

科技部資訊安全技術研發專案計畫

『系統測試計畫書』

System Testing Plan Document

總計畫名稱：雲端物聯網之自動化佈建、動態事件監測、與主動式大
數據服務

子計畫一：雲端物聯網資訊隱密處理與安全資料查詢
(MOST 106-2221-E-011 -003)

子計畫二：雲端物聯網之自動化資料匯集閘道配置
(MOST 106-2221-E-011-002)

子計畫三：雲端物聯網中以出版訂閱為基礎之動態事件監測研究
(MOST 106-2221-E-011-001)

子計畫四：雲端物聯網之行動物件動態大數據分析與主動式服務
(MOST 106-2221-E-259-003)

主 持 人：金台齡副教授（台灣科技大學/資訊工程系）

共同主持人：吳秀陽教授（東華大學/資訊工程系）

共同主持人：邱舉明教授（台灣科技大學/資訊工程系）

共同主持人：項天瑞副教授（台灣科技大學/資訊工程系）

2018/05/11

目 錄

版次變更記錄 (Revision History)

1. 簡介 (Introduction).

1.1 多關鍵字加密查詢系統 (EMAS)

1.1.1 測試範圍 (Scope of Testing)

1.1.2 接受準則 (Acceptance Criteria)

1.2 自動化閘道配置系統 (GDS)

1.2.1 測試範圍 (Scope of Testing)

1.2.2 接受準則 (Acceptance Criteria)

1.3 座標資訊處理子系統(CDPS)

1.3.1 測試範圍 (Scope of Testing)

1.3.2 接受準則 (Acceptance Criteria)

1.4 建置傳輸資料閘道子系統(DTGS)

1.4.1 測試範圍 (Scope of Testing)

1.4.2 接受準則 (Acceptance Criteria)

1.5 建置連通資料閘道子系統(DCGS)

1.5.1 測試範圍 (Scope of Testing)

1.5.2 接受準則 (Acceptance Criteria)

1.6 動態事件監測服務系統 (PSEM)

1.6.1 測試範圍 (Scope of Testing)

1.6.2 接受準則 (Acceptance Criteria)

1.7 子計畫四

1.7.1 測試範圍 (Scope of Testing)

1.7.2 接受準則 (Acceptance Criteria)

2. 測試環境 (Testing Environment)

2.1 多關鍵字加密查詢系統 (EMAS)

2.1.1 操作環境 (Operational Environment)

2.1.2 硬體規格 (Hardware Specification)

2.1.3 軟體規格 (Software Specification)

2.1.4 測試資料來源 (Test Data Sources)

2.2 自動化閘道配置系(GDS)

2.2.1 操作環境 (Operational Environment)

2.2.2 硬體規格 (Hardware Specification)

2.2.3 軟體規格 (Software Specification)

2.2.4 測試資料來源 (Test Data Sources)

2.3 座標資訊處理子系統(CDPS)

- 2.3.1 操作環境 (Operational Environment)
- 2.3.2 硬體規格 (Hardware Specification)
- 2.3.3 軟體規格 (Software Specification)
- 2.4 建置傳輸資料閘道子系統(DTGS)
 - 2.4.1 操作環境 (Operational Environment)
 - 2.4.2 硬體規格 (Hardware Specification)
 - 2.4.3 軟體規格 (Software Specification)
- 2.5 建置連通資料閘道子系統(DCGS)
 - 2.5.1 操作環境 (Operational Environment)
 - 2.5.2 硬體規格 (Hardware Specification)
 - 2.5.3 軟體規格 (Software Specification)
- 2.6 動態事件監測服務系統 (PSEM)
 - 2.6.1 操作環境 (Operational Environment)
 - 2.6.2 硬體規格 (Hardware Specification)
 - 2.6.3 軟體規格 (Software Specification)
- 2.7 子計畫四
 - 2.7.1 操作環境 (Operational Environment)
 - 2.7.2 硬體規格 (Hardware Specification)
 - 2.7.3 軟體規格 (Software Specification)
- 3. 測試時程、程序與責任 (Testing Schedule, Procedure, and Responsibility)
 - 3.1 子計畫一各子系統
 - 3.1.1 測試時程 (Testing Schedule)
 - 3.1.2 測試程序 (Testing Procedure)
 - 3.1.3 整合測試(Integration Testing Cases)
 - 3.1.4 接受測試 (Acceptance Testing)
 - 3.1.5 人員職責分配 (Personnel Responsibilities Assignment)
 - 3.2 子計畫二各子系統
 - 3.2.1 測試時程 (Testing Schedule)
 - 3.2.2 測試程序 (Testing Procedure).
 - 3.2.3 整合測試(Integration Testing Cases)
 - 3.2.4 接受測試 (Acceptance Testing)
 - 3.2.5 人員職責分配 (Personnel Responsibilities Assignment)
 - 3.3 子計畫三各子系統
 - 3.3.1 測試時程 (Testing Schedule)
 - 3.3.2 測試程序 (Testing Procedure)
 - 3.3.3 接受測試 (Acceptance Testing)

3.3.4 人員職責分配 (Personnel Responsibilities Assignment)

3.4 子計畫四系統設計概念與測試方法

4. 測試案例 (Acceptance Testing Cases)

4.1 子計畫一測試案例

4.2 子計畫二測試案例

4.3 子計畫三測試案例

5. 測試案例測試結果與分析 (Test Results and Analysis)

5.1 子計畫一各子系統

5.2 子計畫二各子系統

5.3 子計畫三各子系統

5.4 子計畫四

Appendix A : References

文件版本修正履歷表

編號：

名稱： 雲端物聯網之自動化佈建、動態事件監測、與主動式大數據服務

修訂次序	核准日期	版本	修訂內容
1	2018.05.15	0.1	初版

1 簡介 (Introduction)

隨著現今雲端計算與物聯網技術進步，許多雲端運算服務也逐漸普及。然而這樣的物聯網系統尚有許多關鍵技術待突破，首先在收集資料的過程中，解決閘道設備位置佈建與範圍覆蓋的問題，是確保資料匯集品質的基礎，而在成千上萬的感測裝置中傳遞資料動態監測事件的發生，則有賴於可靠的通訊方法與協定，另外從這些收集來的資料中萃取有用的資訊，需要有效的大數據分析技術，最後讓使用者能夠利用任意裝置於物聯網平台查詢資料，得到精準的服務，則安全與隱私將是重要課題。本計畫將以建構雲端物聯網之自動化佈建、動態事件監測、主動式大數據服務與安全資料查詢等相關應用技術為目標，研究主要目標如下：

子計畫	研究主題	研究及應用發展之價值
1	資訊隱密處理與安全資料查詢	在雲端運算的環境下，數據擁有者將上傳雲端的資料進行加密是必須的，避免私密訊息遭受雲端伺服器第三方服務者的窺探，然而在加密的環境下，如何在不用解密文件的情況下，讓數據使用者以關鍵字查找到雲端資料庫中的資訊，並且在查詢的同時，也保護資料使用者的個人隱私。
2	自動化資料匯集閘道配置	為建立物聯網裝置與雲端伺服器之連結，需考慮靜態與動態的物聯網裝置，以規劃閘道設備的佈署方式。在考慮靜態的物聯網裝置時，基於閘道設備連通的重要性以及設置成本，需探討最少數量閘道設備的設置方式，並確保物聯網裝置皆能與雲端伺服器通訊。
3	以出版訂閱為基礎之動態事件	針對雲端物聯網中以出版/訂閱(publish/subscribe)服務架構為基礎，並且透過建立訂閱者屬性條件間的主宰關係，從中萃取可資利用的優勢，以避免不必要的匹配檢視運作，藉以提供龐大數量之事件監測平台一個高效率的解決方案。
4	行動物件動態大數據分析與主動式服務	基於物聯網的資料量龐大，我們研發共通的行為監測和大數據分析方法，設計開發相關系統模組，提供龐大的計算能力與儲存空間，在面對巨量資料處理時，更能有效利用平行與分散式處理方式，進行位置偵測、軌跡追蹤、移動行為監控、鄰近物件蒐尋、群聚樣式探勘、動態資源共享等。

1.1 多關鍵字加密查詢系統 (EMAS)

隨著雲端運算技術的進步，許多雲端服務也相繼出現，例如 Dropbox 提供的網路儲存服務、Google 資料查詢服務的使用也越來越普及。在使用這些服務的同時，使用者會需要將資料上傳至此類雲端伺服器，然而我們並不知道這類伺服器是不是可以信任，是否會在提供我們服務的同時蒐集我們在雲端伺服器上存放會查詢的資料。本研究在第一年的計畫中，制訂了一套多關鍵字加密查詢系統(EMAS)，目的是要讓雲端伺服器無法得知內部存放的資料及使用者的查詢內容。

本次測試將依據本主系統之下列各子系統進行測試：

- 加密索引建子系統(IE)
- 關鍵字加密子系統(KE)
- 密文查詢子系統(CS)

1.1.1 測試範圍 (Scope of Testing)

本文件主要是描述多關鍵字加密查詢系統的測試計劃。確認在系統整合前，必須先確認所有的設計元件均可正確的輸出，在此我們著重於整合系統測試 (Integration Test) 及接受度測試 (Acceptance Test)。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.1.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則與前一項規定相同。

1.2 自動化資料開道配置系統(GDS)

1.2.1 測試範圍 (Scope of Testing)

本計畫執行項目及預定目標規劃以兩年循序達成。本文件之測試範圍依據此兩年期計畫中第一年所訂定之目標及執行工作項目為基準(表 1)。因各子計畫之整合為本計畫之第兩年之執行項目及目標，故本文件未包含整合測試。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.2.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則與前一項規定相同。

1.3 座標資訊處理子系統(CDPS)

本系統負責接收智慧裝置的位置座標檔案，其座標檔案為 CSV 檔，並將檔案內容轉換為系統所需之格式，以供 DTGS 和 DCGS 運算，以及產生資料開道設置座標檔案。

1.3.1 測試範圍 (Scope of Testing)

本文件主要是使用者群組規劃管理子系統的測試計劃，確認在系統整合前，所有的子系統均可正確的執行，在此我們著重於接受度測試 (Acceptance Test)。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.3.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則

與前一項規定相同。

1.4 建置傳輸資料閘道子系統(DTGS)

根據智慧裝置的座標位置和資料傳輸距離，我們考慮如何設置最少的資料閘道，使智慧裝置都至少能將資訊傳輸至一個資料閘道。

1.4.1 測試範圍 (Scope of Testing)

本文件主要是動態群組服務架構系統的測試計劃，確認在系統整合前，先確認所有的子系統均可正確的執行，在此我們著重於接受度測試 (Acceptance Test)。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.4.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則與前一項規定相同。

1.5 建置連通資料閘道子系統(DCGS)

在物聯網的網路架構中，除了考慮智慧裝置能連結至資料閘道外，資料閘道間也期望能形成一個連通的網路，因此我們除了佈置資料閘道供智慧裝置連結外，也根據已設置資料閘道的位置，設置最少的資料閘道，使得資料閘道能形成一個連通網路。

1.5.1 測試範圍 (Scope of Testing)

本文件主要是使用者群組規劃管理子系統的測試計劃，確認在系統整合前，所有的子系統均可正確的執行，在此我們著重於接受度測試 (Acceptance Test)。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.5.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。

- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則與前一項規定相同。

1.6 動態事件監測服務系統 (PSEM)

動態事件監測服務(Publish/Subscribe-based event monitoring system, PSEM)專案的目的是為了建構一個動態發佈/訂閱系統架構，以持續偵測符合使用者所提出或是有興趣的資料條件之事件的發生情形。並導入主宰觀念，在龐大的訂閱量中，快速且有效率的找出匹配的訂閱者，進而發展有效率的事件監測技術與機制。

本次測試將依據本系統之下列各項子系統，進行測試：

- 同質性訂閱主宰關係建構子系統(DR)
- 事件匹配訂閱子系統(SMP)

1.6.1 測試範圍 (Scope of Testing)

本文件主要是描述動態事件監測服務系統(Publish/Subscribe-based event monitoring system, PSEM)的測試計劃。確認在系統整合前，先確定所有的設計元件均可正確的輸出，在此我們著重於整合系統測試 (Integration Test) 及接受度測試 (Acceptance Test)。本文件內容將依據系統需求規格書與系統設計文件，描述關於整合測試的相關計畫與內容。並希望透過此文件之描述與實踐，達到順利進行測試工作之目的。

1.6.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則與前一項規定相同。

1.7 子計畫四

1.7.1 測試範圍 (Scope of Testing)

