毛利研究室輪講資料　令和 4 年度 第 10 回　令和 4 年 12 月 16 日（金）

Expected Exploitability: Predicting the Development of Functional Vulnerability Exploits

山下 恭平

概要：本稿は 31st USENIX Security Symposium にて掲載された論文「Expected Exploitability: Predicting the Development of Functional Vulnerability Exploits」[1] の内容についてまとめたものである. 既存の脆弱性評価基準の分析を通じて得られる結果は, その脆弱性が悪用されることを予測するのには不十分であるため, ソフトウェア脆弱性の公開時の悪用可能性を評価することは困難である. さらに,「悪用できない」という評価は不確実性が高く, 悪用可能性の評価にはバイアスがかかっていることが問題として挙げられる. これらの問題を解決するために, 機能的なエクスプロイトが開発される可能性を経時的に反映する,Expected Exploitability（EE）と呼ばれる新しい指標を提案する.

1. はじめに

エクスプロイトがセキュリティに深刻な影響を与えた事例として,2017 年に世界中で大流行した WannaCry と NotPetya がある. これらが悪名高い成功を納めた原因として, 武器化されたエクスプロイトの使用が挙げられる. しかし, 武器となり得る脆弱性を利用するプログラムを開発する難易度が上がっていることから, 既知の脆弱性のうち 5%のみを悪用することに注力するようになっている. そういった中で, 着目すべき脆弱性の優先順位をつけることで人々に対して最適な意思決定をもたらし, 悪用防止に向けた研究機会の深い理解のために, 各脆弱性の悪用可能性を評価する必要がある. しかし, 悪用可能性の評価は, どの脆弱性が, どのように利用されるかが不明なため, 困難である. 具体的には,WannaCry や NotPetya によって悪用された脆弱性である CVE-2017-0144 は, 当時の専門家が推奨するパッチから省かれたいた. このことから, エクスプロイトの開発によってエクスプロイト可能性を証明することはできるが, 非エクスプロイト可能性を証明することは困難である. この結果,「悪用不可能」という評価にはバイアスが発生し, 不確実性を持つことが分かる.

この問題を解決するために Expected Exploitability (EE) と呼ばれる新しい指標を提案する. この指標は, 脆弱性を「悪用可能」または「悪用不可能」と決定的に分類するのではなく, 類似の脆弱性に関する過去のパターンに基づいて, 機能的なエクスプロイトが開発される可能性を時系列で継続的に推定するものである. ここで機能的なエクスプロイトとは, 脆弱性が引き起こすセキュリティ上の問題を完全に実現し, 実際の攻撃を容易にするものである. 本稿では,2 章で研究の背景と目的について述べ,3 章では開発にあたっての課題について,4 章では実際に取集するデータについてまとめ, 5 章で EE の評価を行い,6 章でそれらをまとめる.

2. 研究の背景と目的

既存の評価についての問題

エクスプロイトの開発によって, 悪用可能性を証明することができるが, 悪用不可能であることを証明するのは困難である. その代わりに, 脆弱性悪用緩和の取り組みとして, 悪用の難しさを把握することを目的とした脆弱性スコアリングシステムがよく用いられる. 以下にその例をあげる.

- NVD CVSS[2]
  - 脆弱性を悪用することの容易さと技術的手段を反映することを目的とした, 悪用可能性の評価指標を持つスコアリングシステム. 必要なアクセス数, 攻撃の複雑さ, 権限レベルなど, 様々な脆弱性の特性を 0〜4 の数値に落とし込んだもの.
- Microsoft Exploitability Index[3]
  - Microsoft が 0〜3 の 4 段階で悪用可能性を評価し割り当てる,Microsoft 固有のスコア.
- RedHat Severity[4]
  - CVSS を補完し,RedHat 製品に影響を与える脆弱性について専門家の評価によって, 同様に脆弱性の悪用の難易度を評価したもの.

