The Attacker's Dictionary

Auditing Criminal Credential Attacks

Tod Beardsley, Senior Security Research Manager

Roy Hodgman, Data Scientist

Jon Hart, Senior Security Researcher

Harley Geiger, Director of Public Policy

Rapid7, Inc. | March 1, 2016

RAPID7

The Attacker’s Dictionary

Auditing Criminal Credential Attacks

Contents

01

EXECUTIVE SUMMARY

This paper is the product of nearly a year's worth of opportunistic credential scanning data collected from Heisenberg, Rapid7's public-facing network of low-interaction honeypots. Instead of focusing on the passwords that end users typically pick, with this data we can see what opportunistic scanners are using in order to test—and likely compromise— Internet connected point of sale (POS) systems, kiosks, and scamware-compromised desktop PCs which offer the Remote Desktop Protocol (RDP) service for remote management.[1]

Given Rapid7's unique visibility into the credentials that attackers are choosing, we can measure a variety of statistics that are of interest to security practitioners and data scientists. For security practitioners, we report on the frequency and source of opportunistic attacks; the top attempted usernames, passwords, and username:password combinations, and the overlap between these chosen credentials and published password dumps collected from breach data.

For the number crunchers out there, we also measure the complexity of chosen passwords. Our findings show that the majority of passwords attempted are very simple, implying a widespread use of defaults and passwords chosen out of convenience instead of security necessity. We also look at the lifetime of a given password, and assess how often the same passwords are tried over time.

Finally, while this paper offers some general advice for securing RDP endpoints in particular, the information security research community still struggles with the chilling effects on credential-based research due to the language of the Computer Fraud and Abuse Act (CFAA), prohibitions against "trafficking in stolen authentication features," and similar laws as they are written today. Therefore, the authors of this paper hope the data and analyses presented in this paper are useful, but also hope that legislators and policymakers can better understand the current issues in law that are effectively preventing valuable security research.

[1] Yes, malicious actors attempting to compromise the usurped assets of other malicious actors. There may be no better indication of the volatility in the underground/criminal economy.

02
BAD PASSWORDS ARE BAD

At the end of 2015, password management company SplashData released its annual "Worst Passwords" report, once again reminding the world that "123456" and "password" are the worst possible choices to secure user accounts[2]. Millions of passwords were leaked over the course of 2015, and SplashData's most recent report corroborates the statement that "regular humans are terrible at picking passwords." The carbon-based computers we store in our heads are certainly suboptimal password generation devices, since people tend to pick easy-to-remember, easy-to-type passwords.

While reports on the inherent weaknesses of human-chosen passwords are valuable in communicating security shortcomings to the general population, it's difficult to characterize people's regular work, banking, e-mail, and other critical passwords based off of lists of password leaks. After all, the majority of the sources of these passwords are almost necessarily from sites with weak security standards. Most important accounts have passwords that are subject to complexity requirements in the form of minimum lengths, mixes of character sets, and the like, so it would be difficult to register a new account with "starwars" as a password, no matter how much one might love the Disney franchise.

In this report, we take a different approach to surveying the passwords. We focused on the passwords that attackers attempt to use during opportunistic, criminal attacks on high-value, Internet-exposed systems. To collect this data, Rapid7 launched "Heisenberg," a network of low-interaction honeypots.

[2] Slain, M. (2016, January 19). Announcing our Worst Passwords of 2015 [Web log post]. Retrieved from https://www.teamsid.com/worst-passwords-2015/

03
HEISENBERG HONEYPOTS

Heisenberg is a companion project to Rapid7's Project Sonar[3], but the data is gathered in the opposite direction. While Project Sonar actively scans the Internet and helps us identify what services and systems are available based on pre-authentication responses, Heisenberg is designed to attract the attention of scanners and elicit login attempts to the services we purport to provide.

Heisenberg honeypots are custom-engineered, low-interaction honeypots that are distributed geographically across several regions. They are "low-interaction" in the sense that they merely emulate the authentication handshakes of several protocols without attempting to emulate (or actually offer) the full capabilities of the protocol or the underlying operating systems. So, while this honeypot network is an ideal collection system for opportunistic attacker-controlled credentials, it does not offer further insight into the motives of attackers in the event that they guess a "correct" password.

Given that these services are fictional, there is no legitimate reason to attempt to log in to any of the Heisenberg nodes over these emulated protocol stubs. Therefore, the intelligence gathered is necessarily scanning traffic from unknown parties, and can usually be characterized as malicious.[4]

This paper concentrates on one of the more interesting protocols which the Heisenberg honeypots support, the Remote Desktop Protocol (RDP).

> Heisenberg honeypots are custom-engineered, low-interaction honeypots that are distributed geographically across several regions.

Remote Desktop Protocol (RDP)

RDP enables remote desktop-based control of home, office, POS, and kiosk systems, and is often enabled intentionally and legitimately by those systems' owners, since it is sometimes considered as a secure alternative to a Virtual Private Network (VPN) connection. RDP does provide native encryption by default, and is optionally configurable with the Transport

[3] Project Sonar. (n.d.). Retrieved from https://sonar.labs.rapid7.com

[4] Of course, not all of the opportunistic login attempts to the Heisenberg honeypots are "malicious" in the ethical sense, as we will explore later in this paper.

Layer Security (TLS) encryption standard. While we at Rapid7 strongly recommend VPN-secured network access to internal resources (primarily to reduce attack surface), some IT organizations and Virtual Private Server (VPS) providers treat RDP as a viable alternative to traditional VPN access.

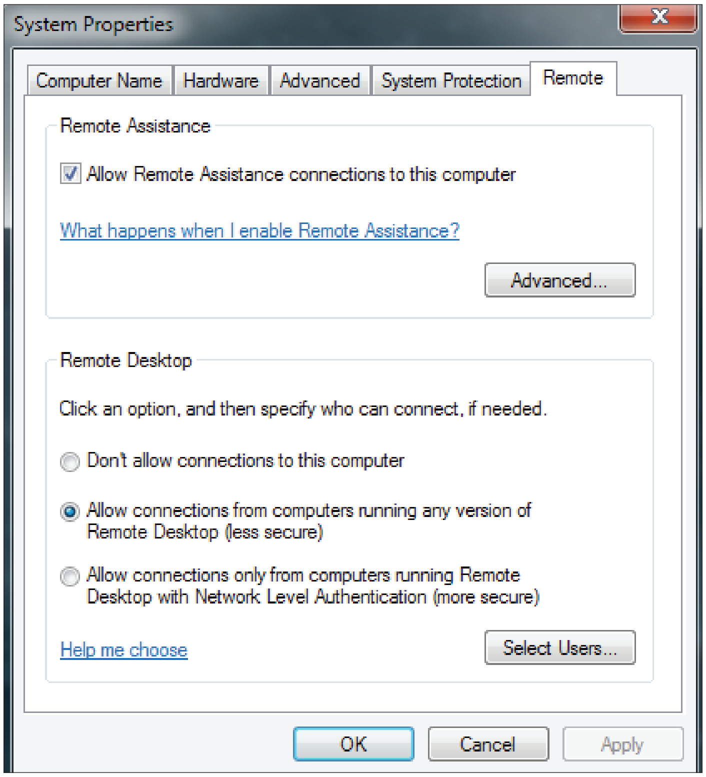
RDP is implemented on the server side as Remote Desktop Services (RDS) and ships with all Microsoft Windows operating systems since Windows XP. According to Sonar scans conducted in February 2016, there were 10,822,679 IP addresses listening for 3389/TCP, the default port for RDP. Attackers have taken notice of these millions of potential targets.

