ACCOUNTS THAT NEVER EXPIRE: AN EXPLORATION INTO PRIVILEGED ACCOUNTS ON WIKIPEDIA

Jonathan Kaufman  
*Christopher Newport University*, jonathan.kaufman.18@cnu.edu

Christopher Kreider  
*Christopher Newport University*, chris.kreider@cnu.edu

Follow this and additional works at: [https://aisel.aisnet.org/sais2022](https://aisel.aisnet.org/sais2022)

**Recommended Citation**

[https://aisel.aisnet.org/sais2022/39](https://aisel.aisnet.org/sais2022/39)

This material is brought to you by the Southern (SAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in SAIS 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
ACCOUNTS THAT NEVER EXPIRE: AN EXPLORATION INTO PRIVILEGED ACCOUNTS ON WIKIPEDIA

Jonathan Kaufman  
Christopher Newport University  
jonathan.kaufman.18@cnu.edu

Christopher Kreider  
Christopher Newport University  
Chris.kreider@cnu.edu

Abstract

Account compromises are an issue that every organization with user accounts has to deal with and mitigate. Depending on the level of privileged access a compromised account has, it is possible that it could cause severe damages to the individual and organization. One site where an account compromise could have potentially significant impacts would be Wikipedia, a free online encyclopedia with information about almost everything. Therefore, the accuracy and reputation of the site is crucial to its success. Administrators on the site have the ability to perform many different actions on the site itself as well as other users. If a threat actor were to get access to an administrator’s account, they could cause damage to Wikipedia as a whole as well as to the reputation of the user. When Wikipedia was first started in 2001, when a user account was promoted to a privileged administrator role, they were an “administrator for life”. However, in June 2011, the community changed the policy where an administrator could have their privileges revoked if they were inactive for at least 12 months. This study explores two questions. First, we analyze the log data to determine key indicators of a compromise. Using this information, we then seek to use a natural experiment to better understand the effects of the inactive privileged account revocation policy on account compromises. By answering these questions, we will gain a better understanding of how this type of policy improves the security of large open source web based communities.

Keywords

Wikipedia, open source communities, authentication, web security

INTRODUCTION

Cybersecurity incidents have increased in recent years with the number of successful attacks increasing threefold between 2010 and 2015, with 76% of organizations experiencing a web-based attack (Walters, 2015). One big concern, especially when using online services, is the risk of our account compromise. There are many ways for a hacker to compromise an individual’s account. Some of these methods include cracking passwords, using backdoors, and social engineering. Once a hacker gains access to an account or system, they can perform any action that the user that was hacked could perform. If a compromised user was a privileged user on the system, that could be detrimental to the organization and the user. This study will explore account compromises on the English language Wikipedia, specifically related to the mandatory revocation of administrator rights when an account has gone inactive. Wikipedia is a free online encyclopedia where almost anyone can create an account and post articles on any topic. As of February 16, 2021, there are over 6 million content articles with over 40,000 registered users (Wikipedia 2021). Out of the total number of users, approximately one thousand of them are administrators on the site. In a study conducted by Rector (2008), it compares the accuracy of articles between Wikipedia, the print and online edition of Encyclopedia Britannica, the Dictionary of American History, and American National Biography. They concluded that Wikipedia has an accuracy of 80%, where the other sources examined had between a 96% and 97% accuracy rate. Even though Wikipedia is not completely accurate, it still has a high accuracy rate and can be used to find information about any topic easily. As stated by Jakob Voss (2005), Wikipedia falls within one of the top 100 most popular websites around the world. Wikipedia receives over 1.7 billion unique visitors to its website each month (2021). In an article by Harrison and Benjakob (2021), it discusses the view of Wikipedia and about its reliability. In the article by Harrison and Benjakob, it mentions how with the number of policies that it has in place and the number of people in the community, it is a more reliable fact-checking source than social media and other news sources.

