雲端日誌驗證稽核系統

劉奕賢張展華盧建同李忠憲
國立成功大學電機工程學系 / 電腦與通信工程研究所
{dannyliu, frankchang, chientung}@hsnet.ee.ncku.edu.tw, jsli@mail.ncku.edu.tw

摘要

近年來，雲端運算迅速發展，目前已成為資訊領域中最受關注的議題之一，但雲端運算所產生龐大資料量的安全性問題與日俱增，因此雲端安全的議題也是資訊產業發展的焦點。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蒐集機制，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，加上協定驗證與視覺化呈現日誌系統，預期增加雲端環境中不同虛擬主機之間資料傳遞可行性，以及提高日誌可讀性。

關鍵詞：雲端運算、網路安全、日誌蒐集系統、協定驗證、資料視覺化

Abstract

Recently, cloud computing is getting fully developed, which has become one of the most popular information technology. However, it generated huge volume of data and led to some security issues, so the cloud security issue was also popular in computing field. Therefore, in this paper we proposed two mechanisms for log collection system, one is the cross-layer distributing log collection of multiple transport protocols, and the second one is the tracking mechanism, which is based on cross-layer distributing log collection system. In addition to these two mechanisms, we add another mechanismfor protocols verification, which was expected to increase the reliability of the current cloud environments with different hosts. Finally, we used the data visualization technology to improve the readability of log collection system.

Keywords: Cloud Computing;Network Security; Log Collection System; Protocols Verification; Data Visualization

1. 前言

隨著網際網路的快速發展，各種服務的資料量日漸增加造成巨量資料(Big Data)的產生，因此為了能快速處理這些巨量資料，雲端運算技術日趨成為目前熱門的 IT 產業趨勢。雲端運算將計算和資料從桌上型電腦與可攜式電腦搬移到雲端資料中心，使用者無需知道實體機器的細部設定與位置，便能透過網路使用雲端資源。雲端供應商按照使用者需求，加以配置並共享雲端資源 (如網路、伺服器、記憶體等)，此舉亦可減少資源閒置。

雲端運算近幾年迅速發展，目前國內外有許多廠商提供雲端的服務，最知名的即為亞馬遜(Amazon)，此外還有許多國外知名廠商，如：谷歌(Google)、奇摩(Yahoo)、微軟(Microsoft)等公司，而在國內亦有中華電信、台灣大哥大等公司提供雲端運算服務。由此可知雲端運算將是一種趨勢，而近年來所重視的個資法更是讓雲端運算資料安全的重要性與日俱增，所以蒐集及統整雲端運算的日誌資料，可更加掌握雲端運算的狀況。

目前在雲端運算有許多的議題，例如：雲端運算的資源分配 [1]、雲端運算模型研究[2]、雲端運算管理架構 [3]等。近年來國內越來越重視個資法，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與存取更顯重要，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，但仍有洩漏的風險疑慮。所以在個資越來越受到保護的現在，雲端系統對於個資的處理必須更加嚴謹。所以如何在雲端環境中，有效保護個人隱私，同時又可兼顧系統資料分析有效性的需求，更成為目前的趨勢 [4] [5]。

因此用戶在使用雲端服務中最大的疑慮就是雲端服務的安全問題，當用戶透過網路對雲端進行運算、處理和存取等行為時，除了提供其必須且可靠的授權與存取安全性之外，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蒐集，紀錄雲端資料的存取情況，並透過資料流向、軌跡與協定認證等相關資訊稽核機制。若日後使用者對於平台服務有任何問題，管理者皆可透過查詢與分析，提供相對應之稽核日誌資料。故本研究運用分散式架構建置日誌蒐集平台[6]，並支援多重傳輸協定，有效蒐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與協定認證的方式，提供追蹤特定資料流向、存取軌跡的管道，並提供管理者一個視覺化介面使其更具可讀性。

2. 文獻探討

本研究主要目的是運用分散式架構建置支援多重傳輸協定的視覺化日誌蒐集平台，當用戶利用網路對雲端進行運算、處理和儲存等行為時，透過資料特徵值與協定驗證，進行特定資料存取軌跡的追蹤。故本研究將藉由日誌服務、日誌所需特性及相關標準等議題進行相關探討，以作為本研究系統設計的基礎。

2.1 日誌服務相關議題及特性

分析系統及除錯前必須先了解系統行為，從相關研究[7][8]中指出透過日誌蒐集而得到的歷史資料，可以提供管理分析之所需，並經由預警等功能來偵測威脅，以便維持系統的安全性及可靠性。此外，透過日誌觀察，可了解系統與各設備的關係[9]。因此日誌蒐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。

