胡雅涵、翁政雄、楊亞澄（2016），『運用關聯規則及改變探勘技術於防火牆政策規則優化』，*中華民國資訊管理學報*，第二十三卷，第三期，頁277-304。

運用關聯規則及改變探勘技術於防火牆政策規則優化

胡雅涵
國立中正大學資訊管理學系

翁政雄*
中臺科技大學資訊管理學系

楊亞澄
國立中正大學資訊管理學系

摘要

防火牆設備是企業最普遍的網路防護設施，隨著網路環境的改變，防火牆政策規則須不斷的更新，才能維持防火牆功能的正常運作。如何從防火牆日誌記錄中挖掘出有意義的規則，並且適時依據防火牆日誌記錄的變動篩選出不同樣式的規則，進而調整防火牆政策規則是一項有值得研究的議題。本研究嘗試整合關聯規則探勘（Association rule mining）及改變探勘（Change mining）技術，提出Change-Based Association Rule Mining（CBARM）方法。首先，從防火牆日誌記錄中挖掘出有意義的規則，進而運用改變探勘技術辨識出新興樣式（Emerging patterns）、新增樣式（Added patterns）及消失樣式（Perished Patterns）等 3 種不同樣式的關聯規則。最後，將具有不同樣式的關聯規則運用於防火牆政策規則的調整，藉以提升防火牆效率。經實驗結果得知：CBARM 方法效能提升（封包比對次數減少）相較於 Apriori 方法約 95.19%至 582.19%。平均而言，效能約提升 212.10%。

關鍵詞：防火牆政策、防火牆日誌、資料探勘、關聯規則、改變探勘

* 本文通訊作者。電子郵件信箱：107090@ctust.edu.tw
2014/11/14 投稿；2015/05/15 修訂；2015/10/27 接受

Hu, Y.H., Weng, C.H. and Yang, Y. C. (2016), 'Applying association rule and change mining techniques for firewall policy optimization', *Journal of Information Management*, Vol. 23, No. 3, pp. 277-304.

Applying Association Rule and Change Mining Techniques for Firewall Policy Optimization

Ya-Han Hu
Department of Information Management, National Chung Cheng University

Cheng-Hsiung Weng*
Department of Management Information Systems, Central Taiwan University of Science and Technology

Ya-Cheng Yang
Department of Information Management, National Chung Cheng University

Abstract

Purpose－A firewall is the network security system most frequently used by enterprises. Because of changes in the dynamic network environment, firewall policy rules must be constantly updated to maintain efficient firewall operation. Thus, the aim of this study is to optimize firewall policy rules and improve firewall efficiency by using association rules discovered in firewall logs.

Design/methodology/approach－This paper proposes change-based association rule mining (CBARM), which integrates association rule mining and change mining techniques, to discover meaningful firewall policy rules in firewall logs. Specifically, CBARM first determines pertinent association rules by using firewall logs from different time periods. Subsequently, the change mining technique is used to identify emerging, added, and perished patterns. Finally, the three types of patterns can be utilized to optimize the firewall policy rules and enhance firewall efficiency. The firewall logs were collected from a technology company in Central Taiwan. The total number of rules matched in the firewall was used as a performance measure.

* Corresponding author. Email: 107090@ctust.edu.tw
2014/11/14 received; 2015/05/15 revised; 2015/10/27 accepted

Findings ─ The experimental results revealed that the proposed CBARM outperformed the Apriori approach, reducing the number of compared network packets with firewall policy rules by approximately 95.19% to 582.19%. On average, the performance of the proposed CBARM was 212.10% more effective than that of the Apriori approach.

Research limitations/implications ─ This study investigated the firewall logs from one company only. Evaluating the logs from other companies is critical for confirming validity. In addition, future studies can integrate other data mining and machine learning techniques to refine the performance of the proposed method.

Practical implications ─ Two practical implications are provided. First, the association rule mining technique is proven to derive useful firewall policy rules in firewall logs. Second, using the change mining technique can facilitate evaluating the generated rules and applying such rules to optimize firewall policy rules.

Originality/value ─ This study is the first to extend association rule mining and change mining techniques to the domain of firewall log analysis, creating a new approach to optimizing firewall policy rules.

Keywords: firewall policy, firewall log, data mining, association rule, change mining

壹、緒論

網際網路近年快速成長，企業越來越依賴各種網路服務來運作業務及開拓市場，網路逐漸成為企業不可缺少的資源。然而，網路安全事件層出不窮，包含駭客入侵、資料竊取、阻斷服務（DoS）及網路犯罪等，企業面對多變的網路攻擊，資訊管理人員對於資訊安全議題面臨更嚴峻的挑戰。由美國國土安全部轄下的電腦資安單位 U.S. Computer Emergency Readiness Team（US-CERT）統計，影響電腦系統與網路運作的威脅事件已經由 2006 年的 5,503 件成長到 2012 年的 48,562 件，總共成長了 782%（US-GAO 2013）。

越來越多企業或政府單位加強資訊安全設備的強化。依據CSI：Computer Crime and Security Survey 2011 調查所示，在眾多資訊安全機制中，已有 90%以上的企業建置防火牆（firewall）功能（CSI 2011）。在一般企業網路資訊安全設備架構上，最接近網際網路的邊際管制設備通常是防火牆，防火牆是一個用來分隔兩個或以上不同網路的網路安全裝置，透過將網路劃分成不同的區域，防火牆通常裝置於企業內部網路（受信任區域）與外部網路（不受信任區域）之間，讓合法使用者正常取得公開於網路上的資料以及阻擋非法使用者取得尚未公開的機密資料或惡意侵入企業內部網路。然而，僅有防火牆並無法擁有保護的功能，還需搭配防火牆政策規則（policy rule）的設定，將網路劃分成不同的區域，而不同區域的網路溝通，則以政策規則表（policy rule table）為依據，以達到管制網路封包進出的目的（Zwicky et al. 2000）。因此，防火牆防護能力的好壞，管理人員的專業才是重要因素，而防火牆政策規則表設定的優劣，也將決定了系統防護的強度，設備功能與政策規則設定相輔相成，才能發揮最大的防護能力。

防火牆政策規則的研究主要分為兩大主題：入侵偵測與政策規則表優化。其中，入侵偵測研究著重防火牆政策規則的效能（effectiveness），而政策規則表優化研究則著重防火牆政策規則的效率（efficiency）。首先，入侵偵測主要運用網路封包或防火牆日誌記錄來進行分析，以期正確地偵測異常封包（Chang et al. 2007; Feng et al. 2014; Hanguang & Yu 2012）。部份學者透過關聯規則（association rule mining）去分析網路封包，用以檢測企業內部網路是否有網路異常現象，並防止網路攻擊行為（Casado et al. 2006; Lee & Stolfo 1998; Mohammad et al. 2011; Vaarandi 2013），其實驗結果證明關聯規則對於入侵偵測是有效的。此外，亦有部份學者是透過分析防火牆日誌記錄來偵測網路威脅，並提供網路管理者調整政策規則表之參考（Hossain et al. 2012; Saboori et al. 2010）。

其次，如何優化防火牆政策規則表來提升防火牆效率是另一個熱門的研究議題，其主要概念為：如何在不影響現有政策規則表的效能（即判斷正確率）下，

透過規則的調整或比對演算法的設計來提升防火牆封包比對的效率。防火牆日誌記錄中，每一筆記錄都記錄著完整的會談（session）資訊，而每筆會談資訊包含 Session ID、封包數、時間戳記（timestamp）、VPN 類型、來源 IP、來源 Port、目的 IP、目的 Port 及通訊協定等資訊。而防火牆政策規則是採用依序比對上述屬性，當封包比對符合之後就不會再往下繼續比對，並回應預先設定好的回應動作。因此，在防火牆政策規則表中，每條規則的先後次序調整與如何設計阻擋的規則將是影響防火牆效率的關鍵因素（Chang & Chang 2009; El-Atawy et al. 2007; Golnabi et al. 2006; Jeffrey & Samak 2009; Mustafa et al. 2013; Salah et al. 2012; Sreelaja & Pai 2010; Winding et al. 2006）。

