

# 線上遊戲帳號盜用偵測研究

中央研究院資訊科學研究所 / 數位典藏與數位學習國家型科技數位  
技術研發計劃

## 作者簡介

### 林建威

中央研究院資訊科學研究所多媒體網路與系統實驗室研究助理，研究方向為線上遊戲安全議題。

E-mail: [masaki@iis.sinica.edu.tw](mailto:masaki@iis.sinica.edu.tw)

### 陳寬達 ( 陳存暘 )

中央研究院資訊科學研究所助研究員及數位典藏與數位學習國家型科技數位技術研發計劃Web 2.0 團隊共同主持人，研究方向為網路量測、網路安全及線上遊戲。網址：

<http://www.iis.sinica.edu.tw/~cychen>

## 摘要

帳號盜用為線上遊戲中網安方面極嚴重的問題。駭客通常會從盜用的帳號中取走值錢的虛擬道具，換取實際的金錢。雖然帳號盜用事件時有所聞，至今市面上仍未發展出有效偵測帳號是否遭到盜用的方法。在受害者抱怨之前，遊戲公司通常無法察覺帳號盜用事件的發生，因此當玩家發現自己受害時往往都為時已晚，駭客早已取走所有值錢的虛擬寶物及道具。

在本文中，我們研究是否可根據玩家的遊戲行為來偵測使用者的帳號是否已遭盜用。我們發現虛擬角色的閒置時段（意味虛擬角色停止不動的時間）的分佈就像指紋一樣，可代表一個人的特徵。根據此發現，我們利用 Kullback-Leibler 距離（KL 距離）計算出兩個閒置時段分佈的距離，提出 RET 法來解決使用者辨識的問題。評估結果顯示，基於 200 分鐘的歷史資料，RET 法的辨識準確率能夠在 20 分鐘內達到 90%。

## 1. 引言

安全性一向是網路遊戲設計者及玩家最為擔心的問題。而在各種安全性的議題中，帳號盜用是最為嚴重且普遍的問題之一。駭客往往利用網路釣魚網站或木馬程式竊取使

用者的遊戲帳號。一旦使用者的電腦被「**帳號收集器**」的程式感染，其便常駐於系統內，隨時監控使用者登入遊戲時所輸入的帳號及密碼。被截取的帳號及密碼會被傳送到駭客所架設的伺服器；隨後駭客就可以竊取的帳號登入遊戲，拿走值錢的虛擬道具，換取實際的金錢。即使帳號盜用在許多線上遊戲都時有所聞，但至今市面上仍未發展出有效偵測帳號是否遭到盜用的方法。在受害者抱怨之前，遊戲公司通常無法察覺帳號盜用已發生，因此當玩家發現自己受害時往往都為時已晚，駭客早已取走所有值錢的虛擬寶物及道具。

此外，為了節省月費或是為方便交換資訊，單一帳號可能由**多位玩家共享**，這將會增加遊戲人口分析的難度。舉例而言，為改善遊戲設計或收費策略，遊戲公司會分析影響消費者行為的可能因素，例如玩家年齡、性別以及職業等等，這些都可能影響他們的消費選擇或是持續消費的時間長度。而帳號共享的現象會影響人口分析的正確性。另外，帳號共享也會使遊戲公司難以提供準確的個人化服務給每位玩家。

在本文中，我們應用「**使用者辨識**」技術來解決以上所談論的線上遊戲帳號盜用及帳號共享的問題。使用者辨識是學術領域一個研究已久的議題，根據使用者的輸入辨識出誰是「**真的**」使用者。直到今日，幾乎所有的網路系統都倚賴使用者輸入帳號及密碼的認證機制來辨識使用者。一旦使用者的帳號密碼洩漏給其他人，單純倚賴以上機制將無法判斷帳號是否已遭盜用。

我們研究根據玩家的遊戲行為來偵測使用者的帳號是否已遭盜用，原理為利用玩家所控制的虛擬角色在一連串動作之間的**閒置時間長度**（意味虛擬角色停止不動的時間）來做為玩家的**行為特徵**。也就是說，我們發現虛擬角色的閒置時段的分佈就像指紋一樣，可代表一個玩家的特徵。舉例而言，急性子、重度沉迷的玩家有可能會在一個動作結束之前，就已想好接下來更多的動作，因此在一連串的行為當中就會有較少頻率且長度較短的閒置時間出現。相反地，空閒、隨性的玩家可能會進行較多社交行為，或利用虛擬角色移動時進行遊戲世界以外的行為（例如趁空檔瀏覽網頁），因此這類玩家在一連串行為之間的閒置時間就會顯得較頻繁且時間較長。我們與宇峻奧汀科技股份有限公司合作，收集「天使之戀」遊戲中 287 位玩家的行為記錄，來檢測 RET 法的效度。我們根據 Kullback-Leibler 距離計算出兩個閒置時間分佈的差異，在使用者辨識的問題上有不錯的成效。評估結果顯示，基於 200 分鐘的歷史資料，我們能夠在 20 分鐘內，達成 **90%** 的辨識準確率。

## 2. 玩家行為分析

本節中，我們檢驗玩家行為模式是否能夠成為使用者辨識的準則。首先敘述收集到的玩家行為資料型態。接下來，觀察活動時段以及閒置時段的機率分佈，來分析一般的玩家行為模式，檢驗是否能夠代表玩家的遊戲行為特徵。最後，我們檢驗是否只靠閒置

時段的分佈就能夠代表玩家的行為特徵。

## 2.1 資料描述

我們定義活動時段以及閒置時段如下：

定義 3.1 虛擬角色的活動時段為在某時段  $(t_1, t_2)$  之間，虛擬角色進行連續的移動，最多可容忍一秒鐘的閒置，即虛擬角色在  $(t_1, t_2)$  之間沒有閒置超過一秒鐘以上。

