IMPERIAL

Protecting User Secrets in Hostile Environments

LSR Poster by Sirvan Almasi | Supervised by Prof. William J. Knottenbelt

1. Introduction

This thesis is about the security of **user secrets**, such as passwords. **Firstly**, we investigate patterns in human-chosen secrets that make hash functions vulnerable to pre-image attacks. **Secondly**, we investigate broader threats to user secrets in their journey. We ask **"What are the key security risks associated with the journey of user secrets, and how can we mitigate them?"**

Chapter 3: Empirical analysis of data breaches and leaks.

Attack Type	Number of Attacks	%
Unsecured DB	19	32%
Other	11	18%
Unknown	10	17%
Man-in-the-Browser	8	13%
Phishing	5	8%
CSA	4	7%
Ransomware	3	5%

Table 1: Analysis of 60 reported data breaches in 2020

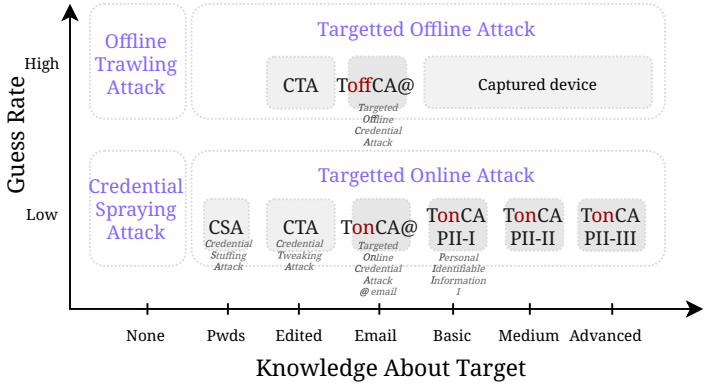
Chapter 4: Why are passwords weak?

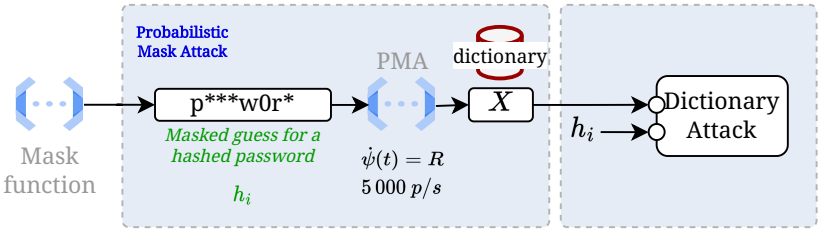
Chapter 5: How are passwords guessed?

Chapter 6: Protecting passwords and user secrets from MitB.

Chapter 7: Alternatives to hash functions and password-less schemes.

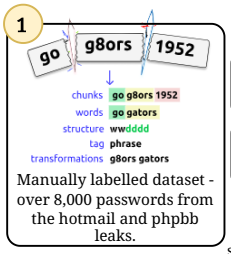
3. Password Guessing Algorithms





2. Password Composition and Complexity

Human-chosen secrets, being predictable, break the pre-image resistance of hash functions by making it computationally feasible to guess the input, thereby inverting the function.



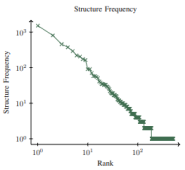
Manually labelled dataset - over 8,000 passwords from the hotmail and phpbb leaks.

First ever large scale labelled dataset.

Transformations:

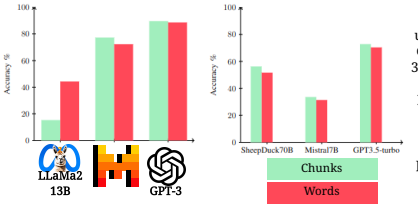
castilleo → leo castillo

9801*well* → 1984 Orwell



Passwords distributions are said to exhibit a power law like distribution. We find here that the password structures and extracted words (from unique passwords) also exhibit a power like distribution--Zipf's law.

2 Using LLMs to expedite the labelling process



Left Fig. The models are fine-tuned using 825 passwords from the phpbb dataset, and are tested on a separate 300 passwords from the same dataset.

Right Fig. The models are fine-tuned using 1 725 manually labelled passwords from the phpbb and hotmail dataset. Test data is 3062 passwords from the hotmail dataset.

