

INSIGHTS INTO THE

I.4 BILLION CLEAR TEXT CREDENTIALS TROVE



COMMUNITY FEEDBACK
Data Validation Response Summary

CONTENTS

Background 2

Executive Summary 3

**Community-Validated
Results 4**

Password Details 11

**Company & Organization
Domains 14**

**Steps To Protect Your
Digital Identity 16**

About ResTech Solutions 17



1. BACKGROUND

On December 8, 2017, ResTech Solutions reported on the discovery of a file with a 41GB database of 1.4 billion clear text credentials, which included usernames and clear text password pairs.

After discovering the trove, we started to analyze the data it contained - not only to validate the information, but also identify key insights that could help individuals and enterprises be smarter about how they approach securing their credentials and, by extension, their digital identity.

Importantly, the trove is not a single database obtained from a one-off breach but a combination of previous breaches with now new decrypted passwords exposed. What's dangerous about this exposure is the ease of access to a sheer volume of clear text credentials in an aggregated, searchable, interactive database. The data was organized in an alphabetic directory tree broken into 1,981 pieces to allow fast searches. That means that even unsophisticated, novice hackers or unauthorized parties could exploit a massive number of usernames and passwords combinations easier than ever before.

To illustrate further, if this trove contains your email username and password, a hacker could log in and review any personal information sent via email to friends, families, doctors, or financial advisors. Using your email account, or accessing online services can lead to serious exposures including financial loss.

We also gleaned information about passwords, providing yet another example of why password standards of yesterday are no longer adequate today.

Our intent with this information is not to scare anyone, but to educate people about identity risk. Consumers know they should be more cautious about changing passwords and using unique login credentials. But it's human nature to think, ***"What's the worst that can happen?"*** or even ***"It won't happen to me."*** This community report demonstrates that a large portion of the passwords in the trove are authentic, and the subsequent community responses we received, are a prime example of why we need to make digital identity security a priority.



2. EXECUTIVE SUMMARY

After a large amount of credential validation requests, we engaged the community to better understand the accuracy of the data. The result was a validation - in numbers, statistics and facts - what everyone believed to be true about poor password security practices. Below are some of the key findings that we will explore in this report:



79.3%

of the passwords are
'true' or authentic

65%

of people are reusing
the same password
amongst different
services

43%

exposed company and
organization domains
were from Educational
Institutes

27.3%

of the passwords were used within
the **LAST SIX MONTHS**,
even though most were created
2 TO 5 YEARS AGO

39.3%

are using "very weak" passwords.
Over **9M** passwords
use "**123456**"

3. COMMUNITY-VALIDATED RESULTS

On **December 12, 2017**, we offered a free service to send obfuscated, exposed passwords to anyone in the community who emailed us a verification request. We started reporting back results to these requests, only to find the demand was so high it required an automated process.

We created a portal where people could enter their email to determine whether their personal credentials were affected. Individuals would then receive confirmation whether or not their email was included in the trove, along with a truncated version of the password if applicable. In return, we requested help to verify the data from the community by answering four questions via email:



1. **Is (are) the password(s) true?
Yes or No.**
2. **When did you create it (them)?**
3. **When did you last use it (them)?**
4. **Which sites did you use it (them) in?**

As of February 15, 2018 **40,836** people have sent requests to check their credentials, with more are coming in every day. Over **1,600** of these people have sent back answers to our questions.

