Extracting Cybersecurity related entities, terms and concepts from text.



RAVENDAR LAL

ADVISER: DR. TIM FININ

M.S. THESIS DEFENSE

EBIQUITY RESEARCH LAB

UMBC

So... what's happening here?

An Excerpt:

"Adobe has released security updates for Adobe Reader and Acrobat X (10.1.3) and earlier versions for Windows and Macintosh. These updates resolve a buffer overflow vulnerability that could lead to remote code execution....."



So... what's happening here?

An Excerpt:

"Adobe has released security updates for Adobe Reader (SOFTWARE) and Acrobat X (10.1.3) (SOFTWARE) and earlier versions (NER_Modifier) for Windows (OS) and Macintosh (OS). These updates resolve a buffer overflow (MEANS) vulnerability in AcrobatReader.exe (FILE_NAME) that could lead to remote code execution (CONSEQUENCES)"



Preview

I. Introduction and Problem

II. My Contribution

III. Results and Evaluation

IV. Future work and Conclusion



Introduction

- Cyber Security Threat
- Attacks ranging Government to large internet population
- Zero Day Attacks
 - YAJO
- Software Vulnerabilities





Too many data sources...

vulnerability

measurement, and

lesource Status NVD contains: 55691 CVE Vulnerabilities 204 Checklists

244 US-CERT Alerts

8140 OVAL Queries

Email List

2702 US-CERT Vuln Notes

Last updated: 03/29/13 **CVE Publication rate:** L4 vulnerabilities / day

NVD provides five mailing

subscription instructions

please visit <u>NVD Mailing</u>

lists to the public. For information and

compliance (e.g. FISMA)

per Security Division/US-CER National Vulnerability Database utomating vulnerability manag ecurity measurement, and compliance checking

Impact Metrics Data Feeds Statistics 800-53/800-53A Product Dictionary Checklists me SCAP SCAP Validated Tools SCAP Events About Contact Mission and Overview

NVD Data Feed and Product Integration NVD is the U.S.

government repository of standards based The entire NVD database can be downloaded from this web page for public use. There are vulnerability management no licensing restrictions on using this data, however, we would appreciate being given data. This data enables credit as is appropriate within products, services, and reports that use our data. automation of

SCAP Data Feeds: management, security CVE vulnerability feeds: security related software flaws

12 Critical Updates for Windows, Adobe Flash, MAR 13 Air

Microsoft and Adobe each released patches today to plug critical security holes in their products. Microsoft issued seven update bundles to address at least 19 20 vulnerabilities in Windows and related software. Adobe released the fourth security update in nearly as many weeks for its Flash Player software, as well as a fix for Adobe AIR.

Microsoft today began pushing out seven security patches, four of them rated "critical," meaning the flaws they fix could be used by malware or bad guys to break into unpatched systems with little or no help from users. The critical patches address bugs in Windows, Internet Explorer, Microsoft Silverlight, Microsoft Office and Microsoft SharePoint. Updates are available for Windows XP, Vista, Windows 7, Windows 8, Windows Server 2003, 2008 and 2012.

More information on the Microsoft patches is available at the Microsoft security response center blog, which also discusses some changes to the way security updates are applied to apps available through the Windows Store.

Continue reading \rightarrow



Web



ta

sicq

Big Picture?

Better auto processing of:

- threats with minimal human effort
- concepts and entities from live data streams
- relations between the entities

Questions like:

Which software is being talked about?What kind of attack they are mentioning?Who and how it is being attacked?





My Mission Statement

"Develop a system that can identify mentions of Cybersecurity-related entities (e.g., software applications) and concepts (e.g., buffer overflow) to assist in extracting Cybersecurity data from text."



