2015 – Year of the Dumps

romcheckfail.com

Mike Hiltz
2015-12-30

Methodology for this analysis:
130 dumps were analyzed from various sources with only 111 containing passwords in a clear text or a hashed manner. In addition 3 handpicked dumps were added to the list due to the relevance in 2015, they are Ashley Madison, Mac Torrents, and 000webhost. Since the stats in this file are based off type of hash found most commonly in the file, not number of hashes these larger lists will not throw off the other 108 randomly selected dumps. The remaining 19 dumps were not included as they did not contain login information.

Once the dumps were captured they were removed of lines that did not contain a username and password then the number of accounts were tallied up by line number. The hash type was determined from software called hashID[1] that reads the length of the hash to give the supposed type such as MD5, SHA1, and various other hashing algorithms. A random sample of hashes were run through hashcat[2] in order to determine what hash type was used in the event of similar length hashes such as ones that are 32 characters in size that can belong to multiple algorithms.

The language section is based off the written site language at the time of viewing. Either in real time or via the use of archive.org's WayBackMachine if the site was closed or down. However in the event the site was not cached by archive.org or currently available in real time then the TLD[3] was used, this only occurred in 2 of the 111 samples.

The country of operations was determined through the address on the Contact US or Social Media accounts tied to the websites. In the event of there being multiple office locations the head office was used.

The services section was sourced based off visible or displayed services and was broken down into a few generic categories listed at the end of this paper for reference.

Things to consider about this paper before you continue.
This is just a small sample of the hundred of leaks and dumps posted publicly on the web every day, this does not include leaks that are traded on private forums or in various dark web circles.

Not all dumps were verified against hashcat to reverse algorithm used, only a small sample so it is possible the numbers are not 100% accurate. This was done in the interest in time.

Some dumps contained multiple algorithms as the site matured. For example, if a site is running an older version phpBB[4] then upgrades not all the accounts are migrated to the newer stronger hash method, only new users that sign up after the phpBB update. Due to this the higher majority of hashes in a file were counted as a whole for that individual dump so this may also shift the numbers slightly.

The anatomy of a dump.

In the last year the release or hacking of private information has been highlighted in the media, gaining the attention of the general public. This was marked by two major breaches this year when the site AshleyMadison.com was hacked and later vTech. This does not mean dumps are new though; since there has been this need to retain data, there has been a want to leak it. The data contained in online dumps vary as they depend on the release.

```
--
-- Database: `hostiles_stresser`
--

-- --------------------------------------------------------

--
-- Table structure for table `users`
--

CREATE TABLE IF NOT EXISTS `users` (
  `ID` int(11) NOT NULL AUTO_INCREMENT,
  `username` varchar(15) NOT NULL,
  `password` varchar(40) NOT NULL,
  `email` varchar(50) NOT NULL,
  `rank` int(11) NOT NULL DEFAULT '0',
  `membership` int(11) NOT NULL,
  `expire` int(11) NOT NULL,
  `status` int(11) NOT NULL,
  `key` text NOT NULL,
  `used` int(2) NOT NULL,
  PRIMARY KEY (`ID`),
  KEY `ID` (`ID`)
) ENGINE=MyISAM  DEFAULT CHARSET=latin1 AUTO_INCREMENT=54 ;

--
-- Dumping data for table `users`
--

INSERT INTO `users` (`ID`, `username`, `password`, `email`, `rank`, `membership`, `expire`, `status`, `key`, `used`) VALUES
(3, 'EggsOnBacon', '5bf79f7d076b9cfcd4cd6b124887887026420f45', 'colbyburke98@hotmail.com', 0, 22, 1553963000, 0, '', 0),
(4, 'unwoundah', '2a32c5f53eda76ba7269563625aedd3427795e5a', 'akber.ahad@gmail.com', 0, 21, 1553986930, 0, '', 0),
(5, 'Arnout', '99efc50a9206bde3d7a8e694aad8e138ca7dc3f7', 'arnout@yoloswag.me', 0, 0, 0, 0, '', 0),
(6, 'Basi88', '9cbdeba523901d4fa68347772739bd44bc1b2b44', 'b.onzenoort@kpnmail.nl', 0, 0, 0, 0, '', 0),
(7, 'illustrious', '532a3258eda304b335e062ebc8ff83cec026d909', 'illustrious@live.co.uk', 0, 0, 0, 0, '', 0),
(8, 'Abashed', '0a0529517c16fe68fe68c5b735948ea116479f74', 'palt.gaming@gmail.com', 0, 0, 0, 0, '', 0),
(9, 'AnonyArme', 'd323500dc3a26a6302ea7e4b27f6847f4d63ebdc', 'anonyarme@gmail.com', 0, 0, 0, 0, '', 0),
(10, 'darkestyank33', '7eeb7557e5f46a360f5191991824cd658a3c7fda', 'crbellflower@gmail.com', 0, 0, 0, 0, '', 0),
(11, 'SamBills', '0ab442d6761100f755a8b08d6410945a6ca940ca', 'joe.rezz12343@gmail.com', 0, 0, 0, 0, '', 0),
```