本研究之主要目的在於如何透過分析防火牆日誌記錄來優化防火牆政策規則表。近年來，雖然有不少研究嘗試從數量龐大的網路日誌中找出目前所需要的政策規則（Golnabi et al. 2006; Hamed & Al-Shaer 2006; Hossain et al. 2012; Saboori et al. 2010; Winding et al. 2006）。這些研究主要的概念乃是運用關聯規則探勘技術產生新的防火牆政策規則。然而，僅運用關聯規則探勘技術將面臨兩個問題：(1)網路封包的存取常會隨著時間而變動，倘若關聯規則未能隨網路封包存取的變化而適當調整，將無法有效阻擋攻擊。例如：殭屍網路來源位址可能改變了，而關聯規則未能即時調整。(2)運用關聯規則探勘技術產生的新關聯規則可能與現有規則差異非常大，若這些規則皆必須套用於正在運作的防火牆規則表上，將耗費更多的時間去核對與驗證規則是否有遺漏，這些反覆調整規則或驗證的步驟，都將成為資安管理人員的另一種負擔。

為了解決上述問題，本研究嘗試整合關聯規則探勘及改變探勘（change mining）技術，提出 change-based association rule mining（CBARM）方法，從防火牆日誌記錄中挖掘與辨識出新興樣式（emerging patterns）、新增樣式（added patterns）及消失樣式（perished patterns）等三種不同樣式的關聯規則。最後，將具有不同樣式的關聯規則運用於防火牆政策規則的調整，藉以動態調整防火牆政策規則並進而提升防火牆效率。

貳、文獻探討

一、關聯規則與改變探勘

關聯規則探勘技術是一項重要的資料探勘技術，可以找出資料間的交互關係，例如：購買者消費某些商品時，同時也會採買其他某一商品的可能性。現今關聯規則探勘已經廣泛應用在商業的預測及決策支援上（翁政雄 2011; 鄭麗珍 & 李麗美 2014; Ahn 2012; Kamsu-Foguem et al. 2013）。除此之外，國內亦有許多學者提出新的關聯規則演算法，用以增進探勘效能（黃仁鵬 & 藍國誠 2007；李瑞庭

等 2012；龔旭陽等 2010）。

改變探勘的目的乃是挖掘兩個不同時間點資料集的改變（差異性），而關聯規則相關的改變探勘乃是針對兩個不同時間點資料集所挖掘得到的關聯規則進行比較。因此，關聯規則相關的改變探勘研究可以粗略分為下列幾種類型：非預期結果改變（unexpected consequent changes）、非預期條件改變（unexpected condition changes）、新興樣式（emerging patterns）、新增樣式（added pattern）及消失樣式（perished pattern）等（Bailey et al. 2003; Chen et al. 2005; Dong & Li 1999; Song et al. 2001）。

改變探勘的研究已經應用於諸多領域，例如：醫學（Wang et al. 2010; Park et al. 2010; Sherhod et al. 2012; Huang et al. 2013）、商業流程（Li et al. 2011; Dam & Ghose 2015）、消費者行為（Song et al. 2001; Chen et al. 2005; Huang 2012）。由於新興樣式乃是某樣式的支持度在現今資料集明顯高於先前資料集，成為資料集中值得注意的式。因此，大部分的研究皆聚焦於新興樣式，新興樣式的相關文獻，如表 1 所示。

表 1：新興樣式文獻整理

研究	樣式	領域
Li 與 Wong（2002）	新興樣式	醫學
Wang 等（2010）	新興樣式	醫學
Park 等（2010）	新興樣式	醫學
Sherhod 等（2012）	新興樣式	醫學
Huang 等（2013）	新興樣式	醫學
Kim 等（2005）	新興樣式	消費者行為
Tsai 與 Shieh（2009）	新興樣式	消費者行為
Shie 等（2013）	新興樣式	消費者行為
Li 等（2015）	新興樣式	消費者行為
Shih 等（2010）	新興樣式	專利

二、防火牆政策規則優化

防火牆政策規則優化的研究可依其是否使用防火牆日誌記錄進行分析而分為兩類。首先，許多過去研究是透過既有的防火牆政策規則進行優化（意即未使用防火牆日誌記錄分析），由於管理者不斷地進行防火牆政策規則調整，長期下來造成規則之間的異常（anomaly）或冗餘（redundancy）情形，進而造成防火牆政策

無法正常執行與效率低落。因此，針對現有防火牆政策規則，許多學者進行相關演算法的設計與開發來自動化識別出異常或冗餘規則。Al-Shaer 與 Hamed（2003）視防火牆規則為政策樹（policy tree）並定義其模型，將防火牆政策逐條透過政策樹的比對定義出錯誤偵測的四種異常狀態，並且使用 JAVA 語言開發出一套名為 Firewall Policy Advisor 的工具。Al-Shaer 與 Hamed（2004）將此演算法成功套用於多部防火牆架構的網路環境上。Yuan 等（2006）以二元決策圖（binary decision diagrams; BDDs）為基礎發展 FIREMAN 演算法，將錯誤偵測規則模型強化，能檢測更廣泛的規則錯誤偵測。Katic 與 Pale（2007）提出概略性的日誌分析方法，透過異常規則的整合及刪除，調整防火牆政策規則，用以提升防火牆效率。然而，此方法僅提醒管理者如何設定防火牆政策規則，避免政策規則的設計錯誤。Liu 等（2008）提出一個針對防火牆政策規則表壓縮的演算法，其研究結果可達到 52.3% 的壓縮比。Abdulmohsin（2009）設計一套 ACL optimization（ACLO）的演算法，透過舊有規則的分析出每一條規則的 IP 範圍，接著透過刪除異常規則（含重複或冗餘等），並且彙整合併規則，重複此步驟直到無法合併為止。該研究運用最少的規則達到控制目的（將存取控制規則表最佳化），成功降低封包過濾的運算延遲。Hu 等（2012）則運用分割技術的演算法發展出 FAME 管理工具，相較於 Firewall Policy Advisor 可提升 70%效能及消除冗餘規則的錯誤。

其次，近年來有部份研究運用防火牆日誌記錄進行分析，嘗試運用資料探勘技術來找出隱含於防火牆日誌記錄之規則，進而提供管理者做為加入或更新防火牆政策規則之參考，以提升防火牆之效率。Golnabi 等（2006）運用關聯規則演算法針對防火牆日誌記錄做分析，該研究從防火牆日誌使用頻率挖掘出關聯規則，並且經過濾規則一般化演算法（filtering rule generalization）將產生的關聯規則整合為較符合為防火牆政策規則表的設定規則，最後經過錯誤偵測來排除這些規則的異常。Rao 等（2011）使用漸進式更新關聯規則（incremental association rule mining）技術來處理新增的日誌紀錄，進而產生即時性的關聯規則給管理者。Lubna 等（2013）實作出一個運用關聯規則技術分析日誌紀錄的系統，可以動態調整防火牆政策規則以解決規則異常偵測問題。

三、改變探勘與防火牆政策規則優化

綜合上節所述，過去防火牆政策規則優化研究均沒有考慮到關聯規則會隨著時間變化（即資料集差異）而造成改變。Ganti 等（1999）認為應用資料探勘技術於兩個不同的資料集合的分析進而評估所產生的結果差異與變化，是一個非常重要的議題。近年來已有不少學者研究使用改變探勘技術來探討商業行為的研究，用以瞭解客戶在大型資料庫中的潛在變化，以提早因應來服務客戶，進而擴大既

有客戶群，並防止客戶流失等（Böttcher et al. 2009; Wu et al. 2005）。Ceci 等（2008）將新興樣式應用於網路封包檢測上而非防火牆政策規則的調整。

由上述的文獻歸納得知：許多學者致力於防火牆政策規則的研究，例如：利用關聯規則探勘技術產生新的防火牆政策規則。然而，即時反應防火牆日誌記錄資料集差異，進而隨時調整防火牆政策規則，仍是值得研究的議題。為了解決上述問題，本研究將整合關聯規則與改變探勘，並且運用新興樣式、新增樣式及消失樣式等三種不同樣式，用於判斷防火牆政策規則的新增、刪除及規則順序調整的動作，做為調整防火牆政策規則之參考，期望防火牆系統持續維持較佳狀態。

參、研究方法

本研究所提出 CBARM 方法流程如圖 1 所示，主要可以分為四個步驟，依序為：資料前處理、關聯規則探勘、改變探勘及防火牆政策規則評估。以下將詳細說明各個步驟的流程與任務。

