Password Requirement Analysis of 100 Major Internet Sites

Keith Major, USMA and Bruce Barnes, USMA

Abstract

Many sites have different requirements for the complexity of the password required to login. There is no prescribed industry standard, just site specific recommendations for passwords. This causes specific sites to have inherently weaker passwords than their similar counter-parts. We have created a comparative analysis of one hundred major on-line websites in thirteen sectors that illustrate the disconnect between information sensitivity stored on the site and the minimum password strength used to protect it. Through this comparative analysis, we have created a web application allowing the user to input a password and be able to identify its strength and compare it to the password requirements in the analysis.

Introduction

Passwords are used all across the Internet today. We do banking, shopping and socializing on the web. Many of these sites require registration and have different requirements for the complexity of the password. There is no industry standard for websites that require a high level of confidentiality, including financial institutions. USAA.com, a banking website, and yahoo.com, an email service, both require passwords over six characters. Security focused websites classify these passwords as weak, but they are acceptable for other websites.[1] It seems logical that a banking site should require a stronger password than an email account. A weak password, in our day and age, is leaving the user vulnerable to an attack. The amount of information that could be compromised as a result of a weak password is immeasurable. Identity theft is becoming a crime that ruins people's lives. Should someone break into an account on the Internet they will have access to very personal and sensitive information. A person can cause mental, emotional and financial harm to others if they have access to a person's account. The login is the first line of defense against harmful attacks against a person. The password is meant to provide authentication and then give that user access to a site they are authorized to see. Any person can get authorization to a site by having the correct username and password.

Passwords are usually coupled with a username; both of which are stored in a website's database. When a user goes through the login process the username is in plain sight. The password is not shown to the user but is usually a series of dots. An attacker can use "shoulder-surfing" to determine a person's username, and on the rare occasion identify keyboard strokes for the hidden password.[2] When a user submits their username and password pair at login, those values are sent to the website's server which has a database that checks to see if an entry exists with that username and password pair. If the two do not exist together the user receives an error, but if the pair does exist in the database, then the user gets authorization for certain information on that website. Therefore, it is critical to have a strong password.

The focus of this paper is on the strength of the password and its correlation to the type of information stored. The password strength is determined by the time it takes to Brute-Force it. Brute-Force attacks are based on the attacker setting up a list of every possible password and systematically attempting each one. There are open-source programs available that will conduct the Brute-Force attack for the user; therefore it is easier for people to do. The best defense against a Brute-Force attack is a website that will freeze an account when the wrong password is entered for a set number of times. As a courtesy, some sites send an email to the user saying that an invalid password has been used to attempt to access the account.[3] However, if an attacker gets a database of the usernames and passwords, they can run a Brute-Force attack offline without restriction. Another example is sites that require mouse input or captcha images in addition to usual login and password requirements. For the purposes of this paper, security is only assessed by the username and password login credentials.

Our analysis of a password is based on the minimum length of the password, case sensitivity, allowance of special characters and numbers. For our study case sensitive will be considered upper and lower case letters in the English alphabet; numbers will be 0-9; special characters are any character not already mentioned that can be directly accessed on the keyboard, such as ':;!@_#$,]}*&'; space will not be considered as a special character. The reason that these requirements affect a Brute-Force attack is because each additional requirement adds more permutations for a possible password so the time taken to get the correct password increases significantly. We believe a strong password is one in which the attacker stops due to time constraints. A weak password is one in which the attacker can go to work and come back with access to the account. For most accounts, we believe a cracker will not spend more than a day attempting to crack it.

We will begin by discussing our collection methods for the 100 website passwords policies, followed by how we analyzed each policy into quantifiable data. Afterwards, we discuss the results found pertaining to top 10 password policy sites, information sensitivity versus time to crack, and average time to crack per sector. We then conclude with why this is important and how to build upon our work.

Collection

Collection of the data consisted of two steps: first deciding which websites to use in the study, and then finding the password requirements for those websites. We were not able to use lists like the "Top 100 Websites for 2007" because these lists contain websites that, for the most part, do not have a login/authentication requirement. Our list came from sites that we use or have used in the past and other sites that we think are common. This list came out to around eighty-five websites which is fifteen short from our goal. We then had to search for websites that required a login that, even though they are not as popular. Many of which we chose to further strengthen our data in a specific sector. Examples of this are the websites for "Sea Coast Medical" and "Churchill Medical", which are not big name medical suppliers, but their password requirements will strengthen our data for our medical sector.