Sample of a DB dump.

But on average a dump will contain the following items:

Releaser name - The hacker or group that released or posted the dump.
Statement of release - This is not always included but can be in the event of *#OPIsreal* or other politically motivated attacks.
Usernames - This is what you use to log into the site or your alias used on the site to identify yourself.
Emails - The email address used to reset passwords, or used to create the account.
Password - This can be hashed, hashed and salted, or simply plaintext (unencrypted).
IP Addresses - This is the Internet Protocol address used by each account to log in. If included it may show IP addresses that the account was both: created from, and last logged in from.
User Level - Displays the amount of access the user has to the site if they act as an admin or general user.
Other important data - This can be sexual preferences or contact phone numbers on the account as well as date of birth. Even secret questions and answers combinations. This can also contain private messages (more commonly known as PMs) between users.
Junk data - Often releases can contain junk data in them, this can be if the user has activated the account by email or if an admin has done it for them or other low level information.
Credit Card Data - This is generally not found in dumps posted publicly unless it is a dox[5]. generally traded in carding forums and sites, rarely found in website dumps.

In the end it comes down to what people are looking for in the dumps. I have included descriptions as to why these fields are important and how they can be used against the site admin and end user of the service.

> **Usernames** - This can be used to cross reference sites for the same user. Most individuals don't use different usernames on different sites.
> **Passwords** - Even when hashed if the algorithm is simple enough the password can be cracked, when combined with the username someone can take over the account and then impersonate the user.
> **IP Address** - IP addresses can be run though a reverse DNS in order to identify government or business entities. In small cases where users have dedicated IP addresses at home it can also identify a specific person.
> **Other Important Data** - In the case of Ashley Madison it was seen that sexual preferences and private messages were leaked allowing people to tie names and emails to peoples personal lives. This can lead to extortion or impersonation.
> **Credit Card Data** - Extortion and financial fraud.

How do people originally obtain dumps?
This is much beyond the analytic scope of this paper, however it is the belief of the writer that it is done most commonly from SQL injection and or insecure servers being compromised. The second method would just be from other dumps of a similar nature with the same or similar admins. Often when I see dumps posted in succession online they are of similar structure. What I mean by this is when 3 dumps are posted by the same team around the same time they often have a motive or unique factor that ties them together. Such as targeting a bio-tech industry and releasing 3 companies at once. Or more commonly a specific version of Apache or PHPAdmin that may have a CVE lodged against it. Even though the sites have completely different content and target user groups they would fall under the same vulnerability scan carried out by the group.

Where dumps are found.
I will not be going into too much detail into this section but the fact is dumps are public, as public as can be. Most are posted on pastebin.com, justpate.it, and various other sites. They remain there indefinitely unless flagged and removed by a admin. Sites like ghostbin.com have a timeout setting that allow the dump to only be available for a specified amount of time and allow for encryption. There are also private forums that contain dumps, earlier this year hell forums[6] (hell2bjhfxm77htq.onion) was shut down and was a common place for users to trade, buy, and sell dumps. More recently there have been a number of onion URLs that offer virtual marketplaces that sell dumps. They deal in bitcoin and prices can range from 50$ CAD for small public dumps to 17,000$ CAD for privately held million user dumps of user names and passwords.