My Contribution

✓ Dataset Collection - Corpus of Cybersecurity related text

Manual Annotation of collected data

✓ Interface between Stanford NER and annotation tool

Cybersecurity Entity and Concept Spotter





Training Methodology





Dataset

- Blogs (50 Various Blog posts)
 - Major contributor (Krebsonsecurity.com)
- Common Vulnerability and Enumerations (300 CVEs)
 - Best for some classes specially like Means and Consequences

- Official Security Bulletins (100 Documents)
 - Adobe
 - Microsoft



Common Vulnerabilities and Exposures (CVE)

Maintained by MITRE (mitre.org)

• Example of a CVE Summary:

"The Firefox sandbox in Adobe Flash Player before 10.3.183.67 and 11.x before 11.6.602.171 on Windows and Mac OS X, and before 10.3.183.67 and 11.x before 11.2.202.273 on Linux, does not properly restrict privileges, which makes it easier for remote attackers to execute arbitrary code via crafted SWF content, as exploited."



Security Bulletins

Excerpt from MS-Bulletin



"This security update resolves a privately reported vulnerability in Visual Studio Team Foundation Server. The vulnerability could allow elevation of privilege if a user clicks a specially crafted link in an email message or browses to a webpage that is used to exploit the vulnerability. In all cases, however, an attacker would have no way to force users to perform these actions. Instead, an attacker would have to convince users to visit a website, typically by getting them to click a link in an email message or Instant Messenger message that takes them to the attacker's website..........."



Annotation Classes

- 1. Software (e.g., Microsoft .Net Framework 3.5)
 - a. Operating_System (e.g., Windows XP SP2)
- 2. Network_Terms (e.g., Firewall)
- 3. Attack (From the work of of Jeffery Undercoffer)
 - a. Means Methods to conduct an attack(e.g., Buffer overflow)
 - b. Consequences final result of an attack (e.g., Denial of service)
- 4. File_Name (e.g., msll_32.dll)
- 5. Hardware (e.g., IBM Mainframe B152)
- 6. NER_Modifier (e.g., Acrobat Reader X and earlier versions)
- 7. Other_Technical_Terms (None of the above. E.g., HTML)





Annotation Task

- Most time consuming and painstaking part
- Asked CS grad students to help me in this regard
- 12 Annotators
- Provided training and guidelines
- •Why Mturk Failed?
 - User selection
 - Domain expert knowledge required



Annotation Task (Brat)

Brat (<u>http://brat.nlplab.org/index.html</u>)

- Easy to setup server and data
- Currently online at: <u>http://swoogle.umbc.edu/brat/</u>
- Easy to annotate
 - Keyboard Shortcuts
 - Highly interactive User Interface







Annotations To Stanford NER (Interface)





Brat System

1. Txt File: This security update resolves a privately reported vulnerability in the implementation of SSL and TLS in Microsoft Windows.

2. Ann File: T1 Network_Terms 90 93 SSL

- T2 Network_Terms 98 101 TLS
- T3 Operating_System 105 122 Microsoft Windows

3. Int File: This security update resolves a privately reported vulnerability in the implementation of **SSL#NETWORK** and **TLS#NETWORK** in **Microsoft@Windows#OPERATINGSYSTEM**



4. Token File

This

security

update

••••

SSL#NETWORK

Privately

TLS#NETWORK

Vulnerability

••••

.

Microsoft@Windows#OPERATINGSYSTEM



5. Stanford NER Input File

This 0 security O 0 а privately O TLS NETWORK, Vulnerability 0 in 0 **OPERATINGSYSTEM**, Microsoft Windows **OPERATINGSYSTEM**,



Stanford NER

- Part of Stanford NLP Toolkit
 - Stanford NER was introduced in 2005
- Provides java based general implementation of Conditional Random Field Sequence models
- Provides well-engineered feature extractors for Named Entity Recognition
- Provides a lot of options for Feature selection





Training New Models From Stanford NER

- Features
 - edu.stanford.nlp.sequences.NERFeatureFactory
 - Easiest option: just add new features here
 - Lots of built in stuff
 - Specifying features
 - edu.stanford.nlp.sequences.SeqClassifierFlags
 - Stores global flags
 - Initialized from Properties file



Features

• Big Problem: Too Little Documentation but great online support

- Most Critical Part: Understanding the data
 - How entities are described?
 - what are the words that we want to extract and any special property of (or surrounding) the word?
- Initial selection is very important



Features that we used...

