行政院國家科學委員會專題研究計畫 期中進度報告

下一代多層級多媒體應用服務匯流網路--子計畫四:分散式視訊偵搜所需整合性網路與計算服務之研究(2/3)

期中進度報告(精簡版)

計畫類別：整合型
計畫編號：NSC 95-2221-E-002-193-
執行期間：95年08月01日至96年07月31日
執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：張時中

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 96年07月14日

下一代多層級多媒體應用服務匯流網路--子計畫四:分散式視訊偵搜所需整合性網路與計算服務之研究(2/3)

計劃類別：整合型計劃

計劃編號：NSC 95-2221-E-002-193

執行期限： 95 年 8 月 1 日至 96 年 7 月 31 日

整合型計劃：總計劃主持人：蔡志宏教授

子計畫主持人：張時中教授

子計畫參與人員:朱紹儀 張家瑋 葉力維 林哲煒 呂文彥

執行單位：國立台灣大學電機工程學系

一. 中文摘要

本計劃的第二年研究工作為探討 IP 網路上以 P2P 社群進行視訊偵搜服務與資訊分享的商務模型。研究進展主要包括以下兩部分:

(一) 有影像品質保證的移動視訊偵搜路由演算法之研究與路由演算法實做

研究目的是在隨意網路上實現有影像品質保證的移動視訊偵搜服務。我們研究以下兩個課題：1.如何週期性取得網路上其他節點的資訊、2. 如何建立並動態調整跳躍數少、能提供足夠的資料傳輸速度傳輸路徑。我們首先利用實驗了解隨意網路傳輸特性後，在 IP 層設計了一個鄰居資訊表路由法協定，包含鄰居資訊列表交換堆疊和鄰居資訊路由演算法 NILRA,來解決所提出的兩個主要研究問題。

我們在 IEEE 802.11 架構下以 4 台的筆記型電腦和運用 visual basic.NET 與 SQL2000 來實做，並與最短路徑法比較影像傳輸品質。

(二) P2P 社群使用者行為模型的建立與分享誘因實驗設計

以同儕對同儕式的教材分享系統為基礎，探討四個促進分享的重要議題:1. 建構教材分享系統會員數成長與教師上傳行為模型、2. 建構以網路社群經驗資料為基礎之使用者分享行為集體模型(collective behavior)、3.評估不同獎勵政策對系統教材質與量的影響、4. 基於無誘因制度下使用者行為模型，設計使用者對獎酬制度反應模型之實驗。藉由上述使用者行為模型的分析與實驗設計，為未來社群管理者提供建立架構處方性(prescriptive)模型以進行誘因設計(incentive design)。

關鍵詞:行動視訊偵搜、鄰居資訊列表、路由演算法、定位、內容與服務分享、誘因、獎酬制度、使用者行為模型、實驗設計、經驗數據分析

Abstract:

In the second yeart, we aim at constructign the business model of video surveillance service and information sharing on P2P community in IP network. Progresses of our research mainly include two parts:

I. Design and implementation of neighbor information-based mobile video surveillance routing over ad hoc networks

The focus of this task has been to explore an innovative mobile video surveillance services, and the quality of video is guaranteed. We have studied two issuesw:i) How to periodically obtain neighbor node information in an ad hoc network, and ii) how to establish transmission path with a low number of hops, enough data rate, and easy to maintain.

We first conducted some experiment to understand the transmission characteristics of an ad hoc network, and then designed a neighbor information list (NIL) routing protocol on the network layer. This protocol contains NIL exchange stack and a NIL routing algorithm (NILRA), which addresses the two main research issues. We used four notebooks and VB.NET ™ and SQL2000 ™ software development environment for a simple prototype implementation of our design and for comparison of NILRA with a shortest path routing algorithm.

II. P2P user behavior modeling and experiment design for incentive scheme:

On top of a P2P teaching material sharing environment, we have studied four issues for encouraging sharing among peers: 1) construction of Membership growth and content sharing model, 2) collective behavior modeling with empirical data from a production TMS, 3) Impact analysis of different reward policies to the quality and quantity of TMS, and 4) experiment design for constructing a model of user response to rewards.

Our research offers TMS community a method and models to construct prescriptive model for incentive design.

Keywords: *mobile video surveillance, neighbor information list, routing algorithm, positioning, content and service sharing, incentive, reward policy, user behavior modeling, experiment design, empirical data analysis*

二. 研究緣由，目的與成果

I. 有影像品質保證的移動視訊偵搜路由演算法之研究與路由演算法實做

I.1 研究目的

本研究想設計的移動視訊偵搜網路是希望在一個沒有任何固定式通訊基設備的戶外環境下(無 AP,如

戰爭或災害發生時)，還是可以立即提供偵搜服務的系統。其中戶外環境的定義是小型的戶外區域範圍(例如 600 乘 600 公尺範圍內)，且其中可能有目視障礙物(如樹木，房子，橋樑等)，我們希望花費最少時間力氣取得想監視的影像畫面。本研究主要研究目的就是在此環境中實現有影像品質保證的移動視訊偵搜服務。

I.2 研究問題探討

要實現移動視訊偵搜服務首先必須要知道偵搜任務的目的地在那(座標位置)，以及是否有 node 在目的地附近。所以我們必須要先有位置資訊，也就是取得隨意網路上自身與其他 node 的座標位置。但是由於隨意網路裡的 node 都是可以自由移動的，不但自身的位置會不停的改變，而且每一個 node 傳輸範圍內的鄰居也會隨時間而不同，而所建立好的路徑也可能因 node 移動而損毀。再者，802.11 架構設計使得多次跳躍(Multi-Hop)的傳輸方式會造成網路的吞吐量(throughput)嚴重下降；且兩個 node 之間距離越遠，802.11 能提供的傳輸速度也會越低，此兩項因素不利於即時多媒體的傳輸。因此我們提出了兩個主要研究問題：

1. 如何取得網路上其他節點的資訊。
2. 如何建立即時影像傳輸路徑(只考路單一傳輸路徑)，此路徑須符合以下要求。
   甲、 影像傳輸不易中斷(降低路徑連結中斷的機率)
   乙、 較少 HOP 數
   丙、 能提供足夠的資料傳輸速度

為了了解第二項研究問題的實際情況，我們利用實驗了解在隨意網路上傳輸即時影像的真實特性。其實驗包括距離遠近與多次跳躍等兩個實驗。由實驗結果和數據的分析，得知了在隨意網路上傳輸影像的特性，如下：

1. 當距離越遠的時候，兩個 node 之間的最大吞吐量會持續的下降(影像越來越差)，圖 1。
2. 當路徑上的 HOP 數量越多時，此路徑的吞吐量值的確會越來越低(影像越來越差)，圖 2。

經過上述的實驗結果分析，我們了解了實際傳輸即時影像會產生的問題，而在設計服務系統時必須考量這些特性存在。

為了提供有影像品質保證的移動視訊偵搜服務(含有位置資訊)，我們利用實驗所得到傳輸即時影像的實際特性，在網路層設計了一個鄰居資訊表路由協定(圖 3)來解決所提出的兩個主要研究問題。鄰居資訊表路由協定包含鄰居資訊列表交換堆疊(NIL exchange stack)和鄰居資訊路由演算法(NILRA)，其中 NIL 交換協定的想法是希望以區域性方式建立拓撲以減少控制訊息的浪費，而其建立方法與更新方法是設定每一個 node 可以週期性的取得鄰居的廣播(broadcast)來保持最新鄰居資訊(包含 IP address，經緯度位置，移動速度，方向和 Data rate)(表 1)，此些資訊可以幫助路由法來建立和維持路徑；而路由演算法(NILRA)裡我們詳細的說明包含尋找路徑，回傳搜尋資訊，建立影像傳輸路徑，降低路徑中斷機率等功能。

尋找路徑演算法的主要想法(圖 4)是接收到搜尋封包的 node，如果自身與目的地之間的距離不在門檻值內，則在自身的鄰居列表裡，尋找符合傳輸速度條件的鄰居中位置最靠近目的地的 node(利用經緯度位置和傳輸速度資訊)，接著將搜尋封包送給此 node，以此類推即可尋找到一條跳躍(HOP)數較少且符合傳輸速度的路徑。路徑搜尋演算法演算法的步驟如下：

符號定義(Notations)

MN	出發端(任務發起端)節點的任務編號 (Mission Number)
i, j, k	節點標籤(node label)，以 IP 位址表示
Li	節點 i 的 NIL，含 1. 經緯度座標，2. 傳輸速度，3. 鄰居節點 IP 位址;
in	節點 i 的鄰居所成的集合
RREQ	路由要求封包(Routing Request packet)
Xd , Yd	目的地經緯度座標
EP	錯誤資訊封包(error packet – 內含發起者的 IP 位址與任務編號)
Xi,Yi	節點 i 的經緯度座標
Rmin	DN 與 Xd,Yd 的距離門檻值
i*	in 集合中的一個節點
DR	傳輸速度的門檻設定值
Disti	節點 i 和目的地(Xd ,Yd 經緯度座標)距離
DRi*	Li 裡 i*的傳輸速度值
SN	RREQ 的發起節點(Source Node)

路徑搜尋演算法

```
輸入：   i，j，k
RREQ(內含 Xd ,Yd，Rmin，SN_s)
error packet(內含發起者的 IP 位址)
         in，jn
         Li
         Xin,Yin
         DR
         Xi,Yi
輸出：   1. 目的地 node 為 i
         2. 無法找到 Rmin 範圍內的 node

Step 1：封包分析
if ( i 接收到 j 的 RREQ )
    紀錄上一個 node → j
if ( MN 重複 )
    發生回圈(loop)，i 產生 EP，傳給 j；
           else
               將 SN_s 記錄下來；
           end if
         else if ( i 接收到 k 的 EP )
    in = in \ {k}； #(將 in 裡的 k 刪除)
    執行 step 3；
else if ( i 是任務發起者 )
```

```
    i 產生一個 RREQ 封包和任務編號 MN
end if
Step 2：目的地檢查
計算 i 與 Xd ,Yd 的距離
Disti = $\sqrt{(Xd-Xi)^2+(Yd-Yi)^2}$
if ( Disti < Rmin )
    DN = i；
    輸出 “ 目的地 node 為 i ”
    Stop；
else
    in = in \ { j}；# (將 in 裡的 j 刪除)
end if
Step 3：尋找鄰居列表裡的 node，檢查傳輸速度門檻值
if ( in = $\phi$ )
    執行 step4；
else
    i* = arg { $\min_{z\in i_n}$ ( $\sqrt{(Xd-Xz)^2+(Yd-Yz)^2}$ ) }
    DRi* = Li 裡的 i*的傳輸速度值
    if ( DRi* < DR )
in = in \ {i*}；   # (將 in 裡的 i*刪除)
    重新執行 Step 3；
    else
    轉送(relay)RREQ 給 i*；
    Stop；
    end if
end if
Step 4：返回上一個 node
if ( j =\= 0 ) # 0 代表不存在
    傳送 EP 給 j；
    Stop；
else
    輸出 “ 無法找到 Rmin 範圍內的 node ”；
    Stop；
end if
```

建立好的路徑後，其 node 也可以利用移動速度與方向這兩個資訊來降低傳輸路徑中斷的機率。假如傳輸路徑上的目的地node以每秒一公尺的速度往北方移動，路徑上的其他 node 也將以同樣速度和方向移動。由此可以維持建立好的路徑 node 相對位置不變，進而降低因為移動而使路徑損毀的機率。在這裡我們不考慮路徑損毀中斷後如何自動修復此路徑。

NILRA 主要的創新為利用鄰居資訊來發送少量的路徑搜尋封包，即可搜尋到足夠傳輸速度的路徑。另外為了讓使用者可以下達需求，在應用層上本研究設計了指令堆疊(圖 3)，包含搜尋、移動、取得影像和訊息回傳等四個功能性方塊，其目的是用來協助使用者指令介面與所設計的路由演算法做溝通，並且同時可利用位置資訊來指揮路徑上的節點移動以完成視訊偵搜任務。

I.3 實作成果

本研究的實做方面，我們使用了 VB.NET™ 和 SQL2000™ 在應用層簡易的實現鄰居資訊表路由協定(如圖 5)。在 NIL 交換堆疊實做方面，利用 VB.NET™設計好的廣播(Broadcast)程式，將自身的資訊週期的廣播(使用 UDP 封包)出去，而其他 node 接收到廣播則將此資訊加入自己的 NIL 資料庫(SQL2000)裡。而在路由法實做方面，使用 VB.NET™程式將建立好的 NIL 讀取出來，使用單點傳送(unit-cast 的 UDP 封包)將搜尋資料(RREQ)依照 NILRA 的搜尋方式傳給符合條件之鄰居。而所找的目的端會依原搜尋路徑將回覆封包(RREP)傳回給出發端(圖 6)。影像傳輸方面，使用 Windows 內建的 route 指令將所找到的路徑建立起來，開啟 webcamXP 即時影像分享軟體將影像傳回給出發端。我們舉出一個例子，如圖 7，紅色為 NILRA 找到的路徑，綠色為最短路徑法找到的路徑，接著比較此兩條路徑，其吞吐量、資料遺失率和人眼評分影像之優劣。由數據結果可得知本研究的路由法的確可以尋找到有足夠傳輸速度之路徑(表 2)。

I.4 能源限制的考慮

為了使無線視訊偵搜系統服務，能在能源有限的戰場中延長它的使用時間，本研究的主要目的為平均的消耗各節點的能源，以使整個系統能加長運作的時間。本系統是架構在 802.11b/g 上的 ad hoc 模式上，運用 ad hoc 的多點傳輸模式，建立起整個系統。由於需要透過多點來傳輸，所以為了避免有些點過度使用，造成過早進入 dead node 的情形，所以要想辦法延長每個 node 的使用時間，使整個系統能運作更久。再者是由於影像的傳輸比一般的聲音傳輸要花更長的時間，所以連結的穩定性也必須要考量。本研究在系統的設計上，主要分為兩個部分：(1)應用層的設計，(2)網路層裡節能路由的設計。其整個系統的架構圖請見(圖 8)。在應用層的設計分為兩個部分，第一部分是影像的傳輸介面，是運用 RTP (Real-time Transport Protocol)/RTCP(Real-time Transport Control Protocol)做為即時影像的傳輸協定，RTP 和 RTCP 配合使用，能以有效的回饋和最小的頻寬開銷，使傳輸效率最佳化，因而特別適合傳送網路上的即時資料。第二部分是命令的堆疊，裡面包含了三個主要的命令，分別是搜尋、移動及監看。而在網路層的路由設計上，我們加入了 neighbor list 的概念，在每個節點上利用定時廣播來取得附近節點的基本資料，包含了 IP address、位置座標、功率強度、剩餘能量，這是為了使在搜尋時能夠快速的找到合適的節點。表 3 為 neighbor list 的格式。而整個路由的路徑搜尋及維持，我們參考了 AODV[PR99]和 ESDSR[TTN05]的想法，提出了一個新的路由方法 Energy Saving Neighbor list Routing (ESNR)，主要是利用上面的 neighbor list 來搜尋目標物附近的節點，再者在回傳的訊息裡加入了利用剩餘能量和傳輸功率所計算出來的 RT(t)值，藉此來判斷出哪一條路徑能夠傳輸較長的時間，這樣就能夠減少能量不足的節點，過度的被使用，造成整個系統太早進入無法傳輸的

