Northwestern | Security & AI Lab

Vulnerability Score and Timing Predictions with VEST

Given the massive number of software vulnerabilities disclosed every year, security officers and administrators are often faced with difficult decisions on devoting time and resources to patching these vulnerabilities. Therefore, it is crucial to know early on whether and when a Common Vulnerability and Exposure (CVE) will be exploited, as well as how severe those vulnerabilities are likely to be. However, today, this information can take months. Research led by NSAIL faculty has previously shown that on average, it takes 132 days between the announcement of a vulnerability by MITRE and the release of severity scores by the US National Institute for Standards & Technology (NIST). During that window of time, multiple exploits can be created and carried out using the announced CVE. To mitigate this, we developed VEST to serve as an early warning system. VEST uses Twitter data about a specified CVE to estimate exploit timing and severity scores. The VEST system was named runner up for the Most Innovative Demo at the 2019 International Joint Conference on Artificial Intelligence.

VEST is the first system to predict when a vulnerability will be exploited and how severe it will be.

VEST (Vulnerability Exploit Scoring & Timing)

The VEST (Vulnerability Exploit Scoring & Timing) system backend consists of a vulnerability database and several predictors. One predictor predicts if and when the CVE will be exploited, while another predictor predicts the severity score. A related set of predictors predict additional, more fine-grained aspects about the CVE's severity. A GUI visualizes scoring and timing predictions, along with the performance of VEST compared to the eventually released scores by NIST. The database is updated daily through a data crawler, a CVE-Author-Tweet (CAT) graph engine, and the predictors. A user can query the updated predictions from the database using the UI, which is integrated with the backend through a Flask framework.

Vulnerability Timing Predictor

VEST predicts when a vulnerability will be exploited after assignment of a CVE number. This project first determines the popularity of a CVE using a novel concept of a CVE-Author-Tweet (CAT) graph which uses three recursively linked popularity measures of "hotness," "expertise," and "availability." Next, retweet volume after an initial training period is estimated with a Hawkes process model.

These two concepts are then used to create ensemble prediction models FEEU (Forecasting Ensemble for Exploit Timing) and FRET (Forecasting Regression for Exploit Timing). FEEU predicts whether a CVE will be exploited within a specified number of months, and FRET predicts the exact date a CVE will be exploited. After evaluation, FRET was found to have a mean absolute error of 11.90 days for real-world exploit prediction. Overall, FRET and FEEU were found to be highly effective at determining timing of vulnerability exploits.

Vulnerability Severity Prediction

The vulnerability score prediction component of VEST predicts Common Vulnerability Scoring System (CVSS) scores on a 1–10-point scale and the 8 CVSS attributes (Attack Vector, Attack Complexity, Privileges Required, User Interaction, Scope, Confidentiality Impact, Integrity Impact, and Availability Impact) using only 3 days of Twitter discussion data after the date when the vulnerability is first mentioned on the platform. VEST builds a graph convolutional network (GCN) along with a unique attention embedding methods for its prediction model, where each node is a CVE and the edges are constructed using the semantic similarities between tweets related to the CVEs. In order to extract the most useful information from a possibly massive and highly varied pool of tweets, VEST introduces a novel concept of an attention-based embedding layer on raw inputs. This embedding layer consists of a bi-directional LSTM to extract a tweet's latent feature vector, an attention layer for sorting out useful tweets, and a moving average layer to handle GPU memory limitations.

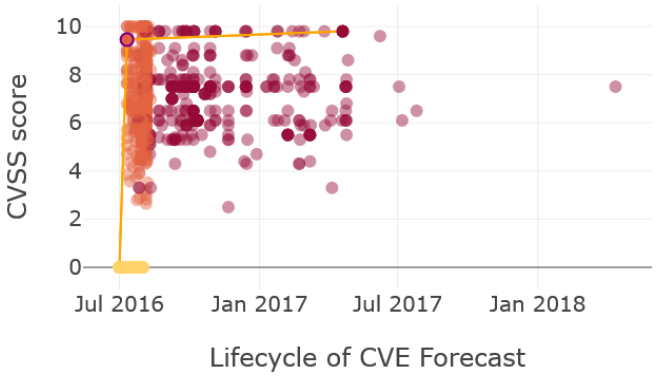
For each CVE, base features are extracted, such as number of tweets, number of verified accounts, number of favorites, etc. as well as a bag of keywords (BoW).

https://sites.northwestern.edu/nsail/projects/vest/

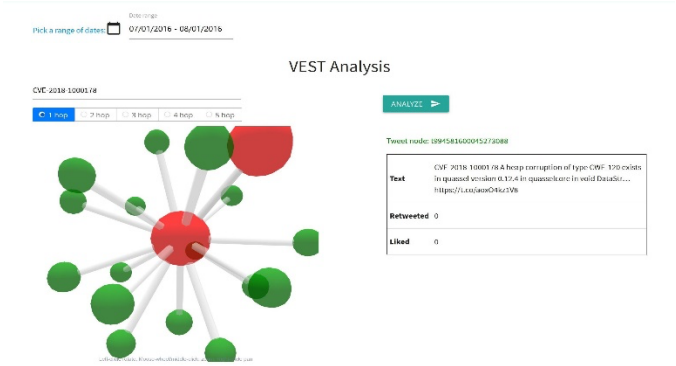
The GCN is then trained by feeding raw input of tweets into the attention-based embedding layer, which is then fed into the GCN layers. The nodes within the CVE graphs are connected based on their base feature similarities.

Our statistics on the 8-months of CVE data show that VEST can predict the CVSS attributes and scores for 37.85% of the CVEs at least one week earlier than the official vulnerability assessments by NIST.

After testing with 8 months of data in the year 2017 from January to August, VEST obtained a mean absolute error of 1.12 for CVSS score prediction, and F1 scores of 0.591 to 0.897 for predicting various CVSS attributes. This demonstrates that VEST is able to serve as a highly accurate early warning system before the detailed evaluations are released by NIST.



VEST demo UI showing CVSS score predictions for a number of CVEs and the true score released by NIST



VEST demo UI showing a graph of CVEs and tweets, along with timing predictions and another details

Video

https://northwestern.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=357af6b3-faa4-496e-944c-aebe015611bc

Additional Information

References

1. Haipeng Chen, Rui Liu, Noseong Park, and V.S. Subrahmanian. Using twitter to predict when vulnerabilities will be exploited. In *Proceedings of the ACM International Conference on Knowledge Discovery and DataMining (SigKDD)*. ACM, 2019.
2. H. Chen, J. Liu, N. Liu, N. Park and V.S. Subrahmanian. VEST: A System for Vulnerability Exploit Scoring & Timing, Demo paper, *Proc. 2019 Intl. Joint Conference on Artificial Intelligence (IJCAI 2019)*, Aug 2019, Macao.
3. Chen, H., Liu, J., Liu, R., Park, N. and Subrahmanian, V.S., 2019, November. VASE: A Twitter-based vulnerability analysis and score engine. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 976-981). IEEE.

PARTICIPANTS

Lead: V.S. Subrahmanian

Haipeng Chen, Jing Liu, Rui Liu, Noseong Park.

Northwestern | BUFFETT INSTITUTE FOR GLOBAL AFFAIRS

Northwestern | McCORMICK SCHOOL OF ENGINEERING