雲端分散式日誌搜集方法

陳怡臻*[b]、劉奕賢[b]、盧建同[b]、方泰鈞[b]、李忠憲[ab]
國立成功大學 電機工程學系[a]
國立成功大學 電腦與通信工程研究所[b]

摘要 — 雲端運算近幾年迅速發展，已成為熱門研究議題，透過虛擬化技術，讓基礎設施與軟體的資源管理得以擴充，比傳統系統建置方式更為彈性。雲端運算資料安全的重要性與日俱增，因此本研究基於雲端日誌與追蹤服務提出兩種機制，一為分散式多重傳輸協定之跨層日誌搜集機制，二為基於跨層日誌記錄的資料軌跡追蹤機制。藉由以上兩種機制，預期改善目前雲端環境中不同層級間日誌搜集的問題，以及掌握資料傳輸的相關情況。[1]

一、 前言

在科技與網路快速的發展情況下，各式服務與資料日漸增加造成巨量資料(Big Data)，為了快速處理巨量資料，雲端運算因而崛起。雲端運算為目前熱門的網路趨勢之一，其想法主要是依照使用者的需求，存取並配置雲端硬體與軟體資源[1][2]。根據 National Institute of Standards and Technology(NIST)的定義[3]，雲端運算共有五個基礎特徵(依照需求服務、任意裝置存取網路、有快速佈署的彈性、共享資源池、量測服務)、四種佈署模式(私有雲、社群雲、公有雲、混合雲)和三種服務模式(基礎設施雲、應用雲、平台雲)。

當用戶利用網路對雲端進行運算、處理和儲存等行為時，如何保障資料的安全，並提高服務的效能及滿足可用需求，已成為雲端服務提供商的一項重大難題。因此本研究試圖藉由日誌資料的搜集，忠實呈現資料的存取情況，並透過資料流向、軌跡等相關資訊，提供事後追查的機制。故本研究運用分散式架構建置日誌搜集平台，並支援多重傳輸協定，以有效搜集來自於如網路設備、作業系統或應用程式等不同層級的相關日誌資料；同時結合資料特徵值的方式，以提供追蹤特定資料流向、存取軌跡的管道。

二、 文獻探討

本研究主要目的運用分散式架構建置支援多重傳輸協定的日誌搜集平台，並運用資料特徵值，進行特定資料存取軌跡的追蹤。故本研究將就日誌服務相關議題、日誌所需特性及相關標準等進行相關探討，以做為本研究系統設計的基礎。

2.1 日誌服務相關議題及特性

分析系統前必須先了解系統行為，從相關研究[4][5]中指出透過日誌搜集而得的歷史資料，可以提供管理分析之所需。其中透過日誌觀察，可了解系統與各設備的關係[6]。因此日誌搜集對管理者而言相當重要。綜合以上所述，日誌系統所儲放的歷史資料，可以協助管理者了解不同設備間的關聯性。

而在日誌搜集的作業中，會面臨下列四種議題：

1. 搜集項目：如何決定要搜集何種資料，應包含那些項目，可以在搜集最少資料的情況下，讓管理者可以有效的依據日誌分析發生的事件，進而提供管理上所需的相關證據。
2. 保存期間：在確認所需搜集的資料項目後，雖然已有取捨，但日誌是種長期累積的資料。隨著時間的增長，日誌保存會面臨巨量資料處理的問題。且隨著設備的增加或更新，其日誌格式不盡相同，如何有效的整合這些異質性的日誌，也成為一大學問。
3. 搜集管道：目前針對日誌搜集，主要有兩種方法；第一種方法是讓各設備或系統回傳資料至管理者的主要伺服器；第二種方法是管理者伺服器主動回收日誌。
4. 管理政策：要如何搜集日誌、如何找出適合的管理政策[1][7]、如何分析[8][9]、如何同步、如何取捨日誌的儲存與成本等，皆為管理者的課題。

日誌服務必需符合可重建性、可說明性、問題偵測及入侵偵測等四項特性[10]，以達成有效的搜集與整合日誌資料，提升對系統情況了解，以妥善管理並降低各式風險的系統管理目的，四項特性詳述如下：

