行政院國家科學委員會專題研究計畫 期中進度報告

Eco-Community：以感測網路建構一個虛境及實境互動式智慧型社區(1/3)

期中進度報告(完整版)

計 畫 類 別 ：整合型
計 畫 編 號 ：NSC 98-2218-E-009-023-
執 行 期 間 ：98 年 11 月 01 日至 99 年 10 月 31 日
執 行 單 位 ：國立交通大學資訊工程學系（所）

計 畫 主 持 人：易志偉
共 同 主 持 人：曾煜棋、張賢宗、孫敏德、江振瑞、葉士青
吳世琳、趙志民、林志明

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 99 年 08 月 31 日

國科會工程處無線感測器網路專案

期中成果報告

計畫基本資料

計 畫 名 稱	Eco-Community： 以感測網路建構一個虛境及實境互動式智慧型社區	
計 畫 編 號	NSC98-2218-E-009-023	
執 行 期 間	民國 98 年 11 月 01 日 ~ 民國 101 年 10 月 31 日	
計畫總主持人	易志偉 副教授	交通大學資訊工程系
共同主持人	曾煜棋 教授	交通大學資訊工程系
共同主持人	吳世琳 副教授	長庚大學資訊工程系
共同主持人	林志明 主治醫師	長庚醫院內科部
共同主持人	張賢宗 助理教授	長庚大學資訊工程系
共同主持人	趙志民 副教授	海洋大學資訊工程系
共同主持人	孫敏德 助理教授	中央大學資訊工程系
共同主持人	江振瑞 副教授	中央大學資訊工程系
共同主持人	葉士青 助理教授	中央大學資訊工程系

年度	經費(元)	參與人力				
		教授	博士後研究	博士生	碩士生	大學部專題生
98	2,012,000	9	0	5	13	5

一、計畫目的、計畫架構與主要內容

1. 計畫目的與預期成效

無線感測網路(Wireless Sensor Network, WSN)為一個智慧型的無線監控系統,運用佈署大量感測器收集例如溫度、空氣品質、動作等資料，並即時監測、反應的運作系統。無線感測網路具有感測、計算及通訊能力，更兼具低成本、低耗電、小體積、易佈建、可程式化、可動態組成等特性，因此無線感測網路非常適合於實際應用。

隨著都市發展計畫及都會社會人口集中，居住環境的生活品質及居家鄰近的生活機能已成為新都市人考量的重點之一，加上全球暖化及石油能源日益短缺，綠化節能等工作未來也和人類生活相連，變成密不可分的一環，故智慧化社區逐漸受到大家的重視，連帶使得各種感測器產品因應而生。無線感測網路技術將原本獨立分散於環境中的各種感測器，藉由無線傳輸技術連結，形成無線感測網路系統。透過該系統可實現很多過去無法達成的生活應用。例如我們的社區感測網路可以提供社區內車輛的追蹤與即時定位、人員安管措施、生活資訊交換平台，這些措施不但可以將人類生活品質獲得大舉的提升，也使我們居住的環境變的更加安全及便捷，也更加環保。

Eco-Community 的目的就是要創造出一個使人感到舒適、優化且兼顧環保、生態觀念的環境

空間。理想中的居住社區，應是多方位的，應考量到環保節能減碳、居住安全、醫療照護及生活便利性，亦即一個對環境、對居民都友善的智慧型生態社區。利用無線感測網路即可達到這些目的。

在本計畫中，我們著眼在人類最重要的生活空間 — 居家及社區，期望透過結合無線感測網路的功能，營造一個有生態概念的溫馨社群生活環境。為了有效利用資源並對各項議題做深入的研究，我們分成四個子計畫來進行：子計畫一為建構 Google Earth 上的虛境及實境互動式平台；子計畫二為互動式智慧型社區之創新應用；子計畫三為感測網路之通訊及感測資料處理技術；子計畫四為建構交叉實境互動平台。本計畫包含了系統平台的建立、創新應用的開發及基礎的研究，是一個深具前瞻性的整合計畫。結合了理論與實務應用，相信其成果及經驗對國內之感測器網路之研究及社區環境的發展改造，有著莫大的助益。

本計畫的預期綜合效益有增進國內研究學者執行整合型計畫的能力與經驗；開發環境監測與控制的先進系統；本系統研究、設計與實作的成果及經驗，可供產官學界未來規劃及開發智慧型生態社區之參考，對國內感測器網路的應用，帶來莫大的助益與經驗；紮實的無線感測網路基礎的研究，將提供多樣化創新應用的基石；互動式平台開發所展示的人機介面可以激發更多的可能，豐富我們的生活；結合各相關領域教授之研究專長，整合設計及開發無線感測器網路之應用；參與此計畫之研究人員得以熟悉無線感測網路、環境監測控制、節能效益評估等相關議題；無線網路應用程式之開發；研究計劃成果可發表學術論文，提昇國內無線通訊之技術與研究水準；參與此一計畫的教授及研究生可以彼此交換研究心得、增進研究成果、並加速學術交流。

2. 計畫架構

要完成智慧型生態社區 Eco-Community，首要之務便是無線感測網路的佈建與資料的蒐集(如溫度、濕度等)，搭配資料整合，將龐大的數據轉成有用的資訊。然而，盲目而隨意的設置感測器，並無法確切取得最需要的資訊，反而會造成系統的負荷。因此，我們必須考慮人們活動，搭配建築思維，在適當的地方設置感應器以擷取資料，並且安裝適當的設備。此外，有了這些資訊和設備，我們還必須結合有效的控制方法來驅動這些裝置，加以調節環境。因此，整合本計畫所提出的四個子計畫，能夠在適當的地方感測各種資訊，並自動啟動設備做出相對應的調節，讓生活、辦公的環境更加舒適宜人，進而成為真正的智慧型生態社區：Eco-Community。

本計畫中的四項子計畫環環相扣、相輔相成(請參見圖 1)。智慧型社區模型提供無線感測網路適當的建物與人車分佈拓樸，讓感測器所取得的數據資料，能夠確切掌握人們活動範圍的環境變化；而無線感測網路將蒐集到的數據交由後端的控制系統分析，並且根據當時環境狀態以及智慧型建築所提供人們對環境舒適度的相關資訊，透過實境與虛境互動平台發送指令，驅動設備，讓各項裝置可以達到預期的環境調節功能。然而，這樣的環境調節也許並未能符合人們的預期與需求，因此，控制端會將人們手動調節裝置的資訊整合交給智慧型社區模型，再加上無線感測網路所蒐集的當時之環境資訊，加以分析、學習，讓智慧型社區模型能夠達到更高的生活舒適滿意度，以及給予控制系統更準確、更符合預期且兼顧生態概念的資訊。

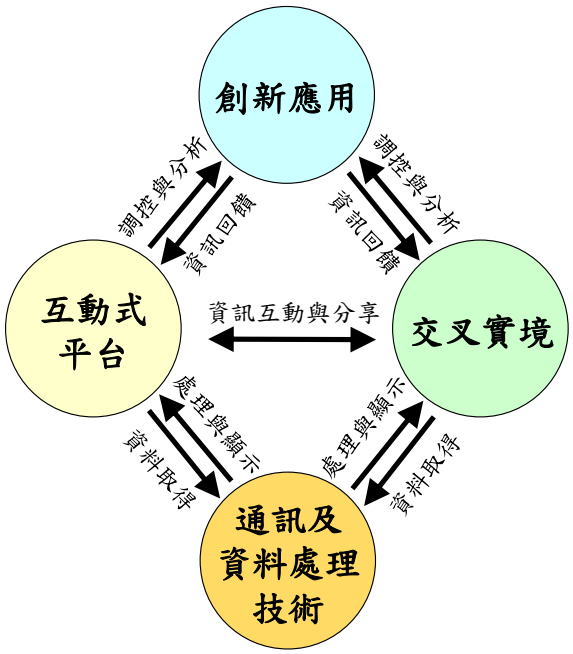


圖 1 四個子計畫之相互關係

3. 計畫主要內容

子計畫一：建構 Google Earth 上的虛境及實境互動式平台

在此計畫中我們選定在 Google Earth 上建立我們的介面，在原有的地圖上建構 3D 物件，如房屋、路燈、花圃、游泳池、籃球場等，營造一個 3D 的虛擬社區，以利我們能以最簡單且直接的方式，對這社區中的任何事物要求我們想知道的相關資訊，甚至可以遠端操縱設備。

為了達到上述的目的，本子計畫擬完成下述四個項目。

(1) 在 Google Earth 上建構一個虛擬 3D 社區

我們將利用 Sketchup 軟體協助設計，將社區內的建築物及設施等，於 Google Earth 上虛擬重現。本研究中使用 Google Earth 來進行虛擬環境之建構，並利用 Google Earth 平台所提供的 KML 或 KMZ 等格式，將資料結合在一起，進行資訊交流以及展現。透過資訊管理服務提供資訊管理、資訊儲存、資訊使用以及通知功能。

(2) 在虛擬社區上建立便於虛境與實境互動之社群(交)平台

我們欲建構一套能夠以無線感測網路為基礎之模組化社群應用服務平台，提供社區上建立虛境與實境互動，加強對銀髮族群的健康關心與追蹤，增強對社區內竊賊的防範與社區人員的安全，增加社區內活動的推展。

(3) 社區環境監控平台

在本計畫中，最終我們是希望能建立一個社區環境監控平台，使社區中每個人都能掌握所在社區的資訊。

子計畫二：互動式智慧型社區之創新應用

隨著生活水準與醫療衛生長足的進步，人類平均壽命因此大幅的延長，高齡化社會已是全球性的現象。中高齡伴隨的生理退化，活動力的降低減少，使得中高齡者容易被邊緣化。在此觀點下，如何增加高齡者的活動，成為重要的一項事務。而三代同堂又是亞洲社會常見的社區住戶模

式，在社區裡，白天青壯年的居民大多外出工作，社區中僅會留下年長與年幼的居民，然而年長與年幼的人通常比較需要被照顧，如果白天居民可以利用本身的力量互相照顧，將可以大幅減少社會成本的浪費與額外的支出。

為了達到上述的目的，本子計畫擬完成下述四個項目。

(1) 建構以社群為中心的無線感測網路定位服務系統

定位在未來生活中扮演相當重要的角色，只有透過定位才可以進行位置相關的服務。社區住戶身上攜帶配備有符合規格的無線感測裝置，在社區空曠處可以接收到 GPS 定位訊號，可以利用現行 GPS 定位方式進行定位，然而在 GPS 遮蔽處之室內及室外環境中，我們將在無線感測網路平台設計與實作一套定位系統。有了社區住戶、物件的位置資訊將可以成為互動式智慧型社區之創新應用的重要根基。

(2) 建構以無線感測網路定位服務系統為基礎的社區管理服務系統

透過社區住戶定位系統所即時更新的住戶位置資訊，社區管理服務系統可以清楚了解每個社區住戶的即時所在位置，當社區住戶有需要服務或是緊急狀況時，社區住戶的位置就會是一個非常有用的資訊，透過此資訊，本系統可以提供許多不同的創新社區管理服務。

(3) 建構以無線感測網路定位服務系統為基礎之導引系統

導引系統與傳統導航系統不同的地方在於，傳統導航系統是透過靜態資料庫存放目的地資訊，然後利用定位系統協助，計算出一條路徑供使用者參考，但是因為導航系統只有靜態資訊，無法得知動態的即時資訊，因此所提供的路徑都是固定的，無法隨著環境的改變，給予不同的服務。而本系設計之導引系統將連接社區之無線感測網路，透過動態即時資訊與靜態資料庫的結合，能夠依據目前社區的狀況導引使用者到正確的位置。

(4) 建構以無線感測網路服務系統為基礎之健康照護系統

本計畫將設計出一個妥善職能照顧的老人養生村：本計畫可搭配智慧型的無線監控系統，依據使用者的心脈律動及體溫變化等生理變化，透過社區網路連結，適時通知社區周遭醫療體系。若監控系統查知使用者有劇烈生理變化，急需醫療救治，感測器利用無線的方式將病患生理資料傳回到中央控制伺服器，中央控制伺服器會根據這些資料，以及病患之前就診記錄，來做出一個適當的決策通知醫療團隊，達到醫療零時差的病患照顧。

子計畫三：感測網路之通訊及感測資料處理技術

為了建立一個智慧型生態社區，我們需要使用相當數量的感測器和其他裝置組成感測網路，藉由這些感測器節點收集環境資訊並加以利用，發展出各種便利的應用。所以，感測網路是整個計畫的基礎，必須建立一套穩定、有效率且牢固的感測網路，才能更有效率、更準確地提供各種服務。一套好的感測網路必須考慮許多因素，如感測器佈點位置、網路初始化步驟以及節點間通訊所使用的網路協定，每個環節都必須相互呼應才能成就一套好的系統。另外，在我們的網路中可能會有手機、PDA 等裝置，因此異質網路間的通訊也是必須考慮的重點。

為了達到上述的目的，本子計畫擬完成下述三個項目。

(1) 網路佈建與初始化

在感測網路初始化的階段希望做到叢聚化(Clustering)和階層化(Hierarchical)的網路架構。叢聚化能有效的壓縮與整合訊息，可節省訊息傳送的資料量，避免過多的訊息傳送造成的碰撞，造成訊息遺失，也可以節省電力的耗損，讓感測器的壽命延長。階層化的目的是為了能夠善用各種不同的感測節點特性，讓一些耗損能量較多的節點能夠減緩其耗損能量的速度，達到系統的穩定。

(2) 網路通訊技術

感測器網路是由許多以電池供電的感測器節點所組成。在網路佈建之後，以人工方式換電池耗時費力，有時甚至不可行，大部分的感測網路在佈置之後，都不會更換電池。因此，延長感測器網路的壽命，就是一個重要的議題。在本計畫中，我們將設計一套省電 MAC 協定來達到延長網路壽命的目標。而在大型的無線感測網路中，封包的傳送可能需要多點轉送(Multihop)才可送達目的地，因此如何尋找路由路徑，以及轉送點如何把封包有效地送往目的地，是無線感測網路中基本且關鍵的任務。

(3) 資料處理技術

資料處理的技術包括 L-Value、G-Value 及 M-Value 的資料淬取。利用 L-Value 資料可以反推光源位置。在 G-Value 及 M-Value 資料的淬取方面，利用篩選過的 G-Value 和 M-Value 來判斷使用者的動作，如靜止、步行、跑步、跌倒等，再把所得資訊傳給中央控制伺服器，中央控制伺服器判斷是否有突發狀況，假使有突發狀況，系統便會利用感測網路，找出相關人士並發出警告訊息，讓有突發狀況的使用者可以在最快時間內得到幫助。

子計畫四：建構交叉實境互動平台

交叉實境(Cross Reality)的目的是把網路虛擬世界中的 3D 場景和現實世界中的感測器網路連接在一起，透過現實世界的 sensor 與虛擬世界進行互動，讓感應到的資訊能夠在虛擬世界相對應的產生動作，創造更多的互動可能。

為了達到上述的目的，本子計畫擬完成以下目標。

(1) 選定中央大學工程五館(實境)，在Second Life的平台上，縮尺寸模擬並建構一個3D虛擬的社區(虛境)

根據本子計畫的其中一個目的，在於能夠提供 Second Life 之廣大使用者，參訪虛境中的 3D 虛擬社區，藉以瞭解並感受真實社區的各式情境，除了即時互動模式的建立，視覺上的真實度(Realism)將影響使用者的感受性(Perception)，進一步決定了使用者與虛境的融入性(Immersion)及使用動機(Motivation)。所以，我們將以所選定之真實社區的實尺寸，建構其縮尺寸的 3D 幾何模型，並根據真實社區的材質及室內或室外裝潢擺設，進行模型紋理的模擬，以達到視覺上的真實度。

(2) 建立實境與虛境之間的訊息交換平台

實境使用者與虛境使用者之間的互動過程及模式，以及經由實境擷取之環境資訊並視覺化於虛境的程度，將影響使用者的感受性、使用者與虛境的融入性以及使用動機。而各式資訊(溫度、濕度、聲音、影像、文字等)傳遞(雙向)的種類及速度，則又主宰了互動模式以及資訊視覺化的多元性和複雜程度。所以，我們將建立一個訊息交換平台，將所有佈建之無線感測器的訊息，進行大量地、多種類地、雙向地以及兼具速度性地訊息交換。

(3) 於虛境中「視覺化」實境之環境參數，為Second Life廣大使用者提供實境之環境資訊

在真實環境中，環境資訊(溫度、濕度等)，通常是以數位化的方式以進行表達，然而我們將在虛擬環境中，利用無線感測器所截取的環境資訊，將環境資訊進行視覺化，運用動畫技術以不同的型態進行表達，這將使得使用者能夠即時感知真實環境的訊息。

(4) 建立實境使用者與虛境使用者之間的社交社群互動模式

使用者的感受性、使用者與虛境的融入性以及使用動機，除了 3D 虛擬社區的視覺真實度的影響，也取決於實境使用者與虛境使用者之間的互動過程及模式，因此，我們將針對實境使用者與虛境使用者，運用無線感測器所傳輸的資料，設計數種不同形式的互動模式：視

覺上的、聽覺上的及文字上的。

(5) 評估實境使用者與虛境使用者之間的社交社群互動模式

我們也希望分別針對實境使用者與虛境使用者，對於不同的互動模式，從使用者角度所可能產生不同程度的感受性、融入性、動機進行評估，以瞭解實境與虛境相對應於不同的互動模式中的使用者的反應，並作為此系統後續設計與改進的重要依據。

二、重要執行成果及價值

1. 實境平台

- Google Maps 平台

我們在 Google Maps 上建立我們的介面，營造一個虛擬社區，以利我們能以最簡單且直接的方式，對這社區中的任何事物要求我們想知道的相關資訊，甚至可以遠端操縱設備。在這個平台上，我們建立了資料庫系統平台，其他應用可以此平台為基礎作應用。

- 社區環境監控平台

我們建立了一個社區環境監控平台，可作環境的監控，如光度、溫度與濕度。圖 2 與 3 為社區環境監控的平台，分別為室外與室內的環境。感測節點偵測到的資料藉由 Multihop 傳回，利用我們設計出適用於目前網路環境的 MAC 及感測網路路由協定，來提高傳遞效率。

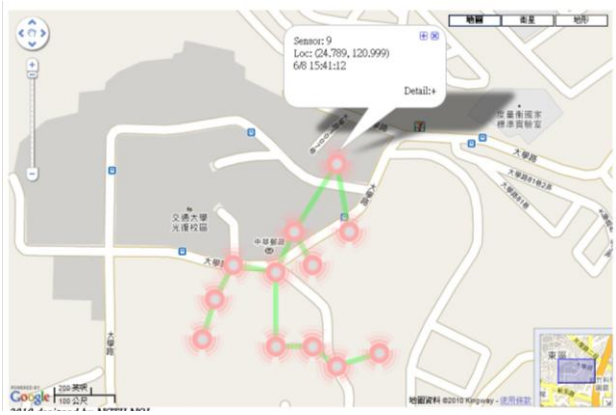


圖 2 社區環境監控平台(室外)