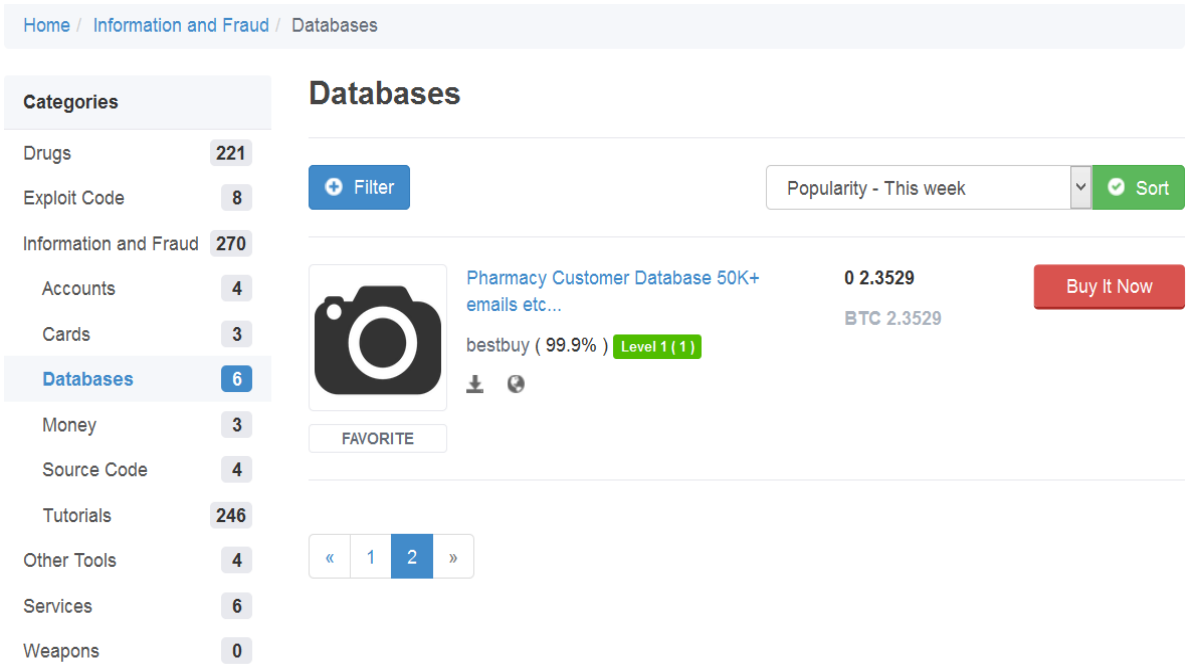
What causes a group to release data?
There are a number of reasons that data is dumped. It is my belief that the number one reason is credit or rep to the team or hacker. This is closely followed by revenge hacks or politically motivated hacks. Embarrassment is also another reason people will dump information about a specific company, if they refuse to pay bug bounties, or a competitor wants an edge on them as was the case in Ashley Madison. [7]

Why would people want dumps?
There are a few reasons why people would want dumps, but I would like to be clear there are legitimate reasons why you would want to collect and analyze them. The majority of reasons would fall into the following categories:

Exploitation - This is common when you look at the types of sites dumped. From our samples, the number one site dumped was dating followed by community. Dating contains a high amount of personal information and generally payment information. Where community would generally be put under the label as hacks of revenge or simply trolling.

