特集論文

ペネトレーションテスト自動化に向けたサイバー攻撃手段の定量的評価法

Cyber Attack Technique Scoring Method toward Automatic Penetration Test

酒井康行* *Yasuyuki Sakai*
河内清人‡ *Kiyoto Kawauchi*
木下洋輔† *Yosuke Kinoshita*
加藤　駿‡ *Shun Kato*
木藤圭亮‡ *Keisuke Kito*

*電子通信システム製作所(博士(工学))
†同製作所
‡情報技術総合研究所

要　旨

ペネトレーションテスト(以下“ペンテスト”という。)は，実際にサイバー攻撃を行うことでシステムの脆弱(ぜいじゃく)性を発見する，サイバー攻撃対策に有効な手法である。ペネトレーションテスター(以下“ペンテスター”という。)が適切な攻撃手段を選択するとき，攻撃手段が成立するか，発覚するか，攻撃で得られる効果の大きさ，の三点を基に判断していると仮定し，攻撃手段をスコア化するCATS(Cyber Attack Techniques Scoring)法を開発した。CATS法によって適切な攻撃手段の選択を自動化でき，ペンテストが自動実行可能となる。ペンテスト自動実行によって，システムの脆弱性発見とサイバー攻撃対策を迅速に行えるようになる。

1. は じ め に

サイバー攻撃対策にはシステムの脆弱性を積極的に発見し修正することが重要である。脆弱性発見手法として，実際にサイバー攻撃を試みるペンテストがある。NIST SP800-115[1]では，ペンテストを“アプリケーション，システム，ネットワークのセキュリティー機能を回避する方法を特定するために評価者が実際の攻撃を模擬して行うセキュリティーテスト”と定義する。ペンテストでは“適切な攻撃手段”の選択と実行を目的達成まで繰り返す。ここで適切な攻撃手段とは，攻撃効果が高く，攻撃成立見込みが高く，発覚しにくい手段のことである。適切な攻撃手段の選択にはペンテスターと呼ばれる専門家が必要である。

この研究では，ペンテスターが適切な攻撃手段を選択するとき，攻撃手段が成立するか，攻撃が発覚するか，攻撃の結果得られる効果が大きいか，の三点を基に判断していると仮定し，攻撃手段をeVc(Evaluation Value of Capture：攻略評価値)，eVd(Evaluation Value of Detectability：発覚評価値)，eVe(Evaluation Value of Effectiveness：効果評価値)としてスコア化する手法であるCATS法を開発した。CATS法の評価実験の結果，ペンテスターによる攻撃手段選択とCATS法による攻撃手段選択がほぼ一致した。また，対象システムに関する構成情報が十分に入手できていない場合でも，ペンテスターが選ぶ攻撃手段とほぼ一致した攻撃手段を選択することができた。これらの結果から，攻撃手段の選択を自動実行可能であることが分かった。今後は，eVc，eVd，eVeがそれぞれ表す攻撃の成立見込み，攻撃検知可能性，攻撃で得られる効果を統合的に取り扱う方式を導入し，より実際のサイバー攻撃に近いペンテストを模擬することを目指す。なお，参考文献[8]に，より詳細な研究内容を記したので参照されたい。

2. 従来のペンテスト自動化研究

従来研究A2P2V[2]は，攻撃ゴール達成までの手順をアタックツリー形式で出力する。アタックツリーのうち一つのパスを攻撃シナリオとして，対応する攻撃手段を実行する。課題は二点ある。第一は，システムに関する完全な情報が必要なことである。テスト対象の脆弱性情報等は実際は完全には入手できない場合が多い。構成情報が欠落している場合，成立しない攻撃手段を実行してしまうことがある。第二は，適切な攻撃手段選択が行えないことである。A2P2Vでは列挙した攻撃手段のうち，実行回数が少ない攻撃シナリオから実行される。攻撃成否見込みは考慮されているが，見つかりにくさや効果を考慮した攻撃手段選択が行われない。適切な攻撃手段は攻撃シナリオにも依存する。例えば見つかりにくさを優先するか，攻撃が検知されてもよいので攻撃の確実性を優先するかで選ぶべき攻撃手段が異なる。これらの攻撃手段選択はペンテスターの暗黙知によって行われる。そのため，従来はペンテストを自動化することは困難であった。

3. 提案手法CATS法

この章では，サイバー攻撃手段をスコア化するCATS法について述べる。本稿で使用する記号の定義を**表1**に示す。

表1 -記号の定義

記号	定義
A	実行可能な攻撃手段の集合
A_i	攻撃手段，$A_i \in$ A（iは自然数）
S	攻撃対象システム
K_S	攻撃者が持つSに関する知識
a_j	Sの属性（jは自然数）
$at_{a_j,k}$	属性a_jが取り得る値（kは自然数）
C_{A_i}	A_iが成功するために必要なa_jの組の集合
D_S	Sを監視する攻撃検知システム
E_{A_i}	A_iによってS上で発生する事象の集合