On Wikipedia, an editor becomes an administrator by being elected by the Wikipedia community (2022). However, some of these administrators become inactive or use their admin rights for malicious activity. There are many reasons why a hacker would benefit from accessing this type of account. First, administrators have many powers on Wikipedia that could be used to facilitate malicious activity. In Das, Lavoie, and Magdon-Ismail’s (2016) work, it discusses how there are thousands
of administrators on the site and that they assist with conflict resolution. Because of the amount of access and the type of activity that administrators participate in, a hacker could use these rights to spread misinformation and include bias on controversial topics. Wikipedia has put into place a set of procedures that users should follow if they are certain that their account has been compromised. The Wikipedia page on Suspected Compromised Accounts states how if an administrator believes that their account has been compromised, to contact a Steward using the emergency method by including “*steward*” at the beginning of the message to request your account to be locked or by contacting an administrator (2021). This Wikipedia page also provides the method to follow to regain account access based on the current status of the account. Having guidelines in place for what happens when an account is compromised is crucial, especially when the type of account that could get compromised allows the hacker to perform high-level tasks that a normal user could not perform. Due to the nature of the type of rights that admins have on Wikipedia, an admin account would be something that certain hackers would find a benefit to gaining access to. In this paper, we will explore whether unrevoked admin rights led to a larger number of account compromises as well as some of the policies in place to prevent unauthorized access.

This research hopes to analyze the demotions that have taken place over the years as well as the policies that have been enacted to determine whether the actions taken by Wikipedia to lessen the number of compromises were adequate and whether they resulted in the decrease in the number of account compromises. This study hopes to try and analyze what might be causing the compromises that occur on Wikipedia. This paper will also explore common indicators of compromise with one of them being an account being demoted as that is a common action when an account becomes compromised.

**Literature Review**

Moira Burke and Robert Kraut (2008) discuss a model to assist in finding new Wikipedia admins as well as determine which factors are a good indication of whether a user will be successful at their RfA. Burke and Kraut (2008) find that one of the most predictive measures used during the RfA process is the diversity of the users experience as well as if they contribute to the creation of policies. They find that the total contribution count for a user is not as much as a predictor since for every 1,000 article edits it only raised the chance of success by 1.8% (Burke and Kraut, 2008). This would mean that with the need of more administrators, that if they are quickly promoting users to admin it could lead to less effective reviews and a higher chance of users that would use their admin rights for malicious activity.

Another work that focused their research on determining what makes a successful RfA and determining possible vandals is Das, Lavoie, and Magdon-Ismail’s work in which they came up with the “Controversy Score” and the “Clustered Controversy Score”, which helps to determine if an editor would most likely use admin rights for malicious activity (Das, Lavoie, and Magdon-Ismail, 2016, p. 24:10). These scores could help to prevent trolling since this model strives to help detect possible vandals before they are granted admin rights, which would assist in maintaining the security and integrity of Wikipedia.

*The Work of Sustaining Order in Wikipedia: The Banning of a Vandal*, by R. Stuart Geiger and David Ribes discuss the use of bots in Wikipedia to assist with detecting and responding to potential vandals. Geiger and Ribes (2016) mentions how the use of bots have become more prevalent and that now they are used to make about 16.33% of all edits (p. 3). One of the most relevant parts of this work for the research we are conducting is the use of these bots in detecting vandals. The article mentions that in order for a user to be temporarily blocked by an administrator that there usually has to be four reversions of an edit due to suspected vandalism (Geiger and Ribes, 2016, p. 6). The process of detecting vandals involves multiple users, either human or bot, reverting the potential vandal’s edits and making warnings on the user’s talk page (Geiger and Ribes, 2016, p. 7). This work is significant since it provides insight into how bots and software can be used to help with detecting vandals as well as it gives insight to criteria that is used to determine when a suspected vandal should be temporarily banned.