情形。而在路徑維持方面，如果中間有節點要移動離開或能量不足所造成的連結中斷，則由前一個節點去尋找替代的路徑，避免從頭開始建立路徑，造成能量的浪費。

I.5 研究成果總結

1. 利用實驗測量實際在隨意網路上傳輸即時影像的特性，包含距離越遠、路徑 HOP 數與吞吐量之間的關係。
2. 鄰居資訊列表(NIL)的設計，可以週期性廣播自身資訊以及動態的取得最新鄰居資訊(包含位置資訊等等)。
3. 設計了一套路由演算法(NILRA)。其利用 NIL 資訊來發送少量的路由封包，即可建立一條跳躍數少且提供足夠傳輸速度的路徑。
4. 設計了移動指令，指揮 node 的移動來完成偵搜任務，且可藉由位置資訊來指揮 node 保持影像傳輸路徑不中斷。
5. 使用了 VB.NET™和 SQL2000™在應用層簡易的實做了 NIL 的建立更新和 NILRA 演算法，且證明 NILRA 的路徑能提供有影像品質保證的傳輸路徑。

目前正在繼續進行如下:

甲、 多個任務造成資源分配不均的問題

乙、 自動修復傳輸路徑的問題

丙、 週期性廣播造成電力浪費之問題

II. P2P社群使用者行為模型的建立與分享誘因實驗設計

II.1 設計目的

在九年一貫教育體系與一綱多本政策下，近年來國內教材分享網站蓬勃發展。但目前缺少鼓勵分享行為的有效誘因。為建構出高品質與持續成長數量的 P2P 網路分享社群，本研究主要目的為提出架構處方性模型之方式以進行誘因設計。由過去的研究中發現，網站資源的量和質與獎勵誘因為影響社群使用者分享行為的主要因素。獎勵誘因對於促使使用者分享有直接的影響，網站資源的量和質則可能間接影響使用者分享行為。

II.2 研究問題探討

本研究探討四項行為：網路社群的加入、離開及社群服務的使用、提供。因此本研究分為四項研究議題：1. 建構教材分享系統會員數成長與教師上傳行為模型、2. 建構以網路社群經驗資料為基礎之使用者分享行為集體模型(collective behavior)、3.評估不同獎勵政策對系統教材質與量的影響、4. 基於無誘因制度下使用者行為模型，設計使用者對獎酬制度反應模型之實驗。

II.3 成果

我們建立使用者加入或離開社群與上傳教材的模型，並討論這些行為之間的相互影響。根據科技接受模型(technology acceptance model)與呂慧甄對影響知識分享因素研究(Lu 2003)，並以 S 曲線建立教材分享系統行為的機率模型。研究中之數學模型建立附於附錄中。此模型掌握影響教師行為的重要因子且符合基本經濟學原理，做為未來實驗的理論基礎(如圖 9)。

基於過去設計之 Bass model 與 S 曲線使用者行為模型，我們提出 S-shaped 曲線以捕捉網站資源的量和質與使用者行為之間的關係。並在無誘因制度的前提下，利用著名網路社群思摩特(SCTNet)的經驗數據，驗證行為機率與網站資源的關係。以統計軟體進行迴歸分析經驗數據後，驗證模型正確性。藉由實際網站之經驗數據，以曲線配適法導出 S-shape 之模型實際參數，以提供未來研究或網路管理之處方性模型。成果如圖 10-13 所示。

根據此模型，我們評估兩種獎勵誘因政策給教材上傳者之影響，分別是相對排名法(relative ranking)與閥值法(threshold)。相對排名法可以讓高品質檔案上傳者得到較多獎勵，讓他們上傳意願提高，增進檔案的平均水準。而閥值法齊頭式地鼓勵教材上傳行為，可以增加上傳的數量(如圖 14、15)。

另外為架構使用者對獎酬制度反應模型之實驗設計，由於在過去文獻中發現，相較分等之獎酬制度下，較能激勵使用者提出高品質的資源。我們以相較分等(relative ranking)的獎酬制度進行探討使用者反應模型之實驗設計。相較分等獎酬制度，將名譽得分予以分等，分數越高者得越多獎勵。我們在固定社群資源下，設計包含獎酬比例與獎酬金額兩項因子之二因子實驗設計。二因子實驗設計不但可評估各因子的主要影響更能探討因子間的交互作用，以探討在不同獎酬制度中使用者行為的反應模型(如圖 16-19)。本實驗環境設定在已架構之 P2P 資源分享社群。於實驗室進行小型實驗，使實驗不僅節省成本更能控制可能產生的不可預期變動。本研究藉由上述使用者行為模型的實證分析與實驗設計，為未來社群管理者提供建立架構處方性模型之方式以進行誘因設計。

II.4 研究成果總結

1. 我們以 S 曲線建立使用者加入或離開社群與上傳檔案的模型，並討論這些行為之間的相互影響。此模型掌握影響教師行為的重要因子且符合基本經濟學原理，做為未來實驗的理論基礎。
2. 我們基於過去研究中之 Bass model 與 S 型使用者分享模型，架構 S-shaped 曲線之使用者集體行為模型以捕捉網路資源的質和量與集體使用者加入、離開社群、分享與利用資源的行為關係。
3. 藉由思摩特往之經驗數據，我們以曲線配適法導出 S-shaped 曲線之參數，提供未來研究與網路管理處方性模型使用。
4. 評估兩種獎勵誘因政策給檔案上傳者之影響，分別是相對排名法(relative ranking)與閥值法(threshold)。相對排名法可以讓高品質檔案上傳者得到較多獎勵，讓他們上傳意願提高，增進檔案的平均水準。而閥值法齊頭式地鼓勵教材上傳行為，可以增

加上傳的數量。

5. 設計雙因子實驗室環境之實驗設計，提供可控制並較為經濟之實驗方法以架構使用者對於不同獎勵制度之反應模型。藉由雙因子實驗室環境之實驗設計，未來研究者可在小型實驗環境中，捕捉在不同獎酬比例與獎酬金額下集體使用者加入、離開社群、分享與利用資源的行為關係，以獲得架構獎勵制度處方性模型之有效資訊。

三. 計劃成果自評

本計劃除了順利完成上述之成果之外。已發表論文如下:

[1] S.-I Chu, S.-C. Chang, “Time-of-Day Internet Access Management: Virtual Pricing Vs. Quota Scheduling,” *Proceedings of IEEE ICCS 2006*, Singapore, Oct., 2006, pp. 1-6.

[2] Shao-I Chu, “Research on Time-of-day Internet Access Management by Quota-based Priority Control.” *PhD Thesis,* Dept. of Electrical Engineering, National Taiwan University, July 2007

[3] Chia-Wei Chang ,“Design and Implementation of Neighbor Information-based Mobile Video Surveillance Routing Over Ad Hoc Networks,” *Master Thesis,* Dept. of Electrical Engineering, National Taiwan University, July 2007

[4] Li-Wei Yeh,” P2P user behavior modeling and experiment design for incentive scheme,” *Master Thesis,* Dept. of Electrical Engineering, National Taiwan University, July 2007

附錄

教材分享行為機率的模型

Membership Growth

At time t, there are N_t teachers already in TMS system. On contrast, (TN- N_t) teachers have not joined the TMS system. The probability to join TMS system is $f_{join}(t)$, so the expected number of new member at time t is $(TN-N_t)f_{join}(t)$. Teachers might leave the system of probability $f_{leave}(t)$, the expected number of remaining member is $N_t(1-f_{leave}(t))$. The total number of teacher at time t+1 equals to the remaining member plus new member.

$$N_{t+1} = N_t(1-f_{leave}(t)) + (TN-N_t)f_{join}(t)$$

Content Growth

In each time slot, teacher might submit one new TM. Because teachers are homogeneous, their probability to upload content is the same. So the probability distribution of x_t new TMs is binomial distribution with parameter N_t and $f_{upload}(t)$. The relation of content variety between t and t+1 is $CV_{t+1} = CV_t + x_t$, where x_t is a positive integer drawn from binomial distribution.

$$P\{x_t = n\} = f(n, N_t, f_{upload}(t)) = C_n^{N_t}(f_{upload}(t))^n(1-f_{upload}(t))^{N_t-n}$$

Quality Change

We use average content quality as an indicator of our system. The average content quality changes while teachers upload their new content, the average content at time t+1 is

$$CQ_{t+1} = \frac{(CQ_t * CV_t + \sum_{j=1}^{n}(x_{tj} * cq_{tj}))}{CV_{t+1}}$$

Joining Probability

We assumed all kinds of reward incentives transfer to monetary reward given to teacher. For example, if they join the TMS for the first time, bureau of education will give bonus to them. The bonus induces teachers to use TMS system, and the bonus given to teacher i at time t is denoted as RIit.

For teacher the “usefulness” is the benefit TM brings, the TM quantity and quality affects that benefit. Teachers will get more benefit if the quantity is large and quality is high. We define the usefulness as expected TM benefit which is a function of product of CV_t and CQ_t.

The joining probability at time t is a function of expected TM benefit and reward incentive.

$$\Pr\{\text{teacher join at time t}\} \equiv f_{join}(CV_tCQ_t, RI_{it})$$

Leaving Probability

Once teachers have used TMS system, they will stay for a period of time and leave afterward. The probability to leave the system reflects opposite willingness to stay. The more willingness teachers have to stay, the lower probability of teachers’ leaving. We believe the probability to leave are affected by new TM benefit and habit.

Teachers are attracted by new TMs, new TMs bring more benefit. Like many portal website, they provide new contents or services everyday to attract Internet users. If no new TMs are uploaded, teachers might leave the community. In each time slot, the more new TMs quantity and quality (x_tcq_t) the more they are attracted to the system.

Researchers found online shoppers’ intentions to continue using a website that they last bought at depend not only on perceived usefulness and perceived ease of use, but also on habit. In [Lu03] authors also found that sustained teacher material sharing behavior is supported by habits.

For reasons mentioned above, we define leaving probability as function of new TM benefit, existing content TMs benefit and a habit factor.

$$\Pr\{\text{a teacher leaves at time t}\} \equiv f_{leave}(x_tcq_t, CV_tCQ_t, S_{it})$$

Content Submission

In [Lu03] the authors find the personality is the main factor that influences the upload behavior. Altruistic teacher contribute their contents without asking reward, they get positive utility from altruism behavior. In contrast, extreme selfish teachers only upload content if they can get reward. So we model teachers upload probability combining these two factors, altruism and reward incentive.

$$\Pr\{\text{a teacher upload one new TM at time t}\} \equiv \mathrm{f}_{upload}(AL_i, RI_{it})$$

四. 參考文獻

[CCL03] I. Chlamtac, M. Conti, Jennifer J.-N. Liu, “Mobile ad hoc networking: imperatives and challenges,” *Ad Hoc Networks*, Vol. 1, No. 1, pp.13-64, July 2003.

[CWK97] B. P. Crow, I. Widjaja, J. G. Kim and P. T. Sakai, “IEEE 802.11 Wireless Local Area Networks,” *IEEE Communications Magazine*, Vol. 35, pp.15-18, September 1997.

[IEE99] IEEE Std. 802.11-1999, Parr 11: Wire/-s U N Medium Access Control (MAC) and Physical Layer (PHY) specifications, Reference number ISO/IEC 8802-11:1999(E), IEEE Std 802.11, 1999.

[XUS02] Shugong Xua and Tarek Saadawi, “Revealing the Problems with 802.11 Medium Access Control Protocol in Multi-hop Wireless Ad Hoc Networks,” *ELSEVIER Computer Networks*, Vol. 38, pp. 531–548, 2002.

[COM99] S. Corson, J. Macker, “Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations,” *IETF RFC 2501*, January 1999.

[Chi06] Yi-Ren Chiou, “TMS behavior modeling, incentive and system design against piracy and collusion” *Master Thesis, NTU IE*, Jul 2006

[Lai05] Jiang-Jang Lai, “P2P Network and Incentive Design for Sharing Teaching Material” *Master Thesis, NTU EE*, Jul 2005.

[Lu03] Hui-Chen Lu, “Study of Influential Factors on Knowledge Educators Sharing of the Case of SCTNet Network Community” *Mater Thesis, National Chung Cheng University*, Jun 2003.

[PR99] Perkins, C.E.; Royer, E.M., “Ad-hoc On-Demand Distance Vector Routing” *IEEE* 1999.

[TTN05] Tarique, M.; Tepe, K.E.; Naserian, M.;, “Energy Saving Dynamic Source Routing for Ad Hoc Wireless Networks” *IEEE* 2005

資訊 鄰居節點 IP 位址	經緯度位置	傳輸速度(data rate)	移動速度	方向
169.254.xxx.xxx	xx	xx	x	xx
169.254.xx.xxx	xxx	xx	xx	xx
169.254.xxx.xx	xx	xxx	xx	x
…. .				
…….				

表 1：NIL example

數據資料 路由方法	路徑吞吐量	資料遺失率	人眼評分
NILRA 之路徑	1.803Mbps	趨近 0	5 分
最短路徑法之路徑	0.852Mbps	8.13%	2 分

表 2：路由法之數據結果比較(實驗結果)

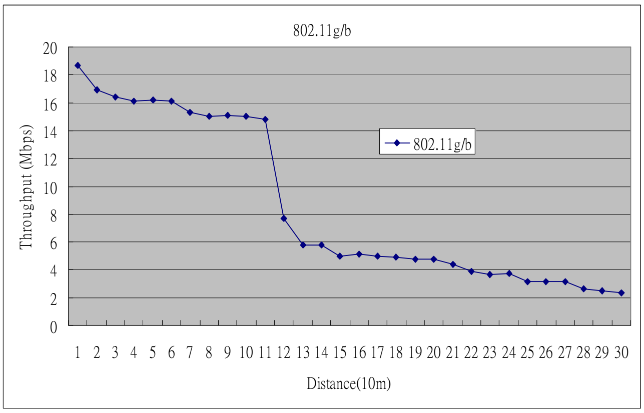


圖 1：距離與吞吐量(實驗數據)

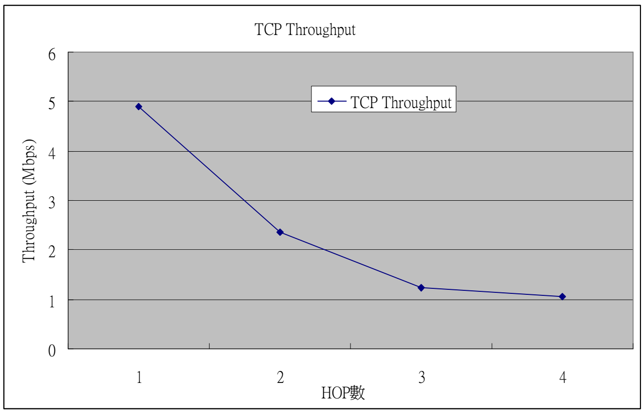


圖 2： Hop 數與吞吐量(實驗數據)