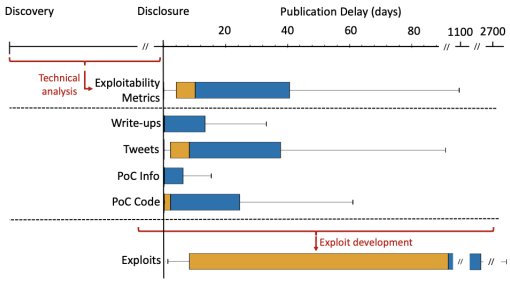

図 1　脆弱性情報のタイムライン

しかし, これらのベンダーなどから提供される指標は, 不正確であることが報告されている. 具体的な事例として,Internet Explorer の悪用可能な脆弱性である CVE-2018-8174 は,CVSS 悪用可能性スコア 1.6 を獲得し, 脆弱性スコアの 91%以下に位置づけられた. 同様に,Windows7 から 10 に影響を及ぼす悪用をされる脆弱性である CVE-2018-8440 は, スコアが 1.8 とされた. これらの指標が悪用可能性を適切に反映できない原因を説明するために, 図 1 に典型的な脆弱性のタイムラインを示す. これらの指標は, 脆弱性が公開される前に行われる技術的分析によって定める. しかし, 脆弱性の公開後, 脆弱性に関する追加の技術情報や PoC などの様々な脆弱性の成果物が公開され, それらについて SNS などで議論が行われることが観察された. これらの成果物から悪用の可能性についての有益な情報が得られることがよく発生する.CVE-2018-8174 は技術的な Write-up の公開がエクスプロイトの開発の直接的な原因になったと報告されており, CVE-2018-8440 の PoC は 2 日以内に悪用を引き起こすと判断されている. これらの例は, 既存の指標が, 公開後にのみ利用可能な有用なエクスプロイト情報を考慮できておらず, 時間の経過と共に更新されていないことを明らかにした.

研究の目的

このことから, 悪用可能性の時間的な変化は, 確立的なプロセスで記述できることが示唆される. ある時点において, 悪用可能性は悪用を観測する確率を符号化した確率変数 E と考えられ, 悪用が不可能とされている脆弱性には確率 0 を, 悪用が確認されている脆弱性には 1 を割り当てる. しかし,E を生成する真の分布 E は利用不可能であり, 実際はノイズを含めた E^{train} を使用する必要がある. これは, 悪用可能性の期待値を推定する尤度を計算することによって, 利用可能なデータから E を近似する必要があることを意味する. この指標のことを Expected Exploitability (EE) と呼ぶ. EE は教師あり機械学習を用いて過去のデータから学習することができ, 新しいエクスプロイトが開発, 発見される前に, 脆弱性に対する悪用の可能性を評価することを可能にする. この研究の目的は, 既知の脆弱性に対して機能するエクスプロイトが開発されるかどうかを予測することで, その脆弱性の悪用可能性を客観的に定量化することである.

3.　EE 開発の課題

EE に教師あり機械学習の技術を適用するためには 3 つの課題がある. ここでは, それぞれについて説明する.

PoC から特徴を抽出する

PoC は脆弱性のトリガーとして設計されており, 直接的な攻撃は行わない. しかし, 機能的なエクスプロイトにおいて必要なステップを満たしている. つまり,PoC コードの構造と複雑さは, 脆弱性攻撃の難しさを直接反映すると考えられる. PoC の持つ予測力を十分に活用するために,PoC の特徴を抽出することが求められる. しかし,PoC は様々なプログラム言語で書かれており, コードと事由形式のテキストを組み合わせた物であることが多く, 既存のプログラム解析技術の適応が制限される. そのため,PoC の特徴抽出には, テキストとコードを分離し, 有用なコード表現を得るための新しい技術を必要とする.

ノイズの把握と軽減

先行研究 [14] によって, 学習に利用できるラベルに偏りがあることが判明している. この問題を機械学習におけるラベルノイズの問題と関連付ける試みはこれまで為されていない. さらに, エクスプロイトの証拠を提供する個々のベンダーが, 脆弱性のカバー率にばらつきがある. このような特徴に依存するノイズの問題はあまり研究されておらず, 実世界のアプリケーションにおけるノイズの特徴を発見することは, 機械学習における未解決の問題と考えられる. このため, ラベルノイズの種類とその影響, およびそれに対処するための学習技術の設計が求められる.

時間的に変化する悪用可能性の評価

脆弱性の公開後に出現する成果物は分類を向上させる可能性があるが, 公開の遅れはタイムリーな予測の有用性に影響を与える. そのため,EE の評価では, リアルタイム性と潜在性という両立困難な指標を使用する必要がある.

4.　収集するデータ

機械学習に必要な特徴量を抽出するために, 必要なデータを集める. 以下は集めるデータの一覧である.

CVE ID

CVE ID は最も広く普及し, 相互参照されている公的な脆弱性識別システムの 1 つであるため, 脆弱性を識別するために CVE ID を使用する.

公開の脆弱性情報

PoC がターゲットとする脆弱性に関する情報を,National Vulnerability Database(NVD)[5] から収集する．NVD には, アナリストが収集した脆弱性情報が掲載されており, 高度な技術的情報を得ることができる. 各脆弱性に対して利用可能な技術情報をより多く把握するために, いくつかの公開情報の参照も行う. 以下に参照を行う公開情報の一覧を示す.