Pnina Shachaf and Noriko Hara’s (2010) work *Beyond Vandalism: Wikipedia Trolls* discusses the type of behavior that trolls show as well as discusses if a troll meets the behavior of a hacker. In the work, one of the primary motivations that was shown by Wikipedia trolls was for entertainment and personal satisfaction (Shachaf and Hara, 2010, p. 9). Even though not much harm would be caused by this type of troll, some harm would occur since they might put inaccurate information on an article. Also if the troll is using a compromised account, more harm would be done since they might hurt that user’s reputation. From their research, Schachaf and Hara (2010) found that they generally fall into the hacker category of script kiddies; however, they said it is possible for some of the Wikipedia trolls to be a hacktivist (p. 10-11). The trolls that fall into the hacktivist category would cause more harm since they are pushing a certain agenda or bias onto Wikipedia or they could be using a compromised account to hurt an editor’s reputation. The research conducted by VanDam, Tang, and Tan (2017) on compromised Twitter accounts will be relevant to this research for a few reasons. One of the most relevant portions of their research is when they define what a compromised account is. They define this as, “...an account accessed by a third party without the knowledge of the original user” (VanDam, Tang, and Tan, 2017, p. 2). This definition will be relevant as we are talking about compromised accounts on Wikipedia. Another part of their work that could be relevant in this research
is when they discuss that thirteen percent of all adults who are online have been compromised (VanDam, Tang, and Tan, 2017).

Evaluation of Password Hashing Schemes in Open Source Web Platforms by Ntantogian, Malliaros, and Xenakis (2019) discusses the strength of password hashing systems used by many web platforms. One important area of this research is the use of guessing a user’s password in order to gain unauthorized access to an account. This is relevant to this paper as guessing a password is one of many ways to compromise an account. In Ntantogian, Malliaros, and Xenakis’s (2019) work, they discuss how using faster compute power would decrease the amount of time needed to crack a password of up to 97%. Another notable part of this research is how they mention that many of the content management systems (CMS) sites use the default password hashing algorithms making it easier to crack the password (Ntantogian, Malliaros, and Xenakis, 2019). This is relevant to this research since many sites use default algorithms. Wikipedia could also be using this which creates a security vulnerability because of how easy it can be to crack passwords using these algorithms.

Security on Wikipedia

One of the key aspects to be able to protect against account compromises is to have effective security policies in place. The English Wikipedia currently enforces a minimum password of 8, except for a privileged account which would require a minimum of 10. One other key aspect about Wikipedia’s password policy is that it is stated as a recommendation and is not actually enforced when passwords are set. This is a significant security vulnerability in Wikipedia because even if an admin knew that a user’s password did not meet their standards and banned them, it still poses a threat to Wikipedia. One aspect of the password policy that is enforced on creation for privileged accounts is the password length (n.d.). Even though Wikipedia requires privileged users to choose a different password if it is not ten characters or longer, the password could still be a relatively weak password, therefore creating a security risk. The next policy that is relevant to this research is the policies on what happens after an account becomes compromised. As it discusses on the Wikipedia: Compromised Accounts page, after an account is compromised the user can contact “Stewards” which has the authority to lock accounts to prevent abuse as well as prevent the account from being modified (n.d.). If the account compromise is noticed early, this policy would be able to stop any significant damage from occurring. Another piece of information on this page would be how they will not unblock an account if the identity cannot be verified, which could cause the user to have to create another account (n.d.).

User Authentication

As defined by CISCO (2021), user authentication is, “... [A] process in which you verify that someone who is attempting to access services and applications is who they claim to be”. In terms of Wikipedia, the user authentication is the username and password used to login to the site. There are four different broad categories of user authentication which are: things you know, things you possess, things you are, and things you do (Stallings and Brown, 2018). Even though using one of these forms of authentication provides security to the system, to make the system even more secure, best practices would be to have at least two-factor authentication (Stallings and Brown, 2018). An example of two-factor authentication would be if you enter a password on a website then enter a code sent to either your email or phone number.