After the list of websites was compiled, we went back to identify the actual password policies for each of the websites. This task was harder than expected, mainly because most websites do not publish their password requirements. Most sites, however, would let the user know of the minimum length of the password. Only through trial and error would the user know if the password that they had was acceptable or not. The tricky thing for us was to find out exactly what the password requirements for the website was without actually signing up for what the website offered. In many cases, the registration page on the website would allow us to test the password without entering in other registration requirements. Upon submission an error would pop up telling us some of the password requirements, namely the length requirement. Depending on how that website did its form validation check, we had to enter in values for the other fields before we could test out our password. The password field on most registration forms was at the bottom, so if the website did a sequential validation check, we would get errors for not entering in our "Name" or "User Name" before we would get feed back on our password. In order to find out the password requirements for these websites we were forced to enter in most of the information the site required. Once we signed up, we checked the password requirements using the change password form. We had to use this method around fifteen of the one hundred websites. The other websites we were able to get the password requirements by just entering in the password field into the registration form. Others by using our own accounts and testing using the change password form.

In order to test the websites validation we used passwords like "aa!!11" with extra 1's at the end to meet the min requirement. This would test to see if the password allowed for special characters and numbers on that site. Once we changed this password, we would attempt to login using "AA!!11." If the site is case sensitive it would state invalid login, however if it was accepted then the password policy did

not include case sensitivity. Some websites did not let the user know what the minimum length for the password was so when we tested the site we would start at 1 character password and then the site would usually respond with the minimum length requirement or accept the password; either way we were able to determine the minimum length required for the passwords. In some of the websites, we were not able to determine if the password allows for numbers, special characters, or if the password is case sensitive so we made an educated guess based on the other password requirements for similar websites. Using these techniques, we were able to determine the password requirements for one hundred websites, which we would use to analyze the strength of each of the passwords.

Analysis

Upon collecting all the data, we had to quantify it in such a way to show the proper results. We separated the data into 4 quantifiable portions: Minimum characters, case sensitive, special characters allowed, and numbers. We then further allowed for special requirements for the few sites that had it. We kept track of the information by creating Booleans for each portion. Once this data was collected, it needed to be compiled to show maximum characters that are available for the password. For example, a password policy that allows numbers and case sensitivity would allow for a-z, A-Z and 0-9, this allows for 62 different character combinations. Password policies that allowed special characters were usually assumed to have 33 possibilities (!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~). Almost every password did not allow for spaces.

After getting these results we needed to then determine how many different possibilities there are for each site. In order to do this we used permutations. This was simple for many of the sites because it required only taking the number of characters available raised to the power of the minimum password length. A site that had a 4 character minimum length with 62 available characters would be 62*62*62*62 or 62^4 or almost 15 million permutations. A few sites required at least one letter and one number in the password. Therefore, assuming a cracker would know the requirements and be able to adjust their password generator they would be able to reduce the guesses in accordance with that policy. For those sites, we calculated the permutations as simply

$$52 * 10 * \mu^{(\alpha - 2)}$$

Where μ = characters available, α = password length, 52 is a-Z, and 10 is 0-9.

Using information provided by Lockdown.Co.Uk, a Class D computer or a Dual-Core Processing Computer can calculate passwords at approximately 10 million a second.[4] Since many PCs today have Dual Core capabilities we used this as our baseline for calculating the time it would take to crack a password via brute-force. We divided the number of permutations by 10 million and then put it in terms of days; essentially the permutation divided by 10000000/60/60/24 to get the time to crack in days. This allowed us to graphically depict easiest to hardest to crack. The disclaimer with this method is that it is the *maximum* time needed to brute force crack the password, in many cases it can be achieved in half the time with some sort of hybrid brute force system.

Another aspect which we analyzed was the information sensitivity. We created a survey, completed by 42 individuals, to help us determine how they would feel on a scale of 1 (negligible) to 5 (devastating) if an account of a given type was compromised. Table 1 contains the data from our survey in terms of average rating from the 42 individuals and ranked ordered from most sensitive to least.

