Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0

為什麼約翰尼不能加密：PGP 5.0 的可用性評估

**Abstract**

用戶錯誤會導致或促成大多數計算機安全故障，但用於安全的用戶界面仍然往往笨拙、混亂或幾乎不存在。這僅僅是由於未能將標準用戶界面設計技術應用於安全性嗎？我們認為，相反，有效的安全性需要不同的可用性標準，並且不會通過適用於其他類型消費軟件的用戶界面設計技術來實現。

為了驗證這個假設，我們對一個安全程序進行了案例研究，該程序確實具有按照通用標準的良好用戶界面：PGP 5.0。我們的案例研究使用認知演練分析和實驗室用戶測試來評估 PGP 5.0密碼學新手可以成功地使用它來實現有效的電子郵件安全。分析發現了許多可能導致安全失敗的用戶界面設計缺陷，用戶測試表明，當我們的測試參與者有 90 分鐘的時間使用 PGP 5.0 對消息進行簽名和加密時，他們中的大多數人無法這樣做成功。

我們得出結論，儘管 PGP 5.0 具有吸引人的圖形用戶界面，但它不足以為大多數計算機用戶提供有效的安全性，這支持了我們的假設，即有效安全性的用戶界面設計仍然是一個懸而未決的問題。最後，我們簡要介紹了我們在用戶界面設計原則和安全技術的開發和應用方面的持續工作。

**2 理解問題**

**2.1 為安全定義可用性**

可用性在不同的上下文中必然具有不同的含義。 對一些人來說，效率可能是一個優先事項，對另一些人來說，是可學習性，對另一些人來說，是靈活性。 在安全環境中，我們的優先級必須是任何需要的東西，才能有效地使用安全性。 我們在下面的定義中捕獲了這組優先級。

定義：安全軟件是可用的，如果預期使用它的人：

1. 可靠地了解他們需要執行的安全任務；

2. 能夠弄清楚如何成功地執行這些任務；

3.不要犯危險的錯誤； 和

4. 對界面足夠舒適以繼續使用它。

**2.2 安全的問題屬性**

安全性具有一些固有屬性，使其成為用戶界面設計的難題。 用於創建可用安全性的設計策略需要明確考慮這些屬性，而通用用戶界面設計不會這樣做。 我們在這裡描述了五個這樣的屬性； 可能還有其他我們尚未確定的。

1. 無動力的用戶屬性

安全通常是次要目標。 人們通常不會坐在他們的電腦前想要管理他們的安全。 相反，他們想要發送電子郵件、瀏覽網頁或下載軟件，並且他們希望在他們做這些事情時有適當的安全措施來保護他們。 人們很容易推遲學習安全性，或者樂觀地假設他們的安全性正在發揮作用，而他們專注於他們的主要目標。 安全用戶界面的設計者不應假設用戶會被激勵去閱讀手冊或去尋找設計為不引人注目的安全控制。 此外，如果安全性太難或太煩人，用戶可能會完全放棄它。

2.抽象屬性

計算機安全管理通常涉及安全策略，它們是用於決定是否授予對資源的訪問權限的抽象規則係統。 此類規則的創建和管理是程序員認為理所當然的活動，但對於更廣泛的用戶群體中的許多成員來說，這可能是陌生且不直觀的。 安全用戶界面設計需要考慮到這一點。

3. 缺乏反饋屬性

防止危險錯誤的需要使得向用戶提供良好的反饋勢在必行，但為安全管理提供良好的反饋是一個難題。 安全配置的狀態通常是複雜的，試圖總結它是不夠的。

此外，正確的安全配置是做用戶“真正想要的”的，因為只有用戶知道那是什麼，所以安全軟件很難執行很多有用的錯誤檢查。

4.穀倉門屬性 The barn door property

關於馬走後鎖上穀倉門是徒勞的諺語描述了計算機安全的一個重要特性：一旦一個秘密被意外遺漏，即使

短時間內，無法確定攻擊者尚未讀取它。 正因為如此，安全用戶界面設計需要非常重視確保用戶充分了解他們的安全性，以免犯下潛在的高成本錯誤。

5.最薄弱環節屬性

眾所周知，聯網計算機的安全性與其最薄弱的組件一樣強。 如果破解者可以利用一個錯誤，那麼遊戲就結束了。 這意味著需要引導用戶注意其安全的各個方面，而不是像使用文字處理器或電子表格那樣進行隨機探索。

