**FITZHUGH**

**CSIA 450 – 8372 CYBER SECURITY CAPSTONE**

**SPRING 2020**

**Project 1**

**Objectives**

• Understand system vulnerability exploitation and potential impact.

• Apply cybersecurity research to solve practical problems.

• Research and apply malware fundamentals and analysis.

**Problems**

**A. CALCULATE KEYS**

Use calculate\_keys.py as an example if needed to answer the following. You don’t need to use

Python. You can do the calculations in Excel or another tool.

**1. If an organization has 570 employees, how many symmetric keys would be needed to**

**securely communicate? How many asymmetric keys would be needed?**

The equation to calculate the number of symmetric keys need is:

Total\_symmetric\_keys = (N(N-1))/2, where n is the number of employees

If an organization has 570 employees:

Total\_symmetric\_keys = (570(570-1))/2=(570(569))/2=324,330/2=162,165

The equation to calculate the number of asymmetric keys needed is:

Total\_asymmetric\_keys = N \* 2, where n is the number of employees

If an organization has 570 employees:

Total\_asymmetric\_keys = 570\*2 = 1,140

If an organization has 570 employees, it will need 162,165 symmetric keys and 1,140 asymmetric keys.

**2. If an organization has 800 employees, how many symmetric keys would be needed to**

**securely communicate? How many asymmetric keys would be needed?**

The equation to calculate the number of symmetric keys need is:

Total\_symmetric\_keys = (N(N-1))/2, where n is the number of employees

If an organization has 800 employees:

Total\_symmetric\_keys = (800(800-1))/2=(800(799))/2=639,200/2=319,600

The equation to calculate the number of asymmetric keys needed is:

Total\_asymmetric\_keys = N \* 2, where n is the number of employees

If an organization has 800 employees:

Total\_asymmetric\_keys = 800\*2 = 1,600

If an organization has 570 employees, it will need 319,600 symmetric keys and 1,600 asymmetric keys.

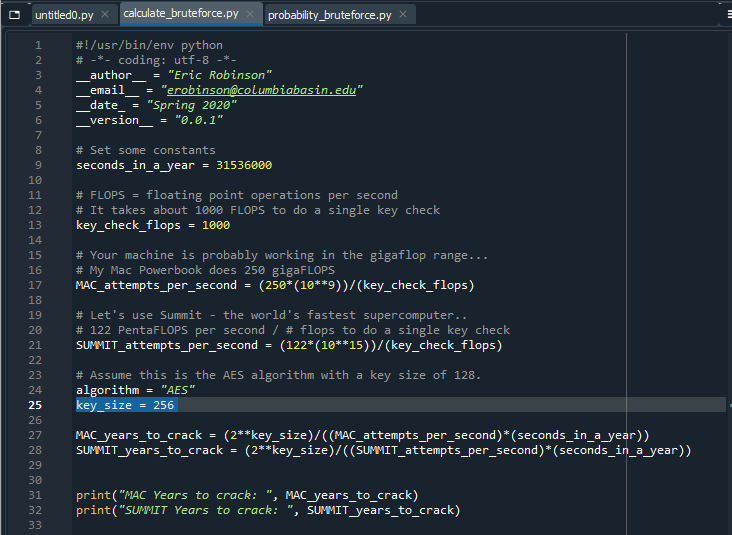
Use calculate\_bruteforce.py and probability\_bruteforce.py as examples if needed to answer the

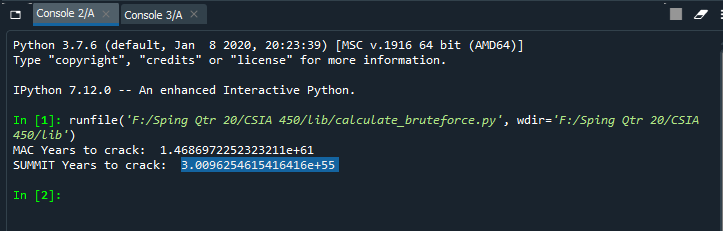
following. You don’t need to use Python. You can do the calculations in Excel or another tool.