Security and Passwords

There are many different topics about passwords that deal with the security of a system. One relevant topic with security is about the reuse of passwords. As stated by Das, Bonneau, Caesar, Borisov, and Wang (2014), “... [A] typical Internet user [is] estimated to have 25 distinct online accounts” (p 1). With many websites requiring passwords with certain requirements, users tend to use the same password because of the difficulty of trying to remember the stronger passwords that are required by the websites (Das, Bonneau, Caesar, Borisov, and Wang, 2014). Another security issue that has to do with passwords is the capability of the password being figured out using a Brute Force Attack. In the book by Stallings and Brown, they define a brute force attack as, “... the brute force attack, is to try every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained” (Stallings and Brown, 2018, p. 55). In terms of passwords, this would mean trying every possible password combination. As stated in the book by Stallings and Brown (2018), a hacker would have to try on average half of all possible passwords for a successful brute force attack on an account. Even though this type of attack exists, due to the password policy for administrators, this is an unlikely attack. Another common type of attack to crack a password is the use of a dictionary attack. This type of attack involves the attacker to try a list of common words or phrases that are in the dictionary (n.d.). If a password contains a word where certain letters are replaced by a symbol, such as ‘a’ with ‘@’ or ‘s’ with ‘S’, this does not make a password more secure as hackers expect this type of action to occur and try those passwords as well. One of the best defenses against this type of attack would be to use memory techniques as well as a series of words as your password. By doing this, this strengthens the complexity of the password, no longer making it susceptible to a dictionary attack.

Methods

For this research, we make use of the natural experiment approach instead of conducting a true experiment. If we conducted a true experiment, there would have been a variable that would be controlled as well as a control group and an
experimental group. However, due to the nature of this research, that was not possible. For a natural experiment, it is where the variable is the experimental manipulation that occurs as a part of a significant event. In our research, this significant event would be the demotion policy that took effect in June 2011. To report how many months of data were available before and after this demotion policy occurred, the data was collected from Wikipedia’s User Rights log, which from there the admin logs was the data that was collected. To collect the data, python scripts were written to take the url of the first page of results and go through it looking for logs with the string “administrator” or “sysop” after going through the first page there was a function that went through the page and got the url for the next page and then collected the data from this page. The script kept looping through and getting the url of the page and getting the data until it went through all pages. After pulling the data from the Wikipedia User Rights Log, the script took those logs and formatting it to put in a text file the following information: date and time of the log entry, who performed the action, who the action was done on, if they were admin before the action, if they were admin after the action, and the comments of the log entry. The scripts then went through the filtered logs and determined the cause for the log entries as either inactive, self requested, deceased, or need to explore. It also converted the data into a file that could be imported into Microsoft excel. At the beginning, there were a total of 99,445 log entries. After all the scripts were executed, it brought the total number of entries to 2,940. After collecting the data, the file was imported into excel, where we investigated the entries that were labeled as need to explore to determine the cause of the demotions. The first log entry that was included in the data set was from December 24, 2004 at 17:42 UTC. The final log entry included in our data set was from September 24, 2020 at 19:50 UTC. The second round of data collection was from the Wikipedia:Bureaucrats’ noticeboard: Revision history log. The python script written to collect the data from this log pulled the “prev” link tag from any log entry that contained “Compromised” or “compromised” in the entry. The earliest entry in the Wikipedia:Bureaucrats’ noticeboard: Revision history log was August 24, 2005 at 20:53 UTC. The most recent entry in the log was on November 7, 2020 at 02:19 UTC. There were a total of 29,925 log entries that were reviewed for compromised entries. From the 29,925 log entries, 135 of them contained either “Compromised” or “compromised” in the entry and from these 135 log entries, 131 of them were reports of compromises. 63 of the 131 were unique usernames, where the rest of them was repeats of the same username.