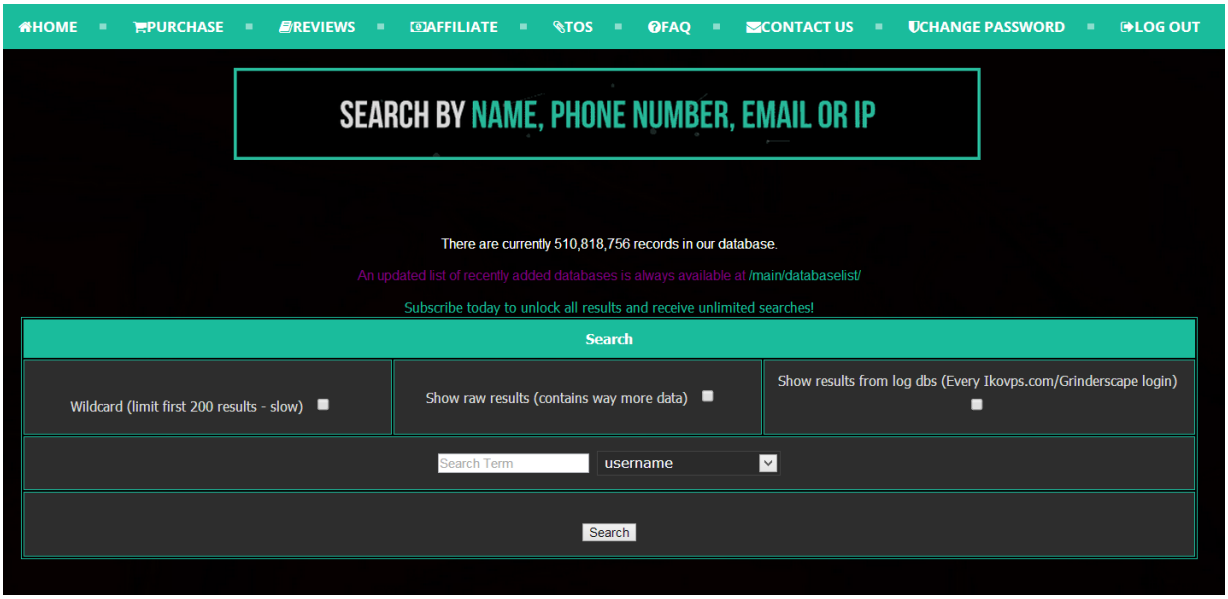
Defacement - This would be the act of gaining administrator level access to a site in order to deface it, or infect unknowing visitors. Generally governments, Police, or other authority in power. This would support the anarchist movement ideology but could also be terrorist motivated.

Revenue - This can been seen in a new trend of sites that have been popping up, explained in more detail below where lists are aggregated into larger lists that are then indexed for search. This can be used to identify people of interest. These sites charge a fee for look up and for removal. Also the sale of database dumps can fetch a high profit.



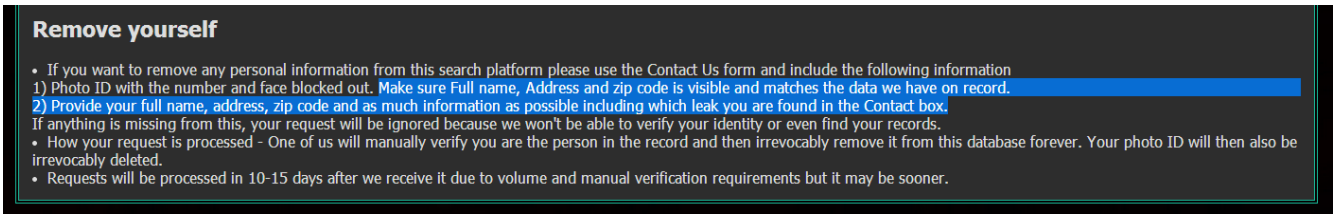
Capture of a darkweb mark selling DBs

The emerging market of selling you back your information.
Since midway in 2015 there has been a frightening trend that has started to appear regarding sites selling users back their leaked data. These sites business model is collecting leaks public and private, and indexing them for search and sale.



Capture of a indexing site for account search.

What that means is they will take a dump, sort the data so that it can be searched and then allow anyone to take a look. Generally on these sites the search is free but viewing the result is not. Prices can be from 2$ a day for access. These markets are clearly marketed towards you looking up someone else not yourself. In the event you look up and find yourself, you can send them 10$ to remove a single. In addition to the fee, you also need to send sufficient identifiable information to them such as a image of a drivers license or ID. The problem is that these sites are springing up all over in an attempt to be the a pioneer of this kind of gray-market service right now. I don't think people realize how frightening of a trend this can become.



