Predicting Software Vulnerability Exploits from Social Media Confabulations

Xi Zhang[1], Akshay Aravamudan[1], Anna Koufakou[2], Chathika Gunaratne[3], Ivan Garibay[3], Georgios C Anagnostopoulos[1]

[1]Florida Institute of Technology, [2]Florida Gulf Coast University, [3]University of Central Florida

Motivation and Objective

Software vulnerabilities are unintentional flaws in software systems that can pose significant security risks, creating an opening to threat actors. Such vulnerabilities have paved the way to an increasing number of malicious campaigns such as ransomware attacks, trojans and botnets to name a few. These campaigns necessitate an effort to better understand software exploits as it relates to vulnerabilities, whose mitigation often requires the mobilization of resources. Understanding which particular vulnerabilities are more likely to be exploited and how soon aids the vendors in prioritizing manpower. To this end, we use a probabilistic generative model to determine the timing of the earliest exploit to be made public.

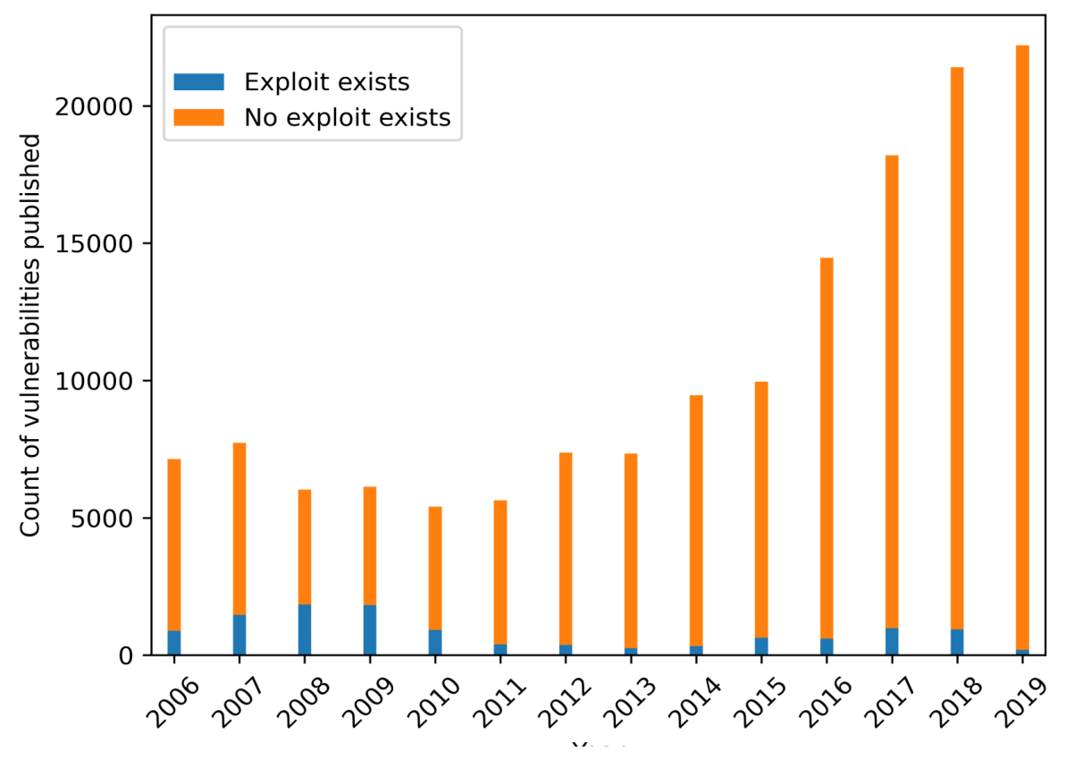

Fig 1. Number of discovered vulnerabilities and exploits per years.