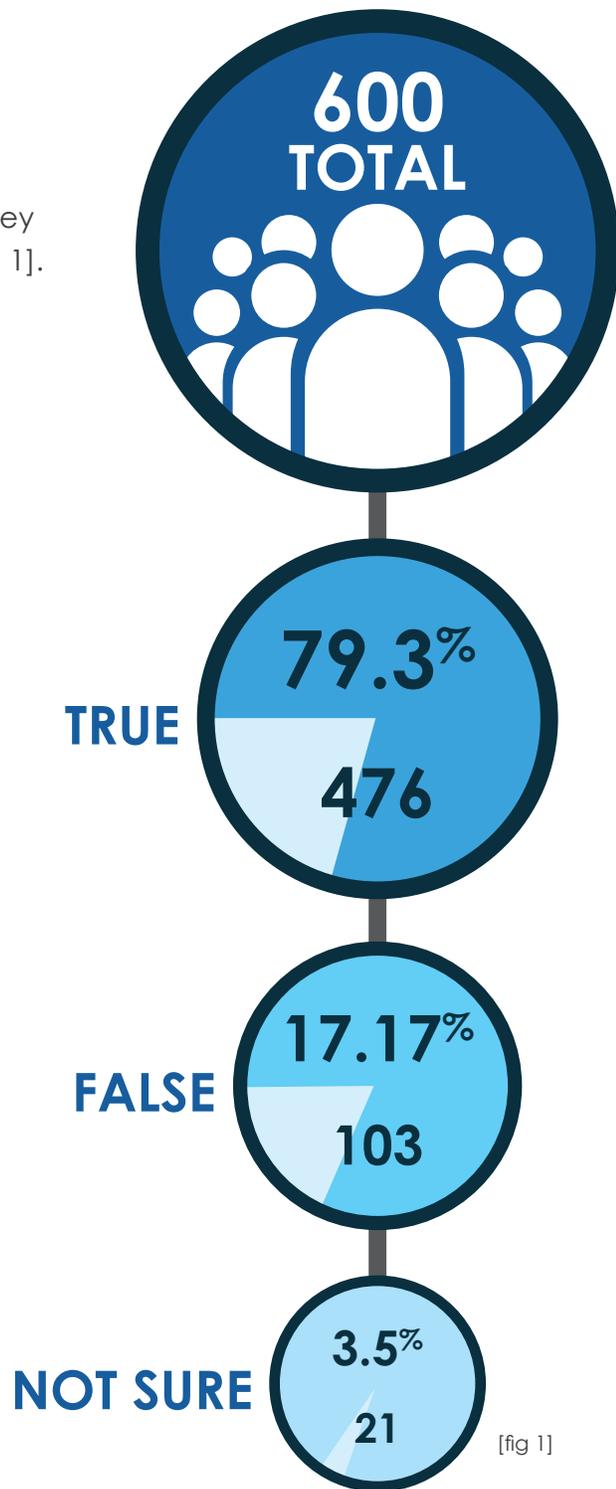
We manually reviewed the first 600 responses to one or more of the questions above. We analyzed their responses to gain insights on real password habits, data freshness and validation of our finding. Additional data from the respondents include how many passwords for each email were exposed.

The statistics and information in this section capture the responses provided by respondents requesting credential verification against the trove findings. We read through each response and tracked each answer to gain insights on user behavior with credentials.

3.1 PASSWORD ACCURACY

Are the passwords authentic?

Most of them are! Almost **80%** of respondents verified that one or more of their exposed passwords were true. Other respondents said they either did not remember or were "not sure." [fig 1].



3.2 WHICH SITES DID YOU LAST USE THEM IN?

Sites where passwords were used in the last 6 months

Some of the respondents gave specific details about sites where they use their passwords. We've compiled the answers into categories and counted the results. Over 25% can't remember all the sites where they used the password.