1. 可重建性(Reconstruction)：根據日誌所記錄的時序或時間，管理者可推斷並重建出系統的事件發生順序，因此各系統的時間同步性相當重要。本研究的分散式多重傳輸協定之跨層日誌搜集機制將以此為出發點。
2. 可說明性(Accountability)：可依據管理者的要求搜集各式日誌，根據日誌所記錄的訊息項目訓練所需的行為，進一步說明系統狀況。
3. 問題偵測(Problem Detection)：當對系統有疑問或系統發生問題，如資源使用率、系統故障等，可察看問題發生時間前後的日誌，了解問題發生的原因。
4. 入侵偵測(Intrusion Detection)：管理者可根據日誌，查看是否有未經授權登入、多次登入失敗、網路攻擊等問題。

根據不同的系統及設備，日誌的記錄項目以及類型將有所改變，在網路設備上常見的日誌記錄是用簡單網

[1] 本研究由經濟部及國科會贊助，計畫編號 101-EC-17-A-02-S1-222 與 NSC100-2218-E-006-029- MY3。.

Corresponding author:jsli@mail.ncku.edu.tw
DOI:10.6159/IJSE.2014.(4-1).96

路管理通訊協定(Simple Network Management Protocol，SNMP)，而在 Linux-based 作業系統上則常用 Syslog，在 Windows 系統則運用作業系統中的事件記錄，而應用程式則採用資料庫的稽核功能或自行記錄等方式。

2.2 簡單網路管理通訊協定

簡單網路管理通訊協定[11]全名為 Simple Network Management Protocol，簡稱 SNMP，其以通訊協定為基礎的網路管理系統。在該協定中，其成員分為三種角色，分別是管理者(Manager)、代理者(Agent)與被管理者(Managed)。其中代理者主要接收管理者所傳達的指令，依照管理者的要求進行各式行為；被管理者指各式網路設備，如交換器、伺服器等。

簡單網路管理通訊協定主要包括五種基本指令，分別為讀取請求(GetRequest)、讀取下一個請求(GetNextRequest)、讀取回應(GetResponse)、設置請求(SetRequest)與異常情況(Trap)。讀取請求主要是由管理者傳送給代理人，用來檢索被管理者之資訊；讀取下一個請求功能如讀取請求，差別在於讀取的項目為下一個資訊，因為 SNMP 一次只能取得一個資訊；讀取回應是由代理人發送，用來回應讀取請求與讀取下一個請求；設置請求，為管理者傳送給代理人，用來設置變數值；異常情況主要用於異常狀況發生時，管理者設定代理者回傳報告。

2.3 系統日誌

在 Linux 家族的作業系統中，Syslog[12]是最常見的日誌伺服器解決方案[13]。Syslog 中最常見的服務有接收日誌資料的 Syslogd、處理核心日誌的 Klogd 與處理日誌更新備份的 Logrotate。基於 Syslogd 這個日誌服務，Syslogd 可使用 Unix Domain Socket 以聽取日誌，亦可經由 UDP 通訊協定傳輸、聽取其他系統的日誌。除了 Syslogd 外，仍需要其他服務來記錄核心所產生的日誌，這個服務即稱之為 Klogd。當日誌大小愈來愈大時，通常需要做備份與更新，則可以使用 Logrotate 來自動化處理。