定義 3.2 虛擬角色的閒置時段為在某時段  $(t_1, t_2)$  之間，虛擬角色超過一秒鐘沒有進行移動，即  $t_2 - t_1 \geq 1$  秒。

在本文中，我們考慮活動時段以及閒置時段的時間長度。

玩家資料來自宇峻奧汀科技股份有限公司所開發的多人線上角色扮演遊戲—**天使之戀**（詳見圖 1）。資料由遊戲伺服器之一取得。為了避免系統負荷過大，只隨機抽取三天資料中部分玩家的資料。另外，為了確保分析上的成效，我們刪除短於 200 分鐘的玩家資料。玩家的行為資料統整如表格一。另外由於我們聚焦於短期間的行為模式，於是刪除活動時段以及閒置時段超過 10 分鐘的部分。

圖 1 天使之戀的遊戲畫面



表 1 玩家資料的統整

玩家編號	資料長度	行為比率	行為週期	閒置週期
287	( 7.0, 50.6, 67.1 ) 小時	( 0.35, 2.28, 5.12 ) 循環/每分鐘	( 3, 6, 9 ) 秒	( 7, 18, 181 ) 秒

\*其中 ( x, y, z ) 分別表示 ( 0.05, 0.50, 0.95 ) 百分位數

\* 循環表示一個活動時段後接一個閒置時段的組合

## 2.2 一般玩家行為

我們將活動時段及閒置時段的機率密度曲線描繪於圖 2。由圖可知閒置時段相較於活動時段有較為分散的曲線。也可看出活動時段通常短於 4 秒，這意味著玩家們通常會先思考一下或是先進行移動以外的行為後才有短暫的移動。

我們進一步檢驗每位玩家活動時段及閒置時段平均長度之間的關係。由於閒置時段有可能非常地漫長（舉例而言，玩家可能沒有下線就離開電腦），我們採用一個界限來界定活動時段以及閒置時段最大的值。當界限定於 30 秒時，如圖 3 所示活動時間與閒置時間有很強的**負相關**關係。換句話說，若玩家有較長的平均活動時間，則他/她會傾向擁有較短的平均閒置時間，反之亦然。因此，積極的玩家相較於隨性的玩家會傾向有較長的活動時段以及較短的閒置時段。