圖 3 社區環境監控平台(室內)

- 定位技術

我們藉由手持裝置上內嵌的微電子機械元件，提升原有定位系統的精準度，並整合不同的定位系統，以提供一個無所不在的定位服務系統。我們利用 GPS、IMU 與 RF，再經由 Kalman Filter 整合，以得到精準的位置資訊。GPS 主要是用在室外定位。而利用 IMU 內的三軸加速度計與三軸磁力計的資料，可以得到相對的位移以及絕對的方向。RF 定位則是即時反應環境中訊號變化的情形，有效率地判斷出使用者靠近於哪些參考點，判斷使用者可能的所在位置，主要是用在室內定位。最後是使用 Kalman Filter 來整合 GPS、RF 定位以及慣性感測元件的定位資訊，提供一個高精確的絕對位置。室內外定位的精確度請見圖 4 與 5，其中藍線為真實路徑、綠線為 GPS 軌跡、紅線則為利用慣性感測元件所得到的軌跡。

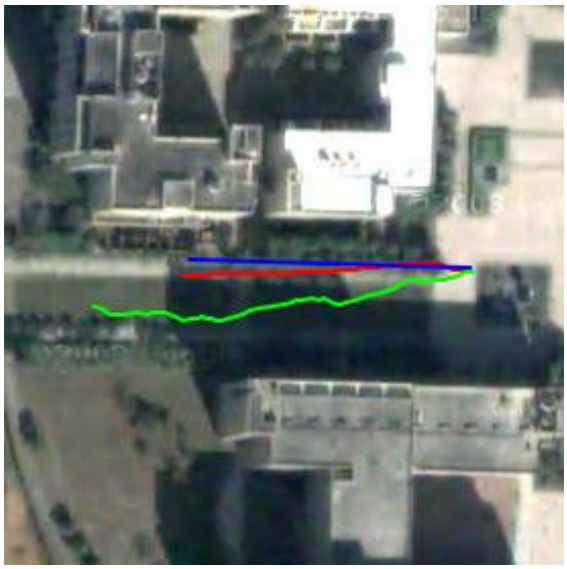


圖 4 社區環境監控平台(室外)

圖 5 社區環境監控平台(室內)

2. 虛境平台

為了讓使用端可以使用方便，虛境的部分我們採用了 JAVA 中的一個專案 — Wonderland Project，3D 圖型主要以 JME (JAVA 以 OpenGL 為核心繪製 3D 物件的引擎)來繪製我們所要的物件。在此虛擬環境之中，我們可以新增虛擬的物件，每個物件我們可以把它當做是一個 Cell，而伺服器端的 Cell 和使用者端的 Cell 可以透過網路連線相互傳遞資訊。使用者端只需要有安裝 JAVA Run Time Package 就可以透過網頁登入到 Wonderland 伺服器，不需要額外下載安裝其它程式。圖 6 為 Wonderland 的介面。

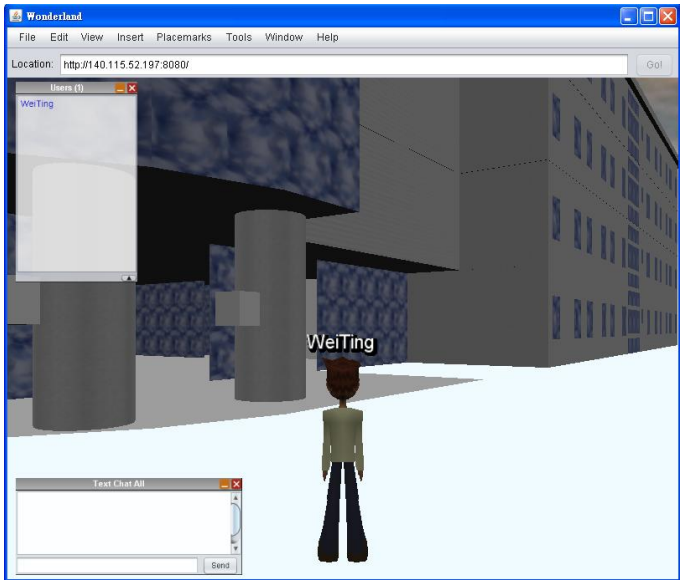


圖 6 Wonderland 的介面

為了將感測器的資料與網路虛擬環境做結合，我們以工程五館為網路虛擬環境的場景，並放置感測器在進門的大門口，當溫度太高時會產生一個需要散熱的訊號傳入 Wonderland 伺服器中顯示，使用者可以利用在遠端 Novint Falcon 的力回饋設備去操作網路虛擬環境中的鼓風器，使它產生氣流，讓網路虛擬環境中的電風扇轉動，讓室內溫度降溫，溫度的高低會決定力回饋的大小，

溫度越高的時候，所產生的回饋就會比較小，當溫度越低，所產生的回饋就越大，也就越難鼓動鼓風器。當鼓風器開始有氣流時，Wonderland 會傳送一個訊號給電風扇風速的控制器，氣流越大，電風扇的轉速也越快，反之則越慢。圖 7 顯示我們將感測器佈建在工程五館的示意圖。

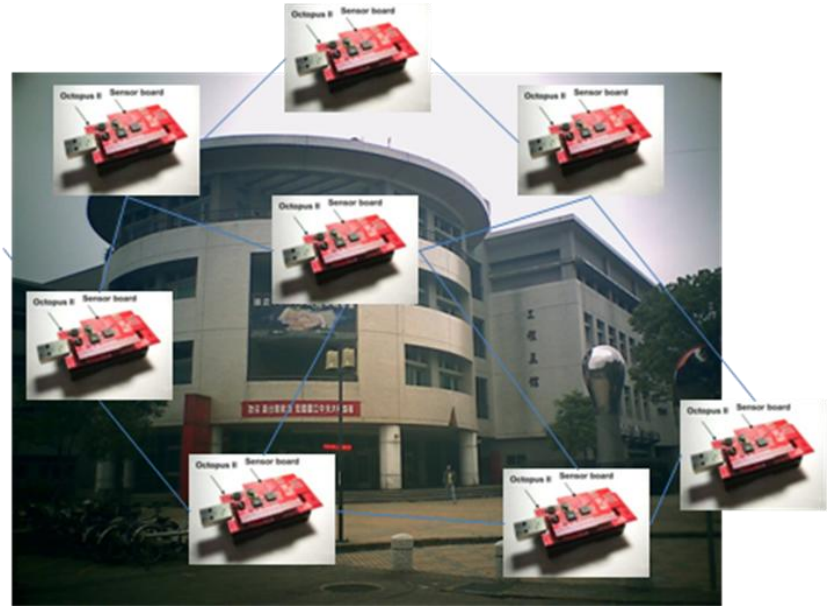


圖 7 感測器佈建示意圖

3. 創新應用

- 個人導航系統

我們開發個人導航系統並在其上實現了以位置為基礎的服務。系統架構包含 Server 端和 Client 端。Server 端主要是以 Google Maps 為基礎，在上面建立圖資，顯示相關區域景點和建築等，並顯示使用者的移動情況和所在行經路線的景點資訊。資料庫中儲存景點的位置、相關介紹資訊，建築的位置、圖資，供使用者使用定位系統查詢相關資料。Client 端配備有 IMU、以高精準的三軸加速度計和三軸磁力計測量提供相對位置，Client 端還配有 GPS，提供絕對位置。室外及室內的以位置為基礎的服務請見圖 8 與 9。

圖 8 以位置為基礎的服務(室外)

圖 9 以位置為基礎的服務(室內)

- 停車位監控系統

我們在 Google Maps 上開發了一個停車位監控系統，使用者可以利用此系統得知停車場可用停車位的數量資訊，方便使用者快速地找到停車位，避免使用者自行搜尋停車位所造成的油耗。首先是 sensor 必須將偵測到的光度資訊回傳至 sink，然而傳統的 sensor 傳送方式為定期傳送，此方法會造成能量上的消耗，故我們改為 sensor 在改變狀態後才回傳資料給 sink，藉此達到省電的效果。系統接收到 sensor 資料後寫入資料庫，將停車位資訊顯示在 Google Maps 上，若 sensor 資料有更新時，Google Maps 上的停車位資訊也會隨著更新。圖 10 與 11 為停車位監控系統的介面，注意到不同的地圖尺寸有著不同的資訊表示方式。

圖 10 停車位監控系統介面(微觀)

圖 11 停車位監控系統介面(巨觀)

- 智慧型手機安全平台

我們整合智慧型手機到社區感測網路中，利用智慧型手機與社區感測網路可以提供婦女與小孩的人身安全機制。在智慧型手機平台上整合了三軸加速器、多媒體(影音)、無線區域網路和定位系統，當使用者遭遇到危險或緊急情況時，使用者可以迅速地搖晃手機來達到通訊的目地，此系統，我們稱之為 i-Mace。

在 i-Mace 中，我們將其切割成四個部分：Sensing Module、Image Acquisition Module、Localization Module 和 Communication Module。Sensing Module 可判別所產生的晃動是單純使用者的一般移動或是緊急事件發生時使用者有意的晃動。當事件發生時，當事人必須在短時間內做出反應，這是我們選擇用手機晃動來啟動 i-Mace 的原因。在此考慮下，我們希望手機晃動後能立即拍攝並直接將照片傳送出去。然而手機在晃動時，所拍攝的影像很可能是模糊不清的。要解決此問題，我們可以讓手機的 Image Acquisition Module 連續拍攝多張相片並自動判斷何張照片有比較清楚的人像來傳送出去。再者，當緊急狀況發生時，當事人的位置資訊對於及時救援是非常重要的，而 Localization Module 便是將智慧型手機上原有的定位系統(GPS 和無線網路定位系統)整合至 i-Mace 上，適時地擷取當事人的位置資訊以方便 Communication Module 可以將資訊送出。圖 12 與 13 為 i-Mace 的執行畫面。

圖 12 i-Mace 執行畫面(警鈴設定)

圖 13 i-Mace 執行畫面(簡訊相關設定)

- 虛擬運動場

目前已經能夠初步掌握 Wonderland 模組的繪製以及物件的位置控制、人物模組的控制，並且能夠將感測器接收到的資料反應在 Wonderland 的虛擬世界之中。利用感測器控制在 Wonderland 之中的物件移動，可以根據感測器移動向量來決定物體該往何處移動，並且知道該移動多少距離，當然也可以做到在社區住戶勞累一天之後，不想外出運動，就可以利用家中的設備和感測器做一些休閒活動，利如跑步、舉啞鈴、仰臥起坐等，配合 Wonderland 裡面的人物動作及場景音樂效果，達到放鬆身心的休閒娛樂，並且有運動的效果。圖 14 顯示虛擬運動場的介面。

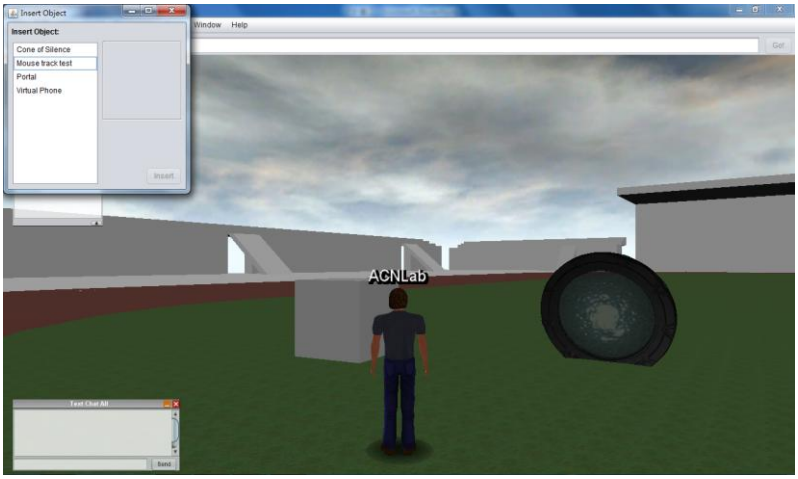


圖 14 虛擬運動場

- My 太極 Book

近年來以網頁為基礎的社交網路，如 YouTube、Blogger、Windows Live、Facebook 與 Twitter 等，引起了廣泛的注意。透過這些社交網路，使用者可以很容易地與朋友分享相片、影片、文章與訊息。然而這些社交網路是虛擬的，若我們能結合實體的 MEMS 設備與無線感測網路，便能在虛擬的世界裡做實體的互動。我們利用 Body Sensor Network 為基礎的(BSN-based)社群式網路活動，來發展社群式網路應用服務元件，讓數量龐大且分佈在各區域的社區使用者，可以打破時間空間上的限制、透過智慧型終端裝置，互相分享、從事共同的運動、分享醫療方面的生理狀態等。我們展示了太極運動社群之成果，圖 15 為其在 Facebook 上的介面。

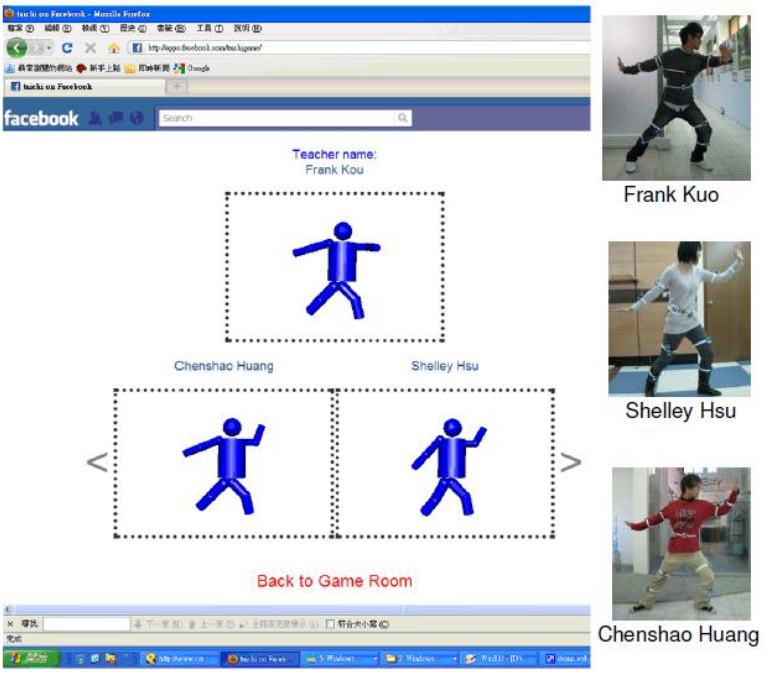


圖 15 太極運動在 Facebook 上的介面

三、整體計畫進度與未來執行重點

1. 整體計畫進度

子計畫一：建構 Google Earth 上的虛境及實境互動式平台

(1) 系統平台的建立

a、 已建立以 Google Maps 為基礎的系統介面。

b、 可在此平台上建立虛擬社區並實現社群功能。

c、 可在此平台上作環境的監測。

d、 相關的應用程式可以此做為平台，如以位置為基礎的服務、感測資料查詢及室內定位服務等。

(2) 資料的初始化

a、 建立資料庫、決定資料庫存放的內容與格式，並初始化資料。

b、 將感測資料存至資料庫。

c、 將資料庫內的人員、車輛與環境資訊顯示於 Google Maps 平台上。

(3) 定位技術

a、 以建築拓樸為基礎的定址機制。

b、 以 IMU 與 GPS 結合來做室外定位；以 IMU 結合 RF 來做室內定位。

子計畫二：互動式智慧型社區之創新應用

(1) 室內定位之無線感測網路佈建

a、 以長庚大學管院四樓為測試模型，進行室內定位之無線感測網路佈建。

b、 利用訊號強度差的方式來判斷 MU (Mobile User)與環境中各 node 的遠近。

c、 定位演算法：利用訊號的強度計算出權重值，而各 node 的座標為已知，故可求得 MU

的座標位置。

(2) 停車位監控系統

a、 實作出停車位監控系統，使用者可以利用此系統得知停車場可用停車位的數量資訊，方便使用者找到停車位。

b、 可協助快速找到可用的停車位，避免使用者自行搜尋停車位所造成的油耗。

子計畫三：感測網路之通訊及感測資料處理技術

(1) 感測網路 MAC 協定設計

a、 已設計出合適的 MAC，可適用於目前所需之應用。

b、 設計具有省電效果且容易實作在真實感測器上的 MAC。

(2) 感測網路路由協定設計

a、 已設計感測器路由協定。

b、 透過 Next-hop Group 觀念確定網路路由方向。

c、 可再利用感測器位置及訊號強度等資訊，找出有效率之路徑。

(3) 智慧型手機與社區的結合

a、 使用者可以因應環境的變化，藉由搖晃手機來傳遞訊息。

b、 手機可透過通訊網路將訊息傳送至社區伺服器。

子計畫四：建構交叉實境互動平台

(1) 交叉實境平台的建立

a、 完成中央大學資工系工程五館室內 Octopus 無線感測網路建構。

b、 熟悉 Wonderland 的使用，並結合數位美術設計人員，以軟體建構縮尺寸之模擬社區，並以動畫技術(雲霧、火焰等)具體視覺化環境參數於虛擬社區中。

c、 以虛擬大廳來展示交叉實境虛擬互動社群平台架構。

(2) 感測器的應用

a、 以多個感測器實作出 Body Sensor Network，以太極運動來展示虛擬互動社群平台架構。

2. 未來執行重點

子計畫一：建構 Google Earth 上的虛境及實境互動式平台

針對目前的成果，我們需要改進的部分主要為增加定位的準確性，這或許可以利用更佳的坐標演算法增加準確性，同時感測器的精準度也會隨著硬體技術的精進而增加，這也會大大地增加我們的定位準確性。另外，也可以考慮藉由多種感測器的合作來增加準確性，因為多個感測器的使用，增加可參考的資料量，所以定位系統可以參考的資料樣本增加，準確性也會有所提升。有了精確的定位資訊之後，再開發出更多與位置相關的應用。

未來的目標主要是著重在網路和手持裝置的管理和維護，如考慮到每個使用者的活動範圍，通常會固定在某些區塊行動，故我們可以將感測器的設置位置最佳化；再者，當區塊內的使用者人數有變化，或者是使用者的身分為單車、機車或汽車，我們需要更改資料收集的頻率，讓感測網路可以更有效地收集到資料；此外，如手持裝置的管理：當使用者在社區行動，感測器對手持裝置作感測，並將所得資訊回傳給後端伺服器，由伺服器判斷此使用者的行動是否合理，也可以有效掌握到手持裝置的位置和狀況；利用有效的資料整合與壓縮，將可以有效減少感網路流量以