**3 評價方法**

我們選擇通過兩種方法評估 PGP 的可用性：一種非正式的認知演練 [17]，其中我們直接審查了 PGP 的用戶界面，並指出其設計中未能滿足第 2.3 節中描述的可用性標準的方面； 以及在實驗室中進行的用戶測試 [15]，測試參與者被選為能夠合理代表電子郵件用戶的一般人群。 兩種方法固有的優點和缺點使它們在不同的方面有用，我們將它們視為互補的評估策略 [7] 比嘗試使用實驗室測試直接驗證提出的觀點更為現實 通過認知演練。

認知演練是一種可用性評估技術，模仿了代碼演練的軟件工程實踐。 為了進行認知演練，評估人員像新手用戶一樣逐步使用軟件，試圖在心理上模擬他們認為新手對軟件的理解在每個點上的情況，並尋找可能的錯誤和區域 困惑。 作為一種評估工具，認知演練傾向於關注用戶界面的可學習性（而不是效率），因此它是評估安全可用性的合適工具

儘管我們的分析被最準確地描述為認知演練，但它也包含了另一種技術的各個方面，即啟發式評估 [11]。 在這種技術中，用戶界面是根據一個特定的高優先級可用性原則列表進行評估的； 我們的原則列表由我們在第 2.1 節中給出的安全可用性定義及其在第 2.3 節中專門針對 PGP 的重述組成。 啟發式評估最好由“雙重專家“double expert”進行，他們非常熟悉應用程序領域以及可用性技術和要求（包括了解預期使用該軟件的人的技能、思維方式和背景）。 我們的評估借鑒了我們作為安全研究人員的經驗，以及在培訓和輔導新手計算機用戶以及戲劇、人類學和心理學方面的額外背景。

使可用性安全性設計成為一個困難和專門問題的一些相同屬性也使測試安全性的可用性成為一項具有挑戰性的任務。要進行用戶測試，我們必須要求參與者使用該軟件來執行一些任務，其中包括使用安全性。但是，如果我們提示他們直接執行安全任務，而在現實生活中他們可能沒有意識到該任務，那麼我們就無法測試該軟件是否設計得足夠好，以便在他們需要時給予他們這種意識.此外，為了測試他們是否能夠在需要時弄清楚如何使用安全性，我們必須確保測試場景為他們提供了一些他們認為值得保護的秘密，與我們期望他們賦予的價值相當自己在現實世界中的秘密。設計測試充分考慮到這些要求是必須小心完成的事情，除了一些關於測試警告標籤有效性的工作[19]，我們發現很少有現有的關於用戶測試的材料可以解決類似的問題。

**5 用戶測試**

**5.1 目的**

我們的用戶測試旨在評估 PGP 5.0 是否符合第 2.3 節中描述的特定可用性標準。 我們為參與者提供了一個既合理又適當激勵的測試場景，然後避免干擾他們執行我們給他們的安全任務的嘗試。

**5.2 說明**

**5.2.1 測試設計**

我們的測試場景是，參與者自願幫助進行政治競選，並被分配了競選協調員的工作（黨派和競選問題留給了參與者的想像力，以免冒犯任何人）。參與者的任務是通過電子郵件將活動計劃更新發送給活動團隊的其他成員，使用 PGP 進行隱私和身份驗證。由於可能自願參加政治競選意味著個人對競選成功的投資，因此我們希望參與者能夠受到適當的激勵，以保護他們信息的機密性。