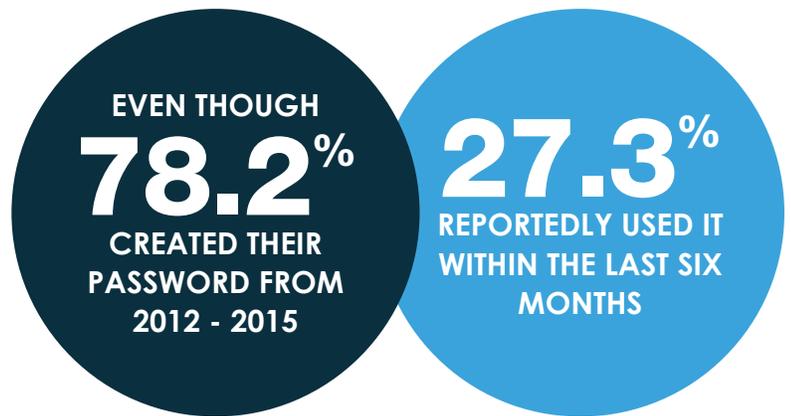
PERCENTAGE	CATEGORIES
26.22%	"Many" "A lot of sites" "Too many to remember"
13.48%	Social Media + Dating Sites (e.g. Facebook, LinkedIn, Twitter, Instagram, Skype, etc.).
13.11%	"Unimportant or throwaway accounts" (e.g. news, blogs, sports, tech, etc.).
10.82%	Entertainment, Gaming and Services (e.g. AT&T, Netflix, Hulu, last.fm, Steam, etc.).
7.12%	eCommerce (e.g. Amazon, Ticketmaster, eBay, etc.).
6.37%	Email
5.62%	Banking, Finance, Taxes, Legal, Payment (Paypal, Docusign, Ameriprise, etc.).
5.24%	Work & Collaboration Tools (e.g. Evernote, Dropbox, Adobe, etc.).
3.75%	Forums (e.g. StackOverflow, Reddit, gaming).
3.37%	Infrastructure (e.g. servers, webhost, gitHub, etc.).
2.62%	Travel: booking and rewards (e.g. AirBnB, Travelocity, Hotels.com, etc.).
1.50%	"Won't say for security reasons."
0.75%	Healthcare, Medical and Pharmacy



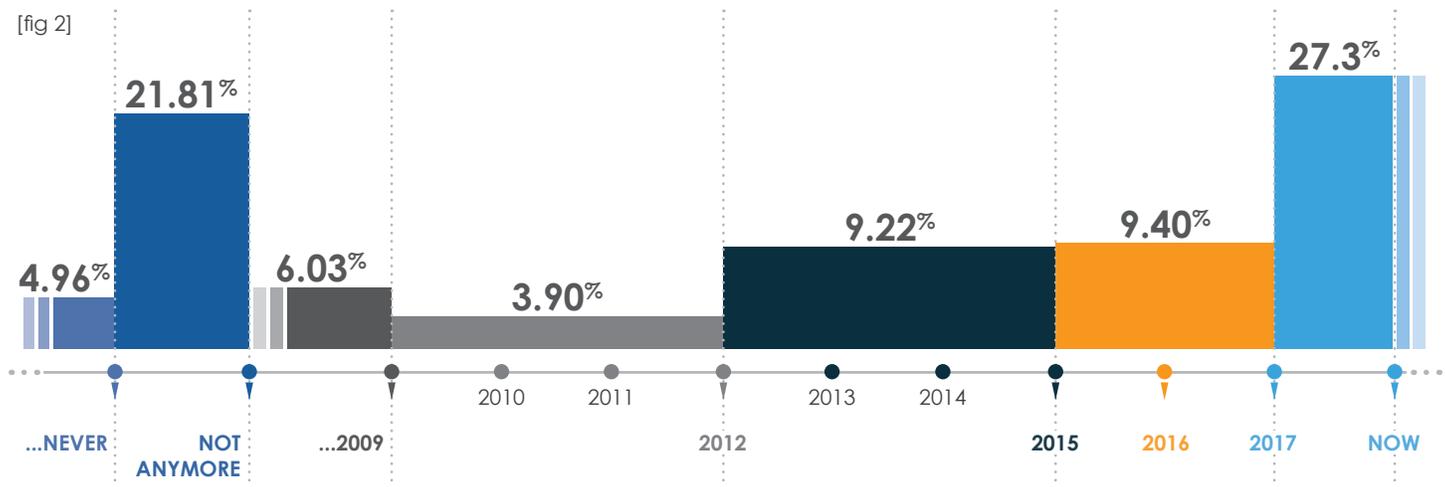
3.3 PASSWORD FRESHNESS

When did you last use the exposed password(s)?

This data indicates password freshness, with **154 (27.3%)** of the respondents stating that they used the password within six months of the day they sent the verification request, and **277 (36.7%)** people are still using or have used these passwords within the last year.



53 (21.81%) gave a general “not using anymore” answer, and **89 (17.38%)** said they didn't remember when they used the password last [see fig 2].