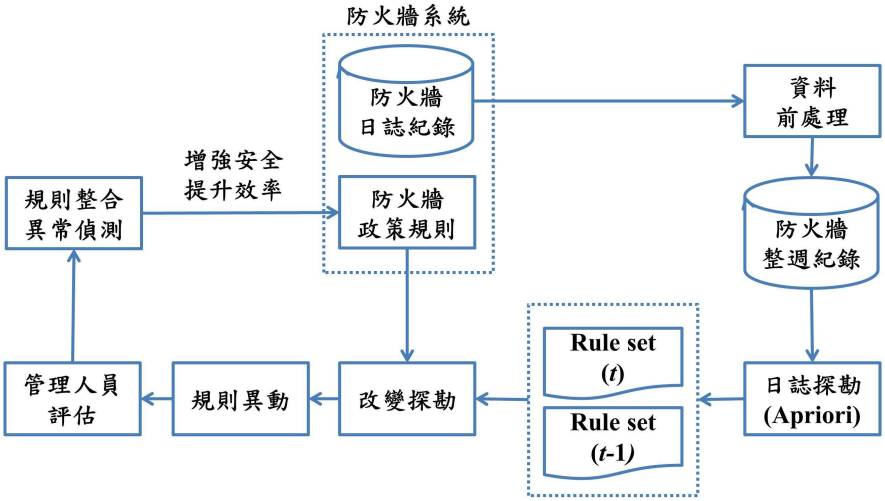

圖 1：CBARM 流程圖

一、資料前處理

本研究從防火牆日誌紀錄中萃取出所需分析的資料欄位，分別是：日期（Date）、時間（Time）、通訊協定（Protocol）、來源位址（Source IP Address）、來源通訊埠（Source Port）、目的位址（Destination IP Address）、目的通訊埠（Destination Port）及狀態（Status）等八項屬性資料。其欄位範例資料，如表 2。

表 2、防火牆日誌記錄資料欄位範例

屬性	值
Date	2013-09-30
Time	05:48:31
Protocol	6
Source IP	192.168.1.23
Source Port	5829
Destination IP	208.91.112.79
Destination Port	443
Status	allow

防火牆系統流量日誌中，通常每一筆資料以一個會談為單位，即兩個位址之間的連線通道，例如撥電話給對方，電話接通時表示此會談建立，接通後的溝通表示封包開始傳送，而掛斷電話則表示此會談結束，而一次會談可視為一筆記錄。本研究將防火牆日誌記錄檔存放於遠端 FTP 伺服器上，並以 Comma-Separated Values（CSV）純文字格式檔案儲存資料。除此之外，由於防火牆系統儲存的時間是以筆數來做匯出儲存，所以無法將每日資料彙整成一個檔案，所以藉由資料前處理先將每份 CSV 檔彙整至資料庫存放，接著匯出一週的日誌記錄由當週第一天（星期日）0 時 0 分 0 秒開始至當週最後一天（星期六）23 時 59 分 59 秒的檔案（如圖 2 所示），並彙整每週的防火牆日誌記錄，以利後續關聯規則探勘之進行。因此，資料前處理步驟中，本研究先透過自動匯出的功能，將防火牆日誌記錄儲存在檔案伺服器或資料庫上，再將原始資料整理成為關聯規則探勘所需的資料格式。

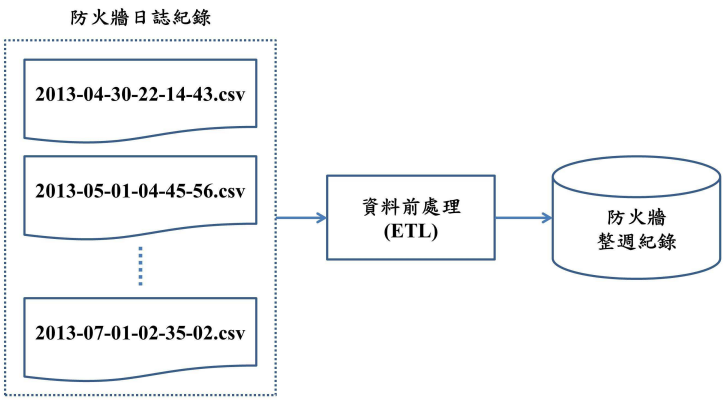

圖 2：資料前處理流程範例

二、關聯規則探勘

關聯規則是用來描述資料項目(item)間關聯性的一種型式。Agrawal 等(1993)提出關聯規則的定義，其目的是用來尋找資料庫中項目之間的關聯性，並分析龐大資料庫裡的交易資料，試圖找出其中的隱含規則。關聯規則的定義如下(Agrawal et al. 1993)：若兩項目集 X、Y 之間有一關聯規則，則以 $X \rightarrow Y$ 的形式表示之，X、Y 為一個或以上之項目所組合而成的集合，且 $X \cap Y = \varnothing$，有兩個條件用來判斷關聯規則 $X \rightarrow Y$ 是否成立：一是支持度（support），其表示為包含有（$X \cup Y$）之交易資料數量佔總交易資料數量的比例值；另一為信心度（confidence），其表示為包含有（$X \cup Y$)之交易資料數量包含有 X 之交易資料數量的比例值。支持度 *suppport*（$X \cup Y$）及信心度 *confidence*（$X \cup Y$）的定義如下：

$$support(X \cup Y) = \frac{|(X \cup Y)|}{|D|}$$

$$confidence(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{sup(X)}$$

其中，$|D|$表示資料庫 D 的總交易資料筆數，$|(X \cup Y)|$表示資料庫 D 同時出現 $X \cup Y$ 的交易資料筆數。

定義 1. 給定最小支持度（minimum support）與最小信心度（minimum confidence)，若關聯規則 $X \rightarrow Y$ 的支持度及信心度分別大於或等於最小支持度及最小信心度，則此關聯規則成立。

一般來說，最小支持度和最小信心度的訂定，可由使用者本身或是專家來訂定。Apriori 演算法是關聯法則中較常使用的方式之一，透過簡單的配對與刪減過程，經由掃描資料庫來找出各種不同項目集之支持度後，產生高頻項目集，之後再進行最小信心度檢測求得關聯規則。Agrawal 與 Srikant（1994）學者所提出的 Apriori 演算法是最常被使用的方法之一，主要包含二個步驟：(1)反覆地產生候選項目集和搜尋整個資料庫，直到找出所有符合最小支持度門檻值的高頻項目集。(2)利用步驟 1 找出的高頻項目集，推導出符合設定最小信心度門檻值的所有關聯規則。以下說明 Apriori 演算法擷取高頻 k-項目集（k>1）並找出關聯規則的步驟：

第一步驟：找高頻項目集

1. 找出所有長度為（k-1）-頻繁項目集 $frequent_{k-1}$，若為∅，則停止執行。
2. 由 1.中找出任兩個有（k-2）項目相同的 $frequent_{k-1}$，組合成 $itemset_k$。

3. 判斷由 2.所找出的 $itemset_k$，其所有包括的 $itemset_{k-1}$ 之子集合是否都出現在 1.中，假如成立就保留此 $itemset_k$；否則就刪除。
4. 檢查由 3.所擷取的 $itemset_k$ 是否滿足最小支持度， 假如符合就成為 $frequent_k$；否則就刪除。
5. 跳至 1.找 $frequent_{k+1}$，直到無法產生高頻項目集為止。

第二步驟：產生關聯規則

1. 將所有高頻 k-項目集（k>1）拆解成 $X \rightarrow Y$，X、$Y \subseteq I$ 且 $X \cap Y = \varnothing$。
2. 判斷所有的規則是否符合最小信心度，若符合則成為關聯規則。

三、改變探勘

一昧地增加關聯規則到防火牆政策中，長久下來規則表將會因為過多的政策規則，而讓防火牆系統效率更差。改變探勘可以應用在防火牆日誌記錄中，針對兩個不同時間區段中，有顯著的網路行為改變及差異進行分析。因此，本研究整合改變探勘的概念，藉以適時調整並更新關聯規則到防火牆政策中，用以解決過多的政策規則問題。關於關聯規則的改變探勘，本研究將運用新興樣式、新增樣式及消失樣式等三種不同樣式於防火牆政策規則的調整。

在介紹變更探勘的定義之前，本研究先介紹定義將使用到的符號：

- t、$t+k$：兩個連續時間，t 表示過去某個時間點，$t+k$ 為以過去時間 t 為基礎的下一個時間點。
- sup^t（$X \cup Y$）、sup^{t+k}（$X \cup Y$）：分別代表時間點 t 及時間點 $t+k$ 時，關聯規則（$X \rightarrow Y$）的支持度。
- R^t、R^{t+k}：分別代表時間點 t 及時間點 $t+k$ 所找到關聯規則集（Rule set）。

