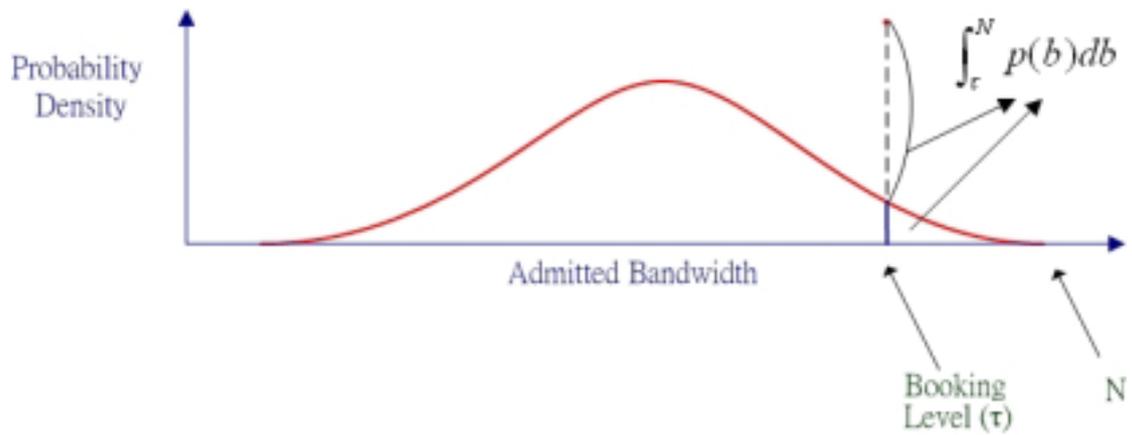


圖 4.15：超過最大訊務量( $\tau$ )的機率總合

如同之前所提，圖 4.15 中虛線的高度為後方由  $\tau$  至  $N$  的實體面積。由於每個 Ingress Router 上最大訊務量的限制，超過  $\tau$  的資料流只能有  $\tau$  的量可以進入網路，所以超過  $\tau$  的機率(實體面積)會累積於  $\tau$ 。所以分佈圖區線會終結於  $\tau$  點， $\tau$  點機率為原分佈圖  $\tau$  點的機率加上上圖後方面積，如圖 4.16 所示。

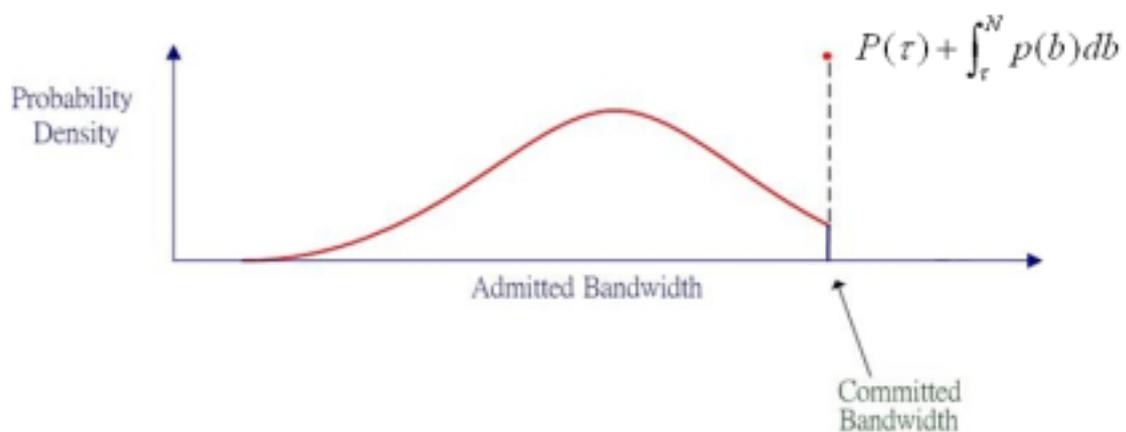


圖 4.16：需求資源為最大訊務量( $\tau$ )之機率

系統收入是以批購出去的頻寬計算，為了避免系統過份超額配置頻寬，使

得 Ingress Router 允許過多的資料進入網路，造成網路品質惡化，因此必須計算因超額配置所造成的品質惡化來訂定損失。其次，我們計算因超額分配法所得的獲利和損失。