由於 PGP 本身不處理電子郵件，因此有必要為參與者提供一個電子郵件處理程序以供使用。我們選擇給他們 Eudora，因為這樣我們也可以評估

PGP 中包含的 Eudora 插件的成功。由於我們對測試 Eudora 的可用性（除了 PGP 插件）不感興趣，因此我們在開始測試之前為參與者提供了一個簡短的 Eudora 教程，並在測試期間如果參與者被卡在某些東西上，我們會提供幫助。與PGP無關。

在向參與者簡要介紹了測試場景並指導他們使用 Eudora 之後，他們獲得了初始任務描述，其中向他們提供了一條秘密消息（候選人的建議行程）、競選經理的姓名和電子郵件地址以及其他四名競選團隊成員，並請求通過簽名和加密的電子郵件將秘密消息發送給五名團隊成員。為了完成這個任務，參與者必須生成一個密鑰對，獲取團隊成員的公鑰，讓團隊成員可以使用他們自己的公鑰，在電子郵件中輸入（短）秘密消息，使用他們的私鑰，使用五個團隊成員的公鑰加密電子郵件，然後發送結果。此外，我們設計了測試，使得其中一名團隊成員擁有 RSA 密鑰，而其他人都擁有 DiffieHellman/DSS 密鑰，因此如果參與者為所有五名團隊成員加密了一份消息副本（這是預期的解釋的任務），他們會遇到混合鍵類型警告消息。參與者被告知，在完成初始任務後，他們應該等待收到來自競選團隊成員的電子郵件並遵循他們給出的任何指示。

五個競選團隊成員中的每一個都由一個虛擬電子郵件帳戶和一個密鑰對代表，測試監視器可以通過聯網的筆記本電腦訪問這些密鑰對。競選經理的私鑰被用來簽署每個團隊成員的公鑰，包括她自己的，所有五個簽名的公鑰都放在麻省理工學院的默認密鑰服務器上，以便參與者請求檢索它們。

在某些情況下，測試監視器會偽裝成競選團隊的成員，並從適當的虛擬帳戶向參與者發送電子郵件。這些情況是：

1. 參與者向該團隊成員發送電子郵件，詢問有關如何做某事的問題。在這種情況下，測試監視器發送了與測試場景一致的最少信息回复，即不會使團隊成員看起來充滿敵意或無知超出合理範圍的最少回答。

2. 參與者通過明文電子郵件發送秘密。然後測試監控員冒充活動經理髮送電子郵件，告訴參與者發生了什麼，強調使用加密保護秘密的重要性，並要求參與者在繼續之前嘗試發送加密的測試電子郵件。如果參與者成功地這樣做了，測試監視器（冒充活動經理）然後通過加密電子郵件向參與者發送更新的秘密，測試從頭開始。

**6。結論**

**6.1 標準接口設計失敗**

在我們的案例研究中看到的結果支持了我們的假設，即此處以 PGP 5.0 為代表的用戶界面設計標準模型不足以使計算機安全性可供尚未熟悉該領域的人使用。我們的 12 名測試參與者普遍接受過電子郵件的教育和經驗，但只有三分之一的人能夠在 90 分鐘內使用 PGP 5.0 正確簽署和加密電子郵件消息。此外，其中四分之一的人意外暴露了他們在此過程中要保護的秘密，他們通過電子郵件發送了他們認為已經加密但沒有加密的信息。

在第 2.1 節中，我們根據四個必要的品質定義了安全可用性，這些品質直接轉化為設計優先級。 PGP 5.0 的用戶界面無法實現有效的安全性，因為它沒有按照這些優先級進行設計：測試參與者對公鑰模型的理解不夠充分，以至於他們知道他們必須為他們希望向其發送安全電子郵件的人獲取公鑰;許多知道他們需要獲取密鑰或加密的人仍然很難弄清楚如何去做；有些人錯誤地以明文形式發送秘密，認為它們已經加密；許多人對嘗試使用 PGP 5.0 的經歷表示沮喪和不滿，以至於他們不太可能在現實世界中繼續使用它。