- useTaggySequences=true (i.e. HMM of classes)
- useNGrams=true (i.e. make substrings of word and see if there is any significant pattern that can be noticed from training data for those "Class" of an entity).
- usePrev= true (i.e. Use the pair of (previous word, class) during training data and make features on that)
- useNext = true (i.e. Use the pair of (next word, Class))
- maxNGramLeng=6 (max size means it can make smaller than that too)
- noMidNGrams=true (n-grams that only contain beginning or end of the word)
- useWordPairs = true (Features for (pw, w, c) and (w, nw, c))
- gazette= OperatingSystem.txt, Softwares.txt
- useGazettes = true
 - cleanGazette=true
 - SloppyGazette = false



Results and Evaluation

5-Fold Cross Validation Results

Strategy for training and testing Systems:

- Divided annotated data in 5 chunks (20% each)
- Trained the classifier with 5 various combinations of chunks
- Training: Testing ratio is (80:20)
- Evaluation score is Based on precision and recall model Precision = TP / (TP+ FP) Recall = TP / (TP + FN) F1 = 2 (Precision x Recall) / (Precision + Recall)



Data Corpus

- 300 CVEs
- 50 Adobe Security Bulletins
- 50 Microsoft Security Bulletins
- 50 Blog Posts

• Processing speed at around 200 words/sec

Task	Tokens	Entities
Training	30,000	3,800
Testing	9,000	1,200



5-Fold Cross Validation Results



Data for previous graph

Classifiers	Р	R	F1	TP	FP	FN	
Classifier 1	0.8273	0.72	0.7699	594	124	231	
Classifier 2	0.8668	0.8127	0.8389	729	112	168	
	0 70 4	0 7440	0 7000	605	4 - 7	200	
Classifier 3	0.794	0.7442	0.7683	605	157	208	
Classifier 4	0.8239	0.7484	0.7844	702	150	236	
Classifier 5	0.8625	0.7862	0.8226	809	129	220	
							F
Average	0.836536	0.76388272	0.798560316	3439	672	1063	
						B	

Average (Class-wise)



Average (Class-wise)

Entity	TP	FP	FN
ATTACK	30	14	27
CONSEQUENCES	299	123	135
FILE	52	0	0
HARDWARE	3	0	2
MEANS	185	94	177
MODIFIER	320	79	147
NETWORK	14	15	45
OPERATINGSYSTEM	920	34	36
OTHER	167	89	230
SOFTWARE	1449	224	268
Total	3439	672	1063

Average (Dataset-wise)





Inter-Annotator Agreement

• To see agreement on various classes between users

• Class level agreement – helps to identify those classes which are confusing

• Same pattern was found on trained classifiers

• Experimented on 10% of total dataset for main classes



Some of the problems in text

"Cross-site scripting (XSS) vulnerability in show_bug.cgi in **Bugzilla before 3.6.13**, **3.7.x and 4.0.x before 4.0.10**, **4.1.x and 4.2.x before 4.2.5**, and **4.3.x and 4.4.x before 4.4rc2 allows remote attackers to inject arbitrary web script** or HTML via the id parameter......"



Some of the problems in text

"Cross-site scripting (XSS) vulnerability in show_bug.cgi in **Bugzilla before 3.6.13**, **3.7.x and 4.0.x before 4.0.10**, **4.1.x and 4.2.x before 4.2.5**, and **4.3.x and 4.4.x before 4.4rc2** allows remote attackers to inject **arbitrary web script** or HTML via the id parameter that can result in delayed service response or even denial of service......."



Inter-Annotator Agreement (classes)



Other Systems for extracting named entities and concept terms

• OpenCalais

- Designed for people, places, organization, facts and events
- Very Poor Performance on Means and Consequences
- Unable to capture version numbers

DBPedia Spotlight

- Identify only those concepts that are present in Wikipedia
- Poor performance on list on means, consequences, file name and even software with low confidence score



OpenCalais

Example:

"Adobe Reader and Acrobat 9.x before 9.5.4, allow r emote attackers to execute arbitrary code or cause a denial of service (memory corruption) via a craft ed PDF document, as exploited in the wild in Februa ry 2013".