RDP is also a popular management interface for some Windows-based Point-Of-Sale (POS) systems. This was discussed extensively in FireEye's mid-2014 report on BrutPOS[5], and US-CERT's alert regarding BackOff[6], both of which target RDP-enabled POS systems and kiosks. Since 2014, successful attacks targeting POSes remain in the headlines as a favored tactic for criminals. Because of its attractiveness to criminals, we feel that this is an ideal protocol to build out an attacker's dictionary of commonly used, rarely changed passwords.

In addition, it is our belief that some fraction of these RDP endpoints are exposed as a result of the recent spike in "Windows tech support" scams[7], where users are tricked into giving control of their desktops to scam perpetrators. On Windows 7, for example, the configuration for RDP is located on a screen for "Remote Assistance," as seen in Figure 1.

As anti-virus and anti-malware consumer products started to identify and flag the third party screen-sharing and remote control software used in these scams, we believe these actors are now adopting Microsoft's built-in tools in order to avoid antivirus (AV) detection.

Figure 1: Remote Assistance configuration panel



Remote Assistance

☑ Allow Remote Assistance connections to this computer

What happens when I enable Remote Assistance?

Remote Desktop

Click an option, and then specify who can connect, if needed.

○ Don't allow connections to this computer

● Allow connections from computers running any version of Remote Desktop (less secure)

○ Allow connections only from computers running Remote Desktop with Network Level Authentication (more secure)

Help me choose

[5] Villeneuve, N., Kyle, W., & Homan, J. (2014, July 9). BrutPOS: RDP Bruteforcing Botnet Targeting POS Systems. Retrieved from https://www.fireeye.com/blog/threat-research/2014/07/brutpos-rdp-bruteforcing-botnet-targeting-pos-systems.html

[6] US-CERT. (2014, July 31). Alert (TA14-212A). Retrieved from https://www.us-cert.gov/ncas/alerts/TA14-212A

[7] VStrohm, M. (2015, September 4). Windows tech support scams are back [Web log post]. Retrieved from http://www.bankrate.com/financing/identity-protection/windows-tech-support-scams-are-back/

```
var td_2F = new td_0g(“4818a8896da3438186bcc1b-
d7a84a986000A030C555B0C0C5406050752065C07510E0C560B-
01121356094C40114A02191B4C5C404F4C5C5B570A0A1D-
575C551E5E46065100065406510200570 35DD0C53520F-
530D545900010554070A56035B525D54”);td_0O = td_2F.
td_f(24,1);td_2P = [‘REF:63333’, ‘VNC:5900’, ‘VNC:5901’, ‘VNC:5902’,
‘VNC:5903’, ‘RDP:3389’];td_0l = td_2F.td_f(50,32);td_1i = td_2F.
td_f(16,8);td_1g = td_2F.td_f(0,16);td_2j = td_2F.td_f(25,25);function
td_2B() {var s=false;if (td_1k.empty()) {return s;}return s;}
```

The belief that RDP is playing a role in scams is bolstered by the recent observation that at least one online banking site, TDBank.com, is actively scanning its customers via Websocket connection attempts to ports associated with RDP (and VNC). Evidence of RDP would help companies like TD Bank prevent connections from potentially risky endpoints. Proof of this intent to portscan customers of TDBank can be seen at https://tmx.tdbank.com/fp/check.js?org_id=i8n5h0pw&session_id=a, as reported by Randy Westergren[8]. The offending javascript snippet is shown above. Notice the tuple, ‘RDP:3389’, indicating the service and port to be scanned.

[8] Westergren, R. [RandyWestergren] (2016, January 18). Anyone else find it odd that @TDBank_US is port scanning customer machines for VNC, RDP, and a UPS? [Tweet]. Retrieved from https://twitter.com/RandyWestergren/status/689117344220721153

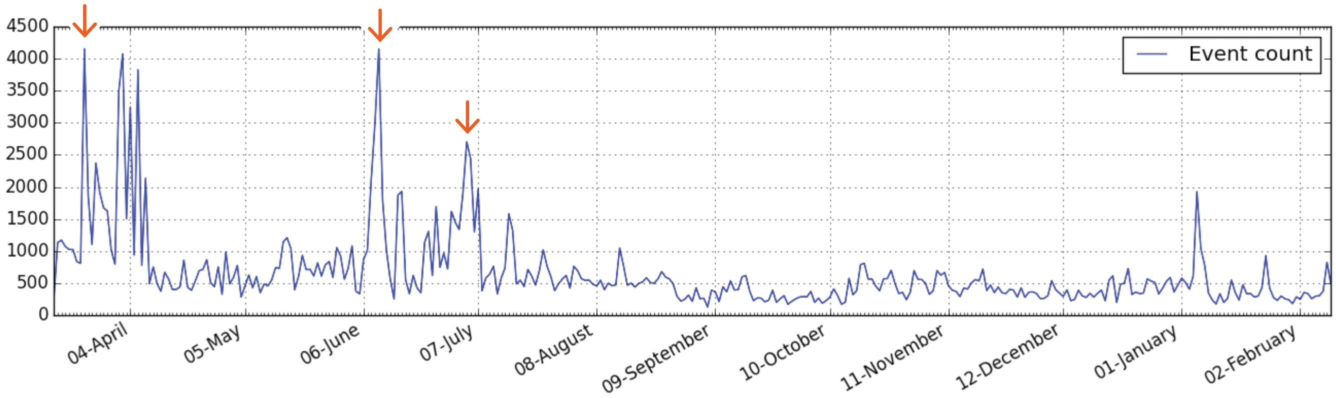
04
EVENT FREQUENCY AND TRAFFIC ANALYSIS

Some days are busier than others for the Heisenberg honeypots. While we cannot say what causes the particular spikes in credential scanning traffic, we can plot when there is increased attention to RDP across all sources. For example, it appears that the spring and summer of 2015 saw much more scanning activity across the board than the following autumn and winter (Figure 2).

However, if we break out our event statistics by source country, we can see that scanning sourced from China is responsible for the vast majority of the RDP activity, with distinct spikes in April, June, and July (Figure3).

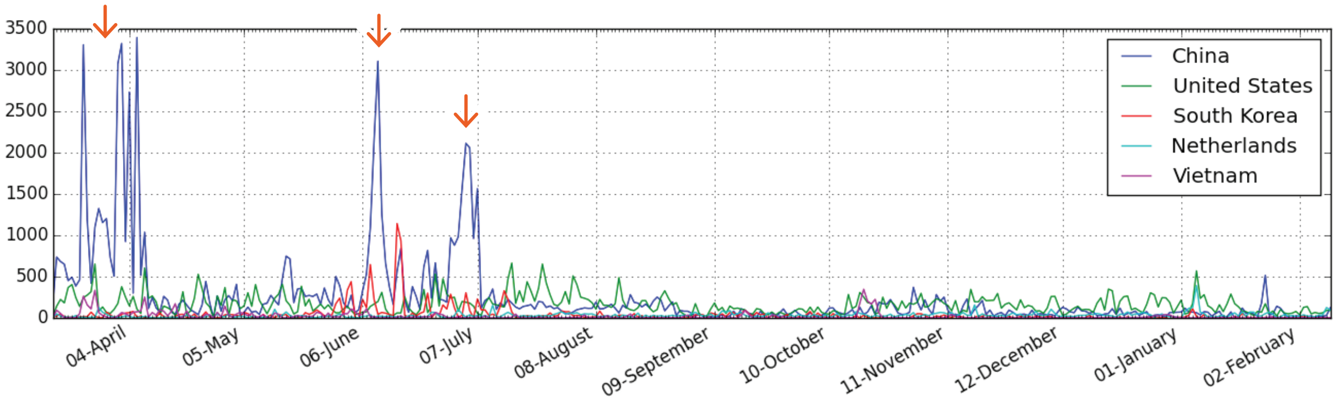
Of course, while some fraction of these "sources" are in fact proxies and artifacts of the limitations inherent in GeoIP databases, the observation that endpoints in networks associated with China are responsible for so much of the traffic is not particularly surprising. China also happens to be the most populous country on Earth, and the most recent data indicates that the Chinese account for nearly 20% of humanity as well as 20% of all Internet users[9]. However, the dominance of endpoints in China-based networks in our scan