Type	Average Rating
Online Banking (Bank of America)	4.48
Credit Card Bill Account (Visa)	4.36
Taxes (Turbo Tax)	3.81
Medical Information (Tricare)	3.74
Shopping Site (Amazon)	3.43
Email (Yahoo)	3.33
Airline Account (JetBlue)	3.14
Social Networking (Facebook)	2.81
Photo Sharing (Flickr)	2.24
Media Site(Youtube)	1.93
Gaming (Second Life)	1.87
Blogging Site (BlogSpot)	1.86

Table 1: Survey Results of Most Devastating to Least

An interesting point to note is our survey sample was small and of mostly college students. If we had sampled more gamers, gaming perhaps would have been higher, or if we sampled people who make their living blogging and have a reputation on the Internet these results would likely have been different.

Table 2 contains data pertaining to what type of information is most sensitive to least sensitive. We used a focus group of 17 Information Technology students and professors to create a consensus of information sensitivity. Using these two tables we went through each website and ranked it from one to five. For example, our banking websites would get a rank of 5 and our blogging sites would get a rank of 2. Some sites were difficult in that they contained relatively more sensitive information than other sites in the sector. An example of this difference is with iTunes. We classified it as entertainment and according to Table 1 should have received about a 2, however since one can charge songs to the account we classified it as a 4.

Most:	5	Banking, Credit Card #, SSN
	4	Phone #
	3	Physical Address, DOB
	2	Social Info, Work Info
Least:	1	Screen Name, Email

Table 2: Information Sensitivity Scale

Using our scales for information allowed us to get at our thesis, the disconnect between information sensitivity and password security. The focus group decided that social security numbers, credit card numbers and banking account were the most sensitive data available online. The next is tying

the online identity to the physical location of the individual. We believed this was sensitive in that most individuals want privacy and anonymity when browsing the web. Below that is social info, marital status, work information, likes, dislikes, and other information that might not readily be made available to others. The lowest information sensitivity category is getting a screen name or email, a minor piece of information, but information nonetheless.

Results

Of the 100 websites, one of the first things we looked at were the top ten hardest to crack sites according to their minimum password policy. Looking at Figure 1, 6 of the 10 sites have the same number of days to crack and thus the same password policy. The password policy for these sites is 8 characters minimum and allows special characters, case sensitivity and numbers. In addition, these sites usually recommended, but did not mandate, using a strong password.

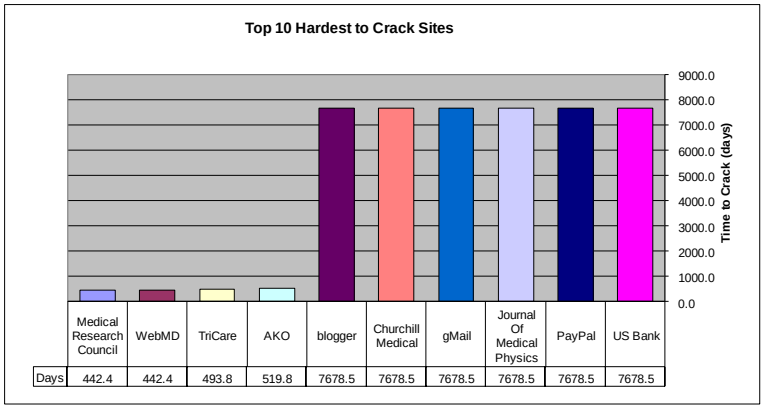

Figure 1. Top 10 Hardest to Crack Sites

As one can see from Figure 2, half of the Top 10 sites we analyzed were Medical in nature. No Shopping sites or Travel sites made it into the Top 10 strongest password policies, both of which usually store your credit card information for future purchases. This is surprising in that although medical information is private in nature, according to Table 1 losing money is more sensitive. However, as Figure 1 and Figure 2 show, Banking and Credit cards only comprise 20% of the Top 10.