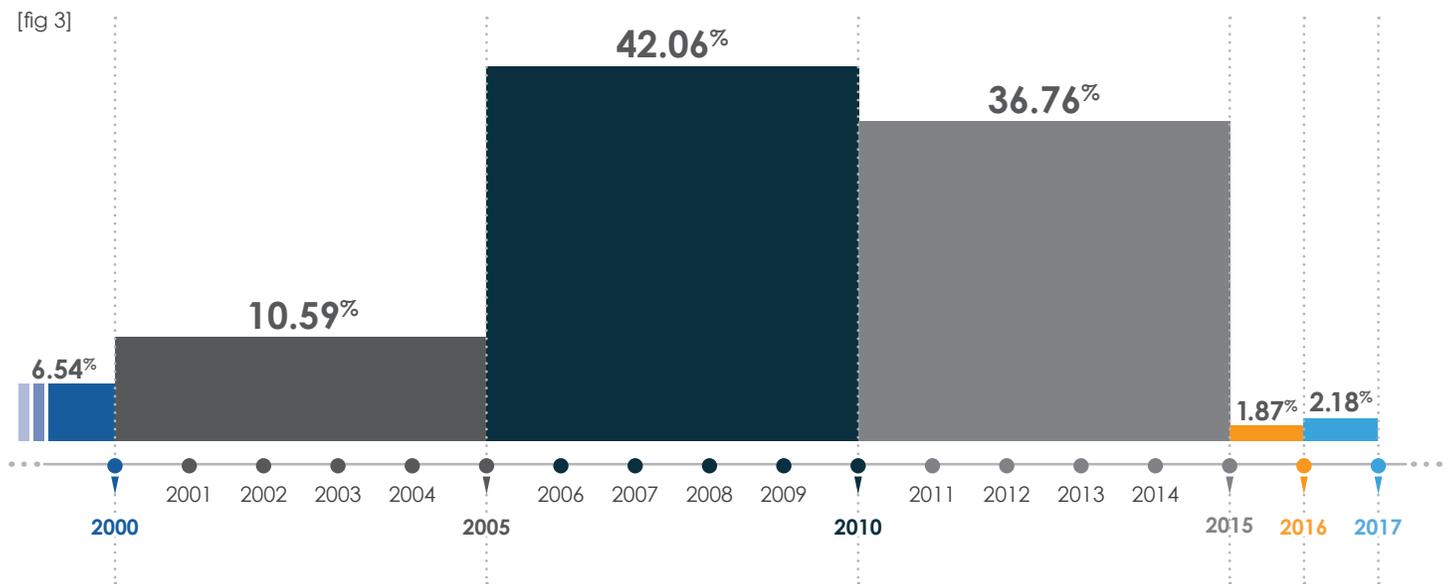


Total: **564** People who don't remember: **98 (17.38%)**

3.4 WHEN DID YOU CREATE AND LAST USE THE PASSWORD(S)?

Date ranges when passwords were created

Almost half of the respondents didn't remember when they created their passwords. From the 321 people that did remember, only **(4.05%)** were created within the last **two years**, **(78.82%)** of them were created from **2005 to 2015** [see fig 3].



Total: **562** People who don't remember: **241 (42.88%)**

3.5 PASSWORD REUSE

We've known that we should be more cautious about changing passwords and using unique login credentials, but this is a prime example of why it's so important. This indicates a trend in password reuse. Out of the **600** respondents:



FEEDBACK ON PASSWORD REUSE

"I USED THAT PASSWORD FOR AMAZON, ATlassian, BITBUCKET, DROPBOX, EVERNOTE, STACKOVERFLOW, AND FACEBOOK."

– Community Response

"NOT PREPARED TO STATE WHICH SITES FOR SECURITY REASONS. MOST ARE UNIMPORTANT BUT ONE IS IMPORTANT SO I WILL BE CHANGING IT."