Figure 2: Events Per Day, 2015-2016



[9] Internetlivestats.com. (n.d.). Internet Users by Country (2014). Retrieved February 19, 2016, from http://www.internetlivestats.com/internet-users-by-country/

Figure 3: Events Per Day by Country, 2015-2016

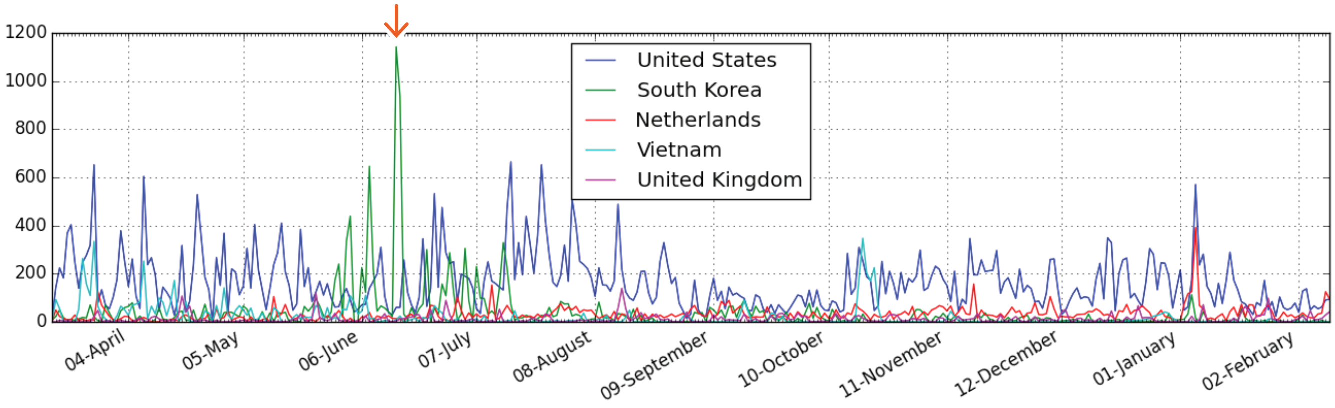


data does make it difficult to see more granular activity reports by other regions, so the plot below (Figure 4) shows the top five countries' activity after endpoints in China-based networks are removed.

Here, we can see that the endpoints in U.S.-based networks were the source of much more consistent scanning through the spring and summer of 2015, with a spike from endpoints in networks associated with the Republic of Korea (South Korea) in mid-June.

Traffic analysis like this can help investigators get a sense of overall opportunistic compromise activity of RDP-enabled systems, so they can look for clues accordingly.

Figure 4: Events Per Day by Country, Excluding Endpoints in China-based Networks



05
COLLECTED CREDENTIALS

We have collected credentials intended for our RDP listening services over the past 334 days, from 2015-03-12 to 2016-02-09. Over this period, we have recorded 221,203 attempts to login, sourced from 119 countries (Figures 5 and 6).

While we have primarily focused on the interesting features of the passwords, our Heisenberg network has also collected usernames. Because of this, we have access to the complete credentials, which can be critical in identifying the intended target.