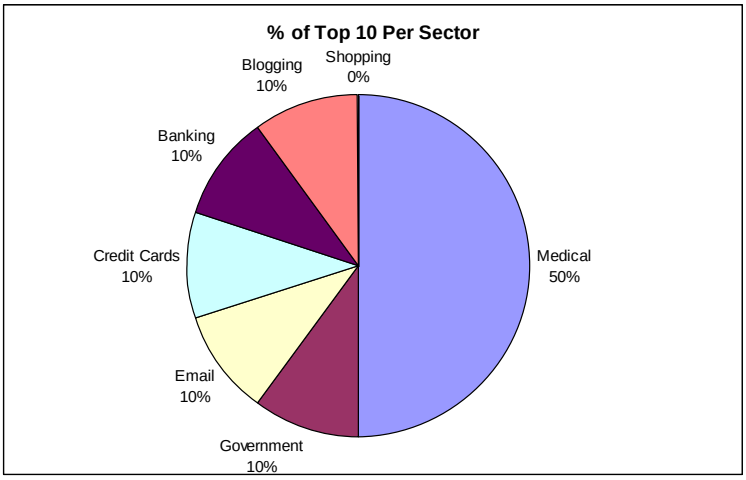

Figure 2. Percent of Top 10 in Each Sector

We classified the Internet into 13 sectors. There is overlap between the sectors, so we had to decide which sector a particular site would fit in. For example, World of Warcraft could have just as easily been placed into Entertainment, but we felt the social aspect and the value of the information behind the account placed it in the social sphere. Table 3 includes information of the top site per sector. An interesting note is the strongest password for taxes is HRBlock and Turbo Tax, both of which can be cracked in under a day. Remember this is the maximum amount of time to crack it. An attacker could potentially crack three of these passwords a day. In order to file taxes a user must use their full social security number, and thus cracking one of these sites would make it easy for identity theft.

Sector	Site	Time to Brute Force (days)
Banking	Bank of America	354.18
Blogging	Blogger	7678.48
Credit Cards/Bills	Paypal	7678.48
Email	gMail	7678.48
Entertainment	iTunes/Stage6/flickr	0.85
Government	AKO	519.82
Job Search	Monster	256.59
Medical	Medical Research Council/Churchill Medical/Journal of Medical Physics	7678.48
P2P	Bit Torrent/Napster	0.85
Shopping	Sustainlane	493.83
Social Network	World of Warcraft	442.42
Tax	HRBlock/Turbo Tax	0.85
Travel	Fed Ex	0.05

Table 3. Top Site per Sector

After finding the average time to Brute-Force a password per sector, we graphed the results as seen Figure 3 below. On Figure 3, we see that five out of thirteen of the sectors have an average time that is greater than 400 days to Brute-Force. These sectors would appear to have strong passwords throughout the sector, but if you look at the individual times, you see that large outliers eschew the data to make the averages larger. Six of the thirteen sectors take too little time to Brute force that they hardly show up on the graph or don't show up at all.

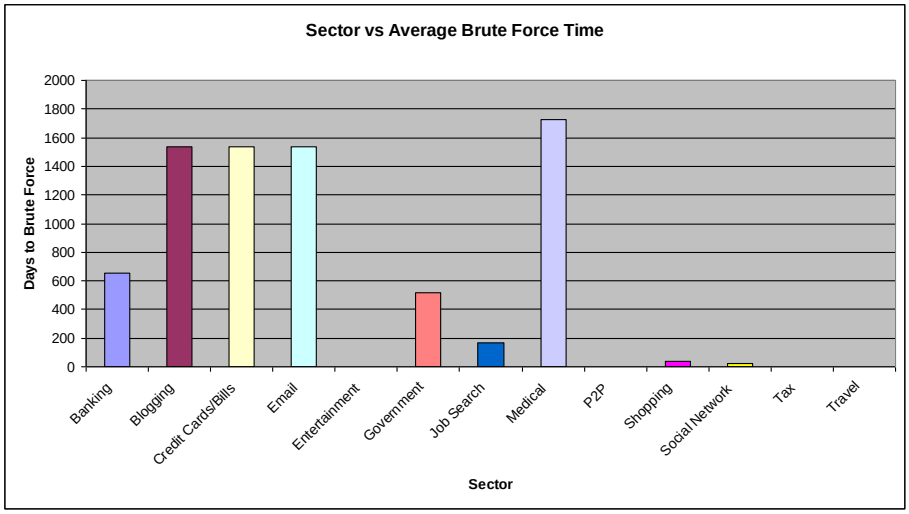

Figure 3. Average Time to Brute Force per Sector

In order to get a better analysis of the password requirements per sector we needed to take out the outliers in the data so we can find a better average, which is shown in Figure 4 and 5. The sectors were split for the purpose of analyzing the data. By taking out the outliers from the data for the graphs, we see that only four sectors; Banking, Government, Job Search and Medical, have average Brute Force time greater than one hundred days. As one can see between Figures 3 and 4 Banking dropped over 500 days on average to Brute-Force the passwords therein.