圖 2 所有玩家的活動時間及閒置時間的分佈

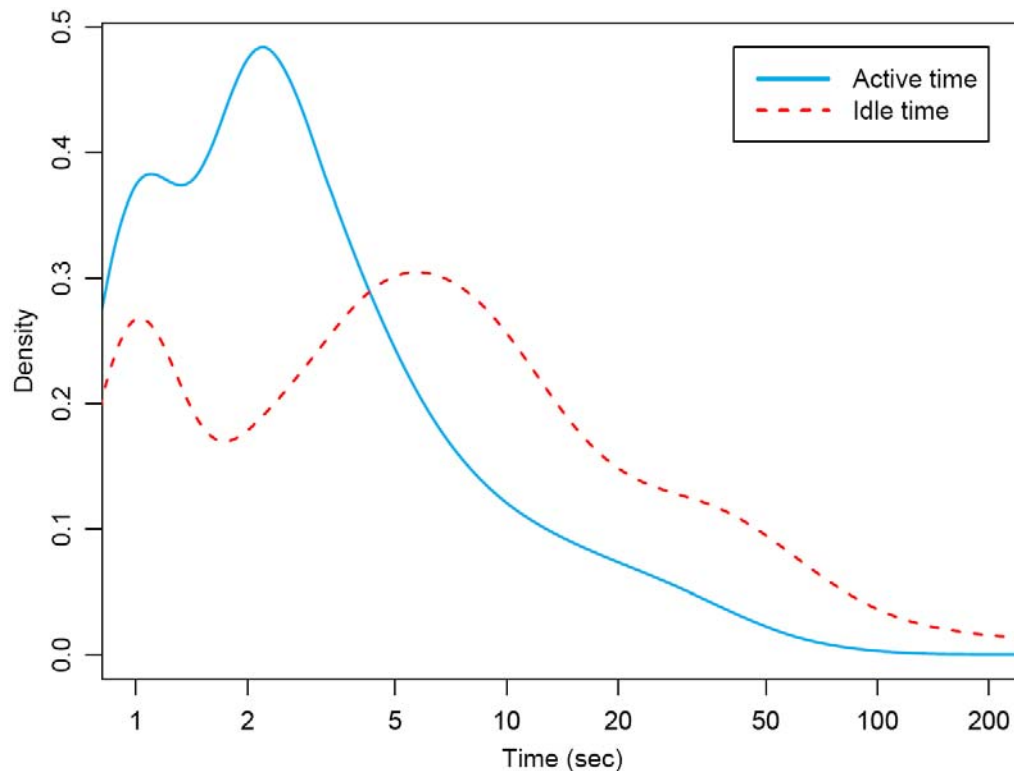


圖 3 所有玩家的平均活動時間 vs. 平均閒置時間

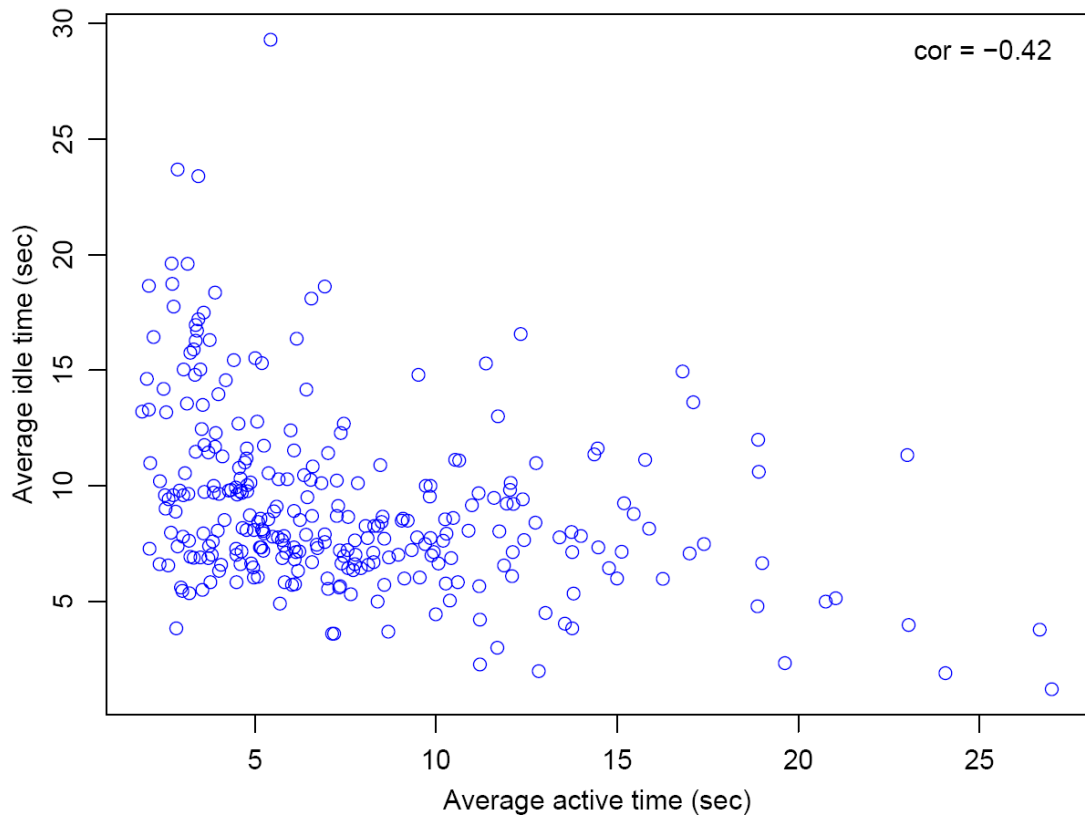
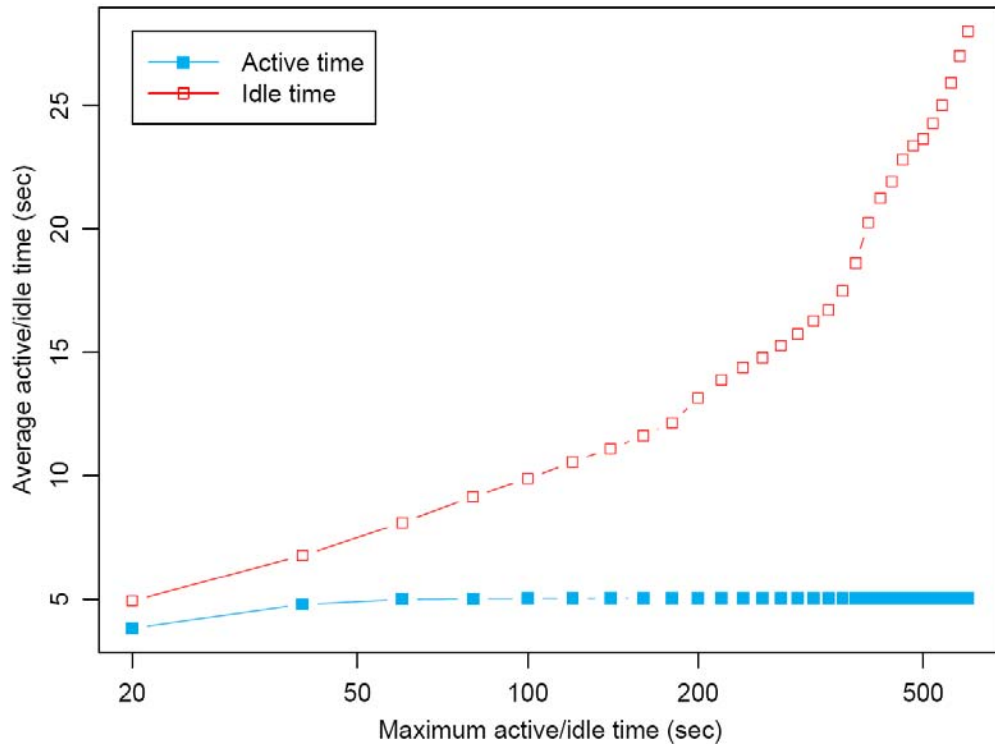


圖 4 隨著最大活動/閒置時間界限的變動，平均活動時間以及閒置時間的變化



## 2.3 活動行為以及閒置行為的比較

我們認為**活動時間**以及**閒置時間**為玩家的遊戲行為特徵。但兩者有著極高的相關性，或許其中之一就能夠充分反應玩家的行為。我們認為閒置時間比活動時間更能夠代表玩家的遊戲行為特徵，理由如下：