以下將依序定義支持度成長率（Growth）、新興樣式、新增樣式及消失樣式。

定義 2. （支持度成長率）令規則（$X \rightarrow Y$）在時間點 t 及時間點 $t+k$ 的支持度分別為 sup^t（$X \cup Y$）及 sup^{t+k}（$X \cup Y$），則規則（$X \rightarrow Y$）的支持度成長率（Growth）定義如下：

$$Growth^{t,t+k}(X \rightarrow Y) = \frac{sup^{t+k}(X \cup Y)}{sup^{t}(X \cup Y)}$$

定義 3. （新興樣式）假設有一關聯規則（$X \rightarrow Y$），在時間點 t 及時間點 $t+k$ 的支持度分別為 sup^t（$X \cup Y$）及 sup^{t+k}（$X \cup Y$），給定一個使用者設定之支持度成長率門檻值 σ_{growth}，倘若關聯規則（$X \rightarrow Y$）同時存在於規則集 R^t 與規則集 R^{t+k} 且 $Growth^{t,t+k}$（$X \rightarrow Y$） > σ_{growth}，則關聯規則

（$X \rightarrow Y$）稱於為新興樣式規則。

根據上述的定義，新興樣式的規則必須同時存在於 t 時間點與 $t+k$ 時間點規則集中，並且該規則的支持度成長率必須高於支持度成長率門檻值。倘若挖掘出屬於新興樣式的規則，此規則將視為新興規則，將可交由管理人員評估是否調整此筆新興規則，即調整防火牆政策規則表中的權重。

定義 4. （新增樣式）假設有一規則（$X \rightarrow Y$）僅存在 R^{t+k} 規則集，而且不存在 R^t 規則集中，則規則（$X \rightarrow Y$）稱為新增樣式規則。

根據上述的定義，新增樣式的規則僅存在於 $t+k$ 時間點規則集中，並且未曾在於 t 時間點規則集中。換言之，該規則在 t 時間點規則集中不曾出現。倘若挖掘出屬於新增樣式的規則，此規則將視為待加入規則，將可交由管理人員評估此筆待加入規則是否可以加入，藉以更新防火牆政策規則表，有利於依照網路現況來保持防火牆之高效率狀態。

定義 5. （消滅樣式）假設有一規則（$X \rightarrow Y$）僅存在 R^t 規則集，而且不存在 R^{t+k} 規則集中，則規則（$X \rightarrow Y$）稱為消滅樣式規則。

倘若挖掘出屬於消滅樣式的規則，此規則將視為待刪除規則，將可交由管理人員評估此筆待刪除規則是否可以刪除，藉以更新防火牆政策規則表。

四、關聯規則與防火牆政策規則之整合評估

企業防火牆政策規則設計皆以開放其所要開放的服務，阻擋未開放或不在開放清單內的服務為主要目標，稱為正向安全模型（Positive security model）或正面表列（White-listing）。以防火牆系統角色來說，也就是允許合法的網路傳輸通過，拒絕所有不受允許的網路傳輸封包。

關聯規則與防火牆政策規則之整合評估流程如圖 3 所示。經由步驟三的改變探勘，可以得到新興樣式、新增樣式及消失樣式等三種不同樣式的關聯規則。其次，不同樣式的關聯規則再與現存的防火牆政策規則（或稱原規則表）進行比較與整合。本研究依照防火牆系統運作的特性，最終比對不到規則的封包都將會拒絕（Deny），而原本允許（Allow）的記錄則表示這些規則已存在於現行防火牆政策規則表中，最後透過調整允許規則及新增、刪除拒絕規則來持續提升並保持防火牆系統的效率。

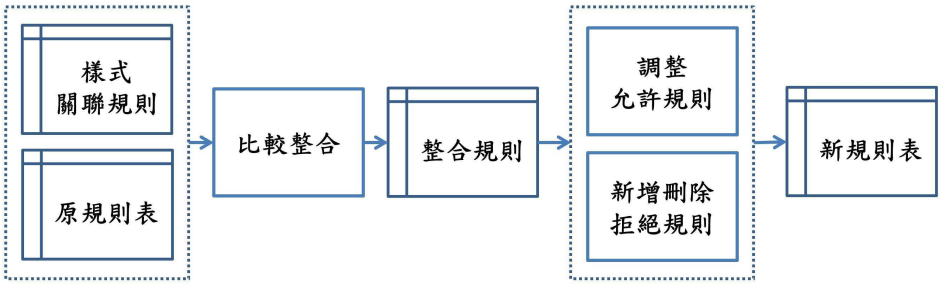

圖 3：防火牆政策規則整合流程圖

本研究的防火牆政策調整動作原則如表 3 所示，此表彙整出任何兩個時間點所產生的關聯規則比對表，符合該表的特徵即執行後續相關動作。例如：在連續兩週所探勘出來的拒絕關聯規則符合新興樣式（即支持度增加），就將此結果通知管理人員並且建議可將該筆拒絕類型的關聯規則加入防火牆政策規則中。同樣地，在連續兩個月所探勘出來的允許類型的關聯規則符合新興樣式（即支持度增加），而且此規則已經存在現有規則表中，即可建議管理人員調整此筆允許類型關聯規則的優先比對順序，使得防火牆系統運作更有效率。

表 3：各種類型規則的執行動作

No	關聯規則類型	樣式	現有防火牆規則表	防火牆政策調整動作
1	允許	新增樣式	有	維持現狀
2	允許	新增樣式	無	Not Available
3	允許	消滅樣式	有	維持現狀
4	允許	消滅樣式	無	Not Available
5	允許	新興樣式	有	依據支持度調整政策順序
6	允許	新興樣式	無	Not Available
7	拒絕	新增樣式	有	維持現狀
8	拒絕	新增樣式	無	維持現狀
9	拒絕	消滅樣式	有	評估為刪除規則
10	拒絕	消滅樣式	無	維持現狀
11	拒絕	新興樣式	有	依據支持度調整政策順序
12	拒絕	新興樣式	無	加入規則

*Not Available：不可能出現的情況。

*若規則不屬於上述類型的規則，則維持現狀（對現有防火牆規則表不做任何異動）。

由於防火牆運採用正面表列的方式，若封包與現有防火牆規則表中的規則都比對不到的話，就會拒絕該封包存取。表 3 中各種規則類型的執行動作詳細描述如下：

1. 因為該規則既然已經存在現有防火牆規則表（有），而且關聯規則類型屬於「允許」，故對現有防火牆規則表不做任何異動，即「維持現狀」。
2. 因為該規則既然未存在現有防火牆規則表（無），代表不可能有任何封包因為該規則而通過防火牆（即不可能出現此規則），故此狀況屬於「Not Available。
3. 因為該規則既然已經存在現有防火牆規則表（有），而且關聯規則類型屬於「允許」，故對現有防火牆規則表不做任何異動，即「維持現狀」。
4. 因為該規則既然未存在現有防火牆規則表（無），代表不可能有任何封包因為符合該規則而通過防火牆（即不可能出現此規則），故此狀況屬於「Not Available」。
5. 因為該規則屬於「Emerging」（即規則的支持度大幅增加），代表符合此規則的封包將大幅增加，為了讓封包與防火牆規則的比對次數減少。因此，將調整該規則在防火牆規則表中的先後順序，即該規則的順序往前調整，讓此符合此規則的封包盡快通過，以降低封包的比對次數。
6. 因為該規則既然未存在現有防火牆規則表（無），代表不可能有任何封包因為符合該規則而通過防火牆（即不可能出現此規則），故此狀況屬於「Not Available」。
7. 因為該規則既然已經存在現有防火牆規則表（有），而且關聯規則類型屬於「拒絕」，故對現有防火牆規則表不做任何異動，即「維持現狀」。
8. 因為該規則既然未存在現有防火牆規則表（無），而且關聯規則類型屬於「拒絕」，故對現有防火牆規則表不做任何異動，即「維持現狀」。
9. 因為該規則屬於「Perished」（即規則的支持度低於最小支持度門檻值），代表符合此規則的封包已經大量減少或消失，為了讓封包與防火牆規則的比對次數減少。因此，可以評估將該規則從防火牆規則表中刪除，以降低封包的比對次數。
10. 因為該規則既然未存在現有防火牆規則表（無），而且關聯規則類型屬於「拒絕」，故對現有防火牆規則表不做任何異動，即「維持現狀」。
11. 因為該規則屬於「Emerging」（即規則的支持度大幅增加），代表符合此規則的封包將大幅增加，為了讓封包與防火牆規則的比對次數減少。因此，將調整該規則在防火牆規則表中的先後順序，即該規則的順序往前調整，讓此符合此規則的封包盡快拒絕，以降低封包的比對次數。
12. 因為該規則既然未存在現有防火牆規則表（無），然而該規則屬於