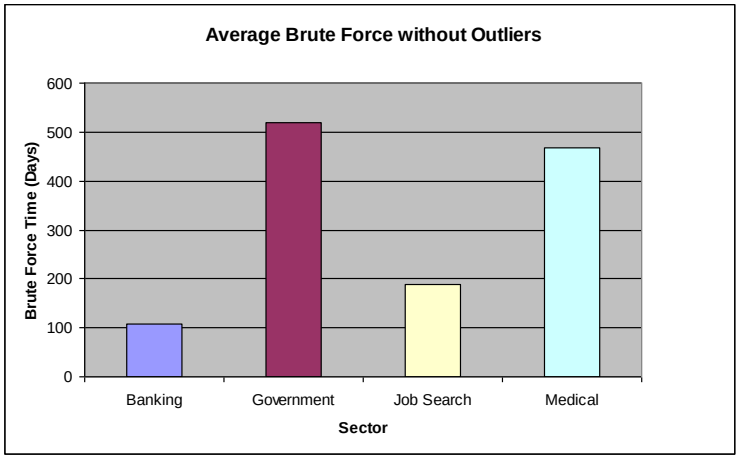

Figure 4. Average Time to Brute Force per Sector without Outliers

As shown in Figure 5, Email, Credit Card, Shopping and Blogging dropped from the strongest passwords as shown in Figure 3 to very weak passwords. The sectors in Figure 5 have weaker passwords in comparison because they can be brute forced in a little over a day or less. Removing the outliers gives us a better view of what the industry standard for each sector would be. The outliers, however, were usually the websites with the strongest passwords and set the top bar for password requirements that the other websites would want to reach.

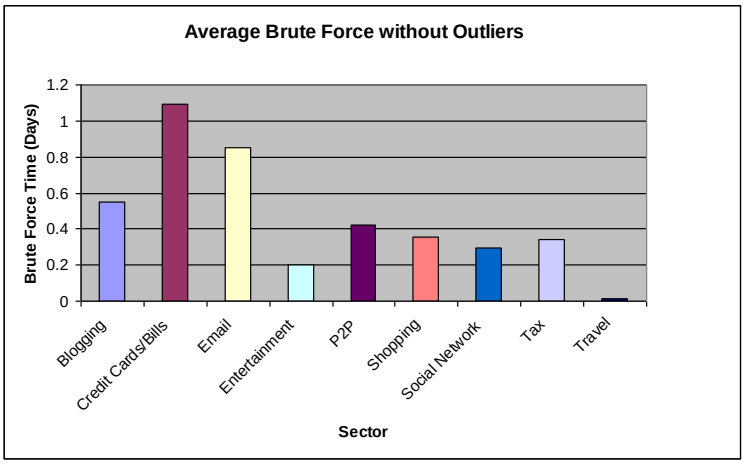

Figure 5. Average Time to Brute Force per Sector without Outliers

A large issue with comparing the different website sectors is that in our data we do not have the same number of data points for each sector. The Government sector only has one data point, which might not accurately reflect the sector as a whole. The Medical Sector, on the other hand, has ten data points to compare, which make the average of those ten a good indicator of what the industry standard is for the Medical websites.

Figure 6 compares the sensitivity of the information stored on the website to time it takes to Brute-Force the site. Figure 6 shows a strong disconnect between the sensitivity of the information stored on the website compared to the strength of the password for that site. We would expect a mathematical relationship between the time to break a password using a Brute-Force attack and the sensitivity of the information on the site. In other words, the more sensitive the information the stronger the password should be. Figure 6 shows that there is no connection between the strength of the password and the sensitivity of the information stored on the site.