- 1) 閒置時間的分佈擁有較多的**變異性**。如圖 4 所示，若我們逐漸增加活動時間及閒置時間的最大界限，活動時間的平均長度將收斂到 5 秒左右，然而閒置時間則是持續地增加。這是合理的，因為活動時間的長度會受到遊戲本身所限制，例如，玩家在進行移動以外的行為時必須先讓虛擬角色停下來。同時也會受到人體上的限制，例如，玩家不可能一直持續進行活動而完全不停下來休息。相對之下，只要玩家沒有下線就離開電腦，閒置時間就可能相當地長。因此，閒置時間並不會受到上述情況的限制，而可能擷取到更多玩家的遊戲行為特徵。
- 2) 閒置時間有較小的**自我相關性**。我們將活動時間及閒置時間的第 1 期差及第 2 期差的自我相關係數畫在圖 5。結果顯示出大部分玩家的活動時間有些許的正自我相關性。相對地，大部分玩家的閒置時間有著非常微弱甚至近乎無的自我相關性，這意味閒置時間較不受玩家處於不同狀況時的影響，較有隨機性。因此，閒置時間較能擷取到玩家穩定的特性。

## 2.4 以閒置時間分佈做為玩家特徵

我們接著檢驗閒置時間分佈 (ITD) 是否能夠區分玩家的不同。圖 6 顯示 9 位隨機抽取的玩家閒置時間分佈。可以看到這 9 位玩家的閒置時間分佈都不盡相同。其中的差異不單只是集中在分佈的中央，包含整體分佈的形狀。其中，玩家編號 4、13、19、26、29 有著**右偏**的 ITD；玩家編號 6、27 有**多個高峰** (資料集中於多個的地方)，其他玩家則為**單峰** (資料只集中於一處)。很明顯地，沒有兩個玩家擁有完全相同的 ITD。這驗證了我們的推測，也就是說，玩家的遊戲行為特徵能夠擷取玩家個人獨有的習慣、特性，可處理使用者辨識的問題。

圖 5 玩家活動時間及閒置時間的自我相關係數

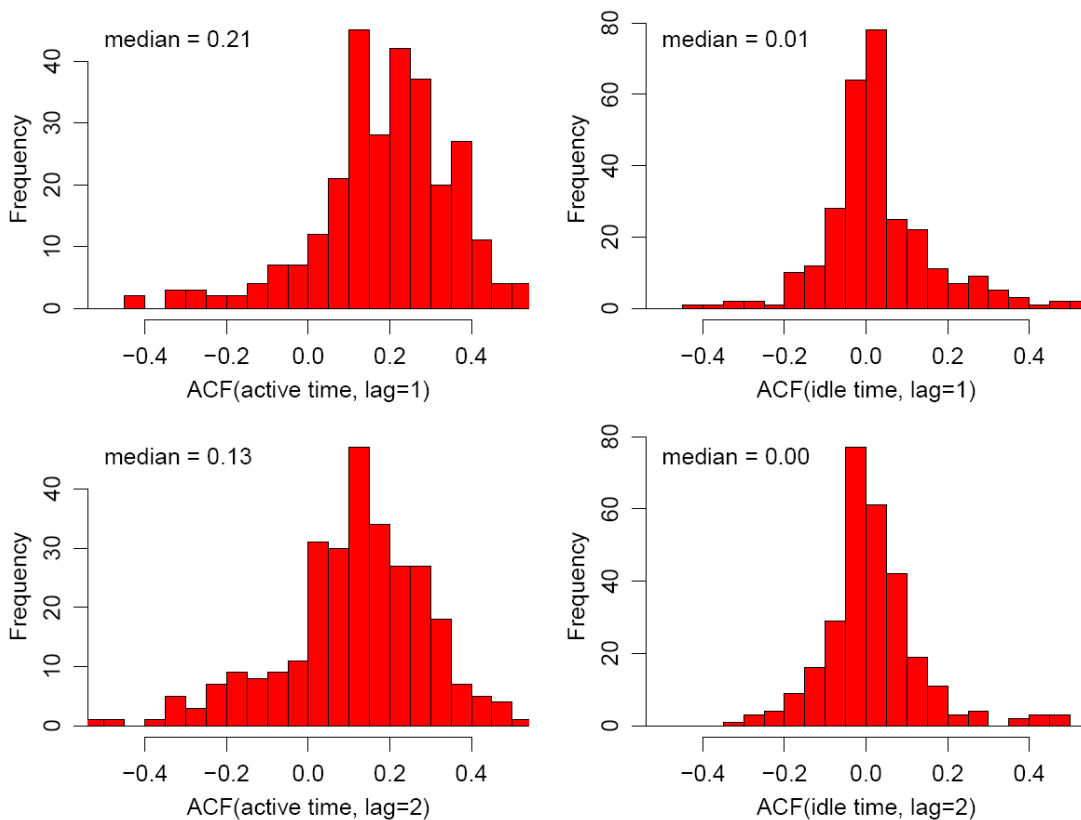
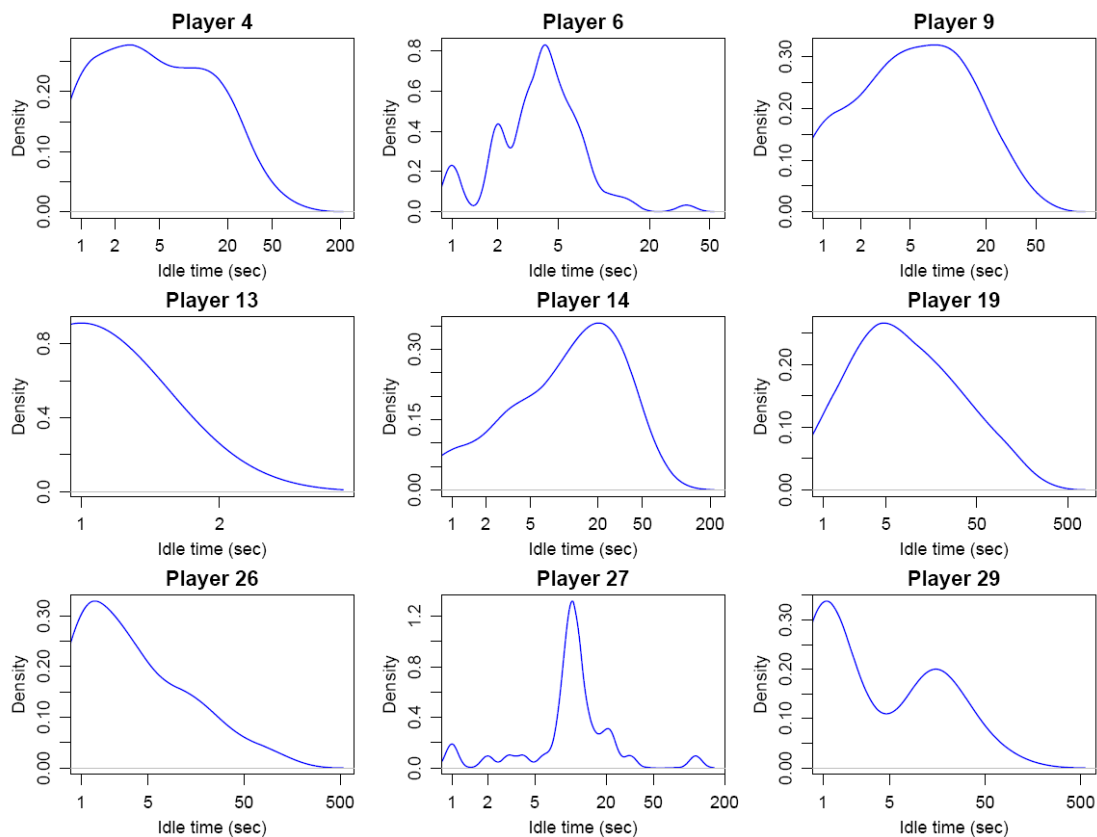