在行動物件動態大數據分析中，循序樣式探勘 (Sequential pattern mining) [1][2]是重

要議題，長久以來也都是資料探勘[3]中的一項熱門議題，目的在於挖掘原始資料之間的先後順序中是否具有循序樣式。然而過去的循序樣式探勘研究大多是著重於分析靜態資料庫，近年來因為資料量的增長速度越來越快，資料庫的更新越發頻繁，動態的循序樣式探勘需求日益提升。以往動態的循序樣式探勘[4]-[6]研究，作法包含每次輸入新資料便重新進行探勘，或是每次輸入新資料時便產生候選子序列並使用時間戳記選定動態範圍，也有採取樹狀結構儲存資料避免重覆掃描資料庫...等等。本系統設計基本想法是已經儲存在資料庫的資料和更新的資料相比，在大多數狀況下佔據了絕大部份比例，而這些資料在上一次探勘時已經進行過計算，再次進行計算顯然有些不合乎經濟效益，若能將儲存在資料庫的資料的計算結果保留，只計算更新的資料部分，對於動態樣式探勘的效能應有進一步的改善。

分散式運算的模式開始被人們廣泛應用，大約是從 MapReduce[7]-[11]的概念被提出作為開端，分散式運算是藉由把一件運算工作分成數個較小的部分，送給各個電腦節點進行平行運算。這種做法可以有效的降低使用者的機器負擔和提升工作效率，使得使用者只需要花費較少的成本便能完成龐大的工作量。而 MapReduce 進一步的簡化了分散式運算的工作，因此可以進行大規模資料集的平行運算。同時 MapReduce 具有可靠性高、效率高、擴展性高...等諸多優秀特性，根據 MapReduce 架構實作而成的 Hadoop[12]更是將分散式運算推上新的高峰，現在 Hadoop 依舊佔據了大規模資料分析工具的主流地位，被廣泛的應用在分散式運算。然而在資料量越發龐大，且各項軟硬體速度不斷提升的現在，Hadoop 已經逐漸無法滿足使用者的需求，仰賴硬碟 I/O 的架構成為拖累效率的最大問題。

「國立東華大學資訊工程系普及計算與資料管理實驗室」曾經發表過靜態的循序樣式探勘演算法[13]，並使用 Hadoop 作為分析大規模資料集的叢集架構。本系統承繼以前的研究，針對原先的循序樣式探勘演算法進行改良，將其發展成為一個漸進式的動態循序樣式探勘演算法，並且改善使用 Hadoop 造成的效率問題。本測試即是針對系統之各項效能進行評估，了解系統性能，提供有興趣的使用者參考。GitHub 專案名稱為「國立東華大學資訊工程系 PCDMLab_基於 Spark 架構之漸進式循序樣探勘系統」，網址為 <https://github.com/ndhupcdmlab/project>。

1.7.2 接受準則 (Acceptance Criteria)

本測試計劃需要滿足下列的測試接受準則：

- 本系統需要對所有列為必要(Critical、Important、Desirable)之需求作完整測試。
- 測試程序需要依照本測試計畫所訂定的程序進行，所有測試結果需要能符合預期測試結果方能接受。
- 以測試案例為單位，當測試未通過時，需要進行該單元的測試，其接受的準則

與前一項規定相同。

2 測試環境 (Testing Environment)

2.1 多關鍵字加密查詢系統 (EMAS)

2.1.1 操作環境(Operational Environment)

本系統的測試環境是在 Windows 作業系統，以 Python 語言測試回傳在資料集中文檔案的查詢結果，整體系統架構如下圖(圖 2-1)所示。Data Owner 會對所有的文檔建立加密索引並上傳至 Cloud Server，此為加密索引建立子系統(IE)負責。當 Data User 有搜尋請求時會發送關鍵字至 Data Owner 進行加密，並回傳暗門 (Trapdoor)，此步驟由關鍵字加密子系統(KE)負責。最後 Cloud Server 的密文查詢子系統(CS)收到 Data User 發送的暗門後會進行查詢分數運算，回傳最終排序結果。

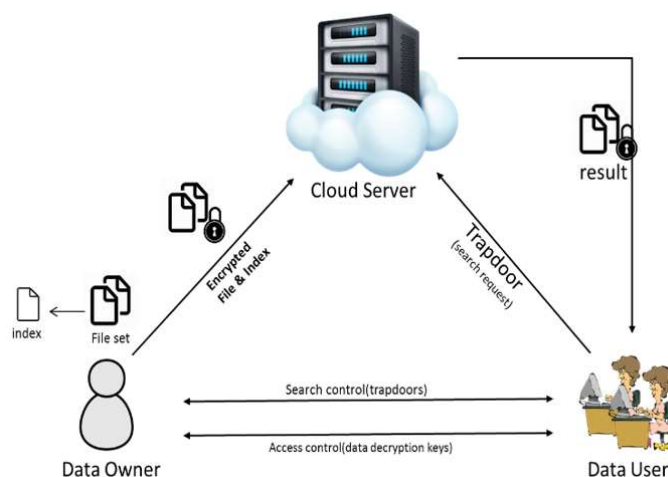


圖 2-1：測試環境架構圖

2.1.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	PC	1	Core i7-6700 16GB RAM

2.1.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 10	1	專業版
2	Anaconda	1	5.1

2.1.4 測試資料來源 (Test Data Sources)

- BBC Dataset(<http://mlg.ucd.ie/datasets/bbc.html>)
- 測試人員撰寫

2.2 自動化資料開道配置系統(GDS)

2.2.1 操作環境 (Operational Environment)

本系統的測試環境是在 Windows 作業系統，以 JAVA 之語言模式測試在固定區域內隨機產生使用者平面座標位置。

2.2.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	Intel Core I7 個人電腦	1	Core i7-4770K 16GB RAM 1 TB SATA HD

2.2.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 8.1	1	專業版 (64bit)
2	Sun Java SE Runtime Environment	1	Version 1.7

2.2.4 測試資料來源 (Test Data Sources)

- 測試人員撰寫

2.3 座標資訊處理子系統(CDPS)

2.3.1 操作環境 (Operational Environment)

本系統的測試環境是在 Windows 作業系統，以 JAVA 之語言模式測試在固定區域內隨機產生使用者平面座標位置

2.3.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	Intel Core I7 個人電腦	1	Core i7-4770K 16GB RAM 1 TB SATA HD

2.3.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 8.1	1	專業版 (64bit)
2	Sun Java SE Runtime Environment	1	Version 1.7

2.3.4 測試資料來源 (Test Data Sources)

- 測試人員撰寫

2.4 建置傳輸資料開道子系統(DTGS)**2.4.1 操作環境 (Operational Environment)**

本系統的測試環境是在 Windows 作業系統，以 JAVA 之語言模式測試在固定區域內隨機產生使用者平面座標位置。

2.4.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	Intel Core I7 個人電腦	1	Core i7-4770K 16GB RAM 1 TB SATA HD

2.4.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 8.1	1	專業版 (64bit)
2	Sun Java SE Runtime Environment	1	Version 1.7
3	lp_solve	1	5.5.2.5

2.4.4 測試資料來源 (Test Data Sources)

- 測試人員撰寫

2.5 建置連通資料開道子系統(DCGS)

2.5.1 操作環境 (Operational Environment)

本系統的測試環境是在 Windows 作業系統，以 JAVA 之語言模式測試在固定區域內隨機產生使用者座標位置。

2.5.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	Intel Core I7 個人電腦	1	Core i7-4770K 16GB RAM 1 TB SATA HD

2.5.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 8.1	1	專業版 (64bit)
2	Sun Java SE Runtime Environment	1	Version 1.7

2.5.4 測試資料來源 (Test Data Sources)

- 測試人員撰寫

2.6 動態事件監測服務系統 (PSEM)

2.6.1 操作環境 (Operational Environment)

對於本系統進行系統測試階段的環境說明，請參考測試環境架構圖(圖 2-2) 所示。本系統的測試環境是在 Windows 作業系統，以 JAVA 之語言測試回傳給訂閱者的匹配結果。

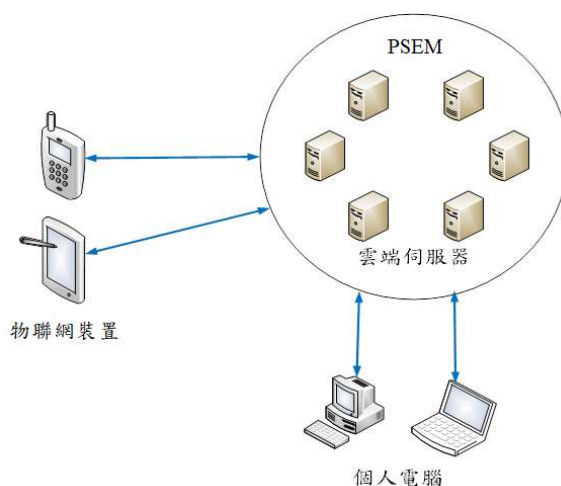


圖 2-2 測試環境架構圖

2.6.2 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	PC	1	Core i5-4440K 16GB RAM

2.6.3 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Windows 7	1	企業版(64bit)

2.6.4 測試資料來源

由資料產生器產生出多個同質性訂閱，並隨機產生多組事件，未來將以實際運作所蒐集到之物聯網裝置微資料來源

2.7 子計畫四

2.7.1 硬體規格 (Hardware Specification)

項目	名稱	數量	規格
1	Cluster	13 PCs	Intel i7 3.3 GHz, 12 cores, 12G RAM

2.7.2 軟體規格 (Software Specification)

項目	名稱	數量	規格
1	Spark	1	2.1.0
2	Hadoop	1	2.7.3
3	Java	1	1.8.0_131
4	Scala	1	2.11

2.7.3 測試資料來源

IBM Quest Synthetic Data Generator 是一個免費的，可以產生人工模擬資料集的工具，以 VS 2008 編譯完成，並運行於 Windows 環境之下。藉由本身提供的參數控制模擬使用者的 transactions，並產生具有關聯性或者循序性的資料集，經常被用來作為大規模資料集的人工資料生產工具，本測試報告亦採用此工具產生實驗測試的數據資料。

(下載來源：<https://sourceforge.net/projects/ibmquestdatagen/>)

ncust	number of transactions
slen	average length per transaction
tlen	average length per item in transaction
nitems	number of different items

3 測試時程、程序與責任 (Testing Schedule, Procedure, and Responsibility)

3.1 子計畫一各子系統：

- 加密索引建子系統(IE)
- 關鍵字加密子系統(KE)
- 密文查詢子系統(CS)

3.1.1 測試時程 (Testing Schedule)

- 加密索引建子系統模組測試(2017/02/01~2017/03/01)
- 關鍵字加密子系統模組測試(2017/02/15~2017/03/15)
- 密文查詢子系統模組測試(2017/03/01~2017/04/01)
- 主系統接受度測試(2017/04/01~2017/05/30)

3.1.2 測試程序 (Testing Procedure)

各子系統開發人員依據接受測試進行並完成測試。若測試結果不符需求或錯誤時，需進行修改後再行測試。

3.1.3 整合測試 (Acceptance Testing)

本系統整合之三階段依序:為 IT1→ IT2→ IT3，如圖 3-1。在 IT1 中，加密索引建立子系統(Index encryption system, IE)會從多關鍵字加密查詢主系統(Efficient Multi-keyword Approximate Search, EMAS)讀取所有的文檔(.txt)，並為所有的文檔建立獨特的加密索引。IT2 中關鍵字加密子系統(Keyword encryption, KE)收到 EMAS 中使用者下達的關鍵字後，會計算並回傳暗門(trapdoor)向量。IT3 則是密文查詢系統(Ciphertext searching system, CS)收到 EMAS 的加密索引及暗門向量後會計算查詢分數並回傳排序結果。

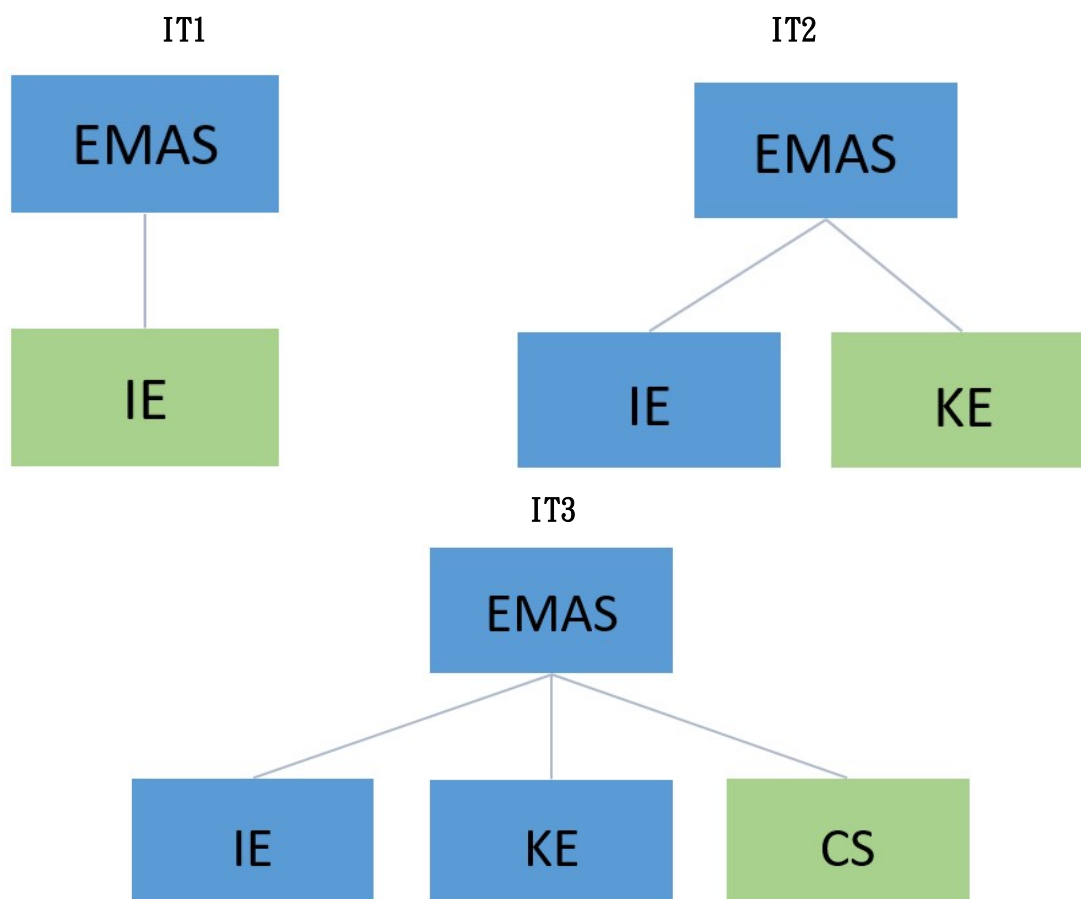


圖 3-1:整合系統測試

3.1.4 接受測試 (Acceptance Testing)