如圖 4.17 之例所示，假設獲利與所配置的資源成正比，但是因進入過多資料導致爭奪傳送資源而造成的品質惡化卻會隨著超額配置的資源增加而大幅上升。

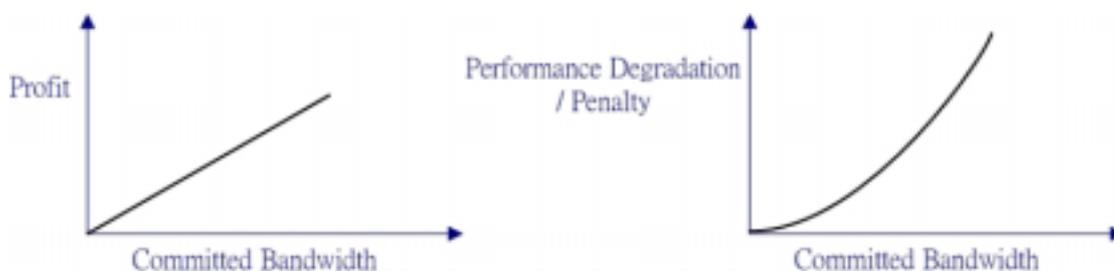


圖 4.17：獲利與品質惡化損失趨勢示意圖

$P_1$  為因 Overbook 而多賣出  $\tau - C$  的頻寬所獲利益，每單位價格為  $m_1$

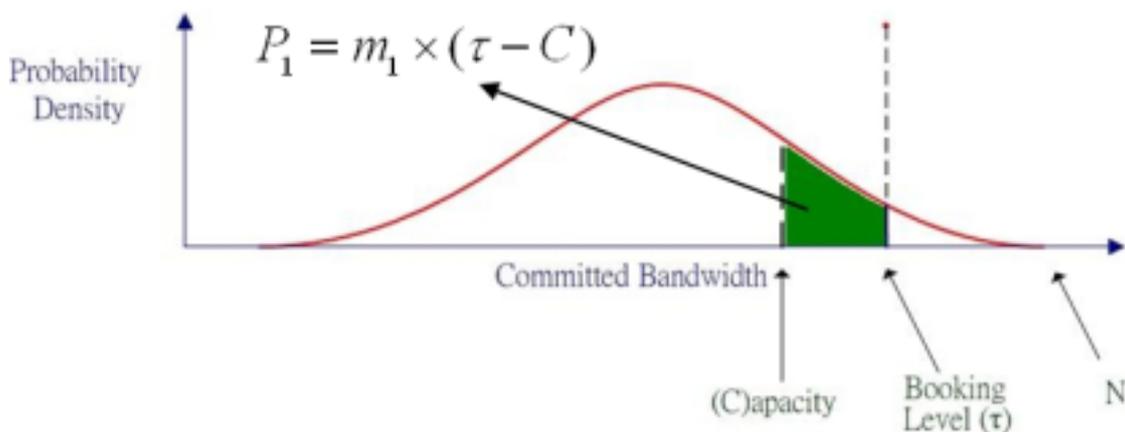


圖 4.18：超額配置獲利

$P_2$  為因超額配置所造成的損失期望值，每單位為  $m_2$

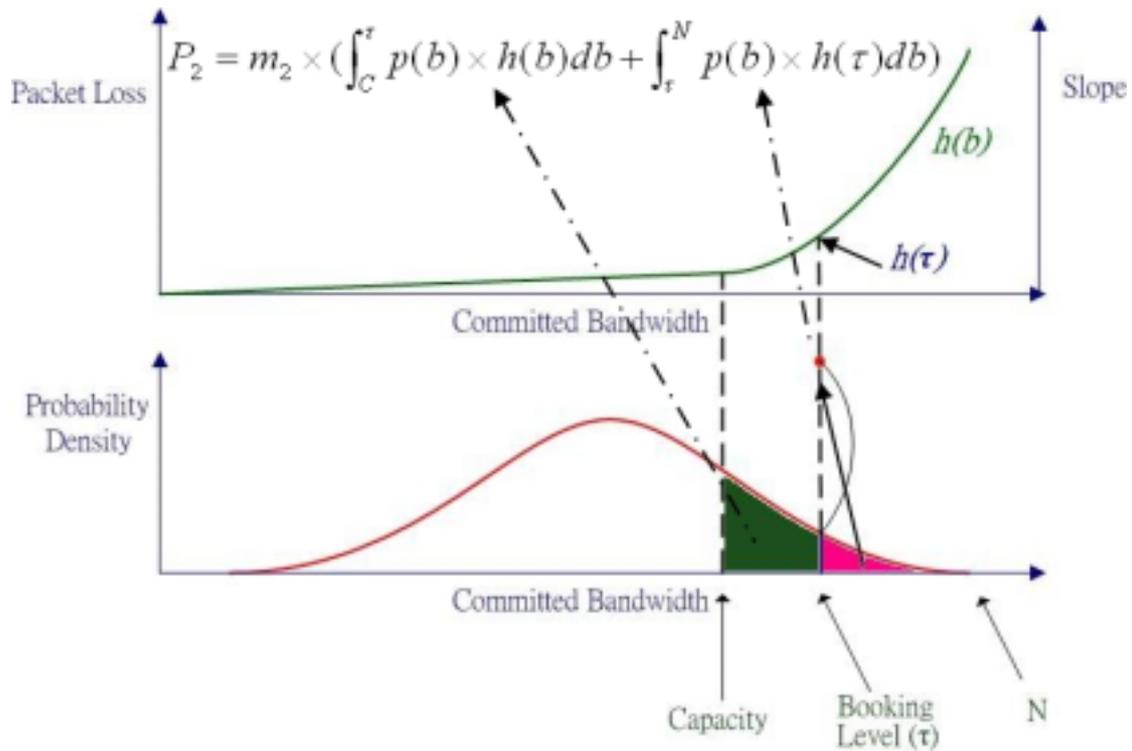


圖 4.19：超額配置賠償

淨利為  $R = P_1 - P_2$

$$R = m_1 \times (\tau - C) - m_2 \times \left[ \int_C^\tau p(b) \times h(b) db + \int_\tau^N p(b) \times h(\tau) db \right] \quad (4-3)$$

令  $L(b) = p(b) \times h(b)$ ：

$$R = m_1 \times (\tau - C) - m_2 \int_C^\tau L(b) db - m_2 \times h(\tau) \times \int_\tau^N p(b) db \quad (4-4)$$

再對  $R$  做一次微分，找出最佳配置值發生的地方：

$$R' = m_1 - m_2 \times L(\tau) - m_2 \times h'(\tau) \times \int_\tau^N p(b) db + m_2 \times h(\tau) \times p(\tau) \quad (4-5)$$

$$= m_1 - m_2 \times L(\tau) - m_2 \times h'(\tau) \times \int_\tau^N p(b) db + m_2 \times L(\tau) \quad (4-6)$$

$$= m_1 - m_2 \times h'(\tau) \times \int_\tau^N p(b) db \quad (4-7)$$

當等式 4-7 為零時，以簡單數學方法計算找到最佳的 Booking Level ( $\tau$ ) 使得因超額配置所帶來的收入最高。

此外當等式 4-7 為零時，可以得到下式：

$$\frac{m_1}{m_2} = h'(\tau) \times \int_{\tau}^N p(b)db \quad (4-8)$$

表示  $m_1$  與  $m_2$  的比值會等於訊務遺失的函數  $h(b)$  在  $\tau$  點的斜率乘上  $\tau$  點發生的機率，如同下圖 4.20 所示。