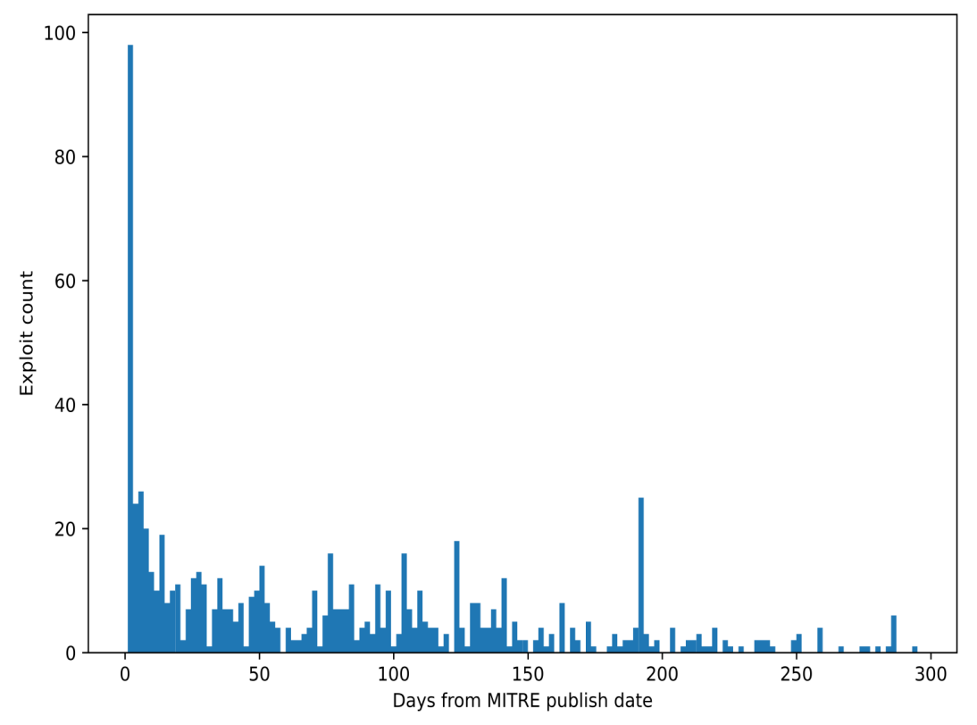

Fig 2. Days it takes for vulnerabilities to be exploited

Assumptions

- Life-cycle of a vulnerability begins after its public disclosure.
- MITRE descriptions largely determine the susceptibility to an exploit.
- Social media stimuli positively drive the availability of an exploit.

Dataset

We collect CVE data from 2016-03-07 to 2018-03-31 (3 years and 2 months). In this time range 20,141 CVE (Common Vulnerabilities and Exploits) Entries had been created by the MITRE corporation.

Exploit-DB data

Maintained by Offensive Security, Exploit-DB serves as one the most comprehensive collection of exploits and proof-of-concepts gathered through a variety of sources. We use the exploit dates for corresponding CVE IDs that are collected in this database. Of the 20,141 CVE-IDs created in our time range, only 1,338 (6.64%) has been exploited after the MITRE published date.

Social Media Data

Social media activity facilitate the spread of awareness of certain vulnerabilities. To understand how discussions on social media influence the availability of an exploit, we query events (in the given time range) associated with each CVE ID from three different social media platforms: Twitter, GitHub, Reddit.

	🐦	👽	🐙	Total
Events Number	547,675	11,933	239,684	799,292
CVE-ID Number	19,995	2,213	8,351	20,141