3.1 ペンテスターのモデル化

ペンテスターは次の考え方で攻撃手段を選択すると仮定する。

(1) 実行したい攻撃手段が攻撃対象で成立するか

(2) 攻撃手段実行時に攻撃対象で攻撃が検知されるか

(3) 攻撃手段が成功したときに得られる効果が高いか

(1)は，対象がWindows(注1)である場合，ペンテスターはWindowsの脆弱性を突くExploitを選択することなどを意味する。攻撃手段成功見込みのスコアを攻略評価値eVcとする。

(2)は，ペンテスターが攻撃手段を実行するとき，ネットワーク上のトラフィックがIDS(Instrusion Detection System)などで検知され得ることを考えて，より検知されにくい攻撃手段を選択することなど意味する。この考え方に基づくスコアを発覚評価値eVdとする。

(3)は，ペンテスターが遠隔でファイル閲覧できる攻撃手段よりも遠隔でコード実行できる攻撃手段の方が攻撃効果が高いと考えて，より攻撃効果が高い攻撃手段を選択することなどを意味する。この考え方に基づくスコアを攻撃効果値eVeとする。

（注1） Windowsは，Microsoft Corp.の登録商標である。

3.2 スコア化方法

(1) 定義1：攻撃対象システムSに関する知識K_S

攻撃者が攻撃対象システムSの構成情報に関して持つ知識K_Sを，$K_S = (f_{\text{estimated}}, \mathrm{I}_S)$と定義する。ここで$f_{\text{estimated}}$は，$S$の属性値に対する確率推定する関数である。また$\mathrm{I}_S$は攻撃者が取得済みの$S$に関する知識の集合である。

(2) 定義2：攻略評価値eVc

eVcを，知識K_Sを持つ攻撃者が，ある攻撃手段A_iを攻撃対象システムSに対して行うとき，A_iが成功する確率と定義する。eVcは次式で与えられる。

$$eVc_{A_i,S,K_S} = P(\mathrm{C}_{A_i}; K_S) \cdot P(A_i | \mathrm{C}_{A_i}; K_S)$$

ここで$P(\mathrm{C}_{A_i}; K_S)$は，Sの知識K_Sを持つ攻撃者が，A_iが成功するための各属性a_jの組の集合C_{A_i}をSが全て満たしている確率である。また$P(A_i | \mathrm{C}_{A_i}; K_S)$は，ある$A_i$が成功するための各属性$a_j$の組の集合$\mathrm{C}_{A_i}$が成立していると仮定したとき，知識$K_S$を持つ攻撃者による$A_i$が成功する確率である。

(3) 定義3：発覚評価値eVd

攻撃手段A_iによって攻撃対象システムS上で発生する事象の集合をE_{A_i}とする。D_SをS上で発生する攻撃を検知するシステムとして，d_{D_S}をE_{A_i}に入力しD_Sによって攻撃が検知される確率を出力する関数とする。Sが検知システムD_Sに監視さ

れているとき，攻撃手段A_iがD_Sによって検知される確率として発覚評価値eVdを次式のように定義する。

$$eVd = d_{D_S}(\mathrm{E}_{A_i})$$

例えば，E_{A_i}は攻撃手段実行後に発生するネットワーク通信や，攻撃対象システムで記録されるログなどを意味する。

⑷　定義４：効果評価値eVe

Sに対して，攻撃手段A_iを実施することで獲得できる効果eVe_{S,A_i}を次式のように定義する。

$$eVe_{S,A_i} = (\text{CVSS影響度}(A_i), \text{CVSS環境値}(S))$$

eVeではCVSS(Common Vulnerability Scoring System) v3[3]を使用する。CVSS影響度は攻撃手段A_iが成功したときに攻撃者がSに与える影響を表して，CVSSの影響度算出法を用いる。CVSS環境値は攻撃手段をSに対して実行したときの影響度を表してCVSS環境値の算出法を用いる。

4．CATS法の評価

4.1　評 価 方 法

CATS法の評価は，次の観点で行う。

⑴　攻撃手段A_iを正しくスコア化できるか

⑵　スコア値で選ぶ攻撃手段とペンテスターが選ぶ攻撃手段が同じか

⑶　知識K_Sが不足する場合もスコア化できるか

評価は次のステップで行う。

S1：評価条件の設定　S，K_S，D_S，評価対象とするA_iを定める。

S2：CATS法によるスコア化　S1で設定したS，K_S，D_Sの条件下でeVcとeVdをA_iごとに算出し，スコア値の高い順で順位を導出する。

S2'：ペンテスターによる順位付け　S1で設定したS，K_S，D_Sの条件下で，評価対象の攻撃手段をペンテスターがeVcとeVdの観点で総当たりで評価し，優れている攻撃手段の順に順位を導出する。