子系統	系統接受測試
加密索引建立子系統(IE)	系統依據資料集中每篇文章檔案的特性建立索引，並產生加密矩陣加以加密。
關鍵字加密子系統(KE)	系統收到用者選擇的關鍵字將以加密形成暗門(trapdoor)，回傳給使用者，作為之後搜尋檔案之依據。
密文查詢子系統(CS)	系統收到使用者的暗門搜尋請求後與加密索引進行搜尋運算，完成後回傳運算分數較高的索引對應之檔案給使用者。

3.1.5 人員職責分配 (Personnel Responsibilities Assignment)

職位名稱	成員姓名	負責任務
專案管理者	金台齡	工作責任的分配和進度的管控。
系統開發者	施宛妮	系統開發撰寫、錯誤排除
系統測試員	施宛妮	執行測試及報告結果

3.2 子計畫二各子系統：

- 座標資訊處理子系統(CDPS)
- 建置傳輸資料閘道子系統(DTGS)
- 建置連通資料閘道子系統(DCGS)

3.2.1 測試時程 (Testing Schedule)

- 各子系統之內部模組測試 (2015/03/01 ~ 2015/05/11)
- 系統接受度測試 (2015/05/11 ~ 2015/05/31)

3.2.2 測試程序 (Testing Procedure)

各子系統開發人員依據接受測試進行並完成測試。若測試結果不符需求或錯誤時，需進行修改後再行測試。

3.2.3 整合測試 (Integration Testing)

自動化資料閘道配置系統之系統整合測試之次序如圖 3-2。在 IT1 中，2.2 自動化資料閘道配置系統(GDS)會先透過座標資訊處理子系統(CDPS)接收智慧裝置的位置座標檔案，並將其檔案內容轉換成建置傳輸資料閘道子系統(DTGS)所需之格式，待座標資訊處理完畢後，DTGS 會根據智慧裝置的位置座標和閘道的通訊範圍，規劃出閘道的設置位置，並將閘道位置提供給建置連通資料閘道子系統(DCGS)，DCGS 會先檢測閘道是否已是一個連通網路，若未連通，則會加入閘道設備，使得閘道能形成一個連通網路。最後，DCGS 會輸出所有的閘道位置。

IT1

IT2

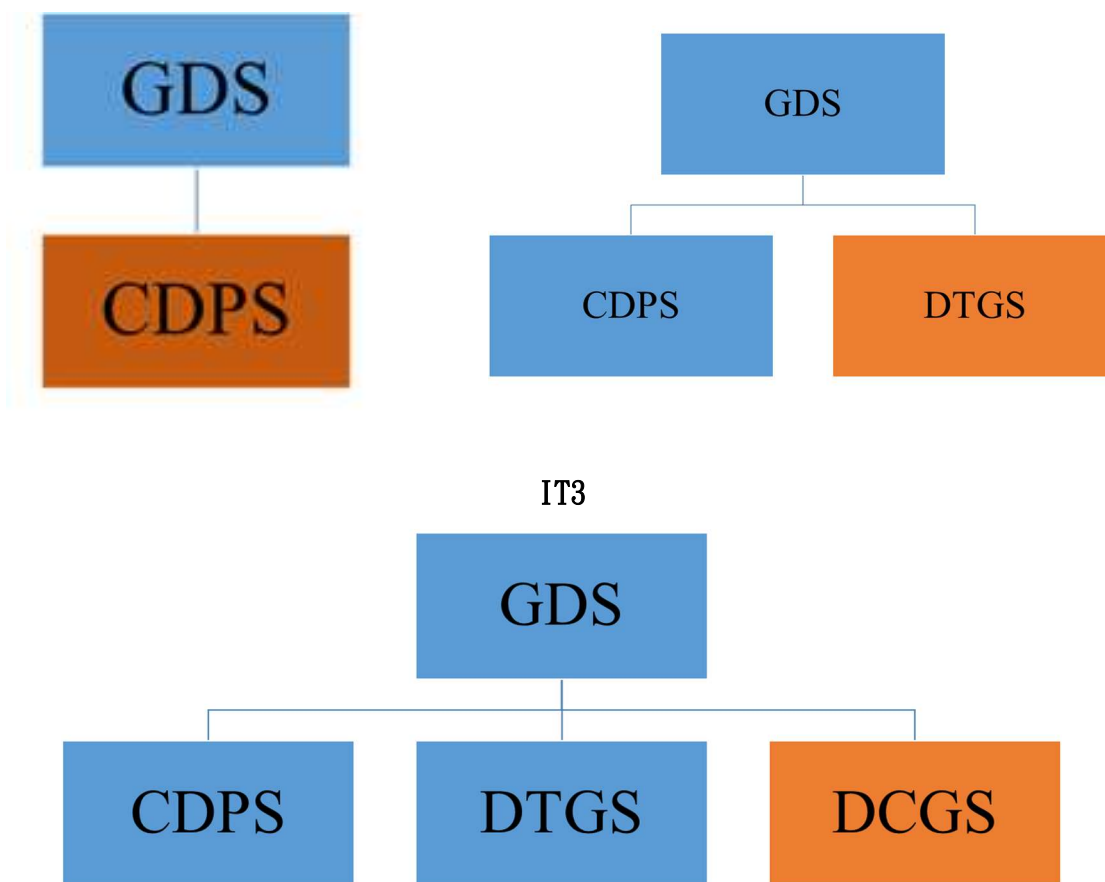


圖 3-2: 自動化資料閘道配置系統之整合系統測試順序

3.2.4 接受測試 (Acceptance Testing)

子系統	系統接受測試
座標資訊處理子系統(CDPS)	負責接收使用者資訊的 CSV 檔，以及產生群組資訊 CSV 檔。
建置傳輸資料閘道子系統 (DTGS)	系統依據智慧裝置之座標資訊，以及通訊範圍，並根據使用者選擇之資料閘道建置策略，提供資料閘道的設置位置，使得智慧裝置皆能至少與一個資料閘道通訊。
建置傳輸資料閘道子系統 (DCGS)	系統根據接收到的資料閘道位置，判斷資料閘道是否連通，若不為一個連通網路，則提供資料閘道的設置位置，以形成連通網路。

3.2.5 人員職責分配 (Personnel Responsibilities Assignment)

職位名稱	成員姓名	負責任務
專案管理者	項天瑞	工作責任的分配和進度的管控。
系統開發者	賴威宇	系統開發撰寫、錯誤排除
系統測試員	賴威宇	執行測試及報告結果

3.3 子計畫三各系統：

3.3.1 測試時程 (Testing Schedule)

- 資料庫模組子系統測試(107/3/23~3/30)
- 同質性訂閱主宰關係建構子系統測試(107/3/23~3/3)
- PSEM 系統整合測試(107/3/23~3/30)
- PSEM 系統接受度測試(107/4/8~5/2)

3.3.2 測試程序 (Testing Procedure)

各子系統之內部元件的整合測試，交由各子系統開發負責人完成。在此我們顯示系統整合測試及系統接受度測試之程序。

3.3.3 整合測試

系統整合測試之次序為：IT1(DR)→ IT2(SMP) (圖 3-3)

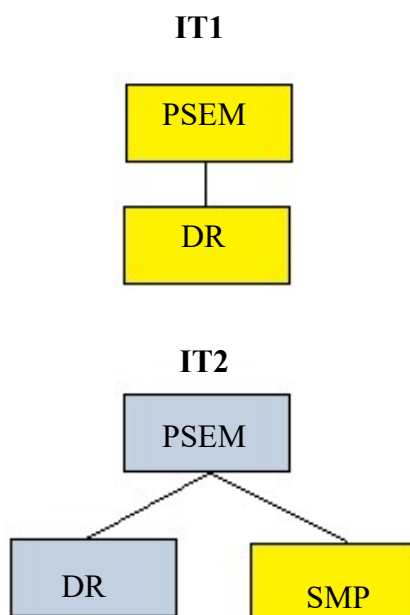


圖 3-3 整合系統測試

3.3.4 接受測試 (Acceptance Testing)

子系統	系統接受測試
同質性訂閱主宰關係建構子系統(DR)	負責接收訂閱者的訂閱資訊，並以訂閱資訊建構訂閱間的主宰關係。
事件匹配訂閱子系統(SMP)	依物聯網裝置提供的事件資訊，有效率的進行訂閱與事件的匹配，並將結果回傳給訂閱者。

3.3.5 人員職責分配 (Personnel Responsibilities Assignment)

成員姓名	負責任務
邱舉明、洪祥仁、 陳君豪	工作責任的分配和進度的管控。
洪祥仁、王云、廖 晨如、羅智晟	系統開發撰寫、錯誤排除
陳君豪、王云、王 群昇	執行測試及報告結果

3.4 子計畫四系統設計概念與測試方法：

本系統提出了一個漸進式循序樣式探勘演算法，可以計算出資料庫的某一個時段的頻繁樣式，並且隨著時間、新資料的輸入，持續進行動態的探勘結果更新，因為是針對一個時段的探勘分析，表示資料會存在時效性。漸進式循序樣式探勘將每一次的探勘資料依照先後順序分成三個區段，過時區段、現存區段和新增區段，過時區段和現存區段的資料是已存於資料庫的資料，在上一次探勘時便已進行過計算，差別在於本次探勘是否依舊處於有效時間。而新增區段是新加入資料庫的資料，未曾進行過計算。

為了解決 Hadoop 仰賴硬碟 I/O 造成的效率問題，本系統嘗試尋找替代 Hadoop 的工具，而 Spark[14]便是一個不錯的方案。Spark 是一個類似於 Hadoop 的開源叢集運算框架，與 Hadoop 每一次完成計算後將資料存入硬碟內的做法不同，Spark 運用了記憶體內的運算技術，資料會在記憶體內部進行運算而不會被寫入硬碟內，只有在需要輸出結果時才會存入硬碟內。因為運算過程中間的 I/O 工作被省略，Spark 官方聲稱，在記憶體內執行程式的速度最快能達到 Hadoop 的 100 倍，即使程式執行於硬碟上，也能達到 10 倍速度。Spark 不僅提供許多豐富的 API，同時也支援 Hadoop，降低了平台轉換的困難度。本系統使用 Scala[15]將漸進式循序樣式探勘演算

法實作完成，最後在 Spark 叢集上進行實驗測試，Scala 是 Spark 的原生語言，是一個簡潔且具有高效率的語言。測試中使用 IBM 提出的資料產生器（IBM Quest Synthetic Data Generator）產生實驗用的人工大規模資料集。我們透過對於資料產生器參數的不同設定，從不同的角度測試系統效能和特性。

4 測試案例 (Test Cases)

4.1 子計畫一測試案例(Test Cases)

4.1.1 加密索引建立子系統(IE)

Identification	AT101-1
Name	Number of documents in dataset: 600, 900, 1350, 1800, 2225
Tested target	IE
Reference	IE-FNR-001、IE-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> ● 資料集文檔(.txt)之文章內容。(ex. Doc1: Verdict delay for Greek sprinters Greek athletics' governing body has postponed by two weeks the judgement on sprinters Costas Kenteris and Katerina Thanou for missing doping ..., Doc2:...) ● 加密向量([0, 1, 1, 0, 0])
Expected result	檔案索引向量集合: Enc1{[15, 21, 14, 32, 19],[30, 13, 28, 11, 15]}
Cleanup	None

Identification	AT101-2
Name	Number of documents in dataset: 1500, 1750, 2000, 2225 The value of r: 250, 500, 750, 1000, 1250
Tested target	IE
Reference	IE-FNR-001、IE-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> ● 資料集文檔(.txt)之文章內容。(ex. Doc1: Verdict delay for Greek sprinters Greek athletics' governing body has postponed by two weeks the judgement on sprinters Costas Kenteris and Katerina Thanou for missing doping ..., Doc2:...) ● 加密向量([0, 1, 1, 0, 0])
Expected result	檔案索引向量集合: Enc1{[15, 21, 14, 32, 19],[30, 13, 28, 11, 15]}
Cleanup	None