圖 6 隨機抽取的玩家閒置時間分佈



### 3. 使用者辨識方法

我們基於 2.4 節的發現，提出 relative entropy test (RET) 法，此方法基於兩個閒置時間分佈之間的差異。其基本精神為利用 Kullback-Leibler 距離計算出兩個機率分佈之間的距離。對於機率分佈  $P$  及機率分佈  $Q$ ，從  $P$  到  $Q$  的 KL 距離定義為

$$D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

為了符合距離的直觀性質—對稱，即  $P$  到  $Q$  的距離等同於  $Q$  到  $P$  的距離，我們定義「 $P$  到  $Q$  的距離」加上「 $Q$  到  $P$  的距離」為  $P$   $Q$  之間的距離，使 KL 距離有對稱的性質。為求表達上的方便，以下所提及 KL 距離皆有對稱的性質。

要注意的是，KL 距離的強弱（遠近）除了  $P = Q$ （KL 距離等於 0）只能做**相對的比較**，KL 距離不能被用來量測兩個分佈到底「有多接近」。只能比較兩個 KL 距離的大小。例如，若  $D_{KL}(P, Q) < D_{KL}(Q, R)$ ，可知  $P$  比  $R$  更接近  $Q$ 。

RET 法即為利用閒置時間分佈的 KL 距離來達到使用者辨識的目的。我們認為玩家隨著時間的不同，所進行的遊戲行為（ITD）不可能完全一致，也就是有所謂的「**變異性**」，但這些不同時間的 ITD 應該會擁有類似的特性（因為來自相同的玩家）。因為機率分佈的中位數為較穩健的趨中量數估計量，所以我們以中位數來描述此特性，也就是說，若 ITD 之間的 KL 距離差異無法收斂到固定的中位數，則 RET 法會檢測出此兩組 ITD 來自不同的兩位玩家。接下來，我們將就**一致性**以及**辨別力**的部分討論 RET 法的演算法以及其效度。

#### 3.1 一致性

假設一位玩家在有  $n$  組的 ITD。我們隨機從中挑選  $k$  組，並且計算  $C(k, 2)$  組 KL 距離，可得到 KL 距離的分佈（為求方便，以 KLD 表示 ITD 之間的 KL 距離分佈）。此流程將會進行  $m = \lfloor n/k \rfloor$  次，因此可得到  $m$  個 KLD。舉例而言，我們將四位玩家的 KLD 畫在圖 7。由圖可見，同一位玩家的 KLD 有類似的形狀並集中在**相同的中位數**。另外，不同玩家之間的 KLD 形狀也不相同。

為了檢測一致性（能否判斷資料來自同一位玩家），我們設定  $k = \lfloor n/2 \rfloor$  使得  $m = 2$ ，即每位玩家有 2 個 KLD。接著，我們用 *Wilcoxon* 假設檢定檢驗同一位玩家的兩個 KLD 是否擁有相同的中位數。此假設檢定為一**無母數**設定的假設檢定，用以檢測兩個分佈是否顯著地有相同的中位數。使用**雙尾**及 **0.05 顯著水準**的設定，並以玩家通過假設檢定的比例做為**正確率**（通過假設檢定的意思代表兩個 KLD 有相同的中位數，則判斷兩筆資料來自相同玩家）。