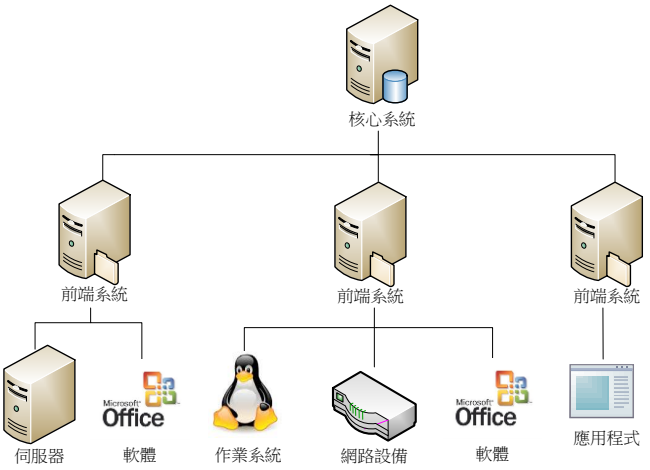
經過 Syslog 所記錄的日誌中，每項資訊都會記錄許多重要資料，其常見資料如：日期、時間、主機名稱、服務名稱、訊息內容等。Syslog 也針對各種服務與日誌記錄定義了：服務的性質、訊息的等級、記錄的所在位置(裝置或檔案)。在服務性質中，Syslog 規範了認證相關機制、例行性工作排程的日誌記錄、與各個程式相關的日誌、或核心產生日誌等；在訊息等級方面，Syslog 則規範基本日誌說明、警示訊息、重大錯誤訊息等。

三、 雲端分散式日誌搜集方法

本研究針對如何建置支援多重傳輸協定的分散式架構日誌搜集平台，同時考量運用資料特徵值，進行特定資料存取軌跡追蹤的目的。本研究提出了分散式多重傳輸協定之跨層日誌搜集機制與基於跨層日誌記錄的資料軌跡追蹤機制兩種機制，詳述如下。

3.1 分散式多重傳輸協定之跨層日誌搜集機制

傳統的日誌搜集，主要多為在特定節點集中搜集的方式，這種方式在效能上較容易遇到瓶頸[14]。而分散式架構(Distributed Architecture)能有效分擔運算，因為分散式架構可分別搜集所分配到的設備日誌，此外更可互相交換，以達到日誌搜集之目的，同時可以避免單一節點效能瓶頸的困境[15]。本研究所提出分散式多重傳輸協定之跨層日誌搜集機制，在系統架構上主要區分為兩個部分，分別為前端系統及核心系統，如圖一所示：

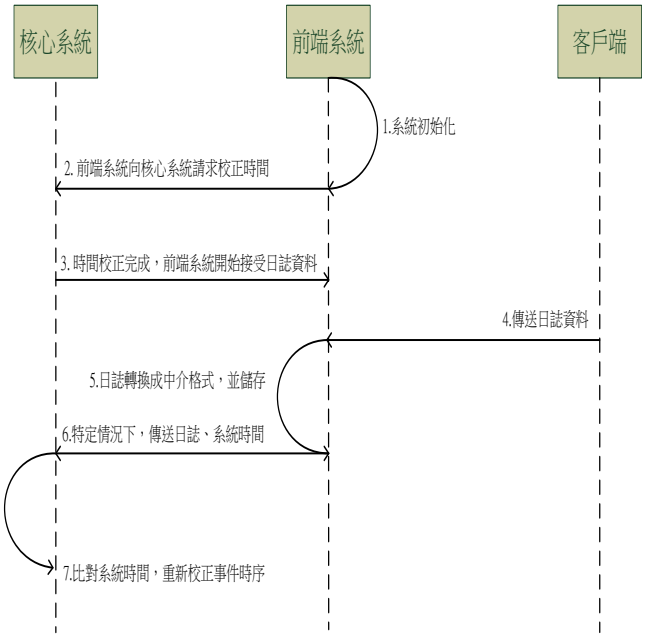


圖一： 分散式多重傳輸協定之跨層日誌搜集機制架構示意圖

該機制的前端系統負責記錄相關系統運用對應的協定傳入的日誌資料，以達成整個機制中對多重傳輸協定支援的目的。前端系統主要會依照所需服務的客戶端使用之協定，搜集其日誌，再將資料轉換成為中介格式，以支援多重協助的日誌資料搜集。而核心系統主要負責彙整前端系統所搜集到的相關日誌資料。然而雲端環境中，前端系統可能位於不同的地區，進而會有時間、日光節約時間等相關的時序問題，故核心系統需比對前端系統的時間與本身的時間，計算其差異，再校正前端系統回傳的每一筆日誌資料發生的時間點，以符合日誌可重建性的要求。