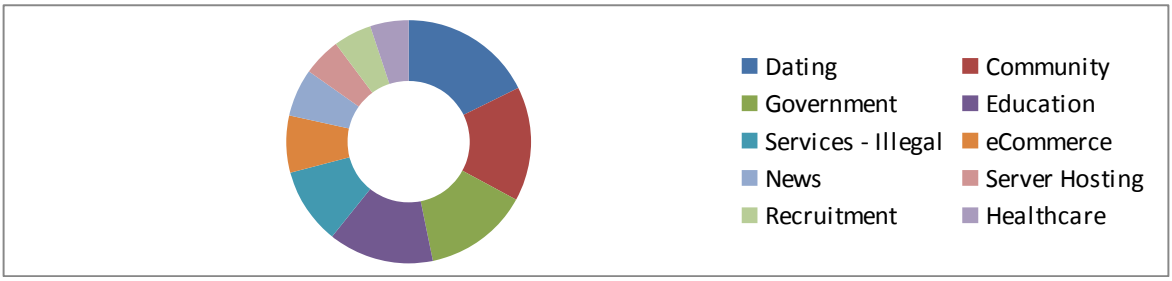
Capture of TOS.

Password Analysis / Analytics / Forensics - Here we get into the lighter side of the spectrum. Password analysis helps us understand how people write passwords, what makes a bad password, and how we can make them better. Without this type of work we would not have papers like these. The passwords that are leaked can also be put into word lists for data forensics to prevent crimes like child exploitation.

Positive Exposure / Awareness -
Troy Hunt has a excellent project on this right now[9]. Without his site HaveIBeenPwnd.com users would have to download the list themselves and see if they are in it. Most lists are large and cannot be opened in notepad or word and can require special SQL tools and regex knowledge to separate the data; Troy makes the process simple. The more important thing is shame through awareness. The site lists all the major leaks, a good amount of pastes, type of information leaked, and if the passwords were encrypted so you know what sites you are on, and what passwords you need to change. In the past, this has always been looked down on, and the sharing of data for use in an analytical product just hinders us from being more secure[9]. This also can stir up a few legal teams.[10]

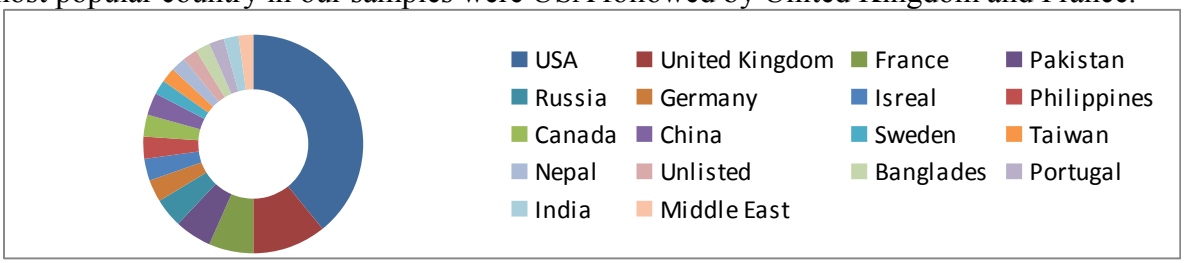
The Numbers.

Here is the break down from our sample of leaks.
The most common industry in our samples were Dating closely followed by Community and Government.