Figure 5: Credentials collected worldwide

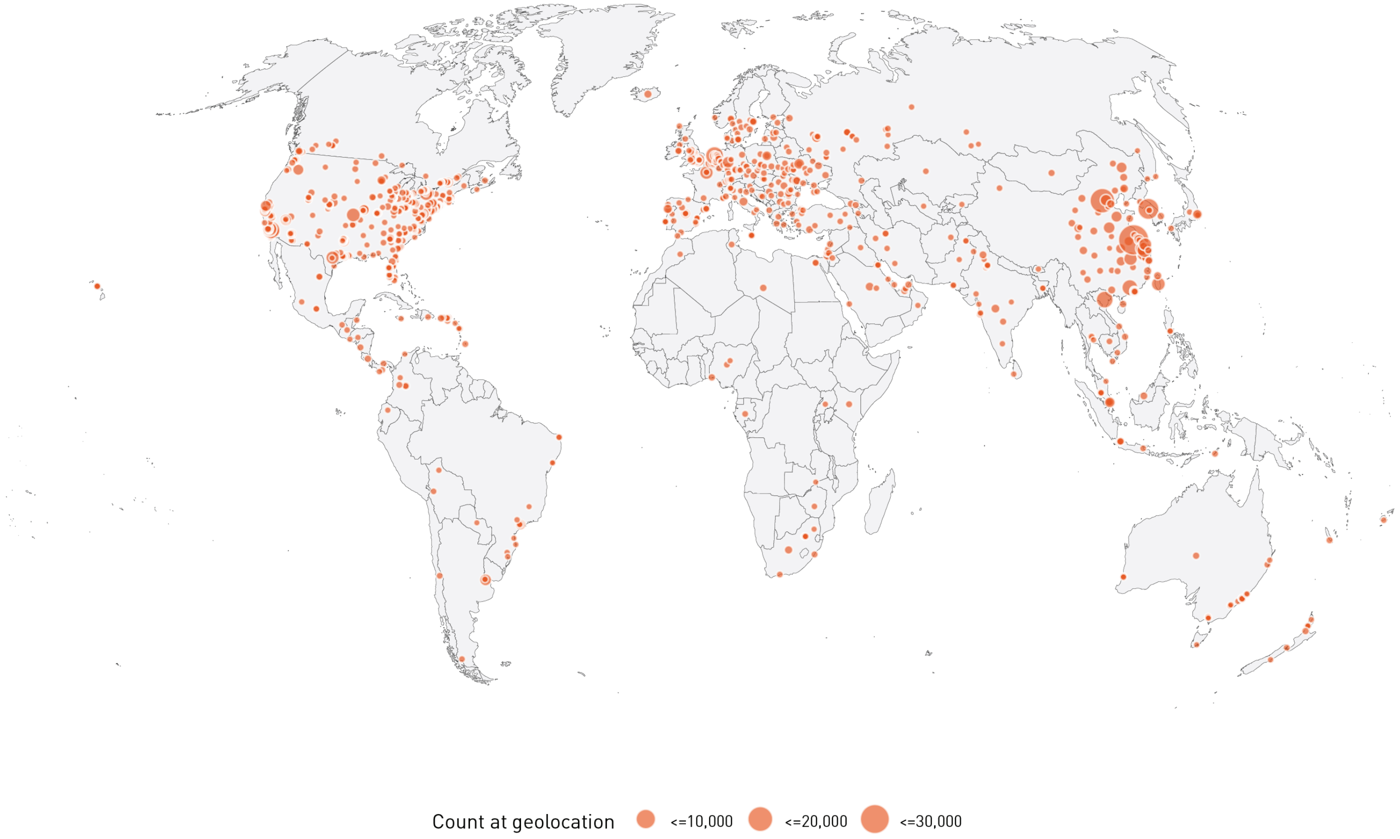
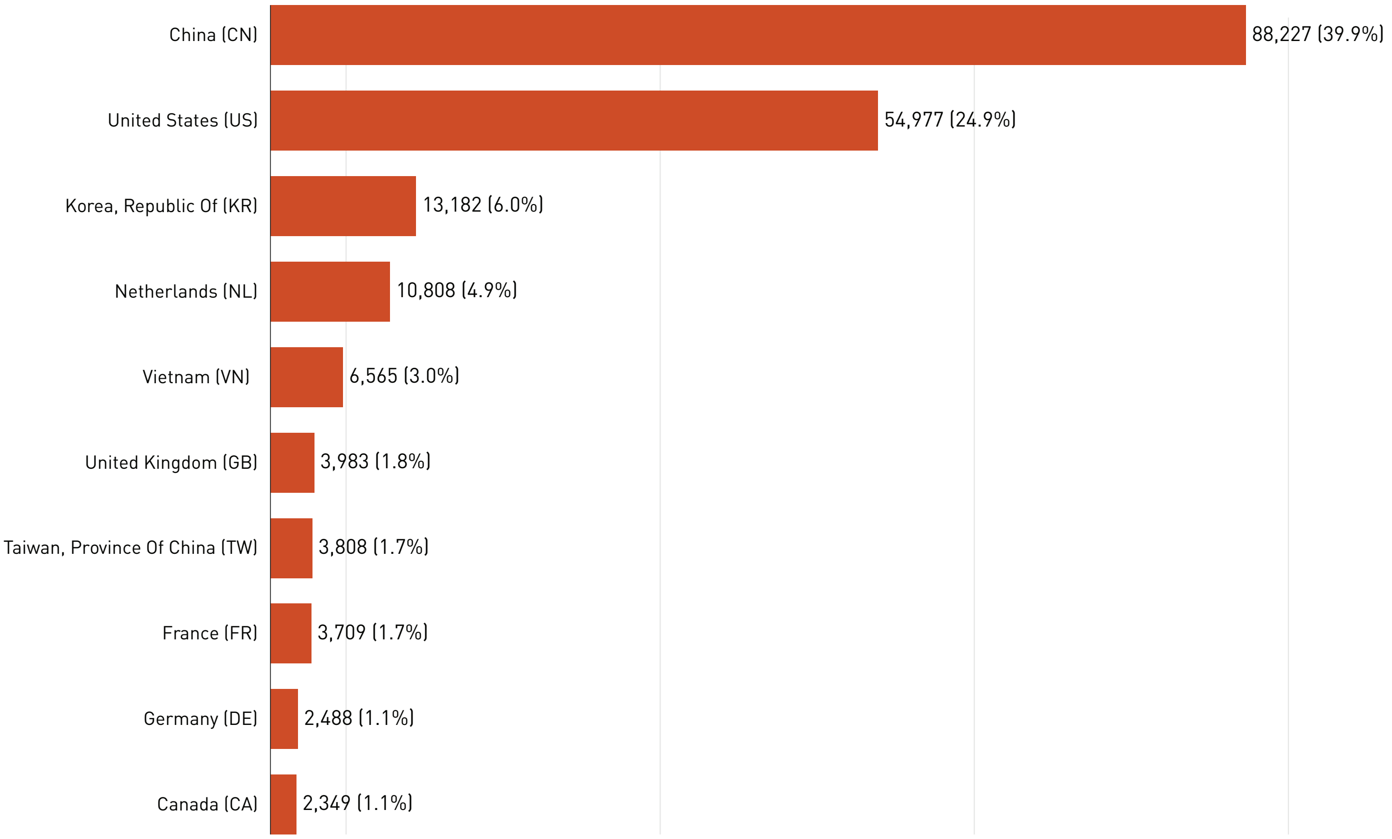


Figure 6: Top 10 Country Network Origins



06
INTERESTING CREDENTIALS

Attackers do not merely pick random strings as passwords (or usernames). Such brute force attacks are process intensive, time consuming, and tend to have very poor performance from the attacker's point of view. Instead, attackers in our data set were clearly conducting dictionary attacks; i.e. they were using chosen usernames and passwords that have an assumed high likelihood of success when applied to a target system.

Despite exhaustive efforts from our team of highly experienced security researchers, we encountered significant trouble in determining the source material for these dictionary attacks. Searching the usual sources of default password lists did not yield satisfying results, particularly for the more unique passwords.

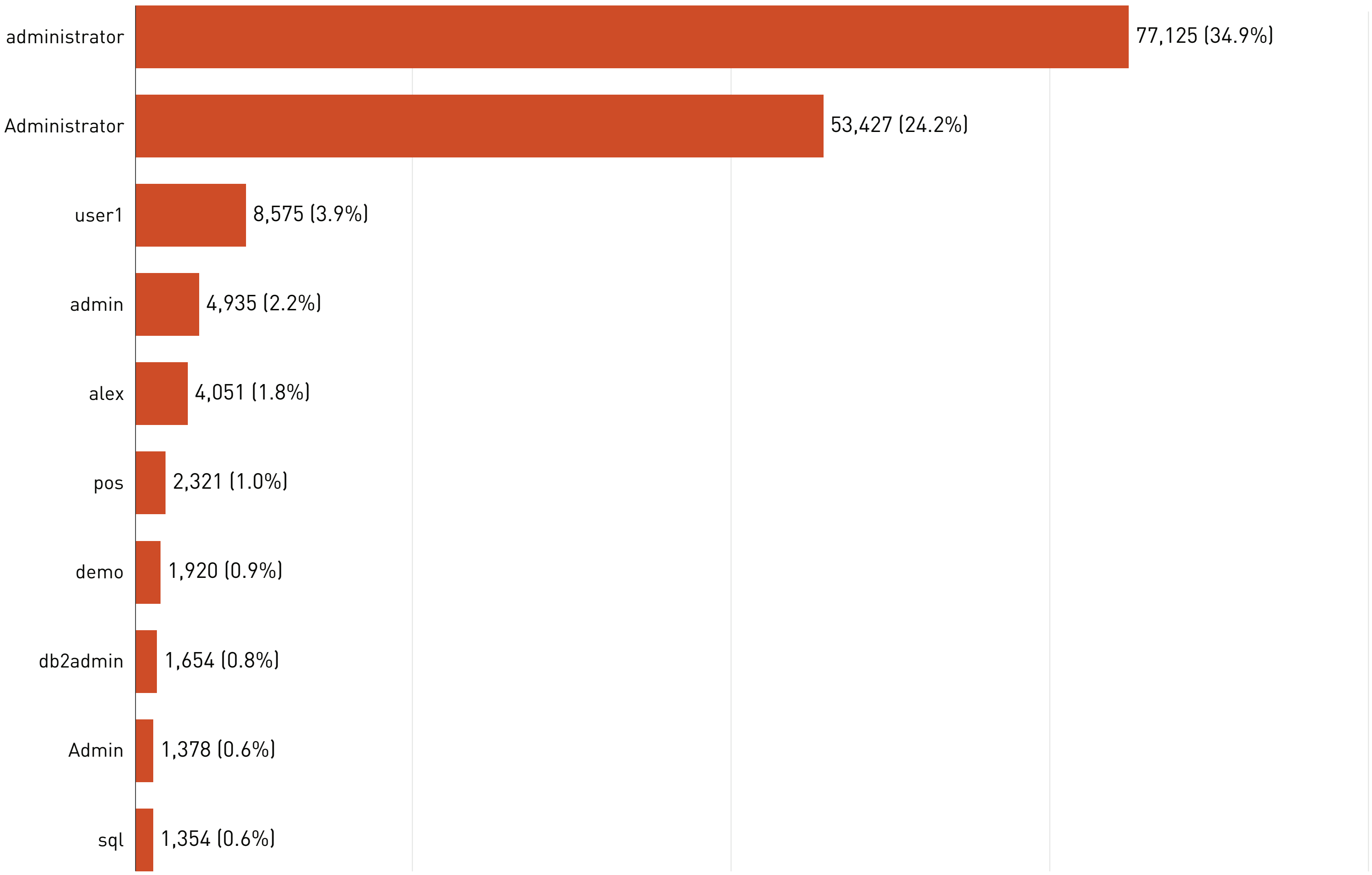
For example, a Google search for "St@rt123," the third most common password attempted, turned up only as a comment on a Krebs on Security blog post suggesting it as a merely hypothetical weak password[10] in connection to a Russian e-payment system breach. It's possible that "St@ rt123," is, in fact, a known-to-some common default. Given the appearance of this password on a blog that is concerned primarily with reporting high-profile data breaches, its possible it's use was merely an error in translation[11].