分散式多重傳輸協定之跨層日誌搜集機制的主要運作流程包含下列七大步驟：

1. 前端系統初始化。
2. 前端系統向核心系統請求校正其時間
3. 前端系統開始接受客戶端傳送日誌資料。
4. 前端系統的客戶端，運用自身的協定將所需記錄的日誌資料傳輸給前端系統。
5. 前端系統接受客戶端傳入之日誌記錄，將日誌記錄轉換為中介格式，再儲存於前端系統上。
6. 前端系統在特定其情況下(如：在特定週期或空間不足等情況)，前端系統會將所儲存的日誌記錄及當前系統時間回傳給核心系統。
7. 核心系統接收到前端系統資料時，會參考前端系統的系統時間，加以比對及校正，依日誌資料實際發生的時間順序，儲存在核心系統中。



圖二：　分散式多重傳輸協定之跨層日誌搜集機制主要運作流程

3.2 基於跨層日誌記錄的資料軌跡追蹤機制

在雲端的環境中，資料被誰存取，一直是雲端服務面臨的困境。因為在商業環境中，資料即代表著商機、金錢；特別是在個資法修正通過後，個人識別資料及機敏性的資料傳輸、存取，在高額的罰則下，在雲端環境中進行資料傳輸，即便有相關的加密、虛擬私人網路(VPN)等資訊安全措施，但仍有受罰的風險疑慮。故本研究結合前述分散式多重傳輸協定之跨層日誌搜集機制，在已妥善搜集所需的相關日誌下，結合資料特徵值的運用，提出以日誌為基礎的事後資料存取、傳輸資料軌跡追蹤機制。

本機制透過分散式多重傳輸協定之跨層日誌搜集機制在結合不同記錄的標準如 SNMP、Windows Event 與 Syslog 等不同設備、系統的日誌記錄來源下，針對各種的資料存取行為，如新增、異動、刪除及修改等，額外運用單向雜湊演算法計算其資料之特徵值，再加以記錄。一來在日誌系統中，沒有額外記錄真實資料內容的疑慮，二來受益於單向雜湊演算法計算的運用，其所產生的資料特徵值具有唯一性，故可於事後證明所傳輸、存取的資料內容是否與樣本相符。以此掌握資料傳輸的管道及相關系統或使用者運用資料的情形。有別於傳統資料外洩的防護方案，以事先定義比對為主。本機制可更明確的證明資料被存取的相關軌跡，並進一步掌握資料被存取的情況。

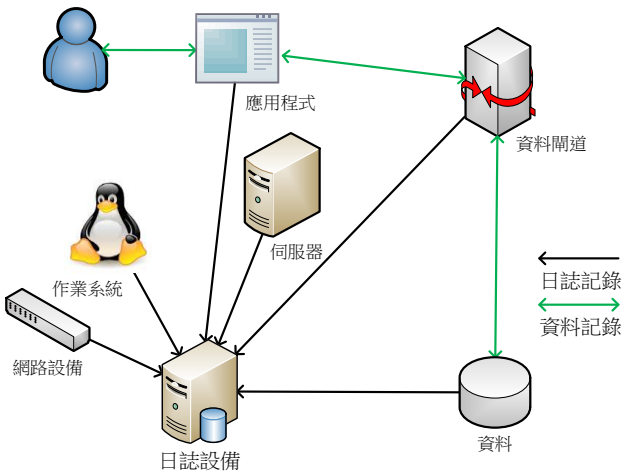
由於本機制透過不同的日誌記錄標準，可突破單一層級，如應用程式或作業系統，掌握更多的資料，更有效的提供資料存取的軌跡，以利相關追蹤。進而更有效的了解資料傳輸的路徑及相關參與到的系統或使用者的操作或運用方式，以供犯罪追查或鑑識相關佐證之用。