「Emerging」(即規則的支持度大幅增加),代表符合此規則的封包將大幅增加,為了讓封包與防火牆規則的比對次數減少,故防火牆規則表中應加入此規則,讓此符合此規則的封包盡快拒絕,以降低封包的比對次數。

本研究以範例說明防火牆政策規則整合流程。假設原始防火牆規則共有 3 項規則,如表 4 所示。經由關聯規則探勘及改變探勘所得到的規則共有 2 項規則,如表 5 所示。表 5 中的第一項規則{如果 X= (Src_IP=140.10.20.3, Dst_IP = 218.89.56.4, Dst_Port = 139),則 Y = (Status = Deny);支持度= 0.1, 信心度= 90%}。假設有 1000 個封包,{支持度 = 0.1}代表 1000 個封包中,有 100 個封包符合{X = (Src_IP = 140.10.20.3, Dst_IP = 218.89.56.4, Dst_Port = 139)且 Y= (Status = Deny)}等項目,其比例為 0.1(100/1000);而{信心度= 90%}代表有 100 個封包符合{X= (Src_IP = 140.10.20.3, Dst_IP = 218.89.56.4, Dst_Port = 139)}的前提下,而且這 100 個封包中有 90 個封包也同時符合{ Y= (Status = Deny) }的情況,其比例為 90%(90/100)。運用於防火牆日誌,其管理上的意涵為:支持度(support)越高代表所有封包資料集中同時符合{X}與{Y}的比例越高。而信心度(confidence)越高代表封包符合{X}的情形下,也同時出現{Y}的比例越高,即{Y}經常伴隨著{X}出現。

經調整後的防火牆規則表,表 6 所示。表 6 中的第一項規則{Src_IP=140.10.20.3, Dst_IP = 218.89.56.4, Dst_Port = 139 → Status = Deny},其規則類型屬於拒絕、規則樣式屬於新興樣式、不存在原始防火牆規則表中,故防火牆政策調整動作為「加入規則」。而第二項規則類型{Src_IP = 192.168.10.5, Dst_IP = 140.89.1.4, Dst_Port = 443→Status = Allow },其規則樣式屬於新興樣式、已存在原始防火牆規則表中,因為規則的支持度提高,故防火牆政策調整動作為「依據支持度調整政策順序」,並將該規則向前調整順序。由上述的範例得知:規則排列的順序應優先考慮支持度,而支持度相同則再比較信心度。

表 4:原始防火牆規則表

ID	Protocol	Src_IP	Src_Port	Dst_IP	Dst_Port	Action
1	TCP	192.168.4.10	*	218.222.122.9	22	Allow
2	TCP	192.168.2.1	*	*	80	Allow
3	TCP	192.168.10.5	*	140.89.1.4	443	Allow

表 5:改變探勘得到的規則

X	Y	支持度	信心度
Src_IP=140.10.20.3, Dst_IP=218.89.56.4, Dst_Port=139	Status=Deny	0.1	90%
Src_IP=192.168.10.5, Dst_IP=140.89.1.4, Dst_Port=443	Status=Allow	0.1	88%

表 6：經調整後的防火牆規則表

ID	Protocol	Src_IP	Src_Port	Dst_IP	Dst_Port	Action	Label
1	TCP	140.10.20.3	*	218.89.56.4	139	Deny	新興
2	TCP	192.168.10.5	*	140.89.1.4	443	Allow	新興
3	TCP	192.168.4.10	*	218.222.122.9	22	Allow	原有
4	TCP	192.168.2.1	*	*	80	Allow	原有

肆、實驗設計與結果

本章節將詳細說明本研究之實驗設計與結果。第一節說明資料來源，第二節說明實驗環境與參數設定，第三節介紹本研究的評估指標，第四節說明實驗程序與結果，最後一節探討實驗結果。

一、資料來源

本研究實驗資料乃是收集某中部科技公司企業防火牆日誌記錄資料，資料收集的時間為 2013/4/14~2013/6/8，共 73,437,062 筆資料。為了避免在實際企業環境上運作分析上，因為日誌記錄太過龐大而造成分析系統效能上的瓶頸，首先在日誌記錄檔的資料前處理時，將大量的日誌記錄檔篩選出研究所需要的資料屬性，並將前處理後的資料分別以每週做為區間來儲存記錄檔，此記錄檔再提供後續演算法來產生新的防火牆規則。

二、實驗環境與參數設定

在實驗環境上，本研究採用 Intel i7 i7-2600K 3.4GHz 處理器，搭配 8GB 記憶體，作業系統為 Microsoft Window 7。演算法採用資料探勘軟體 WEKA 3.7.11 版本的 Apriori 演算法做為找出新政策規則的方法。

防火牆政策優化研究主要使用兩種評估指標：規則比對次數（Hamed & Al-Shaer 2006a; Hamed & Al-Shaer 2006b; Hamed et al. 2006c; Masud et al. 2014）與比對處理時間（Rao et al. 2011; Mustafa et al. 2013），這兩種評估指標在過去研究中基本上已被認定為完全正相關（Masud et al. 2014）。因此，本研究選擇使用規則比對次數做為實驗分析的比較基準，意即累計每一個封包通過防火牆時，依照政策規則優先順序比對時之比對次數，比對次數越多代表封包通過防火牆之時間越長、效率越差；反之則通過防火牆時間越短、效率越高。

本研究透過所資料集中之前兩周的資料來找出最小支持度及最小信心度之最

佳設定值。首先先將第 1 周的防火牆記錄檔做為訓練資料集（Training dataset），並且以第 2 周的防火牆記錄檔做為驗證資料集（Verification dataset），而第 2 周的驗證資料集又進一步被細分為 7 個子資料集（即 DataSet1～DataSet7），分別代表第 2 周每一天的防火牆記錄。我們首先設定最小信心度為 80%，最小支持度則由 0.5%開始，依序使用 1%、1.5%、2%、2.5%及 3%，每次以 0.5%支持度為增量（increment），期望將找出最佳的最小支持度參數。實驗結果，如表 7，與原始政策規則相較之下，發現為最小支持度設定為 1.5%時，可降低比對次數最多。

表 7：比對次數 vs. 最小支持度

資料集	原始比對次數	minsup								
		0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
DataSet1	33239538	44573109	28434892	26899884	27606954	27606954	28021876	27468428	27468428	29275895
DataSet2	53621143	70191808	47613710	45296609	45633988	45633988	45898244	46660576	46660576	47982087
DataSet3	60248017	75282936	51745347	49283227	51121517	51121517	51552964	53200529	53200529	54374940
DataSet4	55507947	70691669	47834693	45747506	46473612	46473612	46787667	48310616	48310616	49597029
DataSet5	52626958	65199710	45705460	43383247	43333866	43333866	43728930	45501543	45501543	46837716
DataSet6	50548737	63500310	43041222	40979946	41465693	41465693	41751435	43659254	43659254	45055518
DataSet7	33976618	43760554	28534438	26436093	28514873	28514873	28930969	28346246	28346246	30148933
比對次數減少（AVG）	—	-27.5%	13.8%	18.2%	16.4%	16.4%	15.6%	13.7%	13.7%	10.7%

除此之外，本研究也發現當最小支持度設定為 0.5%時，雖然所產生的關聯規則數量最多，但新規則整合為防火牆政策規則表後，整體比對次數卻反而增加了 27.5%，其原因為加入太多筆的阻擋規則，反而造成規則比對次數比原防火牆政策規則表更多。因此，一昧地增加太多的防火牆政策規則，對於防火牆運作效率上並不會有幫助。另外，將最小支持度提升至 1%時。之後，新防火牆政策規則表平均比對次數已降低 13.8%。隨著最小支持度的提升，當最小支持度提高至 1.5%時，本研究發現平均比對次數降低最多（18.2%）。當最小支持度提高至 2%到 4.5%時，雖然所產生的可用關聯規則更少了。然而，整合防火牆政策規則表後的封包比對次數卻僅降低至 10.7%。其原因為缺乏合適的關聯規則，導致比對次數無法有效減少。換言之，適當的關聯規則數可以有效減少封包比對次數，而過少或過多的關聯規則數卻反而不利於封包比對次數的減少。最後，本研究以最小支持度為 1.5%為基礎，進而找出最適最小信心度。如圖 4 所示，最小信心度為 90%時，整體的比對次數降低最多。