4.1.2 關鍵字加密子系統(KE)

Identification	AT102-1
Name	Number of documents in dataset: 600, 900, 1350, 1800, 2225 The value of r: 250, 500, 750, 1000
Tested target	KE
Reference	KE-FNR-001、KE-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> ● 字典內單一關鍵字:1.network 2. classical 3. drugs ● 加密向量([0, 1, 1, 0, 0])
Expected result	暗門(trapdoor)向量集合: 1. Trap1{[10,12,24,23,9],[17,15,31,12,25]} 2. Trap2{[13,25,8,17,9],[41,21,33,12,18]} 3. Trap3{[12,24,23,16,5],[10,8,26,34,12]}
Cleanup	None

Identification	AT102-2
Name	Number of documents in dataset: 600, 900, 1350, 1800, 2225 The value of r: 250, 500, 750, 1000
Tested target	KE
Reference	KE-FNR-001、KE-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> 字典內多重關鍵字:1. network, cloud 2. Michael, Johnson, Olympic 加密向量([0, 1, 1, 0, 0])
Expected result	暗門(trapdoor)向量集合: 1. Trap1{[14,12,15,21,19],[17,15,31,12,25]} 2. Trap2{[2,31,16,10,9],[40,21,13,12,38]}
Cleanup	None

Identification	AT102-3
Name	Number of documents in dataset: 600, 900, 1350, 1800, 2225 The value of r: 250, 500, 750, 1000
Tested target	KE
Reference	KE-FNR-001、KE-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> 字典外單一/多重關鍵字:1. Macaw 2. Neisseria, gonorrhoeae 加密向量([0, 1, 1, 0, 0])
Expected result	錯誤訊息
Cleanup	None

4.1.3 密文查詢子系統(CS)

Identification	AT103-1
Name	The value or r: 250, 500, 750, 1000, 1250 Number of K: 10, 20
Tested target	CS
Reference	CS-FNR-001、CS-TSR-001
Severity	1
Instructions	<ul style="list-style-type: none"> ● Trapdoor 向量 ● 檔案索引向量集合
Expected result	搜尋分數排序
Cleanup	None

Identification	AT103-2
Name	Number of documents in dataset(n): 1500, 1750, 2000, 2225 The value or r: 250, 500, 750, 1000, 1250
Tested target	CS
Reference	CS-FNR-001、CS-TSR-001
Severity	1
Instructions	<ol style="list-style-type: none"> 1. Trapdoor 向量 2. 檔案索引矩陣
Expected result	搜尋分數排序
Cleanup	None

Identification	AT103-3
Name	The value of r: 250, 500, 750, 1000, 1250 Number of keywords in a query: 2, 4, 6, 8, 10
Tested target	CS
Reference	CS-FNR-001、CS-TSR-001
Severity	1
Instructions	3. Trapdoor 向量 4. 檔案索引矩陣
Expected result	搜尋分數排序
Cleanup	None

4.2 子計畫二測試案例(Test Cases)

4.2.1 座標資訊處理子系統(CDPS)

Identification	AT201-1
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	1.讀取座標檔案(.txt)，格式如下：x座標,y座標 2.測試資料 6.2992931667009975,3.147075762618673 8.841319984696701,11.576021043635192
Expected result	Point2D.Double[6.2992931667009975,3.147075762618673] Point2D.Double[8.841319984696701,11.576021043635192]
Cleanup	None

Identification	AT201-2
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	測試資料 8.937523978696222,6.426614895829758 13.867252019259215,7.9231413635184404 21.81716111487423,23.256285466711393
Expected result	座標格式錯誤
Cleanup	None

Identification	AT201-3
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	測試資料 8.937523978696222,6.426614895829758 19.516373508600168 13.867252019259215,7.9231413635184404
Expected result	座標格式錯誤
Cleanup	None

Identification	AT201-4
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	測試資料 8.937523978696222,6.426614895829758 13.867252019259215,abc 21.81716111487423,23.256285466711393 2.6399022786866038,2.2807222334274346 5.97640761676719,10.791440780496014
Expected result	座標格式錯誤
Cleanup	None

Identification	AT201-5
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	測試資料 8.937523978696222,6.426614895829758 13.867252019259215,甲 乙 丙 21.81716111487423,23.256285466711393 2.6399022786866038,2.2807222334274346 5.97640761676719,10.791440780496014
Expected result	座標格式錯誤
Cleanup	None

Identification	AT201-6
Name	讀取座標檔案(.txt)
Tested target	CDPS
Reference	CDPS-FNR-001、CDPS-TSR-001
Severity	1
Instructions	測試資料 6,3 8,11
Expected result	Point2D.Double[6.0, 3.0] Point2D.Double[8.0, 11.0]
Cleanup	None

4.2.2 建置傳輸資料閘道子系統(DTGS)

Identification	AT202-001
Name	number of things: 100, 200, 300, 400, 500 number of sub-region: 1, 4, 9, 16, 25 transmission range of things: 2 transmission range of gateway: 4
Tested target	DTGS
Reference	DTGS-FNR-001、DTGS-TSR-001
Severity	1
Instructions	1. 資料閘道設立位置 2. 資料閘道設立座標 格式如下： Point2D.Double 3. 測試資料 6.2992931667009975,3.147075762618673 8.841319984696701,11.576021043635192
Expected result	資料閘道座標 Point2D.Double[1.099520007205589, 9.090734588308973] Point2D.Double[2.099520007205589, 0.09073458830897163]
Cleanup	None

4.2.3 建置傳輸資料閘道子系統(DCGS)

Identification	AT203-001
Name	number of things: 100, 200, 300, 400, 500 number of sub-region: 1, 4, 9, 16, 25 transmission range of things: 2 transmission range of gateway: 4
Tested target	DCGS
Reference	DCGS-FNR-001、DCGS-TSR-001
Severity	1
Instructions	1. 資料閘道設立位置 2. 資料閘道設立座標 格式如下： Point2D.Double 3. 測試資料 Point2D.Double[1.099520007205589, 9.090734588308973] Point2D.Double[2.099520007205589, 0.09073458830897163]
Expected result	資料閘道座標 Point2D.Double[1.099520007205589, 9.090734588308973] Point2D.Double[2.099520007205589, 0.09073458830897163]
Cleanup	None

4.3 子計畫三測試案例(Test Cases)

4.3.1 IT1 Test Case

Identification	IT1-001
Name	整合 DR 至 PSEM
Tested target	DR、PSEM
Reference	DR-FNR-001、PSEM-FNR-001
Severity	1 (Critical)
Instructions	1. 同質性訂閱主宰關係建構子系統藉由系統所得到的訂閱資訊，建立訂閱之間的主宰關係。
Expected result	正確的訂閱之間主宰關係
Cleanup	

4.3.2 IT2 Test Case

Identification	IT1-002
Name	整合 SMP 至 PSEM
Tested target	SMP、PSEM
Reference	SMP-FNR-001、PSEM-FNR-001
Severity	1 (Critical)
Instructions	1. 同質性訂閱主宰關係建構子系統傳送事件通知給相對應的訂閱者。
Expected result	正確將匹配的事件回傳給訂閱者
Cleanup	

4.3.3 同質性訂閱主宰關係建構子系統(DR)

Identification	AT301-1
Name	Number of subdscriptions(million) : 0.2 , 0.4 , 0.6 , 0.8 , 1 Number of dimensions: 4 , 7 , 10 , 13 , 16 Value space: 50 , 200 , 800 , 3200 , 12800
Tested target	DR
Reference	DR-FNR-001 、 DR-TSR-001
Severity	1 (Critical)
Instructions	1. 訂閱資訊 S = [dimensions d] : {value1, value2...value d}
Expected result	主宰關係的 CDG 結構
Cleanup	

4.3.4 事件匹配訂閱子系統(SMP)

AT301 Test Case:

Identification	AT302-1
Name	Number of Event: 10000 Number of subscriptions (million) : 0.2 , 0.4 , 0.6 , 0.8 , 1 Number of dimensions: 4 , 7 , 10 , 13 , 16 Value space: 50 , 200 , 800 , 3200 , 12800
Tested target	SMP
Reference	SMP-FNR-001 、 SMP-TSR-001
Severity	1 (Critical)
Instructions	1. 訂閱資訊 S = [dimensions d] : {value1, value2...value d} 2. 事件資訊 E = [dimensions d] : {value1, value2...value d}
Expected result	匹配的訂閱
Cleanup	

5 測試結果與分析 (Test Results and Analysis)

5.1 子計畫一各子系統

5.1.1 加密索引建立子系統(IE)

5.1.1.1 Test Results for AT101-1

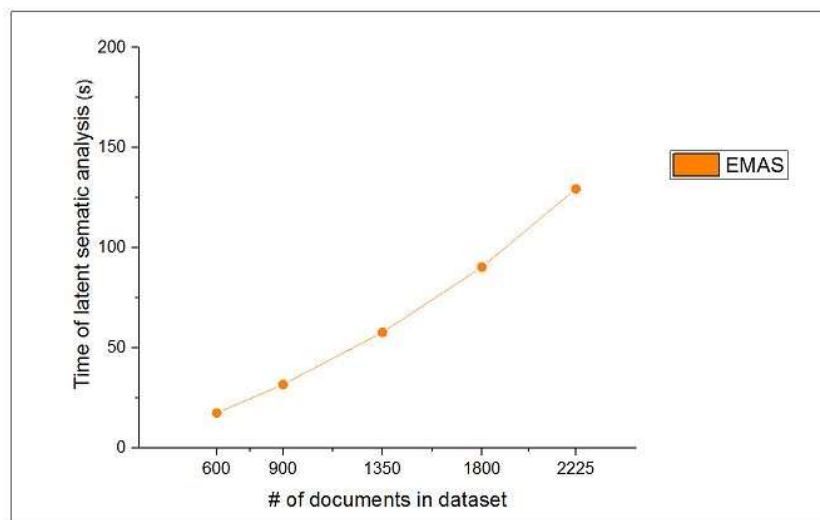


圖 5-1: 建立索引時間

由此實驗可以看出隨著讀取文檔之數量增加，索引建立的時間(Time of latent semantic analysis)也會成長。因為索引的大小取決於資料集中的檔案數和這些檔案中的單字數量。

5.1.1.2 Test Results for AT101-2

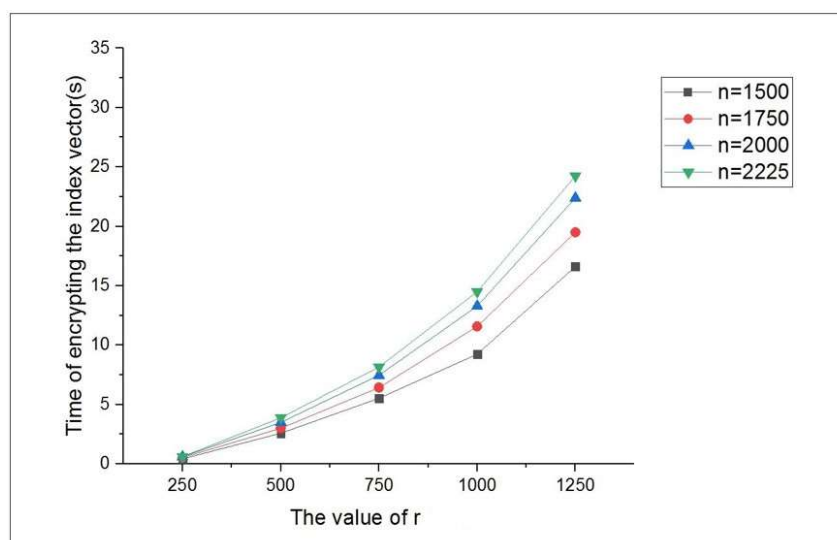


圖 5-2: 加密索引時間

圖 5-2 我們同時比對讀取的資料集數量為 1500、1750、2000、2225 時，所選取的奇異值數量(r)越大時，完成加密索引向量的時間也會跟著增加。因為索引向量加密之時間與向量之長度有關，而索引之長度又與所選的 r 有關聯。

5.1.2 關鍵字加密子系統(KE)