Despite these unknowns, some inferences can be drawn from the tables that follow, especially when it comes to the relationships of passwords to usernames.

[10] "Russ." (2010, December 29). Re: Russian e-Payment Giant ChronoPay Hacked [Web log comment]. Retrieved from http://krebsonsecurity.com/2010/12/russian-e-payment-giant-chronopay-hacked/

[11] The majority of attackers appear to originate from regions of the world where English is not the dominant language.

Figure 7: Top 10 Usernames



Top 10 Usernames

By far, the most common username is "administrator," followed by "Administrator" (Figure 7). Note that because RDP is strongly associated with Windows, usernames are not case-sensitive. While apparently omnipresent, it's good to know that these attackers are not omniscient, too.

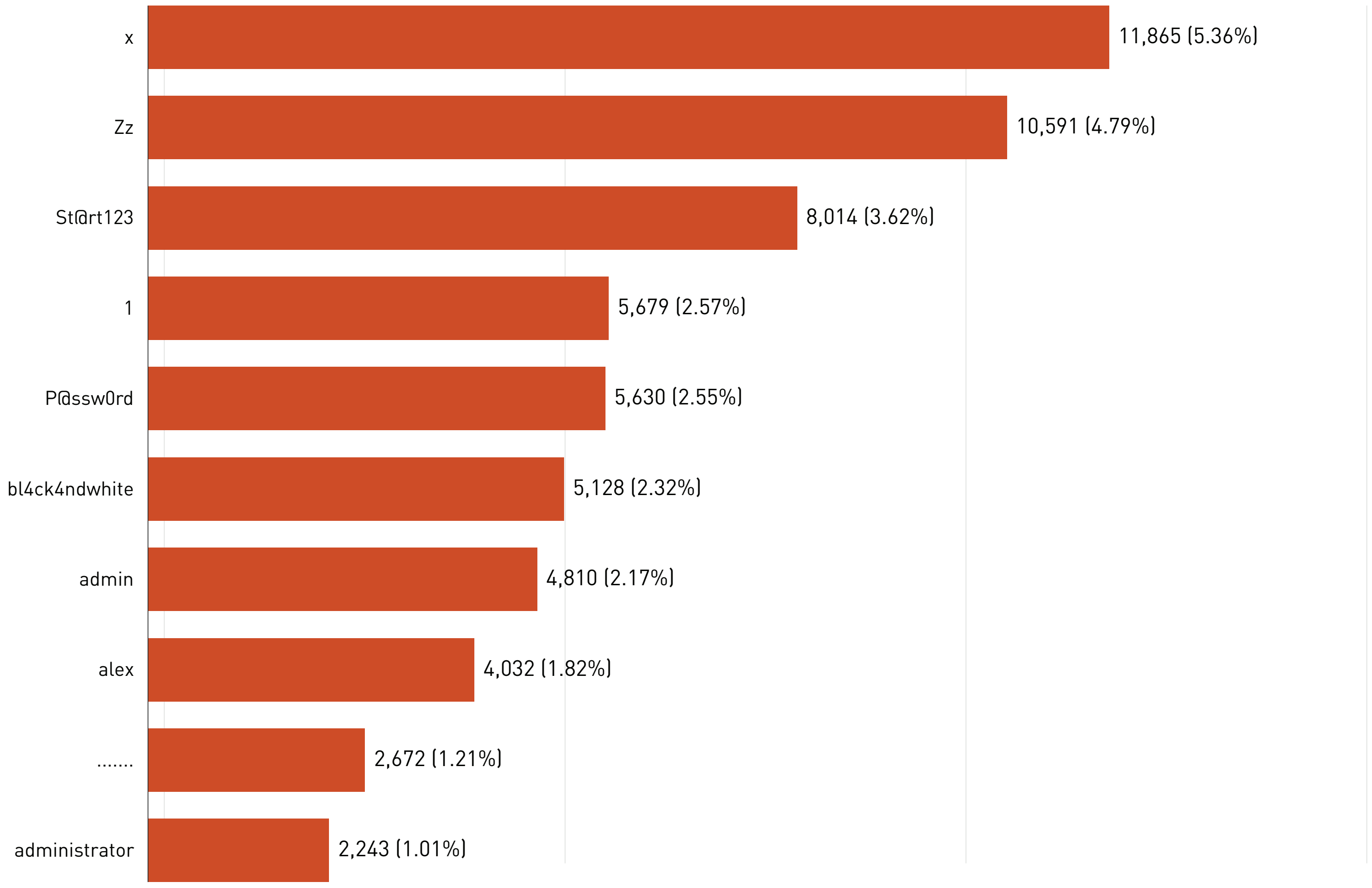
The sixth most common username, "pos," has a strong implication for point of sale systems, while the usernames "db2admin" and "sql" imply a hunt for Internet-facing database instances.

Top Ten Passwords

The most surprising aspect of the top ten passwords is that only the third, fifth, and sixth most common passwords of Sta@rt123, P@ssw0rd, and bl4ck4ndwhite (respectively) are even mildly complex, made up of simple letter substitutions which many, many dictionaries are likely to contain (Figure 8).

Truly, the surprising detail to be uncovered here is just how weak these passwords are. One or two characters, easily guessed strings, and a strange appearance of a series of dots. Since these passwords were deliberately chosen by the various scanners which ran up against Heisenberg, it implies that the default and common passwords to several POS and kiosk systems are chosen out of convenience, rather than security.

Figure 8: Top 10 Passwords



Top Passwords for the Top Ten Usernames

Taking a look at the top passwords per username can give a much clearer picture of what the scanners are after. We can see, for example, the lowercase "administrator" account is associated with two very strong passwords, as well as the very weak "x" and "Zz" passwords (Figure 9).

We can also see that, for some usernames, default passwords seem to be the most common target rather than merely weak passwords. For example, the "db2admin:db2admin" credential is the default credential for many versions of IBM's DB2 database, and it is one of only two passwords that are attempted with that user account.