**3. Research the length of time it would take to brute-force an entire AES keyspace for a 256-**

**bit key size**.

If we run calculate\_bruteforce.py, that was edited for a key-size of 256, we find that it takes a MAC 1.4686972252323211e+61 years to crack and takes Summit 3.0096254615416416e+55 years to crack.

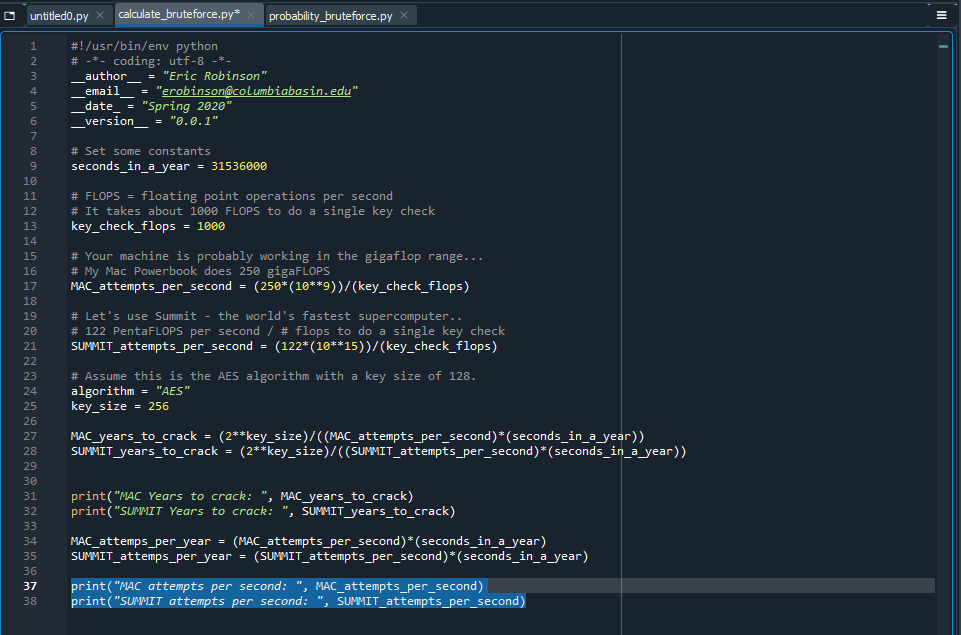


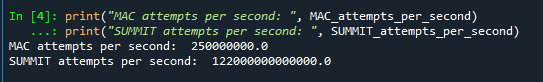


**4. If an attacker was really, really lucky, how long, in seconds, would it take to brute-force a**

**256-bit key?**

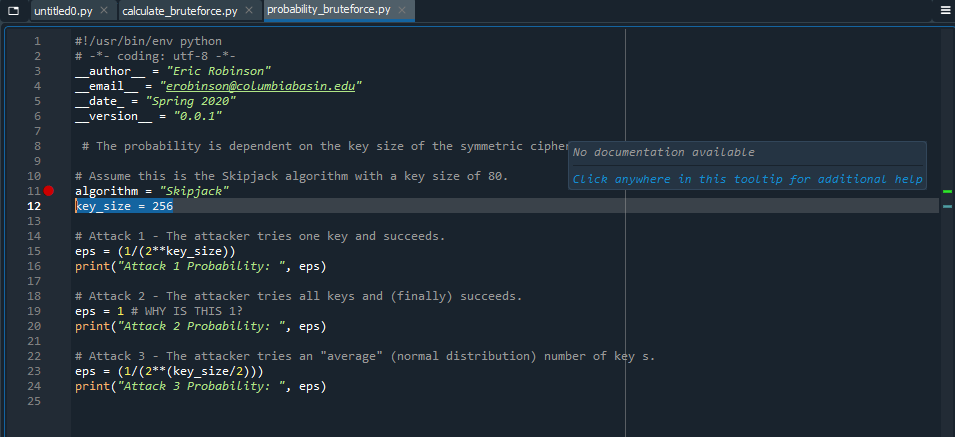
If we run calculate\_bruteforce.py, that was edited for a key-size of 256 and to print attempts per second, we can find that a MAC tries 250000000.0 keys a second, while Summit tries 122000000000000.0 keys a second. If an attacker was really, really lucky, it would take less than 1 second for them to brute-force a 256-bit key. This is because if they were really, really lucky they could find a key as fast as the first key, within the first million keys, or even the first 100 million keys, all of which either computer try within less than a second.