- Bugtraq[6]
- IBM X- Force Exchang[7]
- Vulners[8]

これらから 278,297 の文書、102,936 の脆弱性を参照した. これらの文書は脆弱性に関して公開されている技術情報の全体像を示すものであり,「write-up」と呼んでいる.

Proof of Concepts (PoCs)

ExploitDB[9] , Bugtraq[6] , Vulners[8] の 3 つのデータベースを参照し, 公開 PoC のデータ収集を行なった. 重複した収集データは除去し,CVE ID にリンクされている 48,709 件の PoC を対象とした. 結果 21,849 件の異なる脆弱性に対応した PoC の収集に成功した.

SNS の情報

2014 年 1 月から 2019 年 12 月までの間に CVE ID に言及したツイートの収集を行なった.Twitter Filtered Stream API[10] を用いて 52,551 件の脆弱性に関する 140 万のツイートを収集することに成功した.

開発済みのエクスプロイト情報

開発済のエクスプロイトに関する包括的なデータセットがないため, 複数の公開情報から証拠を集約した. 以下は集約に使用した公開情報一覧である.

- CVSS
- IBM X- Force Exchang[7]
- Bugtraq[6]
- Tenable Nessus[11]
- Skybox[12]
- AlienVault OTX[13]

5. 評価

既存の 2 種類の悪用可能性予測である EPSS[15] と Social Media Classifier(SMC) と比較を行ったところ. 全ての実験において精度を上回った. また, 時間によって精度が向上するのも確認された. また, 緊急性の高い脆弱性, 重要な脆弱性についての実験も行い, 結果, 既存の評価指標よりも高い精度を得ることができた. 新しい指標 EE は次の url で公開されており,CVE ID によって検索をかけ, 実際の指標を閲覧することが可能である.

`https://exploitability.app/`

6. おわりに

既存の脆弱性情報やエクスプロイトの情報など, 様々な情報を取得し機械学習によって学習させ, 悪用可能性を予測することに成功していた. 既存の指標は一度公開されると, 更新されないことを問題にあげ, リアルタイム性が極めて高いシステムを開発していた. また, 特徴量を的確に抽出するための数学的思考や SNS 文字列の分析など, 多岐にわたる知見が使用されており, 極めてレベルの高い研究だと感じた. 今後は, 学習データのノイズをさらに減らすことで更なる精度を期待できると考えられる.

参考文献

[1] Octavian Suciu , Connor Nelson , Zhuoer Lyu , Tiffany Bao , and Tudor Dumitras. Expected Exploitability: Predicting the Development of Functional Vulnerability Exploits. In 31th USENIX Security Symposium (USENIX Security 22), pages 377-394, 2022

[2] A complete guide to the common vulnerability scoring system. `https://www.first.org/cvss/v3.0/specification-document` . (15/12/2022)

[3] Microsoft exploitability index. Microsoft. `https://www.microsoft.com/en-us/msrc/exploitability-index` . (15/12/2022)

[4] Severity Rating. RedHat. `https://access.redhat.com/security/updates/classification/` . (15/12/2022)

[5] National vulnerability database. `https://nvd.nist.gov/` . (15/12/2022)

[6] Bugtraq. Accenture. `https://bugtraq.securityfocus.com/archive` . (15/12/2022)

[7] IBM X- Force Exchang. `https://exchange.xforce.ibmcloud.com/` . (15/12/2022)

[8] Vulners. Vulners vulnerability database. `https://vulners.com/` . (15/12/2022)

[9] ExploitDB. The exploit database. `https://www.exploit-db.com/` . (15/12/2022)

[10] Twitter. Filtered stream. `https://developer.twitter.com/en/docs/twitter-api/tweets/filtered-stream/introduction` . (15/12/2022)

[11] Tenable Network Security. Nessus vulnerability scanner. `https://www.tenable.com/products/nessus` . (15/12/2022)

[12] SkyBox. Vulnerability center. `https://www.vulnerabilitycenter.com/#home` . (15/12/2022)

[13] Alienvault otx. AlienVault. `https://otx.alienvault.com/` . (15/12/2022)

[14] M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker. Beyond heuristics: learning to classify vulnerabilities and predict exploits. In KDD, Washington, DC, Jul 2010.

[15] J. Jacobs, S. Romanosky, B. Edwards, I. Adjerid, and M. Roytman. Exploit prediction scoring system (epss).

Digital Threats: Research and Practice, 2(3), July 2021.