Figure 9: Top Passwords for the Top Ten Usernames

Username	Password	Count	Percent
administrator	x	9623	4.35
	Zz	6724	3.04
	P@ssw0rd	3371	1.52
	&Tf56tUR@28	2024	0.91
	SPbgZBc.gjU	1888	0.85
	bl4ck4ndwhite	4877	2.20
	Zz	3851	1.74
	1	2327	1.05
	qaz	2184	0.99
	admin	2121	0.96
user1	St@rt123	8014	3.62
	St@rt1234	229	0.10
	safetypay123#	141	0.06
	.	82	0.04
	;DuTYmwUxM-	61	0.03
admin	admin	1007	0.46
	admin@123	324	0.15
	XZxzxz123	275	0.12
	Password	273	0.12
	qiaoge54.64	262	0.12
alex	alex	4030	1.82
	P@wss0.A1rd	8	0.00
	ABC.123456	7	0.00
	AAaa3344	6	0.00

chart continued on following page >

Figure 9: Top Passwords for the Top Ten Usernames (continued)

Username	Password	Count	Percent
pos	focus	716	0.32
	pos	490	0.22
	x00x	176	0.08
	1	167	0.08
	12345	161	0.07
demo	Welcome!	1689	0.76
	demo	231	0.10
db2admin	db2admin	1544	0.70
	spider	110	0.05
Admin	Pp123456	181	0.08
	Admin	158	0.07
	1	142	0.06
	admin@123	69	0.03
	123456	67	0.03
sql	Lassword32	705	0.32
	.....	139	0.06
	....	86	0.04
	....	65	0.03
	....	63	0.03

Top Usernames Associated with the Top Ten Passwords

We can also look at the reverse: how many passwords are reused across several usernames? This is a common tactic for quick dictionary testing, especially in environments that may have account lockouts in place. An attacker may only have five guesses for a particular username, but given 10,000 users, that amounts to 40,000 guesses with four passwords each (leaving the last guess off to avoid triggering the lockout).

In Figure 10, we can see that "x" is, by far, the most commonly guessed password. Would-be intruders are likely banking on the fact that POS and kiosk administrators may not realize their device is reachable from the Internet, and would rather not set a password at all. We can also see that the "St@rt123" password is associated with exactly one user account, "user1," just as the password "alex" is almost entirely associated with the username "alex." This would imply that perhaps "user1:St@rt123" and "alex:alex" are default credentials to a particular brand of device, or even a particular botnet default.

Figure 10: Top Usernames Associated with the Top Ten Passwords

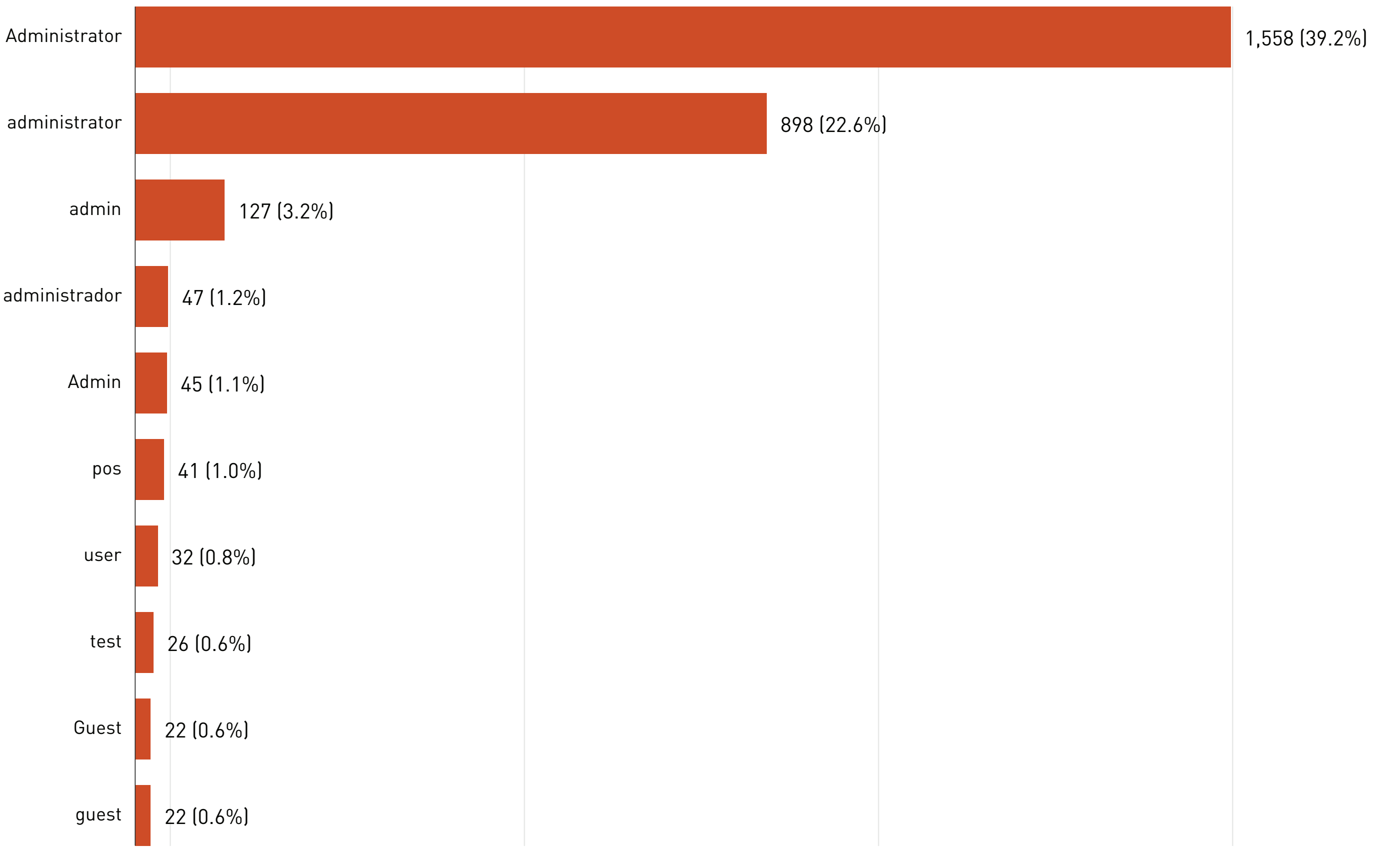
Password	Username	Count	Percent
x	administrator	9623	4.35
	Administrator	2120	0.96
	ETB user	93	0.04
	SPex	12	0.01
	nmelrose	5	0.00
Zz	administrator	6724	3.04
	Administrator	3851	1.74
	123	16	0.01
St@rt123	user1	8014	3.62
1	Administrator	2327	1.05
	administrator	1572	0.71
	1	562	0.25
	pos	167	0.08
	Admin	142	0.06

chart continued on following page >

Figure 11: Top Usernames Associated with the Top Ten Passwords (continued)

Password	Username	Count	Percent
P@ssw0rd	administrator	3371	1.52
	Administrator	1778	0.80
	well	141	0.06
	pratol	41	0.02
	admin	36	0.02
bl4ck4ndwhite	Administrator	4877	2.20
	administrator	251	0.11
admin	Administrator	2121	0.96
	administrator	1419	0.64
	admin	1007	0.46
	Administraator	89	0.04
	superior	37	0.02
alex	alex	4030	1.82
	administrator	2	0.00
....	administrator	375	0.17
	SUPPORT_388945a0	280	0.13
	Administrator	217	0.10
	IUSR_WWW	200	0.09
	Guest	178	0.08
administrator	administrator	1505	0.68
	Administrator	459	0.21
	administrador	215	0.10
	admin	36	0.02
	Admin	22	0.01

Figure 12: Distinct Passwords per Username



Distinct Passwords per Username and Distinct Usernames per Password

When studying online dictionary attacks, it's helpful to realize that while the password is "private" and the username is "public," they are really two equal halves of a valid credential. If you happened to write an e-mail password on a piece of paper, but never reveal the associated e-mail address, you haven't effectively compromised your account. You are simply reversing the roles of username and password, as long as you keep the username secret. Of course, keeping usernames secret is far more difficult, since usernames tend to be publicly disclosed many, many times in normal operations. Passwords, on the other hand are "disclosed" only upon authentication, and ideally over an encrypted channel.