圖 8 為正確率隨著觀察時間長度變化的情形。顯示出不論觀察時間多長，正確率皆普遍高於 **0.95**，同時驗證了 RET 法有著不錯的一致性性質。

圖 7 在不同的時間區段之下，玩家閒置時間分佈的 KL 距離

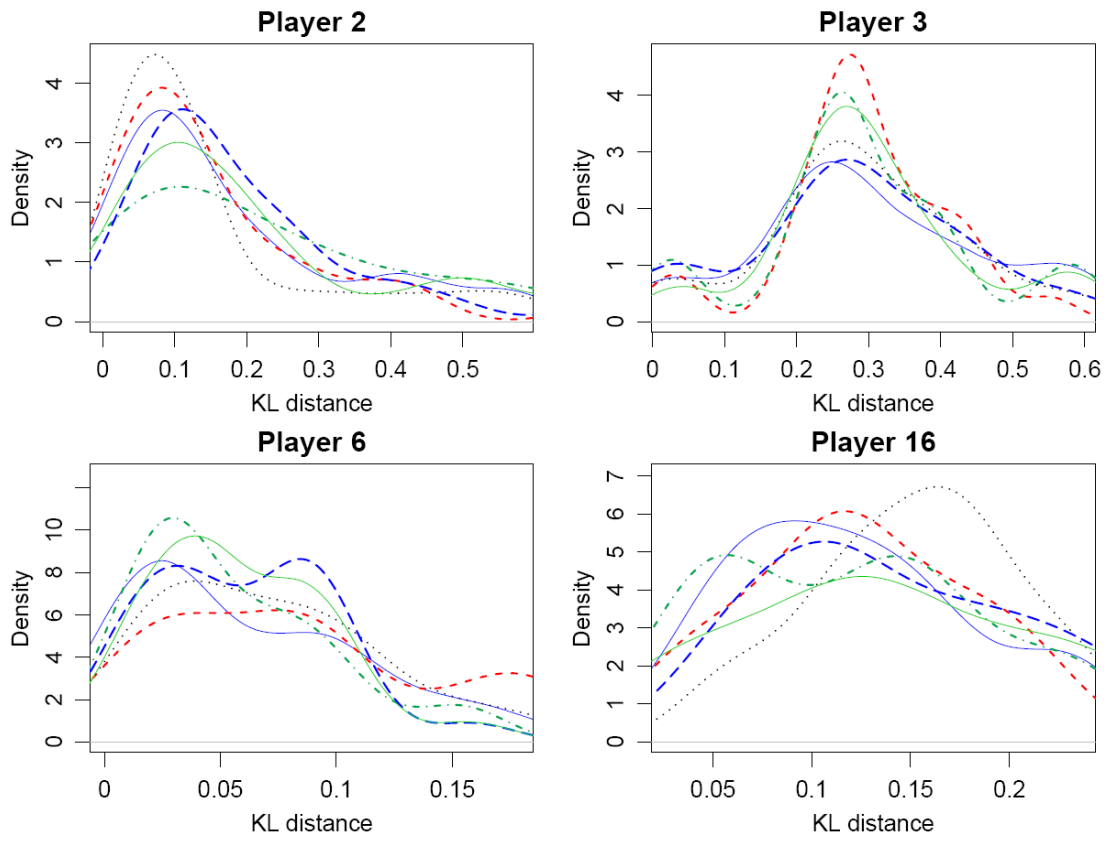
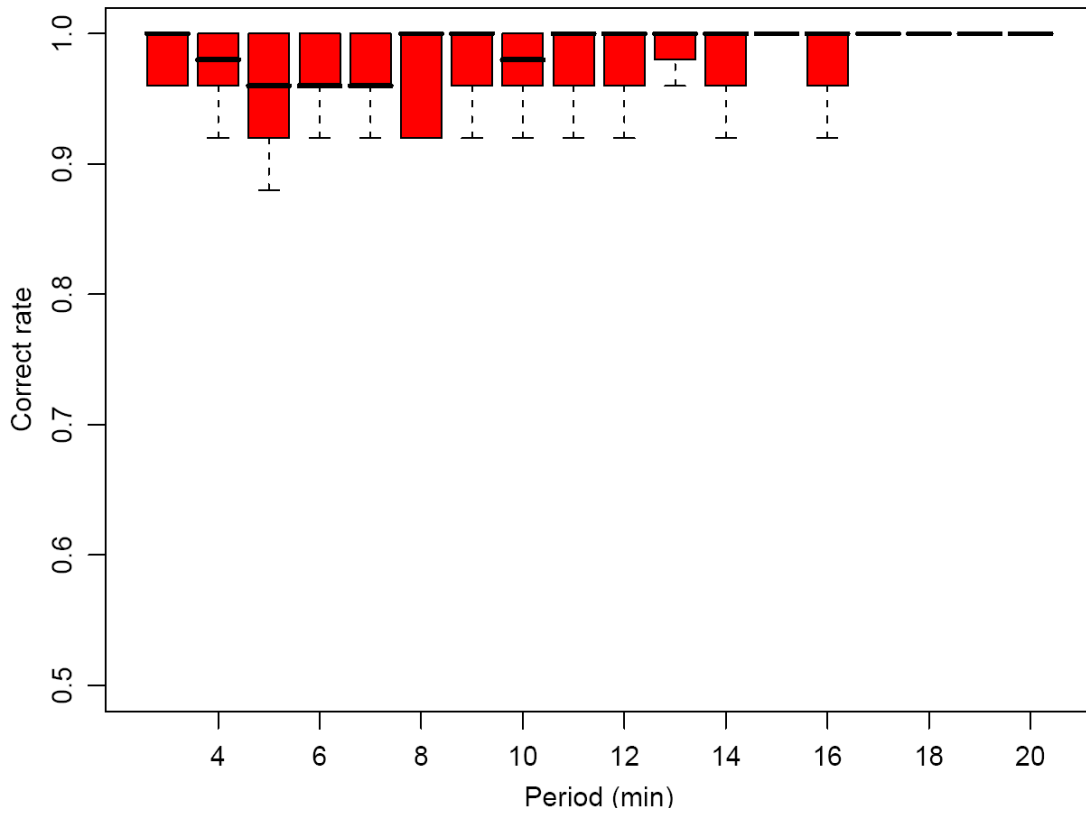


