CCS'17-Seiler-Hwang-"I don't see why I would ever want to use it": Analyzing the Usability of Popular Smartphone Password Managers

优点：①首个针对移动端PM可用性的调研

②本文对讨论问题的分析和建议的论述比较好

问题：①在实验中应当使用最新(稳定)版本的OS和PM，否则对于一个应用的多版本研究不够客观和科学.

②应充分调研关于其它APP的可用性的研究，探究可用性问题是PM特有的还是普遍存在的 ③需要对CCS和WWW的in situ研究更准确地研究可用性

▲研究背景

在移动端使用文本吃不方便(屏幕小，键盘是多层布局)，PM可以帮助提供可用性，但采用率不高，目前还没有针对于移动端PM的工作(在移动端交互的最佳实践和准则不同)

▲研究问题

1> ①对移动端PM通过用户调研进行定量和定性分析

②根据收集的数据给出可用性问题的具体建议.

2> 总体思路.

①基于PAC-MAD模型 定性评估移动端PM的可用性.

②采用标准SUS进行定量分析，同时可以与其它认证方案作对比

③从与外部应用的集成、安全性，用户指导和交互三个方面给出了具体的建议.

④另外与其它相关研究的结论做了比对分析.

▲采用的方法

1> 4个在Android market和Apple Store上流行的PM：1Password, Dashlane, keeper, LastPass

2>评估框架

①定量分析采用标准SUS (System Usability Scale)评估，在参与者对一组10个声明确定同意程度后，产生0~100的可用性分数. (>70为可接受 70s~80s为更好 90s是例外基于测的课)

SUS允许与使用同样方法的研究进行对比

②使用PACMAD(People At the Center of Mobile Application Development)从7个角度评估移动应用：Effectiveness, Efficiency, Satisfaction, Learnability, Memorability, Errors, Cognitive Load. NASA Task Load Index(TLX) 自动的时间测量 SUS 其余采用开放或固定选项的问题.

3>研究结构

①Pre-study Questionnaire：获取身份数据和对PM的认识；提供PM的定义，指导用户安装并完成预定义的任务(基于Chicosson 06年的探索性实验)

(i) Initialization：安装并注册PM应用 (ii) Account migration：在PM中存储一个现有的帐户

(iii) Login：使用PM存储的帐户开启会话 (iv) 使用手机创建新帐户存储在PM中并登录网站

(v) 与主页UI交互：下载上述网站app并登录 (vi) 更改吃：使用PM的吃生成器生成，并重新登录

(vii) 安全设置：更改PM的安全设置

②在每个任务后进行After-task Questionnaire，对PACMAD进行评估.

③最后进行Post-study Questionnaire 用于量化SUS分数

4>每个PM选择了20名参与者(不得选用自己使用的PM)

4>局限性

①从MTurk取样未满足样本的多样性

②研究层面：未隐藏研究目的可能造成偏见；参与者更多是对PM感兴趣的；按照规定的流程执行操作可能与实际不符；研究基于用户的自我报告和远程测试可能有一定误差.

③基于流行度选择PM，且未规定具体的OS、PM的版本，可能产生一定的差异.

▲研究结果(平均花费48min完成任务 中位数为44min)

1>基本信息

①71.25%的参与者使用Android.

②PM的采用率：声称：83.75% (>7)已知PM，但17.5% (14)真正使用.但后续表明参与者对PM定义不一

2>PACMAD属性

①Effectiveness：所有任务的平均成功率超过90%，任务5(下载APP并登录)成功率较低，用户可能需要授权或者发现autofill不能在每个app上成功

②Efficiency：通过ANOVA测试发现各个PM在各个任务中的时间无明显差异 任务1、4和6时间较长

③Satisfaction(采用开放问题使用thematic analysis)

(i) Likes：Effectiveness 和 Simplicity

(ii) Dislikes：Lack of guidance, Lack of features, Mistrust (例不愿使用吃生成器), Performance

(iii) 继续使用的意愿：大约一半愿意

-积极：convenience, enhanced security

-消极：usability issues (40%), No perceived need (40%), Already use another PM (26%)

④Learnability (SUS评估：需专业人员支持才可使用，需学习许多知识才能使用).

ANOVA发现分数存在显著差异，采用Tukey HSD进一步发现除了1Password分数均可接受(超70分)

⑤Errors (通过询问用户失败的原因)

总体成功率较高，且错误的类型是有限的，主要3类：(i).与auto-fill功能匹配 最多 (ii).与password generation and update相关 (iii). 与performance相关

PM未在用户出错时提供足够的帮助，且与用户的mental model不匹配

⑥Cognitive Load.(只有任务1和6存在显著的差异)

采用5分制的Likert scale评估用户完成每个任务的工作量，1和7最低，6最高(且出错最多)

3>SUS分数(平均值 Dashlane (76.5±17.89) > keeper (71±16.98) > LastPass (69±19.66) > 1Password (52.6±21.83))

①采用单向ANOVA测试和Tukey HSD测试确定1Password的平均SUS分数显著低于其三者

②评估不同身份信息的影响(对每个PM进行多个线性回归测试)：无统计显著的信息

总体看，PM的SUS分数远远达不到优秀(excellent)，之间研究表明移动端吃认证方案无

4>进一步观察（探索设计层面的可用性）

①In-app browser：用户可能选择In-app浏览器登录，可能由于不兼容无法完成生成和重置

②Password Generation：在任务中，只有28%的参与者使用，不使用的原因包括：
(i) lack of awareness (ii) lack of interest (iii) distrust of PM.
▲发现使用吃生成器得到的吃更长。

③Preferred features.（给用户8个选项，选6个并根据喜爱程度给出排序）
autofill最高，吃生成和自动更新吃也较高，吃共享最低。

④Security settings.
分为谋：(i) 使应用可用性更高例：使用指纹替代主吃；禁用lock-on-exit和clear clipboard
(ii) 使应用更安全：固定时间锁定，隐藏吃的功能
(iii) 最小化设置的精力：不改动设置。
75%依赖于初始的设置，因此使PM的默认设置更安全是较好的选择。
最多的更改是使用指纹而非主吃解锁，部分原因是coldness

▲建议

1>User Guidance and Interaction

①告知用户PM的概念：如何工作以及为什么安全，主吃的概念；为什么强吃很重要以及如何选择
②告知用户如何使用基本的功能 ③解释安全设置中不同选项的含义
另外指导文件的格式、出现的位置也会影响
从用户交互角度：①Performance ②Additional features

2>Integration

用户和现有指南都认为应减少用户输入，而使用autofill
但现有的接口不够好，需PM和OS的共同努力使得接口在大多数平台可用，且应提供尽可能准确的指导（另外还有吃生成功能）

3>Security

高安全性可以提高用户对PM的信任，建议对主吃的吃策略要求提高，且根据用户需要使用2FA等技术；为用户提供吃强度反馈；
一些设置会减弱安全性：如记住主吃，和允许禁用锁定功能。

▲实现对可用性和安全性之间的关注不同，PM的配置选项也在权衡，应设计成使用户更容易理解如何达到想要的安全性
→一种方法：采用adaptive context-based security，根据传感器推断的安全修改设置
→默认设置应尽可能安全。