5.1.2.1 Test Results for AT102-1~AT102-2

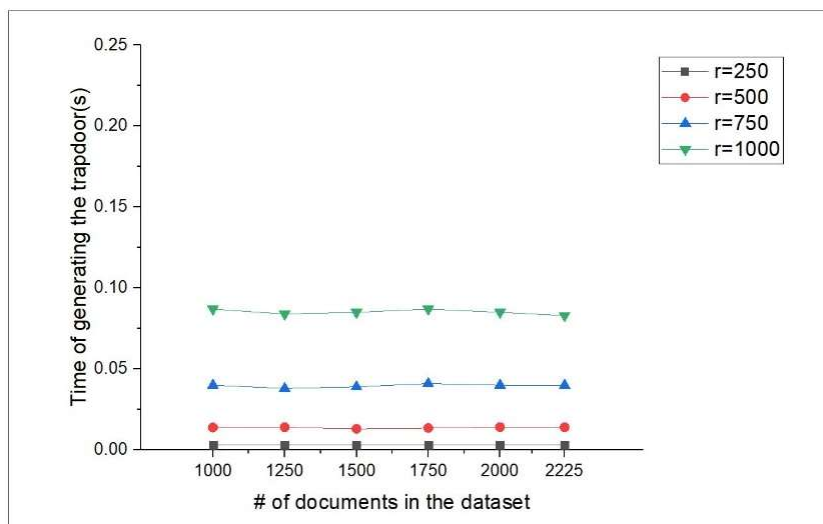


圖 5-3:建立暗門時間

上圖 5.3 同時比較 r 為 250、500、750、1000 時，產生暗門(trapdoor)的時間。可以看出不論資料集中文檔的多寡，皆不會影響暗門產生的時間。因為不論資料集有多少文檔，一旦暗門向量降維成長度 r 之向量，暗門向量的維度皆固定。

5.1.3 密文查詢子系統(CS)

5.1.3.1 Test Results for AT103-1

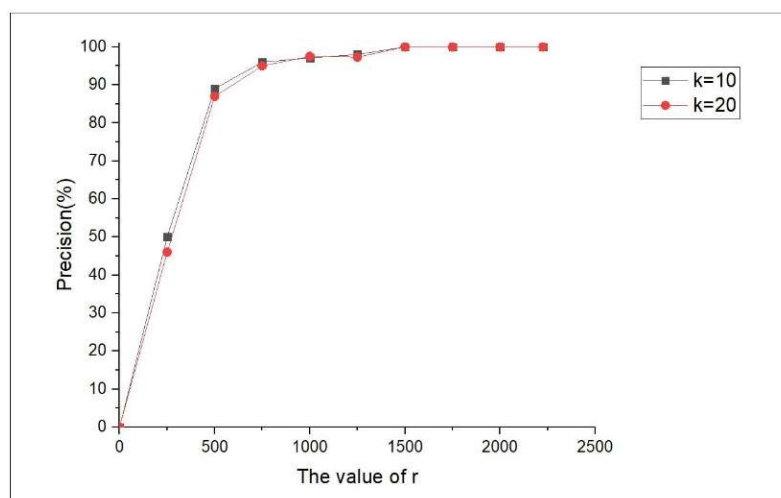


圖 5-4:搜尋精準度時間比較

此實驗為探討在選擇的 r 不同時，對於計算搜尋排序分數前 k 大的精確度之影響。可以看出當 r 大於 500 時，我們同時比對在回傳個數 K 為 10、20 的結果，其精確度皆有 90% 以上的表現。

5.1.3.2 Test Results for AT103-2

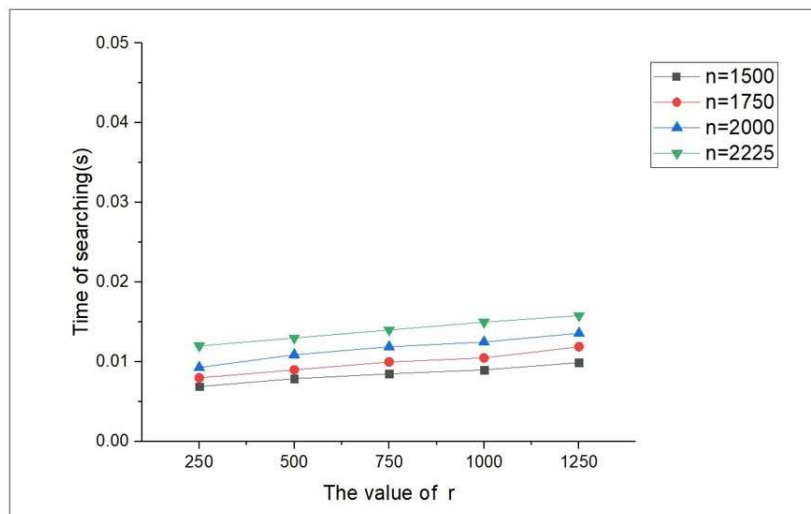


圖 5-5: 奇異值個數影響搜尋時間

本實驗同時比較文檔數為 1500、1750、2000、2225 時，選擇不同的 r 對於計算搜尋分數的時間之影響。可以看出搜尋時間會隨 r 增加呈線性成長。

5.1.3.3 Test Results for AT103-3

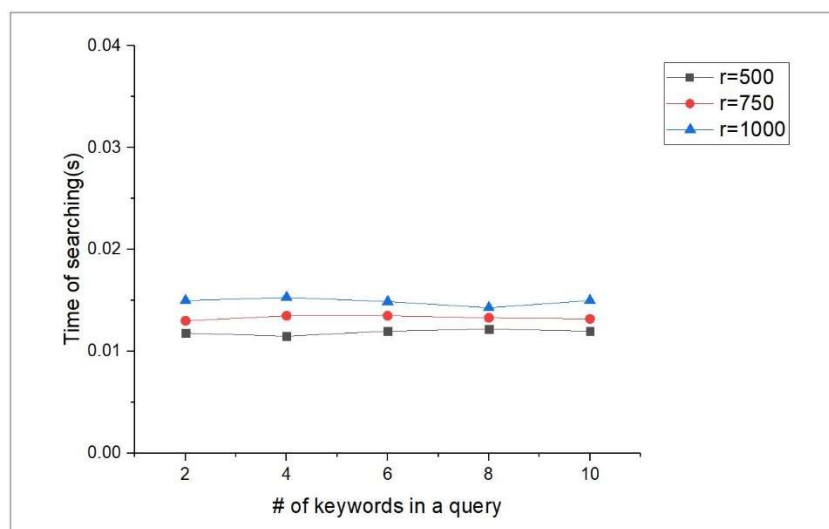


圖 5-6: 關鍵字個數影響搜尋時間圖

本實驗同時比較 r 為 500、750、1000 時，使用者下達的關鍵字數量與搜尋時間的關係。由實驗可以發現關鍵字數量並不影響搜尋時間。因為搜尋時間主要取決於暗門以及索引向量維度的大小。

5.1.3.4 接受測試案例

Test Case #	Results (PASS/FAIL)	Comment
AT101-1	PASS	
AT102-1	PASS	
AT201-1	PASS	
AT301-1	PASS	
AT302-1	PASS	
AT303-1	PASS	

5.2 子計畫二各子系統

5.2.1 自動化資料閘道配置系統(GDS)

5.2.1.1 Test Results for AT202-1:

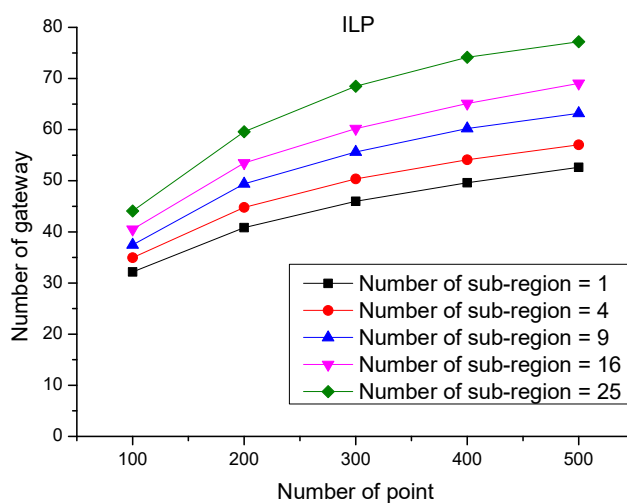


圖 5-7: 以 ILP 規劃部屬負責與智慧裝置通訊之閘道裝置數量

在圖 5-7 中，我們根據不同的子區域數量與智慧裝置數量，比較閘道裝置數量。在子區域數量為 1 時，可視為未將環境分割之狀態，因此閘道裝置的設置數量，如預期般低於將個子區域分別設置的方式。當子區域的分割數量越多，也如預期的設置更多閘道裝置。透過比較閘道數量的設置數量與輸入參數之差異，我們可進一步分析子區域數量對於閘道裝置數量的影響。

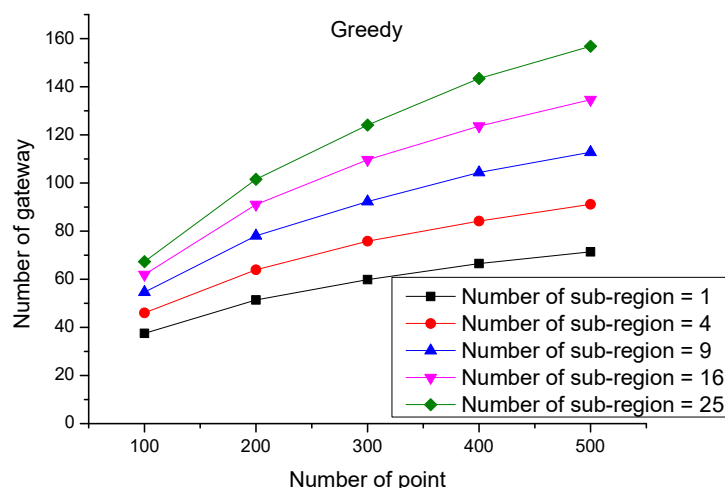


圖 5-8: 以貪婪演算法規劃部屬負責與智慧裝置通訊之閘道裝置數量

除了以 ILP 規劃閘道裝置的設置位置外，我們也比較以貪婪演算法為設置策略產生之結果，如圖 5-8 雖然貪婪演算法的規劃結果不如 ILP，但我們可以先了解兩種不同設置策略之效果差異，並找出效果相似之參數，再分析系統的計算時間，藉此能較快得到一個符合預期效果的規劃結果。

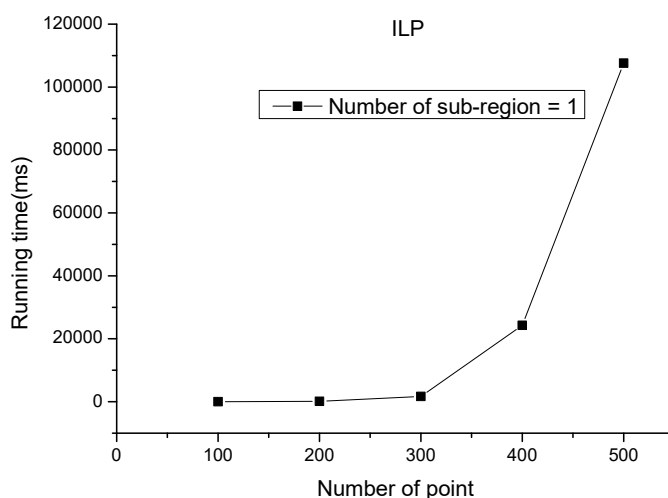


圖 5-9: 未分割環境區域之 ILP 的計算時間

除了比較閘道裝置的設置數量外，我們也比較系統的計算時間。在未分割環境時，圖 5-9 顯示 ILP 對於不同數量智慧裝置的計算時間。雖然，未分割環境能設置最少數量的閘道裝置接至智慧裝置，但我們可以發現系統需要的計算時間，隨著智慧裝置數量增加而大幅成長。

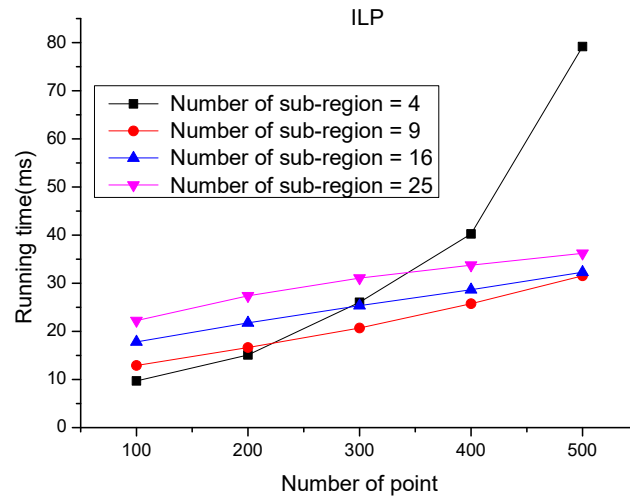


圖 5-10:分割環境區域之 ILP 的計算時間

圖 5-10 顯示在不同子區域的數量下，系統以 ILP 規劃開道設置位置的計算時間。在分割成 4 個子區域，且共有 500 個智慧裝置的情形下，系統需要約 80ms 的計算時間，而其它情形的計算時間不超過 40ms。