– Community Response

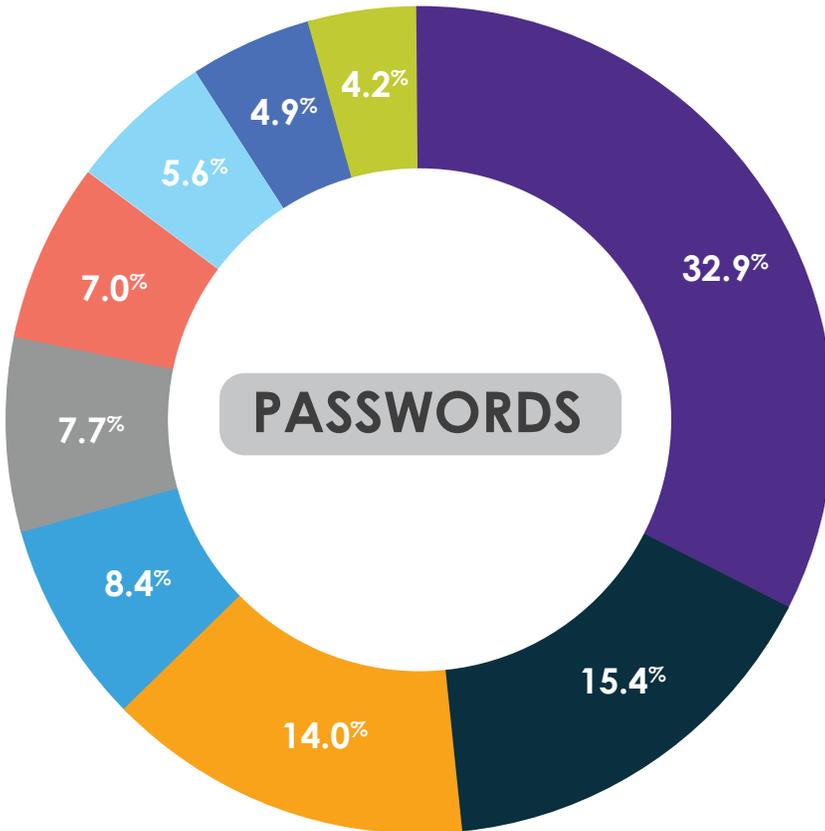
"USED THIS PASSWORD UNTIL THE DAY I RECEIVED VERIFICATION FOR SITES REQUIRING PASSWORDS FOR CREATING ACCOUNTS...BUT NEVER ACCESSING ANY PAYMENT BUT FOR AMAZON. AS SOON AS I RECEIVED YOUR EMAIL, I CHANGED THE PASSWORD."

– Community Response

3.6 PASSWORD FRESHNESS

Using and Mixing where Passwords were Reused within the last 6 Months of verification

Some of the respondents gave specific details in where they reuse their passwords.



- Multiple Site 32.9%
- Social & Other Sites 15.4%
- Unimportant Sites/ Throwaway Accounts 14.0%
- eCommerce with other Sites 8.4%
- Too Many, Don't Remember 7.7%
- ALL – Email, Banking, Social... 7.0%
- Work Tools & Applications 5.6%
- Email & other sites 4.9%
- Banking & Finance 4.2%

FEEDBACK ON PASSWORD REUSE

2018: "PRETTY MUCH EVERY SITE I NEED TO USE A PASSWORD."