Modeling

Exploit Process Overall Model Sturcture

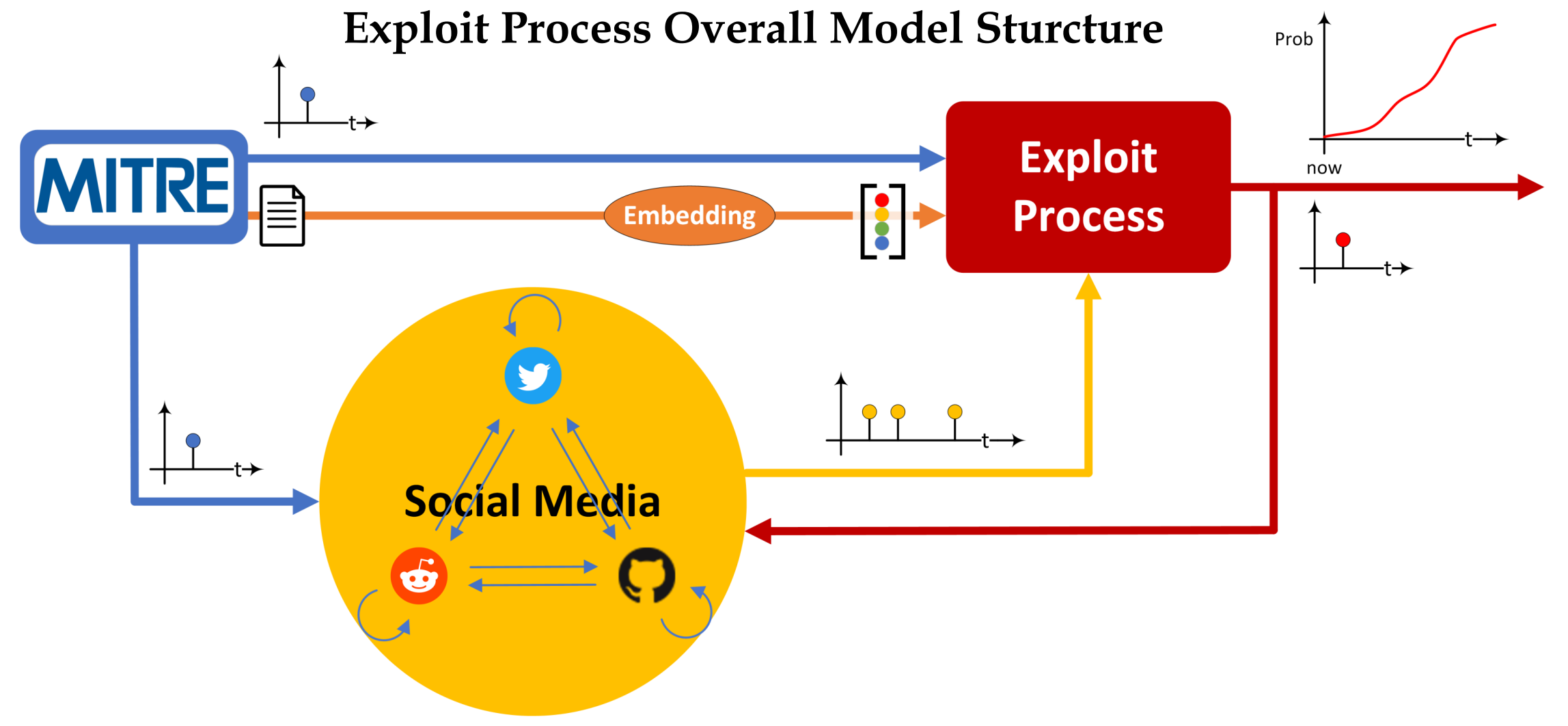

Fig 3. Overall structure of exploit process model: First, the vulnerability feature/embedding is extracted from MITRE description. Then, if it exist, social media events will pose a positive influence, i.e., increase the chance of a vulnerability seeing an exploit. The probability of observing an exploit event is modeled as a survival process with a split population setup.

Survival Process: The appearance of the earliest discovered exploits is modeled as a survival process.

Susceptibility: Two distinct groups of vulnerabilities based on feature vectors: the group that will be exploited, and the group will never be exploited

Base hazard: The base hazard rate of exploit process. The hazard rate of vulnerabilities without any social media events.

SM Excitation: We assume every social media events pose a positive excitation on the exploit event.

Hazard Rate: The instantaneous rate of occurrence of the exploit event for susceptible vulnerability:

$$h_e(t|r=1) = \alpha_e \phi_e(t) + \sum_{s \in \mathcal{S}} \alpha_s \sum_{t_i^s < t} \phi_s(t - t_i^s)$$

Model Training: Training our model amounts to minimizing a negative log likelihood, which is a difference of convex functions. To this end, we use a hybrid Expectation-Maximization and Convex-Concave procedure.

Prediction

Given a previously unexploited software vulnerability and a future timeframe, the probability of witnessing an exploit is given by:

$$\mathbb{P}\{t_c, \Delta_t | x, \mathcal{H}\} = \frac{\pi(x|w)\mathbb{P}_s(t_c, \Delta_t)}{\mathbb{P}_{ne}(t_c)}$$

$\pi(x|w)\mathbb{P}_s(t_c, \Delta_t)$ is the probability of a susceptible vulnerability surviving another Δt

$\mathbb{P}_{ne}(t_c)$ is the probability of not observing exploit event up to current time.

This material is based upon work/research supported in part by the
Defense Advanced Research Projects Agency Grant No. FA8650-18-C-7823

Results

Trained Model

Best model uses	α
Weibull base kernel	0.0018
Exponential GitHub kernel	0.0054
Exponential Reddit kernel	0.0095
Exponential Twitter kernel	0.0037

- The base hazard rate has a Weibull base kernel, whose shape resembles the histogram of exploit times as shown in fig 1.