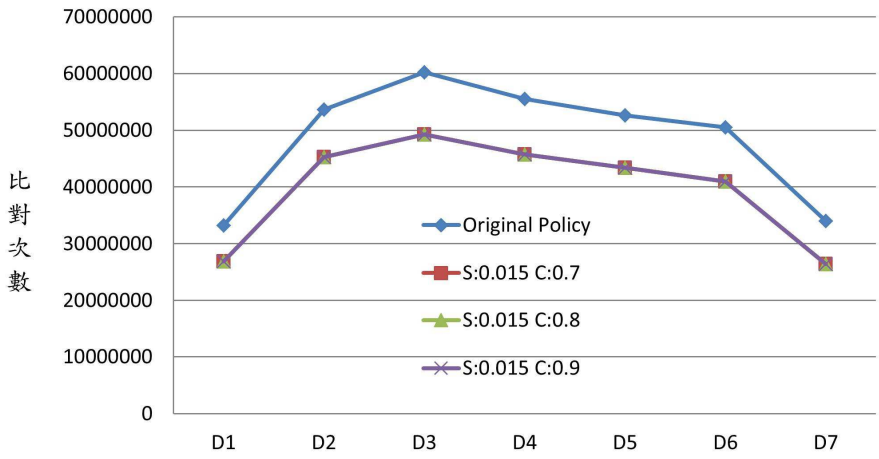

圖 4：比對次數 vs. 最小信心度

由上述的實驗結果得知：將適當的關聯規則數量納入防火牆政策規則表中，可以有效減少封包比對次數進而提升防火牆效能。因此，本研究將以最小支持度為 1.5%、最小信心度為 90%及支持度成長率>1（即 σ_{growth} >1）的參數條件，進行後續實驗。

三、實驗結果

由前章節的參數設定中得知：適當的關聯規則數量納入防火牆政策規則表中，可以有效減少封包比對次數進而提升防火牆效能。然而，如何即時因應資料封包的變化（即網路攻擊的變化，如電腦 IP 不同），適當地新增與刪除防火牆政策規則，或適當地調整防火牆政策規則的權重（或稱比對順序），將有利提升防火牆政策規則的效能，更能協助資安管理者對防火牆政策規則表的管理績效。本章節將說明 CBARM 方法如何結合改變探勘來降低防火牆政策規則表中規則的異動，並且以 Apriori 演算法為比較基礎，進而呈現 CBARM 方法的績效。

在資料部分，本研究從第一周的防火牆記錄檔資料挖掘出關聯規則，並且以後續 4 周的防火牆記錄檔資料（即 Dataset-A, B, C, D）做為驗證資料集。換言之，本研究亦將從 Dataset-A, B, C, D 資料集中挖掘出關聯規則，並運用改變探勘技術分析出那些規則需要異動至防火牆政策規則表。表 8 為以 Dataset-A 為驗證資料集並運用 Apriori 與 CBARM 方法所挖掘出必須異動至防火牆政策規則表中的規則。其中，Apriori 方法所挖掘出的規則有 5 項規則必須異動至防火牆政策規則表中。而 CBARM 方法僅有 1 項規則必須異動至防火牆政策規則表中，即規則#2。以

Dataset-B、Dataset-C、Dataset-D 為驗證資料集所挖掘出必須異動至防火牆政策規則表中的規則請分別參閱附錄表 A、表 B 及表 C。

表 8：Dataset-A 的規則異動

No	src	src_port	dst	dst_port	proto	status	sup.	pattern	Apriori	CBARM
1			60.249.234.226		6	accept	4.75%	Emerging	調整	
2	10.2.23.135			443	6	deny	2.28%	Emerging	調整	調整
3	10.102.23.152			80	6	deny	1.74%	Added	新增	
4	10.102.23.156			80	6	deny	1.73%	Added	新增	
5	10.2.25.11					deny	3.42%	Perished	移除	

Apriori 與 CBARM 方法從驗證資料集的防火牆記錄檔資料(即 Dataset-A, B, C, D）中，所挖掘出的規則且必須異動至防火牆政策規則表的規則數，如表 9 所示。由規則異動次數比較得知，CBARM 方法所挖掘出的規則造成防火牆政策規則表規則異動的次數低於 Apriori 方法，其原因為 CBARM 方法加入改變探勘後，對於封包傳送屬於經常性且屬於特定樣式，才會將該項規則納入防火牆政策規則表中，用以避免無意義規則造成防火牆政策規則表異動過於頻繁。

表 9：規則異動次數比較

資料集	Apriori	CBARM
Dataset-A	5	1
Dataset-B	4	3
Dataset-C	1	1
Dataset-D	5	1

本研究進一步分析規則異動次數的差異是否會對封包比對次數有顯著的影響。表 10 為 Apriori 與 CBARM 的封包比對次數比較，CBARM 方法的封包比對次數在四個驗證資料集中較原始規則表平均減少 15.6%，而 Apriori 方法的封包比對次數較原始規則表平均減少 15.2%。因此，在平均封包比對次數效能上，CBARM 方法略優於 Apriori 方法約 0.4%。

由表 9 的規則異動次數及表 10 的封包比對次數比較得知：雖然 Apriori 與 CBARM 方法的規則皆能降低封包比對次數，但 CBARM 方法所挖掘出的規則在變更防火牆政策規則次數上遠低於 Apriori 方法。如表 11 所示，以 Dataset-A 的防

火牆記錄檔資料集為例，Apriori 方法平均每一項規則異動可以減少 11,459,702 次的封包比對次數，而 CBARM 方法卻能減少 60,106,936 次的封包比對次數，其效能提升（即封包比對次數減少）比例約為 60,106,936/11,459,702 =524.51%。由上得知，CBARM 較 Apriori 方法在效能（封包比對次數減少）上提升 95.19%~582.19%，平均效能提升 212.10%。因此，Apriori 與 CBARM 方法在比對效率（即每條異動規則造成比對次數減少的次數）提升上有明顯的差異。

表 10、Apriori 與 CBARM 的封包比對次數比較

資料集	原始規則	Apriori	CBARM
Dataset-A	330906280	273607768	270799344
Dataset-B	341426229	286779276	290152004
Dataset-C	350824306	301990005	304339094
Dataset-D	327475404	282656142	275288946
平均比對次數減少百分比	—	15.2%	15.6%

表 11：Apriori 與 CBARM 的封包比對效率比較

資料集	比對減少次數/規則異動數		效率
	Apriori	CBARM	
Dataset-A	11,459,702	60,106,936	524.51%
Dataset-B	13,661,738	17,091,408	125.10%
Dataset-C	48,834,301	46,485,212	95.19%
Dataset-D	8,963,852	52,186,458	582.19%
平均	20,729,899	43,967,504	212.10%

由表 8（Dataset-A 的規則異動）得知：Apriori 方法總共產出 5 項異動規則，包含 2 項調整規則、2 項新增規則及 1 項移除規則。而 CBARM 方法僅產出 1 項異動規則（即 1 項調整規則）。倘若採用 Apriori 方法所產出異動規則，則防火牆政策規則表中的總規則數將增加 1 項。然而，倘若採用 CBARM 方法所產出異動規則，則防火牆政策規則表中的總規則數維持不變。假設現有防火牆政策規則表中已經有 9 項規則，由於 Apriori 方法產出 2 項新增規則及 1 項移除規則，倘若採用 Apriori 方法所產出異動規則，則防火牆政策規則表中將有 10（9+2-1）項規則。然而，倘若採用 CBARM 方法所產出異動規則，則防火牆政策規則表中仍維持 9 項規則。假設此一新增規則列於防火牆政策規則表中的第一順位，則每一個封包都會增加一次規則比對次數，即 Apriori 方法產出的新增規則都迫使新封包的比對次