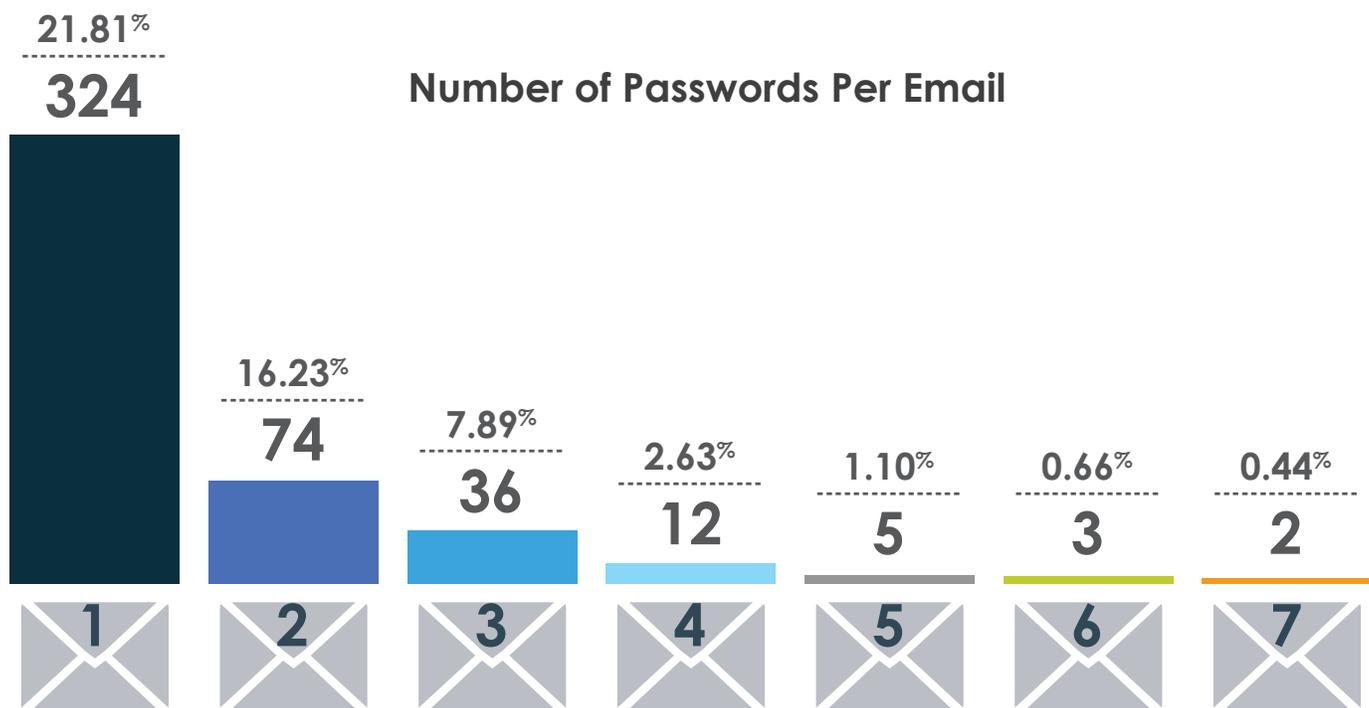
– Community Response

"I USE IT A LOT. I USUALLY HAVE TO LOG INTO MY ACCOUNTS FROM DIFFERENT COMPUTERS. DON'T KNOW EXACTLY WHERE."

– Community Response

4. PASSWORD DETAILS

The following section provides details on the number of passwords per email. For the respondents (sample set of **600**), the majority of respondents had just one password exposed per email, **324 (71.95%)**. The number drops significantly with two passwords being the second highest, **74 (16.23%)**.



INSIGHTS ON RESPONSES FOR NUMBER OF PASSWORDS

9 = HIGHEST NUMBER OF PASSWORDS EXPOSED: ALL WERE VERIFIED AS TRUE.

– Community Response

“I STILL USE THAT PASSWORD FOR NONCRITICAL ACCOUNTS, BUT I WASN’T AWARE THE PASSWORD WAS AVAILABLE IN CLEAR TEXT. RETHINKING WHERE I USE IT NOW.”

– Community Response

4.1 THE TOP 40 PASSWORD PATTERN

We ran a count to detect the most commonly used password patterns. Here are the **top 40**:

RANK	COUNT	PASSWORD	RANK	COUNT	PASSWORD
1	9218720	123456	21	370652	666666
2	3103503	123456789	22	354784	123
3	1651385	qwerty	23	347187	monkey
4	1313464	password	24	343864	dragon
5	1273179	111111	25	311371	1qaz2wsx
6	1126222	12345678	26	300279	123qwe
7	1085144	abc123	27	299984	121212
8	969909	1234567	28	298938	myspace1
9	952446	password1	29	291132	a123456
10	879924	1234567890	30	276473	qwe123
11	866640	123123	31	270488	1q2w3e4r
12	834468	12345	32	268121	zxcvbnm
13	621078	homelesspa	33	263605	zxcvbnm
14	564344	iloveyou	34	255079	7777777
15	527158	1q2w3e4r5t	35	250732	123abc
16	470562	qwertyuiop	36	241721	qwerty123
17	468554	1234	37	241495	qwerty1
18	417878	123456a	38	227701	987654321
19	398114	123321	39	226785	222222
20	371627	654321	40	220363	112233

PASSWORD PATTERNS & STRENGTH

“THIS IS AN AUGMENTED PASSWORD BASED ON AN ORIGINAL ONE FOR SITES WITH STRICTER REQUIREMENTS.”