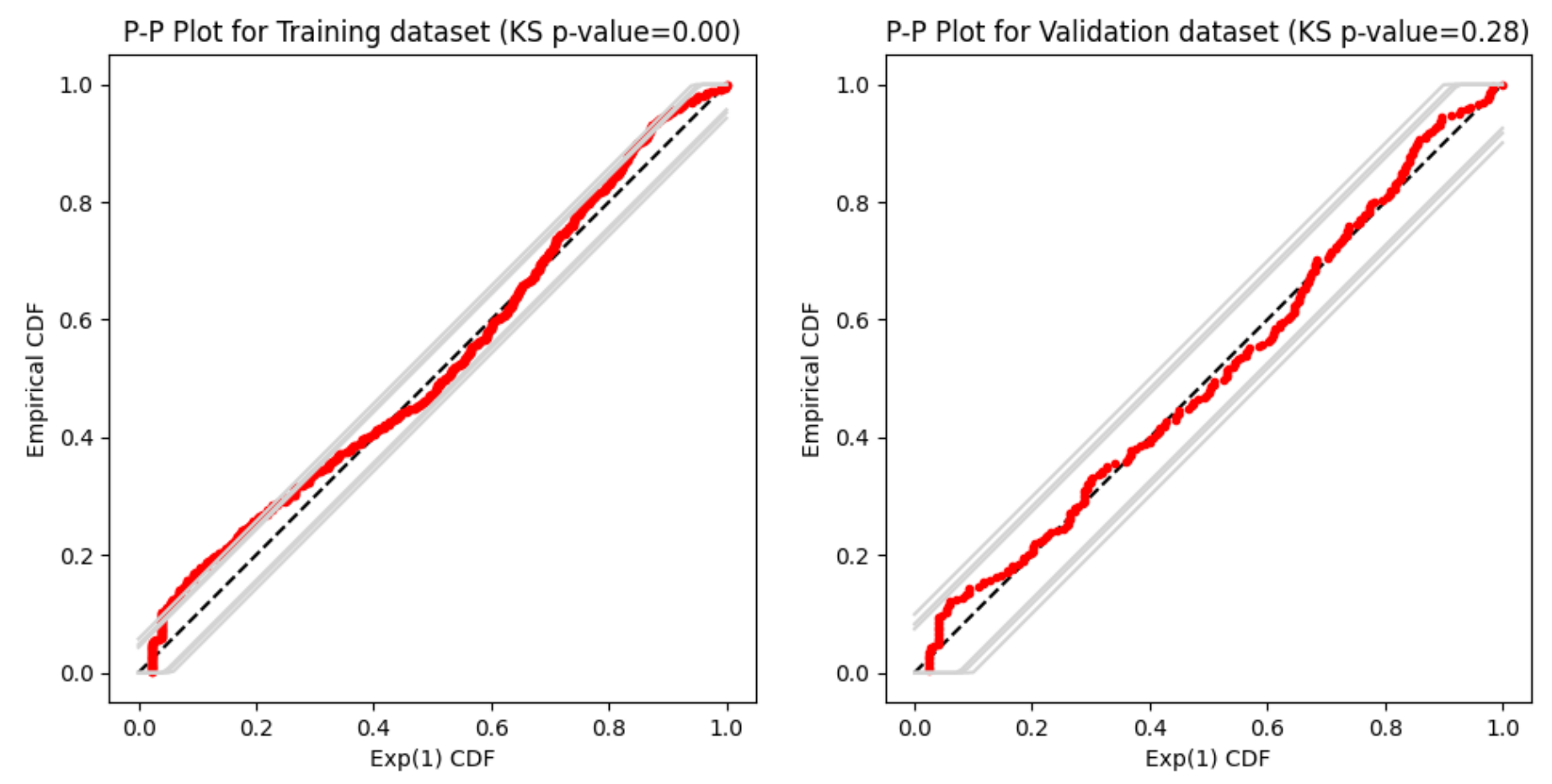

Fig 4. Probability-Probability plot (P-P plot) for training (left) and validation (right) dataset. This pair of plots show a well fitting model with potentially good generalization performance.

Exploit Predictions

- Forecasts deteriorate for prediction intervals greater than 120 days.
- An observation period of 90 days produces the best average accuracy.
- Our model outperforms models that do not leverage social media activity.