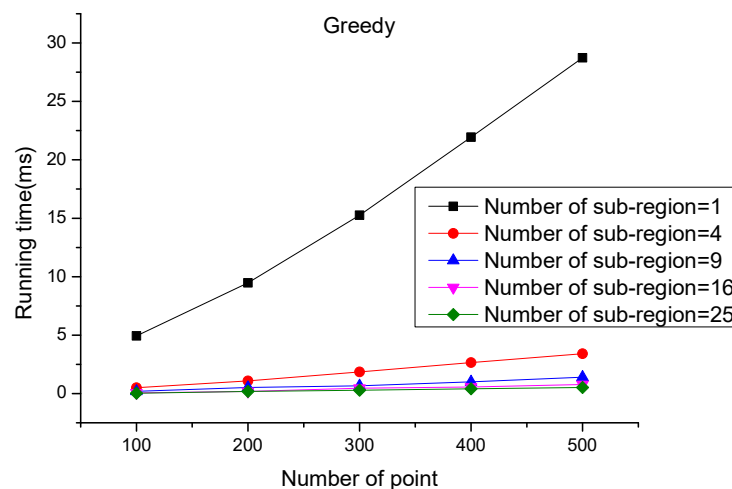


圖 5-11:貪婪演算法的計算時間

在我們的系統中，貪婪演算法的計算時間如圖 5-11。在未分割子區域的情形下，且考慮有 500 個智慧裝置時，貪婪演算法需要約 30ms 的時間即可完成設置，雖然這是圖 5-11 中最長的算時間，但以低於所有 ILP 的計算時間。由上述的統計圖表，我們可以找出貪婪演算法和 ILP 相識

5.2.2 Test Results for AT202-1:

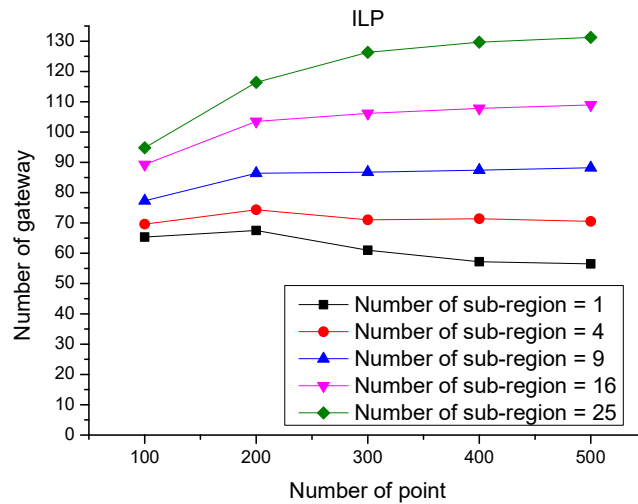


圖 5-12:根據 ILP 的設置結果加入負責建立連通網路之閘道裝置

我們先以 ILP 規劃閘道裝置，再加入閘道裝置以形成連通網路，以圖 5-12 顯示閘道裝置的總數量。當智慧裝置數量較少時，雖然智慧裝置都可連結至閘道裝置，但閘道裝置並不是一個連通的網路，因此會在額外加入閘道裝置。在加入建立連通網路之閘道裝置，我們找出對於兩個最近且未連通的的閘道裝置，並加入閘道裝置使其連通，因此可能我們加入較多的閘道裝置，如圖中未分割子區域的情形。在多數的情形，閘道裝置的數量並未因智慧裝置增加而增加。

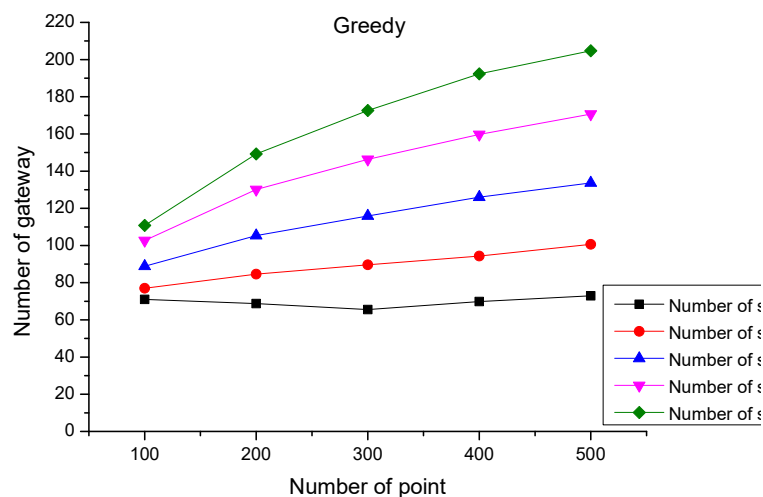


圖 5-13: 根據貪婪演算法的設置結果加入負責建立連通網路之閘道裝置

若先以貪婪演算法規劃閘道裝置，再加入閘道裝置形成連通網路，所需的閘道裝置數量圖 5-13。由圖中我們可以明顯看出，閘道裝置數量會隨智慧裝置數量和子區域分割數量增加。雖然在相同的子區域分割數量和智慧裝置數量的情形下，貪婪演算法的規劃結果不如 ILP，若相同的智慧裝置數量，卻有不同子區域分割數量時貪婪演算法的設置數量和 ILP 相差

無幾。例如，在 500 個智慧裝置時，若貪婪演算法不分割子區域時，則會設置約 72 個閘道裝置，而 ILP 分割為 4 個子區域時，則會設置約 70 個閘道裝置，但貪婪演算法和 ILP 的計算時間分別為 28.73ms 和 79.13ms。

5.3 子計畫三各子系統

5.3.1 整合測試案例 (Integration Testing Cases):

同質性訂閱主宰關係建構子系統可以正確的傳送事件通知給相對應的使用者測試分析：

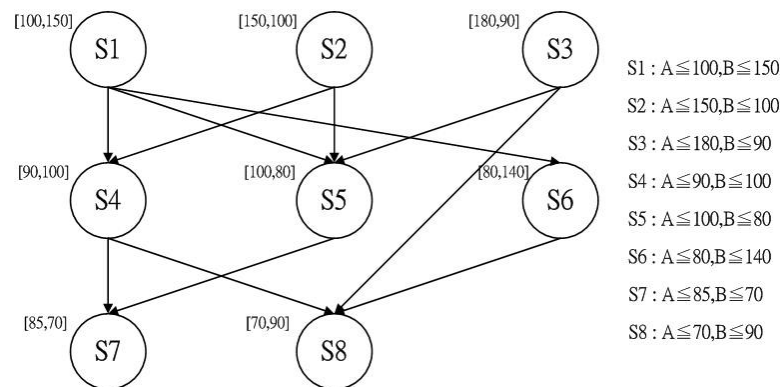


圖 5-14

同質性訂閱主宰關係建構子系統透過設計的演算法，建立訂閱者與訂閱者之間的主宰關係，以一層層的 Layer 建構，並採用廣度優先搜尋法 BFS(Breadth-first search) 對主宰關係結構做搜尋符合事件的訂閱，此一方式得以有效的縮減搜尋時間，避免不必要的搜尋過程。我們主要透過實驗模擬的方式，以實驗結果呈現本建構子系統在不同的資料量下效能的影響，並與現存的 OpIndex 方法做比較。其中主要模擬參數，分別介紹如下：

- N: 訂閱者個數
- d: 維度總數，與訂閱者的謂語總數相同

以下，我們以不同參數對本建構子系統在不同的資料量下，匹配時間的影響。

- N 對於同質性訂閱主宰關係建構子系統之影響

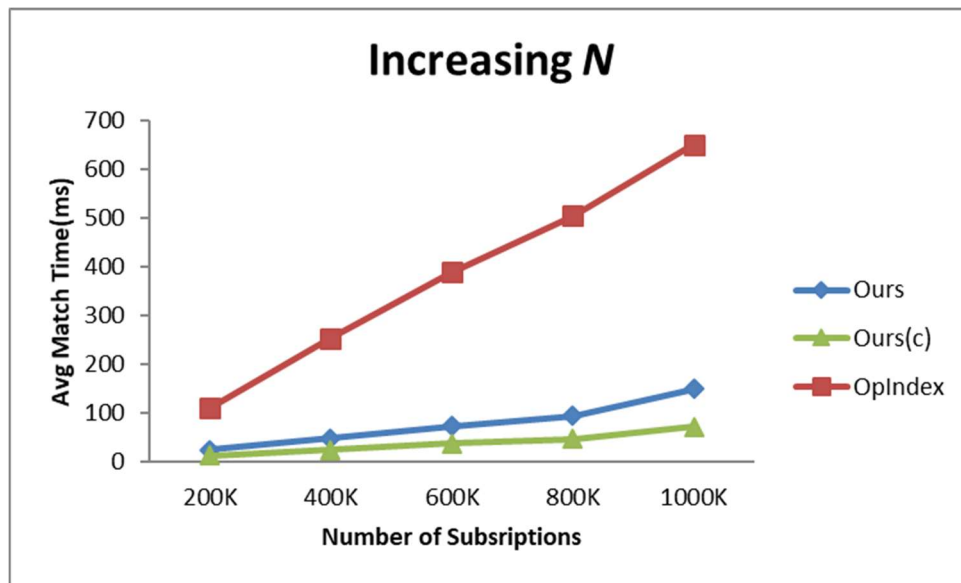


圖 5-15

針對此一實驗，由圖中(圖 5.14)可以看出隨著訂閱者個數的增長，我們的方法(藍線 Ours)平均匹配時間(Avg Match time)的成長趨勢比起現存的 OpIndex 方法(紅線 OpIndex)平緩甚多，另外我們加入 Counter 的概念優化匹配演算法(綠線 Ours(c))能縮短更多的匹配時間，而相關的演算法在結案報告中進一步闡述之。

- d 對於同質性訂閱主宰關係建構子系統之影響

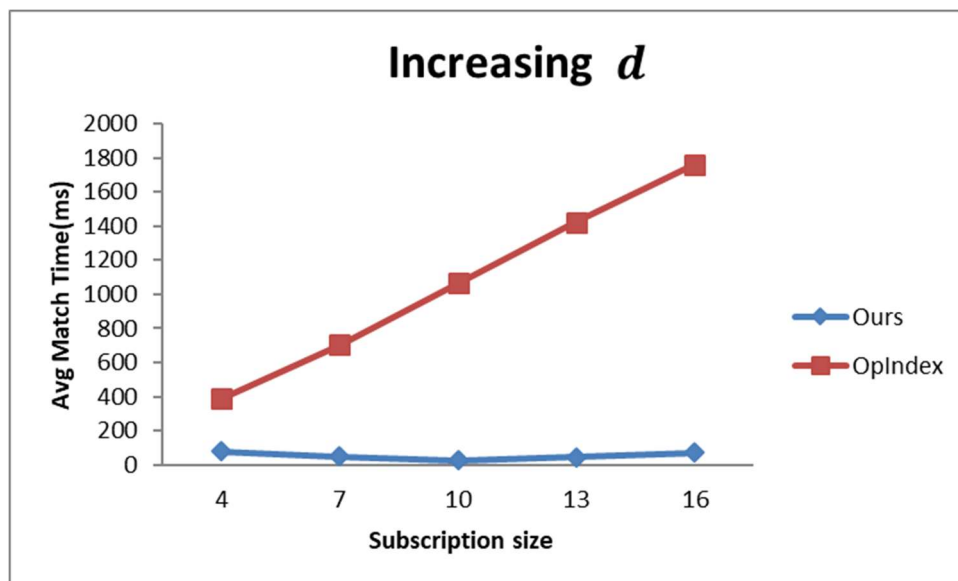


圖 5-16

在此我們固定 0.6M 資料數量做資料庫不同維度總數的評估，如圖(圖 5-16)所示，OpIndex 方法(紅線 OpIndex)因維度 d 的增加而進行匹配的資料量也隨之增加，導致匹配時間增長，而我們設計出的方法(藍線 Ours)並不會因此成長，是因為主宰關係結構對於維度的增加，主宰(dominance)的關係會減少 Layer 層數也會漸少，整個關係結構會變得平坦，並且第一層 Layer 1 的訂閱者會隨著維度總數 d 的擴大，訂閱者的數量會越來越多，正常來說，維度總數 d 的增加，事件匹配到的訂閱者結果會減少，換句話說，訂閱者必須匹配事件的需求條件增加，會較為困難匹配事件 E，所以維度總數 d 從 4 ~ 7 個維度可以觀察到因為維度的增加，造成匹配結果集合的減少，進而降低平均匹配時間(Avg Match Time)，然而，維度總數 d 從 10 ~ 16 個維度開始，平均匹配時間(Avg Match Time)卻開始攀升，原因就在於，雖然匹配結果集合的減少會降低平均匹配時間(Avg Match Time)，但相對的卻也花費龐大的時間在檢查 Layer 1 的訂閱者集合是否匹配事件。

5.3.2 接受測試案例 (Acceptance Testing Cases)