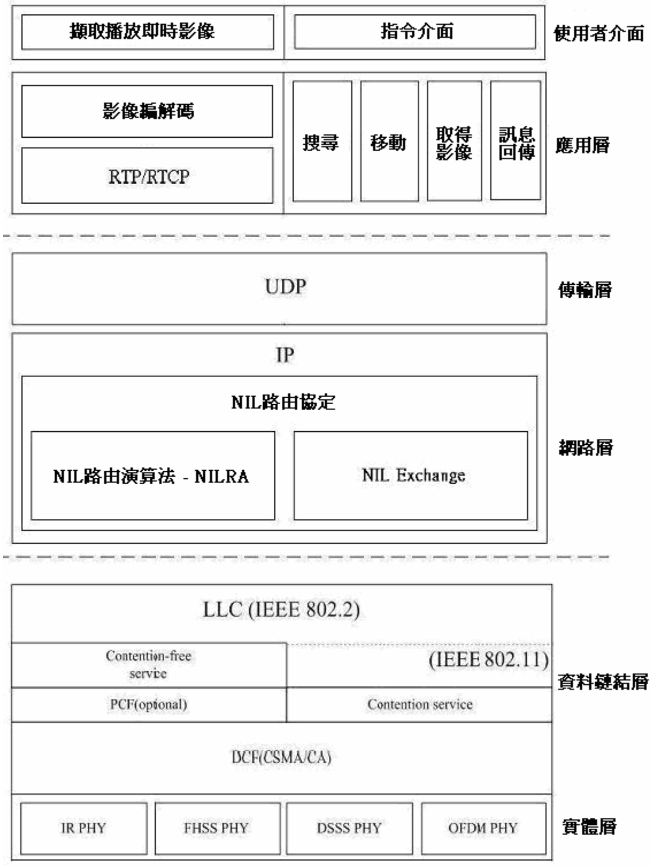


圖 3：The complete stack figure

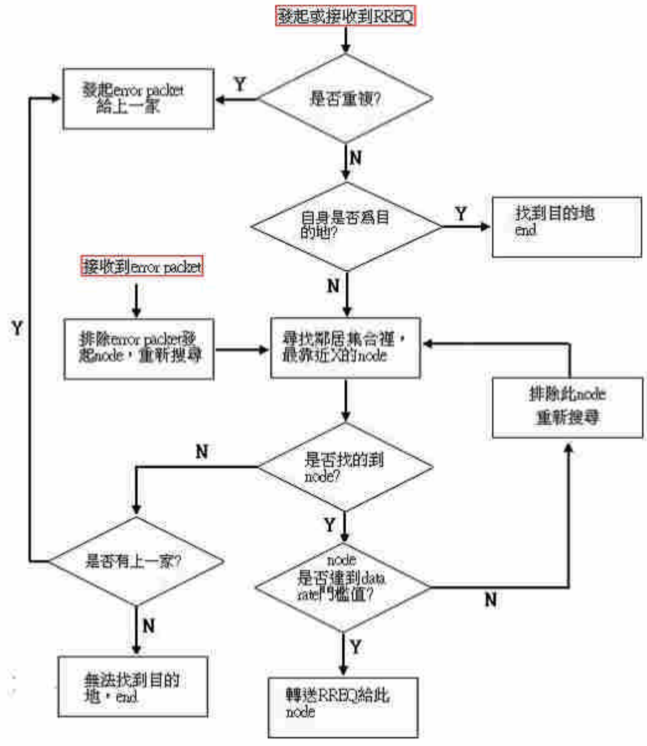


圖 4：個別 node 的搜尋步驟邏輯流程圖

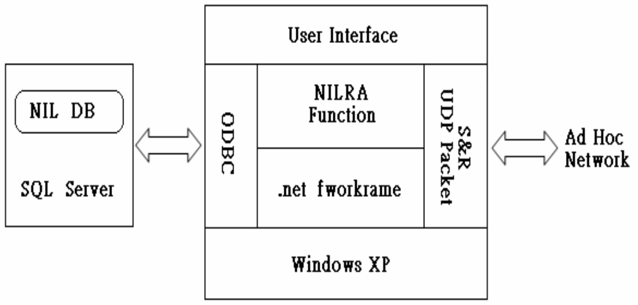
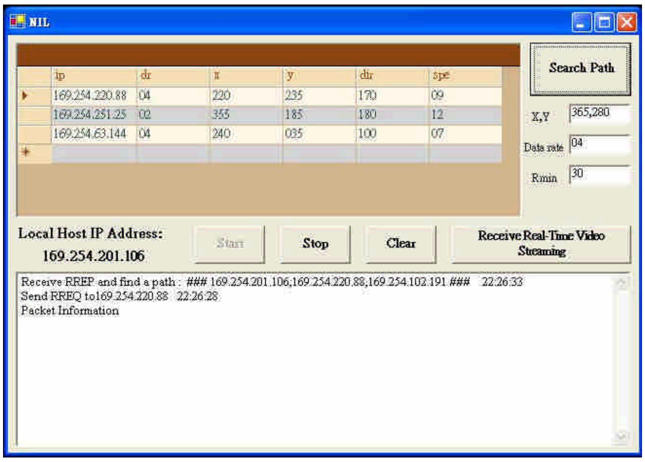


圖 5: 軟體架構堆疊



	ip	dr	x	y	dir	spe
▶	169.254.220.88	04	220	235	170	09
	169.254.251.25	02	355	185	180	12
	169.254.63.144	04	240	035	100	07
*						

X,Y 365,280

Data rate 04

Rmin 30

Local Host IP Address:
169.254.201.106

```
Receive RREP and find a path :  ### 169.254.201.106;169.254.220.88;169.254.102.191 ###    22:26:33
Send RREQ to169.254.220.88   22:26:28
Packet Information
```

圖 6：發送 RREQ 與接收 RREP 搜尋結果

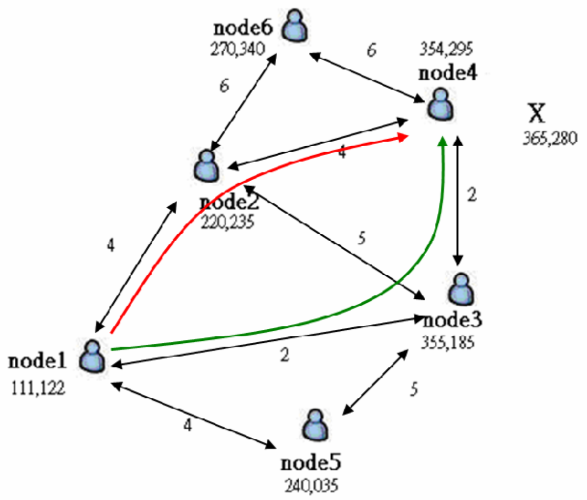


圖 7: NILRA 與最短路徑法之搜尋結果

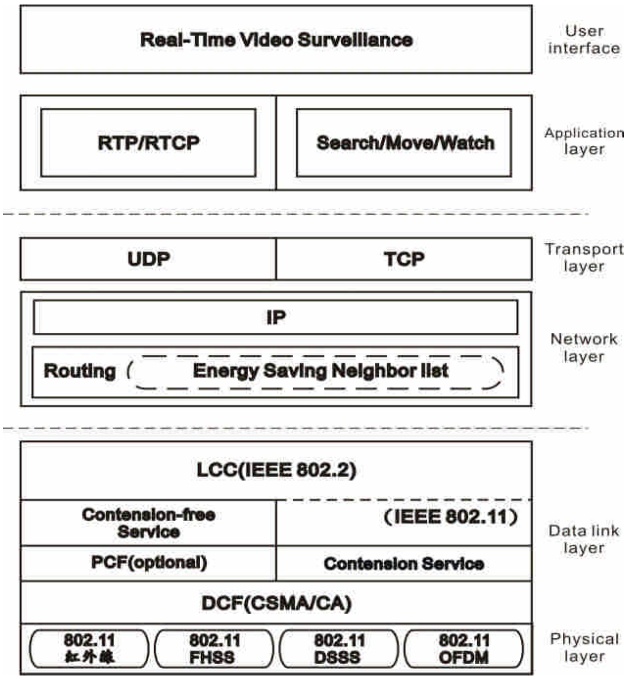


圖 8: 系統的架構圖

表 3: Neighbor list 的格式

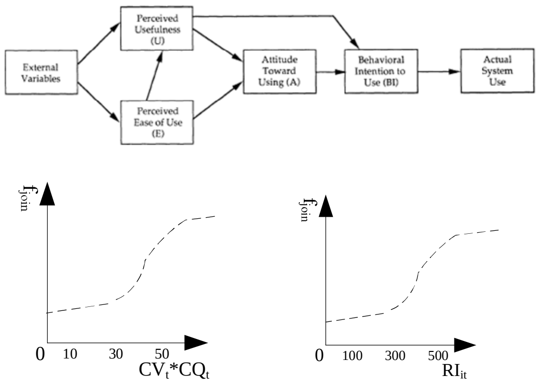


圖 9: Technology acceptance model and s-curve join probability model

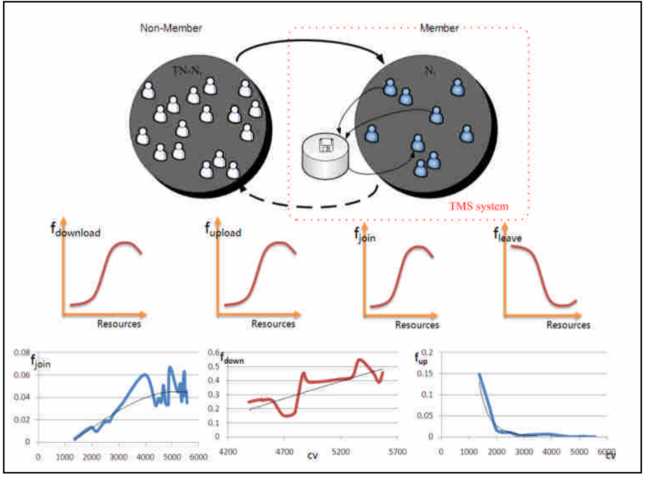


圖 10:使用者行為模型分析結果

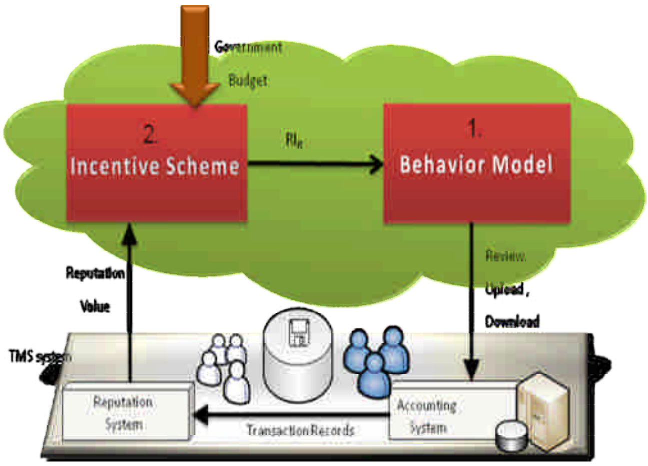
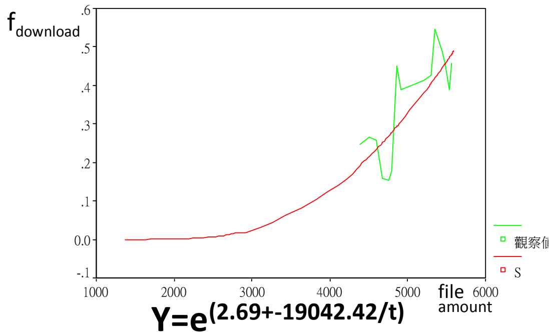


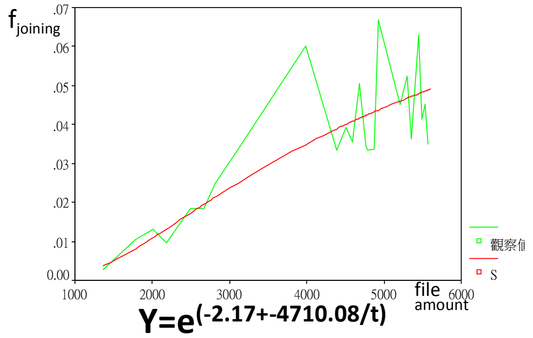
圖 11： Design issues

	IP address	位置座標	功率強度	剩餘能量
Peer2	xx	xx	xx	x
Peer3	x	xxx	xx	xx



R Square	Sig. F-value	b0	b1
0.53024	0.0021	-19042.428146	2.690284

圖 12: S-shape curve fitting for download probability



R Square	Sig. F-value	b0	b1
0.88409	0.0000	-4710.0840	-2.173498

圖 13: S-shape curve fitting for joining probability

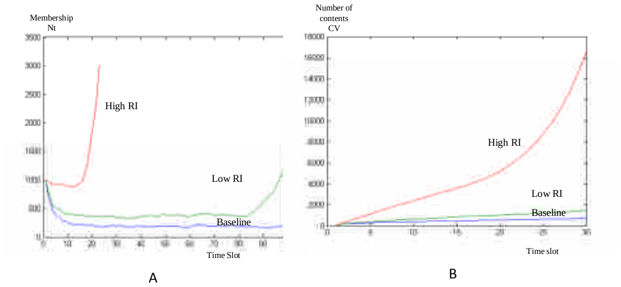


圖 14: (A) Membership in different RI scheme (B) Content variety in different RI scheme

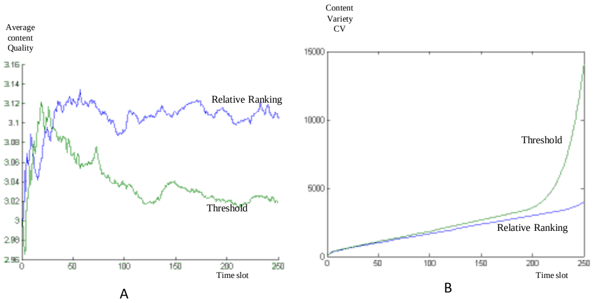


圖 15: (A) Content quality under reward policy (B) Content variety under reward policy

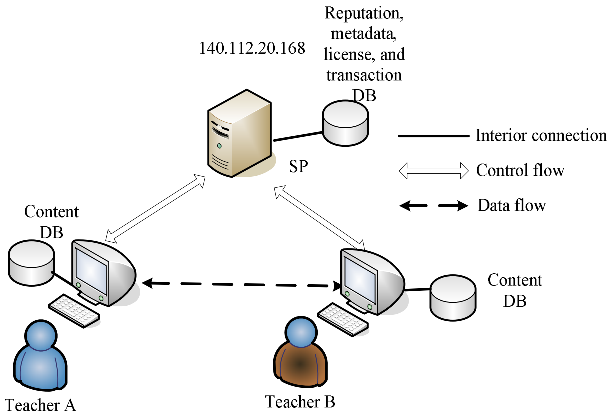


圖 16: Experiment environment

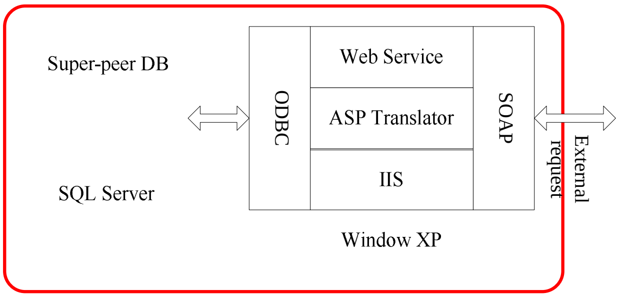


圖 17: Super-peer software stack

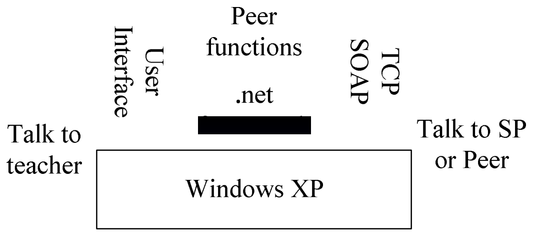


圖 18: Peer software stack

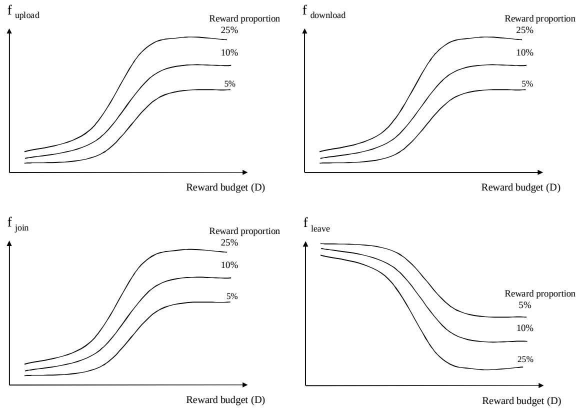


圖 19: Expected behavior model under different reward parameters

博士班研究生出席國際會議報告書

報告人姓名	朱紹儀	所屬系所	電機工程學系
會議時間地點	Holiday Inn ATRIUM SINGAPORE, Singapore, 30 Oct -1 Nov 2006		
會議名稱	The Tenth IEEE International Conference on Communication Systems (ICCS 2006)		
發表論文題目	Time-of-day Internet Access Management: Virtual Pricing vs. Quota Scheduling		