圖 8 一致性檢定，RET 法的正確率

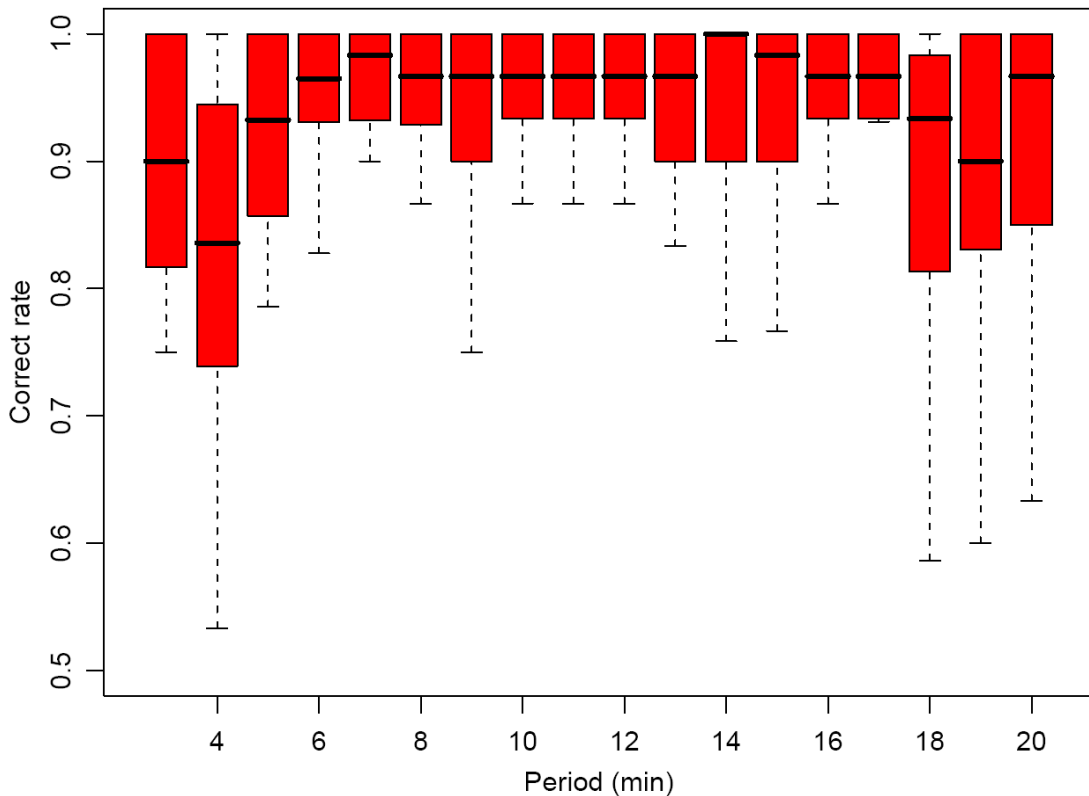


### 3.2 辨別力

接著檢驗 RET 法的**辨別力** ( 能否正確區分不同玩家 )。假設有  $N$  位玩家，第  $i$  位玩家有  $n_i$  個 ITD，我們計算第  $i$  位玩家的  $n_i$  個 ITD 以及第  $j$  位玩家的  $n_j$  個 ITD 之間的  $n_i n_j$  個 KL 距離，並以  $KLD_{i,j}$  表示其分佈。以**單尾**的 Wilcoxon 假設檢定檢驗玩家  $i$  與玩家  $j$ ，**虛無假設**為  $KLD_{i,j}$  的中位數較  $KLD_{i,i}$  大。**正確率**定義為  $C(N, 2)$  個假設檢定中，有多少比例接受了檢定。

如圖 9 所示，我們發現辨別力的正確率不受時間長度影響普遍地高 ( 比 0.9 高 )。結合圖 8 的結果，RET 法在處理辨識使用者的問題上，不論是一致性或是辨別力都有不錯的結果。