及延長感測網路壽命；感測網路的架構、資料庫的設計也值得進一步地研究；最後，資料的整理與顯示與無線感測網路之管理與維護等都是可以改善的方向。

子計畫二：互動式智慧型社區之創新應用

目前只利用光感測來判斷停車格的狀態，還有不足之處，為了提高停車格判斷的準確率，未來我們考慮到用具有超音波感測或是紅外線感測的 sensor 來做進一步的感測，超音波方面利用其音波反射的原理來判斷車位是否有被使用，或是以紅外線直線前進的特性進行判斷，加入上述兩種方式與原來光感的方法相輔相成進行實作，以提高判斷的準確率。

未來還要導入停車格的導引系統，讓使用者可以知道從目前的位置如何到達停車場停車，導引系統與傳統導航系統不同的地方在於，傳統導航系統是透過靜態資料庫存放目的地資訊，然後利用定位系統協助，計算出一條路徑供使用者參考，但是因為導航系統只有靜態資訊，無法得知動態的即時資訊，因此所提供的路徑都是固定的，無法隨著環境的改變，給予不同的服務。而本系設計之導引系統將連接社區之無線感測網路，社區之無線感測網路可以動態知道社區附近的情況，因此除了靜態資料庫以外，更可佐以無線感測網路所得到之動態即時資訊。透過動態即時資訊與靜態資料庫的結合，能夠依據目前社區的狀況導引使用者到正確的位置。

除此之外，我們希望將此定位系統與醫療照護相互結合，在老人養生村建置健康照護系統，透過無線感測裝置隨時記錄使用者的生理變化，此系統與社區用戶密切的互動，可以發展出社區用戶安全警示及緊急事故通報的系統，最後將社區導向更具有安全性與便利性的社區，達成醫療零時差的照顧。

子計畫三：感測網路之通訊及感測資料處理技術

目前手機安全平台已具有所期待的功能，但仍然有可改進的空間，其一是關於手機搖晃的靈敏度偵測，目前對於使用者搖晃手機的靈敏度偵測，還未能穩定的做出判斷，可以考慮加上回饋機制讓偵測更趨於穩定。還有，由於事發現場不一定是在室外，而目前手機上的定位系統，例如 GPS，無法做準確的室內定位，故我們希望可以結合其他子計畫所發展的定位系統來取得室內定位的資訊。再者，我們也考慮將位置資訊和 Google Maps 結合，讓使用者可以在 Google Maps 上清楚地看見自己或求救者的位置。最後，經過測試發現使用者在拍照後，將照片上傳至伺服器時會有壅塞的情況，未來可以考慮將 Communication Module 放到背景執行。

將來，希望手機可以和社區網路內的其他感測器互相溝通以整合其功能，提供更完善的服務。例如，當婦女持有手機，手機可定時發出所在位置及目的地資訊，藉由社區的伺服器，提高婦女周遭路燈的亮度，讓婦女處在較安全的情境下；此外，我們希望伺服器可調亮通往目的地之路線的路燈，藉此導引婦女走往正確的路線。婦女通過之後，照明設備可以恢復原來的亮度以達節能省碳的目標。

在網路的通訊部分，未來將驗證模型的正確性與可行性、分析並研究 MAC 與 Routing 協定中的相關參數、整合實驗測試空間之無線感測網路、致動器與控制器，並驗證 MAC 與 Routing 協定。從文獻探討中可以發現，無模型的控制器設計方法，僅在小型空間進行過實驗，控制目標也僅是減少單一感測器所量測的誤差值(量測溫度與設定溫度之差值)。而且測試空間單純，無人員來往進出或其他熱源，使問題相對簡化，是以能夠透過實驗與試誤法決定模糊推論法則。然而當所欲控制的空間變得複雜，致動器與感測器數目也增加時，試誤法便難以有效地設計出合適的控制器。因此對於應用在真正建築物的控制系統，仍需要參考其數學模型做為設計的依據。

子計畫四：建構交叉實境互動平台

未來將佈建室外無線感測器網路、建立行動式交叉實境互動平台、建立精確定位機制，讓實體使用者能夠以虛擬分身在虛擬社區中與其他虛擬玩家互動。

此外，研究並設計人因量表，量測以使用者為中心對於視覺上的互動、聽覺上的互動、文字上的互動以及虛擬介面使用上的感受性，這包括了視覺或聽覺舒適度、系統的功能性和使用性的評估；更進一步，研究並設計存在量表，量測並比較使用者置身於實境以及虛境的融入性，並藉此評估使用者繼續使用此一平台之動機的強烈程度。還有，進行真人實驗，以評估實境使用者與虛境使用者對於不同互動模式(視覺、聽覺、文字)的反應，並比較使用者對於資訊視覺化的反應，以做為系統改善的方向。

四、計畫之實際執行情形與預期工作之差異分析

1. Google Maps 與 Google Earth

Google Maps 與 Google Earth 皆為 Google 公司所提供的服務，兩者的商業模式有所不同，然而大致上說來都提供免費的地圖服務與開放的 API，掀起了公眾參與地理資訊系統的熱潮，因此，提供了個人加值或是企業加值的空間。Google Maps 與 Google Earth 共用 KML 地標資訊，在資料的轉換上難度不高。使用 Google Earth 瀏覽地圖，較為順暢，但其所提供的地理資訊較少；而使用 Google Maps 來瀏覽地圖，則顯示速度略為緩慢，但是資訊較為豐富。也因此，在初步的開發上，我們選擇圖資較為豐富且使用率較高的 Google Maps 來當作開發與顯示的平台。

2. Wonderland 與 Second Life

本計畫原先預定用 Second Life 來當作虛境的平台，然而在 Second Life 中租借一塊虛境需要每年固定付費給其公司，耗費相當大的維護資金，且若是要使用 Second Life 作為虛境顯示介面，使用者端必須要額外安裝其他程式才能使用，所以我們虛境的部分不採用 Second Life 為顯示虛境的介面。因此，我們採用了使用者端只要能夠執行 JAVA 的程式就可以使用的 Wonderland Project，雖然沒有營利公司為其撐腰，但表現一點也不遜色。而其又為 Open Source，故為一經濟的解決方案。Wonderland 以 Dark Star 為基礎，再加上 Voice Bridge 和一些其他的應用程式伺服器所構成，3D 圖型主要以 JME (JAVA 以 OpenGL 為核心繪制 3D 物件的引擎)來繪製我們所要的物件，感測器的資料或是其他設備上的感測裝置(如 G-Sensor)可以透過網路傳輸 UDP 的協定將資料傳送到 Wonderland 之中，並且可以看到使用者感測資料顯示在虛擬環境之中。

五、早期參與廠商績效

目前正與工研院討論技轉的相關事宜。

六、目前所遭遇之困難與因應對策

1. Octopus II 的相關支援

在使用 Octopus II 模組的平台時，遭遇了進入障礙，原因是必須在 TinyOS 的環境下開發 nesC 程式。而針對不同種類的感測器，必還須自行撰寫 lib。此外，在感測器的搭配與選擇上，並不是很便利。希望 Octopus 的開發團隊能提供協助，以減少大家摸索的時間，提高開發系統的效率。前陣子，清華大學與淡江大學有感於使用者的需求，特別架設網站來幫助使用者，其網址為 http://163.13.128.59/，該網站提供相當豐富的資料。感謝清華大學與淡江大學團隊的用心。

七、從計畫辦公室是否已能連接至貴研究成果網站已供參訪人士觀賞

是的，網址為 https://nol.cs.nctu.edu.tw/~zlchen/Project/Eco-Community.html。

八、是否連接至 Google Maps 或 Google Earth

是的，此計畫以 Google Maps (Google Earth)為平台。

Leveraging Spatio-Temporal Redundancy for RFID Data Cleansing

ABSTRACT

Radio Frequency Identification (RFID) technologies are used in many applications for data collection. However, raw RFID readings are usually of low quality and may contain many anomalies. An ideal solution for RFID data cleansing should address the following issues. First, in many applications, duplicate readings (by multiple readers simultaneously or by a single reader over a period of time) of the same object are very common. The solution should take advantage of the resulting data redundancy for data cleaning. Second, prior knowledge about the readers and the environment (e.g., prior data distribution, false negative rates of readers) may help improve data quality and remove data anomalies, and a desired solution must be able to quantify the degree of uncertainty based on such knowledge. Third, the solution should take advantage of given constraints in target applications (e.g., the number of objects in a same location cannot exceed a given value) to elevate the accuracy of data cleansing. There are a number of existing RFID data cleansing techniques. However, none of them support all the aforementioned features. In this paper we propose a Bayesian inference based approach for cleaning RFID raw data. Our approach takes full advantage of data redundancy. To capture the likelihood, we design an n-state detection model and formally prove that the 3-state model can maximize the system performance. Moreover, in order to sample from the posterior, we devise a Metropolis-Hastings sampler with Constraints (MH-C), which incorporates constraint management to clean RFID raw data with high efficiency and accuracy. We validate our solution with a common RFID application and demonstrate the advantages of our approach through extensive simulations.

1. INTRODUCTION

Radio Frequency Identification (RFID) is an electronic tagging technology that allows objects to be automatically identified at a distance without a direct line-of-sight, using an electromagnetic challenge/response exchange [27]. An increasing number of major retailers such as Wal-Mart, The Home Depot, Kroger, and Costco have installed RFID based inventory management systems in their warehouses and distribution centers. However, practitioners are facing a challenging problem: the raw data collected by RFID readers are inherently unreliable [23, 17]. Therefore, middleware systems [16] are required to correct readings and provide cleansed data. Most previous solutions [11, 17, 21, 15, 8] for cleansing RFID raw data focused on smoothing the readings generated by a group of readers. However, these existing solutions suffer from three major limitations:

- Data redundancy introduced by overlapping detection regions of multiple stationary readers (spatial redundancy) or continuous readings over time of a single mobile reader (temporal redundancy) is not utilized to improve reading accuracy.
- Prior knowledge about tagged objects and RFID readers is not effectively utilized to improve reading accuracy.
- Constraints in target applications (e.g., the maximal capacity of a room or a shelf) are not effectively utilized to cleanse the data.

In this paper, we propose a method to address these limitations. We focus on taking full advantage of data redundancy, prior knowledge, and application constraints to elevate the accuracy of data cleansing.

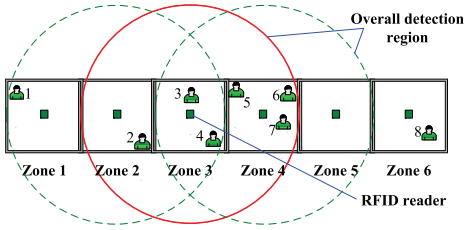


Figure 1: Spatial overlapping of detection regions.

	Zone 1 Reader	Zone 2 Reader	Zone 3 Reader	Zone 4 Reader	Zone 5 Reader	Zone 6 Reader
Obj 1	1	0	0	0	0	0
Obj 2	0	1	1	0	0	0
Obj 3	0	0	0	1	0	0
Obj 4	0	0	1	1	0	0
...	...	...	...	...	...	...

Table 1: RFID readings.

1.1 Data Redundancy

Two types of redundancy may arise in RFID related applications: spatial redundancy, where an object is detected by multiple readers in its neighborhood, and temporal redundancy, where an object is detected multiple times by a single reader over time.

Spatial Redundancy: In order to reduce the complexity of data analysis, previous works [11, 17, 21, 15, 8, 28] assume that each object is read once, and read by one reader only. Clearly, this assumption is difficult to enforce, but more importantly, it oversimplifies the reality. Because RFID readings are of low quality, many applications have to employ redundant readers to cover the target area completely to improve localization accuracy, which means objects are read by multiple readers simultaneously.

Indeed, in RFID systems, spatial redundancy is very common. Figure 1 shows an example of spatial redundancy where the target area is divided into six zones (using one dimensional model) and an RFID reader is located in the center of each zone. Spatial overlap of readers' detection regions leads to duplicate readings, i.e., an object is in the detection regions of multiple readers. A possible set of readings is shown in Table 1 wherein 1's denote successful detections and 0's otherwise. The table shows two effects of redundancy:

- Object 2 is detected by the reader in Zone 2 and also the reader in Zone 3, which makes it difficult to tell the exact location of Object 2. However, since an object cannot appear in more than one zone at the same time, at least one of the readings belongs to spatial redundancy.

- Object 3 is detected in Zone 4 only. However, it does not necessarily mean that Object 3 is in Zone 4 for sure. It is possible that the reader in the zone where Object 3 is located simply fails to detect it.

On the first look, spatial redundancy causes confusion as it introduces inconsistent information (e.g., about the location of Object 2). However, a redundant reading may supply the necessary information for the system to derive the location of an object when its intended reader fails to detect it (e.g., Object 3). Thus, the challenge is how to take advantage of redundancy while avoiding its undesirable effect in data cleansing.

Temporal Redundancy: Besides employing multiple stationary readers, many applications monitor the target area using a mobile reader (e.g., a handheld or a robot-mounted reader [24]) to take continuous readings on its route. Because the exact location of the mobile reader is always changing when the reader reports raw data, the detection regions at different time points may overlap, introducing temporal data redundancy of readings. However, if we treat the same reader at different time points as different readers, e.g., as shown in Table 1, when a mobile reader traverses from zone 1 to zone 6, the reader can be considered as the zone 1 reader while moving in zone 1 and the reader can be treated as the zone 2 reader while moving in zone 2, the temporal redundancy problem can be reduced to the spatial redundancy problem. Therefore, we will mainly focus on spatial redundancy in this paper.

1.2 Prior Knowledge

As false negatives and false positives abound in raw RFID readings [23, 17], in order to recover the true information, the data cleansing system should take prior knowledge into account. Prior knowledge may include information such as, for example, the detection areas of readers in Zone 2 and Zone 3 have significant overlapping, the positioning of the reader in Zone 4 makes it more likely to detect objects in Zone 3 than objects in Zone 5, or the reader in Zone 3 has high false negative rate, etc. Such information, when properly integrated with the readings, is extremely valuable for data cleansing.

1.3 Constraints

Environmental constraints can be utilized to improve data cleansing. For example, the maximal capacity of each zone (the number of objects that can reside in the same zone) is a constraint. If each zone represents a rack or shelf in a warehouse, one possible constraint is the total size or weight of the objects which the rack can hold. In addition to these physical constraints, information obtained from other channels can be translated into constraints. For instance, if an extra source indicates that two certain objects are in the same zone, it may help cleanse the data of these two as well as other objects when the information is integrated with readings and other constraints.

1.4 Overview of Our Approach

In this paper, we propose an innovative approach of cleansing RFID raw data which is able to take full advantage of duplicate readings and integrate prior knowledge as well as environmental constraints. Our approach is based on Bayesian inference. We introduce an n-state detection model and prove by entropy analysis that the 3-state detection model can maximize the system performance. Furthermore, we devise a Metropolis-Hastings sampler with constraints to efficiently approximate the posterior (Metropolis-Hastings sampling is a Markov Chain Monte Carlo method [22, 2]). Consequently, our approach enables two important types of queries against RFID raw data: the location query and the aggregate query (e.g., about remaining capacity of each zone). The contributions of this study are as follows:

- By using Bayesian inference, we derive a universal framework for computing the posterior probabilities (of the location of each object).

- Based on the physical characteristics of RFID readers, we propose an n-state detection model to capture likelihoods, which enables us to take full advantage of duplicate readings.

- We analyze the relationships between the system entropy and the read rate of RFID readers under the 2-state and 3-state detection models, respectively.

- By investigating the impact of the number of states in a detection model on the system entropy, we formally prove that the system entropy can be minimized if the 3-state model is adopted compared with other state models. In other words, having even more states (greater than 3) can in fact deteriorate the overall system performance.

- We devise MH-C, an improved Metropolis-Hastings sampler, to sample from the posterior while taking the environmental constraints into consideration.

- We demonstrate the efficiency and effectiveness of our approach by comparing the performance of MH-C with the Sequential Importance Sampling (SIS) based solution [28] through extensive simulations.

1.5 Paper Organization

The rest of this paper is organized as follows. The Bayesian inference-based framework of our approach is presented in Section 2. In Section 3, we propose the n-state detection model to take full advantage of duplicate readings. We prove that the 3-state detection model can maximize the system performance in Section 4. In Section 5, we introduce a Metropolis-Hastings sampler with constraints. The experimental validation of our design is presented in Section 6. Section 7 surveys the related work. Finally, Section 8 concludes this paper.

2. BAYESIAN INFERENCE

In this section, we develop a Bayesian inference-based approach to handle redundant readings and prior knowledge, and we analyze the challenges when applying this approach. Table 2 summarizes the notations used in this section.

2.1 A Bayesian Inference Based Approach

Bayesian inference is a statistical inference technique that estimates the probability of a hypothesis (x) based on observations (y). Bayesian inference shows that posterior is proportional to the multiplication of likelihood and prior, which can be represented as $p(x|y) \propto p(y|x)p(x)$.

Suppose there are m zones and n objects in our monitoring environment, each zone with a reader mounted in the zone center. Let o_i represent the object with ID i. For each o_i, its location is represented by a random variable h_i. Therefore, a possible distribution of n objects in m zones can be denoted as an instance of the random vector $\hat{H} = (h_1, h_2, \ldots, h_n)$. h_i represents the zone ID where object o_i is in. For example h_1 = 2 denotes that object o_1 is in zone 2 in the current instance. For the reader in zone j, the raw data (0 or 1) it receives from the RFID tag of object o_i is denoted as z_{ij}. The raw data matrix for each complete scan from m readers can then be represented as an $n \times m$ matrix $\mathbb{Z} = [z_{ij}]$. Thus the Bayes' theorem can be represented as Equation 1, where $post(\hat{H}|\mathbb{Z})$ denotes the posterior probability of location vector $\hat{H}$ given the raw data $\mathbb{Z}$, and a valid hypothesis means the hypothesis satisfies all constraints:

$$
\begin{array}{ll}
post(\hat{H}|\mathbb{Z}) = 0 & : \hat{H} \text{ is not valid} \\
post(\hat{H}|\mathbb{Z}) > 0 & : \hat{H} \text{ is valid} \\
post(\hat{H}_1|\mathbb{Z}) > post(\hat{H}_2|\mathbb{Z}) & : \hat{H}_1 \text{ is more likely than } \hat{H}_2
\end{array}
$$

In particular, if z_{ij} = 1 in a raw data matrix and the actual location of object o_i is not in zone j, then z_{ij} is a false positive. Take Table 1 as an example, at least one of z_{22} and z_{23} is a false positive because object 2 cannot be in zone 2 and zone 3 simultaneously. Similarly, at least one of z_{43} and z_{44} is a false positive.

To compute $post(\hat{H}|\mathbb{Z})$, we make some independence assumptions of random variables. RFID reader transmissions or tag transmissions may lead to collisions because readers and tags communicate over a shared wireless channel. Reader collisions happen when neighboring readers communicate with a tag simultaneously [9] and tag collisions occur when multiple tags transmit to a reader at the same time [10]. However, the two kinds of collisions can be effectively prevented by arbitration protocols (e.g. by scheduling adjacent readers to operate at different times) [25, 13, 20]. Therefore, we assume each reader detects the tags of different objects independently (i.e., whether a reader can successfully detect the tag of a certain object does not interfere with whether the reader can successfully detect that of another object) in this research. Based on the assumption, we can derive Equation 2.