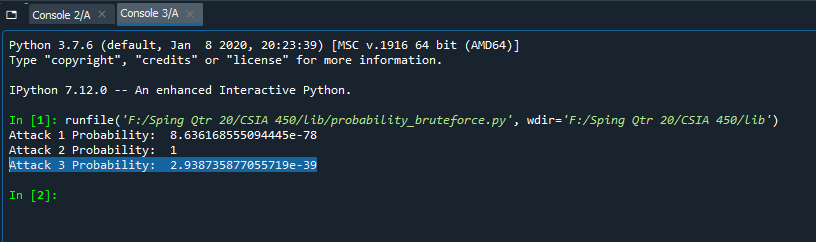




**5. What is the average probability for an attacker to break an AES-256 cipher?**

If we run probability\_bruteforce.py, edited for a key-size of 256, we find the average probability for an attack to break an AES-256 cipher as 2.938735877055719e-39.





**B. DLP ANALYSIS**

Rebound Security (http://reboundsecurity.info) is a local company that makes a data-loss

prevention appliance (Rebound\_DLP) that your company has used for years to protect its data

from unauthorized exfiltration.

The appliance is reporting a large spike in UNKNOWN outbound data every day at 2 PM.

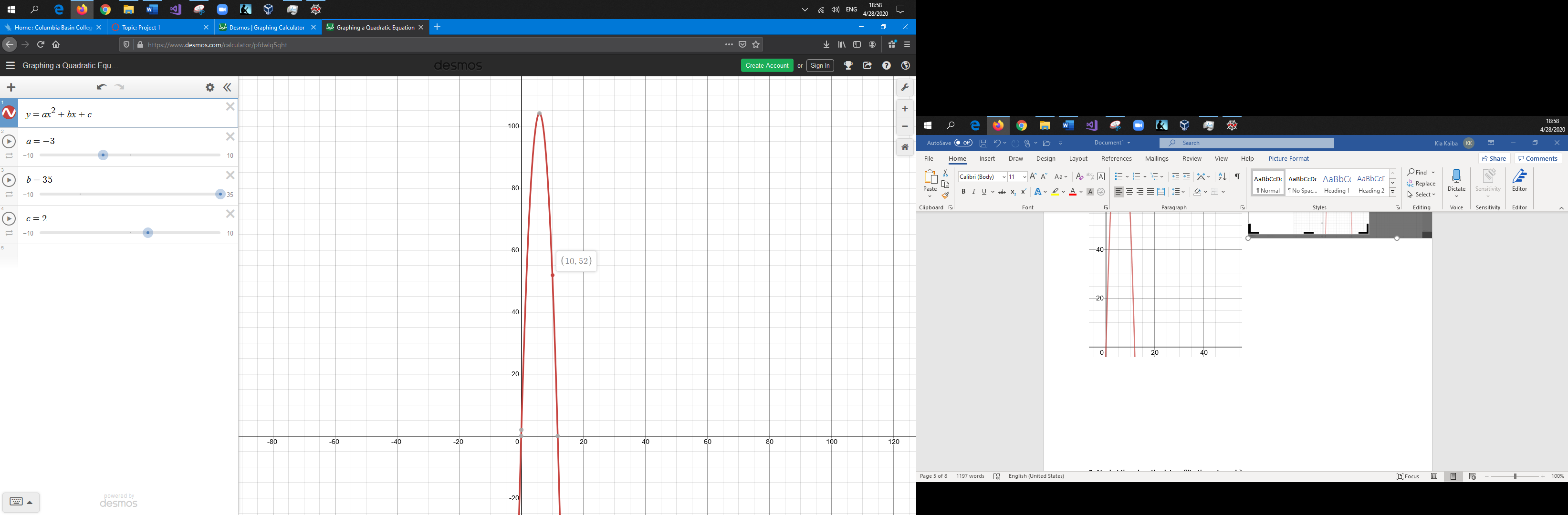
It can be modeled using the following function where x is minutes, and k(x) is outbound data