This is a real problem for online attacks where very little about the system or its users is known. As seen in Figure 12, the scanners are clearly targeting well-known, non-secret default Windows accounts, such as "administrator," default usernames associated with a particular product, such as "pos," or guessing at a commonly chosen username, such as "admin." Figure 13 is dominated by these easy guesses.

Conversely, the guessed passwords do not clump together as strongly as guessed usernames. The top password, "password," is only associated with 2.95% of the recorded events. This indicates that attackers are not spending a lot of effort guessing one password across many usernames, but instead are guessing many usernames with relatively unique passwords for each.

Well-Known Passwords

In order to assess the commonality of the collected passwords across traditional password dumps found in other research reports, we referenced two primary sources of leaked password data.

Figure 13: Distinct Usernames per Password

Password	Username count	Username percentage
password	50	2.77
1	49	2.71
....	42	2.32
123abc!@#	39	2.16
P@ssw0rd	28	1.55
123456	26	1.44
admin	22	1.22
1x	21	1.16
......	20	1.11
123	19	1.05

07
PASSWORD COMPLEXITY AND PROVENANCE

In 2012, Dan Wheeler at Dropbox, Inc. wrote a fascinating piece on password strength estimators[12], and how Dropbox measures password complexity. Most importantly, the complexity algorithm and implementation was released as zxcvbn[13].

Zxcvbn

Zxcvbn is hosted on a GitHub repository and was released by Dropbox with a permissive open source license. Rapid7 data scientists and software engineers absolutely love well-cared-for open source projects, so we have adopted zxcvbn as a means to measure the complexity of Heisenberg-collected passwords. By running our collected passwords through zxcvbn, we can approximate "complexity" with zxcvbn's crackability score. Each password, therefore, is assigned a complexity level ranging from zero to four, with four being the most complex. The complexity level accounts for length, entropy, and common keyboard patterns.

Password Dumps

In analyzing the possible sources and patterns of the collected passwords, we also consider if these passwords are present in common, well-circulated password dumps. In order to do this, we referred to two primary password dump sources (Figure 14).

The first is the CrackStation's 1.5 billion wordlist[14], assembled over the course of 2010 to 2014 from various sources, including not only password dumps but the unique words found in Wikipedia and Project Gutenberg, both circa 2010.

The second is Mark Burnett's ten million credential sample, released February 9, 2015 in response to the January 2015 sentencing of journalist Barrett Brown[15].

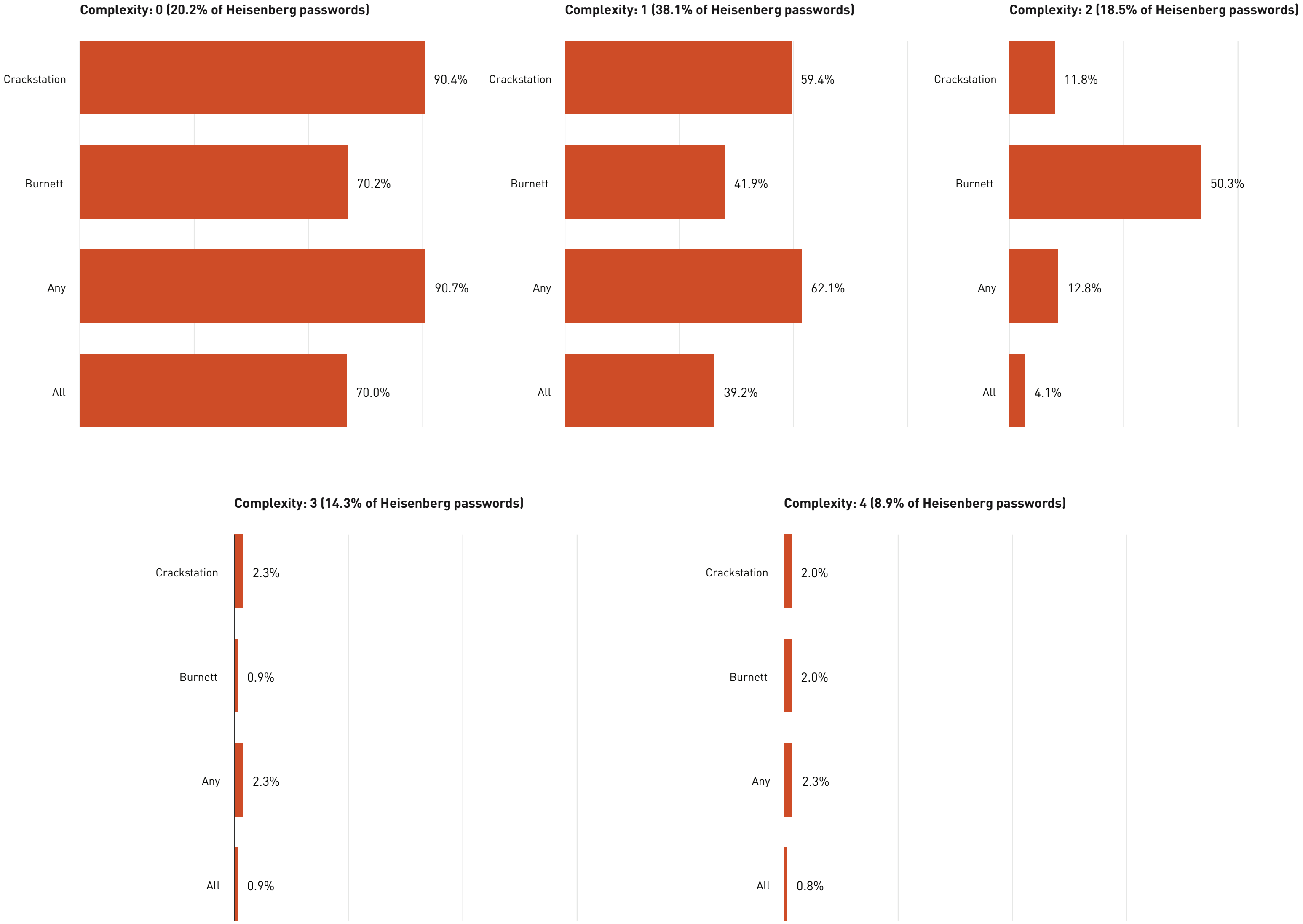
[12] Wheeler, D. (2012, April 10). Zxcvbn: Realistic password strength estimation [Web log post]. Retrieved from https://blogs.dropbox.com/tech/2012/04/zxcvbn-realistic-password-strength-estimation/

[13] Wheeler, D. (2012). Dropbox/zxcvbn. Retrieved February 19, 2016, from https://github.com/dropbox/zxcvbn

[14] Hornby, T. (n.d.). CrackStation's Password Cracking Dictionary. Retrieved February 19, 2016, from https://crackstation.net/buy-crackstation-wordlist-password-cracking-dictionary.htm

[15] Goodin, D. (2015, February 9). Fearing an FBI raid, researcher publishes 10 million passwords/usernames. *Ars Technica*. Retrieved from http://arstechnica.com/security/2015/02/fearing-an-fbi-raid-researcher-publishes-10-million-passwordsusernames/

Figure 14: Password Complexity



Ranked Password Complexity

Figure 14 illustrates how complex the collected passwords are, as a body. For each complexity rating, it shows whether they appear in the CrackStation or Burnett password lists.