**Results**

After collecting all the data from the user rights log that was collected by the python script, we went through the data to determine the cause for each demotion that occurred. Chart 1 shows the number of account demotions. From 2007 through 2009, there was a total of 1 demotion each year. From 2007 through 2020, the average number of demotions was approximately 49 demotions, the range was 115, and the standard deviation was 36.1. This data is consistent with the policies that were in place during these time periods since before June 2011, there was no policy to demote administrators for inactivity. However, when you look at the average number of demotions between the years 2011 and 2020, it was an average of approximately 68 demotions, with the range being 82 and the standard deviation is 22.8 between 2011 and 2020. This data is consistent with the policies that were in place during these periods of time. Until June of 2011 the policy where an admin can be demoted for inactivity after 12 months did not exist and administrators, once promoted, were an “administrator for life”. Before the inactive administrator policy took place, there were a total of 3 demotions. All 33 demotions in 2011 took place after the new policy was put in place. Now that we have looked at the total number of demotions by year and analyzed these numbers based on the policies in effect at the time, we will now look at the number of demotions by the cause of demotion. Looking at the chart to the right, it shows that the cause with the greatest number of demotions is for inactivity with a total number of 365 demotions. Another cause of demotion that had a large number fall under that category would be self-requested. Looking at the number that were demoted for compromise, which was 8, is not a lot, but this number is only the number that were demoted and not the total number of compromises. In order to better understand the effects of the inactive account policy, let us look at the number of inactive accounts that got demoted each year. From 2011 to 2020, the average number of demotions due to inactivity was 33.9 accounts with a standard deviation of 21.6. The highest number of demotions under this policy occurred in 2012, which was the first year where accounts could be demoted under the policy. Looking at the data, it shows that 2 years after the policy took place or 1 year after accounts could be demoted under the policy, there was a significant drop from 73 demotions to 26 demotions. In addition, for the following 2 years after that it continued to drop significantly from 26 to 1 and then 0. After these significant drops, it then spiked back up from no demotions in 2015 to 45 demotions in 2016. From 2016 through 2019, there was an average of 43.5 account demotions with a standard deviation of 1.1. This demonstrates that the number of accounts demoted under the policy during these years were around the same. From January 2020 through around mid-november 2020, when we stopped collecting data, there were a total of 31 demotions under the inactivity policy.

As shown above, starting in 2011 the number of administrators demoted increased significantly due to the new policy relating to inactive administrators taking effect on the site. By having this many inactive users on Wikipedia provides for potentially easier targets for hackers who would want to get access to this type of account. This study focuses on whether this policy has affected the overall security posture on Wikipedia. Looking at Chart 4 to the left, we notice that the number of compromises that occurred prior to June 2011 was only 2. It is also shown that after June 2011, there were a total of 11
account compromises, which makes a total of 13 compromises in our data set. This chart demonstrates that there was a significant increase in the number of compromises after the policy where administrators can be demoted for inactivity. Now that we have analyzed the number of account compromises before the policy change and after the policy change, it is also important to look at the number of account compromises to better visualize the distribution of the account compromises that occurred. Based on the data, there was no reported compromise between the years 2001 and 2006. However, in 2007 there were a total of 2 compromises, which is a small jump from 2006. In the dataset from 2007, when the first compromise was, through November 2020 the average number of account compromises that occurred was approximately 1, with the range being 4, and the standard deviation is 1.2. However, the average number of account compromises for the years where there was at least 1 compromise would be approximately 2. When looking at the data, it shows that there were only compromises reported in the years 2007, 2011, 2012, 2015, 2018, 2019, with the most compromises occurring in 2018. Out of the 14 years of data shown in the chart above, 8 of those years did not have a single compromise reported.