一、　會議簡介

ICCS 會議是每兩年在新加坡舉辦一次有關通訊系統和工程方面的國際會議，今年收到大量的論文投稿(自於 43 個國家總共約 480 篇的論文)，經過嚴格的審核過程，最後約有 200 篇左右的論文被接受。在這次會議期間亦舉辦了四個 tutorials，分別是 Iterative Receiver Design，Emerging Wireless Standards for WRAN, WiFi, WiMedia and ZigBee，WiMax Systems and Mesh Networks 和 Wireless Sensor Networks － Research vs Reality。

二、　參加會議經過

第一天以一個 Keynote Speaker 的演講作為會議的開始，題目為 Universal Communications – Towards Ubiquitous Network Society，演講者為 Shingo OHMORI, Vice president, Member of the Board National Institute of Information and Communication Technology (NICT)。印象頗為深刻的部份為演講者播放一段未來日常生活跟網路結合的影片。生活所遭遇到的各種問題，均可由各種網路的相連而得到解決，帶給人類便利的生活。接下來即為各領域的報告，學生參與研究相關的網路資源管理（Network Resource Management）的 Session，一方面聽取他人的新研究方向，另一方面報告亦報告自己的研究，發表過程中聽眾發問的相當踴躍，學生於報告結束後與在 Institute for Infocomm Research 工作的研究學者有充分的討論和交換資訊。

三、　與會心得

雖然有近 200 的論文的作者需要與會發表，但第一天 Keynote Speaker 的演講參與程度並非相當熱烈，頗為遺憾。此外，由此次參與新加坡的 ICCS2006 會議中，可窺見國際化的視野對學術界的重要性，並深深感覺到英文聽力的重要，尤其是須適應不同人種的英文口音。

四、　攜回資料名稱及內容

1. FINAL PROGRAM in a book.
2. PROCEEDINGS in a CD

TIME-OF-DAY INTERNET ACCESS MANAGEMENT: VIRTUAL PRICING VS.QUOTA SCHEDULING[*]

Shao-I Chu, Shi-Chung Chang

Department of Electrical Engineering
National Taiwan University

ABSTRACT

There exists abusive and unfair Internet access during peak hours over a free-of-charge or flat-rate network even under a quota-based priority control (QPC). To design effective management over time based on QPC, this paper compares and analyzes two classes of schemes: quota scheduling (QS) and time-of-day pricing (TDP). The TDP design captures both myopic and prudent user behaviors by exploiting empirical demand data. The load balancing-based quota scheduling (LB-QS) intends to equalize traffic over time by proportional quota allocation to time periods of control, while the peak shaving-based quota scheduling (PS-QS) is designed to reduce total traffic during peak hours through a rough empirical data-based user model. Performance evaluation demonstrates that TDP significantly outperforms both LB-QS and PS-QS. This is because TDP exploits user demand modeling and pricing to induce user behavior over time. The TDP design is more complicated than QS. Recommendations are given for selecting an effective Internet access scheme based on data availability.

1. INTRODUCTION

There often exists abusive and unfair usage of Internet access, especially during peak hours, over a network environment where the service charge is free or flat rate. For example, consider the dormitory network of National Taiwan University (NTU), where the network management adopts a quota-based priority control (QPC) scheme [1] to control its Internet access traffic. When a user's Internet-access volume exceeds the daily regular-service quota, the user's traffic is directed to a lower priority service. Statistics shows that the drop rate during peak hours (2.5Mbps above) is much higher than that of off-peak hours. The average usage of heavy users (8% of the user population) is 12.08 times more than that of all other users. Such observations imply that even under QPC, heavy users still abuse the network resource, and unfair usage is still significant. It is because QPC does not take the temporal effect of user demands into consideration.

The temporal issues, such as peak-shaving and load-balancing over time, are important for network management. To deal with them, some network management tools, like Cisco P-Cube [2] and Packeteer Packet-Shaper [3], are getting more widely adopted. They offer the needful functions of traffic control, similar to the system in [1]. However, they are not only costly in installation and maintenance, but also require additional policy design for effective operation. This paper focuses on the policy design and intends to give network managers guidelines for regulating the Internet access over time.

Time-of-day pricing or peak-load pricing [4, 5] offers a simple and indirect load management mechanism for public utilities. They meet the dual objectives of 1) reducing peak load, and 2) shifting a portion of the peak load to the base load. Over communication networks, Paschalidis and Tsitsiklis [6] showed through their simulations that static time-of-day pricing is almost as good as congestion pricing if user reactions to changing prices are well modeled. In [7], Shih et al concluded that static time-of-day pricing can influence user behavior more effectively than congestion pricing by conducting pricing experiments for a computer-telephony service.

Aside from time-of-day pricing, the quota scheduling is another approach to enrich QPC capability over time. It allocates the given daily quota to individual time periods to directly and forcedly limit the maximum volume usage of each user during peak hours. The quota of one time period cannot be carried over to another. There is no need of pricing mechanism in quota scheduling design.

This paper studies the policy designs and evaluates the control effects of quota scheduling and virtual time-of-day pricing on abuse and fairness improvements, peak shaving and load balancing over a network with QPC. Time-of-day pricing (TDP) design adopts a general user utility function and requires two data collections for user classification and modeling to maximize the bandwidth utilization while keeping the total demand below the link capacity. Two schemes for quota scheduling (QS) are proposed and studied. One is to equalize the traffic of peak and off-peak hour without any empirical data, called load balancing-based quota scheduling (LB-QS). The

[*] This work was partly supported by the National Science Council, Taiwan, Republic of China, under grants NSC-91-2219-E-002-033, NSC-92-2212-E-002-060 and NSC-93-2212-E-002-082.

1-4244-0411-8/06/$20.00 ©2006 IEEE.

other, called peak shaving-based quota scheduling (PS-QS), aims to reduce the peak-hour traffic by adopting an aggregated user behavior model. The model is constructed from the measurements of a QPC network.

TDP, LB-QS and PS-QS are evaluated on the empirical data of a 5000-user network. Results show that TDP significantly outperforms LB-QS and PS-QS in abuse and fairness reduction and load balancing and peak shaving effects. PS-QS is better than LB-QS because of more knowledge from empirical data. If the peak hours scatter over all time slots, TDP is the best scheme because either LB-QS or PS-QS results in the severe congestion at the time of quota renewal. All these schemes can be easily integrated with a QPC network from the viewpoint of practical implementation. As a network manager, TDP is the best choice if the needed data of user behavior is available.

2. QUOTA-BASED PRIORITY CONTROL AND ITS DEFICIENCY

2.1. Quota-based Priority Control

Lin et al in [1] combined the ideas of quota limitation and priority differentiation (QPC) to assure every user's basic demand, regulating heavy users' abusive usage by homogeneous quota. The prioritized service consists of two service classes: regular and custody. The regular service has a higher priority and each user is given a daily quota. Through a prioritized service, a user can still access the Internet at a lower priority after exhausting his quota.

Lin et al experimented with QPC over the NTU dormitory network with 5355 dormitory users, as shown in Figure 1. Only 54Mbps was allocated to the outbound traffic from the dormitory networks. Experimental results show that abusive Internet access by the top 2% heavy users is reduced by 57.82%. Accordingly, the congestion is alleviated with a 48.9% reduction of average packet drop rate. Fairness is also improved in users' daily usage.

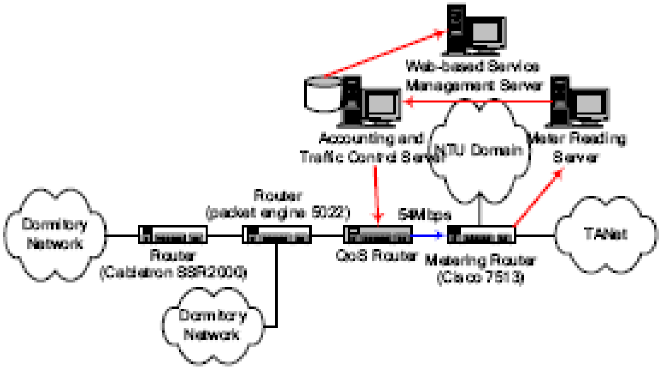


Figure 1: Network architecture of quota-based priority control

2.2. Deficiency of QPC

In spite of the effectiveness of the QPC as demonstrated in [1], the QPC design does not involve the network management of Internet access over time, especially during peak hours. Our further analyses show that the Internet-access bandwidth is highly utilized and the drop rate is higher than 2.5Mbps during peak hours of 9 a.m. to 3 a.m. when the daily quota is 1 gigabyte (GB). The ratio of heavy users' usage to normal and light users' usage is 1,308%. Heavy users still occupy most bandwidth. There is an obvious need for a finer management scheme to regulate users' Internet access over time.

For effective management, two ideas are motivated. Given a daily quota, quota scheduling (QS) is an intuitive measure, which allocates different quotas to different time periods and directly controls user usage over time. Instead, the virtual time-of-day pricing (TDP) takes advantage of differentiated prices to give users incentive to allocate their demand over time. This paper evaluates and compares QS and TDP, giving network managers suggestions for selecting an effective Internet access scheme based on data availability.

Specific challenges are as follows:

(C1) What QS schemes to consider for fair comparison with TDP;

(C2) How to quantitatively assess the applications of QS and TDP to a production network in abuse reduction, fairness improvement and load balancing and peak shaving effects;

(C3) How to choose an appropriate scheme according to the available empirical data, traffic pattern over time, design complexity and the network performance.

3. VIRTUAL PRICING

A virtual time-of-day pricing (TDP) policy [11] was designed for a network service provider (NSP) to use in conjunction with QPC. It captures myopic and prudent behaviors by exploiting empirical data for the effectiveness of TDP design. It also intends to give users the flexibility in allocating their quota, inducing users to shift part of their peak-hour demands to off-peak hours.

3.1. User Demand Model under Pricing

To summarize the mathematical model of user behaviors, let us first define some notations.

Notations:

B: bandwidth of internet access;

T: length of a time slot;

$v_{i,k}$: internet-access volume submitted by user i for regular service at time slot k, i=1,2,…,I, and k=1,2,…,K;

v^b_i: daily internet-access demand of user i, obtained from the baseline network, where there is no quota control, i=1,2,…,I;

Q: daily quota allotted to each user;

$Q_{i,k}$: remaining quota of user i at time slot k, i=1,2,…,I, k=1,2,…K; note that $Q_{i,1}$=Q;

p_k: price (number of quota per byte) of regular service at time slot k, k=1,2,…,K;

$\omega_{i,k}$: preference value of user i at time slot k, i=1,2,…,I, k=1,2,…K;

S_j: set of type j users and $j\in\{m, p\}$, where m corresponds to the myopic type, while p is the prudent type.

3.1.1. User Classification Mechanism

Whether a user is myopic or prudent depends on the given price profile and personal daily demand. The user demand is estimated from a baseline network, where there is no traffic control. If the daily demand of a user can be satisfied at the maximal price, there is no need to be prudent in quota allocation. Such a user is therefore regarded as a myopic user. If a user's demand cannot be met at the minimal price, the user should be prudent and will carefully allocate the quota. Let x_i be the probability of user i being a prudent user.

$$x_i = \begin{cases} 1, & v_i^B \geq B; \\ \dfrac{v_i^B - A}{B - A}, & A < v_i^B < B; \\ 0, & v_i^B \leq A; \end{cases} \qquad (1)$$

where $A = Q/\max_k\{p_k\}$ and $B = Q/\min_k\{p_k\}$.

3.1.2 Myopic User Model

At time slot k, a myopic user i with $Q_{i,k}$ >0 determines the internet-access volume of regular service, $v_{i,k}$, to maximize user i's own benefit at that time slot only (short-term benefit), without considering the available quota value of $q_{i,k}$ and the future demand. User i's benefit function is

$$J_i^k(v_{i,k}) = U_i^k(v_{i,k}) - p_k v_{i,k}\text{, for all } i\in S_m, \qquad (2)$$

where the first term represents the utility from submitting $v_{i,k}$, and the second term represents the amount of quota deducted if $v_{i,k}$ is successfully transmitted.

Experimental results of [8, 9] suggest that users' utility follows diminishing returns to scale. Hence, we assume that

$$U_i^k(v_{i,k}) = \omega_{i,k} F(v_{i,k}), \qquad (3)$$

where $F(v_{i,k})$ is strictly increasing, concave and continuously differentiable.

A user's quota accounting is based on the actually transmitted volume, which equals the submitted volume minus the dropped volume of a time slot. The drop ratio of regular service at time slot k is defined by

$$d_k = \max\left[\left(\sum_{i=1}^{I} v_{i,k} - BT\right)/\sum_{i=1}^{I} v_{i,k}, 0\right], \qquad (4)$$

At the beginning of time slot k+1, user i's quota is therefore updated by

$$Q_{i,k+1} = Q_{i,k} - p_k v_{i,k}(1-d_k). \qquad (5)$$

The myopic user decision problem (MUDP) is formulated as

(MUDPi) for k=1,2,…..,K,

$$\underset{v_{i,k}}{\text{Max}}\, U_i^k(v_{i,k}) - p_k v_{i,k}$$

subject to Eqs. (4) and (5) with $Q_{i,k}$ >0.

3.1.3. Prudent User Model

At time slot k, a prudent user i considers the daily demand and allocates the available quota and determines the internet-access volume of regular service to maximize user i's total benefit from time slot k to time slot k (long-term benefit), which is

$$\sum_{t=k}^{K} J_i^t(v_{i,t}) = \sum_{t=k}^{K}\left[U_i^t(v_{i,t}) - p_t v_{i,t}\right]. \qquad (6)$$

When planning for quota allocation, a user presumes that the planned submission would be transmitted, and the total submission should satisfy

$$\sum_{t=k}^{K} p_t v_{i,t} = Q_{i,k}. \qquad (7)$$

The prudent user decision problem (PUDP) is then

(PUDPi) for k=1,2,…,K,

$$\underset{(v_{i,t},\, t=k,\dots,K)}{\text{Max}} \sum_{t=k}^{K}\left[U_i^t(v_{i,t}) - p_t v_{i,t}\right]$$

subject to constraint (7) and $Q_{i,k}$ >0.

(MUDPi)and (PUDPi) are separable convex programming problems. A Lagrangian relaxation method [10] may be applied to solve them.

User Preference Estimation

Chu and Chang [11] utilize user submitted volumes collected from a QPC network to estimate $\{\omega_{i,k}\}$ for user model construction. Since pure QPC corresponds to TDP with p_k=1 for each time slot k, the preference of user i at time slot k is estimated as

$$\omega_{i,k} = 1/F'(v_{i,k})\,|_{v_{i,k}=v_{i,k,QPC}} \qquad (8)$$

by exploiting the optimality conditions of (MUDPi)and (PUDPi). Here, $v_{i,k,QPC}$ be the submitted volume of user i for regular service at time slot k over a network with QPC.