而在日誌蒐集的作業中，會面臨下列四種議題：

1. 蒐集項目：決定蒐集何種資料與應包含哪些項目，可以在蒐集最少資料的情況下，讓管理者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：在確認所需蒐集的資料項目後，雖然已有取捨，但日誌是種長期累績的資料。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. 蒐集管道：目前針對日誌蒐集，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：要如何蒐集日誌、如何找出適合的管理政策[10][11]、如何分析[12][13]、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[14]，以達成有效的蒐集與整合日誌資料，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蒐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蒐集各式日誌，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路管理通訊協定(Simple Network Management Protocol，SNMP)，而在 Linux-based 作業系統則常用 Syslog，在 Windows 系統則運用作業系統的事件記錄，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 雲端運算

「雲端運算」就是使用分散式運算的方式，讓不同電腦透過網路同時協助你處理資料運算。根據 National Institute of Standards and Technology(NIST)的定義[15]，雲端運算可以依照使用者的需求，便利的透過網路提供資源，讓使用者共同享用。

雲端運算五個基本特徵，包含：依照需求提供服務、隨時隨地透過任何網路裝置存取、多人共享資源池、快速且有彈性的使用服務資源、可被監控與量測的服務。另有私有雲、社區雲、公共雲及混合雲等四種部署模型，服務模式則有基礎設施即服務、平台即服務、軟體即服務等三種 [16]。

從服務模式來說，基礎設施即服務(Infrastructure as Service)為使用者可根據需求租用所需的基礎設施，這種透過虛擬化提供客戶端的服務，使硬體的使用率大幅提昇。平台即服務(Platform as a Service)為提供一個平台讓使用者可以在上面部署及執行應用程式等雲端運算服務。軟體即服務(Software as a Service)則為透過網路操作雲端應用軟體。從部署模型來說，私有雲(Private Cloud)主要用於運算或儲存資料隱密性較高的情況。社區雲(Community Cloud)通常為多個群組組成，主要用於共享資料的情況。公有雲(Public Cloud)則常為服務供應商提供，適用於資料隱密性較低的情況，可提供給一般民眾使用。綜合上述這三種部署模型，在考量資料的情況下加以選擇，而同時混合兩種或兩種以上的部署模型，則稱為混合雲(Hybrid Cloud)。

2.3 系統日誌

Syslog[17]為 Linux 家族作業系統中最常見的日誌伺服器解決方案[18]，其中最常見的服務有處理日誌更新備份的 Logrotate、處理核心日誌的 Klogd 與接收日誌資料的 Syslogd。Syslogd 除可使

用 Unix Domain Socket 以聽取日誌之外，亦可透過 UDP 通訊協定傳輸聽取其他系統的日誌。Klogd 則是用來記錄核心所產生的日誌。當日誌日漸增加時，需做備份或更新時則可以透過 Logrotate 來自動化處理。

Syslog 所記錄的日誌紀錄中，每筆資訊除了記錄一些常見的重要資料如：日期時間、主機名稱、服務名稱、訊息內容等。此外，Syslog 也會針對各種服務與日誌記錄分別定義：服務的性質、日誌的等級、日誌的所在位置(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；在日誌等級方面，Syslog 則規範基本日誌說明、警示訊息、重大錯誤訊息等。

2.4 協定驗證

協定驗證是指網路中任兩實體裝置在互相溝通時所需遵循的通訊協定，也就是說不同電腦之間若要互相傳送資料就必須要有共同的語言與規範，因此其規範包含資料的傳輸格式、編號、資訊控制、錯誤處理以及時序結構等等。

TCP/IP 所使用的三向式握手(three-way handshake)是一個典型的通訊協定認證，其工作機制為先從來源端發出 SYN 封包到伺服器端要求建立 TCP 連線，並進入 SYN_SEND 狀態，當伺服器端接受到連線請求後會回傳 SYN/ACK 封包，並進入 SYN_RECV 狀態，在來源端收到伺服器回傳的 SYN/ACK 封包後，便再發送 ACK 封包進行確認，此時雙方都進入 ESTABLISHED 狀態，完成三向式握手並開始傳送資料，若在這期間發生封包遺失而未收到 ACK 封包，則會在限定時間後重新發送確認封包，若持續未收到回覆，則會放棄建立連線。因此可知網路是雙向的，並且在雙方建立連線前必須先經過授權請求，才能進行後續個別對應的資料傳輸策略。

3. 具程序驗證機制之雲端分散式日誌系統

本研究將探討如何建置支援多重傳輸協定的分散式架構日誌蒐集平台，同時考量運用資料特徵值與協定驗證，進行特定資料存取軌跡追蹤的目的，因此提出了分散式多重傳輸協定之跨層日誌蒐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 分散式日誌蒐集系統

傳統日誌蒐集大多採用集中式，而此方法有個致命的缺點就是在效能上容易遇到瓶頸[19]。而分散式架構(Distributed Architecture)能有效地避免此情況發生，分散式架構可分別蒐集其所分配到的設備日誌，此外更可透過互相交換達到日誌蒐集之目