Furthermore, because we employ MH-C to take into account constraints (i.e., to ensure that each generated sample satisfies all the constraints), here we can simply assume independence between different h_i (i.e., the locations of objects). In addition, we assume that each reader's detection of the same object is independent. Besides, the prior distribution of each object does not depend on that of other objects. Therefore, we can obtain Equation 3. If we rewrite Equation 3 using the normalizing constant, denoted as α, we can reach Equation 4, which shows how to compute the posterior of each sample. To be specific, $p(z_{ij}|h_i)$ reflects the corresponding *likelihood*, which is the probability that the reader in zone j reports the value of z_{ij} about object o_i given that object o_i is actually in the zone with ID h_i. Furthermore, $p(h_j)$ denotes the *prior probability* that the object o_j is in the zone with the ID of h_j. The prior probability can be interpreted as the assumed distribution before the acquiring of the RFID raw data.

Symbol	Meaning
$\hat{H}$	The random vector that represents the locations of all the objects
h_i	The random variable that represents the location of object o_i
$\mathbb{Z}$	Raw data reported by RFID readers
z_{ij}	The raw data (0 or 1) reported by the reader in zone j for object o_i
$post(\hat{H}\|\mathbb{Z})$	The posterior probability of the location vector $\hat{H}$ given the raw data $\mathbb{Z}$
$p(z_{ij}\|h_i)$	The likelihood that the zone j reader reports the value of z_{ij} for object o_i given that object o_i is in the zone with ID h_i
$p(h_j)$	The prior probability that object o_j is in the zone with ID h_j

Table 2: Symbolic notations of Section 2.

2.2 The Goal and the Obstacles

Based on Equation 4, given the raw readings $\mathbb{Z}$ and a hypothesis $\hat{H}$ (the location of each object), we can derive the probability of the hypothesis. However, finding just one valid hypothesis will provide nothing more than a biased answer to queries against the uncertain data. To address this issue, we need to query against all valid hypotheses. However, this is unrealistic because there are numerous valid hypotheses in most cases. Thus, our goal is to create a large sample set of valid hypotheses, each associated with a weight computed by Equation 4: $(\hat{H}_1, w_1), (\hat{H}_2, w_2), \cdots, (\hat{H}_n, w_n)$. The sample set of valid hypotheses as a whole enables us to answer queries with high credibility. To achieve this goal, we must overcome the following obstacles:

- A prerequisite for effective hypothesis sampling is to be able to compute the posterior probability of each hypothesis precisely. Therefore, we propose the n-state detection model in Section 3 to capture likelihoods in an affordable and accurate way.
- The hypothesis space is high dimensional, and the posterior probability is difficult to sample from. Thus, we need a sampling technique that has desirable efficiency. In Section 5, we apply a Markov Chain Monte Carlo method (MCMC) because MCMC can maintain the correlation between samples, resulting in an improved sampling efficiency.
- We need to incorporate constraint management in sampling. We propose a sampler called Metropolis-Hastings sampler with Constraints (MH-C), which improves the naive Metropolis-Hastings (MH) sampler. Each sample generated by MH-C automatically satisfies all the constraints.

3. RFID READER DETECTION MODELS

The major difficulty in computing the posterior of each sample (Equation 4) lies in how to accurately estimate the likelihood $p(z_{ij}|h_i)$. To do so, we introduce the n-state detection model of RFID readers to capture likelihoods in an affordable and precise way.

3.1 Physical Characteristics

An RFID reader sends RF signals to communicate with (passive) tags to retrieve a list of IDs in its detection range. However, RFID data acquisition and transmission are unreliable [10, 15, 17, 21]. In our experiments, we investigated the change of the read rate over distance using regular RFID readers and tags. The results are as

$$post(\hat{H}|\mathbb{Z}) = post(h_1, h_2, \ldots, h_n| \begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}) \propto p(\begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix} |h_1, h_2, \ldots, h_n) \cdot p(h_1, h_2, \ldots, h_n) \quad (1)$$

$$post(\hat{H}|\mathbb{Z}) \propto \prod_i p(z_{i1}, z_{i2}, \ldots, z_{im}|h_1, h_2, \ldots, h_n) \cdot p(h_1, h_2, \ldots, h_n) \quad (2)$$

$$post(\hat{H}|\mathbb{Z}) \propto \prod_i p(z_{i1}|h_i) \cdot p(z_{i2}|h_i) \cdot \ldots \cdot p(z_{im}|h_i) \cdot \prod_j p(h_j) \quad (3)$$

$$post(\hat{H}|\mathbb{Z}) = \alpha \cdot \prod_i p(z_{i1}|h_i) \cdot p(z_{i2}|h_i) \cdot \ldots \cdot p(z_{im}|h_i) \cdot \prod_j p(h_j) \quad (4)$$

illustrated in Figure 2. The model of the tags is Gen2 RFID Smart Label and the model of the reader is MPR-6000 (antenna 902-928 MH), provided by WJ Communications Inc. Our test environment is a lab with many metal objects (tables, desks and computer equipment), representing a noisy environment.

As Shown in Figure 2, the overall detection range of a reader can be separated into the major detection region and the minor detection region, where in the major detection region from 0 to almost 5 feet, the read rate can keep a level of around 95% and in the minor detection region approximately from 5 to 13 feet, the read rate drops off almost linearly. Furthermore, the read rate deteriorates to zero in the region more than 13 feet away from the reader, which is defined as beyond the overall detection range [17].

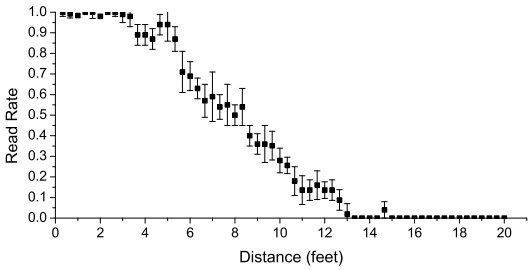


Figure 2: An illustration of the relationship between read rate and distance.

3.2 Problems of the 2-State Detection Model

Taking the scenario in Figure 1 as an example, one way to estimate likelihood is as follows:

$$p(z_{ij} = 1|h_i) = \begin{cases} r & \text{if } h_i \in \{j-1, j, j+1\} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$p(z_{ij} = 0|h_i) = \begin{cases} 1-r & \text{if } h_i \in \{j-1, j, j+1\} \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

where r is the average read rate. Intuitively, it means that an object o_i holds the same probability to be detected by a reader whether o_i is in the zone (j) the reader is associated to, or in any of neighboring zones ($j-1$ or $j+1$). Apparently, if compared with Figure 2, this 2-state detection model is inherently inaccurate as it fails to capture any change of the read rate in the overall detection region. Consequently, when predicting the location of an object, the resulting system is unable to differentiate between its own zone and all its neighboring zones because all the above zones are with an identical read rate.

In order to solve this problem, current works are forced to adopt a simplified 2-state detection model, which is shown in Figure 3. This simplified 2-state model assumes readers' detection regions do not overlap, i.e., a reader is only able to detect the objects in its own zone. Then the likelihood can be estimated as follows:

$$p(z_{ij} = 1|h_i) = \begin{cases} r & \text{if } h_i = j \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$p(z_{ij} = 0|h_i) = \begin{cases} 1-r & \text{if } h_i = j \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

This simplified 2-state model, however, has two problems. First, it is unrealistic to assume that we can divide the space into non-overlapping detection regions. Second, the model does not support applications that use redundant information (such as redundant readings) to offset the unreliability of raw RFID readings.

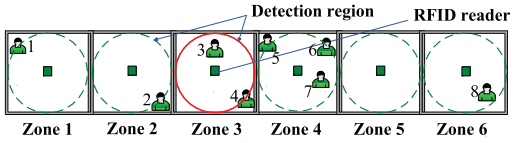


Figure 3: An illustration of the simplified 2-state detection model.

3.3 The n-state Detection Model

In order to take full advantage of duplicate readings, we propose an n-state detection model, which is illustrated in Figure 4. In Figure 4, the overall detection region of an RFID reader is divided into several sub-regions, each of which corresponds to a zone associated with a unique read rate. As far as a specific detection model is concerned, the difference in the read rate over any two adjacent sub-regions is a constant, i.e., the read rates for different states constitute an arithmetic sequence. Take the 4-state model as an example. Suppose the highest read rate in the model is x. The first state (counted with the increase of the detection distance) holds a read rate of x, the second state keeps a read rate of $\frac{2x}{3}$, the third state maintains a read rate of $\frac{x}{3}$, and the fourth state eventually has a read rate of zero. Thus, as for a specific reader, by employing the n-state detection model, each correlated zone is assigned a distinct read rate according to its distance to this reader. In particular, the simplest model in the family of the n-state detection model is the 2-state model, where an identical read rate is assumed in the overall detection region of each reader.

Notice that in practice, the n value depends on how zones are divided in overall detection regions of RFID readers. This is because an n-state model in fact implies that every $2(n-2)+1 = 2n-3$ zones correlate with each other (assuming that all the zones are in a 1-dimensional distribution). For example, if it is known as prior knowledge that one object can be read simultaneously by up to five readers, we have to choose $n = 4$ to incorporate the correlation

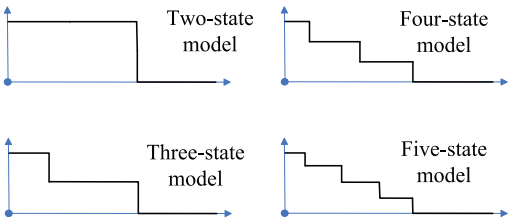


Figure 4: The family of the n-state detection model.

among every 5 successive zones. A formal derivation of the location distribution in terms of the probabilistic mass function can be found in Section 4.

3.4 A Case Study: The 3-State Model

We elaborate on the 3-state model as a case study. Suppose one reader can only detect its own zone and the two neighboring zones. This assumption implies that, as for a particular reader, there are three distinct location-based states of a object: in the same zone as the reader, in the neighboring zones, and in all the other zones. To capture this correlation, we have to choose $n = 3$ and the resulting model is the *3-state detection model*, where the overall detection range of a reader is divided into two sub-regions, as shown in Figure 5. Specifically, the major detection region, the minor detection region and the zero read rate region in Figure 5 correspond to the zone where the reader locates, neighboring zones and all the other zones, respectively. Therefore, the motivating scenario (Figure 1), if interpreted by the 3-state model, can be illustrated in Figure 6.

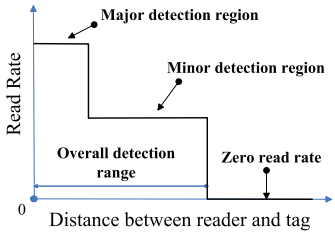


Figure 5: The 3-state detection model of RFID readers.

In Figure 6, by using the 3-state detection model, not only the duplicate readings can be incorporated, but also a zone and all its neighboring zone can be differentiated because they are with distinct read rates. To be specific, if object o_i is in zone j, not only z_{ij} but also $z_{i(j-1)}$ and $z_{i(j+1)}$ should have a considerable chance to be 1 (false positives). In the meantime, other readers are unable to detect the tag of object o_i. The reason is that object o_i may be in the major detection region of the reader in zone j while it may be also in the minor detection region of both readers in zones $j-1$ and $j+1$. As for the other readers, object o_i is totally beyond their overall detection regions, leading those readers to report 0 for object o_i. If we denote the mean value of the read rate in major detection region as r_{major} and the mean value of the read rate in minor detection region as r_{minor}, the estimate of the likelihood using the 3-state model can be represented in Equation 9 and Equation 10.

$$p(z_{ij} = 1|h_i) = \begin{cases} r_{major} & \text{if } h_i = j \\ r_{minor} & \text{if } h_i \in \{j-1, j+1\} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$p(z_{ij} = 0|h_i) = \begin{cases} 1 - r_{major} & \text{if } h_i = j \\ 1 - r_{minor} & \text{if } h_i \in \{j-1, j+1\} \\ 1 & \text{otherwise} \end{cases} \quad (10)$$

The advantage of the 3-state detection model over the 2-state detection model can be measured by the system entropy (Section 4). We also answer the question whether having even more states (more than 3) can further benefit the system.

4. ENTROPY ANALYSIS

We use entropy to measure the uncertainty in a system after invalid system states have been eliminated by a data cleansing method. Generally, applying an efficient data cleansing method will lead to systems with smaller entropy. In this section, we firstly show the advantage of the 3-state model over the 2-state model and then prove that the 3-state model can maximize the system performance compared with other detection models with even more than 3 states. A snippet of the RFID raw data is shown in Table 3 and the actual location of an object i is denoted as a random variable L.

4.1 Entropy versus Read Rate

Entropy of the 2-State model: Suppose that y denotes the read rate in the 2-state detection model. According to the right side of Equation 4, the probabilistic mass function of L in the 2-state model can be represented as:

$$p(L = l) = \begin{cases} \alpha(1-y)y(1-y)\beta & \text{if } l = j \\ \alpha(1-y)(1-y)y\beta & \text{if } l \in \{j-1, j+1\} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where α is the normalizing constant and β represents the prior probability (we assume the prior distribution as a uniform distribution) in Equation 4. Thus, we can calculate the entropy of the distribution of L as:

$$\begin{aligned} H(L) = \; & -\alpha(1-y)(1-y)y\beta \cdot \ln(\alpha(1-y)(1-y)y\beta) \\ & -\alpha(1-y)y(1-y)\beta \cdot \ln(\alpha(1-y)y(1-y)\beta) \\ & -\alpha(1-y)(1-y)y\beta \cdot \ln(\alpha(1-y)(1-y)y\beta) \end{aligned} \quad (12)$$

Because the probabilities on all the locations sum to 1, we can derive Equation 13. By applying Equation 13 to Equation 12 (α and β are canceled out), we can obtain Equation 14.

$$\alpha\beta = \frac{1}{3(1-y)^2 y} \quad (13)$$

$$H(L) = -3 \cdot \frac{1}{3} \cdot \ln \frac{1}{3} = 1.098 \quad (14)$$

Entropy of the 3-State model: Figure 6 corresponds to the 3-state model scenario. Suppose x is the read rate in the major detection

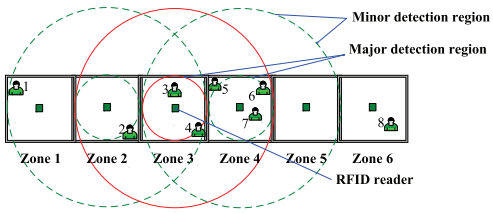


Figure 6: The detection-region overlap interpreted by the 3-state detection model.

	...	Zone $j-1$ Reader	Zone j Reader	Zone $j+1$ Reader	...
...	...	...	...	...	...
Obj i	...	0	1	0	...
...	...	...	...	...	...

Table 3: A snippet of the RFID raw data.

region. Then the read rate in the minor detection region can be denoted as $x/2$. Thus, according to the right side of Equation 4, the probabilistic mass function of L can be represented as follows:

$$p(L=l) = \begin{cases} \alpha(1-\frac{x}{2})x(1-\frac{x}{2})\beta & \text{if } l=j \\ \alpha(1-\frac{x}{2})(1-x)\frac{x}{2}\beta & \text{if } l \in \{j-1, j+1\} \\ 0 & \text{otherwise} \end{cases}$$

Similarly, α is the normalizing constant and β represents the prior probability in Equation 4. Therefore, we can calculate the entropy of the distribution of L as:

$$\begin{aligned} H(L) = \quad & -\alpha(1-\frac{x}{2})(1-x)\frac{x}{2}\beta \cdot \ln(\alpha(1-\frac{x}{2})(1-x)\frac{x}{2}\beta) \\ & -\alpha x(1-\frac{x}{2})^2\beta \cdot \ln(\alpha x(1-\frac{x}{2})^2\beta) \\ & -\alpha(1-\frac{x}{2})(1-x)\frac{x}{2}\beta \cdot \ln(\alpha(1-\frac{x}{2})(1-x)\frac{x}{2}\beta) \end{aligned} \tag{15}$$

Since probabilities on all locations sum to 1, we can obtain Equation 16.

$$\alpha\beta = \frac{1}{x(1-\frac{x}{2})(2-\frac{3x}{2})} \tag{16}$$

Combining Equation 16 and Equation 15, we have:

$$H(L) = -2 \cdot \frac{1-x}{4-3x} \cdot \ln\frac{1-x}{4-3x} - \frac{2-x}{4-3x} \cdot \ln\frac{2-x}{4-3x}$$

In Figure 7, we plot the relationship between the reconstruction entropy and read rate under the 2-state and 3-state models, respectively. As Figure 7 illustrates, the entropy will decrease accordingly with the increase of read rate, which indicates that the system will have less uncertainty with more reliable readers. Moreover, Figure 7 shows that the entropy in the 3-state model is always smaller than that in the 2-state model. For example, if $x = 0.95$, the entropy in the 3-state model is 0.395 while the entropy in the 2-state model is 1.098. This observation reveals that the 3-state model can be more informative in object localization than the 2-state model.

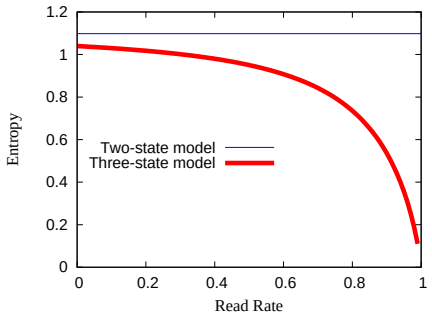


Figure 7: Relationship between entropy and read rate under the 2-state and 3-state models.

4.2 Entropy versus Number of States

Here we investigate the relationship between system entropy and the number of states in a detection model. Suppose an n-state model is adopted with the highest read rate of x. Thus, the read rate in the i^{th} state (counted with the increase of the detection distance), can be represented as $\frac{(n-i)\cdot x}{n-1}$. Combined with Equation 4 and Table 3, we can obtain the probabilistic mass function of L, represented as Equation 17. Equation 17 shows that in an n-state model, a successful reading "1" of a certain reader about an object in fact implies that this object may exist in any of the $2(n-2)+1 = 2n-3$ correlated zones (including the zone which the reader is associated to), each with a non-zero probability.

Based on Equation 17, we plotted the relationship between entropy and the number of states in a detection model in Figure 8, where we assume x equals to 0.95 (the most common case). According to Figure 8, the 3-state detection model can minimize the system entropy and lead to the maximum system performance. In other words, having more states (more than 3) in a detection model can even deteriorate the system performance. Therefore, our experiments are mainly focused on the 3-state model.