3.2. Tim-of-day Pricing Design

The goal of price setting by the NSP over a free-of-charge or flat-rate network is to maximize the total bandwidth utilization over a day while the average total demand does not exceed the capacity. The bandwidth utilization of regular service at time slot k is defined as

$$\frac{1}{BT}\sum_{i\in A_k} v_{i,k}(1-d_k), \qquad (9)$$

where $A_k = \{i\,|\,Q_{i,k} > 0, \forall i\}$ is the set of users whose available quota at time slot k is nonzero, and $\sum_{i\in A_k} v_{i,k}(1-d_k)$ represents the total transmitted volume of regular service at time slot k. The constraint that the total expected volume of submission cannot exceed the link capacity at a time slot is expressed as

$$\sum_{i\in A_k} v_{i,k} \le BT, k = 1,2,\ldots,K. \qquad (10)$$

Taking users' behaviors characterized by (MUDP*i*) and (PUDP*i*) into consideration, the NSP has a pricing problem (PP) formulated as a Stackelberg game [12]:
(PP)

$$\underset{\{p_k\in\Omega, k=1,2,\ldots K\}}{\text{Max}} \frac{1}{BT}\sum_{k=1}^{K}\sum_{i\in A_k} v_{i,k}(1-d_k)$$

subject to constraints (4), (10), the user classification mechanism, and MUDP*i* (PUDP*i*) if user *i* is myopic (prudent).

Although the solutions to individual user's optimization problems have closed forms, the analytic solution of (PP) is not available. A numerical method is thus adopted to solve (PP). Such a TDP design method applies to problems of user utility function with the property of diminishing returns to scale.

4. QUOTA SCHEDULING

To response challenge (C1), two QS schemes are proposed. One is load balancing-based quota scheduling (LB-QS); the other is peak shaving-based quota scheduling (PS-QS). They both directly curb user usage over time rather than adopting an incentive control like TDP.

4.1. Load Balancing-based Quota Scheduling

LB-QS aims to equalize the average traffic of peak and off-peak hour. Let Q_{peak} and $Q_{off\text{-}peak}$ be the quotas allocated by the NSP to individual users for peak and off-peak hours. Let T_{peak} and $T_{off\text{-}peak}$ be the corresponding total time lengths, and I be the total number of users. The average traffic for peak and off-peak hour is estimated as IQ_{peak}/T_{peak} and $IQ_{off\text{-}peak}/\ F_{off\text{-}peak}$, respectively. To balance the traffic load, i.e.,

$$\frac{IQ_{peak}}{T_{peak}} = \frac{IQ_{off-peak}}{T_{off-peak}} \qquad (11)$$

, it is concluded that the allocated quota of one time period is proportional to its corresponding time length. As a result, the quotas for peak and off-peak hours are calculated as

$$Q_{peak} = \frac{T_{peak}}{T_{peak}+T_{off-peak}}Q\text{, and}$$

$$Q_{off-peak} = \frac{T_{off-peak}}{T_{peak}+T_{off-peak}}Q\,. \qquad (12)$$

4.2. Peak Shaving-based Quota Scheduling

The goal of PS-QS is to reduce the traffic of peak hours. We regard all users as an aggregate one, i.e. all users have the same quota allocation behavior. In the aggregate user model, the quota allocation of a user in one time period is proportional to the total submitted volume collected from a QPC network. Accordingly, the quotas of a user allocated for peak and off-peak hours can be calculated by

$$Q'_{peak} = \frac{\sum_{k\in\text{peak hours}}\sum_{i=1}^{I} v_{i,k,QPC}}{\sum_{k\in\text{peak hours}}\sum_{i=1}^{I} v_{i,k,QPC} + \sum_{k\in\text{off-peak hours}}\sum_{i=1}^{I} v_{i,k,QPC}}Q\text{, and}$$

$$Q'_{off-peak} = \frac{\sum_{k\in\text{off-peak hours}}\sum_{i=1}^{I} v_{i,k,QPC}}{\sum_{k\in\text{peak hours}}\sum_{i=1}^{I} v_{i,k,QPC} + \sum_{k\in\text{off-peak hours}}\sum_{i=1}^{I} v_{i,k,QPC}}Q\,. \qquad (13)$$

At time slot *k*, there is 100* d_k percent of total submission exceeding the link capacity. In the worst case, there is 100*max{$d_k|k\in$peak hours} percent of packet loss. To level off the peak-hour traffic below the link capacity, the network manager should conservatively reduce 100*max{$d_k|k\in$peak hours} percent of total submission. As a result, the quota scheduled for peak hours is designed as

$$Q_{peak} = Q'_{peak}(1-\max\{d_k \mid k\in \text{peak hours}\})\,. \qquad (14)$$

The quota for off-peak hours is thus

$$Q_{off-peak} = Q'_{off-peak} + Q'_{peak}\cdot\max\{d_k \mid k\in\text{peak hours}\}. \qquad (15)$$

5. COMPARISONS OF VIRTUAL PRICING AND QUOTA SCHEDULING

To address challenge (C2), we quantitatively evaluate and compare the performances of LB-QS, PS-QS and TDP in abuse and fairness reduction and load balancing and peak shaving effects by exploiting the empirical data of NTU dormitory networks with 5535 users. Peak hours are from 9 a.m. To 3 a.m. The bottleneck bandwidth for outbound traffic is 54Mbps. As the NTU network management system collects metering data once every 10 minutes, accordingly, the length of a time slot is set to 10 minutes. The daily quota for each user is 1G. The form of user utility is assumed as $F(v_{i,k})=\log v_{i,k}$ to satisfy the diminishing returns to scale [13, 14].

It is hypothesized in this numerical experiment that the peak-hour congestion is more effectively alleviated under PS-QS than under LB-QS since PS-QS grasps user quota allocation through empirical data to limit user usage of peak hours. TDP outperforms QS because TDP utilizes more empirical data (two data collections) for user behavior modelling and gives users incentive to shift the peak-hour demand to off-peak hours.

Under LB-QS, the scheduled quotas for peak and off-peak hours are calculated as (Q_{peak}, $Q_{off\text{-}peak}$)=(750MB, 250MB), and under PS-QS, (Q_{peak}, $Q_{off\text{-}peak}$)=(620MB, 380MB). The optimal price profile of TDP is computed as $(p^*_{peak}, p^*_{off-peak}) = (1.3, 1.1)$. Note that under QS, a user may have different usage behaviors (myopicity or prudence) during peak and off-peak hours, which are based on the corresponding demands. However, under

TDP, whether a user is myopic or prudent depends on the daily total demand.

5.1. Load Balancing and Peak Shaving

TABLE I presents peak shaving and load balancing indices under LB-QS, PS-QS and TDP, where peak shaving index (PSI) is defined as the average total submission rate of peak hours, and the difference of average total submission rates between peak and off-peak hours serves as load balancing index (LBI). It reveals that TDP has average improvements of 24% in LBI and 9% in PSI over QS schemes. LBI and PSI under PS-QS are improved by 37.6% and 4.7% respectively as compared to LB-QS. Amazingly, LB-QS is worse than PS-QS in LBI. It is because LB-QS does not take actual user preferences over time into consideration (empirical user data) even though the goal of the LB-QS design is to make the traffic load balanced.

TABLE I
PSI AND LBI UNDER LB-QS, PS-QS AND TDP

	LB-QS	PS-QS	TDP
LBI (Mbps)	16	9.98	9.28
PSI (Mbps)	57.42	54.75	51.35

The total submission rates of regular service under LB-BS, PS-QS and TDP are depicted in Figure 2. It is discovered that under LB-QS and PS-QS, there exists a spike at 9 a.m. The reason is that QS leads to a huge amount of user usage at the time of quota replenishment, while TDP utilize the prices to adjust user usage. In comparison with LS-QS, PS-QS encourages more user usage during off-peak hours. Figure 3 shows the patterns of drop rates over time. TDP reduces the drop rate to 0 in average all the time. In QS, the peak-hour drop rate under PS-QS is significantly reduced by 70% over LB-QS.

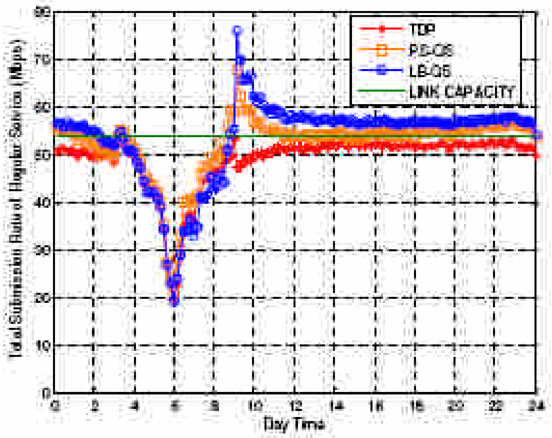


Figure 2: Total submission rate under LB-QS, PS-QS and TDP

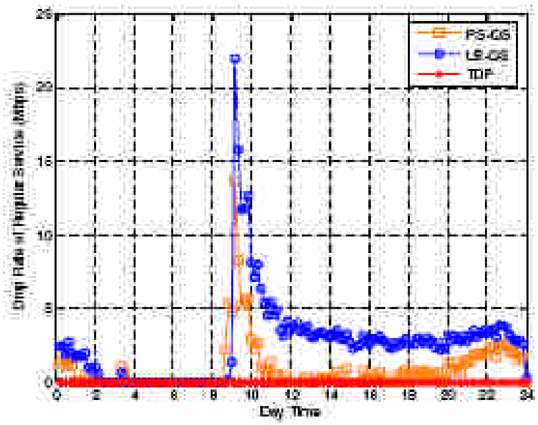


Figure 3: Drop rate under LB-QS, PS-QS and TDP

5.2. Abuse and Fairness Improvement in Peak Hours

Let the Internet access volume by top 5 users be the abuse index (AI). The standard deviation among all users' usage for Internet access is chosen as fairness index (FI). TABLE II demonstrates AI and FI under LB-QS, PS-QS and TDP during peak hours, indicating that TDP outperforms LB-QS and PS-QS by at least 14%. The reason is that TDP give users incentive to shift their peak-hour demand to off-peak hours through differentiated prices. AI and FI under PS-QS are reduced by 9% and 7% respectively over LB-QS.

TABLE II
AI AND FI UNDER LB-QS, PS-QS AND TDP IN PEAK HOURS

	LB-QS	PS-QS	TDP
AI (bytes)	226964566	206758615	173572422
FI (bytes)	2631251	2440906	2107364

Figures 4 and 5 depicted the AI and FI over time. Although TDP has the best performances during peak hours, AI and FI under TDP are the worst during off-peak hours. However, it does not matter for a network manager because there is still bandwidth left during off-peak hours.

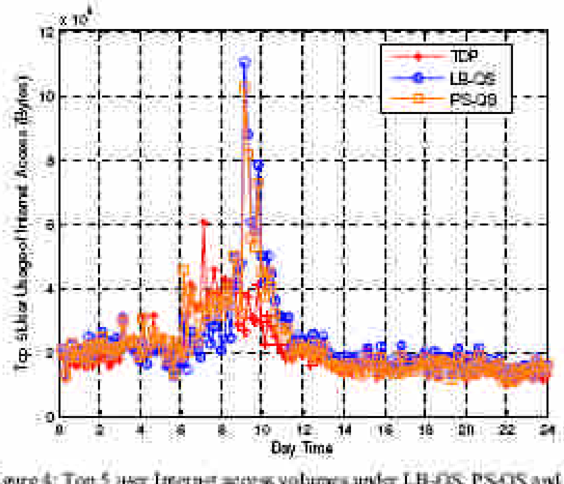


Figure 4: Top 5 user Internet access volumes under LB-QS, PS-QS and TDP

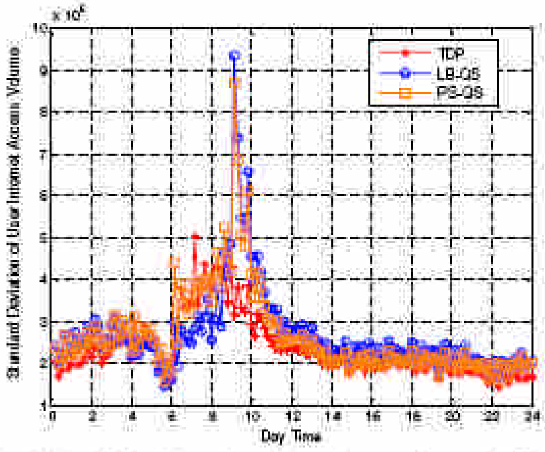


Figure5: Standard deviations of user Internet access volume under LB-QS, PS-QS and TDP

5.3. Design and Implementation Related Issues

In view of challenge (C3), the issues about policy design in practice are discussed as follows.

Measurement Requirement

The TDP design requires measurement data from a baseline network and a network with QPC. The former is used to characterize individual user demand for classification. The latter estimates user preferences over time for constructing user demand model. When designing PS-QS, the NSP only needs empirical data measured from a QPC network to construct an aggregate user quota allocation model. However, there is no measurement data needed for the LB-QS design.

Complexity of Calculation

The quota calculation of LB-QS and PS-QS only needs simple mathematical operations. The TDP design has to solve an optimization problem. It takes more the computation time than LB-QS and PS-QS, but is acceptable.

Implementation Requirements

Over an existing QPC network, TDP requires a pricing calculation module, while LB-QS and PS-QS need quota scheduling modules. They are easily implemented and integrated on the accounting and traffic control server in Figure 1.

Applicability to Traffic Pattern

If the peak hours are not contiguous but scatter over all time slots, TDP is better than LB-QS and PS-QS. It is because the NSP needs to design more quotas for more time periods, which leads to sever congestion at quota replenishment time from the observations of Figure 2.

6. CONCLUSIONS

In this paper, we designed LB-QS and PS-QS for comparison and evaluation with TDP for time-of-day Internet access management over a QPC network. The key idea of LB-QS is to equalize the average traffic of peak and off-peak hours. PS-QS requires the measurement data collected from a QPC network to approximate an aggregate user model for peak-hour traffic reduction. TDP adopts a general utility and exploits two data sets gathered from a baseline network and a QPC network, respectively to model user behavior for the optimal price design. Evaluations show that the peak-hour abuse and fairness under TDP are improved by at least 14% over QS. TDP has also the great improvements on load balancing and peak shaving, and avoids the congestion at the time of quota renewal. In QS, the peak-hour drop rate under PS-QS is significantly ameliorated by 70% over LB-QS. The TDP design requires solving an optimization problem, but its computation time is acceptable. A network manager should gauge data availability and actual traffic pattern over time to make the appropriate decision. Generally speaking, TDP is suggested as the most effective way for time-of-day traffic management.

REFERENCES

[1] T.-C. Lin, Y. S. Sun, S.-C. Chang, S.-I Chu, Y.-T. Chou, and M.-W. Li, "Management of Abusive and Unfair Internet Access by Quota-based Priority Control," *Computer Networks*, vol. 44, pp. 441-462, 2004.

[2] www.p-cube.com

[3] www.packeteer.com

[4] E. D. Farmer, B. J. Cory, and B. L. P. P. Perera, , "Optimal Pricing of Transmission and Distribution Services in Electricity Supply," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 142, issue 1, pp. 1-8, Jan. 1995.

[5] N. V. Pillai, "A Contribution to Peak Load Pricing Theory and Application," http://ideas.repec.org/p/ind/cdswpp/346.html.

[6] I. Ch. Paschalidis, and J. N. Tsitsiklis, "Congestion-dependent Pricing of Network Services," *IEEE/ACM Transactions on Networking*, vol. 8, (no. 2), pp. 171-84, Apr. 2000.

[7] J. S. Shih, R. H. Katz and A. D. Joseph, "Pricing Experiments for a Computer-telephony-service Usage Allocation," *Proc. of IEEE GLOBECOM'01*, vol. 4, pp.2450-2454, Nov. 2001.

[8] C. Lambrecht, and O. Verscheure, "Perceptual Quality Measure Using a Spatio-temporal Model of Human Visual System," *Proc. of IS&T/SPIE*, Feb. 1996.