Test Case #	Results (PASS/FAIL)	Comment
AT1	PASS	
AT2	PASS	

5.4 子計畫四

以下針對不同實驗參數的變化，分成各個小節做討論，所有實驗數據圖中的線段名稱皆為演算法名稱加上每一個節點的設置記憶體。

5.4.1 交易量可擴充性測試

本測試討論 data 量增加造成的影響，參數設置參照 Table 1。測資為資料庫內的 transactions，本測試使用 One-Phase 演算法對 unit 進行計算，對於一個 transactions 平均長度為 k，unit size 為 T 的 unit，會產生數量為 $|T| \times (2^k - 1)$ 的子序列。漸進式演算法在計算 window 之前需要將 transactions 依照先後順序切分成不同的 units，因此若是每一個 unit 內的 transactions 數量增加，即 unit size 變大，表示每一次計算 window 時的 data 量增加，則直接計算演算法和漸進式演算法的時間皆會呈現非線性增加。總時間參照圖 5-17 和 Table 2，平均 window 時間參照圖 5-18 和 Table 3，兩個演算法都會呈現非線性成長，當 transactions 數量越大時，兩者的時間差異將會越明顯，漸進式演算法改善的效能越好。

Table 1:交易量可擴充性測試之參數設定

items	window size (#units)	slide (#units)	transaction length	total units
1000	40	10	6	600

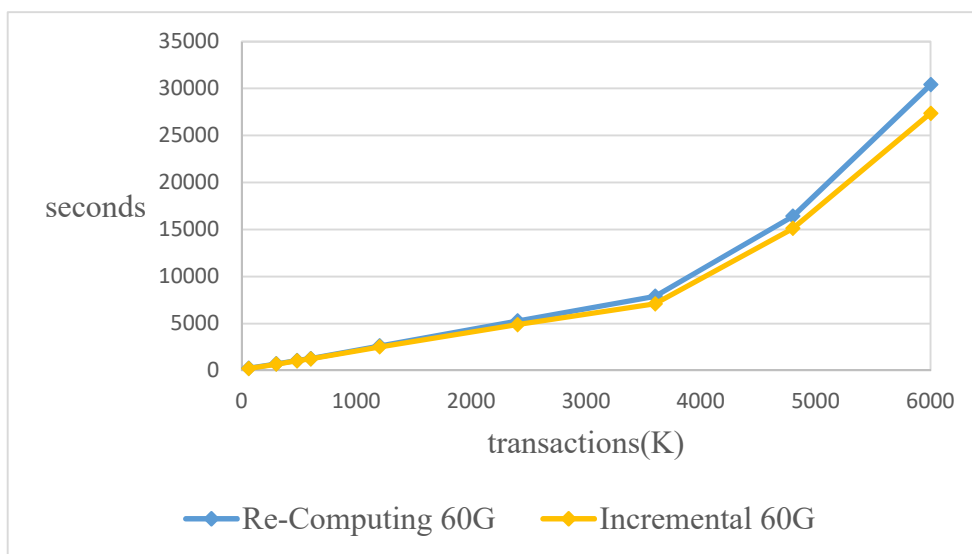


圖 5-17:交易量可擴充性測試之執行時間比較

Table 2: 交易量可擴充性測試之執行時間結果

transactions(K)	unit size (#transactions)	Direct-Computing (seconds)	Incremental (seconds)	Incremental improve
		60G	60G	
300	500	713.81	687.3	3.86%
480	800	1083.06	1043.55	3.79%
600	1000	1285.63	1241.2	3.58%
1200	2000	2626.32	2531.48	3.75%
2400	4000	5295.34	4890.13	8.3%
3600	6000	7901.23	7111.11	11.1%
4800	8000	16419.57	15133.04	8.5%
6000	10000	30411.1	27370.00	11.1%

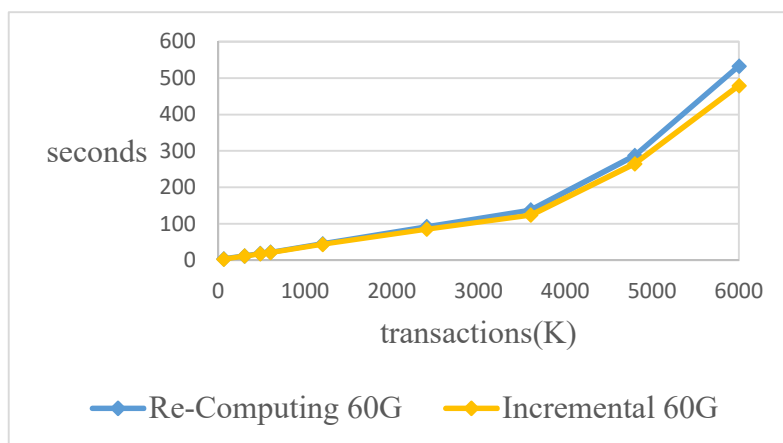


圖 5-18: 交易量可擴充性測試之平均視窗執行時間比較

Table 3:交易量可擴充性測試之平均視窗執行時間結果

transactions(K)	unit size (#transactions)	Direct-Computing (seconds)	Incremental (seconds)
		60G	60G
300	500	12.293	11.83
480	800	18.77	18.08
600	1000	22.32	21.53
1200	2000	45.85	44.19
2400	4000	92.63	85.56
3600	6000	138.40	124.56
4800	8000	287.87	265.31
6000	10000	533.31	479.97

5.4.2 探勘區段大小測試

這一項測試討論 window size 增加造成的影響，參數設置參照 Table 4。Window size 增加表示每一次計算 window 時，計算範圍包含的 units 數量增加。直接計算演算法每一次計算 window 時都需要重新計算整個 window，而漸進式演算法計算 window 時，是依據 slide 值來決定需要計算的 units 範圍，而 slide 值這裡並沒有改變，表示每一次計算 window 的推移 units 數量依然不變，因此 window size 增加會造成直接計算演算法的時間增加，漸進式演算法的時間大致上會持平。總時間參照圖 5-19 和 Table 5，平均 window 時間參照圖 5-20 和 Table 6，直接計算演算法會呈現線性增加，而漸進式演算法不受影響，當 window size 越大時，漸進式演算法改善的效能越好。

Table 4:探勘區段大小測試之參數設定

items	unit size (#transactions)	slide (#units)	transaction length	total units
1000	100	10	4	600

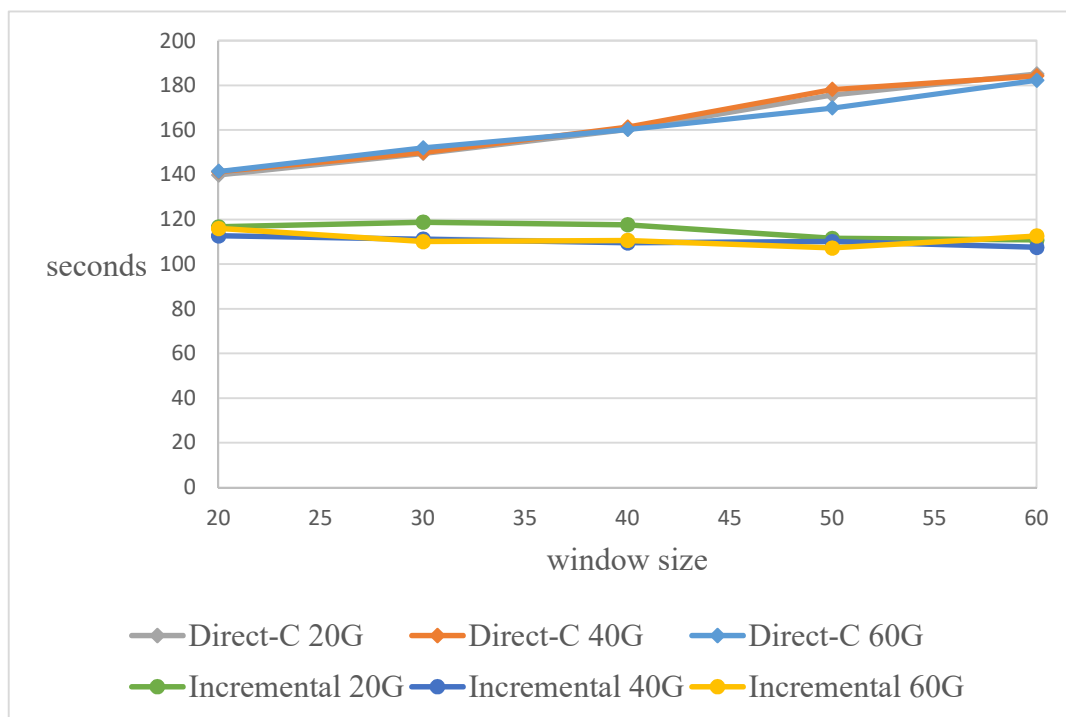


圖 5.19:探勘區段大小測試之執行時間比較

Table 5:探勘區段大小測試之執行時間結果

window size (#units)	Direct-Computing (seconds)			Incremental (seconds)			Incremental improve
	20G	40G	60G	20G	40G	60G	
20	139.88	141.30	141.53	116.70	112.69	115.98	22.4%
30	149.59	149.93	152.06	118.77	111.16	110.06	33.0%
40	160.33	161.36	160.33	117.58	109.53	110.58	42.9%
50	175.70	178.22	169.89	111.49	110.09	107.24	59.3%
60	185.20	184.25	182.35	110.78	107.47	112.57	66.9%

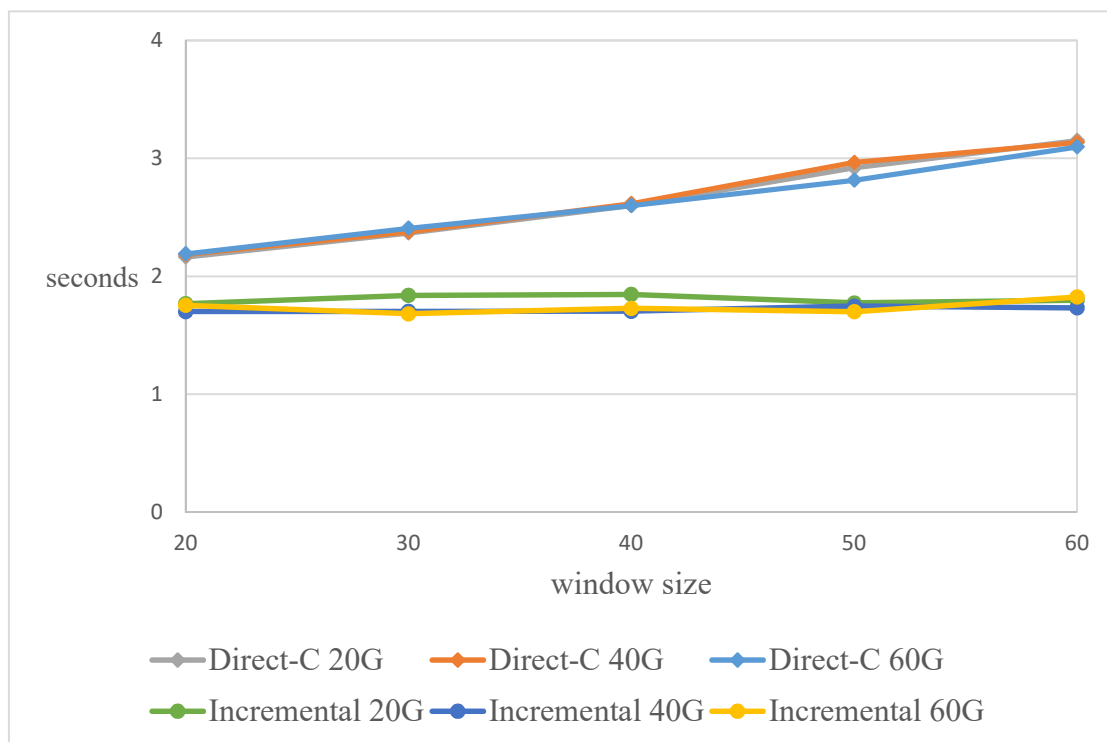


圖 5-20:探勘區段大小測試之平均視窗執行時間比較

Table 6:探勘區段大小測試之平均視窗執行時間結果

window size (#units)	Direct-Computing (seconds)			Incremental (seconds)		
	20G	40G	60G	20G	40G	60G
20	2.16	2.18	2.19	1.77	1.70	1.75
30	2.37	2.38	2.41	1.84	1.70	1.68
40	2.60	2.62	2.60	1.85	1.70	1.73
50	2.92	2.97	2.82	1.77	1.75	1.70
60	3.15	3.14	3.10	1.80	1.73	1.82

5.4.3 漸進式位移量測試

這一項測試討論 slide 值增加造成的影響，參數設置參照 Table 7。Slide 值增加表示在兩次 window 計算之間的 units 推移數量增加，units 數量相同時，slide 值越大相對 window 數量將越少，參照圖 5-21，window 數量的下降幅度相當的大。因為 window 數量大幅下降的關係，直接計算演算法和漸進式演算法的總時間皆會降低，參照圖 5-22 和 Table 8。直接計算演算法在計算 window 時與 slide 值無關，因此平均 window 時間大致上持平，而

漸進式演算法每一次進行 window 計算的 units 將會增加，因此平均 window 時間將會增加，參照圖 5-23 和 Table 9。另外要注意的是當 slide 值越接近 window size 時，漸進式演算法的執行時間反而會超過直接計算演算法，這是因為漸進式演算法需要處理新增的 units 和過時的 units，執行的操作較直接計算演算法來的更多，以本實驗來說，當 window size = 40，大約在 slide = 30 時可以明顯看出這個趨勢，當 slide 較小時，漸進式演算法改善的效率較好，當 slide 值越接近 window size 時，漸進式演算法的效能反而不如直接計算演算法。