S3：比較　S2とS2'の結果を比較する。

4.2　S1：評価条件の設定

MITRE ATT&CK Enterprise[4]のTacticsの一つである，Lateral Movement(以下"LM"という。)に属するTechniquesを評価対象の攻撃手段とした。LMには攻撃手段が多数ありLMの攻撃手段を適切に選択できれば他のTacticsの攻撃手段も適切に選択できる。実際の攻撃手段として，ペンテストツールのデファクトであるMetasploit[5](注2)から，LMのTechniquesに相当するモジュールを選択した。Sは一般的なIT環境として，二つのS_1，S_2を考える。S_1，S_2の詳細を**表２**に示す。

表２−攻撃対象システムSの設定

S	Sの詳細(a_1＝"OS"，a_2＝"開放ポート"，a_3＝"脆弱性"，a_4＝"稼働サービス"とする)
S_1	$at_{a_1,1}$＝Windows10 1909 $at_{a_2,1}$＝445，$at_{a_2,2}$＝3389，$at_{a_2,3}$＝5900，$at_{a_2,4}$＝5985，$at_{a_2,5}$＝135，$at_{a_2,6}$＝8030 $at_{a_3,1}$＝CVE−2020−0796，$at_{a_3,2}$＝CVE−2009−3028，$at_{a_3,3}$＝CVE−2009−4655 $at_{a_4,1}$＝SMB 3.1.1，$at_{a_4,2}$＝Altiris 6.9.x，$at_{a_4,3}$＝eDirectory 8.8.5，$at_{a_4,4}$＝VNC
S_2	$at_{a_1,1}$＝linux(注3) $at_{a_2,1}$＝445，$at_{a_2,2}$＝22，$at_{a_2,3}$＝5901，$at_{a_2,4}$＝8888，$at_{a_2,5}$＝8030 $at_{a_3,1}$＝CVE−2021−25646，$at_{a_3,2}$＝CVE−2009−4655 $at_{a_4,1}$＝samba(注4)，$at_{a_4,2}$＝SSH，$at_{a_4,3}$＝Apache(注5) Druid 0.20，$at_{a_4,4}$＝eDirectory 8.8.5，$at_{a_4,5}$＝VNC

（注３）　Linuxは，Linus Torvalds氏の登録商標である。
（注４）　Sambaは，Software Freedom Conservancy, Inc.の登録商標である。
（注５）　Apacheは，The Apache Software Foundationの登録商標である。

攻撃者が持つK_Sは，ネットワークスキャンなどの偵察行為を段階的に行うことを想定し，次の４段階を対象にした。

⑴　何も知識を持っていない

⑵　ネットワークスキャナーで得られる情報に相当する知識

⑶　脆弱性スキャナーで得られる情報に相当する知識

⑷　⑶に加えて，ドメイン管理者の認証情報

⑴はSに関する情報がない状態である。攻撃者は客観的な情報を基にK_Sを定める。Sの属性値に対する確率推定に$f_{\text{estimated}}$を使用する。例えば開放ポート番号推定にはOSの確率分布とデフォルトの空きポート番号を基に特定のポート番号が開いている確率が計算できる。

⑵はネットワークスキャナー(6)で得られる情報の知識を持つ状態である。例えばIP(Internet Protocol)アドレス，動作ソフトウエアのバージョン，OS情報である。

⑶はNessus(7)(注6)などの脆弱性スキャナーで得られる情報に相当する知識とCVE番号のような具体的脆弱性情報を持つ状態である。

⑷は⑶に加えてドメイン管理者の認証情報を持っている状態である。評価では何も知識がない状態K_0と，S_1，S_2それぞれに対して３段階の知識を持つ状態，計７種類をK_Sの条件とした。K_Sの詳細を**表３**に示す。