儘管 PGP 5.0 很有吸引力，但所有這些失敗都是由帶有標籤和圖標的按鈕整齊地表示的，其餘的都是下拉菜單，儘管對於那些已經了解它的人來說使用起來很簡單。公鑰密碼學和基於數字簽名的信任的基本模型。因此，設計對那些尚未了解它的人來說足夠有用的安全性必須需要更多的東西。

**6.2 安全可用性評估**

由於可用安全性要求用戶界面設計優先級與一般消費軟件不同，因此它同樣需要適用於測試這些優先級是否已充分實現的可用性評估方法。簡單應用的標準可用性評估方法可能會將安全功能視為用戶的主要目標而不是次要目標，從而導致錯誤的結論。在安全環境中進行可用性評估的公共工作將非常有價值，並且幾乎肯定必須來自研究資源，因為軟件開發人員並不急於公開他們在自己的產品中發現的可用性缺陷。

在我們自己的工作中，它專注於對個人計算機的初步了解很少的用戶安全性，我們對可學習性賦予了很高的價值，因此發現認知演練是一種自然評價技術。其他技術可能更適合企業或軍事用戶，但可能需要對適合安全的優先級進行類似的調整。在設計適當的用戶測試時，著眼於對消費者安全負有既定責任的其他領域可能是有價值的；這些領域更有可能對如何最好地確定產品設計是否成功促進安全使用模式進行了大量研究。

**6.3 走向更好的設計策略**

我們案例研究中的詳細發現提出了幾種設計策略，以提高可用的安全性，這是我們在正在進行的工作中所追求的。首先，很明顯需要盡快將準確的安全概念模型傳達給用戶。概念模型越小越簡單，我們就越有可能成功地做到這一點。因此，我們正在研究將安全功能縮減為真正必要且適合特定人群需求的實用方法，同時又不犧牲提供給用戶的安全完整性。

在建立了一個最小但有效的安全概念模型之後，它必須比其他類型軟件的概念模型更快更有效地傳達給用戶。我們正在研究實現這一目標的幾種策略，包括精心設計界面隱喻以以更高要求的準確性匹配安全功能的可能性。

此外，我們正在尋找當前對教育軟件的研究，以了解如何最好地指導用戶學習管理他們的安全性。我們不相信家庭用戶可以配合大量的教程，但我們正在研究更溫和的方法，以便在正確的時間為用戶提供正確的指導，包括如何最好地利用警告消息、嚮導和其他交互式工具.

**七、相關工作**

迄今為止，我們幾乎沒有發現關於安全可用性問題的已發表研究。 在確實存在的東西中，最突出的例子是 Adage 項目 [12, 20]，它被描述為一個旨在處理分佈式應用程序和組的授權策略的系統。 可用性是 Adage 的一個主要設計目標，但它旨在供已經擁有高水平專業知識的專業系統管理員使用，因此它沒有解決使更一般人群有效使用安全性所帶來的問題。 對安全關鍵系統 [10] 的可用性相關問題也進行了研究，例如控制飛機或製造工廠的系統，但我們可能希望，與個人計算機安全用戶不同，這些系統的用戶將經過仔細選擇和 受過訓練。

Ross Anderson 在 [1] 中討論了用戶不合規對安全性的影響，Don Davis 在 [3] 中分析了基於公鑰的安全系統通常對用戶施加的不切實際的期望。

除此之外，我們只知道一篇關於數據庫身份驗證例程可用性測試的論文 [8]，以及對計算機支持的協作工作中固有的安全和隱私問題的一些簡要討論 [16]。 John Howard 的論文 [6] 對 1989 年至 1995 年間向 CERT5 報告的安全事件進行了有趣的分析，但更多地關注攻擊的類型，而不是這些攻擊所利用的漏洞的原因，並且僅代表足夠複雜的實體所經歷的事件向 CERT 報告。