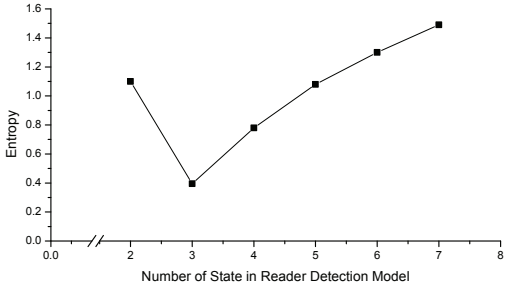


Figure 8: Relationship between entropy and the number of states in a detection model.

5. SAMPLING

By using Bayesian inference, we derive the posterior, as shown in Equation 4. Since Equation 4 is easy to compute but hard to sample from, we need an efficient method to draw samples from the posterior distribution. In this section, we firstly focus on the Metropolis-Hastings (MH) and Gibbs samplers. Next, we show why MCMC is chosen in our solution. Finally, we propose a Metropolis-Hastings sampler with Constraints (MH-C) method.

5.1 Metropolis-Hastings and Gibbs Sampling

The Metropolis-Hastings (MH) sampler and the Gibbs sampler are the two most common MCMC samplers. MH conducts a sequence of random walks using a proposal distribution and decides whether to reject the proposed moves using the rejection sampling. In the applications of Bayesian inference, the normalizing constant is usually extremely difficult to compute. MH avoids the computation of the constant. It approximates the posterior by using only the ratio of the posterior, where the constant is canceled out.

Recall that the random vector representing the locations of objects is denoted as $\hat{H}$ and the posterior distribution is $post(\hat{H}|\mathbb{Z})$. Suppose $\hat{H}_{t-1}$ is the immediate previous state before the state $\hat{H}_t$ in the formed Markov chain. According to the MH algorithm, at first, a proposal sample, $\hat{H}_q$, is drawn from a proposal distribution, $q(\hat{H}_q|\hat{H}_{t-1})$, i.e., $\hat{H}_q$ is a random deviation from $\hat{H}_{t-1}$. In our research, we use a uniform proposal distribution whose support is defined as the step length. The proposal sample $\hat{H}'$ can be denoted as $\hat{H}_{t-1}+\hat{H}_q$. Then MH accepts $\hat{H}'$ as the next state $\hat{H}_t$ with the probability of $\frac{post(\hat{H}'|\mathbb{Z})}{post(\hat{H}_{t-1}|\mathbb{Z})}$.

Here we compare MH sampler with the Gibbs sampler in brief. The Gibbs sampler requires that conditional (marginal) distributions for each variable are known and easy to sample from. MH relies on the ratio of the posterior, and does not require to sample from any distribution. Because we have already derived the closed form of the posterior as Equation 4 and are able to calculate likelihoods easily according to the proposed n-state model, it will be much more straightforward to use MH sampler rather than the Gibbs sampler in our design.

5.2 Sample Correlation

$$p(L=l)=\begin{cases}\alpha(1-\frac{x}{n-1})(1-\frac{2x}{n-1})...(1-\frac{(n-2)x}{n-1})\cdot x\cdot(1-\frac{(n-2)x}{n-1})...(1-\frac{2x}{n-1})(1-\frac{x}{n-1})\beta & \text{if } l=j\\ \alpha(1-\frac{x}{n-1})(1-\frac{2x}{n-1})...(1-\frac{(n-2)x}{n-1})\cdot\frac{(n-1-k)x}{n-1}\cdot(1-\frac{(n-2)x}{n-1})... & \\ (1-\frac{(n-1-k+1)x}{n-1})\cdot(1-x)\cdot(1-\frac{(n-1-k-1)x}{n-1})...(1-\frac{2x}{n-1})(1-\frac{x}{n-1})\beta & \text{if } l\in\{j-k,j+k\}, k\in\{1,2,...,n-2\}\\ 0 & \text{otherwise}\end{cases} \tag{17}$$

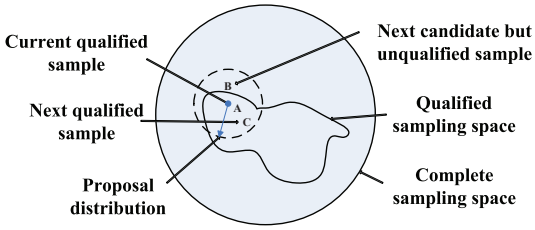


Figure 9: Taking advantage of the correlation between samples to improve sampling efficiency.

In our design, we choose MCMC instead of other sampling technique because MCMC maintains the correlation among samples. In MCMC, the next sample depends on the current sample. Before we elaborate on how we can take advantage of sample correlation to improve the efficiency of sampling in our scenario, we define two terms as follows.

Definition We call any sample generated by the sampler a *candidate sample*. A *qualified sample* is a candidate sample that satisfies all constraints.

The existence of constraints leads to the uniqueness of our problem. Samples must satisfy all the constraints to become qualified. Note that in most sampling problems, we prefer independent samples, that is, the current draw of a sample is independent from the previous draw. In our scenario, however, the sampling techniques which generate independent samples (e.g., importance sampling) may suffer from low sampling efficiency due to the loss of correlation between adjacent samples. Figure 9 illustrates how the correlation between samples can be utilized to improve the sampling efficiency. The qualified sampling space is a subset of the complete sampling space. Suppose sample point A is the current sample in the qualified sampling space. For an independent sampler, the next sample could be any point in the complete sampling space. However, the next sample is useful only if it happens to fall into the qualified sampling space. On the contrary, if a MCMC-based sampler is employed, the next sample will be chosen according to the proposal distribution at point A, i.e., the next sample will be in the area denoted by the dotted circle centering at point A. Therefore, compared to other independent samplers, the probability that the next sample generated by MCMC falls into the qualified sampling space is considerably increased.

Note that although MCMC improves sampling efficiency, a sample generated by MCMC may not necessarily be a qualified sample. As in Figure 9, point B is a sample generated by MCMC after point A. However, point B is outside the qualified sampling space. Consequently, we have to sample again to acquire point C, which is a qualified sample, and then add it into the Markov chain as the next state. Afterward, the Markov chain moves from point A to point C.

5.3 Metropolis-Hastings Sampler with Constraints

Although the naive MH algorithm can evaluate the posterior by forming a Markov chain in the sampling space, it does not take constraints into account. If we impose constraints to samples, many of them should be rejected because they are inapplicable, i.e., they are not qualified samples.

Symbol	Meaning
$\mathbb{Z}$	The raw data matrix from RFID readers
$\mathbb{S}$	The sample set
$\vec{C}$	The current sample in the Markov chain
$\vec{P}$	The proposal sample in the Markov chain
C_j	The j^{th} dimension of $\vec{C}$
P_j	The j^{th} dimension of $\vec{P}$
E	The number of effective samples
B	The number of samples in the burn-in phase
S	The step length for the uniform proposal distribution
D_{object}	The total number of monitored objects
D_{zone}	The total number of zones
$Jitter$	A random number between 0 and 1
$Rand(a,b)$	Generate a random integer between a and b based on uniform distribution
$Post(\hat{H}\|\mathbb{Z})$	The posterior probability of the sample $\hat{H}$ given raw data $\mathbb{Z}$

Table 4: Symbolic notations used in Algorithm 1.

To incorporate constraints in sampling, we propose a Metropolis-Hastings sampler with Constraints (MH-C). With MH-C, each zone is associated with multiple variables called *resource descriptors*. The current value of a *resource descriptor* represents how much the associated resource is still available. Suppose we have a variable, denoted as $Descriptor_{zone_i}$, to keep track of the current available vacancy in zone i. The initial value of $Descriptor_{zone_i}$ is set to the maximal capacity of zone i. Thus, whether an object j with the volume $Volume_{object_j}$ is able to be stored in zone i can be examined by:

$$Descriptor_{zone_i} = Descriptor_{zone_i} - Volume_{object_j} \tag{18}$$

The proposed resource allocation is feasible only if $Descriptor_{zone_i}$ is no less than zero. Otherwise, we have to re-sample until a new allocation meets all the constraints. Consequently, the problem of whether an allocation is feasible (or compatible) can reduce to the problem of monitoring the value of each descriptor.

With MH-C, because each sample is a D_{object}-dimensional vector, a proposal sample is generated iteratively dimension by dimension. If any descriptor for the current allocation is less than zero, there will be no chance for the current partial sample to become a qualified sample. Therefore, we can discard the current value and then choose another value for that dimension by re-sampling. As far as the proposal distribution is concerned, we construct a random walk chain by choosing a uniform proposal distribution within the step length. A detailed description of MH-C algorithm is illustrated in Algorithm 1 and the related notations are summarized in Table 4.

In Algorithm 1, line 1 initializes the sample set and takes the RFID raw data. Line 2 loads the n-state detection model of readers with the objective of computing the likelihood. Line 3 initializes all the resource descriptors and line 4 randomly chooses a qualified sample as the first state of the Markov Chain. Lines 6 to 17 generate a random sample as the proposal sample dimension by dimension (object by object). Lines 8 to 13 correspond to the sam-

Algorithm 1 Metropolis-Hastings Sampler with Constraints

1: Set $\mathbb{S} = \emptyset$ and take raw data $\mathbb{Z}$
2: Load the n-state detection model
3: Initialize all the resource descriptors to their maximal capacity.
4: Initialize $\vec{C}$ by randomly choosing a qualified sample within the support of $Post(\hat{H}|\mathbb{Z})$ as the starting point.
5: **for** $Cycle$ = 2 to E+B **do**
6: **for** j = 1 to D_{object} **do**
7: **repeat**
8: P_j = C_j+ Rand(-S,S)
{Generate a new integer based on the current value and a proposal value within the step length}
9: **if** $P_j < 1$ **then**
10: P_j = 1+ (1-P_j)
{Overflow and Reflection}
11: **end if**
12: **if** $P_j > D_{zone}$ **then**
13: P_j = D_{zone}-(P_j-D_{zone})
{Overflow and Reflection}
14: **end if**
15: **until** The value of any resource descriptor related to the referred zone is no less than zero after the proposed allocation on the current object is committed
16: $j \leftarrow j + 1$
17: **end for**
18: Generate a random number between 0 and 1: $Jitter$
19: **if** $Jitter \leq \min(1, \frac{Post(\vec{P}|\mathbb{Z})}{Post(\vec{C}|\mathbb{Z})})$ **then**
20: $\vec{C}$=$\vec{P}$
{Metropolis-Hastings}
21: **end if**
22: Add $\vec{C}$ into $\mathbb{S}$ as the next sample
23: Resetting all the resource descriptors
24: $Cycle \leftarrow Cycle + 1$
25: **end for**

pling process. First, we obtain a random integer based on the current value and the random proposal value. The correct range of the value on each dimension of the sample vector, represented as h_i, is [1, D_{zone}]. Therefore, if the proposal value P_j overflows in the range, we need to make the value reflect into the range. Then, we check all the related descriptors to make sure their updated values are no less than zero. If any value is less than zero, it means that the current allocation will violate the corresponding constraints. Consequently, we go back to line 8 to re-sample until an allocation on that dimension is feasible, as shown in line 15. Note that our algorithm guarantees each proposal sample is also a qualified sample. After a complete proposal sample is generated, lines 18 to 21 accept this proposal sample as the next state of the Markov chain with the probability of the posterior ratio of the proposal sample over the current sample. Line 22 adds the next state into the sample set. Line 23 resets all the resource descriptors to make sure that the examination on descriptors for the next proposal sample is correct. Note that line 5 is to set the upper bound of the sampling loop to E + B in order to guarantee that the final number of samples in sample set is E + B. It is because the first B samples should be excluded as burn-in samples and consequently only the remaining E samples will be taken into account to reconstruct the posterior to answer the related queries.

6. EXPERIMENTAL VALIDATION

In this section, we applied our Bayesian inference-based method to a warehouse application, i.e., each object corresponds to a case and each zone corresponds to a rack. To capture the likelihood, without loss of generality, we assumed that the 3-state detection model can be adopted in this application. Also, we implemented MH-C (we use the terms MH-C and MCMC interchangeably in this section) to sample from the posterior. For comparison with our MCMC-based solution, we extended the SIS-based approach in [28] to incorporate duplicate readings because their approach does not consider duplicate readings and only focuses on the distribution of the missing cases. To show the scalability of our approach, our experiments focus on two settings. In the small-scale warehouse experiment, we demonstrated the results of the related queries returned by MCMC and SIS based on a specific true distribution matrix. On the other hand, in the large-scale warehouse experiment, we randomly generated 100 times the true distribution matrix and the corresponding noisy RFID input matrix to investigate the performance of MCMC and SIS in terms of reconstruction efficiency and accuracy.

6.1 Types of Issued Queries

In our simulations, we investigate two common types of queries: location queries and remaining capacity queries. Location queries are used to retrieve the location of a specific object while remaining capacity queries are used to find the leftover volume of a certain resource in a specific zone. Their definitions are as follows.

- *Location query:* Given an object and a zone, a location query returns the probability that the object is in the zone.
- *Remaining capacity query:* Given a zone, a remaining capacity query returns the leftover volume of a certain resource in the zone.

6.2 Simulator Implementation

Our simulator consists of seven components as displayed in Figure 10. The true matrix generator randomly produces distribution matrices as true distributions. The rows represent cases (objects) and columns represent racks (zones). On the contrary, noisy matrix generator provides the noisy matrices as the RFID raw data in the same format. Then MCMC and SIS modules reconstruct the distribution for each case using the noisy matrix as the input. Our simulator generates the synthetic RFID raw data with the duplicate readings according to the physical characteristics of RFID readers [17]. The 3-state detection model was used to capture the likelihood. Also, we assume as the prior distribution that each case exists on each rack with the same probability. Table 5 lists the parameters used in our simulator. For the simplicity of presentation, we used the small-scale parameter set to visualize query results. On the contrary, the large-scale parameter set was used to investigate the advantage of our MCMC-based approach over the SIS-based approach [28]. All our experiments were conducted on a Linux machine with an Intel Pentium 4 2.4GHz processor with 2GB of memory.

We employed K-L divergence, the top-1 success rate, and the top-2 success rate to evaluate the reconstruction accuracy. Specifically, K-L divergence is a metric commonly used to evaluate the difference between two distributions. In our research, we calculated the K-L divergence from the reconstructed distribution to the

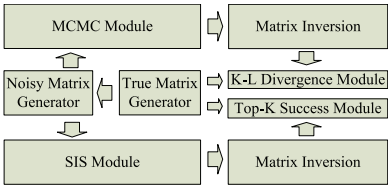


Figure 10: The simulator structure.

Parameter	Small scale	Large scale
D_{object}	20	5000
D_{zone}	5	200
B	50	50
S	1	30
$Volume_{object}$	1	1
$Capacity_{zone}$	7	50

Table 5: The parameters for our simulations.

true distribution, i.e., the smaller value of K-L convergence indicates the higher accuracy of the reconstructed distribution. The recovered matrices (reconstructed distributions) were inverted by the matrix inversion module to facilitate the computation of K-L divergence. Then the K-L divergence module was used to compare the reconstructed distributions with the true distributions and compute the corresponding K-L divergence values for MCMC and SIS. On the other hand, the top-k success rate reflected how many cases can be located precisely at a certain resolution using data cleansing techniques. The top-k analysis module was responsible for calculating the top-k success rate.

Definition The *top-k success rate* is a percentage of the number of cases whose true locations match the top k predicted locations of the reconstructed distribution over the total number of cases.

6.3 Visualization of Query Results

To visualize the results of location and remaining capacity queries, for the simplicity of presentation, we issued queries against the small-scale warehouse setting. As Table 5 shows, we assume that there are twenty cases with identical length and five racks with a length limit to accommodate at most seven cases. True distributions of the first six cases are shown in Table 6(a) where the rows represent cases and columns represent racks. To be specific, if a case is on a rack, the corresponding position is 1 and 0 otherwise. Therefore, according to Table 6(a), the first case is on the first rack, the second case is on the fourth rack and so on. On the other hand, the corresponding noisy RFID data matrix, generated by our simulator, is shown in Table 6(b) in the same format. By comparing the two matrices, we can easily see that there is notable difference (existence of noise and duplicate readings) between these two matrices. Specifically, duplicate readings of the fourth and fifth case (consecutive 1's) occur in the noisy raw data. Afterward, we employed MCMC and SIS to cleanse the noisy distribution matrix to recover the true distributions. In this experiment, we drew 5000 qualified samples for MCMC and SIS, respectively, and we assume that the read rate in the major detection region of readers is 95%.

6.3.1 The Location Query

The results of the location queries returned by MCMC and SIS for the first six cases are demonstrated in Figure 11. For the first case, the correct location is the first rack. MCMC predicted a probability of 0.47 on that rack while SIS predicted a probability of only 0.01. Similarly, for the second case, the true location is the fourth rack. MCMC predicted a probability of 0.70 on that rack while SIS predicted a probability of 0.98. For the third case, the accurate location is the third rack. MCMC predicted a probability of 0.53 on that rack while SIS predicted a probability of 0.98. For the fourth case, the exact location is the second rack. MCMC predicted a probability of 0.51 on that rack while SIS predicted a probability of 0.32. For the fifth case, the precise location is the fifth rack. MCMC predicted a probability of 0.53 on that rack while SIS predicted a probability of 0.99. At last, for the sixth case, the correct location is the first rack. MCMC predicted a probability of 0.20 on that rack while SIS predicted a probability of only 0.05. In summary, MCMC tends to generate a much smoother probability distribution than SIS. Consequently, MCMC provides a superior overall prediction of the distribution of all the objects monitored by an RFID system compared with SIS.

$$\begin{bmatrix} 1&0&0&0&0\\ 0&0&0&1&0\\ 0&0&1&0&0\\ 0&1&0&0&0\\ 0&0&0&0&1\\ 1&0&0&0&0 \end{bmatrix} \quad \begin{bmatrix} 1&0&0&0&0\\ 0&0&0&1&0\\ 0&1&0&0&0\\ 1&1&0&0&0\\ 0&0&0&1&1\\ 0&1&0&0&0 \end{bmatrix}$$

(a) (b)

Table 6: The matrix of the first six cases: (a) true distribution and (b) noisy raw data.

6.3.2 The Remaining Capacity Query

We applied the available length on a rack as the acquirable volume of that rack to demonstrate the remaining capacity query. The results of the remaining capacity queries answered by MCMC for the first five racks are demonstrated in Figure 12. Note that the remaining capacity on each rack is 3 because each rack exactly accommodates 4 cases according to the true distribution matrix used for this experiment (the first six rows of the matrix are as shown in Table 6(a)). As illustrated in Figure 12, the remaining volumes on racks 1 and 4 were reported correctly. On the contrary, the remaining volumes on racks 2 and 3 were underestimated and the remaining volume on rack 5 was overestimated.

6.4 The Performance Analysis of MH-C

In this section we focus on the performance of MCMC and SIS with respect to the reconstruction efficiency and accuracy in the large-scale warehouse setting where there are 5000 cases and 200 racks in total as shown in Table 5. For each result in this section, we randomly generated the true distribution matrix and the corresponding RFID noisy matrix 100 times. We reconstructed case distributions, recorded the average sampling time, and computed the average K-L divergence, the top-1 success rate and the top-2 success rate of all the involved 5000 cases.

