Usenix Sec'18 - Lyastani - Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse

概述：(S)用户通常将PM视为口令重用和脆弱性问题的解决方案，但PM在实际中对口令强度和重用情况的影响尚不可知；(T)首次进行PM对用户真实口令的影响的大规模调研。(A)通过476名在线调研参与者的口令创建和管理策略的定性分析，以及通过浏览器插件对170名用户真实行为的监测，分析了PM是否增强了口令强度并改善了口令重用问题。(R)发现(详见最后)如果只使用PM存储口令，reuse会更严重(chrome)；而如果使用生成口令的技术，即使手动输入，强度和reuse都会改善。

优点：①3.4章的论述很好，如何为自己的操作证明合理性，如何区分与前人研究的区别

②给出了本文方案的 limitation，实则论述自己方案的合理性

问题：① 对于与前人相悖的结论如 用户可识别口令强度应给出一定的对比

② 对170名用户的实验 极有可能 因为记录用户行为造成较大 bias，应进一步探索更好的方案。

③ 检测 entry method 应对说是否全面包含了所有 method，例如 typing speed 慢于用户速度。

▲Password Manager对用户真实实践口令的影响：口令强度和口令重用

▲引出研究目标：

增多的帐户数以及记忆能力有限导致了用户对口令的不安全实践(例如重用口令)，最好就利用这一点对用户发动攻击。→许多措施被提出，例如 PSM 和 PM → PM可以同时提供安全性和可用性(autofill, generate, storage) → PM是否真正地改善了不安全的口令实践尚不清楚：

① PM真正存储了由口令生成器等生成的强口令还是用户选择的弱口令？

② 使用了PM的用户是为每个网站服务生成唯一的口令还是仍然重用口令？

▲关键结果：

1> 采用的方法

▲ guessability metric 是评估口令强度的更实际的方法。(由N15T推荐)本文使用研的 zxcvbn

▲雇佣MTurk的用户进行数据收集，数据用于两个不同的阶段。

① a survey sampling → password survey

(i)调问参与者的隐私态度，对口令的态度，创建和管理的技能和策略。(Westin index.) ——了解口令创建及存储策略 / 辨识潜在的bias

(ii)由两个researchers进行coding.

从MTurk收集了476份有效的响应，且询问用户是否愿意参与接下来的studies(24人无兴趣)

② collection of password metrics → Chrome plugin-based data collection.

(i)为收集口令强度、重用，entry method和domain的信息，创建Chrome插件监控表单中口令字段的值，在用户登录后将信息发送到服务器。

(ii)要求参与者以Chrome为主要浏览器且随时使用移动设备浏览网页，且保证取样 unbiased.

最后共170人完成此阶段的任务。

▲ plugin收集的指标数据：

(i). Composition：口令的长度和每种字符的长度。

(ii) Strength：本文使用 zxcvbn 检测口令的强度：0(weakest) → 4(strongest)

包括 pattern matching (repeats, sequences, keyboard patterns,...)

common password dictionaries (leaked passwords, names, English dictionary words...)

mangling rules (leetify...)

(iii) Website category：在Alexa Web Information Service中网站所属类别

(iv) Entry method：输入口令的方式，包括 human, Chrome auto-fill, copy& paste, 3rd party password manager plugin 以及 external password manager program.

(v) In situ questionnaire：对于输入口令和网站的问卷调查，例如网站对应的隐私程度

(vi) Hashes：收集口令的哈希值以及每4个字符的哈希值。包括不同类型的重用：

▲Exactly reused passwords：与另一个口令相等　▲partially-and-exactly reused passwords：均有

▲Partially reused passwords：与另一个口令共享子串　→不足参与者出校口令

▲Detecting the entry method

(A) Typing detected? Y → (B) Typing speed human? Y → Human. N → External PM program. (too fast)

N → (C) Paste event? Y → Copy&Paste

N → (D) Chrome autofill? Y → Chrome built-in. PM. N → known installed PM plugin.

此处假设用户不会采用多种方式混合。

▲当plugin检测到有口令提交时，会询问用户3个问题：

① 对网站价值的估计　② 对输入口令强度的估计　③ 登录是否成功

▲保证用户隐私安全：

① 解释调研的动机和内容　② 以认证的方式发放plugin，不会使代码难理解

③ 只收集足以完成研究的最少的数据。

2> 研究PM的影响

① 通过使用 Mann-Whitney rank test 测试测试两个研究用户群体的区别，未发现明显区别

人群的身份特征与大学中定性调研的参与者的身份特征相近。

▲由于需安装 password logger，担心隐私意识差的用户会会使研究产生偏差

通过询问隐私态度、对口令的态度和是(或安全)否有过口令泄漏的经历，发现大多数(365)个参与者认为口令是一种有力的措施，且多的参与者经历过口令泄漏。

② General password statistics

在第2个研究中，170名参与者收集到1054个不同的口令，共1767个口令条目。

▲平均每个参与者在10.39个domain上输入口令(在问卷中参与者平均由口令保护的帐户数为29.95个，其中61%为估计数字，部分低估)。

▲只考虑不同的口令，平均每位参与者6.15个口令，以2.24种不同方法输入。总体来看，平均重用3.56%的口令重用的最小值和最大值为0%和100%，即有参与者完全重用或不重用口令。(Exact-and-partial 最常见，占比36.46%)

▲平均口令长度为9.61，包括2.12种字符类型，平均 zxcvbn score为2.20.