– Community Response

4.2 PASSWORD STRENGTH

We ran a subset of passwords through some algorithms to check for strength of the credentials. The following checks were performed.

Entropy: Measures diversity of the characters and level of difficulty to bruteforce crack the passwords.

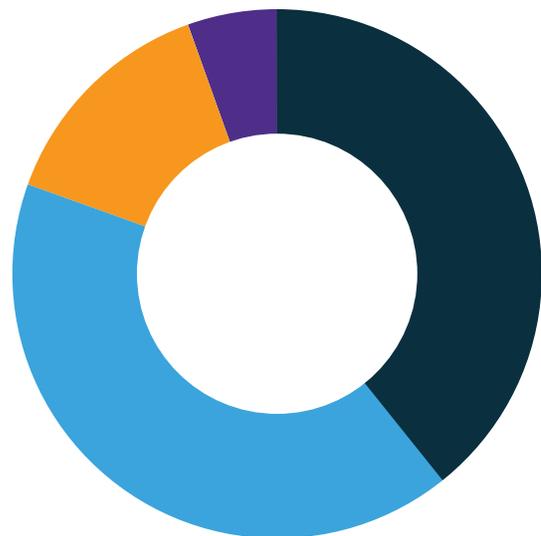
Format: Factors usage of digits, uppercase letters, and symbols within passwords.

Username & Dictionary Similarity: Checks for the username or dictionary words being part or similar to the password.

We took a random sample size of **292,661** from the trove with the above algorithms.

Here are the results for password strength:

PASSWORD STRENGTH	PERCENTAGE
● Very Weak	39.30%
● Weak	41.33%
● Medium	14.10
● Strong	5.27%



PASSWORD USAGE

“I USE IT FOR LOW SENSITIVITY ACCOUNTS AND ANNOYINGLY PAYPAL SINCE THEIR PASSWORD POLICY DISALLOWS “;” WHICH PREVENTS ME FROM USING MY HIGH SENSITIVITY PASSWORD.”

– Community Response

“AFTER THE CREDENTIALS DUMP, I STARTED USING A PASSWORD MANAGER. I RECOMMEND ANYONE WHO HAS USED DUPLICATE PASSWORDS TO DO THE SAME BECAUSE YOU NEVER KNOW WHEN YOUR ACCOUNT INFORMATION COULD BE BREACHED.”

– Community Response

5 COMPANY & ORGANIZATION DOMAINS

We went through the data to get insights into company and organization domains with the highest exposures and grouped them into industry categories. In order to get accurate and clean results we removed webmail and ISP provided domain exposures, including Facebook.com which was a free service provided by the social network from 2012 - 2014 for anyone with a Facebook profile.

We also removed nonexistent and discontinued addresses that are now redirected or are no longer in service, as well as fake domains. For example, some domains are used solely for phishing attacks and other scams (e.g, dr.com, writeme.com, etc.)

It is important to note that the data does not imply that companies within these categories were themselves breached. It merely indicates that people have used the company or organization domain emails for their credentials to access other websites and services. The actual exposure of the emails can be from a variety of breached sites, including LinkedIn, Exploit.in, Adobe, Dropbox, etc.

5.1 TOP 75 INDUSTRY DOMAIN EMAILS EXPOSED

Industries are broken up into 10 categories for the top 75 company or organization domain-based emails.

INDUSTRY	# EXPOSED DOMAINS	# EXPOSED RECORDS	PERCENT
Education	33	791,243	43%
Technology	10	263,696	14%
Consumer Goods, Services, Automotive, Real Estate, Travel	10	216,547	12%
Banking & Finance	9	206,468	11%
Military & Government	4	161,959	9%
Consulting & Services	3	80,957	4%
Insurance	2	51,266	3%
Pharmaceutical & Healthcare	2	44,132	2%
Social Network	1	17,992	1%
Other: (e.g. News)	1	17,484	1%
	75	1,851,744	100%

5.2 TOP 10 COUNTRIES OF EXPOSED COMPANY AND ORGANIZATION DOMAINS

Countries for the top 75 company or organization domain-based emails.