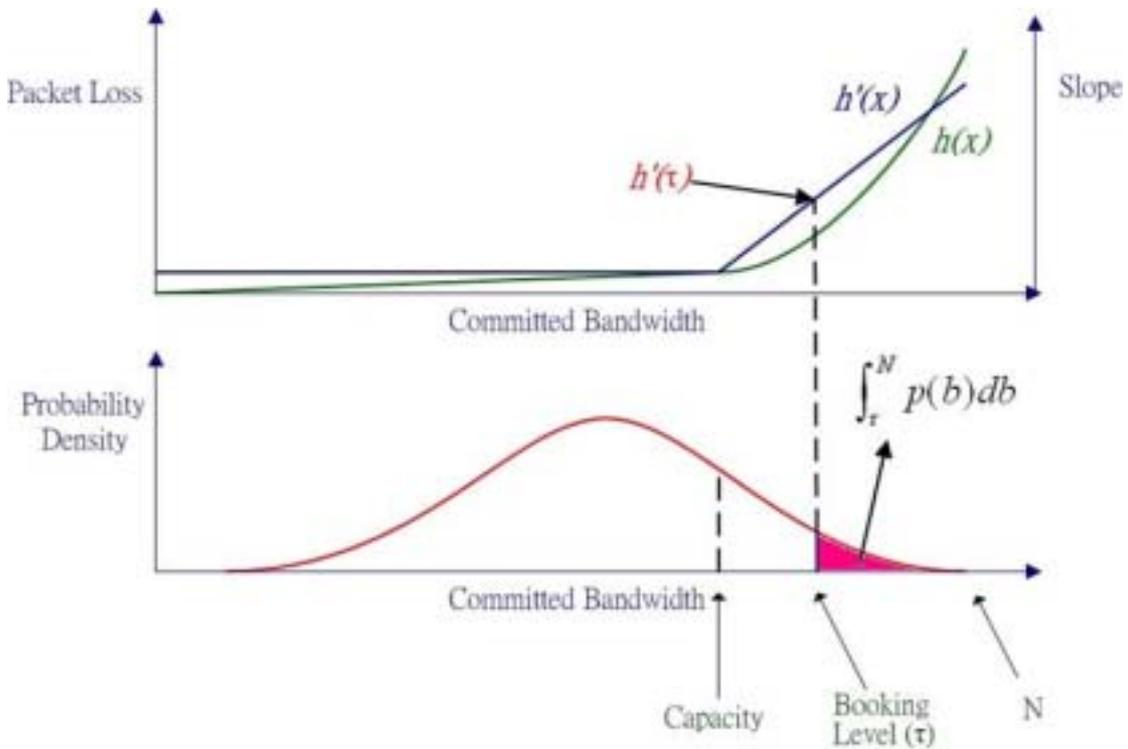


圖 4.20：最佳超額配置量

當需求分佈為一不連續函數時，獲利函數如下，不可微分，

$$R = m_1 \times (\tau - C) - m_2 \times \left[ \sum_{b=C}^{\tau} p(b) \times h(b) + \sum_{b=\tau}^N p(b) \times h(\tau) \right] \quad (4-9)$$

等式 4-9 為一個  $\tau$  的函數，求最佳解時前半部  $m_1 \times (\tau - C)$  的複雜度為  $O(\tau)$ ，後半部  $m_2 \left[ \sum_{b=C}^{\tau} p(b) \times h(b) + \sum_{b=\tau}^N p(b) \times h(\tau) \right]$  的複雜度為  $O(\tau^2)$ ，不同情況下之最佳解可以簡單的數學方法求出。

由上面的討論，可以知道影響超額配置量位置的因素為 Ingress Router 對頻寬需求的機率分佈，訊務遺失函數  $h(b)$ ，獲利的參數  $m_1$  與賠償的參數  $m_2$ 。

### 4.3 資源配置法分析比較

透過上面的分析比較，中央保留資源法會增加即時資源配置的負荷，但資源浪費與被拒絕服務的訊務量較容易控制，超額配置法則是將彌補預測誤差的部份隱含在方法中，不僅沒有即時資源配置的額外負荷，而且方法簡單，網路元件並不需要有特殊的功能與強大的即時運算能力。

由於系統是以獲利做最佳化運算，系統業者的環境或是策略會影響到資源的使用情況。但不同的環境下，兩個性質迥異的資源配置法表現出來的處理能力與彌補預測誤差的效果也大不相同，在下一章的資源配置方法效能評析中，以實驗模擬不同情況下上述兩種方法的效能，並提出建議，幫助系統業者於兩者間做出選擇。