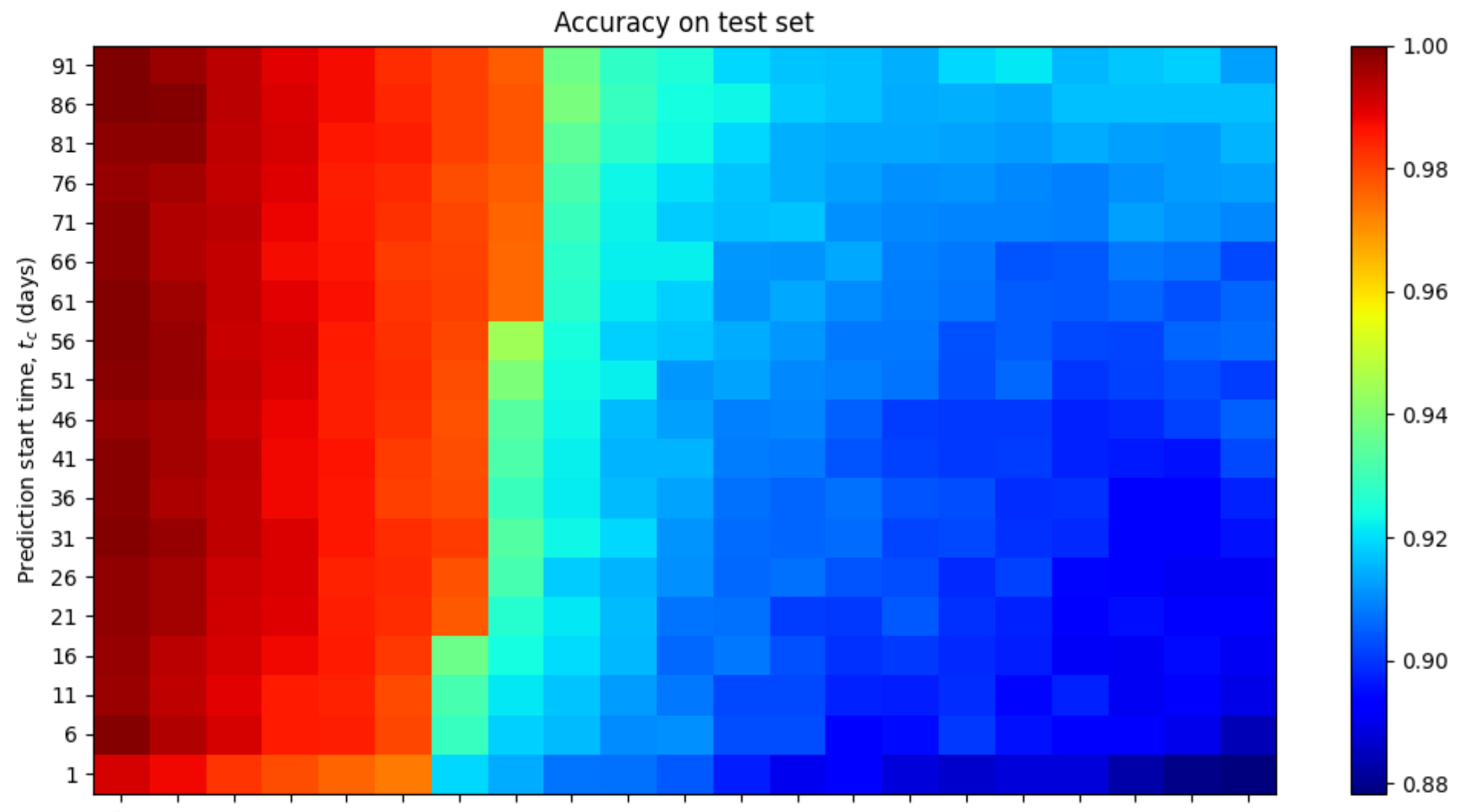

Fig 5. Test accuracy of our model depicted in color. Each patch represents a tuple (t_c , δ_t) where t_c is the start time of the prediction and δ_t is the prediction interval. The accuracy increases as the observation period t_c increases since the model has observed more social media events.

Conclusion

1. The MITRE description of a vulnerability may contain sufficient information to determine the susceptibility to exploitation.
2. Timing, intensity and volume of social media activity about a susceptible vulnerability influences the timing of its exploitation.

References

[1]Mehran Bozorgi, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyondheuristics: Learning to classify vulnerabilities and predict exploits. InProceedings of the16th 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 predictwhen vulnerabilities will be exploited. InProceedings of the 25th ACM SIGKDD Interna-tional Conference on Knowledge Discovery & Data Mining, KDD ’19, pages 3143–3152,New York, NY, USA 2019. ACM.