ML in practice: Malware detection

JSTAP: A project does malicious JS detection

Premise/problem JS can cause:- bitcoin mining, abuse browser vulnerabilities

Abstract Syntax Tree – Derived from grammar of programming language

JSTAP Principle : Perform static analysis with abstract syntax trees and random forests

Static Analyses

Static analysis - we don't run the code at all, Reverse analysis of the code

dynamic analysis - run the code in virtual machine or debugger. Malware writers deliberately obfuscate to defeat static tools. Example: GozNym runs trivial infinite loop in thread, then suspends thread and overwrites code with jump to previously dead code

Dynamic Analysis pitfalls - 1. Easy to detect you are in a debugger, VM, or running Anti-virus – Query registry – IsDebuggerPresent – VM specific instructions. 2. Do long delay in hopes simulator will give up and go away

Control Flow Graph – Shows program flow (calls, selection, loops)

Program Dependence Graph – Includes data and control dependencies

Token - splitting a program into lexical units (words in sentences for English)

ML in practice: Malware detection (cont)

N-gram - simple way to analyze token sequences

JSTAP n-grams

- Depth-first pre-order traversal of AST

- For CFG, also traverse AST, but only nodes linked by control flow edge. Traverse sub-AST for each node with control flow once

- Similar for PDG, considering data flow

- Independent n-grams for tokens, AST, CFG, PDG-Data Flow and PDG-Control Flow

- 4 is the best value.

- Use chi-squared test to check for correlation(check the ngram in benign or malcious), keep x^2(chi squared) geq 6.63 (confidence of 99%)

- if ngram in both (benh and malc), throw ngram away

JSTAP Dataset - 131448 malicous, 141768 benign

JSTAP Classifier Training • Select 10,000 malicious and benign randomly for training – Additional 5,000 of each for validation • Repeat 5 times and average detection results

JSTAP results • Two step process • First phase – Unanimous voting, classifies 93% of data with 99.73% accuracy • Second phase – Unanimous voting, classifies 6.5% of data with accuracy still over 99%

ML in practice: Malware detection (cont)

Evasion techniques - Add more benign features • Copy malicious into larger benign file

Extremely avstract OS- learn the sample without implementing the underlying OS. Over-approxi-mation has more behaviors than system S, under-approximation has fewer. Less precise than virtualization or emulation

Abstract execution - A Technique for Efficiently Tracing Programs. In a dynamic analysis, it has Emulator, Extremely avstract OS and paths, Less precise than virtua-lization or emulation

ML in practice: Phishing detection

Phishing Websites - Often used to collect credentials. Fake website to induce personal info.

Techniques for finding Phish:

- Industrial toolbar-based: Eg SpoofGuard, TrustWatch, Netcraft (found these ineffective)

- User-Interface-based: Eg provide custom image per user, Password manager (Only provides password to certain domains)

- Web page content-based: Use web page info (URL, links, terms, images, forms) to detect phishing

ML in practice: Phishing detection (cont)

–- CANTINA: compute term frequency-inverse document frequency for terms, then Google a few terms to see if current website is a top result – B-APT: Bayesian based on tokens from DOM

Some defination: Surface level content-URL, hyperlinks, Textual content-Terms or words, Visual content- Color, font size, style, location of images

Textual and visual classification: text classifiers work by examining text within a page to detect whether certain words are more likely in a fraudulent page or not. Image classifiers transform webpage to images and then compares similarity to genuine webpages.

Step of baye analysis: 1. Obtain webpage and normalize 2. Compute signature 3. Calculate EMD and similarity between website and protected web page 4. Classify via threshold

Overall framework 1. Train text and image classifier, collect similarity measurements for different classifiers 2. Partition similarity into sub-intervals 3. Estimate probs for text classifier 4. Estimate probs for image classifier 5. Classify each test image 6. If different from two classifiers, calculate decision factor 7. Return final classific-ation

High quality dataset:

ML in practice: Phishing detection (cont)

accessibility: publicly available;**completeness**: encompass all the breadth within phishing; **consistency** : range and variance of dataset to make sure data won't be substantively changing; **integrity**: data and labels is correct, non-corrupted; **Validity**: data is properly representative; **interpretability** : data is understandable; **Timeliness:** data is updated or still valid today and future

Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction

boosting classifier is random forests build each tree independently while gradient boosting builds one tree at a time. This additive model (ensemble) works in a forward stage-wise manner, introducing a weak learner to improve the shortcomings of existing weak learners.

Social network security - Spam

Spam - irrelevant messages sent to many, Spamming is the use of messaging systems to send multiple unsolicited messages (spam) to large numbers of recipients

Social network security - Spam (cont)

Criminal accounts tend to be socially connected, Maybe less discriminating in who they follow – Maybe intentional

Criminal hubs are more inclined to follow criminal accounts

K-anonymity - Publisher decides which attributes public/private – Public are “quasi-identifiers” • Every quasi-identifier tuple appears in at least k records in anonymized DB

Determine if a database is k-anonymous for a particular value of k - for quasi-identifier, if it appears in at least k records in the db. Every public tuples appears at least twice. We can't uniquely identify someone. A database is 2-anonymous if no click trace is unique

how an attacker might deanonymize a database with auxiliary information(background info related to record)

- Amplification of background knowledge - Uses Aux(r) close to r on subset of attributes to find r’ close to r on all - Extended to a subset

1. Compute score(aux, r’) for each r’ in sample 2. Apply matching criteria 3. Output record or probability distribution for records

Social network security - Spam (cont)

Bystander - Someone who is “present but not taking part” in the photo, Someone who is “not a subject of the photo and is thus not important for the meaning of the photo”

How bystander detection could improve privacy: this can stop bystanders from being recorded without knowing or let them know. Self-centered photos can put bystanders in awkward situations, poor posture, or reveal information they don't want on record,

Unicity - Proportion of unique pieces of information U =0 is k-anonymous and k>=2. U =0.25 means 1/4 of the click traces are unique.