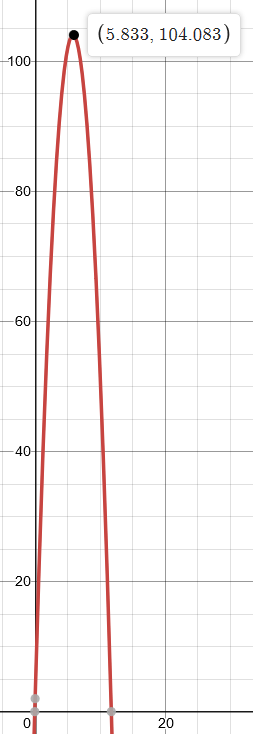
record counts:

k(x) = -3x2 + 35x + 2

**6. Plot the outbound data function for a period of 10 minutes.**

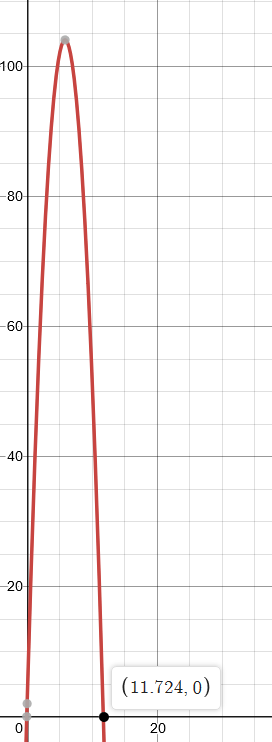
If we plot the outbound data function, k(x) = -3x2 + 35x + 2, we get the graph to the left. At 10 minutes, we find 52 unknown data, seen on the graph below.



**7. At what time does the data exfiltration rate peak?**

The exfiltration rate peaks at 5.833 minutes, as seen in the graph to the left.

**8. At what time does the data exfiltration end?**

The data exfiltration ends at 11.724 minutes, as seen in the graph below. 

**C. SECURITY RESEARCH**

Security vulnerabilities are system bugs, irregularities or other issues that provide mechanisms

for exploitation and unintended system use. The larger and more complex a system is, the

greater the likelihood that vulnerability exists. Tesla, for example, is a well-recognized player in

the automobile industry. Yet in 2016 the company was subject to security hacks as a result of

system/device vulnerabilities. Use the Internet as needed to answer the following.

**9. What types of security breaches did Tesla face in 2016?**

According to Golson, a Chinese Security Lab was able to hack into a Tesla Model S web browser via a malicious hot spot. They were able to take advantage of this vulnerability to unlock the doors, open the trunk, fold the side mirrors and activate the brakes while the car was in motion (Golson 2016).

According to Brewster, Zhejiang University and the University of South Carolina, were able to use spoofing and jamming techniques to hack into a Tesla Model S’ autopilot. Using these techniques, they were able to fool the autopilot into believing objects were not nearby, which would cause the car to collide with obstacles (Brewster 2016).

**10. How was the organization's reputation and brand name affected?**

Tesla’s reputation and brand name was not affected negatively by these breaches. Tesla has a bug bounty program, where they pay security researchers for any new bugs found (Tesla). According to Tesla’s security page, they work to verify and reproduce vulnerabilities so they can respond to them accordingly. I believe having the bug bounty program helps to improve Tesla’s organization.

**11. What did Tesla do to restore customer confidence and address these security issues?**

According to Greenberg, Tesla not only implemented a bug fix, but provided their vehicles with a security update to help block further attempts at hacking. This update included a feature that “requires any new firmware written to components on the CAN Bus. . . be digitally signed with a cryptographic key only Tesla possesses” (Greenberg 2017).

Using the Internet, find a recent example of an organization that encountered a security breach

due to malware. Answer the following.

**12. Provide an overview of the incident.**

According to Winder, North Carolina city and country government systems were attacked by ransomware on March 6, 2020. The malware was detected early, and the networks were shut down to avoid further spread, including the city’s phone network. However, the 911 network remained operational. Winder further stated that 80 servers were contaminated, as well as 1,000 computers.