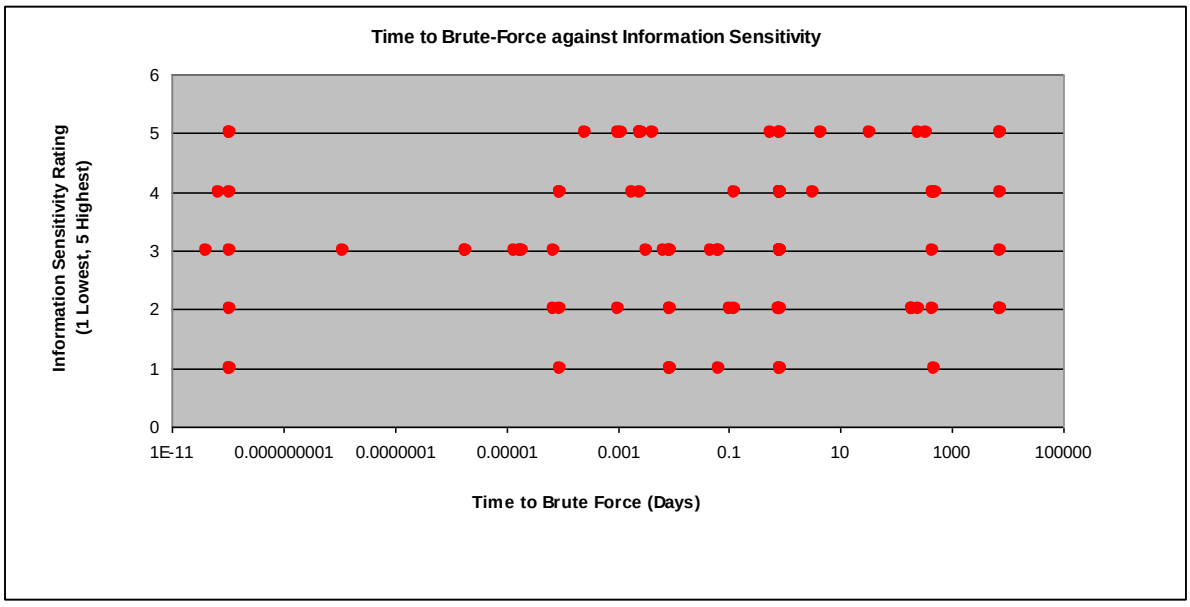

Figure 6. Scatter Plot of Information Sensitivity vs. Time to Brute-Force

All of our results show us that there is no relationship between the sensitivity of the information stored on a site and how strong the password is to protect that information. Therefore, it is imperative for users to not use the minimum password policy on the Internet, but instead use a stronger password. Our results assumed the user is using the minimum password length available. If the password required is only 6 characters and a user decides to use 8, the permutations would increase from 95^6 to 95^6+95^7+95^8, or a 9000% increase in permutations.

Conclusions

According to a CNN poll, 54% of approximately fifteen thousand individuals stated that they use 1 -5 passwords on the internet.[5] This is alarming. If a cracker is able to brute force one easy password, he has a good chance of guessing the password of the next site. Let us say the attacker cracks a victim's MySpace account, which can potentially be cracked in less than five minutes. This password could be something such as "love1!" Now using this password and the information gained the attacker might find out the victim uses Bank of America. The attacker determines the user id and can adapt the MySpace password towards Bank of America and find that the new password is "ILove1!!" If that does not work, information gained from the MySpace account, likes/dislikes/hobbies, can be used because 50% of passwords use family, pet or significant others according to a British study.[5] Therefore, an attacker can escalate the sensitivity of the information gained by first Brute-Forcing a weaker site and then use that password to break into a site with more sensitive information.

There is no relationship between information on the site and the minimum password policy. Therefore it is on the user to create their own password policy that above and beyond minimum requirements for sites. Also as a minimum, given the advancing technology, users should have passwords of minimum length of 8 if the passwords can be case sensitive and include special characters. This password would take a maximum of 7600 days to crack at 10 million password guesses a second. If the policy only allows numbers and non-case-sensitive letters, 10 characters should be used at a minimum. 10 characters would take a maximum 4200 days to crack at 10 million password guesses a second. The user risks information loss if they use a password that is weaker than these recommendations. Users should also be aware the assumption of 10 million passwords a second is a base level for attackers. Attackers can potentially array many dual-core computers to vastly decrease the time to Brute-Force a given password. Imagine 1000 computers whose sole purpose is to crack your 8 character password. 7600 days to crack is now decreased to 7.6 days and now feasible for some hackers.

By increasing the password policies for these websites there will be increase in requests for forgotten passwords. This can be mitigated using a secure password manager. However, we would avoid managers that are freeware or have internet connectivity. Password policies should also have a strong level of encryption. However, the overall security is dependent on the "Master" password for the manager; therefore, one should use a stronger password than normal.