6.4.1 The Reconstruction Efficiency

Here we investigated the performance of MCMC and SIS in terms of the average sampling time. Compared to SIS, the average sampling time of MCMC is remarkably reduced over different number of qualified samples as illustrated in Figure 13. For example, with 5000 qualified samples the sampling time of MCMC is 11.58 seconds while the sampling time of SIS is 230.78 seconds. This is because MCMC takes advantage of the current qualified sample to generate the next qualified sample (i.e., keeping the relevance of samples). Consequently, MCMC takes less time than SIS to come up with the same number of qualified samples .

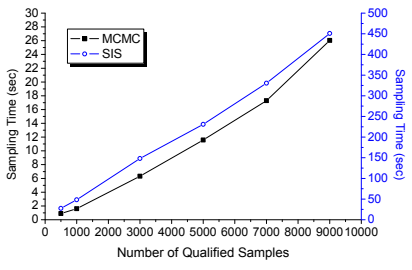
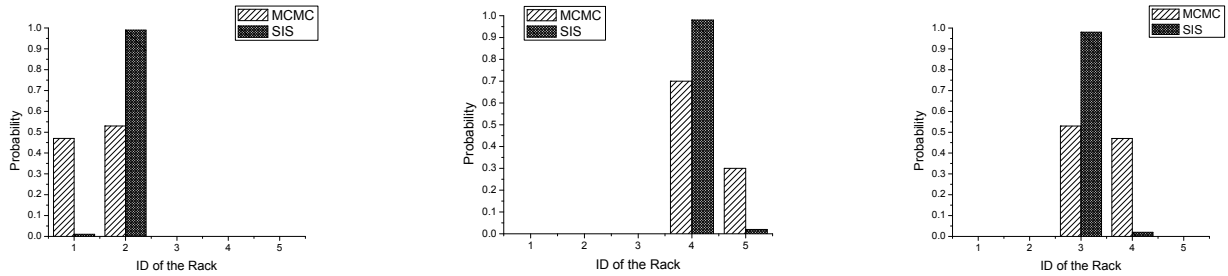
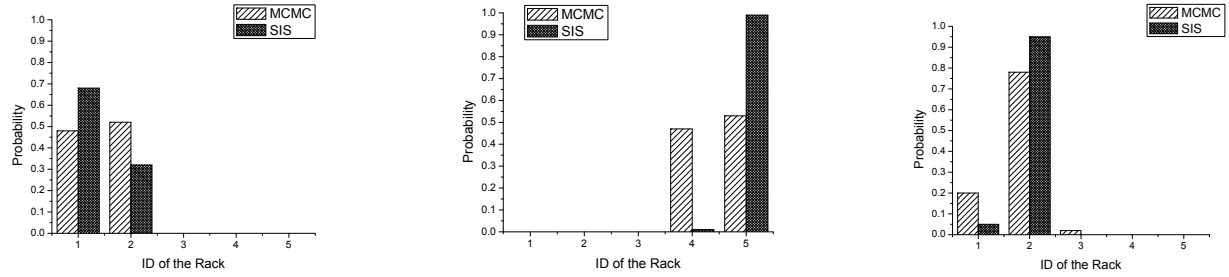


Figure 13: The comparison of MCMC and SIS on sampling time.

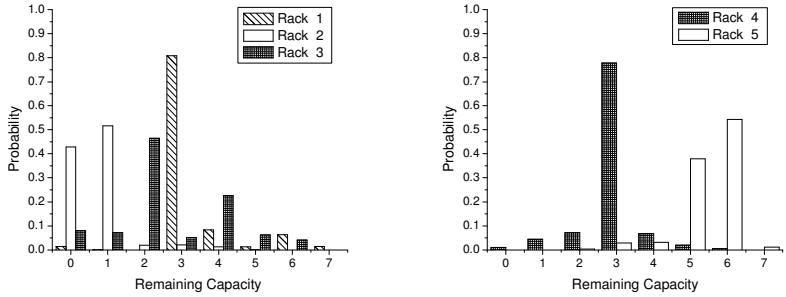


(a) First case (true location: rack 1). (b) Second case (true location: rack 4). (c) Third case (true location: rack 3).



(d) Fourth case (true location: rack 2). (e) Fifth case (true location: rack 5). (f) Sixth case (true location: rack 1).

Figure 11: The results of the location queries answered by MCMC and SIS.



(a) Racks 1, 2 and 3 (true value: 3). (b) Racks 4 and 5 (true value: 3).

Figure 12: The results of the remaining capacity queries answered by MCMC.

6.4.2 The Reconstruction Accuracy

In this experiment, we varied the number of qualified samples, data redundancy degree, and the number of managed racks per reader to investigate their effects on the reconstruction accuracy.

The Impact of the Number of Qualified Samples

We first increased the number of qualified samples from 500 to 9000 to investigate the performance of MCMC and SIS on reconstruction accuracy. Here we assumed that the read rate in the major detection region is 95% and the number of racks managed by a reader is 1. As demonstrated in Figure 14(a), with the increase of the number of qualified samples, the K-L divergence values of both approaches kept decreasing. However, MCMC always outperformed SIS with all experimented sample numbers. Particularly, when we drew 500 qualified samples, the K-L divergence of MCMC was 0.86 while the K-L divergence of SIS was 3.78. When we picked 9000 qualified samples, the K-L divergence of MCMC reduced to a remarkable value 0.64 comparing with 2.77 of the SIS solution. Also, as far as the top-1 success rate is concerned, with the increase of the number of qualified samples, the top-1 success rate of MCMC increased from 0.50 to 0.70 while the top-1 success rate of SIS extended from 0.36 to 0.55 as shown in Figure 14(b). In addition, the top-2 success rate of MCMC raised from 0.70 to 0.89 while the top-2 success rate of SIS changed from 0.60 to 0.76 as demonstrated in Figure 14(c).

The Impact of the Redundancy Degree

Next, we studied the performance of MCMC and SIS on the reconstruction accuracy by varying the data redundancy degree. Because the false positives are actually the successful readings about the objects in the minor detection regions of readers, we use the read rate in the minor detection region to define the data redundancy degree. The larger redundancy degree indicates the higher probability that a reader can detect an object in the neighboring zones (or racks).

Definition The *data redundancy degree* is the probability that a reader successfully detects a object in the minor detection region of that reader.

Here, we varied the data redundancy degree from 0.325 to 0.475, corresponding to the read rate in the major detection regions from 65% (the least reliable reader) to 95% (the most reliable reader).

For each experiment, we drew 5000 qualified samples and the number of racks managed by a reader is 1. Figure 15 illustrates the results. With the enlargement of data redundancy degree, both the performances of MCMC and SIS on reconstruction accuracy are elevated. Specifically, as demonstrated in Figure 15(a), MCMC always maintained a lower K-L divergence value than SIS, reflecting a more precise prediction. Furthermore, as shown in Figure 15(b), the top-1 success rate of MCMC increased from 0.54 to 0.65 with the increase of the data redundancy degree while the top-1 success rate of SIS expanded from 0.42 to 0.51. Figure 15(c) demonstrates how the top-2 success rate increased when we raised the redundancy degree for MCMC and SIS.

The Impact of the Number of Managed Racks per Reader

We evaluated the performance of MCMC and SIS by varying the number of managed racks per reader. In order to deploy readers in a warehouse more efficiently, users may want to assign multiple racks to be managed by a single reader. Taking into account the fact that the overall detection region of a regular RFID reader has little chance to be more than 20 feet [17], we changed the number of racks managed by a reader from 1 to 6. As Figure 16(a) demonstrates, when each rack had its own reader, the K-L divergence values of MCMC and SIS were 0.68 and 3.11, respectively. When a reader monitored more racks, both the K-L divergence values of MCMC and SIS deteriorated. When a reader was responsible for detecting cases on six racks, the K-L divergence values raised to 1.66 of MCMC and 4.01 of SIS. Moreover, as demonstrated in Figure 16(b), with the enlargement of the number of managed racks per reader the top-1 success rate of MCMC decreased from 0.65 to 0.55. On the other hand, the top-1 success rate of SIS dropped from 0.51 to 0.41. Also, Figure 16(c) depicts how the top-2 success rate of MCMC and SIS decreased correspondingly.

7. RELATED WORK

Many systems have been developed to manage data with uncertainty. The usual approach to address uncertainty is to augment the classical relational model with attribute-level or tuple-level probability values [1, 3, 4, 6, 7]. On the other hand, a more general approach based on sampling is proposed for managing incomplete and uncertain data [14, 28]. The idea is simple and intuitive: we construct random samples while observing the prior statistical knowledge and constraints about the data. Thus, each sample is one possible realization in the space of uncertainty, and the entire set of samples reveals the distribution of the uncertain data we want to model. Queries and inferences are then conducted against this distribution.

An important application that drives the recent surge of interest in managing incomplete and uncertain data is RFID data management. Chawathe et al. [5] proposed a system architecture of a distributed RFID system and discussed related data management challenges such as inferences and online warehousing. An expressive temporal data model for RFID data is defined by Wang and Liu [26] to support tracking and monitoring queries. Gonzalez et al. [12] designed a warehousing model which provides RFID data compression and path-dependent aggregates. Based on the proposed warehousing framework, they also developed techniques for summarizing and indexing RFID data and various query methods. Because RFID raw data usually contains anomalies [10], solutions have been proposed to clean the input data from readers. Jeffery et al. [15] presented a framework for building data cleaning infrastructure to support pervasive applications. An adaptive smoothing filter (SMURF) for RFID data cleaning was proposed in [17]. SMURF focuses on a sliding-window aggregate that interpolates for lost readings. Our paper, however, is motivated by how to leverage data redundancy to elevate the localization accuracy for all the tagged objects in a target area.

The works in [18, 19, 28] are the most relevant research to this paper. For correcting erroneous RFID raw data, Khoussainova et al. [18, 19] presented a system for correcting input data errors automatically using application defined global integrity constraints. The system corrects the input data by inserting missing tuples when necessary and assigning to each one the probability that it is correct for groups of conflicting tuples. This maximum entropy based solution is practical. However, it is unable to capture all application related prior knowledge and dependency compared with sampling-based approaches. Useful information can be recovered from noisy RFID data by exploiting constraints and sequential importance sampling methods [28]. Nevertheless, the work in [28] failed to consider the duplicate readings caused by the overlapped detection regions of RFID readers.

8. CONCLUSIONS

The data reported by RFID devices are known to be unreliable. In this research, we propose a Bayesian inference based approach for cleansing RFID raw data which can take advantage of duplicate readings. In order to evaluate the location and aggregate queries, our approach employs prior knowledge to quantify the degree of uncertainty on the location of each object and the remaining capacities in each zone. Furthermore, we propose the n-state model to capture likelihood and validate that the 3-state model can maximize the system performance. Finally, we devise MH-C to efficiently sample from the posterior distribution under environmental constraints.

9. REFERENCES

[1] P. Agrawal, O. Benjelloun, A. D. Sarma, C. Hayworth, S. Nabar, T. Sugihara, and J. Widom. Trio: A System for Data, Uncertainty, and Lineage. In *VLDB*, pages 1151–1154, 2006.

[2] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An Introduction to MCMC for Machine Learning. *Machine Learning*, 50(1-2):5–43, 2003.

[3] P. Andritsos, A. Fuxman, and R. J. Miller. Clean Answers over Dirty Databases: A Probabilistic Approach. In *ICDE*, page 30, 2006.

[4] L. Antova, C. Koch, and D. Olteanu. Query Language Support for Incomplete Information in the MayBMS System. In *VLDB*, pages 1422–1425, 2007.

[5] S. S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. E. Sarma. Managing RFID Data. In *VLDB*, pages 1189–1195, 2004.

[6] R. Cheng, S. Singh, and S. Prabhakar. U-DBMS: A Database System for Managing Constantly-evolving Data. In *VLDB*, pages 1271–1274, 2005.

[7] N. Dalvi and D. Suciu. Efficient Query Evaluation on Probabilistic Databases. *The VLDB Journal*, 16(4):523–544, 2007.

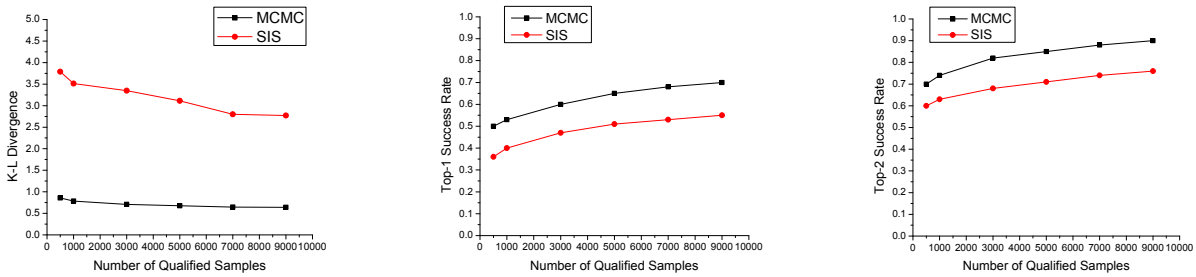
[8] A. Deshpande, C. Guestrin, and S. Madden. Using probabilistic models for data management in acquisitional environments. In *CIDR*, pages 317–328, 2005.

[9] D. W. Engels and S. E. Sarma. The Reader Collision Problem. In *IEEE SMC*, 2002.

[10] C. Floerkemeier and M. Lampe. Issues with RFID Usage in Ubiquitous Computing Applications. In *Pervasive*, pages 188–193, 2004.

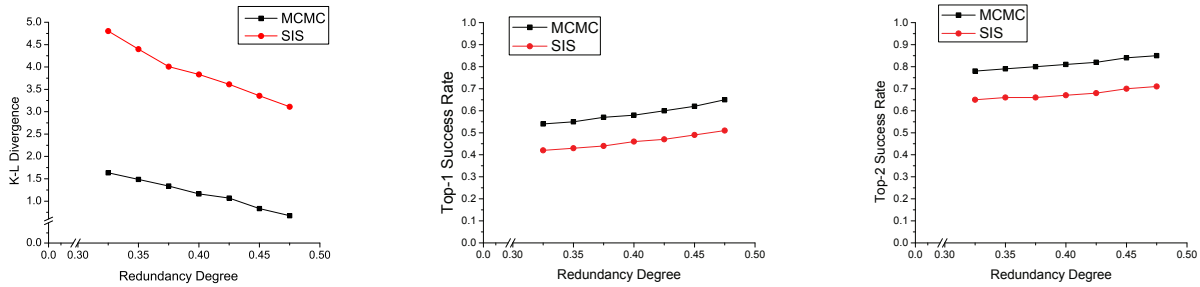
[11] M. J. Franklin, S. R. Jeffery, S. Krishnamurthy, F. Reiss, S. Rizvi, E. Wu, O. Cooper, A. Edakkunni, and W. Hong. Design Considerations for High Fan-In Systems: The HiFi Approach. In *CIDR*, pages 290–304, 2005.

[12] H. Gonzalez, J. Han, X. Li, and D. Klabjan. Warehousing



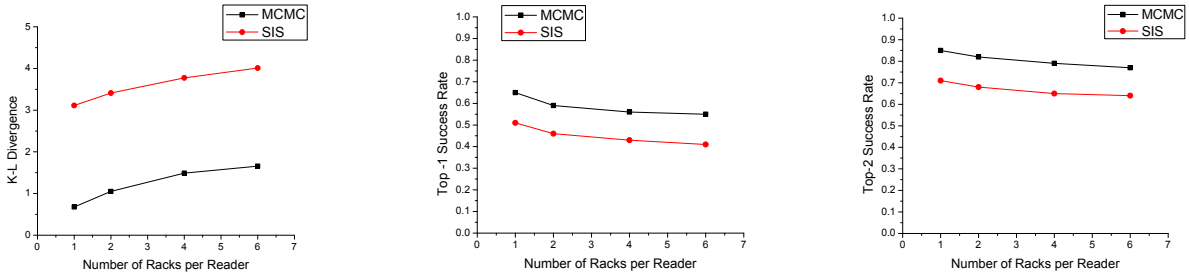
(a) K-L divergence of 5000 cases. (b) Top-1 success rate of 5000 cases. (c) Top-2 success rate of 5000 cases.

Figure 14: The impact of the number of qualified samples.



(a) K-L divergence of 5000 cases. (b) Top-1 success rate of 5000 cases. (c) Top-2 success rate of 5000 cases.

Figure 15: The impact of RFID redundancy degree.



(a) K-L divergence of 5000 cases. (b) Top-1 success rate of 5000 cases. (c) Top-2 success rate of 5000 cases.

Figure 16: The impact of the number of managed racks per reader.

and Analyzing Massive RFID Data Sets. In *ICDE*, page 83, 2006.

[13] J. Ho, D. W. Engels, and S. E. Sarma. HiQ: A Hierarchical Q-learning Algorithm to Solve the Reader Collision Problem. In *SAINT Workshops*, pages 88–91, 2006.

[14] R. Jampani, F. Xu, M. Wu, L. L. Perez, C. Jermaine, and P. J. Haas. MCDB: A Monte Carlo Approach to Managing Uncertain Data. In *SIGMOD*, pages 687–700, 2008.

[15] S. R. Jeffery, G. Alonso, M. J. Franklin, W. Hong, and J. Widom. Declarative Support for Sensor Data Cleaning. In *Pervasive*, pages 83–100, 2006.

[16] S. R. Jeffery, M. J. Franklin, and M. N. Garofalakis. An Adaptive RFID Middleware for Supporting Metaphysical Data Independence. *VLDB J.*, 17(2):265–289, 2008.

[17] S. R. Jeffery, M. N. Garofalakis, and M. J. Franklin. Adaptive Cleaning for RFID Data Streams. In *VLDB*, pages 163–174, 2006.

[18] N. Khoussainova, M. Balazinska, and D. Suciu. Towards Correcting Input Data Errors Probabilistically Using Integrity Constraints. In *MobiDE*, pages 43–50, 2006.

[19] N. Khoussainova, M. Balazinska, and D. Suciu. Probabilistic Event Extraction from RFID Data. In *ICDE*, pages 1480–1482, 2008.

[20] J. Myung, W. Lee, J. Srivastava, and T. K. Shih. Tag-Splitting: Adaptive Collision Arbitration Protocols for RFID Tag Identification. *IEEE Trans. Parallel Distrib. Syst.*, 18(6):763–775, 2007.

[21] J. Rao, S. Doraiswamy, H. Thakkar, and L. S. Colby. A Deferred Cleansing Method for RFID Data Analytics. In *VLDB*, pages 175–186, 2006.

[22] S. M. Ross. *Introduction to Probability Models, Ninth Edition*. Academic Press, 2006.