數增加一次。由上述範例得知：Apriori 方法容易造成規則異動，因此 Apriori 方法容易比 CBARM 方法產生更多封包比對次數。

伍、結論

近年來，許多研究嘗試透過關聯規則技術從數量龐大的網路日誌中找出目前所需要的政策規則，然而，過去的研究只考量單一時間區間的關聯規則，卻忽略了在不同時間區間關聯規則的變化趨勢將可能顯著影響防火牆效率，若將單一時間區間的關聯規則全部套用於正在運作的防火牆規則表上，將耗費更多的時間去核對與驗證規則是否有遺漏，同時亦會增加比對次數，造成防火牆效率低落。為解決上述問題，本研究嘗試整合關聯規則探勘及改變探勘等技術，提出 CBARM 方法。我們首先挖掘出關聯規則，進而運用改變探勘技術辨識出新興樣式、新增樣式及消失樣式等 3 種不同樣式的關聯規則。最後，將具有不同樣式的關聯規則運用於防火牆政策規則的調整，進而提升防火牆效率。實驗結果亦證實 CBARM 方法的效能優於原始規則表及 Apriori 方法。

在未來研究方面，本研究所使用的支持度、信心度及支持度成長率乃是由專家事先指定。因此，未來研究可嘗試由資料中自動設定支持度、信心度及支持度成長率等門檻值，以解決需專家事先指定的瓶頸。此外，防火牆的政策規則的辨識率仍是非常值得研究的資安議題，未來希望能整合資料探勘技術，用於提升防火牆的政策規則的辨識率，期能同時提升防火牆比對效能及辨識率。

誌謝

感謝審查委員提供許多的寶貴建議，使本論文之內容更臻完美；本研究承蒙科技部專題研究計畫經費補助，計畫編號為 NSC 102-2410-H-194-104-MY2，謹致謝忱。

參考文獻

李瑞庭、楊富丞、李偉誠，(2012)，『探勘封閉性多維度區間樣式』，*資訊管理學報*，第十九卷，第一期，頁 161-184。

翁政雄，(2011)，『從購買意願資料中挖掘高度相關性的關聯規則』，*資訊管理學報*，第十八卷，第四期，頁 119-138。

黃仁鵬、藍國誠，(2007)，『高效率探勘關聯規則之演算法－EFI』，*資訊管理學報*，第十四卷，第二期，頁 139-167。

鄭麗珍、李麗美，(2014)，『探勘不平衡資料集中之突顯樣式－以國道事故資料為

實證研究』，*資訊管理學報*，第二十一卷，第二期，頁 161-183。

龔旭陽、林美賢、林靖祐、賴威光，(2010)，『針對重要稀少性資料之一種有效率關聯式探勘方法設計』，*資訊管理學報*，第十七卷，第一期，頁 133-155。

Agrawal, R., Imieliński, T. and Swami, A. (1993), 'Mining association rules between sets of items in large databases', *ACM SIGMOD Record*, Vol. 22, No. 2, pp. 207-216

Agrawal, R. and Srikant, R. (1994), 'Fast algorithms for mining association rules', *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, Santiago, Chile, September 12-15, pp. 487-499.

Ahn, K.I. (2012), 'Effective product assignment based on association rule mining in retail', *Expert Systems with Applications*, Vol. 39, No. 16, pp. 12551-12556.

Al Abdulmohsin, I.M. (2009), 'Techniques and algorithms for access control list optimization', *Computers & Electrical Engineering*, Vol. 35, No. 4, pp. 556-566.

Al-Shaer, E.S. and Hamed, H.H. (2003), 'Firewall policy advisor for anomaly discovery and rule editing', *Proceedings of IFIP/IEEE Eighth International Symposium on Integrated Network Management (IM 2003)*, Colorado Springs, USA, March 24-28, pp. 17-30.

Al-Shaer, E.S. and Hamed, H.H. (2004), 'Discovery of policy anomalies in distributed firewalls', *Proceedings of Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004)*, Vol. 4, Hong Kong, China, March 7-11, pp. 2605-2616.

Bailey, J., Manoukian, T. and Ramamohanarao, K. (2003), 'A fast algorithm for computing hypergraph transversals and its application in mining emerging patterns, *Proceedings of the Third IEEE International Conference on Data Mining (ICDM* 03), November 19-22, Melbourne, Florida, USA, pp. 485.

Böttcher, M., Spott, M., Nauck, D. and Kruse, R. (2009), 'Mining changing customer segments in dynamic markets', *Expert Systems with Applications*, Vol. 36, No. 1, pp. 155-164.

Casado, M., Garfinkel, T., Akella, A., Freedman, M. J., Boneh, D., McKeown, N. and Shenker, S. (2006), 'SANE: a protection architecture for enterprise networks', Proceedings of the 15th *USENIX Security Symposium*, Vancouver, B.C., Canada, July 31-Aug 4, pp. 137-151.

Ceci, M., Appice, A., Caruso, C. and Malerba, D. (2008), 'Discovering emerging patterns for anomaly detection in network connection data', *Proceedings of the 17th International Symposium (ISMIS 2008)*, Toronto, Canada, May 20-23, pp.

179-188

Chang, R.I. and Chang, K.W. (2009), 'C-SWF Incremental Mining Algorithm for Firewall Policy Management'. *Journal of Information, Technology and Society*, Vol. 9, pp. 45-62.

Chang, R.I., Lai, L.B., Su, W.D., Wang, J.C. and Kouh, J.S. (2007), 'Intrusion detection by backpropagation neural networks with sample-query and attribute-query', *International Journal of Computational Intelligence Research*, Vol. 3, No. 1, pp. 6-10.

Chen, M.C., Chiu, A.L. and Chang, H.H. (2005), 'Mining changes in customer behavior in retail marketing', *Expert Systems with Applications*, Vol. 28, No. 4, pp. 773-781.

CSI (2011), 'Computer Crime and Security Survey 2011', available at http://www.ncxgroup.com/wp-content/uploads/2012/02/CSIsurvey2010.pdf (accessed 7 December 2013).

Dam, H.K. and Ghose, A. (2015), 'Mining version histories for change impact analysis in business process model repositories', *Computers in Industry*, Vol. 67, pp. 72-85.

Dong, G. and Li, J. (1999), 'Efficient mining of emerging patterns: discovering trends and differences', *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, CA, USA, August 15-18, pp. 43-52.

El-Atawy, A., Samak, T., Wali, Z. and Al-Shaer, E. (2007), 'An automated framework for validating firewall policy enforcement', *Proceedings of the Eighth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2007)*, Bologna, Italy, June 13-15, pp. 151-160.

Feng, W., Zhang, Q., Hu, G. and Huang, J.X. (2014), 'Mining network data for intrusion detection through combining SVMs with ant colony networks', *Future Generation Computer Systems*, Vol. 37, 127-140.

Ganti, V., Gehrke, J. and Ramakrishnan, R. (1999), 'CACTUS—clustering categorical data using summaries, *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, CA, USA, August 15-18, pp. 73-83.

Golnabi, K., Min, R.K., Khan, L. and Al-Shaer, E. (2006), 'Analysis of firewall policy rules using data mining techniques', Proceedings of the *10th IEEE/IFIP Network Operations and Management Symposium (NOMS 2006)*, Vancouver, Canada, April 3-7, pp. 305-315.

Hamed, H. and Al-Shaer, E. (2006a), 'On autonomic optimization of firewall policy

organization', *Journal of High Speed Networks*, Vol. 15, No. 3, pp. 209-227.

Hamed, H. and Al-Shaer, E. (2006b), 'Dynamic rule-ordering optimization for high-speed firewall filtering', *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security (ASIACCS'06)*, Taipei, Taiwan, March 21-24, pp. 332-342.

Hamed, H., El-Atawy, A. and Al-Shaer, E. (2006c), 'On dynamic optimization of packet matching in high-speed firewalls', *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 10, pp. 1817-1830.

Hanguang, L. and Yu, N. (2012), 'Intrusion detection technology research based on apriori algorithm', *Physics Procedia*, Vol. 24, pp. 1615-1620.

Hossain, S.M.S., Rahman, S.M. and Kabir, M.F. (2012), 'Network proxy log mining: association rule based security and performance enhancement for proxy server', *Computer Science and Engineering*, Vol. 49, pp. 9852-9857.

Hu, H., Ahn, G.J. and Kulkarni, K. (2012), 'Detecting and resolving firewall policy anomalies', *IEEE Transactions on Dependable and Secure Computing*, Vol. 9, No. 3, pp. 318-331.

Huang, T.C.K. (2012), 'Mining the change of customer behavior in fuzzy time-interval sequential patterns', *Applied Soft Computing*, Vol. 12, No. 3, pp. 1068-1086.

