Analyzing the Perceived Severity of Cybersecurity Threats Reported on Social Media

THE OHIO STATE UNIVERSITY

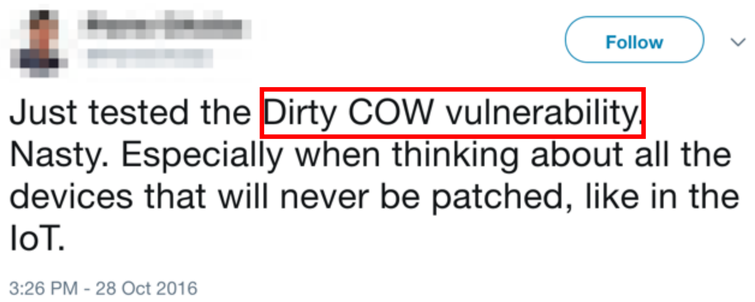
Shi Zong, Alan Ritter, Graham Mueller, Evan Wright

leidos FIREEYE

Introduction

Task

- Analyze perceived severity of cybersecurity threats reported online



Contributions

- **A corpus** of 6,000 tweets annotated with users' opinions towards threats' severity
- **Automatic classifiers** for analyzing users' opinions about threats' severity with high precision
- **A pipeline** for forecasting high severity vulnerabilities, including real-world exploits

Dataset

Data Collection

- Track keywords "ddos" and "vulnerability" in Twitter

Data Annotation – 6,000 Tweets Annotated

- Two-phase annotation by Amazon MTurk
  First annotate for vulnerability mentions, then for severity
- Agreement with experts
  0.66 Cohen's kappa for vulnerability mentions (0.52 for severity)

> [nerID_21413] New post: "Adobe Security Advisory : A critical vulnerability in the Adobe flash player" https://t.co/RnKFmIwc96
>
> Based on the text above, does the author think the cybersecurity threat to **adobe flash player** is exploitable and could affect many users? Does the author feel users should be worried about this threat?
>
> ○ There is a severe cybersecurity threat towards **adobe flash player**
> ○ There is a moderate cybersecurity threat towards **adobe flash player**
> ○ Above two choices don't apply

Annotated Corpus Statistics

1st Annotation (by 5 workers)			2nd Annotation (by 10 workers)		
Label	# Tweets	%	Label	# Tweets	%
With Threat	2,543	42.4	Severe Threat	506	25.7
			Moderate Threat	1,460	74.3
Without Threat	3,457	57.6	/		

A Live Demo

Date: Mar-04-2019		
Entity	**Representative Tweet**	**Severity (Est.)**
macos	Google Project Zero discloses high-severity zero-day vulnerability in MacOS after Apple fails to patch within discl... https://t.co/iHVXmfMsR2	Probably Severe 0.492
xml	CVE-2018-13798 Siemens SICAM A8000 Series suffers from an XML injection denial of service vulnerability.... https://t.co/vMaFFM3XO9	0.078
apple	RT @stickypassword: Google's Project Zero has publicly disclosed a zero-day vulnerability in Apple macOS software after a deadline to resol...	Probably Severe 0.479
google	RT @whizsec: Google discloses High-Severity Flaw in MacOS Kernel A high-severity vulnerability has been found in Apple's MacOS which was r...	Probably Severe 0.34

WIRED

APPLE
Google Finds "BuggyCow," a Rare MacOS Zero-Day Vulnerability
ANDY GREENBERG

Check our paper Check our demo

Analyzing Perceived Severity

Task Setup

- For a given tuple <ENTITY, TWEET>, identify (1) if there exists a cybersecurity threat towards ENTITY, and (2) if the threat is severe

Model Performance

Task	Model	Dev AUC	Test AUC
Vulnerability Detection	Logistic Regression	0.88	0.85
Threat Severity	Logistic Regression	0.62	0.54
	CNN	0.70	0.65

Our text-based method can accurately identify vulnerability mentions and opinions about their severity

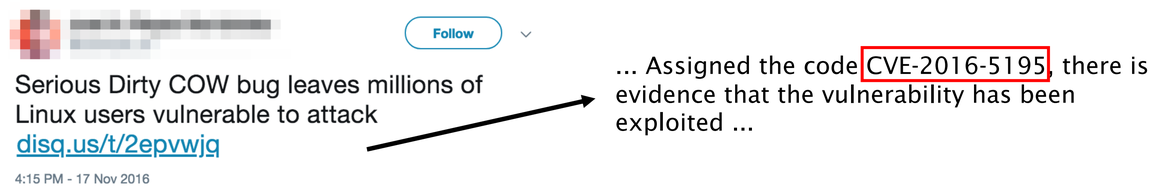
Top Ranked Features

Features	Weight	Features	Weight
ddos attack	1.40	lets attackers	0.95
hackers to	1.11	<TARGET> users	0.91
a massive	1.07	a critical	0.91
critical vulnerability	1.03	of a	0.89
0 billion	0.96	many <TARGET>	0.89

Forecasting Severe Threats

Linking to the National Vulnerability Database (NVD)

- Check CVE numbers in linked webpages



Forecasting Setup

- Consider CVEs where >= 2 associated tweets were posted at least 5 days ahead of official NVD publication date
- **13,942** tweets for **1,409** unique CVEs are gathered for evaluation

Forecasting NVD's Severity Ratings (CVSS Scores)

	P@10	P@50	P@100	AUC
Random	59.0	61.2	58.8	0.595
Volume Model	70.0	68.0	70.0	0.583
Our Model	**100.0**	**86.0**	**78.0**	**0.658**

Our model achieves P@50 of 86% when forecasting severe vulnerabilities (CVSS score >= 7.0)

Forecasting Real Exploitable Threats

	Top 10		Top 50		Top 100	
	P	R	P	R	P	R
True CVSS	10.0	0.7	16.0	6.0	16.0	11.9
Volume Model	60.0	4.5	22.0	8.2	19.0	14.2
Our Model	**70.0**	**5.2**	**28.0**	**10.4**	**21.0**	**15.7**

Our model outperforms baselines for predicting real-world exploits (as identified antivirus signatures)

Sample System Output

CVE-2016-0728
Date in NVD:
2016-02-08
CVSS scores:
7.2 HIGH (v2.0)
7.8 HIGH (v3.0)

Tweets Matched to CVE-2016-0728	Score	Date
Vulnerability in the Linux kernel could allow attackers to gain access to millions of Android devices! <URL>	0.98	2016-01-20 (**+19**)
A Serious Vulnerability in the Linux Kernel Hits Millions of PCs, Servers and Android Devices <URL>	0.89	2016-01-20 (**+19**)
Millions of PCs and Android devices are at risk from a recently discovered critical zero-day vulnerability. <URL>	0.89	2016-01-20 (**+19**)

Users' opinions can be an early indicator to help prioritize severe threats