其中口令的强度与重用有显著关系▲

▲ Entry method中，Chrome auto-fill：53.71%，Manual Entry 33.39%，LastPass plugin：7.47%，3.11% Copy&paste，2.32% Unknown plugin，0.23%(4位) External manager

▲对于口令重用，除了 LastPass plugin 和 copy&paste，partial-and-exact为最常见的重用方式其中 Unknown plugin 54%，Human Manual 44%，Chrome autofill 45%，而不重用的比例中较高的为 LastPass Plugin 53%，Copy&Paste 78%，但 Copy&Paste 的口令强度最弱。

▲基于创建口令的策略分组

1>分组

①Password managers/generators (PWM) 26.47% 45/170

使用password manager program or an extra service.; 存储: 使用PM或另外独立的硬件

②Human-generated (Human) 71.18% 121/170

使用自己的方式创建口令(且尽量便其随机)或使用analog tools创建口令或用passphrases

存储: 易记忆; 使用analog或digital storage; 重用口令.

2>不同组的人群信息 (使用 Mann-Whitney test 检测两组人群的差异)

PWM组性别多为男性, 计算机背景和对口令的乐观态度高于Human组.

3>不同组口令强度和重用的比较

① PWM共输入522个口令, Human组共输入1245条口令, 均包含重用的口令.

② Human组比较倾向于使用更弱的口令, score为1的数(3%)为牛的4.9倍多2倍.

PWM组 2分最多(158), 但其余几组较为均衡.

PWM组中, 通过LastPass' plugin 输入的数目(n=93或17.82%)比Human (n=35或2.81%)多

且PWM组中通过LastPass输入的口令分数多数高于2(82个), 4分的最多为32个. 44.58%

③ 口令重用: 对于最频繁的 exactly-and-partially reused, PWM组(189 or 36.21%), Human(555 or

且PWM中不重用口令的几乎与exactly-and-partially reused的一样多.

▲Chrome auto-fill在几种重用分类都存在

在PWM组中, 使用LastPass输入的多于一半未重用 (n=49, 52.69%)

▲重用最多的输入方式为 manual entry 和 Chrome auto-fill: PWM组, 335(64.18%)重用, 其中778(82.99%)为以上两种方式; Human组, 979(78.63%)重用, 其中926(94.59%)为以上两种方式

▲最影响口令强度和重用概率的因素. (包括entry method以及与用户个人相关的因素)

→分析方法: multilevel analysis (属于regression analysis, 考虑数据的层次模型, 评估几种regression model, 比起单独考虑某个因素更适合.

1>可用因素的相关性分析.

▲As multi-level models are highly vulnerable to multi-collinearity, detecting and potentially removing strongly correlated variables is essential to prevent inaccurate model estimations which could lead to false positive results.

①检测到zxcvbn分数与password composition特别是口令长度有强相关关系 (长的口令更高)

②口令重用与口令中存在小写字母相关 (更深入的发现是因为数据集中存在PIN, 故有此特征)☆

2>构造模型

首先构造base model, 随后包含了① entry methods, self-reported value, and strength. ② the number of individually submitted passwords per participant, the creation and storage strategy of the user. ③ the interaction between creation strategy and detected entry method.

best trade-off of complexity, stability, and fitness

在每次迭代后计算model fit并使用log likelihood model fit comparison验证新模型是否比原模型拟合data

3>Zxcvbn score

ordinal model with all predictors and also the mentioned interaction可以最好地描述数据

▲口令创建策略和检测到的entry method中的Chrome autofill, copy&paste以及Lastpass为模型中显著的predictors. 但entry methods和creation strategy与口令强度的关系不显著

▲只有当使用口令管理工具时一同使用创建口令的技术才能显著提高口令强度.

▲自己报告的口令强度是测量口令强度的显著的predictor, 即用户可以正确认识自己创建的口令的强度

4>Password Reuse.

logistical model with all predictors but without interactions 可最好地描述数据.

①entry method: reuse的odds值, 相对于人类输入, 若用LastPass则低2.85倍, 若用copy&paste则低1.29倍, 而Chrome autofill则高了1.65倍.

②创建口令的方法.

若使用了工具创建口令, 不重用的几率高了.

▲未发现entry method与password reuse之间有显著关系

▲用户输入口令数与口令重用相关, 每多1个口令, 重用概率多6% ☆.

▲对网站占的评估价值和口令强度是重用的predictor. higher value则重用少, 用户认为较强的口令也重用少;

▲使用analog password storage. 重用口令的概率也减小.

▲①几乎所有参与者使用多于一种entry method, 且每一种entry method都会重用口令, 但Chrome autofill的重用率超过80%, 而由LastPass plugin则只有47%, copy&paste只有22%

②所有entry-method都可能产生弱口令. LastPass的zxcvbn的平均值最高

manually entered passwords及Chrome auto-filled passwords的口令强度处于平均水平, 但重用率高于平均值.

③使用password generator更可能得到强口令, 但Chrome的生成功能默认是disabled. 而Safari默认是开启的; 网站的value也可改善口令重用.

④用户通常可以知道自己口令的强度.

⑤当不使用口令生成功能时, Chrome autofill可能会加剧口令重用.

▲应设计更好的PM的集成功能, 并探索用户一开始不使用PM的原因.