本研究提出的基於跨層日誌記錄的資料軌跡追蹤機制，主要是運用日誌搜集系統，配合特殊的網路設備及程式，結合已有的一般網路設備及程式、作業系統來進行，特點分述如下：

1. 由雲端運算環境中所佈署的應用程式(App)，加入相關功能或運用其原有記錄功能，在相關操作時，記錄請求者的應用程式識別資料，如用戶帳號、請求時間，請求項目及回傳項目等。
2. 於平台即服務(PaaS)層中，記錄相關的連線請求資訊(網路層、傳送層、會談層、表現層、應用層)資料，輔以基礎架構即服務(IaaS)層內記錄相關如 MAC、IP 位址等(實體層、資料鏈結層、網路層)資料。
3. 系統中資料對外傳輸的管道，有特定的管道，其需經由特定的設備，以追蹤其所傳輸資料的行為及內容，在此系統中稱為資料閘道。
4. 於日誌系統中，結合上述資料，提供相關比對功能，以識別出潛在的違反規範行為。

基於跨層日誌記錄的資料軌跡追蹤機制完整的運作流程如下所示：

1. 應用程式與相關網路設備，依相關設定回傳其日誌記錄予特定設備。
2. 當應用程式接收到資料請求時，會轉向如資料閘道的特定元件處理；該特定元件負責所有資料的存取行為。
3. 特定元件(如資料閘道)在接受請求及回應時，將相關日誌記錄回傳予指定設備。
4. 被指定設備接收回傳日誌記錄，運用該記錄功能，以日誌記錄方式集中於特定設備群，作為日後相關追蹤與分析比對之用。
5. 管理者可分析與比對該設備的日誌資料，確認資料傳輸的行為與內容。



圖三：　基於跨層日誌記錄的資料軌跡追蹤機制的架構示意圖

雖說基於跨層日誌記錄的資料軌跡追蹤機制有上述僅需儲存單向雜湊演算法產生之特徵值，但在額外運用單向雜湊演算法計算資料內容時，需額外的處理時間。該系統相關要求，如下列所示：

1. 資料閘道中，記錄各日誌項目的識別資料，可識別相關操作行為及異動前後的資料情況。
2. 應用程式(App) 記錄相關的使用者及系統狀態。
3. 記錄支援管理功能的網路設備的狀態。

四、 相關機制比較

本研究將常見的簡單網路管理通訊協定、系統日誌(Syslog)與本研究所提出的兩種機制進行比較。傳統的SNMP 主要適用於監控各式網路設備，並搜集設備日誌，但日誌搜集後並沒有對時序進行校正，且沒有支援各設備的系統應用程式的日誌記錄。系統日誌(Syslog)則由 Syslogd、Klogd 與 Logrotate 三大元件組成，分別由 Syslogd 負責接收日誌、Klogd 記錄核心日誌、Logrotate 職司更新與備份，但仍缺乏日誌資料的時序校正與多重標準支援等特色。因此本研究中所提出的分散式多重傳輸協定之跨層日誌搜集機制，在時序校正與多重標準上可獲得改善。

表I: 現有技術與本研究機制比較表

方法名稱	SNMP	Syslog	本研究提出機制
網路傳輸資訊	支援	部分支援	支援
作業系統訊息	無	部分支援	支援
應用程式訊息	無	部分支援	支援
時序校正	無	無	支援
多重標準支援	無	無	支援
記錄資訊	網路事件	系統/程式事件	資料存取軌跡相關事件

以支援多重傳輸協定為目的的日誌搜集機制中，先前有[16]等研究提出其作法。[16]的主要方式為，透過產生集中式的日誌存放系統，將各個網路裝置或安全裝置產生的日誌資料統一蒐集，這些未經改動的日誌資料檔案會被存放在一個原始日誌資料庫，以便日後重新取出分析使用。又因為每個網路裝置所產生的日誌資料格式不盡相同，為了提升管理與分析的效率，此系統會針對每一筆日誌資料，重新格式化為統一的日誌格式，並集中存放於一個中央儲存的資料庫中，藉以提升資料的可維護性，也方便管理者對於當前網路安全狀況做更完整的評估。相較於[16]，本研究更可校正日誌事件的時間點，以正確重新還原其事件發生順序，避免日誌因時序錯亂而失去可還原性。此外本研究亦針對資料存取相關的應用場景，運用單向雜湊演算法，於發生各種資料存取行為，如新增、異動、刪除及修改等操作時，產生當時資料內容的資料特徵值，以利事後追蹤、稽核的證據需求，並避免過度蒐集資訊進而導致外洩管道增加的問題。