For instance, there were 352 passwords classified as being complexity level 4, 7 of which were found in the crackstation dump, and 4 of which were found in the Burnett sample. Futhermore, 8 of the passwords were found in at least one of the password dumps, meaning that if we were to use all the available leaked passwords, we would only find 2.27% classified with a complexity value of 4. Similarly, we would find 3 (0.85%) passwords in each of the dumps.

Figure 15: Password Lifetimes

Complexity	Number of passwords	Percent	Average timespan (in days)
0	803	20.23	65.42
1	1512	38.10	48.10
2	735	18.52	31.51
3	567	14.29	23.40
4	352	8.87	27.56

Password Lifetimes

The final statistic we can measure about the passwords themselves is their individual and average lifetimes; that is, how often we see these passwords show up on opportunistic scans.

Over the 334 days we have been collecting this data, we found that a given password has an average lifetime of 43.18 days. Now, just like with human average lifespans, there are some consistent outliers with very long lifespans, such as "P@ssw0rd," "x," and the like. But more interestingly, some are extremely short-lived. 796 of our collected passwords were seen once and only once, which account for 20% of the total unique passwords collected. Figure 15 illustrates the distribution of complex passwords over their average lifespans.

While many of these passwords are typical of the complexity and patterns of the overall set, some are quite specific to POS vendor and brand names, and may represent newly learned default credentials acquired by the criminal groups seeking out these devices.

08

SOLVING THESE PROBLEMS

Evidence from our Heisenberg honeypot RDP credentials analysis shows that there are definite risks associated with exposing RDP to the Internet. Opportunistic attackers search for these services on a daily basis and have developed playbooks for how to find ones that are most vulnerable. Here are some steps defenders can take to mitigate the exposure associated with these services.

Securing RDP

One of the most straightforward actions any enterprise can take is to assess the need to expose RDP to the Internet. If these services are not absolutely critical, immediate action should be taken to block access to port 3389 (the most common and default port for RDP services). If access is required, a little security by obscurity will go a long way; simply changing RDP's listening port, and configuring the corresponding clients to attempt this alternate port, can alleviate most opportunistic scanning.

Employing a VPN solution can also help significantly improve your "cyber hygiene." While RDP does offer very reasonable encryption capabilities, it is still an exposed service that is commonly targeted. Cordoning off all resources behind a VPN – if they are not intended for the general public – does introduce some complexity for the authorized end user, but it virtually eliminates the threat from casual scanners.

Many other solutions are available, ranging from enforcing account lockouts to more complex, adaptive user behavior solutions that recognize and react to credential-based attacks in progress. However, we don't believe this advice will be heeded by the majority of victim organizations. That may sound harsh, but remember that it is likely that many (likely, most) of the RDP endpoints discovered by Project Sonar are exposed unintentionally, so no amount of security advice will reach the people responsible for these systems. Unfortunately, it's difficult to know for sure without further investigation. Periodic credential sweeps in order to size the problem appropriately would be most helpful but are blocked by legal barriers.

Chilling Effects and Legislative Bug-fixing

Ideally we would include a section in this paper that assessed the actual threat exposure to the credentials collected by Heisenberg, which might help organizations manage their risks. Now that we are armed with the sources, usernames, passwords, complexity patterns, and password provenance – all associated with real-world scanning evidence – we should expect that the security research community would be in an excellent position to determine exactly how effective these tactics are against Internet endpoints on a continuous basis. We would immediately start working with regional CERTs and other coordinating bodies to clean up the easily-compromised RDP space, which we have also identified via Project Sonar. After all, RDP-based break-ins are strongly associated with both point of sale and personal banking attacks, as indicated earlier in this paper, and is a continuous problem for many companies, large and small. We strongly believe that routine credential scans can help protect both individual and organizational sensitive information, prevent opportunistic fraud and crime, and ultimately make the Internet a safer and more trustworthy platform for commerce, communication and expression.

However, current U.S. computer crime laws pose barriers to testing the passwords on systems connected to the Internet. As a result, it is risky for researchers with good intent to undertake this cleanup effort. Specifically, attempting to validate these collected credentials against a sample on the Internet, no matter how controlled to prevent harm, would almost certainly run afoul of the Computer Fraud and Abuse Act since it would necessarily involve "intentionally access[ing] a computer without authorization."[16] Analogous laws in jurisdictions around the world similarly prohibit such activity. In addition, even "trafficking" in these credentials is somewhat legally risky under the CFAA.[17]

Given that there is no information security research organization body in the world that is authorized to perform such research, and sharing a corpus of raw credential data appears to be prohibited even for beneficial research purposes, we are unfortunately at an impasse with the current legal landscape. However, these laws are not preventing others from undertaking these credential scanning efforts and sharing their own results, likely with malicious motives, and we believe this paper demonstrates this reality.

Actively assessing the hygiene/exposure of systems takes a fair amount of resources that not every computer owner has or is willing to dedicate. A community of security experts seeks to help, but overbroad and outdated laws often chill the research needed to identify the problems and develop solutions – all while criminal activity continues and systems remain exposed. The law should not drive this positive activity underground – we need to engineer an effective regulatory system that manages controlled security research in a manner that minimizes harm to both computer owners and the researchers themselves. Unfortunately, that is not the system we have today.

At Rapid7, we understand that it is tricky to draft, debate, and pass legislation that distinguishes between security researchers and malicious actors. Simplistic solutions tend to break down in complex real world settings, and overly complex solutions can result in confusion and unintended consequences. This is not an easy problem to solve, but Rapid7 is dedicated to addressing it. That is why we are currently working with attorneys, lobbyists, and legislators to help craft legal and policy solutions that both deter crime and encourage research.

Government officials are starting to publicly acknowledge the value of independent security research more openly and frequently than they did even five years ago. For example, the United States Library of Congress did approve a DMCA exemption for "good-faith security research" in its 2015 round of rule-making[18], and the Bureau of Industry and Security (BIS) in the U.S. Department of Commerce is now actively seeking input from Google and other organizations concerned with more open security research[19]. There is growing collaboration between researchers, government agencies and legislators; we believe in this approach and are optimistic that it will continue to yield positive outcomes. But acknowledgment and collaboration are not law, and legislation is needed to modernize regulations that too often construe research as criminal behavior and ultimately undermine cybersecurity. We will continue working to improve cybersecurity law in a balanced and responsible way, and we hope others will join us. It may take a long time to see significant change, but we will only get there by working together.

[16] 18 USC 1030(a)(2).

[17] 18 USC 1030(a)(6)(B).

[18] Ellis, J. (2015, October 28). [Web log post]. Retrieved from https://community.rapid7.com/community/infosec/blog/2015/10/28/new-dmca-exemption-is-a-positive-step-for-security-researchers

[19] Peterson, A. (2015, July 15). The government is headed back to the drawing board over controversial cybersecurity export rules. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2015/07/29/the-government-is-headed-back-to-the-drawing-board-over-controversial-cybersecurity-export-rules/

09 ABOUT RAPID7

Rapid7 is a leading provider of security data and analytics solutions that enable organizations to implement an active, analytics-driven approach to cyber security. We combine our extensive experience in security data and analytics and deep insight into attacker behaviors and techniques to make sense of the wealth of data available to organizations about their IT environments and users. Our solutions empower organizations to prevent attacks by providing visibility into vulnerabilities and to rapidly detect compromises, respond to breaches, and correct the underlying causes of attacks. Rapid7 is trusted by more than 5,100 organizations across 99 countries, including 37% of the Fortune 1000. To learn more about Rapid7 or get involved in our threat research, visit www.rapid7.com.