[9] A. Watson, and M. A. Sasse, "Evaluating Audio and Video Quality in Low-cost Multimedia Conferencing Systems," *Interacting with Computers*, vol. 8, pp. 255-275, 1996.

[10] M. Minoux, *Mathematical Programming: Theory and Algorithms*, Wiley, Chichester, 1986.

[11] S.-I Chu and S.-C. Chang, "Time-of-day Internet Access Management by Combining Empirical Data-based Pricing with Quota-based Priority Control," submitted to *IEE Proceedings-Communications*.

[12] Martin J. Osborne, *An Introduction to Game Theory*, Oxford University Press, 2003.

[13] F. P. Kelly, "Charging and Rate Control for Elastic Traffic," *European Trans. Telecom*, vol. 8, pp. 33-37, 1997.

[14] X. Wang and H. Schulzrinne, "Pricing Network Resources for Adaptive Applications in a Differentiated Services Network," *Proc. of IEEE INFOCOM'01*, vol. 2, pp. 943-952, Apr. 2001.

出席國際會議報告書

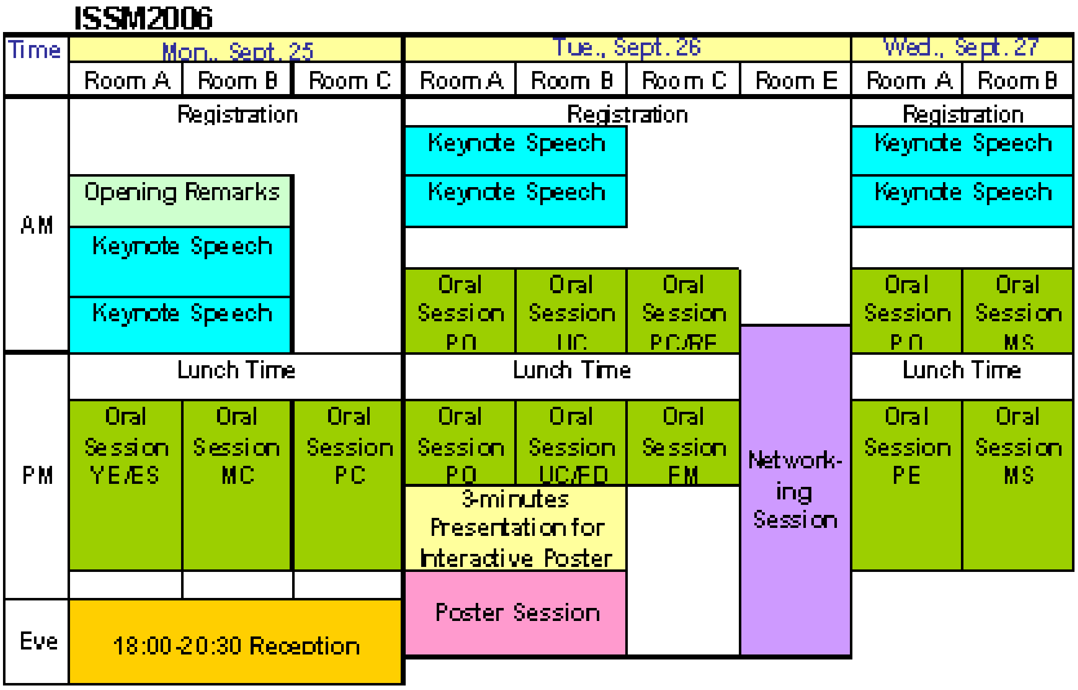
報告人姓名	張時中	所屬系所	電機工程學系
會議時間地點	Century Hyatt Tokyo, Japan. Monday, September 25 – Wednesday, September 27, 2006		
會議名稱	INTERNATIONAL SYMPOSIUM ON SEMICONDUCTOR MANUFACTURING (ISSM 2006)		
發表論文題目	1. Priority X-Factor Modeling for Differentiated Manufacturing Service Planning 2. Priority Behavior Modeling of Fab for Supply Chain Management		

五、　會議簡介

ISSM is the industry's largest forum of semiconductor manufacturing professionals dedicated to sharing technical solutions and opinions on the advancement of manufacturing science. The highlight topics of ISSM 2006 include, process control maturation, application of Taguchi Method, DFM-total optimization for 65nm and beyond, systematic productivity improvement, fab extendibility and flexibility, application-specific semiconductor manufacturing, SiP, 3D modules, Environmental and safety, nanometer-level contamination control, challenges for 450mm fab, and new business model to meet with time-to-market. The Society of Applied Physics of Japan, IEEE Electron Devices Society, and Semiconductor Equipment and Materials International (SEMI) offer ISSM as a forum to broaden semiconductor manufacturing knowledge.

六、　參加會議經過

Shi-Chung Chang served as a technical program committee member, presented one oral session paper in the 09/25 afternoon session of Manufacturing control and one poster session paper in the afternoon of 09/26, and served as a session chair. The program attended is attached below.　Shi-Chung Chang met the new TPC chair, Mr. Thomas Sounderman about ISSM2007.



ISSM2006

Time	Mon., Sept. 25			Tue., Sept. 26				Wed., Sept. 27	
	Room A	Room B	Room C	Room A	Room B	Room C	Room E	Room A	Room B
AM	Registration			Registration				Registration	
				Keynote Speech				Keynote Speech	
	Opening Remarks			Keynote Speech				Keynote Speech	
	Keynote Speech								
	Keynote Speech			Oral Session PO	Oral Session UC	Oral Session PC/RE	Networking Session	Oral Session PO	Oral Session MS
PM	Lunch Time			Lunch Time				Lunch Time	
	Oral Session YE/ES	Oral Session MC	Oral Session PC	Oral Session PO	Oral Session UC/FD	Oral Session FM		Oral Session PE	Oral Session MS
				3-minutes Presentation for Interactive Poster					
Eve	18:00-20:30 Reception			Poster Session					

- Factory Design (FD)
- Manufacturing Control and Execution (MC)
- Manufacturing Strategy and Structure (MS)
- Process Control and Monitoring (PC)
- Process and Metrology Equipment (PE)
- Yield Enhancement Methodology (YE)
- Ultraclean Technology (UC)
- Environment, Safety and Health (ES)
- Process and Material Optimization (PO)
- Final Manufacturing (FM)
- Robust Engineering (RE)

七、 與會心得

Researchers from Taiwan continue to show a strong presence in ISSM. Design for manufacturing (DFM) and Supply Chain Management have gain significant growth in attendees' interest. Industrial participation from Taiwan was largely from TSMC.

八、 攜回資料名稱及內容

1. PROCEEDINGS in a CD

ISSM2006
MC- 231

Priority X-Factor Modeling for Differentiated Manufacturing Service Planning

Shi-Chung Chang Ke-Ju Chen
scchang@cc.ee.ntu.edu.tw r93921070@ntu.edu.tw
Dept. of Electrical Engineering, National Taiwan University, Taipei, Taiwan, ROC, 10617
Phone: +886 –2 –2362-5187 Fax: +886 –2 –2363-8247

Abstract – This paper addresses the X-factor modeling needs in fab capacity and release rate planning for differentiated manufacturing service provision. A priority X-Factor constrained planning problem is first formulated that describes the relation among profit, release rates of individual priorities, and capacity utilization. Modeling priority X-factors is key to the formulation. We deign a novel M/G/1:PR queue approximation-based network modeling methodology to capture in a scalable way how operation priority, production flow variations, and capacity utilizations may affect individual PXFs of overall fab and tool groups. Numerical studies demonstrate the potential applications of our PXF models.

I. INTRODUCTION

Effective provision of manufacturing services in multiple priority levels has been one critical aspect to the competitiveness of wafer fabs. A customer order with a higher priority level demands a shorter cycle time than orders of a lower priority. Wafers of lower priority orders have elongated cycle times because they need to wait in line until wafers of higher priority orders finish processing. Machine capacity loss may occur when processing a high priority order by a batching machine without requiring a full load policy. Priority mix percentage significantly affects the variations of fab performance such as throughput, cycle time, wafer-in-process (WIP) and bottleneck location [1].

Among the many fab performance indices, cycle time has a significant impact on productivity learning and customer serviceability. There is a basic relationship among capacity utilization (U), throughput (T) and cycle time [2]. The cycle time of a fab increases exponentially with the increase of U/T when U/T goes beyond a high level, say, 90%, while it is proportional to U/T at a lower level. To measure and manage cycle times, the notion of X-factor (XF), where

XF = cycle time/raw processing time (RPT)

has been introduced to provide a sensitive performance indicator and is standardized across different products [3].

It has been shown that many fab opration problems can be effectively identified through the analysis of X-factors. Customized X-factor targets can be set for short cycle time manufacturing (SCM) that not only allow the performance differentiation among toolsets of different characteristics but also guarantee the overall fab objective [3-6].

In production planning of a fab, there are different XF target (XFT) specifications for individual priority levels of manufacturing services [7]. The XF of each priority (PXF) is a function of release rates of individual priorities and the total utilization of the bottleneck tool group, which we shall refer to as a PXF behavior model. Note that different XFs require different levels of resources and hence lead to different costs and manufacturing services of different XFTs should be priced differently. Given a pricing policy, capacity cost structure, and a set of XFT, a PXFT constrained production planning decides the priority mix (or wafer release rats) of products in individual manufacturing service priorities for profit maximization subject to machine capacity and PXFT constraints. Key to this planning problem is the behavior modeling of the relationship between PXF and priority mix and capacity utilizations.

Motivated by the problem of PXFT constrained production planning. In this paper, we design and develop an M/G/1:PR queue approximation-based network modeling methodology with a focus on capturing how operation priority, production flow variations, and capacity utilizations may affect individual PXFs of a fab. The M/G/1:PR queue model is adopted to model the behavior of a service node (tool group). On top of the single node model, we derive a PXF contribution theory that relates PXFs of individual service nodes to the overall fab PXF and provides a novel priority network model. Model fitting is then adopted to fit the M/G/1:PR-based approximation to empirical fab data. The key idea of fitting is to add a parameter in calculating the mean residual service time, which compensates, for each priority, the effect of arrival process variation to the mean service time at a node. Figure 1 depicts the concept.

The remainder of the paper is organized as follows. Section II gives a problem formulation of PXFT constrained production planning and identify the need for a PXF behavior model. Section III then presents the M/G/1:PR queue approximation and priority contribution theory-based network modeling methodology. Model analysis and applications are given in Section IV. Finally, Section V concludes the paper.

II. PRIORITY X-FACTOR CONSTRAINED PLANNING

Consider a fab with a given set of prespecified PXFTs for individual priorities that need to be achieved by fab operations. Prices and WIP costs of individual priorities and capacity costs are also given. A fab manager can control XFs of the fab, individual stages and priorities at each stage by adjusting priority mix and utilization levels while maximizing the profit.

To formulate the XF constrained planning problem, let us first define some notations, where we assume for simplicity of discussions that there is only one type of products in each priority.

Notations

XF_{ij}: X-Factor of j-th priority product at processing step i;

λ_j: Mean release rate of jth priority product;

τ_{ij}: Mean service time of a jth priority wafer at step i;

RPT_j: Average processing time of a wafer at step i;

$Var[S_{ij}]$: Variance of service time of jth priority at step i;

ρ_{ij}: Utilization of jth priority product in step i;

P_j: Per wafer price of jth priority;

c_j^M: Per wafer manufacturing cost of jth priority;

c_{ij}^I: Per wafer and per unit time WIP cost of jth priority at ith step.

The revenue rate of manufacturing wafers of one priority comes from its offered price and wafer release rate of the priority, while the WIP cost at a step is proportional to the release rate and cycle time at the step and the capcity cost is proportional to the release rate. In deciding on the release rates or mix of individual priorities to maximize one's manufacturing profit, a fab manager must consider constraints of capacity and PXF targets. A PXFT constrained planning problem is formulated as follows:

$$\underset{\{\lambda_j\}}{Max}\ J = \sum_j [Revenue_j(\lambda_j) - Cost_j(\{XF_{ij}\}, \lambda_j)]$$

$$= \sum_j [\lambda_j \times P_j - (c_j^I \times \lambda_j \times \tau_{ij} \times XF_{ij} + c_j^M \times \lambda_j)]$$

subject to

$$\Phi_{p2f}(\{\lambda_j\}) \leq XFT_{fab};$$

$$\Phi_{p2s\_j}(\{\lambda_j\}) \leq XFT_{\mathrm{Pr}\_j}, \forall j; \text{ and}$$

$$\sum_{(i,j)\ uses\ \text{tool group}\ m} \lambda_j \times \tau_{ij} < C_m, \forall m;$$

with targets XFT_{fab}, $\{XFT_{\mathrm{Pr}\_j}, \forall j\}$ and capacity $\{C_m, \forall m\}$ given.

In the formulation, function Φ_{p2f} respresents a model of the relationship beween priority release rates and XF of the wole fab while function Φ_{p2s_j} respresents a model of the relationship beween priority release rates and the XF of priority j at a step s. These functions have to be constructed for PXFT constrained planning.

III. MODELING PRIORITY X-FACTORS

Exploiting available closed-form results, we first adopt a M/G/1:PR queueing approximation to model the PXF of a service node (tool group). As the Poisson arrival assumption of the M/G/1:PR model may not be a good approximation, the model needs to be modified for a closer match to fab data. Network relationship among nodes is then needed for modeling fabwide PXF of each priority based on nodal models. Our modeling approach is described as follows.

III.1 Single Node Model: M/G/1:PR approximation

Key to the fab PXF behavior model . is the single-node priority behavior modeling. We modified the M/G/1:PR results of [8] by adding a compensation parameter α_j to account for the variance of a general, non-Poisson arrival process. The approximation model of XF for priority j is then

$$XF_j \equiv \Phi_{p2s\_j}(\{\lambda_k\})$$

$$\approx \frac{1}{\tau_j} \frac{\frac{1}{2}\sum_{k=1}^{J}(Var[S_k] + \tau_k^2) \times \lambda_k + \alpha_j}{(1 - \sum_{k=1}^{j-1}\rho_k)(1 - \sum_{i=1}^{j}\rho_k)} + 1,$$

where $\tau_j \equiv \sum_i \tau_{ij}$ and $\rho_j = \lambda_j \times \tau_j$.

Figures 3 and 4 depict numerical studies of the properties of the modified M/G/1: PR approximation over a two-priority service node example, where wafer processing requirements are the same for both priority. Model data setting refers to that provided by ITRS [7]. There is not much difference in PXF of priority-1 when priority-1 mix varies in the low percentage range of 1-10%. Note that under a fixed mix percentage, the PXF of priority-1 increases almost linearly with respect to overall capacity utilization while the PXF of priority-2 increases drastically when utilization is higher than 90%. The lower priority is much more sensitive to capacity utilization and priority mix than the higher priority.

II.2 Priority XF Contribution Theory

In [3], D.P. Martin derived a contribution theory that describes the contribution by XFs of individual processing step/tools to fab XF as following:

$$
\begin{aligned}
XF_{fab} &= \sum_{i=1}^{I} XF\,Contribution_{step\_i} \\
&= \sum_{i=1}^{I} \frac{RPT_i}{RPT_{fab}} \times XF_i \\
&= \Phi_{s2f}(\{RPT_i, XF_i\})
\end{aligned}
$$

where $RPT_{ij} = \tau_{ij}$ and Φ_{s2f} respresents the relationship beween XF at a tool group to fab XF.