五、 結論與未來展望

雲端運算被視為目前最熱門、高成長的產業，國內外的各大企業，皆大量投入資源，從事相關研發，由此可推估未來雲端運算市場的商機不可限量。此外近年來開始重視個人資料的保護，在各式法規的要求下，雲端運算的安全與隱私成為重要議題。所以日誌的重要性更是不言而喻。因此雲端服務廠商、網通服務廠商或應用服務提供商，可參考運用本研究所提出的這兩種機制，搜集其系統的相關日誌記錄，以供資料存取行為稽核或其他管理佐證需求之運用。

誌謝

感謝經濟部計畫 101-EC-17-A-02-S1-222 及國科會計畫 NSC100-2218-E-006-029- MY3 提供經費支持本研究的進行。

參考文獻

[1] C. Basescu, A. Carpen-Amarie, C. Leordeanu, A. Costan and G. Antoniu, “Managing data access on clouds: A generic framework for enforcing security policies,” IEEE International Conference on Advanced Information Networking and Applications, Mar. 2011, pp. 459-466.
[2] C. H. Lin, C. T. Lu, Y. H. Chen, and J. S. Li, “Resource Allocation on Cloud Virtual Machines Based on Empirical Service Data Traces,” International Journal of Communication Systems, doi: 10.1002/dac.2607, 2013.
[3] P. Mell and T. Grance, The NIST Definition of Cloud Computing (Special Publication 800-145), NIST, 2011.
[4] K. E. Nawyn, A Security Analysis of System Event Logging with Syslog, SANS Institute, 2003.
[5] P. Jackson, Introduction to Expert Systems, Addison-Wesley, 1986.
[6] J. Stearley, “Towards Informatic Analysis of syslogs,” IEEE International Conference on Cluster Computing, 2004, pp. 309-318.
[7] D. Lin and A. Squicciarini, “Data protection models for service provisioning in the cloud,” Proceeding of the 15th ACM symposium on Access control models and technologies, 2010, pp. 183-192
[8] J. Zhou, M. Heckman, B. Reynolds, A. Carlson and M. Bishop, “Modeling Network Intrusion Detection Alerts for Correlation,” ACM Transactions on Information and System Security, Vol. 10, No.1, pp. 1-31, 2007..
[9] 許宏名, 日誌簡化與相關性整合下發展有效偵測網頁入侵策略, 國立中正大學通訊工程研究所 碩士論文, 2010.
[10] G. Spafford, “The Importance of Audit Logs,” Retrieved 2013/08/13 from http://www.datamation.com/columns/article.php/3578916/The-Importance-of-Audit-Logs.htm, 2006.
[11] J. Case, M. Fedor, M. Schoffstall and J. Davin, RFC 1157 - A Simple Network Management Protocol (SNMP), IETF Network Working Group, 1990.
[12] J. Schönwälder, “On the Impact of Security Protocols on the Performance of SNMP,” IEEE Transactions on Network and Service Management, Vol. 8, NO. 1, pp. 52-64, 2011.
[13] 陳嘉玫, 林孝忠, 洪瑞麟, 吳惠麟, “以Linux 系統為基礎之日誌檔樣式化研究,” TANET 2010, Tainan, Taiwan, 2010.
[14] 李亮寬, 結合防毒與入侵偵測之網路阻斷系統研究, 大同大學 資訊工程系 碩士論文, 2009.
[15] 呂崇富, 網路規劃與管理實務, 學貫出版社, 2007
[16] S. Shengyan, S. Xiaoliu, Z. Jianbao, M. Xinke, “Research on System Logs Collection and Analysis Model of the Network and Information Security System by Using Multi-agent Technology,” Fourth International Conference on Multimedia Information Networking and Security (MINES), Nov. 2012, pp.23,26.