How to get < 10% unicity • Remove all info pertaining to clients and website visits • Coarsen time to at least hours

Strategic manipulation, propaganda, and fake news

fake news - news that is itentionally false, published by news outlet.

challenges in defining “fake news” - apart from validity of information, is it satire, actual misinformation, intended for deception, clickbait, rumor etc.

automatic fact-checking - compare with knowledge/expert base (references); use base of SFO triples: subject, predicate object

Strategic manipulation, propaganda, and fake news (cont)

fact extraction: redundancy(Donaldjohntrump vs donald-trump), timeliness(Britain, joinIn, EuropeanUnion), conflict, unreliability(TheOnion), incompleteness(May need to infer if something is missing)

Why temporal analysis may help with fake news detection: time can change the validity of information Why source analysis may help with fake news detection: is the news satire or credibile

Explain how textual and visual analysis may help with fake news detection

Cond

-textual can determine fake news by Quantity, Complexity, uncertainty, subjectivity, sentiment, informality, specificity and readavlity

- visual content can clarity, coherence, similarity distribution, diversity and clustering score.

- using SVM’s and CNNs for text analysis

mixed code - Use of different languages, symbols, scripts, shapes to avoid detection. Text on Document – Defined from standard alphabetic characters • Text in Visual Media – Text in pictures • Text as Art Form** – Use symbols not part of the alphabet to depict a simple code

Cond (cont)

frequency-inverse document frequency - tfidf is used to reflect how important a word is to a document in a collection tfidf

bi-clique - bipartite graph where every vertex of first is connected to every vertex of second

Label bi-partite graph with nodes as articles and users, Edge if user mentions article, Find maximal bi-cliques,

Find temporal cohesion, And textual cohesion, And created weighted sum, For an article, average its score in all bi-cliques,

top 5% of these are seeded fake, Bottom 5% are seeded true

Spread labels if – Part of same bi-cliques – Have a lot of common users – Are textually similar,

Spread labels based on – Common users – Textually similar

Dark Web

Deep Web: (password) consists of internet not indexed on search engines (such as social media)

Dark Web: (Tor) overlay networks that use the Internet but require specific software, configurations, or authorization to access -Behind password logins – Encrypted – Not linked – Tor Hidden Servcies

ransomware: threatens to publish victims data or holds data hostage unless paid

Dark Web (cont)

Tor browsing: use many(3) different machine to create onion networks. Each connection is encripted beside of the exit.The exit will appear to be browsing.

Tor hidden service - introduction points , directory service () and rendezvous point.

1.pick introduction points to build encrpted tunnels 2. announce the service into db. 3. User get back to 3 introduction points and create rendezvous points (3 steps from) and 4. send msg to intro point. 5. now the rend point is 6 hops away from intro.

beneficial uses of Tor and anonymous browsing: can prevent control from author-itarian regimes; people cannot be banned from accessing information

socially detrimental uses of Tor and anonymous browsing: can be used as a harbor for illegal/ illicit things

how Tor traffic could be deanon-ymized by a large organization: they with the computational power can get both a entry and exit point and then be able to decrypt what goes on in between

Dark Web (cont)

how researchers have crawled the dark web: first get access by identifying dark web forms. Then get data thru anon access, then process and identify relations-hips/ link data sources etc. then visualization and reports

why dark web crawling is beneficial for security practi-tioners - are able to limit the damage of a data breach and take the necessary steps to protect business, employees, customers, etc. from potential attacks. Can be used to detect/ collect any leaked information

Information gain - reduction of entropy gained by knowing feature x: IG(y|x) = H(y) – H(y|x)

Stemming - remove suffixes to get stem word can be use to handling-misspellings with 3-7 ngrams

REST

Abstract execution records a small set of events during the traced program's execution. These events serve as input to an abstract version of the program that generates a full trace by re-executing selected portions of the original program.

insider threat and accidental insider threat: threats from within (employees, associates) weak passwords, unlocked devices intentional can be injecting rogue software

REST (cont)

Techniques for host-based user profiling on Unix and Windows: Markov chain codel; bayers factor to determine if transition is consistent (command A-> command B); windows measures “properties” which vote with weights wether an intrusion has occurred

Advantage of a hidden Markov model over an SVM for classi-fying command sequences: Markov model creates probab-ility of each transition; this can easily grow very big; pick a K that is small; svm can be very accurate but it does not address concept drift very well

honeypot: a computer security mechanism set tro detect deflect or counteract attempts at unauthorized use of info systems. Generally consists of data that appears legit with info but is isolated and monitored and blocks or analyses attackers