6^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2020, MIT Media Laboratory, Boston, MA, USA

Predicting Software Vulnerability Exploits from Social Media Confabulations

Keywords: Cyber-Security; Software Vulnerability Exploits; Social Media; Split-Population Modeling; Multi-Variate Temporal Point Processes

Extended Abstract

Software vulnerabilities are unintentional flaws in software systems that can pose significant security risks, if exploited by a threat actor. Such vulnerabilities fuel nowadays' proliferation of ransomware attacks, banking and crypto-currency trojans, botnets and other malicious campaigns. Ideally, once made known, affected software vendors will promptly address them and, subsequently, distribute corresponding patches to shield their customers. In reality, pursuing such remedies is a resource-demanding endeavor and often, for vendors with expansive suites of products, vulnerabilities have to be prioritized. Hence, understanding which particular vulnerabilities will be most likely exploited and how soon is of significant value to vendors' vulnerability management.

Aside from zero-day exploits, vulnerability discoveries are commonly communicated first to the National Cyber-security Federally-Funded Research and Development Center operated by the Mitre Corporation, which maintains the Common Vulnerability & Exploits (CVE) system for referencing publicly-disclosed vulnerabilities and assigns a unique identification number to each vulnerability tracked. From there, social media play a key role in spreading vulnerability awareness to vendors, users and cyber-security specialists alike. Moreover, online discussions and activities markedly contribute to the full appraisal of a vulnerability's potential impact. However, steering up of such attention may well motivate ethical as well as malicious actors to produce corresponding exploits.

There is only a handful of recent prior works, such as [1] and [2], which attempt to predict exploit timing based on a vulnerability's social media popularity. While such approaches have met reasonable success in short-term forecasting of exploit timings, they employ discriminative classification and regression models that are opaque, in the sense that the models themselves are difficult to be interpreted and resist providing clear insights into the potential causal nature that social media discussions may impart on the development and eventual publishing of exploits.

In this work, we utilize a comprehensive, probabilistic generative model that, by utilizing a system of interacting temporal point processes, aims at jointly modeling the dynamics between social media activities and a process that determines the timing of when the first exploit is made public. Most importantly, by virtue of its nature and design, our modeling framework is highly explainable. In terms of social media platforms, apart from Twitter (re)tweets (which have been considered in past studies) and Reddit posts and comments, we also consider GitHub repository events linked to patching of vulnerabilities, since we view the latter as being reflective of patching priority. Fig. 1 provides an overview of our model. For each CVE, we consider that our system of processes is kick-started by the Mitre Corporation's entry creation of the vulnerability, which is accompanied by a short description about its nature. We model the ensuing online

confabulations and activities as a subsystem of self- and mutually-exciting Hawkes processes, whose event volume and intensity drives a survival process. This latter process models the timing of the first exploit to appear for the given CVE. Furthermore, a key, innovative aspect of this process is its use of split-population modeling, which postulates that not all vulnerabilities are susceptible to exploitation. In particular, we theorize that a given CVE's susceptibility can often be determined by tell-tale keywords in its Mitre description. Towards this end, we use a deep neural architecture to learn description embeddings, which our model subsequently uses to determine such susceptibility.

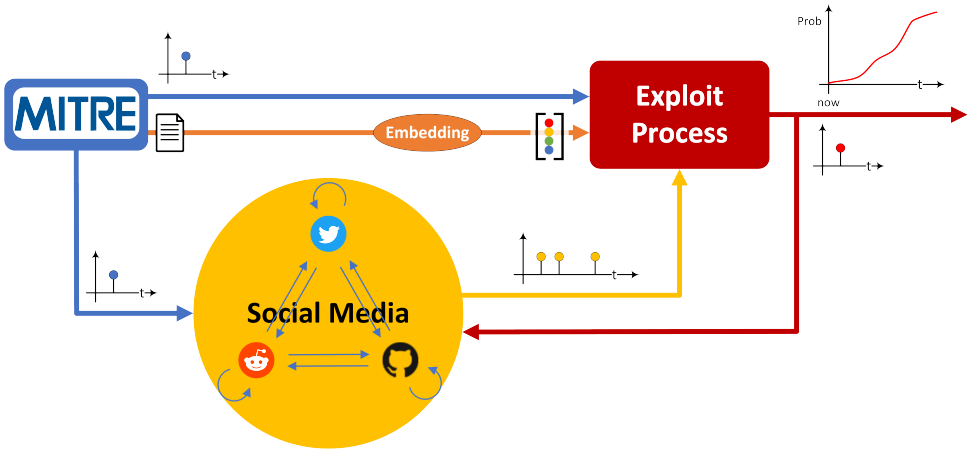

Figure 1: Interactions between processes of our model.

Our work examines 20,283 unique CVEs that were documented by the Mitre Corporation during the two-year period of 2016-03-07 to 2018-03-31. In order to determine which of these CVEs were exploited, we relied on data hosted by the Exploit Database (exploit-db.com), which curates probably the most comprehensive, freely-available collection of exploits; we discovered that only 1,254 ($\sim$ 6%) of them were exploited in our time frame of study. Finally, for the same aforementioned time frame, we identified 679,322 relevant social media events.

Our experimental results showcase, that, for a given CVE, the information encapsulated in its Mitre description in combination with the amount and intensity of social media attention it receives allows for reasonable short- and long-term prediction of whether it is going to be exploited or not within a given time frame. One can trace back this success to our ability to derive effective CVE description embeddings that inform the model about a vulnerability's susceptibility, as well as to the parsimony of our overall framework in modeling all the relevant social media event times.

References

[1] Mehran Bozorgi, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyond heuristics: Learning to classify vulnerabilities and predict exploits. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 105–114, New York, NY, USA, 2010. ACM.

[2] Haipeng Chen, Rui Liu, Noseong Park, and V.S. Subrahmanian. Using twitter to predict when vulnerabilities will be exploited. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 3143–3152, New York, NY, USA, 2019. ACM.