**Discussion**

The data collected during this research has shown some very important points. First, when looking at the data from the number of demotions and comparing them to the demotion policies, the number that was demoted each year due to suspected compromise is inconsistent with these policies. Before 2011, we identified a total of 3 demotions due to suspected account compromise. In 2011, when the inactive account privilege revocation policy was put in place, there was a large increase in the number of users who had their privileges revoked. This increase in privilege revocation is likely related to accounts that were sitting inactive. In addition, the number of demotions per year remained higher every year starting in 2011. This demonstrates that there are many users that are being demoted due to inactivity. When looking at the number of demotions due to suspected account compromise per year compared to the number of demotions due to inactivity per year, we found that demotions due to inactivity were a majority of the total demotions per year, especially in 2012, accounting for 63.5%. However, between the years of 2013 through 2015, the average percentage of inactive account demotions was 13.1%, the highest in these years being 37.7% in 2013. From 2016 through November 2020, the average percentage was 59.3%. As shown through the data, the first year of demotions from inactivity had a high percentage, then the number went down for a few years and then spiked again in 2016. From 2012 to 2020, which is the whole period of time from our dataset where demotions from inactivity occurred, the percentage of inactive account demotions out of total demotions is 47%.

The most important question that we were looking at with this research was if the policy change that occurred in June 2011 had any effect on the number of compromises that occurred on the site. Through the data we collected, it shows the opposite. It demonstrated that there was an increase in the number of compromises after the policy changed. This is an interesting finding since it would be expected that the number would decrease by demoting any users that sit inactive for at least a year. One idea to consider is how the Wikipedia community handles account compromises, especially when it relates to a privileged account. As of now, most of the time when an account is compromised it is demoted from administrator to regular editor. However, it might be more appropriate to completely block the user as opposed to revoking their privileges, as non-privileged users are still capable of performing harmful activities. The most likely reason that this policy was implemented on Wikipedia was to help lessen the number of compromised privileged accounts on Wikipedia.

When it comes to indicators of account compromise that resulted in accounts being demoted or blocked, there are a few key ones. The first of these is when the user directly reports the compromise. Everytime a compromise was reported to administrators, the account was always demoted to a non-privileged account. This is an effective tactic to use since even if the account is not compromised, it will allow the administrators to determine if it was compromised and confirm it is back in control of the actual user before re-granting the administrative privileges to the account. The second indicator is the identification of suspicious edits by the account. Suspicious edits generally include contributions to the project that do not fit with the reputation of the user. It is important that accounts that have suspicious edits get demoted or blocked as soon as possible in order to prevent further damage to Wikipedia articles.

With the increase in compromise, it leaves a new question of what might have caused the increase in compromises? With the growth of technology over the years, that might have contributed to the increase in the amount of compromises.

**Conclusions**

This study focused on determining whether the demotion due to inactivity policy that took place in June 2011 had any effect on the number of account compromises that occurred on the site. When analyzing the data, we found unexpected results, specifically that the policy did not have the expected effect, as the number of compromises before the policy was put in place and the number after were different. Based on the data that we collected about compromises on the site, it appears that the new policy on demoting administrators who sit inactive had an opposite effect when compared to what was expected. This was shown since there were only 2 identified compromises before the policy change and 11 after. When analyzing how an account compromise is determined on Wikipedia, there are a few key aspects that we determined. One of the key aspects is if there are suspicious edits made by a user that does not fit with their reputation. A majority of the cases that we looked at on Wikipedia had to do with this type of activity. The other time where accounts were demoted for suspected compromise...
was when the user reported it. The one case that stands out is when a user reported that his account could be compromised since his Gmail was compromised. However, there were no indications of a compromise on Wikipedia.

In future works, we plan to further explore the data collected as a part of this research to better understand the unexpected findings. Additional analysis will allow us to better understand compromises that took place on Wikipedia during the period of time that we collected data for. This information will provide better understanding of indicators to look for in order to detect compromises.

Approved for Public Release; Distribution Unlimited. Public Release Case Number 21-4036

REFERENCES