Table 7:漸進式位移量測試之參數設定

items	window size (#units)	unit size (#transactions)	transaction length	total units
1000	40	100	4	600

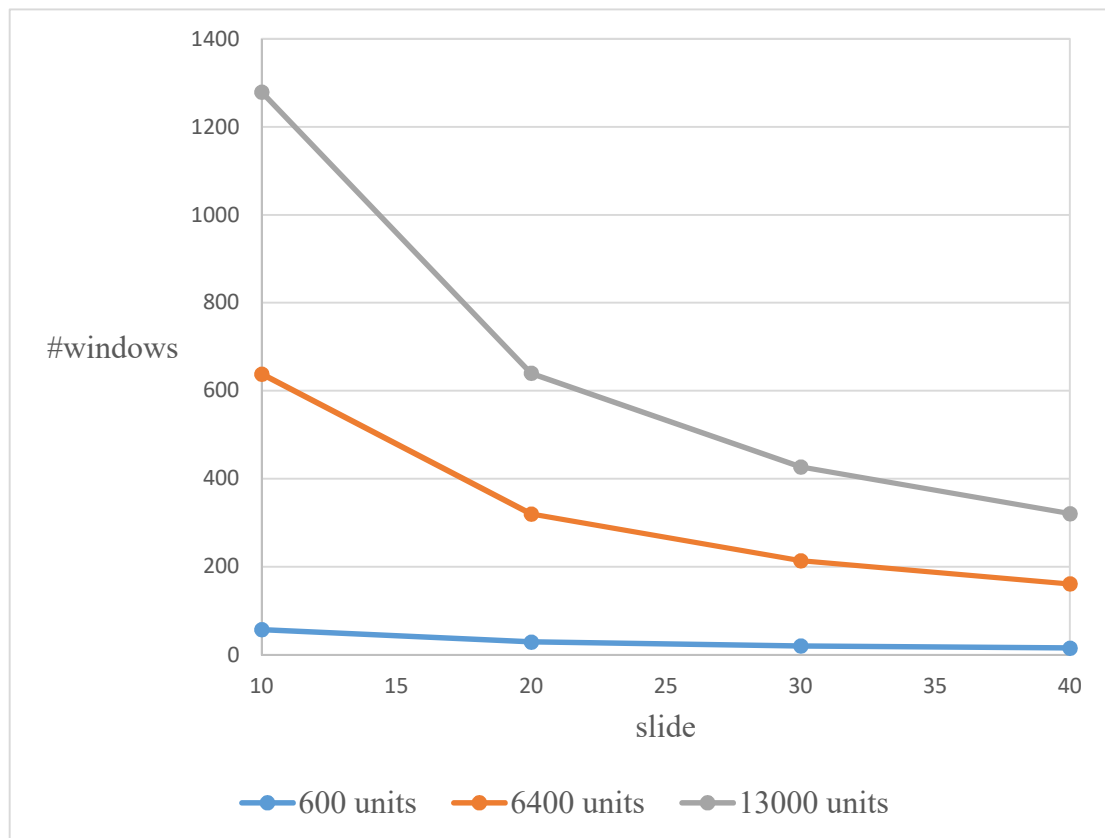


圖 5-21:漸進式位移量與視窗大小之關係

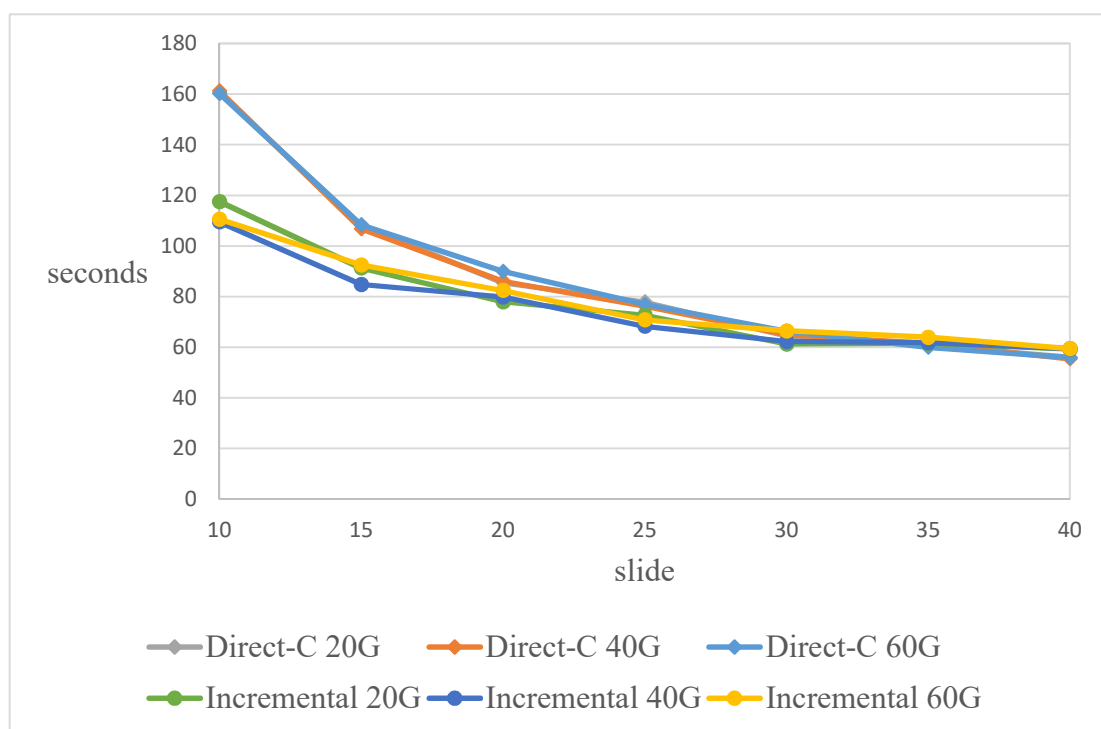


圖 5-22: 漸進式位移量測試之執行時間比較

Table 8: 漸進式位移量測試之執行時間結果

slide (#units)	Direct-Computing (seconds)			Incremental (seconds)			Incremental improve
	20G	40G	60G	20G	40G	60G	
10	160.33	161.36	160.33	117.58	109.53	110.58	42.9%
15	108.41	106.79	108.30	91.28	84.82	92.49	20.6%
20	85.47	86.10	90.01	77.96	79.86	82.48	8.9%
25	77.90	76.25	76.97	72.72	68.30	70.79	9.2%
30	64.42	64.94	66.38	61.28	62.28	66.45	3.1%
35	60.69	61.40	59.92	61.52	61.76	63.94	-3.7%
40	55.92	55.35	55.91	59.26	59.37	59.43	-6.1%

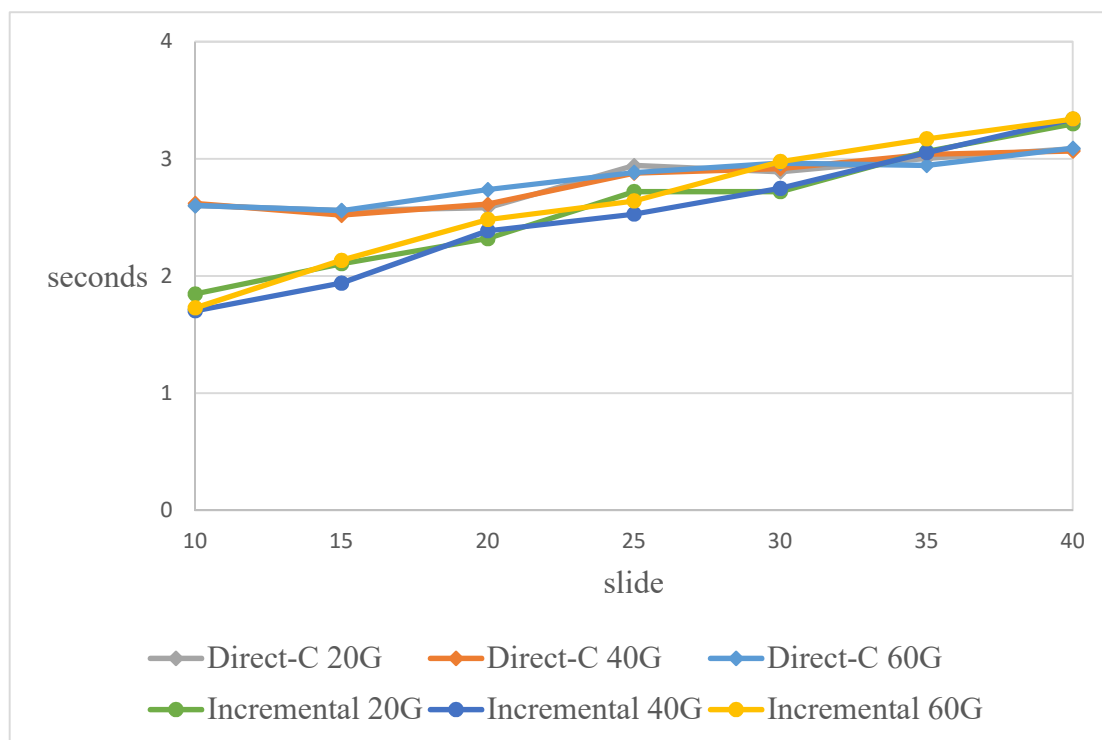


圖 5-23: 漸進式位移量測試之平均視窗執行時間比較

Table 9: 漸進式位移量測試之平均視窗執行時間結果

slide (#units)	Direct-Computing (seconds)			Incremental (seconds)		
	20G	40G	60G	20G	40G	60G
10	2.60	2.62	2.60	1.85	1.70	1.73
15	2.56	2.52	2.56	2.10	1.94	2.13
20	2.58	2.61	2.74	2.32	2.39	2.48
25	2.94	2.88	2.88	2.72	2.53	2.64
30	2.89	2.92	2.96	2.72	2.75	2.98
35	3.00	3.04	2.94	3.06	3.05	3.17
40	3.08	3.07	3.09	3.30	3.34	3.34

Appendix A : References

- [1] R. Agrawal, R. Srikant, “Mining Sequential Patterns”, *Data Engineering, 11th International Conference on*, pp. 3-14, 1995.
- [2] R. Agrawal, R. Srikant., “Fast algorithms for mining association rules”, *Very Large Data Base, 20th International Conference on*, Vol. 1215, pp. 487-499, 1994.
- [3] Ming Syan Chen, Jiawei Han, Philip S. Yu, “Data Mining : an overview from a database perspective”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8(6), pp. 866-883, 1996.
- [4] Hong Cheng, Xifeng Yan, Jiawei Han, “IncSpan :Incremental Mining of Sequential Patterns in Large Database”, *Knowledge Discovery and Data Mining, 10th ACM SIGKDD International Conference on*, pp. 527-532, 2004.
- [5] Ming-Yen Lin, Suh-Yin Lee, “Incremental update on sequential patterns in large databases by implicit merging and efficient counting”, *Information Systems*, Vol. 29(5), pp. 385-404, 2004.
- [6] Florent Masseglia, Pascal Poncelet, Maguelonne Teisseire, “Incremental mining of sequential patterns in large databases”, *Data and Knowledge Engineering*, Vol. 46(1), pp. 97-121, 2003.
- [7] Jeffrey Dean, Sanjay Ghemawat, “MapReduce: Simplified Data Processing on Large Clusters”, *Communications of the ACM*, Vol. 51(1), pp. 107-113, January 2008.
- [8] Xin Yue Yang, Zhen Liu, Yan Fu, “MapReduce as a programming model for association rules algorithm on Hadoop”, *Information Sciences and Interaction Sciences, 3rd International Conference on*, pp. 99-102, 2010.
- [9] Ling juan Li, Min Zhang, “The strategy of Mining Association Rule Based on Cloud Computing”, *Business Computing and Global Information, 2011 IEEE International Conference on*, pp. 475-478, 2011.
- [10] Chun-Chieh Chen, Chi-Yao Tseng, Ming-Syan Chen, “Highly Scalable Sequential Pattern Mining Based on MapReduce on the Cloud”, *Big Data Congress, 2013 IEEE International Congress on*, 2013.
- [11] Zahra Farzanyar, Nick Cercone, “Efficient mining of frequent itemsets in social network data based on MapReduce framework”, *Advances in Social Networks Analysis and Mining, 2013 IEEE/ACM International Conference on*, ACM, 2013.
- [12] The Apache Software Foundation, Apache Hadoop, <https://hadoop.apache.org>, October 2017.
- [13] Yen-hui Liang, Shiow-yang Wu, “Sequence-Growth : A Scalable and Effective Frequent Itemset Mining Algorithm for Big Data Based on MapReduce Framework”, *Big Data, 2015 IEEE International Congress on*, pp. 393-400, 2015.
- [14] The Apache Software Foundation, Apache Spark, <https://spark.apache.org>, October 2017.
- [15] Scala, <https://www.scala-lang.org/>, October 2017.