圖 9 辨別力檢定，RET 法的正確率



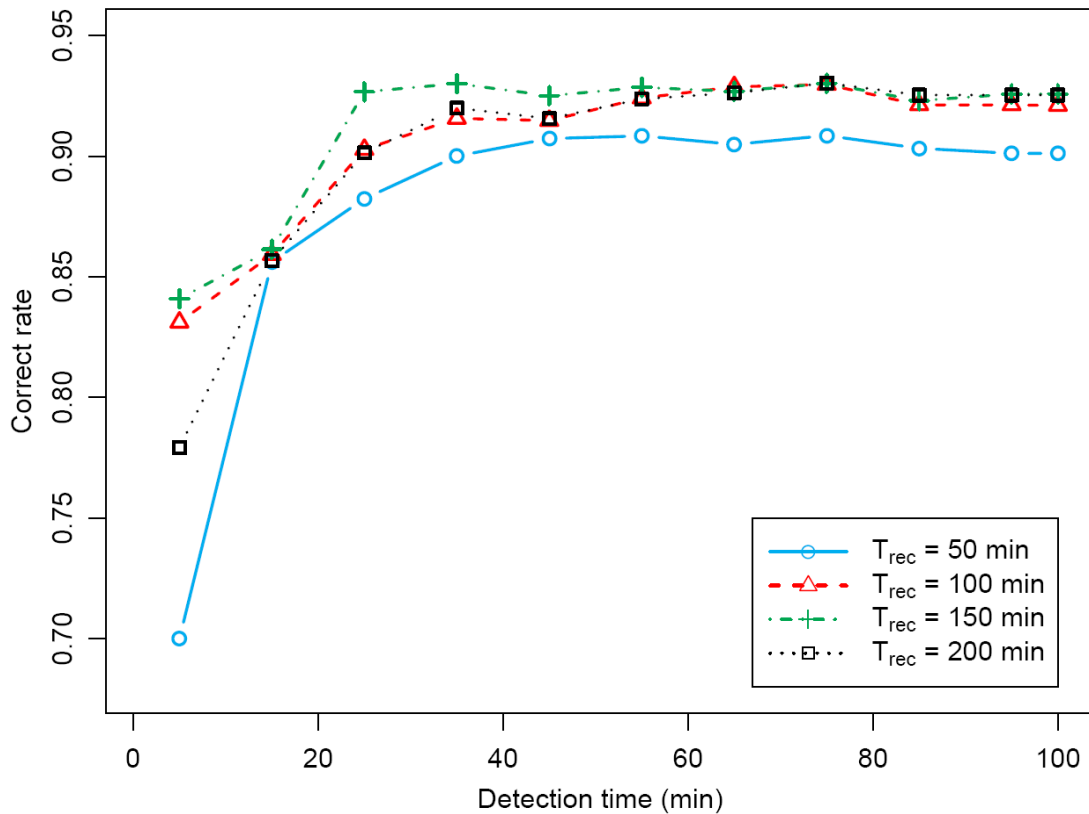
### 4. 效果評估

在前一節當中，我們證明了 RET 法在一致性以及辨別力方面皆有不錯的結果。在本節中，探討不同的**偵測時間**及不同的**歷史資料**長度會對 RET 法有何影響。我們假設遊戲系統觀察玩家  $T_{rec}$  分鐘並儲存玩家在這期間的閒置時間做為歷史資料。當此玩家重新登入  $T_{obs}$  分後即執行使用者辨識的程序。其中兩個重要的因子  $T_{rec}$  及  $T_{obs}$  扮演的角色如下：

- $T_{rec}$  決定資料庫中保留多長的玩家歷史資料。根據收集到的資料顯示，一個行為循環（一個活動時段以及一個閒置時段的組合）平均約為一分鐘，因此  $T_{rec}$  的最大值將會受到其中包含的閒置時段數量的影響。若我們有一百萬的玩家， $T_{rec}$  為 200 分鐘，而每段閒置時間需要 4 bytes 的空間儲存，則總共需要 800MB 的儲存空間。
- $T_{obs}$  決定一位新加入的玩家在多短的時間會再次被檢驗其身分（根據玩遊戲的行為）。 $T_{obs}$  有兩個缺點：
  - 1) 若目前的帳號正被駭客所使用，在時間未超過  $T_{obs}$  之前系統無法偵測出此帳號正遭到盜用。
  - 2) 閒置時間必須儲存於主記憶體之中，這對於系統而言將會造成不小負擔。假設同時一萬名玩家在線上，並且  $T_{obs} = 30$  分，則大約需要 1.2MB 的主記憶體來儲存所有玩家在此偵測時間內的閒置時間。

我們將不同的偵測時間以及不同的歷史資料長度，相對於 RET 法的辨識正確率畫於圖 10。如圖中所示，較大的歷史資料以及較長的偵測時間都能夠增加辨識的正確率。一般來說我們會傾向採用較小的  $T_{obs}$ ，因為若偵測時間過長，有可能會錯失掉機會辨識出遊戲時間較短的玩家們。因此我們可能會選擇固定歷史資料為 200 分鐘，如此在 20 分鐘長的偵測時間即可達到高於 0.9 的辨識正確率。

圖 10 在不同的歷史資料大小以及偵測時間長度的組合下，RET 法的成效



## 5. 結論

在本文中，我們提出利用使用者辨識技術來解決線上遊戲帳號盜用及帳號共享的問題。我們的貢獻有以下三點：

- 1) 提出利用玩家的遊戲行為特徵處理使用者辨識問題的概念。
- 2) 證明線上遊戲虛擬角色閒置時間分佈可做為玩家的行為特徵。
- 3) 提出 RET 法進行使用者辨識。

評估結果顯示，基於 200 分鐘的歷史資料，RET 法能夠在 20 分鐘達到 90% 辨識準確率。雖然只要觀察玩家 20 分鐘，本方法即有不錯的效果，但在實際情況下，駭客不太可能會在線上待這麼久的時間，而且或許只需幾秒鐘即可盜走值錢的虛擬道具。我們相信利用更多的玩家行為，例如滑鼠或鍵盤的操作方式，結合本方法使用的閒置時間分佈，能有更佳的偵測效果。

帳號的盜用，乍看之下「竊賊」偷走的虛擬道具無傷大雅，但是實際上卻對許多玩家而言是不可取代的寶物，是意義非凡的「象徵」，例如，跟夥伴們一同打敗魔王得到的裝備、好朋友送的道具、辛苦收集到的道具材料。一旦玩家遭遇駭客的洗劫，有可能因為失去這些象徵，而喪失繼續玩遊戲的動力，而選擇永遠離開遊戲世界。從遊戲公司

的角度，若無法處理帳號盜用的問題，玩家將無法感到安全感，又怎麼可能死忠於該公司的遊戲呢？這無疑等於是放任玩家不斷地流失。無論對於玩家或遊戲廠商，帳號盜用都是迫切且急需解決的問題。我們希望能夠以本研究為一個出發點，繼續在這個問題上努力不懈，也希望能夠藉由學術界與遊戲產業界的無間合作，在不久的將來提供給玩家更優質、安全的遊戲環境。