To further academia research in Internet password studies we would suggest improving on the following areas: a larger and broader spectrum survey group for information sensitivity, equal number of sites per sector, using hybrid password cracking techniques as a corollary to Brute-Forcing, and the time to Brute-Force a password using current hacker trends i.e. botnets. These additional areas of improvement would improve the study and give a better understanding to users and administrators alike in the field of Internet password studies. (Research Foundation Texas A&M University n.d.)

Bibliography

Brown, A. (2002). "UK study: Passwords often easy to crack". *CNN.com*. March 13, 2002. Available at <http://www.cnn.com/2002/TECH/ptech/03/13/dangerous.passwords/>. Accessed on 14 February 2008.

Kumar, Manu, Tal Garfinkel, Dan Boneh and Terry Winograd. "Reducing Shoulder-surfing by Using Gaze-based Password Entry." *ACM International Conference Proceeding Series*, vol. 229 (2007). Available at < http://delivery.acm.org/10.1145/1290000/1280683/p13-kumar.pdf?key1=1280683&key2=0409614021&coll=GUIDE&dl=GUIDE&CFID=56939704&CFTOKEN=32557802>. Accessed on 19 February 2008.

Lucas, Ivan N. "Password Recovery Speeds." *Lockdown.co.uk*. 22 January 2007. Available at <http://www.lockdown.co.uk/?pg=combi&s=articles>. Accessed on 13 February 2008.

Rajput, Saeed A., Jihong Chen, and Sam Hsu. "State based authentication." *ACM Southeast Regional Conference: Proceedings of the 43rd annual Southeast regional Conference*. Vol 2. (2005). Pgs 160-165. Avaiable at <http://0-delivery.acm.org.usmalibrary.usma.edu/10.1145/1170000/1167292/p160-rajput.pdf?key1=1167292&key2=8498614021&coll=Portal&dl=GUIDE&CFID=18063467&CFTOKEN=33538498.> Accessed on 14 February 2008.

"Weak Passwords." *Research Foundation Texas A&M University*. Available at <http://rf-web.tamu.edu/security/SECGUIDE/V1comput/Password.htm>. Accessed on 28 February 2008.

References

[1] "Weak Passwords." *Research Foundation Texas A&M University*. Available at <http://rf-web.tamu.edu/security/SECGUIDE/V1comput/Password.htm>. Accessed on 28 February 2008.

[2] Kumar, Manu, Tal Garfinkel, Dan Boneh and Terry Winograd. "Reducing Shoulder-surfing by Using Gaze-based Password Entry." *ACM International Conference Proceeding Series*, vol. 229 (2007). Available at < http://delivery.acm.org/10.1145/1290000/1280683/p13-kumar.pdf?key1=1280683&key2=0409614021&coll=GUIDE&dl=GUIDE&CFID=56939704&CFTOKEN=32557802>. Accessed on 19 February 2008.

[3] Rajput, Saeed A., Jihong Chen, and Sam Hsu. "State based authentication." *ACM Southeast Regional Conference: Proceedings of the 43rd annual Southeast regional Conference*. Vol 2. (2005). Pgs 160-165. Avaiable at <http://0-delivery.acm.org.usmalibrary.usma.edu/10.1145/1170000/1167292/p160-rajput.pdf?key1=1167292&key2=8498614021&coll=Portal&dl=GUIDE&CFID=18063467&CFTOKEN=33538498.> Accessed on 14 February 2008.

[4] Lucas, Ivan N. "Password Recovery Speeds." *Lockdown.co.uk*. 22 January 2007. Available at <http://www.lockdown.co.uk/?pg=combi&s=articles>. Accessed on 13 February 2008.

[5] Brown, A. (2002). "UK study: Passwords often easy to crack". *CNN.com*. March 13, 2002. Available at <http://www.cnn.com/2002/TECH/ptech/03/13/dangerous.passwords/>. Accessed on 14 February 2008.

Biographical Information

Bruce Barnes is an Information Technology Major with Honors at the United States Military Academy. CDT Barnes interned for 5 weeks at the National Security Agency in Maryland. Upon graduation he will commission as a 2LT in the Medical Service Corps and his first duty station will be in South Korea.

Keith Major is also an Information Technology Major with Honors at the United States Military Academy. CDT Major completed a 3 week internship with the 1st Information Operations Unit in Virginia. On 31 May 2008, CDT Major will be commissioned as a 2LT in the Army Corps of Engineers and his first duty station will be Fort Shafter, HI.