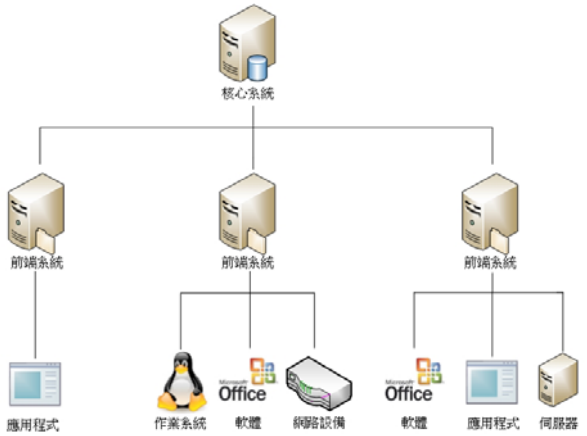

圖 1. 分散式多重傳輸協定之跨層日誌蒐集機制架構示意圖

的，因此可以避免單一節點效能瓶頸的困境[20]。本研究所提出分散式多重傳輸協定之跨層日誌蒐集機制，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 1 所示。

此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，以達成整個機制中支援多重傳輸協定的目的。前端系統會依照客戶端使用之協定，進行日誌蒐集，再將資料轉換成為中介格式。核心系統負責彙整前端系統所蒐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。

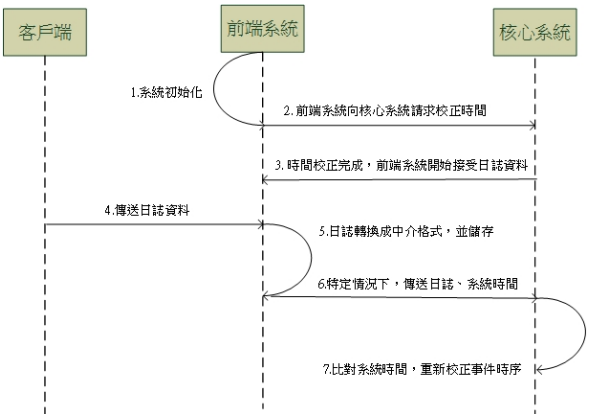

圖 2. 分散式多重傳輸協定之跨層日誌蒐集機制主要運作流程

分散式多重傳輸協定之跨層日誌蒐集機制的主要運作流程包含下列七大步驟 (如圖 2 所示)：

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. 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。

3.2 日誌活動程序驗證機制

前兩節主要是針對多筆日誌同時寫入系統所造成時間差問題，並結合不同記錄的標準而提出解決方法。本節將規範兩實體裝置互相溝通所需之通訊協定的驗證程序，如圖 3 所示。

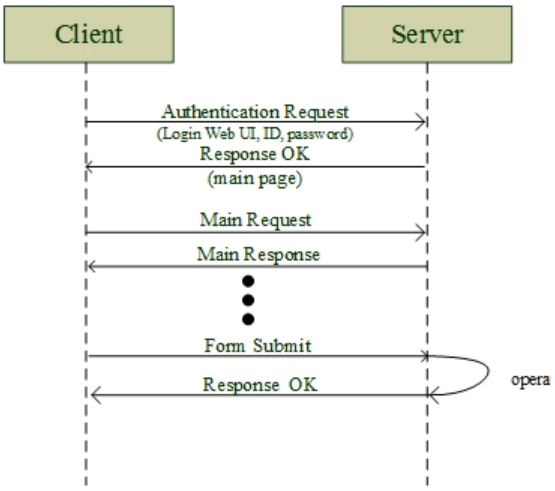

圖 3. 實體裝置間互相溝通所需之通訊協定驗證的架構示意圖

因不同模組間有著不同的動作請求，但前置步驟都是必須先經過授權認證，其詳細運作流程如下所示：

1. 客戶端(來源端)發送授權認證到伺服器端(目的端)
2. 伺服器端收到授權請求後進行資料確認
3. 伺服器端將確認結果回傳到客戶端
4. 客戶端收到授權許可便開始傳送請求，否則將不繼續進行動作

在雙向溝通途中，若有一方出現問題，未依循既定協定格式發送訊息，則無法完成任務，並且紀錄在異常日誌紀錄中。一般日誌紀錄可提供給網管人員監控系統的工作狀態，一旦發生問題時透過日誌查詢分析，便能在短時間內解決問題。對於一般使用者，有時也會有需要察看日誌記錄的需求，因此為了提高日誌的可讀性，我們從架設在雲端伺服器上的日誌資料庫中提取日誌等級、日期、時間、代理者與訊息內容，並呈現在前端網頁界面上。此外，因考量到原始訊息內容的可讀性，我們也將訊息內容簡化處理，並且在網頁下方加上對應的通訊協定架構圖，如圖 4 所示。