DBpedia Spotlight

Example:

"Buffer Overflow in the Broker Service in Adobe Flash Player 10.x and earlier versions on Windows XP SP2 and Mac OS X on Linux, allows attackers to execute arbitrary code via unspecified vectors."



Challenges

• Consequences vs. Means

• Is it a consequence? Or Is it a means? Or May be none?

• Data Noise (i.e., "(Parenthesis)")

Domain Experts for annotations

Stanford NER documentation



Conclusion

- There is a lot of information on Internet on software vulnerabilities
- Need to build automated service to process data streams
- Early detection, early prevention
- This Cybersecurity Entity and Concept spotter is the first step and can be used for multiple applications requiring processing of technical texts
- We believe this work can be used as foundation for analyzing Cyber security web text







Questions?



Thank You.

Back Up Slides

3-Fold Cross Validation Results

Strategy for training and testing Systems:

- Divided annotated data in 3 chunks (33% each)
- Trained 3 classifiers with various combinations of chunks
- Trained on 2/3rd and Tested on 1/3rd
- Evaluation score is Based on precision and Recall model



Average (Class-wise) for 3-Fold



Average (Dataset-wise) for 3-Fold



Classifier 2





Classifier 2 (Data-wise)

Classifier 2 (In detail)

Entity	ТР	FP	FN
ATTACK,	4	1	6
CONSEQUENCES,	59	31	27
FILE,	9	0	0
MEANS,	45	27	44
MODIFIER,	46	27	21
NETWORK,	1	1	7
OPERATINGSYSTEM,	156	8	4
OTHER,	20	12	40
SOFTWARE,	265	50	50
Totals	605	157	208



Classifier 2 (Adobe)

Entity	Р	R	F1	ТР	FP	FN
ATTACK,	0.7273	0.7273	0.7273	8	3	3
CONSEQUENCES,	0.5556	0.3846	0.4545	5	4	8
FILE,	1	1	1	2	0	0
OPERATINGSYSTEM,	1	1	1	128	0	0
OTHER,	0.6667	0.6667	0.6667	2	1	1
SOFTWARE,	0.9167	1	0.9565	121	11	0
Totals	0.9301	0.9466	0.9383	266	20	15

Classifier 2 (CVE)

Entity	Р	R	F1	TP	FP	FN
ATTACK,	1	0.5	0.6667	1	0	1
CONSEQUENCES,	0.7407	0.8163	0.7767	40	14	9
FILE,	1	1	1	7	0	0
HARDWARE,	1	0.5	0.6667	2	0	2
MEANS,	0.7778	0.5714	0.6588	28	8	21
MODIFIER,	0.8816	0.6907	0.7746	67	9	30
NETWORK,	0.3333	0.3333	0.3333	1	2	2
OPERATINGSYSTEM,	1	0.9348	0.9663	43	0	3
OTHER,	0.8333	0.625	0.7143	35	7	21
SOFTWARE,	0.8182	0.8438	0.8308	81	18	15
Totals	0.8402	0.7457	0.7902	305	58	104

Classifier 2 (Microsoft)

Entity	Р	R	F1	ТР	FP	FN
CONSEQUENCES,	0.875	0.875	0.875	14	2	2
MEANS,	0.7273	0.8889	0.8	8	3	1
NETWORK,	0.8571	0.4	0.5455	6	1	9
OPERATINGSYSTEM,	0.9744	0.9268	0.95	38	1	3
OTHER,	0.7333	0.7333	0.7333	11	4	4
SOFTWARE,	0.7708	0.881	0.8222	37	11	5
Totals	0.8321	0.8261	0.8291	114	23	24

Classifier 2 (Blogs)

Entity	Р	R	F1	ТР	FP	FN
MEANS,	0	0	0	0	1	2
NETWORK,	0	0	0	0	1	2
OPERATINGSYSTEM,	0.875	0.9333	0.9032	14	2	1
OTHER,	0	0	0	0	1	1
SOFTWARE,	0.8108	0.6667	0.7317	30	7	15
Totals	0.7857	0.6377	0.704	44	12	25

Classifier 2



Classifier 2 (Data-wise)



Classifier 3



Classifier 3 (Data-wise)



Classifier 4



Classifier 4 (Data-wise)



In Numbers

