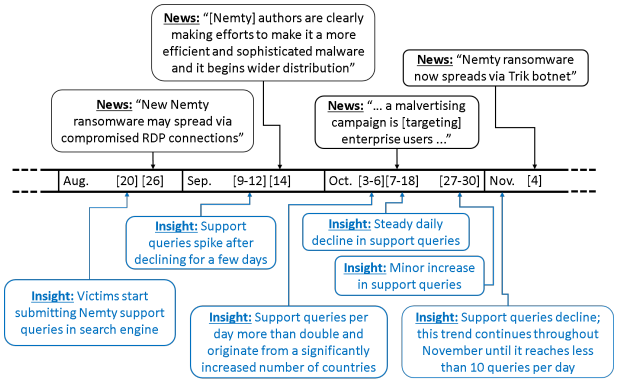

Figure 5: Timeline of news and query logs insights about Nemty.

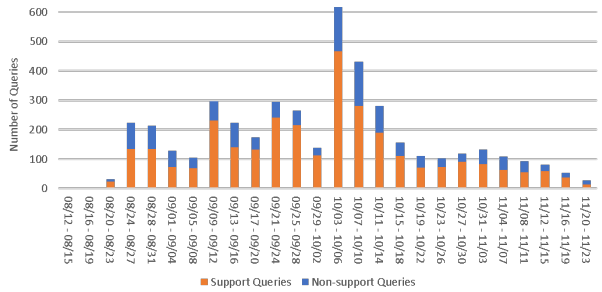

Figure 6: Volume of queries about Nemty per 4-day periods.

through distribution campaigns during September, October and November 2019, as seen in the timeline of Figure 5 (black boxes above the timeline show published news about Nemty [8]).

We gathered attack-related search engine query logs (with English locale) about Nemty between the start of August to end of November 2019 from all countries. We then classified the queries as support and non-support using our ML model (see Section 2.3). Finally, we analyzed the results to gain insights about Nemty, such as when the ransomware started infecting users and how its distribution evolved over time (see blue boxes below the timeline in Figure 5). Our insights found trends related to the distribution of Nemty that start days before they are reported in the news [8]. This result shows that our query log analysis technique could be used to timely learn about the origin and effectiveness of a ransomware attack, even in the early days of its distribution.

In Figure 6, we show the number of such queries per 4-day periods between August and November 2019. We see that until 08/19, there were no queries about Nemty. However, users started submitting support queries on 08/20, which is likely when the ransomware started first spreading. Indeed, on 08/26 the first news about Nemty are published [8] (see Figure 5). On 09/14, an article [8] discusses how the Nemty authors are enhancing the ransomware with the goal of achieving a wider distribution. Our analysis indeed found that although the support queries started declining early September, on 09/09 there was a spike in such queries (see Figure 6). The insight we gained from this corresponds to the news about the ransomware becoming more efficient and sophisticated.

There was a tremendous increase in support and non-support queries about Nemty between 10/03 and 10/06 (see Figure 6). Published news confirm this finding, as there was a new distribution campaign during October that was targeting enterprise users [8]. We found that Nemty was now spreading worldwide, as support queries started being submitted by an increasing number of countries. After this period of time, the support queries started decreasing, most likely because more people became aware of the ransomware and how to protect from it. In early November there was a minor increase in the queries (see Figure 6), which corresponds to a new distribution method of Nemty via Trik botnet. However, this method was not very efficient, because after this minor increase in queries, the query volume continued to decline.

6 CONCLUSION

In this work, we did the first study to find insights about ransomware attacks using web search logs. We analyzed query logs from a major web search engine using a machine learning classifier to extract support queries by users who were attacked by ransomware. We did a correlation analysis and found that certain features such as query volume and click counts are correlated with attacks. Further, we analyzed geographical and temporal trends and validated our findings from publicly available information. Lastly, we did a case study on the Nemty ransomware and showed that with query log analysis, it is possible to mine key insights about the origin and spread of specific attacks.

REFERENCES

[1] 2019. 2018 Cyber Incident & Breach Trends Report. https://www.internetsociety.org/resources/ota/2019/2018-cyber-incident-breach-trends-report/
[2] 2019. McAfee Labs Threats Report August 2019. https://www.mcafee.com/enterprise/en-us/assets/reports/rp-quarterly-threats-aug-2019.pdf
[3] 2019. The Number of Cyber Attacks Against the North Dakota Government Triples in a Year. https://www.kxnet.com/news/local-news/the-number-of-cyber-attacks-against-the-north-dakota-government-triples/
[4] 2020. Ransomware Attacks Grow, Crippling Cities and Businesses. https://www.nytimes.com/2020/02/09/technology/ransomware-attacks.html
[5] 2020. Ransomware attacks hit schools, but experts question intent, whether trend affects Arkansas. https://www.arkansasonline.com/news/2019/jul/22/ransomware-attacks-hit-schools-20190722/
[6] Chetan Bansal, Thomas Zimmermann, Ahmed Hassan Awadallah, and Nachiappan Nagappan. 2019. The Usage of Web Search for Software Engineering. *arXiv preprint arXiv:1912.09519* (2019).
[7] Davide Canali, Leyla Bilge, and Davide Balzarotti. 2014. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*. 171–182.
[8] NJ Cybersecurity & Communications Integration Cell. 2019. Nemty Ransomware. https://www.cyber.nj.gov/threat-profiles/ransomware-variants/nemty
[9] Stevie Chancellor and Scott Counts. 2018. Measuring employment demand using internet search data. In *Proceedings of the 2018 CHI*. 1–14.
[10] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.
[11] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. 2005. Evaluating implicit measures to improve web search. *ACM TOIS* 23, 2 (2005), 147–168.
[12] Shagun Jhaver, Justin Cranshaw, and Scott Counts. 2019. Measuring professional skill development in US cities using internet search queries. In *Proceedings of the International AAAI CWSM*, Vol. 13. 267–277.
[13] Amin Kharraz, William Robertson, Davide Balzarotti, Leyla Bilge, and Engin Kirda. 2015. Cutting the gordian knot: A look under the hood of ransomware attacks. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 3–24.
[14] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *NIPS'13 - Proceedings(Volume 2)*. USA.
[15] John Paparrizos, Ryen W White, and Eric Horvitz. 2016. Detecting devastating diseases in search logs. In *Proceedings of the 22nd ACM SIGKDD*. 559–568.
[16] Danqing Xu, Yiqun Liu, Min Zhang, Shaoping Ma, Anqi Cui, and Liyun Ru. 2011. Predicting epidemic tendency through search behavior analysis. In *22nd IJCAI*.
[17] G. U. Yule and K. M. Kendall. 1968. An introduction to the theory of Statistics. 14 (1968).