**13. Describe what you believe led to or allowed the breach to occur.**

Winder reports that the breach was able to occur because “both city and county employees were clicking links in emails.”

**14. Explain what could have been done to prevent this breach.**

User awareness is one of the top measures that could have prevented this breach. User awareness training is important because it enables all employees to understand and help prevent security measures, without becoming as versed in security as the security teams. If employees were taught about the dangers of clicking links from emails, and how to identify malicious links, the breach would’ve been avoided.

**D. SHAM PHARMACY ANALYSIS**

Attackers sell counterfeit pharmaceuticals online using sham pharmacies. One way these

pharmacies are advertised is by compromising high-ranking websites and redirecting visitors

from search engines to these fake pharmacies via the compromised websites. These search redirection attacks are described in a research paper

(https://www.usenix.org/legacy/event/sec11/tech/full\_papers/Leontiadis.pdf) if you are

interested.

You will need the pharmterms.csv file from Canvas for your analysis. This file contains a table

recording the top 32 search results from Google and Bing for 218 drug-related search queries

collected over two weeks in November 2011.

Each row in the table corresponds to one search result, including the date, search engine,

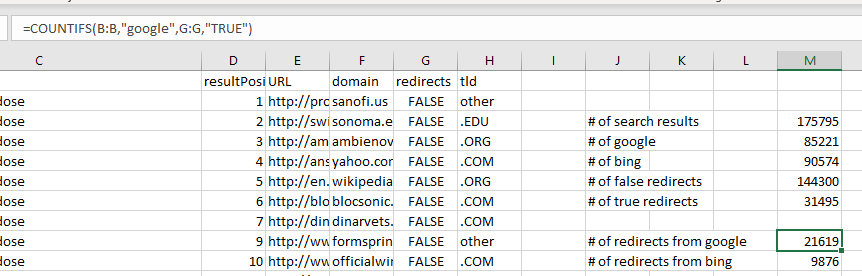
search term, position in the search results (1–32), the URL, second-level domain, whether the

result redirected to an illicit pharmacy (True/False), and the top-level domain for the result

(.COM, .EDU, .ORG, .GOV, .NET, or other).

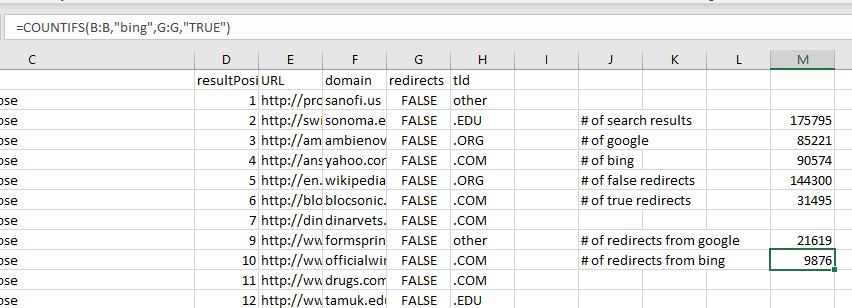
**15. What percentage of search results redirect to illicit pharmacies on Google?**

Using excel, we can use the COUNTIFS formula to see how many search redirects there was on Google. The formula is: =COUNTIFS(B:B,"google",G:G,"TRUE"). This produces 21,619 redirects. There are 175,795 search results. So, the percent of search results redirected to illicit pharmacies on Google is 21,619/175,795, which is equal to 12.3%.