[23] L. Sullivan. RFID Implementation Challenges Persist, All This Time Later. *InformationWeek*, October 2005.

[24] T. Tran, C. Sutton, R. Cocci, Y. Nie, Y. Diao, and P. Shenoy. Probabilistic Inference over RFID Streams in Mobile Environments. In *ICDE*, 2009.

[25] J. Waldrop, D. W. Engels, and S. E. Sarma. Colorwave: An Anticollision Algorithm for the Reader Collision Problem. In *ICC*, pages 1206– 1210, 2003.

[26] F. Wang and P. Liu. Temporal Management of RFID Data. In *VLDB*, pages 1128–1139, 2005.

[27] R. Want. The Magic of RFID. *ACM Queue*, 2(7):40–48, 2004.

[28] J. Xie, J. Yang, Y. Chen, H. Wang, and P. S. Yu. A Sampling-Based Approach to Information Recovery. In *ICDE*, pages 476–485, 2008.

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：99 年 6 月 15 日

計畫編號	NSC98－2218－E－009－023		
計畫名稱	Eco-Community： 以感測網路建構一個虛境及實境互動式智慧型社區		
出國人員姓名	孫敏德	服務機構及職稱	國立中央大學
會議時間	2010 年 6 月 6 日至 2010 年 6 月 11 日	會議地點	Indianapolis, Indiana, USA (印第安納波里,美國印第安納州)
會議名稱	(中文) (英文)2010 ACM SIGMOD/PODS Conference: Indianapolis, Indiana, USA		
發表論文題目	(中文)利用時空重複的特性達成 RFID 的資料清洗 (英文)Leveraging Spatio-Temporal Redundancy for RFID Data Cleansing		

一、參加會議經過

本次會議於 2010 年 6 月 6 日至 2010 年 6 月 11 日在美國印第安納波里斯舉行，本人於 6 月 5 日早上到達當地。在會期中，除出席會議外，亦積極參與主辦單位所舉辦之 Demo session。在會議結束後，於 6 月 12 日一早搭乘班機離開印第安納波里斯先飛至亞特蘭大訪問明年將短期訪問的喬治亞理工學院電機系，並於 6 月 14 日下午搭機返回台灣。

二、與會心得

SIGMOD 是資料庫研究領域中最頂級的會議之一。我這次有幸參與 SIGMOD 會議，除了報告自己的研究成果外，也盡量多聽各方先進對資料庫研究最新的進展。我最近比較偏重空間資料庫相關的研究，所以在會議 SESSION 中，選擇比較多這方面論文的，如"Advanced Query Processing", Graph Data & Query"，和"Graph Mining"。在聆聽了這些 SIGMOD 論文後，我的想法是(1)現今因為資料庫越變越大，而且許多的資料來源也越來越不穩定，這使得統計的工具變得越來越重要。(2)相對於亞洲別的國家(如香港，新加坡，和印度)，台灣在資料庫研究明顯不足或落後。

三、考察參觀活動(無是項活動者略)

無。

四、建議

在此頂尖的國際會議中，會場上只見少數來自台灣或台裔的學者及研究生，在人數上遠少於來自香港與中國的研究學者，期望國內的教育當局能在政策面上加以鼓勵並提高補助，以增加台灣在學術研究上的重要性。

五、攜回資料名稱及內容

會議論文集。

六、其他

無。

A Power Saving Scheduling Algorithm for Multiple MSSs in Large-Scale IEEE 802.16e Environments

Huai-Sheng Huang and Yu-Jen Chen
Dpt. of Information Management
Chang Gung University,
Kweishan Taoyuan, Taiwan
m9744008@stmail.cgu.edu.tw,
cyr@mail.cgu.edu.tw

Shih-Lin Wu
Dpt. of Computer Science and Imformation Engineering,
Chang Gung University,
Kweishan Taoyuan, Taiwan
slwu@mail.cgu.edu.tw

Jen-Jee Chen
Department of Computer Science,
National Chiao Tung University,
Hsinchu, Taiwan
chencz@cs.nctu.edu.tw

ABSTRACT
Power Saving Class (PSC) is an essential issue on IEEE 802.16-2009. In previous research, many algorithms had been proposed to reduce the consumption of power, but most of them only considered multiple connections in a Mobile Subscriber Station (MSS); in fact, it does not fit in with the situation of real world. On the contrary, others proposed algorithms considering the situation of multiple MSSs with multiple connections; nevertheless, it is difficult to increase the amount of MSSs. In this paper, we propose an efficient algorithm, which refers to both categories and avoids state transitions. When packet size is much smaller or delay bound is more loosening, the result shows that our scheduling algorithm can serve almost double multiple MSSs with multiple connections and still maintain high sleep ratio for energy efficiency.

Categories and Subject Descriptors
C.2.1. [**Computer-Communication Networks**]: Network Architecture and Design – Wireless communication

General Terms
Algorithms, Design, Performance

Keywords
IEEE 802.16-2009; power saving class (PSC); energy saving; multiple MSSs; state transitions

1. INTRODUCTION
IEEE 802.16 Worldwide Interoperability for Microwave Access (WiMAX) is used for the internet of Broadband Wireless Access (BWA), and its characteristics of widespread coverage area in metropolitan areas and high-speed bandwidth are much better than those of personal communication networks we use today. By adopting the structure of Point-to-Multipoint (PMP), a Base Station (BS) can serve several Mobile Subscriber Stations (MSSs) with multiple connections simultaneously, but the scheduling algorithm can not violate each Quality of Service (QoS) demand of connections.

"IWCMC'10, June 28– July 2, 2010, Caen, France. Copyright © 2010 ACM 978-1-4503-0062-9/10/06/...$5.00"

Because the scheduling algorithm is an open issue on the Media Access Control (MAC) layer in the IEEE 802.16-2009 standard (including the specification of IEEE 802.16e) [1], an efficient scheduling algorithm must increase the capacity of BS in order to serve more MSSs in large-scale environments. Besides, the algorithm attempts to make MSSs work longer based on the battery-powered energy. In the standard, there are four QoS mechanisms: Unsolicited Grant Scheme (UGS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS) and Best Effort service (BE) [2]. UGS, like VoIP, requests consistent bandwidth for a guaranteed period of time. Unlike UGS, rtPS, like streaming audio or video, requests variable bandwidth and strict delay bound. nrtPS, like FTP service, provides efficient service of non-real-time traffic with minimum reserved rate. Like web browsing, BE allocates bandwidth only when the bandwidth is not exhausted. In this paper, we focus on UGS applied to multiple connections among multiple MSSs

The standard also establishes three types of Power Saving Class (PSC): type I, type II and type III. Even though the process of each type is different, the goal is identical to less energy consumption for each MSS. In the PSC, there are two states of transceiver within the time unit of frame: on-state and off-state. The continuous frames with on-state are called listening period, and those with off-state are called sleep period. We can find that an association between listening period and sleep period is a periodical pattern called sleep cycle. When starting the sleep period, the transceiver will go into off-state; when ending the sleep period, it will revert back to the on-state. In type I, usually applied to BE and Non-Real-Time Variable Rate (NRT-VR), sleep period will double itself in the next time when listening period has no packet to deliver. When applying to UGS and Real-Time Variable Rate (RT-VR), type II will repeat sleep cycle sequentially; i.e. the sleep cycle has a fixed listening period and sleep period. When the purpose of multicast or management operations is known beforehand, type III only enters the sleep period once. Fig. 1 presents a comparison between these types [3]. Because we focus on the UGS of QoS type, we adopt the Type II of PSC in this paper.

In the real world, a MSS usually has multiple connections simultaneously; furthermore, a BS also serves many MSSs at the same time. The problem is that the BS can only deliver packets to a MSS in a time slot, the unit of minimal time for scheduling. When PSC is adopted, the whole frame, the unit of minimal time for PSC, must awake if a slot wants to deliver a packet. In past research, many algorithms had been proposed to reduce the consumption of power, but most of them only considered multiple connections in a MSS, which does not fit in with real world

situations. On the contrary, others proposed algorithms considering the situation of multiple MSSs with multiple connections; nevertheless, it is difficult to increase the amount of MSSs. In this paper, we propose an efficient algorithm, which refers to both categories and avoids state transitions. The result shows that our scheduling algorithm can serve more multiple MSSs with multiple connections and still maintain high sleep ratio for energy efficiency.

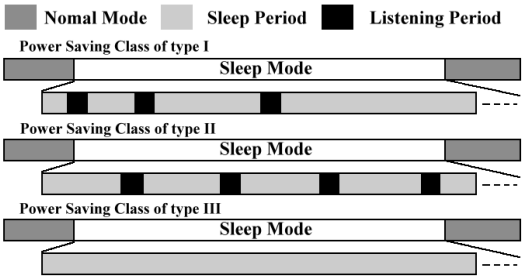


Figure 1. Power saving classes defined in the IEEE 802.16-2009

The remainder of this paper is organized as follows. In section II, we not only review the related work but also classify it into two categories; furthermore, we summarize the problems. In section III, we propose a new scheduling algorithm to match the real world situations. The simulation results are presented in section IV, and in conclusion, we summarize this paper in section V.

2. MOTIVATION AND PROBLEM DEFINITION

We categorize the related work into two categories: the first category concentrates on power saving with multiple connections in only a single MSS, and the second category deals with multiple connections among multiple MSSs. There are some different properties between these two categories, particularly in combining or separating the overlapping frames. In the related work, we find that [4-6] and [10] focused on the first category and [7-9] and [3] emphasized the second one. Both categories consider three basic parameters. The first parameter is packet inter-arrival time representing the interval period, in which two packets are delivered between BS and MSS. In type II of PSC, inter-arrival time must be the same in the sleep cycle repetitively. The second parameter is bandwidth, which means the maximal traffic that BS grants for MSS in a frame. If the number of MSS increases, the bandwidth for each MSS decreases. The third parameter is delay bound, which is used to concern the maximum tolerable period of inter-arrival time. In this paper, we only consider the scenario that inter-arrival time is smaller than delay bound. In fact, it is almost true in the real world.

References [4-5] propose the Maximum Unavailability Interval (MUI), and apply the Chinese Remainder Theorem to select the best start frame for each connection. The purpose of increasing the overlapping number of frames in the sleep period among connections is to decrease the energy consumption of MSS. In addition, [5] proposes Intelligent Table Consulting (ITC) to reduce the computational complexity, but it does not consider the bandwidth and how sleep cycle among numerous connections is scheduled. While [5] only changes the start frame on each connection, the result represents to improve the limited consumption of power. All of [3-10] consider delay bound on scheduling algorithms with a single PSC; more importantly, [6] attempts to consider with multiple PSCs. When the size of packet is smaller than bandwidth, [6] derives the most efficient power saving rate among all other papers; nonetheless, when scheduling, the algorithm of [6] does not consider the state transitions, a phenomenon proposed in [10]. In fact, [10] is the first paper to discuss this new parameter that inherited from [11] to improve Aperiodic on-off Scheme (AS) [3] and Minimum Wakeup Time (MWT) [7] in 802.16. Although the scheduling of [10] can work more efficiently than [3] and [7], sleep cycle changes from periodic to aperiodic. In other words, the scheduling of [10] still causes the issue of state transitions. We consider this new parameter in our algorithm and try to avoid it.

Reference [7] proposes an algorithm separating different transmission time among MSSs, but it is deficient. Reference [8] fixes this problem by adopting Ford-Fulkerson algorithm. Although [8] can work more efficiently than [7], it has the same problem as [10]. Reference [9] proposes a much easier and more efficient algorithm to interleave different MSSs and still maintains each connection with periodic sleep cycle; in other words, each connection among MSSs can still fit in with UGS. Unfortunately, this algorithm brings the unnecessary cost because of additional connection to MSS and does not discuss power saving. Besides, [9] also assumes that each MSS only has one connection, but it is not real world situation. Reference [3] proposes Periodic on-off scheme (PS) and AS and considers to combine not only the packets from different connections but also the pattern from different MSSs under the maximum bandwidth of minimum delay bound. Before [6] and [10] are proposed, [3] is more outstanding in power saving than previous research; nonetheless, [6] and [10] do not consider the situation of multiple MSSs that [3] does. In the simulation result, we compare PS with a new scheduling algorithm called Power Saving Class Management Scheme based on CAGE (PSS-CAGE), which we propose in this paper, with PS. In Fig. 2, we can observe that PS has a limited BS pattern to the minimal MSS pattern; on the contrary, PSS-CAGE has a limited BS pattern to the maximal MSS pattern. For this reason, we can estimate that the number of MSSs, which BS can maintain, increases when the length of MSS patterns is much more variable. For example, the BS pattern in Fig. 2(b) can serve an additional MSS pattern 1 and an additional MSS pattern 2, but that in Fig. 2(a) can only serve an additional MSS pattern 1.

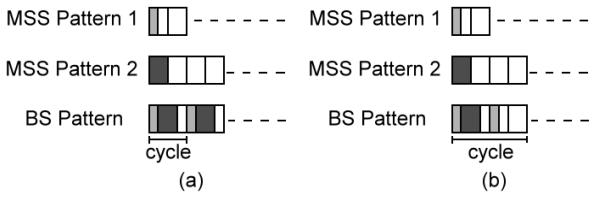


Figure 2. Example of difference between (a) PS and (b) PSS-CAGE

In this paper, we want to combine the issues of two categories. We define that B_f is the maximum bandwidth per frame granted from BS in an OFDM frame for a MSS. We attempt to aggregate potential frames from each connection as more as possible, and still keep each delay bound of connections, as shown in Fig. 3(a), and the maximum bandwidth of MSS pattern under control. In different MSSs, we interleave the overlapping frame by rearranging the start frame of each type II connection but not violate the delay bound. By the research of [9], we can find that if

the sleep cycles of two connections are multiples or factors of each other, the sleep cycles do not have overlap with each other, and they can work together in the same MSS. The same property can apply to two MSS patterns in the same BS. Therefore, we must maintain each MSS pattern to conform to type II in order to interleave with other MSSs as shown in Fig. 3(b).

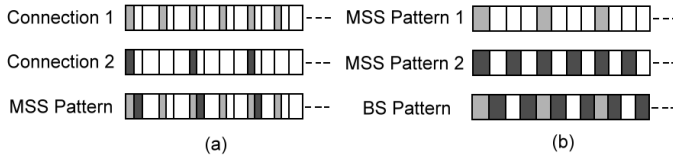


Figure 3. Example of ideas

Because the probability that sleep cycle of connections or MSS pattern becomes the multiple or factor of each other will increase, the bandwidth utilization rate of BS pattern will increase, when the number of connection and MSS increases. On the other hand, in some special cases such as shown in Fig. 4(a), MSS pattern is not a multiple or factor of each other but can still interleave successfully. Nonetheless, in some special cases such as shown in Fig. 4(b), MSS patterns will overlap in some frames. This means that those MSSs can not work simultaneously in the same BS. In order to simplify the situation of Fig. 4(a) and avoid the situation of Fig. 4(b), we can use the relation between multiple and factor to schedule the BS patterns like in Fig. 3(b). Furthermore, we must schedule the connections of each MSS and consider all four parameters, especially state transitions, referred in the previous paragraph.

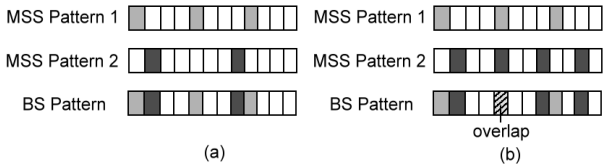


Figure 4. Example of special case

3. PROPOSED SCHEMES

We define that T_i is the length of MSS i, T_s is the minimal length of MSS pattern in BS pattern, and T_m is the maximal length of MSS pattern in BS pattern. M_u, which means how many T_s are in T_m, is that T_m divided by T_s. S_x is the remaining resource in the X^{th} CAGE. TFR_i is the request resource of MSS i. $C_{i,j}$ means connection j of the MSS i. $D_{i,j}$ is the delay bound of $C_{i,j}$. CHECK is used to indicate if the first two different MSS patterns are successfully scheduled in the BS pattern. In the below, we propose an algorithm to check whether T_i can join to the BS pattern.

$$T_i = \frac{2^n}{F} \leq minD_{i,j}$$

$$i = 1 \dots a; j = 1 \dots b; a, b, n \in N \qquad (1)$$

Algorithm 1: Scheduling BS pattern

Input:
Sorting T_i from minimum to maximum;
$T_s = T_1$; $T_m = T_2$; $M_u = T_m / T_s$; $S_1 = (T_1 * B_f) - TFR_1$; $CHECK=1$;
Join MSS pattern 1 to BS pattern;
Output:
Checking whether T_i can join to the BS pattern
For i = 2 to a do
  If $CHECK$ <> 2 and Mu == 1
    If $S_1 > TFR_2$
      $S_1 = S_1 - TFR_i$; Join T_i to the BS pattern;
    Else
      Can't join T_i to the BS pattern;
    End If
  Else If $CHECK$ <> 2 and Mu <> 1
    If $S_1 > TFR_i$
      For k=2 to Mu do
        $S_k = S_1$; $S_1 = S_1 - TFR_i$; $T_s = T_i$; $CHECK$ = 2;
      End For; Join T_i to the BS pattern;
    Else
      Can't join T_i to the BS pattern;
    End If
  Else
    For r = 1 to Mu do
      For z = 1 to ((T_i / T_m) – 1) do
        $S_r = S_{r+z*Mu}$;
      End For
    End For
    $T_m = T_i$;
    If any S_x can contain TFR_j;
      Join T_i to the BS pattern;
    Else
      Can't join T_i to the BS pattern;
    End If
  End If
End For

Because we restrict that T_i is an exponent of 2, as shown in Fig. 5, we can avoid the problem presented in Fig. 4(a) and Fig. 4(b).