In this paper, we exploit the relationship between total queue size and queue sizes of individual priorities, the relation between total cycle time and cycle times of individual priorities, and Little's formula, to obtain the contribution by XFs of individual priorities to the XF of step i as

$$
\begin{aligned}
XF_i &= \sum_{j=1}^{m} (\text{relative utilization})_j \cdot XF_j \\
&= \sum_{j=1}^{J} \frac{\mathrm{WL}_{ij}}{\sum_{j=1}^{m} \mathrm{WL}_{ij}} \times XF_{ij}(\mathrm{WL}_{ij}) = \Phi_{p2s}(\ \mathrm{XF}_{ij})
\end{aligned}
$$

where $\mathrm{WL}_{ij} = \lambda_j \times\ \tau_{ij}$.

We then obtain the contribution by XFs of individual priorities to fab XF:

$$
XF_{p2f\_j} = \Phi_{p2f\_j}(\lambda_j)
$$

$$
= \sum_{i=1}^{I} \frac{RPT_{stage\_i}}{RPT_{fab}} \times \frac{WL_{ij}}{\sum_{j=1}^{m} WL_{ij}} \times XF_{ij}(WL_{ij})
$$

The Priority XF contribution theory above assumes given PXF behavior models at individual nodes in a fab. Figure 2 shows the relationship among PXFs in the behavior model.

IV. APPLICATIONS OF PXF MODEL TO PLANNING

Construction of the PXF behavior model completes the PXF target constrained production planning problem. Figures 5 gives results of numerical studies of the planning problem. The same two-priority example for Figures 3 and 4 is examined, where release rate of individual priorities are changed. Figure 5 shows that profit increases with capacity utilization under different priority mixes (PM). Under the given PXFTs, there is a maximum profit level among various mixes for each utilization level as indicated by a red solid curve in Figure 5. Our analysis show that the XF of the lower priority is very sensitive to slight variation of higher priority relase rate. When the bottleneck capacity utilization goes beyond 82% in this example, the WIP cost of the lower priority surpassses the gain in revenue and the maximum profit drops. We further apply the model to address the following two questions.

Q1: How total utilization should be adjusted with respect to PM change under given PXF targets?

Optimization tools can be applied to compute optimal capacity allocations and release rates of individual priorities. A fab manager may also utilize PXF charts for decision-making. For example, consider Figures 2 and 3 as fab models with priority-1 at 10% and the total capacity utilization at 90%. If there is a new order demanding an increase of priority-1 to 20% with bottleneck capacity still at 90% utilization, then PXF1 = 1.34 and PXF2 = 3.58 and profit increases 2.86% from Figure 5. To maintain PXF1<1.3 and PXF2<3.2, the fab manager may want to reduce the fab utilization to 84% so that PXF targets are achieved at a price of 3.8% total profit decrease. In this case, the tradeoff is between the revenue gain from increase in high priority orders and the loss in reducing low priority wafer release. Besides capacity reduction, one may want to increase the price of priority 1 to recover profit loss in priority 2. So, the model also provides a link for pricing consideration.

Q2: How to find cycle time bottleneck of tool groups?

A two-priority, say *P1* and *P2*, experiment is designed for investigation. The numbers of operation steps of *P1* and *P2* products are 32 and

60, respectively. There are 12 service nodes (tool groups). Release processes of the two priority wafers are Poisson while service time distributions contain uniform, erlang-k and exponential distributions. This example basically follows that in [9].

Figures 6 and 7 show comparisons of XFs between simulation results and M/G/1: PR + PXF contribution theory for priority 1and 2 under 12 tool groups with different service time distribution. Table 1 shows the error percentages of each tool group. Figure 8 shows XF contributions of each node to fab XF and we can identify that tool group 5 is the cycle time bottleneck but the capacity bottleneck is tool group 3.

V. CONCLUSIONS

In this paper, we have designed and developed an M/G/1:PR queue approximation-based network modeling methodology. The PXF models obtained captures how operation priority, production flow variations, and capacity utilizations may affect individual PXFs of a fab. We have also demonstrated good application potential of these models to the planning of priority manufacturing services..

Figure 1: PXF model fitting

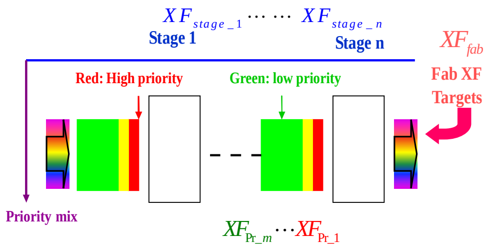


Figure 2: the architecture of the Fab behavior mode

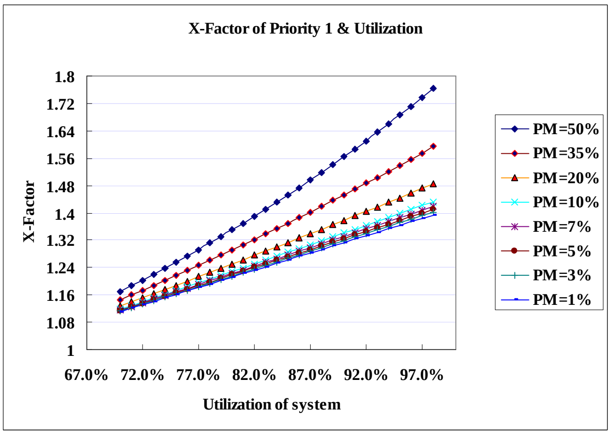


Figure 3: Relation between PXF and utilization for priority 1

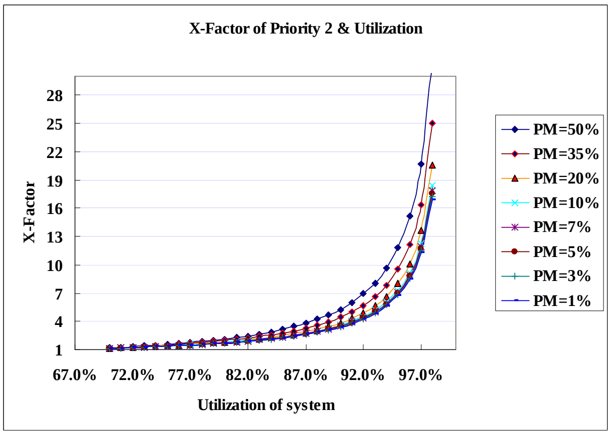


Figure 4: Relation between XF and utilization for priority 2

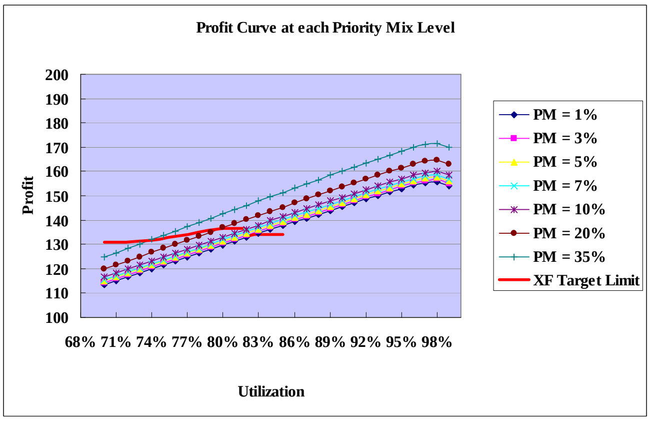


Figure 5: Relation between Profit and utilization under different priority mix

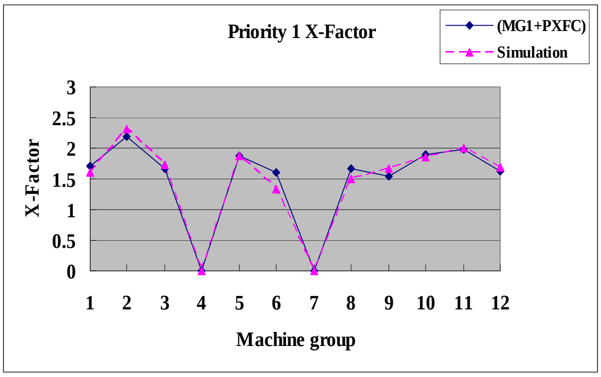


Figure 6: Comparison of X Factor between Simulation and MG1+PXFC for priority 1

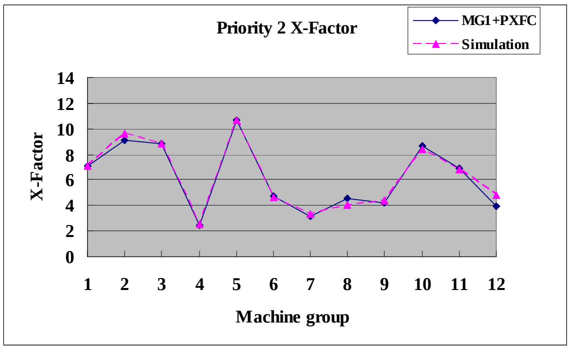


Figure 7: Comparison of X Factor between

Simulation and MG1+PXFC for priority 2

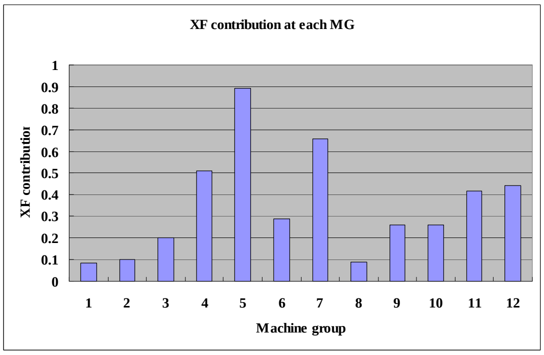


Figure 8: XF contributions of each node to fab XF

ACKNOWLEDGEMENT

* This work was supported in part by the Semiconductor Research Corporation and International Semiconductor Manufacturing Initiative under FORCe-II project 1214 and by the National Science Council, Taiwan, ROC, under grants NSC-93-2213-E-002-043 and 94-2213-E-002-015

REFERENCES

[1] Amy H.I. Lee, He-Yau Kang, Wen-Pai Wang "Analysis of priority mix planning for the fabrication of semiconductors under uncertainty," *Int. J. Adv. Manuf. Tech.* (2006) 28: 351–361.

[2] W. Hopp, M. Spearman, *Factory Physics*, 2^{nd} *ed.*, MacGraw-Hill Higher Education, 2000.

[3] D. P. Martin," The Advantage of Using Short Cycle Time Manufacturing (SMC) Instead of Continuous Flow Manufacturing (CFM)," *Proc. of ASMC*, 1998, pp.89-94.

[4] Y. Narahari and L. M. Khan" Modeling the Effect of Hot Lots in Semiconductor Manufacturing Systems," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no.1, Februnary 1997.

[5] M. Kishimoto, K. Ozawa, K. Watanabe, and D. Martin, "Optimized Operations by Extended X-Factor Theory Including Unit Hours Concept" *IEEE Transactions on Semiconductor Manufacturing*, vol. 14, no. 3, Aug. 2001.

[6] J. Robinson and F. Chance," Dynamic X-Factor application for Optimizing Lot Control for Agile Manufacturing,"*Proceedings of ISSM*, Tokyo, Oct. 2002.

[7] *International Technology Roadmap for Semiconductors 2005 edition*, *Factory Integartion*, http://www.itrs.net/reports.html.

[8] J. Virtamo," 38.143 Queueing Theory / Priority queues," Helsinki University of Technology, Fall 2001.

[9] Ming-Der Hu and S.C. Chang, "Translating Delivery Specifications into Distributed Flow Control Requirements for Re-entrant Production Lines" *PhD dissertation*, EE, NTU 2000.

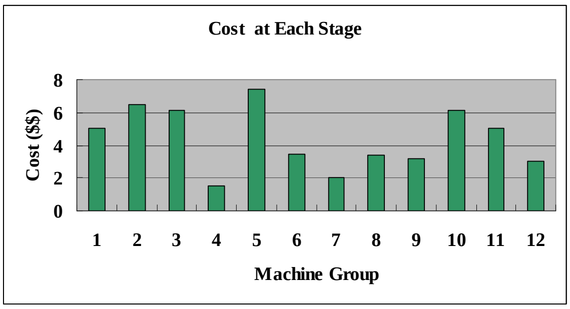


Figure 9: Cost at each node

Priority Behavior Modeling of Fab for Supply Chain Management*

Shi-Chung Chang scchang@cc.ee.ntu.edu.tw
Bo-Jiun Liao r93546017@ntu.edu.tw
Argon Chen achen@ntu.edu.tw

Graduate Institute of Industrial Engineering, National Taiwan University
No.1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 10617
Phone: +886 -223625187 Fax: +886-23638247

Abstract – *This paper develops modeling methods and fab behavior models with predictability and scalability to capture variability and manufacturing service differentiation in semiconductor supply chain management. A novel, hybrid decomposition approximation-based priority queueing network model is designed for fab behavior modeling. The model characterizes the relationship between output performance metrics of cycle time, wafer-in-process, throughput, and capacity utilization and input factors of priority mix, wafer release, capacity allocation and machine characteristics. Model evaluation results over some fab models demonstrate that the hybrid decomposition approximation-based network model yields very quick and good quality estimations of mean and variability of tool group and fab performance metrics.*

I. INTRODUCTION

A supply chain is a system of nodes that provides manufacturing services—in fact, a variety of services. Services differentiation, namely, prioritization, is common in operations of semiconductor supply chains (SSC, Figure 1). It affects how to allocate resources and charge prices. Such a new paradigm of manufacturing services requires new methods of operation control. The grand challenges will be scalability and predictability with respect to differentiation of services and variability that are exacerbated by rapidly increasing product varieties and process variations in the chains.

To provide service with differentiable ensures that the quality of service (QoS), allocation of manufacturing capacity and pricing of services have to be dependent of and differentiated by QoS requirements. Product, process and operation variability affects the performance of individual service nodes such as tool groups and fabs and chains/networks of service nodes. In order to predict the behavior of the SSC that provide differentiated services, research is needed in follow aspects: predictable and scalable performance metrics with respect to the chain structure, and fundamental understanding of the behavior of service nodes and chains under variability.

Among the SSC service nodes, fab is the most expensive, complicated, and important. So this study is aimed at the behavior modeling of a fab and provides a cornerstone model for supply chain management. In behavior modeling, the output performance metrics consist of cycle time, wafer-in-process (WIP), throughput, and capacity utilization. And input options to a fab include priority mix, wafer release, capacity allocation and machine characteristics. The behavior model describes the relationship between output performance metrics and inputs, which are characterized by not only mean values but also variability.

In this paper, we develop behavior models and modeling methods that enhance the scalability and predictability of the semiconductor supply chains with respect to varieties, and differentiability of services. We aim at fab behavior modeling that provides a cornerstone model for SSC management. The network-based fab behavior models describe how priority, resource allocation and sources of variations affect fab performance metrics such as mean and variability of cycle time, wafer-in-process, throughputs, and machine utilizations as shown in Figure 2. The input/output relationship of a fab will be modeled as a network of priority service nodes [1]. Such a priority network captures factors and effects of variations throughout the network model. The model is scalable and allows chain/network performance metrics to be decomposed into per node and per priority metrics. It has also predictability that allows very quick evaluation of mean and variability of both nodal and system output performance metrics with various priority input options.

II. FAB BEHAVIOR MODELING BY HYBRID DECOMPOSITION OF PRIORITY NETWORK

Consider a fab with multiple part priorities, failure prone machines, and re-entrant process flows. We model the fab as a failure-free, batch-free, deterministic-feedback and priority open queueing network (OQN) [4]. Then we design and develop an innovative, hybrid decomposition approximation-based approach for such a priority OQN with a focus on capturing operation priority and variations in a fab.

Key ingredients of the hybrid decomposition approach for modeling a priority fab OQN are as follows:

1. Decompose the fab network into many independent service nodes and their networking relationship by adopting the decomposition approximation of queueing network analyzer (QNA, [3]).
2. Model single service node behavior by sequential decomposition approximation (SDA) of coupling interactions among priorities in one node ([2], Figure 3;
3. Combine the networking relationship among service nodes with a fixed point iteration to approximate re-entrant flow line performance.