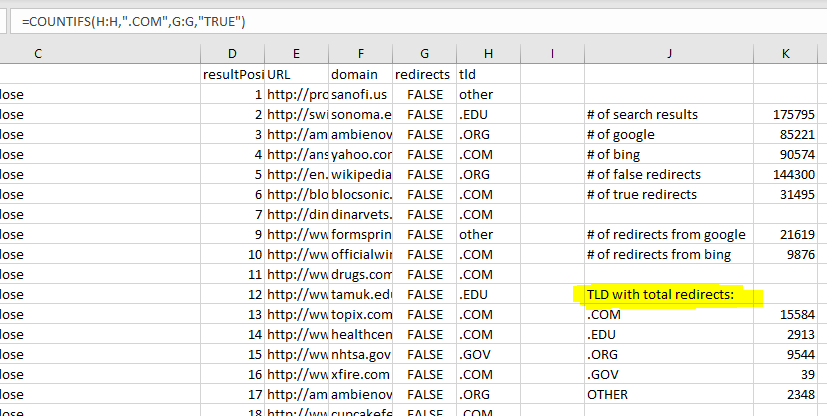
**16. What percentage of search results redirect to illicit pharmacies on Bing?**

Using excel, we can use the COUNTIFS formula to see how many search redirects there was on Bing. The formula is: =COUNTIFS(B:B,"bing",G:G,"TRUE"). This produces 9,876 redirects. There are 175,795 search results. So, the percent of search results redirected to illicit pharmacies on Bing is 9,876/175,795, which is 5.6%.



**17. Which top-level domain (TLD) has the most TOTAL redirects?**

Using excel, we can use the COUNTIFS formula to determine how many redirects each TLD had. The formula is as follows: =COUNTIFS(H:H,".TLD",G:G,"TRUE"), with TLD being substituted for each of the 5 TLDs. The full results are shown in the screenshot below. The TLD with the most total redirects is .COM, with 15,584 redirects.



**18. Which top-level domain (TLD) has the least TOTAL redirects?**

Using the same search results and findings as found in question 17, the TLD with the least total redirects is .GOV, with 39 redirects.

**19. What is the main factor driving the fact that one TLD has more TOTAL redirects?**

I believe the main factor driving the fact that one TLD has more total redirects is because there is one TLD that is considered the “main” TLD. According to Wikipedia, .COM is the main TLD used, it’ss an open TLD and is a commercial entity, that was originally used for any for-profit organizations. While .ORG and .NET are also open TLDs, they were originally used for non-profit organizations and networks of computers (Wikipedia). Because of this, I believe .COM became the most popular and most used TLD.

**E. ADVANCED MALWARE ANALYSIS**

Rebound Security (http://reboundsecurity.info) is a local, independent consulting and training

company focused on providing premium service to organizations. Rebound Security’s CEO, Eric

Robinson, wants to find out if it's possible to PREDICT malware infections using his employees'

endpoints as a baseline.

You will need the following files from Canvas for your analysis:

• mw\_data.txt – This file is a description of the data elements found in known.csv and

unknown.csv.

• mw\_known.csv - This file contains a list of endpoints (100), their endpoint attributes,

and a 0 (no) or 1 (yes) for malware infection state. The data structure of the file is as:

<endpoint>, <department>, <memory>, <storage>, <network>, <application>, <risk>,

<OS>, <internet>, <malware>

• mw\_unknown.csv – This file contains a list of endpoints (10) and their endpoint

attributes. The malware infection state is blank since it is unknown. The data structure

of the file is as:

<endpoint>, <department>, <memory>, <storage>, <network>, <application>, <risk>,

<OS>, <internet>, <malware>

**20. Since most endpoints have signature-based anti-malware installed, is this even a useful**

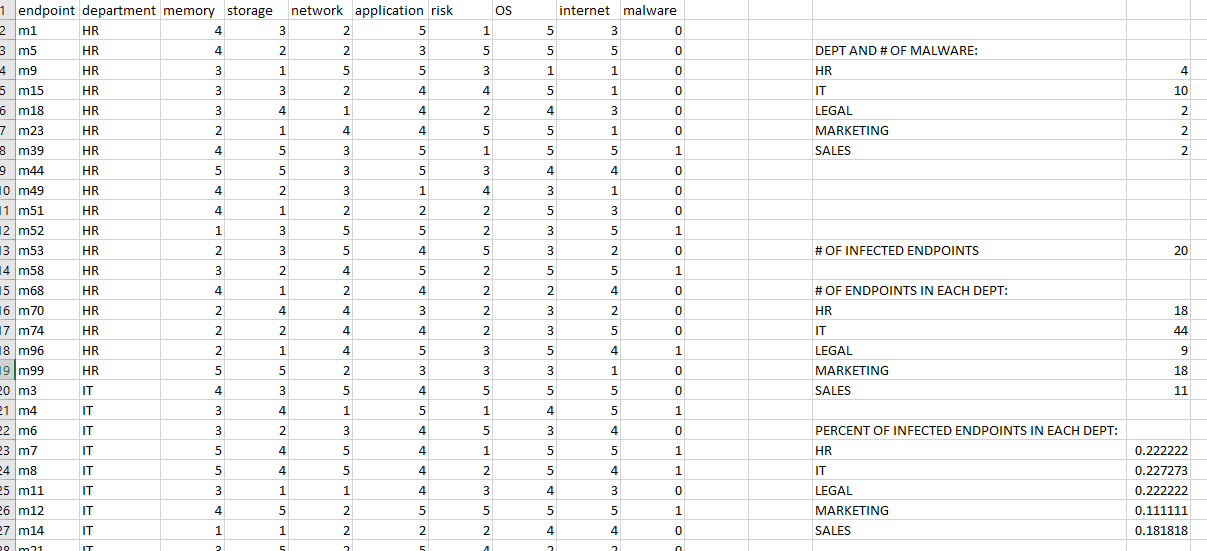
**problem to solve? Why or why not?**

I believe this is a useful problem to solve. This is because there are tons of new malware released every day. While anti-malware software is great at detecting malware, there may be some time before an update is released to account for new malware. Even if those updates become available in a super timely matter, there is also the risk that the user may not update their anti-malware software in a timely matter. It is important to know which endpoints have the highest risk of being infected, that way those endpoints can be monitored to help prevent malware infection.

**21. What department PROPORTIONALLY is the worst offender with malware infections? Why**

**do you think that is?**

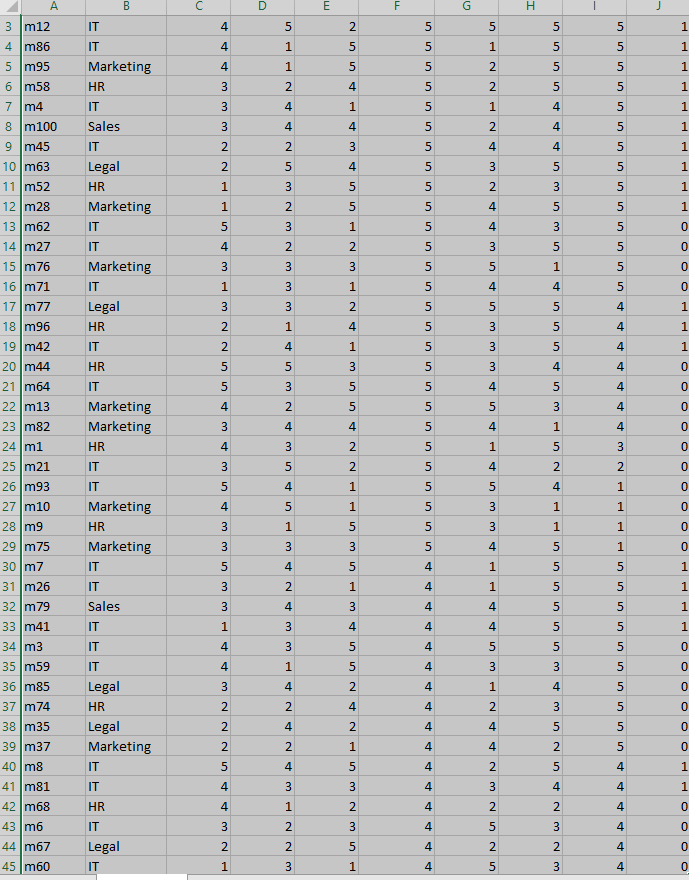
In excel, using the mw\_known.csv, we can use the COUNTIFS formula to find how many end points in each department have malware infections. The formula is =COUNTIFS(B:B,"DEPT",J:J,"1"), with DEPT being changed to one of the corresponding 5 departments. We can divide the results from this formula to how many endpoints are in each department (which is found by using =COUNTIF(B:B,"DEPT"). The data is seen in the screenshot below. The department that is proportionally the worst offender with malware infections is almost a three-way-tie between HR, IT and LEGAL, all with a 22% rate. IT is slightly worse than the other two departments, where rounded up the percent becomes 23% (from 0.227). I think this may be because endpoints in those departments house data that is targeted by attackers.