3 Password Complexity

$$H = \log_2 R^L$$

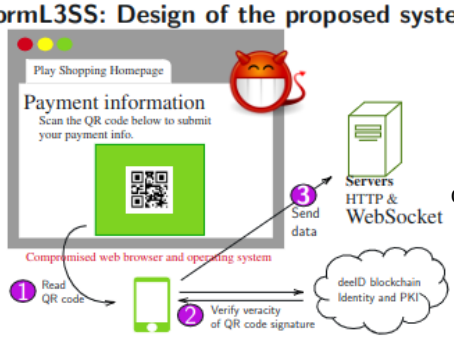
Structure	*log* perm'	Frequency Count	%	Hashcat using Nvidia RTX 4090 MD5	PBKDF2-sha256	scrypt	bcrypt-sha512	zxcvbn [25] Mean Score	Guesses
w		1520	21.0	~0	0.1 s	1 min	4 min	1.3	5.36
n		811	11.2	~0	0.1 s	1 min	4 min	1.0	4.72
nd		58	0.8	~0	0.6 s	12 min	42 min	1.4	5.68
wd		91	1.3	~0	0.6 s	12 min	42 min	1.5	5.97
nl		33	0.5	~0	1.5 s	30 min	2 h	1.2	5.09
wdd		294	4.1	~0	5.6 s	2 h	7 h	1.8	6.69
ndd		221	3.1	~0	5.6 s	2 h	7 h	1.6	6.33
nddd		71	1.0	~0	56.4 s	19 h	3 day	1.7	6.37
wddd		91	1.3	~0	56.4 s	19 h	3 day	1.9	6.84
wsdd		29	0.4	~0	3 min	3 day	9 day	2.0	7.28
ddddw		35	0.5	~0	9 min	8 day	29 day	2.4	7.95
wdddd		174	2.4	~0	9 min	8 day	29 day	2.3	7.59
ndddd		205	2.8	~0	9 min	8 day	29 day	2.0	6.90
nn		376	5.2	1.5 s	8 h	1 yr	4 yr	2.0	6.91
ww		457	6.3	1.5 s	8 h	1 yr	4 yr	2.0	7.04
nw		30	0.4	1.5 s	8 h	1 yr	4 yr	2.2	7.51
ndddddd		51	0.7	3.0 s	16 h	2 yr	8 yr	2.5	8.05
wdddddd		33	0.5	3.0 s	16 h	2 yr	8 yr	2.8	8.84
wwdd		58	0.8	3 min	33 day	111 yr	402 yr	2.9	8.97
nndd		38	0.5	3 min	33 day	111 yr	402 yr	3.2	9.62
wwddd		33	0.5	4 h	9 yr	1.11×10^4 yr	4.02×10^4 yr	3.3	9.81
nwn		41	0.6	9 day	447 yr	5.56×10^5 yr	2.01×10^6 yr	3.0	8.88
www		165	2.3	9 day	447 yr	5.56×10^5 yr	2.01×10^6 yr	3.0	9.26
nww		35	0.5	9 day	447 yr	5.56×10^5 yr	2.01×10^6 yr	2.7	8.18
nnn		27	0.4	9 day	447 yr	5.56×10^5 yr	2.01×10^6 yr	3.3	9.95
wwww		55	0.8	171 day	8.70×10^3 yr	1.08×10^7 yr	3.91×10^7 yr	3.5	11.24

Table 2: The table presents the top 25 labelled password structures from the **hotmail** dataset, accounting for 5032 (69.5%) of the labelled dataset. **Hashcat** benchmarks indicate the performance of consumer hardware in a hybrid attack. The dictionary size for w (words) and n (numbers) is set at 0.5 million records . **zxcvbn** serves as a password strength meter, with scores ranging from 0 to 4, and guess estimates are log based.

4. Browser and Server Exploits

Protecting User Secrets from MitB Malware

FormL3SS: Design of the proposed system

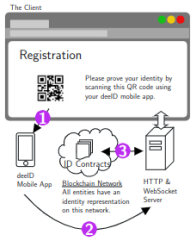


Form grabbers and Man-in-the-Browser (MitB) malware attacks are an effective method of stealing sensitive information. One method of circumventing compromised browsers and operating systems is by using another device to submit the data (an out-of-band system).

Identity and Password-less Authentication

deeID: Design of the Proposed System

Web Authentication API (**WebAuthn**) uses public-key cryptography to authenticate users. **Passkeys**, a platform implementation (Apple WWDC22) of WebAuthn. Automatic cloud backup of passkeys raises privacy issues.



Research Overview

Research Overview:

- S. Almasi and W.J. Knottenbelt. **Protecting Users from Compromised Browsers and Form Grabbers.** NDSS Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb 2020), San Diego, California, USA, Feb 2020.
- S. Almasi and W.J. Knottenbelt. **PasswordNinja: Password Composition and Complexity Insights from Manually Labelled Datasets.** *2024* ***Drafted.***
- S. Almasi and W.J. Knottenbelt. **SoK: Password Guessing Algorithms.** 2024 ***In Progress.***

Working Papers:

- S. Almasi and W.J. Knottenbelt. **Human Identity and the Quest to Replace Password Continued.** 2019
- S. Almasi and W.J. Knottenbelt. **Security First API Modelling Techniques.** 2021
- S. Almasi and W.J. Knottenbelt. **Research Opportunities and Lessons Identified from Recent Web Security Attacks.** 2021