The Tangled Web of Password Reuse

R07944058
網媒碩一 陳鵬宇

1 Summary

這篇論文主要想探討在網路世界中，使用者傾向重複使用相同或是近似的密碼，因為要記住長密碼不大容易，因此運用一組已知的密碼再去推算其它密碼是很常見的，並且論文也對一些洩漏的密碼庫做了一些分析，發現了有將近 43 至 51 % 的使用者會在不同的網站使用一樣的密碼。文中也透過各種不用網站、各項實驗數據去列出了存取的資料集是 $Hash(plaintext)$ 或是單純的 $plaintext$、使用者數量的規模大小、網站類型 (社交、部落格、電子信箱、購物、金融)。論文在最後也對了一些常用的密碼轉換方式做了討論，其實就是字串操作，像是 edit distance 的插入、刪除、轉換等等，並且發展出了一套跨網站的密碼猜測演算法，經實驗可以在 100 次嘗試中達到 30% 的成功率，大勝了傳統密碼猜測演算法的 14%。

2 Strength(s)

論文很詳細的介紹了使用者選擇密碼的背景和一些網站為了使用者安全性所訂定的政策，例如密碼不得含有使用者相關資訊、密碼長度應至少大於多少字元、是否有大小寫字元等等。現今已經有許多可以取代純文字的密碼登入方式，舉凡：生物辨識 (臉部辯試、指紋辨識)、公鑰認證、token、二階段認證等方式，但純文字還是現在的主流，並且也不需要額外的硬體支援，論文特別針對了許多不用的資料集進行分析，論文有提到為什麼這些資料集是受到了哪些攻擊以至於資料外洩，大多是 SQL injection、釣魚等等。另外論文一個很棒的地方是，他運量了大量的折線圖和表格，並且在分類部分採用條列式的撰寫方式，這幫助了讀者快速理解統計數據。

3 Weakness(s)

使用者為了記憶方便，通常不會擁有太多的密碼，這裡指的「擁有」是，在不同網站中使用了幾種不同種類的密碼，透過調查洩漏出來的密碼，我們發現有 97.75% 的使用者只使用了 2 種密碼，這是一個很明顯的警訊，而且使用者在不同的網站傾向使用不同複雜度的密碼，在與金錢有關的網站容易使用比較複雜的密碼，這和有些網站會強迫用户要遵循一定「較複雜」的政策很像，反過來看，若這些政策不夠完善，某種程度也是告訴了駭客能從這些政策下去破解密碼。

4 Reflection

這篇論文滿多地方都是在做實驗數據上的探討，比較少仔細地去講方法上的細解，只有透過簡單的例子來讓讀者看前後比較的差異，雖然宣稱他們的密碼猜測演算法會比較傳統方法更好，但其實從 pseudo code 中不難發現，他也只是將傳統字串操作方式做些拼湊罷了。本文比較像是導讀，當中整理了許多密碼重複使用的情況，也讓我們知道人性是懶惰的，當然如何一方面讓使用者不用人腦記憶太過複雜的密碼，同時又能保有一定的安全性，是很值得探討的，像生物辨識就是一個還不錯的解決方式，但他仰賴了一定程度的硬體裝置，同時如果是雙胞胎的話，還是有可能在現今流行的臉部辨識騙過電腦。透過增加密碼強度，是能提高計算上時間複雜度，但這始終不是一個根治的方法，而硬體也會隨著時間成長。目前有些公司採用了更為安全的二階段驗證，像 Google、GitHub，兩者採用的方式稍有不同，Google 是透過 SMS，用户必需在手機上確認一次性的登入來驗證身份，而 GitHub 則是可採用 SMS 或是管理密碼專用的密碼本應用程式來驗證。如何在網路這個世界訂出一套合理的遊戲規則來驗證個人身份，還是有許多發展空間的。