HQ LOCATION	# EXPOSED DOMAINS	# EXPOSED RECORDS/ COUNTRY
United States	60	1,452,480
Russia	2	76,907
Germany	2	73,904
United Kingdom	3	59,980
Ireland	1	39,600
Netherlands	2	39,929
Sweden	1	25,204
India	1	23,864
Switzerland	1	22,422
Korea	1	19,462
Italy	1	17,992
	75	1,851,744

COMPANY DOMAIN USAGE

“MY AMERIPRISE EMAIL ADDRESS IS TYPICALLY USED FOR A LOGIN ID ON VENDOR WEBSITES, FINANCIAL PLANNING VENDOR SITES, ETC. I MAY HAVE USED THIS COMBINATION WHEN SIGNING UP FOR WEBSITES LIKE GOOGLE, FACEBOOK, LINKEDIN, MAPQUEST AND ANY WEBSITES USED FOR WORK.”

– Community Response

6. STEPS TO PROTECT YOUR DIGITAL IDENTITY

This trove presents significant data to support the need for consumers to take preventative action to protect their usernames and passwords - the keys that unlock the door to your digital identity.

This is particularly important as risks evolve.

Therefore, consumers can benefit from the following key takeaways:

- **PASSWORD STRENGTH:** Using passwords that are too easy to crack, increases an individual's risk greatly, which is why the **National Institute of Standards Technology** (NIST) has outlined password guidelines. Consumers should become aware of the recommendation to create stronger passwords that are harder for hackers to crack.
- **DON'T USE A PASSWORD FOR MORE THAN ONE SITE.** Also, password Managers can be used to secure and protect credentials and consumers should be using these applications to prevent identity theft and account takeover.
- **SECURE YOUR INFORMATION WITH 2-FACTOR AUTHENTICATION.** 2-factor authentication provides an extra layer of security that requires not only a password and username but also something that only that user has on them.
- **MONITOR YOUR DIGITAL IDENTITY, NOT JUST YOUR CREDIT SCORE.** In the event information is compromised, consumers should use a service that monitors the deep and dark web for leaked credentials to ensure they can act quickly to prevent identity theft.
- **TAKE ACTION WHEN YOU LEARN ABOUT A BREACH.** If you hear about a social media platform you use or company you buy from being breached, you should change your password in that service, as well as all services that you think share the same or similar password.

**14 OUT OF THE 600 COMMUNITY RESPONSES SAY
THEY HAVE SWITCHED TO A PASSWORD MANAGER.**

6. ABOUT RESTECH SOLUTIONS

Hello, we are ResTech Solutions, an IT Services company that protects your information and helps you get things done more efficiently. Think about us as your better and more affordable in-house IT department.

On the surface, we help businesses and individuals keep their information safe, and systems and processes running smoothly and efficiently. To do so, we monitor, diagnose, and address the root causes of difficulties and nagging problems with technology to help our clients seize opportunities and serve their people with peace-of-mind. We are not a temporary fix; we are a holistic and permanent solution.

At our core, we exist to create harmony between humans and machines to help people do what they do better. We envision a world where technology enables us to interact genuinely with one another and make more meaningful human connections. An environment where people thrive because the technology just works, a world that is simpler, more enjoyable, and more human.

Whether you need general IT services, cybersecurity, or IT consulting, we are here to help.

Please give us a call at (713) 936-6855, we'd love to talk to you.

**1.4 BILLION
CLEAR TEXT
PASSWORDS**



RESTECH
S O L U T I O N S

ResTech Solutions Headquarters

8715 Meadowcroft Dr. #102

Houston, TX 77063

ILLUMINATING THE DARK WEBSM

© 2020 ResTech Solutions, LLC. All right reserved. ResTech Solutions and the ResTech Solutions logo are registered trademarks of ResTech Solutions, LLC. Other names may be trademarks of their respective owners.

www.restech.solutions