**22. Develop an “algorithm” for detecting malware based on the data in mw\_known.csv.**

If you sort the data in Excel by Infected, it becomes clear that every endpoint infected with malware has a risk of 4 or higher in application and internet. If you sort that data, you can find there are 15 endpoints that have a 5 in both application and internet, and 11 out of those 15 endpoints are infected with malware. If you sort the data by all endpoints that have at least a 4 in both application and internet, you find 44 endpoints, of which 20 endpoints are currently infected. My algorithm would be to sort the data by highest in the application and internet columns. All endpoints that have a 4 or higher in both categories I would predict to be infected with malware.

The screenshot below shows there are 44 endpoints that have a 4 or higher in both application and internet. Running the COUNTIF formula, =COUNTIF(J2:J45,1), against the malware column down to row 45) shows us there are 20 infected endpoints.

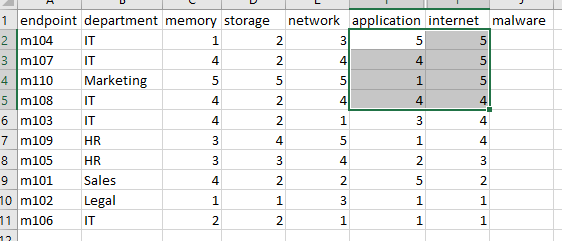


**23. Are there technical reasons for these endpoint attributes being stronger indicators?**

The technical reasons for those attributes being strong indicators for detecting malware is a higher risk score for application shows there are applications installed on the endpoint that are not fully updated and patched, and a higher risk score for internet shows the user of that endpoint engages in risky behavior online, such as accessing social media, personal email, watching porn. These reasons were based on how the data is given a risk score in mw\_data.txt. Unpatched applications provide vulnerabilities attackers can take advantage of. Risk behavior online can lead to downloads that contain malware.

**24. Use your algorithm to make malware predictions for mw\_unknown.csv.**

If we sort the data by application and internet, we find there are 3 endpoints that have a 4 or higher in both application and internet. It would predict that m104, m107, and m108 are infected with malware.



**25. What is your PERCENT confidence in your results?**

I’d say I am about 75% confident in my results. This is because the data shows an almost 75% infected rate of all endpoints that have a risk score of 5 in both application and internet. While having a risk score of 5 in both categories does not guarantee the endpoint to be infected at the time the data is gathered, there is still a very high probability it will be infected at a later date if the risks are not lowered.

**Works Cited**

Brewster, T. (2016, August 4). Hackers Fool Tesla Autopilot Into Making Obstacles 'Disappear' -- But Don't Panic About Crashes Yet. Retrieved from https://www.forbes.com/sites/thomasbrewster/2016/08/04/tesla-autopilot-hack-crash/#2fed3352b873

Golson, J. (2016, September 20). Car hackers demonstrate wireless attack on Tesla Model S. Retrieved from https://www.theverge.com/2016/9/19/12985120/tesla-model-s-hack-vulnerability-keen-labs

Greenberg, A. (2017, June 3). Tesla Responds to Chinese Hack With a Major Security Upgrade. Retrieved from https://www.wired.com/2016/09/tesla-responds-chinese-hack-major-security-upgrade/

Tesla. (n.d.). Product Security: Tesla. Retrieved from https://www.tesla.com/about/security

Wikipedia. (2020, April 26). List of Internet top-level domains. Retrieved from https://en.wikipedia.org/wiki/List\_of\_Internet\_top-level\_domains

Winder, D. (2020, March 10). Two 'Russian' Ransomware Attacks Take Down North Carolina City And County Government Systems. Retrieved from https://www.forbes.com/sites/daveywinder/2020/03/10/two-russian-ransomware-attacks-take-down-north-carolina-city-and-county-government-systems/#29123fbd588f