Huang, Z., Gan, C., Lu, X. and Huan, H. (2013), 'Mining the changes of medical behaviors for clinical pathways', *Studies in Health Technology and Informatics*, Vol. 192, pp. 117-121.

Jeffrey, A. and Samak, T. (2009), 'Model checking firewall policy configurations', *Proceedings of the IEEE International Symposium on Policies for Distributed Systems and Networks (POLICY 2009)*, London, UK, July 20-22, pp. 60-67.

Kamsu-Foguem, B., Rigal, F. and Mauget, F. (2013), 'Mining association rules for the quality improvement of the production process', *Expert Systems with Applications*, Vol. 40, No. 4, pp. 1034-1045.

Katic, T. and Pale, P. (2007), 'Optimization of firewall rules', *Proceedings of the 29th International Conference on Information Technology Interfaces (ITI 2007)*, Cavtat/Dubrovnik, Croatia, June 25-28, pp. 685-690.

Kim, J.K., Song, H.S. and Kim, H.K. (2005), 'Detecting the change of customer behavior based on decision tree analysis', *Expert Systems*, Vol. 22, No. 4, pp. 193-205.

Lee, W. and Stolfo, S.J. (1998), 'Data mining approaches for intrusion detection', *Proceedings of the 7th USENIX Security Symposium*, San Antonio, Texas. January

26-29, pp. 79-94.

Li, C., Reichert, M. and Wombacher, A. (2011), 'Mining business process variants: Challenges, scenarios, algorithms', *Data & Knowledge Engineering*, Vol. *70*, No. 5, pp. 409-434.

Li, G., Law, R., Vu, H.Q., Rong, J. and Zhao, X.R. (2015), 'Identifying emerging hotel preferences using Emerging Pattern Mining technique', *Tourism Management*, Vol. *46*, pp. 311-321.

Li, J. and Wong, L. (2002), 'Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns', *Bioinformatics*, Vol. 18, No. 5, pp. 725-734.

Liu, A.X., Torng, E. and Meiners, C.R. (2008), 'Firewall compressor: an algorithm for minimizing firewall policies', *Proceedings of the 27th Conference on Computer Communications (INFOCOM 2008)*, Phoenix, AZ, USA, April 13-18, pp. 691-699.

Lubna, K., Cyiac, R. and Karun, K. (2013), 'Firewall log analysis and dynamic rule re-ordering in firewall policy anomaly management framework', *Proceedings of the International Conference on Green Computing, Communication and Conservation of Energy (ICGCE 2013)*, Chennai, India, December 12-14, pp. 853-856.

Masud, M.M., Mustafa, U. and Trabelsi, Z. (2014), 'A data driven firewall for faster packet filtering', *Proceedings of the International Conference on Communications and Networking (COMNET 2014)*, Hammamet, Tunisia, March 19-22, pp. 1-5.

Mohammad, M.N., Sulaiman, N. and Muhsin, O.A. (2011), 'A novel intrusion detection system by using intelligent data mining in weka environment', *Procedia Computer Science*, Vol. 3, pp. 1237-1242.

Mustafa, U., Masud, M.M., Trabelsi, Z., Wood, T. and Al Harthi, Z. (2013), 'Firewall performance optimization using data mining techniques', Proceedings of the 9th International *Wireless Communications and Mobile Computing Conference (IWCMC 2013)*, Cagliari, Sardinia, Italy, July 1-5, pp. 934-940.

Park, J.H., Lee, H.G. and Park, J.H. (2010), 'Real-time diagnosis system using incremental emerging pattern mining', *Proceedings of the 5th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2010)*, Sanya, Hainan, China, December 16-18, pp. 1-5.

Rao, C.S., Rama, B.R. and Mani K.N. (2011), 'Firewall policy management through sliding window filtering method using data mining techniques', *International Journal of Computer Science & Engineering Survey*, Vol. 2, No. 2, pp. 39-55.

Saboori, E., Parsazad, S. and Sanatkhani, Y. (2010), 'Automatic firewall rules generator for anomaly detection systems with Apriori algorithm', *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE 2010)*, Vol. 6, Chengdu, China, August 20-22, pp. V6-57.

Salah, K., Elbadawi, K. and Boutaba, R. (2012), 'Performance modeling and analysis of network firewalls', *IEEE Transactions on Network and Service Management*, Vol. 9, No. 1, pp. 12-21.

Sherhod, R., Gillet, V.J., Judson, P.N. and Vessey, J.D. (2012), 'Automating knowledge discovery for toxicity prediction using jumping emerging pattern mining', *Journal of Chemical Information and Modeling*, Vol. 52, No. 11, pp. 3074-3087.

Shie, B.E., Yu, P.S. and Tseng, V.S. (2013), 'Mining interesting user behavior patterns in mobile commerce environments', *Applied Intelligence*, Vol. 38, No. 3, pp. 418-435.

Shih, M.J., Liu, D.R. and Hsu, M.L. (2010), 'Discovering competitive intelligence by mining changes in patent trends', *Expert Systems with Applications*, Vol. 37, No. 4, pp. 2882-2890.

Sreelaja, N.K. and Pai, G.A. (2010), 'Ant Colony Optimization based approach for efficient packet filtering in firewall', *Applied Soft Computing*, Vol. 10, No. 4, pp. 1222-1236.

Song, H.S., kyeong Kim, J. & Kim, S.H. (2001). Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications*, 21(3), 157-168.

Tsai, C.Y. and Shieh, Y.C. (2009), 'A change detection method for sequential patterns', *Decision Support Systems*, Vol. 46, No. 2, pp. 501-511.

US-GAO (2013), 'CYBERSECURITY: National strategy, roles, and responsibilities need to be better defined and more effectively implemented', available at http://www.gao.gov/assets/660/652170.pdf (accessed 7 December 2013).

Vaarandi, R. (2013), 'Detecting anomalous network traffic in organizational private networks', *Proceedings of IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2013)*, San Diego, CA, USA, February 25-28, pp. 285-292.

Wang, G., Zhao, Y., Zhao, X., Wang, B. and Qiao, B. (2010), 'Efficiently mining local conserved clusters from gene expression data', *Neurocomputing*, Vol. 73, No. 7-9, pp. 1425-1437.

Winding, R., Wright, T. and Chapple, M. (2006), 'System anomaly detection: mining

firewall logs', *Proceedings of Securecomm and Workshops*, Baltimore, MD, USA, August 29-September 1, pp. 1-5.

Wu, R.C., Chen, R.S. and Chen, C.C. (2005), 'Data mining application in customer relationship management of credit card business', *Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC 2005)*, Vol. 2, Edinburgh, UK, July 26-28, pp. 39-40.

Yuan, L., Chen, H., Mai, J., Chuah, C.N., Su, Z. and Mohapatra, P. (2006), 'FIREMAN: a toolkit for firewall modeling and analysis', *Proceedings of IEEE Symposium on Security and Privacy (S&P'06)*, Oakland, CA, USA, May 21-24, pp. 199-213.

Zwicky, E.D., Cooper, S. and Chapman, D.B. (2000), *Building internet firewalls*, O'Reilly Media, Inc.

附錄一

表 A：Dataset-B 的規則異動

No	src	src_port	dst	dst_port	proto	status	sup.	pattern	Apriori	CBARM
1	10.2.1.122				6	deny	2.40%	Added	新增	
2	10.2.23.135			443	6	deny	2.01%	不動作	調整	調整
3	10.102.23.152			80	6	deny	1.86%	Emerging		新增
4	10.102.23.156			80	6	deny	1.84%	Emerging		新增
5	10.2.25.11					deny	1.75%	不動作	新增	
6	10.2.35.30				6	deny	1.58%	Added	新增	

表 B：Dataset-C 的規則異動

No	src	src_port	dst	dst_port	proto	status	sup.	pattern	Apriori	CBARM
1	10.2.25.11					deny		Perished	刪除	刪除

表 C：Dataset-D 的規則異動

No	src	src_port	dst	dst_port	proto	status	sup.	pattern	Apriori	CBARM
1	10.2.25.11					deny	2.83%	Added	新增	
2	192.168.8.77				6	deny	2.06%	Added	新增	
3	192.168.8.62					deny	2.03%	Added	新增	
4	10.2.23.135				6	deny	1.65%	不動作	調整	調整
5	10.2.35.30				6	deny	1.52%	Perished	刪除	