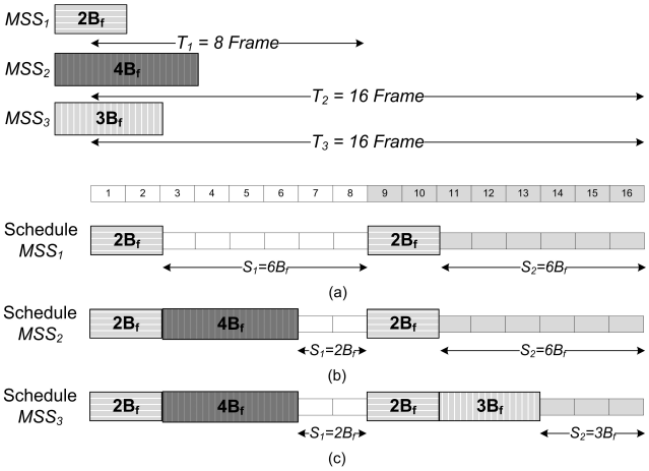


Figure 5. Example of Algorithm 1

The remaining work is how to schedule each connection to a sleep cycle of type II in the same MSS and maintain each pattern of MSS with the exponent of 2. Furthermore, we have to consider the basic three parameters together with state transitions but not violate those rules. We define that $PI_{i,j}$ is the packet inter-arrival time of $C_{i,j}$. The frame duration F is assumed to be an exponent of 2ms. Firstly, we sort $C_{i,j}$ of the same MSS i by delay bound of $C_{i,j}$ like $D_{i,1} \leq D_{i,2} \leq \cdots \leq D_{i,a}$ and then choose the $D_{i,1}$ and $CAGE$, which is the length of sleep cycle for $C_{i,1}$, qualified by (2).

$$2^{\lfloor log_2 D_{i,1} \rfloor} = 2^n \le D_{i,1} < 2^{n+1}$$

$$2 \le CAGE \le \frac{2^n}{F} \quad (2)$$

Secondly, we can set each length of sleep cycle for $C_{i,2...a}$ like $T_{i,j}$.

$$T_{i,j} = T_{i,j-1} \times \left\lfloor \frac{D_{i,j}}{T_{i,j-1} \times F} \right\rfloor \quad (3)$$

Therefore, we can obtain each $T_{i,j}$, a multiple of *CAGE*. Assuming $P_{i,j}$ is the expected packet size of $C_{i,j}$, we can compute the required bandwidth $R_{i,j}$ of $C_{i,j}$ per $T_{i,j}$ by (4).

$$R_{i,j} = \left\lceil \frac{T_{i,j} \times F}{PI_{i,j}} \right\rceil \times P_{i,j} \quad (4)$$

In the next step, Fig. 6 indicates four connections, $C_{i,1}$, $C_{i,2}$, $C_{i,3}$ and $C_{i,4}$, in the same MSS i with $T_{i,1}$=*CAGE*, $T_{i,2}$=2**CAGE*, $T_{i,3}$=2**CAGE* and $T_{i,4}$=4**CAGE* and $R_{i,1} = B_f$, $R_{i,2} = 2B_f$, $R_{i,3} = B_f$ and $R_{i,4} = 2B_f$ when *CAGE* = 4 frames.

We break each $R_{i,j}$ into fractions as $FR_{i,j}$ by (5).

$$FR_{i,j} = \frac{R_{i,j}}{\frac{T_{i,j}}{CAGE}} \quad (5)$$

Then we schedule $FR_{i,j}$ continuously from the first frame. We can observe that each $FR_{i,j}$ is limited in the period of CAGE, just as a CAGE contains all $FR_{i,j}$ following (6).

$$\sum_{j=1}^{a} FR_{i,j} \le CAGE \times B_f \quad (6)$$

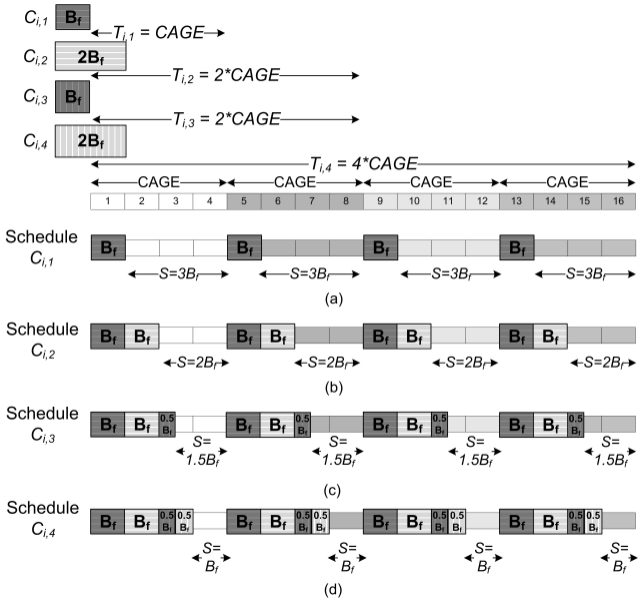


Figure 6. Example of PSS-CAGE

Finally, we can obtain a MSS pattern, which sleep cycle is *CAGE*, the listening period is $\left\lceil \frac{\sum_{j=1}^{a} FR_{i,j}}{B_f} \right\rceil$, and the sleep period is $CAGE - \left\lceil \frac{\sum_{j=1}^{a} FR_{i,j}}{B_f} \right\rceil$. In fact, we can use exhausted search to find every possible *CAGE* from (2) and calculate every possible sleep period. We will choose the best *CAGE* of minimal power consumption. The length of MSS pattern will be the exponent of 2, and we can use Algorithm 1 to check whether this MSS pattern can join to BS pattern.

4. SIMULATION RESULTS

We set up our simulation environment by referring to VoIP parameters and considering only downlink traffic. The length of an OFDM frame is assumed to be 4ms, and each MSS has four connections simultaneously. Our algorithm can still work by multiplying by five in (2) when the duration of frame is 5ms, even though it has slightly decreasing performance. Besides, our scheduling algorithm is able to gain much better performance in the simulation result when assuming that the duration of frame is an exponent of 2ms.

Each packet inter-arrival time of connections is picked between 20ms and 32ms randomly, and delay bound is chosen between 48ms and 200ms randomly. All of the simulation results are averaged from 100 rounds. Three metrics are used to evaluate the performance: total number of MSSs, sleep ratio and successful scheduling rate. The number of MSSs that can work simultaneously in the same BS is counted. The sleep period divided by the sleep cycle is sleep ratio. The successful scheduling rate is the ratio of the number of successful scheduling MSSs to the number of MSSs need to schedule.

4.1 Effects of Packet Size

In this simulation, we not only fix the B_f to 2500 bytes per frame but also randomly set packet size between 20 and 40 bytes per frame by increasing the interval gradually. In Fig. 7, we can observe that the number of MSSs decreases when the packet size increases, and PSS-CAGE can always maintain more MSSs simultaneously than PS. When packet size is much smaller between 20 and 40 bytes per frame, the number of MSSs by PSS-CAGE is almost 100% more than that by PS. This phenomenon happens because PSS-CAGE adopts the pattern calculated by an exponent of 2 with type II for each MSS. Unlike PS that always adopts the same minimal length of MSS pattern for each MSS, PSS-CAGE can use the maximum length of MSS pattern to schedule much more MSSs without overlapping or violating each delay bound. Furthermore, the results shown in Fig. 7 can response to the situation of Fig. 2.

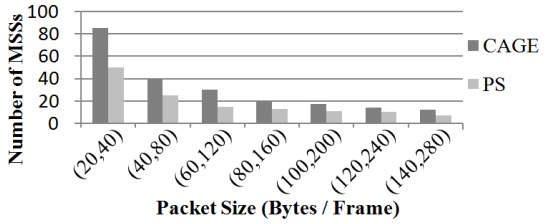


Figure 7. Effects of packet size on the number of MSSs

4.2 Effects of Delay Bound

We fix the B_f to 2500 bytes per frame in this simulation and randomly pick the packet size of connection between 40 and 80 bytes per frame. We gradually loosen the delay bound of connection with a 25 frame increment. In Fig. 8, although PS can contain more MSSs than PSS-CAGE when the delay bound fixes in an integer but not randomly picked from an interval, the gap can be somewhat neglected. The reason is that maximum length of MSS pattern and minimum length of MSS pattern are the same in this situation, and PSS-CAGE can schedule each sleep cycle no more than that of PS. However, each sleep cycle of connection is not usually the same in the real world, so that we simulate some

situations randomly picked the delay bound from an interval. The simulation result presents that when the interval is more loosening, PSS-CAGE can contain more MSSs as a result of bigger gap between the maximum length of MSS pattern and the minimum length of MSS pattern. As the same, the results shown in Fig. 8 can response to the situation of Fig. 2.

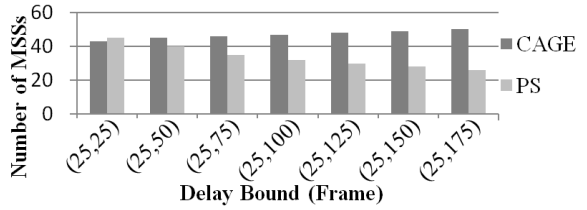


Figure 8. Effects of delay bound on the number of MSSs

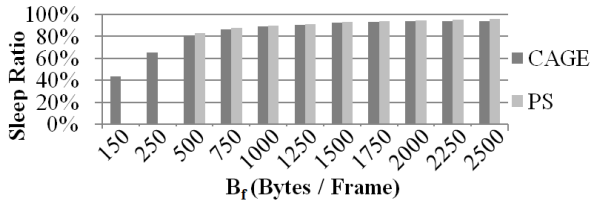


Figure 9. Effects of B_f on the sleep ratio

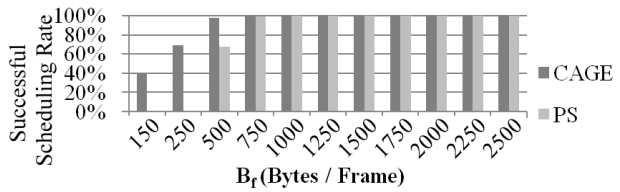


Figure 10. Effects of B_f on the successful scheduling rate

4.3 Effects of Maximum Bandwidth (B_f)

We randomly pick the packet size from 40 to 80 bytes per frame and gradually increased the B_f from 150 to 2500 bytes per frame. Furthermore, we set five MSSs working simultaneously in this simulation. Through Fig. 9, we can observe that PS can not activate the sleep mode when B_f is lower than 500 bytes per frame because the total requested bandwidth of all MSSs exceeds B_f; nevertheless, PSS-CAGE can activate even though the B_f is very small. This phenomenon happens because PSS-CAGE can abandon some MSSs in order to successfully schedule others when B_f is very small; however, PS can only either successfully schedule all MSSs with all connections or schedule none of them. In other words, PSS-CAGE gets better performance than PS when the BS lay in high loading. We can also observe that when B_f is lower than or equal to 500 bytes per frame, PSS-CAGE can successfully and completely schedule as many MSSs into BS as PS after B_f is greater than or equal to 750 bytes in Fig. 10. Although PS still obtains higher sleep ratio when B_f exceeds 500 bytes per frame, PSS-CAGE is very close to PS as shown in Fig. 9. The gap between PS and PSS-CAGE is only 3% and lower than 1% when B_f is greater than or equal to 500 bytes per frame.

5. CONCLUSIONS

In this paper, we propose a new power saving algorithm called PSS-CAGE that can let more MSSs work simultaneously than previous work and still maintain high sleep ratio. Besides, PSS-CAGE also avoids state transitions issue by adopting type II of PSC. Thus, each connection and MSS can maintain a regular pattern and does not violate its delay bound. In other words, PSS-CAGE can decrease additional listening period and try its best to approach the delay bound. According to Fig.7, the result indicates that PSS-CAGE can schedule more MSSs simultaneously than PS, and PSS-CAGE still maintains high sleep ratio as shown in Fig. 9.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Science Council, Taiwan, Republic of China, under grant NSC96-2221-E-182-007-MY3.

7. REFERENCES

[1] IEEE Std 802.16-2009 “IEEE standard for local and metropolitan area networks part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems”, May 2009.

[2] C. So-In, R. Jain, and A.-K. Tamimi, “Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey”, IEEE Journal on Selected Areas in Communications, vol. 27, pp. 156-171, Feb. 2009.

[3] S.-L. Tsao and Y.-L. Chen,, “Energy-efficient packet scheduling algorithms for real-time communications in a mobile WiMAX system”, Computer Communications, vol. 31, pp. 2350-2359, June 2008.

[4] T.-C. Chen, Y.-Y. Chen, and J.-C. Chen, “An Efficient Energy Saving Mechanism for IEEE 802.16e Wireless MANs”, IEEE Transactions on Wireless Communications, vol. 7, pp. 3708-3712, October 2008.

[5] T.-C. Chen, J.-C. Chen and Y.-Y. Chen, “Maximizing unavailability interval for energy saving in IEEE 802.16e wireless MANs”, IEEE Transactions on Mobile Computing, vol. 8, pp. 475-487, April 2009.

[6] J.-J. Chen, S.-L. Wu and S.-W. Wang, “Power Saving Class Management for Energy Saving in IEEE 802.16e Wireless Networks”, in Proc. IEEE Int. Conf. Mobile Data Management, pp.490–495, Taipei, Taiwan, 2009.

[7] S.-C. Huang, R.-H. Jan, and C. Chen “Energy efficient scheduling with QoS guarantee for IEEE 802.16e broadband wireless access networks ”, in Proc. Int. Wireless Communications and Mobile Computing Conf., pp. 547-552, Honolulu, Hawaii, 2007.

[8] S.-C. Huang, C. Chen, R.-H. Jan and C.-C. Hsieh, “An energy-efficient scheduling for multiple MSSs in IEEE 802.16e broadband wireless ”, IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications, pp. 1-5, Cannes, France, 2008.

[9] W.-H. Liao and C. Liu, “An Efficient Scheduling Algorithm for Multiple MSSs”, National Computer Symposium, Taipei, Taiwan, 2009, in press.

[10] L.-H. Yen, C.-Y. Cheng, and C.-H. Wang, “Sleep scheduling that minimizes state transitions for IEEE 802.16e mobile subscriber stations”, Proc. Int. Conf. New Trends in Information and Service Science, pp. 1341-1346, Beijing, China, 2009.

[11] P. Havinga and G. Smit, “Energy-efficient TDMA medium access control protocol”, Asian Int. Mobile Computing Conf., pp. 1-9, Penang, Malaysia, 2000

國科會補助專家學者出席國際會議報告

99 年 7 月 8 日

第六屆ACM國際無線通訊與行動計算研討會 (The 6th International Wireless Communications and Mobile Computing Conference, ACM IWCMC 2010) 在法國巴黎近郊CAEN市的City Conference Center舉行，時間從6月28日到7月2日共計五天。會議主辦單位包含有卡昂大學(Université de Caen Basse-Normandie)與ACM，協辦單位包含有卡昂市政府、飛利浦、法國電信等政府與國際通訊大廠。會議由IBM公司 Dr. Peter Mueller 與 Florida 大學 Prof. Ahmed Helmy 共同主持。此國際會議是ACM組織在通訊網路中知名且重要的會議，目前已有六屆，本次會議出席人員約有五百人。

會議主題包含Broadband Wireless Access, Communication and Information Theory Symposium, Computer and Network Security, Cooperative and Cognitive Networks, Cross-Layer Optimized Wireless Networks, MIMO Systems, Mobile Computing, Multimedia over Wireless , Next Generation Mobile Networks, Vehicular Communication Technology, Wireless LANs and Wireless PANs Symposium, 及 Wireless Sensor Networks。會議包含有45 sessions，將近三百篇的論文發表。另外，會議主辦單位特別邀請三位國際間知名學者分別舉行三場符合時宜且精采專題演講。其中第一天的主題 “The Next Generation of High Performance Computing” 與本人的研究主題 mobile computing 相關，Invited

Speaker 是來自Illinois Urbana-Champaign大學的國際級大師Prof. Bill Gropp，他深入淺出的精湛演說，使我獲益良多。第二天的演講題目是 “On Cyber-Physical Systems Challenges and Research Opportunities”，由Pittsburgh大學的Prof. Tieb Znati 主講，他把未來設計Cyber-Physical Systems所面臨的挑戰以淺顯易懂的方式介紹給大家了解。最後一天則邀請University of Florida著名的學者Prof. Ahmed Helmy 談 “Data-driven Modeling and Design of Networked Mobile Societies: A Paradigm Shift for Future”。從他的演說中，我們可以了解到網路不再是漫無秩序的使用，取而代之是有品質保證的服務。

這次發表的論文被安排在第四天上午Session ThM-1: Energy Awareness and Scheduling in WiMAX Networks中發表，論文的英文題目是 “*A Power Saving Scheduling Algorithm for Multiple MSSs in Large-Scale IEEE 802.16e Environments*”。 本篇論文在 IEEE 802.16e 通訊系統中，提出省電的方法借以改善行動主機的續航力。為了減少行動裝置的電量消耗，IEEE 802.16e 提供三種省電類別 (PSCs)，PSC of type I, PSC of type II 及 PSC of type III，讓行動裝置使用，在每一種省電類別中可包含一個或多個不同的連線。然而，標準中對於不同的連線如何歸屬到相同的省電類別以及省電類別中的相關參數如何設定並沒有加以描述。針對即時性連線 (real time traffic)，先前的文獻中提出了以單一PSC of type II 囊括行動裝置中所有不同的real time flows，進而決定行動裝置的睡醒模式。在本論文中，我們提出兩個在IEEE 802.16e無線網路環境中以多個PSC為基礎並能節省耗電之省電類別管理機制。此兩種方法分別依據最大容許延遲時間 (delay bound) 及封包到達間隔

時間 (packet interarrival time) 來決定PSC的參數。在模擬結果的部份顯示出我們所提出的方法在省電效益及資源利用率方面，均優於先前文獻中所提的方法。會議中許多學者提供了需多不同的看法，我們已將這些寶貴的意見全部紀錄下來以便日後改善時可以參考。

本人此次參加會議收穫頗多，除了論文發表時獲得與會專家學者提供的寶貴建議外，還可以聽到其他學者所發表的論文，並且可以當場向其本人請教心中疑惑。除此之外，session 與 session 之間休息時刻，可與其他專家學者相互討論各種不同領域的議題及未來可能合作的議題，這對我個人來說，更是一項難得的寶貴經驗，也可以趁此機會結交國際間不同領域的優秀學者，方便日後遇到問題時能有更多討論或諮詢的對象，並可使日後之研究更具寬廣性與國際觀。

本次會議攜回資料共三份。第一份為大會會議議程表;第二份是研討會註冊收據; 第三份是本次會議 IWCMC 2010 的會議論文集光碟片一張。其中第三份資料是研究上難得可貴的資訊，可供我與學生做更進一步研究的參考工具。

無研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

計畫主持人：易志偉	計畫編號：98-2218-E-009-023-
計畫名稱：Eco-Community：以感測網路建構一個虛境及實境互動式智慧型社區(1/3)	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	2	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	13	0	100%	人次	
		博士生	5	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	3	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	7	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 ■洽談中 □無

其他：(以 100 字為限)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）(以 500 字為限)

本計畫的預期綜合效益有增進國內研究學者執行整合型計畫的能力與經驗；開發環境監測與控制的先進系統；本系統研究、設計與實作的成果及經驗，可供產官學界未來規劃及開發智慧型生態社區之參考，對國內感測器網路的應用，帶來莫大的助益與經驗；紮實的無線感測網路基礎的研究，將提供多樣化創新應用的基石；互動式平台開發所展示的人機介面可以激發更多的可能，豐富我們的生活；結合各相關領域教授之研究專長，整合設計及開發無線感測器網路之應用；參與此計畫之研究人員得以熟悉無線感測網路、環境監測控制、節能效益評估等相關議題；無線網路應用程式之開發；研究計劃成果可發表學術論文，提昇國內無線通訊之技術與研究水準；參與此一計畫的教授及研究生可以彼此交換研究心得、增進研究成果、並加速學術交流。