More details will be given in the following subsections.

II.1 Nodal Model: Sequential Priority Decomposition

The behavior modeling of priority single service node provides a cornerstone for the fab behavior modeling. We develop the nodal behavior model with priority by treating each service node independently as a GI/G/1 non-preemptive priority queue, and then adopt the sequential decomposition approximation (SDA) proposed by [2] among priorities for our behavior modeling. SDA decomposes the coupling among priorities into approximately independent queues for individual priorities as shown in Figure 4 by the notion of equivalent service time.

SDA determines "equivalent" service parameters for each queue by taking the interactions with other queues into consideration. The most important feature to consider in each individual queue is that the service time of a job arriving into an empty queue differs from one arriving into a non-empty queue. If a job arrives into an empty queue, the equivalent service time is measured from its arrival; else if a job arrives into a non-empty queue, its equivalent service time is measured from the departure of the previous job having the same priority to its departure. Given the equivalent service time parameters and part release parameters, the means and variances of the overall nodal performances and the departure process of individual priorities can then be obtained.

II.2 Decomposition Approximation-based Queueing Network Modeling of Fab

In combining SDA with QNA [3], a priority open queueing network (OQN) [4] is first developed for a fab with multiple part types, multiple priorities, failure prone machines, and re-entrant process flows. This re-entrant OQN is analyzed by using a class of approximate decomposition methods. For better handling of uncertainties, the aspect of variability, i.e., second order statistics, as well as mean values is adopted to model the characteristics of this fab system. The decomposition methods decompose an OQN into individual network nodes and use two types of parameters to characterize the stochastic arrival, service and departure processes of each node: one describing the rate and the other describing the variability. Various stationary network performance measures, such as cycle time, WIP, and machine utilization, can then be derived based on these two types of parameters.

There is an OQN for each priority. OQNs of individual priorities are coupled through competition of service node resources. The priority coupling is handled by application of the SDA procedure to sequentially solving the equivalent service times from the highest priority in a node. We apply QNA to deal with interactions among nodes: splitting, merging, and deterministic feedback with priority. If a fab has no re-entrant flows, this fab network is tandem queue with flow always in a single direction from one node to the other. Figure 4 depicts a two node-example. The departure parameters of one node are equal to the arrival parameters of next node in tandem queues. We can directly use QNA to separate nodes in the network, and apply SDA to sequentially analyze the performances of each priority in each node. But in a fab, there are re-entrant flows and the arrival parameters of one node are affected by the departure parameters of many other nodes. Figure 5 depicts a simple re-entrant example of two nodes and two priorities. We deal with the re-entrant flows in a priority fab network by combining fixed point iteration [] over SDA and QNA.

III. MODEL EVALUATION

To evaluate the hybrid priority network model, we consider the small example of Figure 5. Figure 6 contrasts the cycle times obtained by hybrid SDA+QNA and simulation and there are good fits. Although there exists some difference in standard deviation for priority 2, the difference does not increase as the number of nodes increases.

Numerical experiments are also conducted on simple but full-scale fab models [6] to examine the efficiency, accuracy and application potential of hybrid decomposition approximation-based queueing network modeling. Also discrete event simulations are developed for validation of the fab behavior model. These fab models have two parts with two priorities classes, and the numbers of processing steps of P1 and P2 are 32 and 60, respectively shown in Figures 7 and 8. In a special fab model (SFM), all the service times of a node have exponential distributions. Other service node data is the same that shown in Table 1. Node and system level cycle times of the SFM are listed in Figure 9 and Table 2 respectively. The differences of mean cycle times of two priorities and the cycle time standard deviation of P2 are mostly within 3%. Although the difference of cycle time standard deviation of P1 is high (up to 45%), the absolute error is only 1.627.

In a general fab model (GFM), the service times of individual nodes have general distributions, such as uniform, erlang and exponential (see Table 1). Node and system level cycle times of the GFM are given in Figure 10 and Table 3 respectively. The differences of mean cycle times of two priorities and the cycle time standard deviation of P2 are mostly within 10%. The cycle time standard deviation of P1, has a high relative error of 55%. But again, the absolute error, 1.214, is still very small as compared to the mean.

Comparisons of numerical results with simulation in these two fab models show that our network modeling methodology has good approximations in most nodal and system performances. However, application of hybrid decomposition approximation to each model (listed in Table 4) only requires less than 4 seconds of CPU time on a 2.8 GHz personal computer. This leads to fast calculation with respect to changes of system input options. Consequently, both the accuracy and computing efficiency of hybrid decomposition approximation-based network modeling support its potential for applications of real fab with service differentiation.

IV. REMARKS ON APPLICATIONS

Responding to rapidly changing complex SSC requirements, SSC planners/managers need an effective performance evaluation/prediction tool to do what-if analysis between SSC inputs and outputs. Given a set of priority mix, capacity, mean and variability of wafer release, mean service time and variability of tools, the hybrid decomposition approximation-based network model allows very quick evaluation of mean and variability of nodal and fab performance metrics with good accuracy.

Effective evaluation of various input options in terms of capacity allocation, priority mix, wafer release policy, tool adjustment, etc. may serve as a behavior model for SSC performance optimization. For example, the mean and variability statistics at individual tool groups may also be utilized by six-sigma management for on time and quick delivery, where the mean values can be used to derive control targets while variability values for calculation of control limits.

ACKNOWLEDGMENT

This work was supported in part by the Semiconductor Research Corporation and International Semiconductor Manufacturing Initiative under FORCe-II project task 1214.001 and by the National Science Council, Taiwan, ROC, under grants NSC-93-2213-E-002-043 and 94-2213-E-002-015.

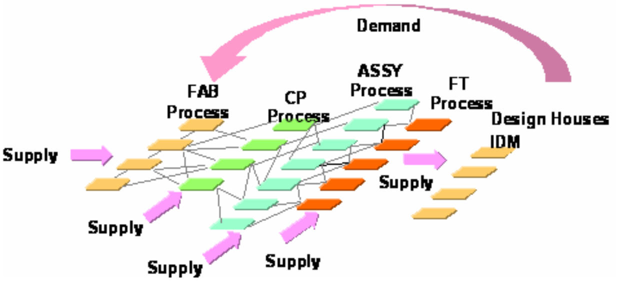


Figure 1: Semiconductor supply chain

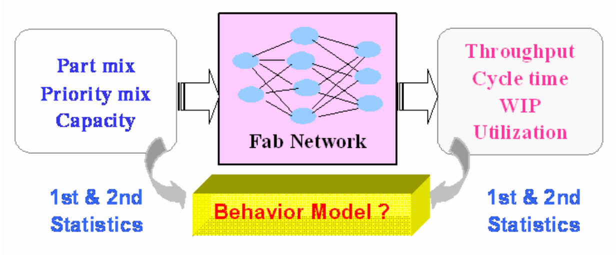


Figure 2: Fab behavior modeling

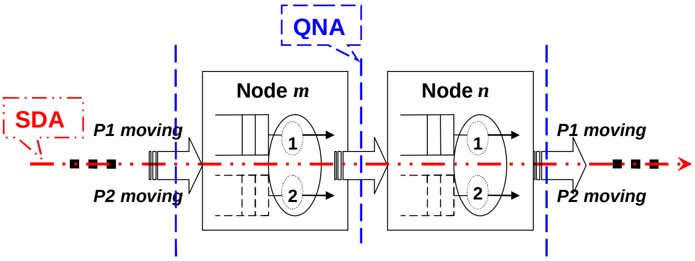


Figure 3: Decomposition approximation-based queueing network modeling

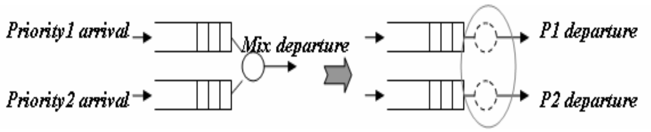


Figure 4: Decomposition among priorities

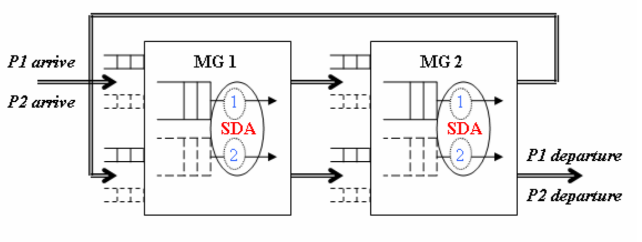


Figure 5: Two stations with reentrant line

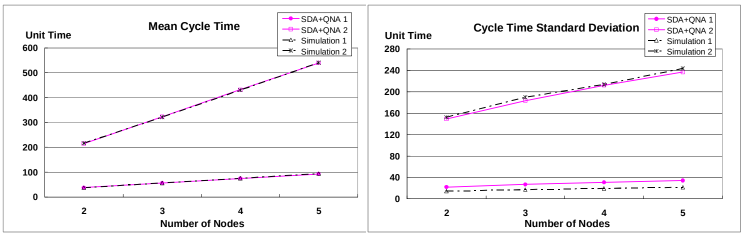


Figure 6: Mean and standard deviation of cycle time for multi-nodes (2~5) with reentrant line

Enter →	1 →	2 →	10 →				
	1 →	2 →	3 →	8 →			
	1 →	2 →	10 →				
	1 →	2 →	11 →	5 →	2 →	3 →	8 →
	1 →	2 →	3 →	8 →	10 →		
	1 →	2 →	3 →	8 →			
	1 →	2 →	6 →	2 →	9 →	12 →	Exit

Figure 7: Process Flow of P1 in 2-PR fab model

Enter →	1 →	2 →	3 →	8 →	10 →		
	1 →	2 →	6 →				
	1 →	2 →	3 →	8 →	9 →		
	1 →	2 →	3 →	8 →	10 →		
	1 →	6 →	11 →	1 →	2 →	5 →	
	1 →	2 →	3 →	8 →	9 →		
	1 →	2 →	3 →	8 →	10 →		
	1 →	6 →	11 →	1 →	2 →	5 →	
	1 →	2 →	3 →	8 →	9 →	10 →	11 →
	1 →	2 →	4 →	8 →	3 →		
	1 →	2 →	7 →	8 →	10 →		
	1 →	2 →	12 →	Exit			

Figure 8: Process Flow of P2 in 2-PR fab model

Table 1: Service processes of nodes for general fab model

Node	# of Machines	Service Time Distribution	MPT (hr/lot)	Utilization %*
1	1	Erlang Order 4	0.125	88.38
2	1	Exponential	0.125	85.23
3	1	Uniform	0.25	91.54
4	1	Erlang Order 3	1.8	68.18
5	1	Erlang Order 2	0.9	90.91
6	1	Erlang Order 4	0.6	83.33
7	1	Exponential	1.8	68.18
8	1	Erlang Order 3	0.2	80.81
9	1	Uniform	0.6	83.33
10	1	Erlang Order 2	0.3333	88.38
11	1	Exponential	0.6	83.33
12	1	Uniform	1.25	78.91

$$^{*}\text{Utilization \%} = \left[\frac{0.2525(\#of\ \ P1\ Visits)(MPT)}{\#of\ \ Machines} + \frac{0.3788(\#of\ \ P2\ Visits)(MPT)}{\#of\ \ Machines}\right] \times 100$$

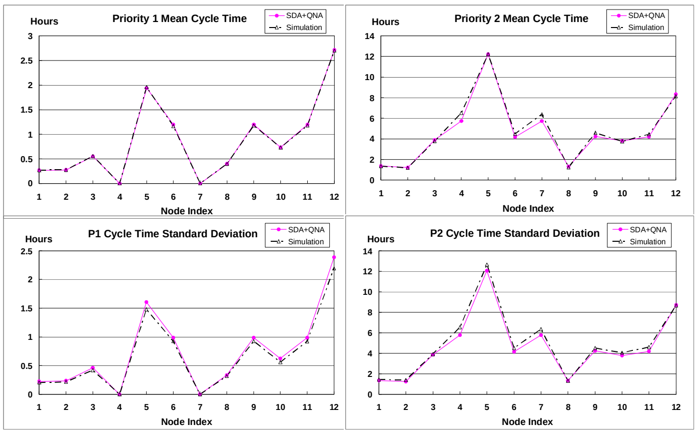


Figure 9: Mean & std. dev. of nodal cycle times (SFM)

Table 2: System level performance comparisons (SFM)

		Mean Cycle Time	Cycle Time Standard Deviation
1	SDA+QNA	18.613	5.190
	Simulation	18.572±0.014	3.5622±0.008
	Absolute Error	0.041	1.627
	Relative Error	0.22 %	45.68 %
2	SDA+QNA	172.78	54.674
	Simulation	174.695±1.159	53.088±0.899
	Absolute Error	-1.915	1.586
	Relative Error	-1.10 %	2.99 %

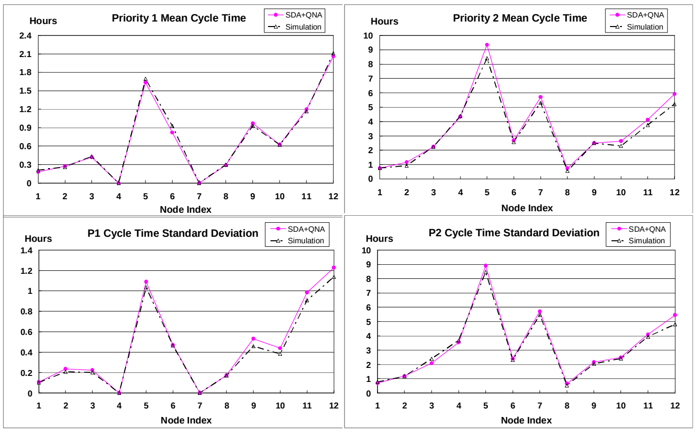


Figure 10: Mean & std. dev. of nodal cycle times (GFM)

Table 3: System level performance comparisons of general fab model

		Mean Cycle Time	Cycle Time Standard Deviation
1	SDA+QNA	15.137	3.503
	Simulation	15.451±0.009	2.289±0.005
	Absolute Error	-0.314	1.214
	Relative Error	-2.03 %	53.02 %
2	SDA+QNA	122.22	36.392
	Simulation	112.015±0.668	38.133±0.424
	Absolute Error	10.205	-1.740
	Relative Error	9.11%	-4.56%

Table 4: Comparisons of CPU times in fab models

	Average CPU time (seconds)	
Two Priorities Fab Models	Hybrid SDA+QNA	Simulation (one run)
Special case	3.606	2466
General case	3.622	2484

REFERENCES

[1] D. P. Connors, G. E. Feigin, and D. D. Yao, A queueing network model for semiconductor manufacturing, *IEEE Trans. Semi. Manuf.*, vol. 9, pp.412-427, 1996.

[2] G. Horvath and M. Telek, "Approximate Analysis of Priority Queues" *Technical Report, Technical University of Budapest*, 2000.

[3] W. Whitt, "The Queueing Network Analyzer", *The Bell System Technical Journal*, Vol. 62, pp2779-2815, 1983.

[4] M. D. Hu, S. C. Chang, "Translating Overall Production Goals into Distributed Flow Control Parameters for Semiconductor Manufacturing", *Journal of Manufacturing Systems*, Vol. 22, No. 1, 46-63, 2003.

[5] J. D. Faires, and R.L. Burden, *Numerical Methods*, *3rd ed.*, Brooks/Cole Pub. Co., Nov. 2002.

[6] S. H. Lu, D. Ramaswamy, and P. R. Kumar, "Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Times in Semiconductor Manufacturing Plants", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 7, No. 3, August 1994.