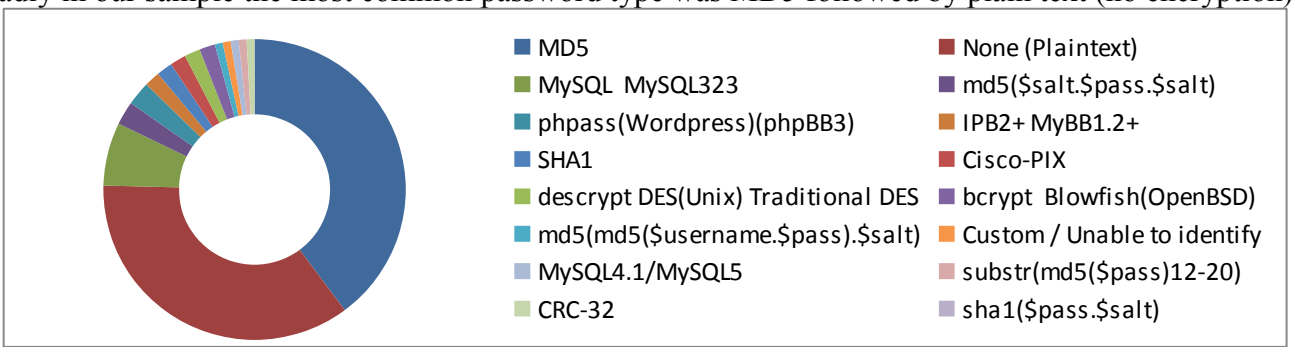
Dating	14	News	5	Music	3	Gaming	2	Religion	1	Software	1		
Community	12	Server Hosting	4	Infrastructure	3	Legal	1	Engineering	1	Sports	1		
Government	11	Recruitment	4	Games - Kids	3	Electronics	1	Hotel	1				
Education	11	Health care	4	Gambling	2	Manufacturing	1	Auto Indus	1				
Services - Illegal	8	Banking	3	Classifieds	2	Religion	1	Supplies	1				
eCommerce	6	Commerce	3	Sightseeing	2	Engineering	1	SEO	1				

The most popular country in our samples were USA followed by United Kingdom and France.



USA	36	Canada	3	India	2	Greece	1	Mexico	1
United Kingdom	10	China	3	Middle East	2	Delhi	1	Venezuela	1
France	6	Sweden	2	Burma	1	Spain	1	New Zeland	1
Pakistan	5	Taiwan	2	Brazil	1	Denmark	1	Ireland	1
Russia	4	Nepal	2	Mumbai	1	Venezuela	1	Cayman Islands	1
Germany	3	Unlisted	2	Yemen	1	Iraq	1		
Isreal	3	Bangladesh	2	Vietnam	1	Hungary	1		
Philippines	3	Portugal	2	Bangkok	1	Poland	1		

Sadly in our sample the most common password type was MD5 followed by plain text (no encryption).



MD5	47	phpass(Wordpress)(phpBB3)	3	Cisco-PIX	2	Custom / Unable to identify	1
None (Plaintext)	42	IPB2+ MyBB1.2+	2	descrypt DES(Unix) Traditional DES	2	MySQL4.1/MySQL5	1
MySQL MySQL323	8	IPB2+ MyBB1.2+	2	bcrypt Blowfish(OpenBSD)	2	substr(md5($pass)12-20)	1
md5($salt.$pass.$salt)	3	SHA1	3	md5(md5($username.$pass).$salt)	1	CRC-32	1

The average number of accounts per dump was 5074 , this did not include 30 Million from Ashley Madison or 13 Million from 000Webhost as they were hand picked and not part of the sample.

Now you know a little more about dumps, how they work, and where they come from. This brings us to what can be done about it. The solution to privacy online is this: use 2FA when possible, don't use the same password on multiple sites, and be aware there is nothing you can do to prevent something you typed into the computer from becoming public at some point in your life.

Outlook for next years report:
Statistics on prices of dumps vs freshness and number of passwords.
Fluidity of dump site markers.
Update on emerging market of selling access to aggregated dump information.

[1] HashID – Software used to identify the algorithm used to store a password.
https://github.com/psypanda/hashID

[2] hashcat – Software used in brute forcing hashes to find a match.
http://hashcat.net/oclhashcat/

[3] TLD – Top Level Domain – Example .ca – Canada , .us – United States

[4] phpBB – A common community forum software similar to Invision Powerboards

[5] dox – internet-based practice of researching and broadcasting personally identifiable information.
https://en.wikipedia.org/wiki/Doxing

[6] http://www.romcheckfail.com/hell-forum-closed-administrator-ping-arrested-hell2bjhfxm77htq-onion/

[7] http://krebsonsecurity.com/2015/08/leaked-ashleymadison-emails-suggest-execs-hacked-competitors/

[8] https://haveibeenpwned.com/

[9] http://techcrunch.com/2015/02/10/a-security-researcher-just-dumped-10-million-real-passwords/

[10] https://www.thecthulhu.com/legality-of-dumped-data/