圖 4. 前端網頁介面之使用者日誌紀錄檢索系統

4. 相關機制比較

本研究將 Linux 家族的系統日誌(Syslog)與本研究所提出的機制進行比較，詳見表 1。系統日誌(Syslog)是由 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 執行更新與備份，但缺乏日誌資料的時序校正與多重標準支援等特色。因此本研究所提出的分散式多重傳輸協定之跨層日誌蒐集機制，在時序校正與多重標準上可獲得明顯的改善。

表 1. 現有技術與本研究機制比較表

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

先前有研究[21]提出在支援多重傳輸協定為目的的日誌蒐集機制，其方法為集中式的日誌存放系統。因為每台網路或安全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，其系統會將每筆資料重新格式化為統一的日誌格式，並將其所產生的日誌資料集中存放於一個中央儲存的原始資料庫中，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。本研究增加可校正日誌事件的時間點功能，可正確還原其事件之發生順序，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各種資料的存取行為，如新增、異動、刪除及修改等操作，並產生當時資料內容的資料特徵值，以保留事後追蹤、稽核的證據，亦可避免資訊被過度蒐集而導致資訊外洩機率增加的可能性。

5. 結論與未來展望

近年來，因個資保護法逐漸被重視，因此雲端運算的安全與隱私性成為重要的議題，這也致使日誌的重要性更是不言而喻。未來將運用日誌系統分析的結果，提供稽核及預警的功能作為主要訴求，針對系統行為提供稽核功能，並對異常行為進行相關的預警。

本研究建置此日誌系統與研發稽核和預警機制，以期協助雲端服務供應商所需的相關資訊，在有限的資源中，提昇服務的效能，同時增進其可用性。並透過日誌活動程序驗證機制，於事後稽核相關活動是否符合預期，確保管理規則、政策確實被遵循。

誌謝

感謝經濟部計畫 102-EC-17-A-02-S1-222 及科技部計畫 MOST 103-2221-E-006 -146 -MY3 提供經費支持本研究的進行。

參考文獻

[1] E. Elghoneimy, O. Bouhali and H. Alnuweiri, “Resource allocation and scheduling in cloud computing,”International Conference on Computing, Networking and Communications (ICNC) 2012, Jan. 30-Feb. 2,Maui, 2012.

[2] C.-H. Lin, C.-T. Lu, Y.-H. Chen and J.-S. Li, “Resource allocation in cloud virtual machines based on empirical service traces,”International Journal of Communication Systems, DOI: 10.1002/dac.2607, 2013.

[3] Z. Liu, W. Tong, Z. Gong, J. Liu, Y. Hu and S. Guo, “Cloud Computing Model Without Resource Management Center,” International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) 2011, Oct. 10-12, Beijing, China, 2011.

[4] S. Hamouda, “Security and Privacy in Cloud Computing,”International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM) 2012, Dec. 8-10, Dubai, 2012.

[5] S. M. Rahaman and M. Farhatullah, “PccP: A Model for Preserving Cloud Computing Privacy,”International Conference on Data Science & Engineering (ICDSE) 2012, Jul. 18-20, Cochin, Kerala, 2012.

[6] 陳怡臻, 劉奕賢, 盧建同, 方泰鈞, 李忠憲, “雲端分散式日誌搜集方法,” 2013 全國電信研討會, Jan. 15-16, Tainan, Taiwan, 2013.

[7] K. E. Nawyn, “A Security Analysis of System Event Logging with Syslog,” SANS Institute InfoSec Reading Room, 2003.

[8] P. Jackson, Introduction to Expert Systems, Addison-Wesley, 1986.

[9] J. Stearley, “Towards Informatic Analysis of syslogs,” IEEE International Conference on Cluster Computing 2004, Sep. 20-23, San Diego, USA,2004.

[10] 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 (AINA) 2011, Mar. 22-25, Biopolis, 2011.

[11] D. Lin and A. Squicciarini, “Data protection models for service provisioning in the cloud,” 15th ACM Symposium on Access control models and technologies, Jun. 9-11, Pittsburgh, USA, 2010.

[12] 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..

[13] 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.

[14] 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.

[15] P. Mell and T. Grance, The NIST Definition of Cloud Computing (Special Publication 800-145), NIST, 2011.

[16] P. Mell et al., “Cloud Computing: Recommendations of the National Institute of Standards and Technology,” NIST, 2011.

[17] 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.

[18] 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, Oct. 27-29, Tainan, Taiwan, 2010.

[19] 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.

[20] 呂崇富, 網路規劃與管理實務, 學貫出版社, 2007

[21] 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,” 4th International Conference on Multimedia Information Networking and Security (MINES) 2012, Nov. 2-4, Nanjing, 2012.