表３-知識K_Sの設定

知識K_S	知識K_Sの詳細(a_1＝“OS”，a_2＝“開放ポート”，a_3＝“脆弱性”，a_4＝“稼働サービス”とする)
K_S0	知識なし
$K_{S1}1$	$at_{a_1,1}$＝Windows $at_{a_2,1}$＝445，$at_{a_2,2}$＝3389，$at_{a_2,3}$＝5900，$at_{a_2,4}$＝5985，$at_{a_2,5}$＝135，$at_{a_2,6}$＝8030 $at_{a_3,1}$＝SMB 3.1.1，$at_{a_3,2}$＝eDirectory，$at_{a_3,3}$＝VNC 保有クレデンシャル＝一般ユーザー
$K_{S1}2$	$at_{a_1,1}$＝Windows10 1909 $at_{a_2,1}$＝445，$at_{a_2,2}$＝3389，$at_{a_2,3}$＝5900，$at_{a_2,4}$＝8888，$at_{a_2,5}$＝8030 $at_{a_3,1}$＝CVE-2020-0796，$at_{a_3,2}$＝CVE-2009-3028，$at_{a_3,3}$＝CVE-2009-4655 $at_{a_4,1}$＝SMB 3.1.1，$at_{a_4,2}$＝Altiris 6.9.x，$at_{a_4,3}$＝eDirectory 8.8.5，$at_{a_4,4}$＝VNC 保有クレデンシャル＝一般ユーザー
$K_{S1}3$	$at_{a_1,1}$＝Windows $at_{a_2,1}$＝445，$at_{a_2,2}$＝3389，$at_{a_2,3}$＝5900，$at_{a_2,4}$＝5985，$at_{a_2,5}$＝135，$at_{a_2,6}$＝8030 $at_{a_4,1}$＝SMB 3.1.1，$at_{a_4,2}$＝eDirectory，$at_{a_4,3}$＝VNC 保有クレデンシャル＝ドメイン管理ユーザー
$K_{S2}1$	$at_{a_1,1}$＝linux $at_{a_2,1}$＝445，$at_{a_2,2}$＝22，$at_{a_2,3}$＝5901，$at_{a_2,4}$＝8888，$at_{a_2,5}$＝8030 $at_{a_4,1}$＝samba，$at_{a_4,2}$＝SSH，$at_{a_4,3}$＝Apache Druid，$at_{a_4,4}$＝eDirectory，$at_{a_4,5}$＝VNC 保有クレデンシャル＝一般ユーザー
$K_{S2}2$	$at_{a_1,1}$＝linux $at_{a_2,1}$＝445，$at_{a_2,2}$＝22，$at_{a_2,3}$＝5901，$at_{a_2,4}$＝8888，$at_{a_2,5}$＝8030 $at_{a_3,1}$＝CVE-2021-25646，$at_{a_3,2}$＝CVE-2009-4655 $at_{a_4,1}$＝samba，$at_{a_4,2}$＝SSH，$at_{a_4,3}$＝Apache Druid 0.20，$at_{a_4,4}$＝eDirectory，$at_{a_4,5}$＝VNC
$K_{S2}3$	$at_{a_1,1}$＝linux $at_{a_2,1}$＝445，$at_{a_2,2}$＝22，$at_{a_2,3}$＝5901，$at_{a_2,4}$＝8888，$at_{a_2,5}$＝8030 $at_{a_4,1}$＝samba，$at_{a_4,2}$＝SSH，$at_{a_4,3}$＝Apache Druid，$at_{a_4,4}$＝eDirectory，$at_{a_4,5}$＝VNC 保有クレデンシャル＝ドメイン管理ユーザー

Sを監視するD_Sも一般的IT環境に導入されているものとして，D_Sを次の３段階に設定する。

①ウイルス対策ソフトウエア：ウイルス対策ソフトウエアだけをSに導入する。シグネチャーにマッチするファイルがSに書き込まれたとき検知される。

②ウイルス対策ソフトウエア，IDS(①)に加えてIDSを導入する。シグネチャーにマッチする攻撃通信が確認された場合に検知される。

③ウイルス対策ソフトウエア，IDS，EDR(Endpoint Detection and Response)(②)に加えてEDRをSに導入する。EDRはS上の挙動が正常から逸脱する場合に検知される。

(注２)　Metasploitは，Rapid7 LLCの登録商標である。
(注６)　Nessusは，Tenable, Inc.の登録商標である。

4.3　S2：CATS法によるスコア化と，S2’：ペンテスターによる順位付け

S1で設定した前提条件を基に，CATS法による攻撃手段A_iのスコア化を行う。今回の評価ではeVc及びeVdについてスコア化を実施した。S2と独立してS1で設定した前提条件を基に，ペンテスターによる攻撃手段の順位付けを行った。今回eVcの観点，eVdの観点での順位をそれぞれ２名のペンテスターによって評価した。（スコア値は参考文献(8)を参照）

5. む　す　び

評価結果をまとめる。評価の結果，eVcの観点では，A_{11}及び知識$K_{S2}3$，K_S0の場合を除いて上位５位は同じA_iであった。したがってeVcのスコア値に基づいて攻撃手段A_iを選択すれば，実際に攻撃が成功しやすい攻撃手段を選択できる。またネットワークスキャン結果相当の知識である$K_{S1}1$，$K_{S2}1$に着目すると，CATS法のスコアとペンテスターの順位付けは１～５番目までの攻撃手段は一致した。Sの知識K_Sが不足しても正しくスコア化できることが分かった。また，Sの情報が全く得られていない場合でも適切にスコア化できることも分かった。これらの結果から，攻撃手段の選択を自動実行可能であることが分かった。

参 考 文 献

(1) NIST, SP800-115 Technical Guide to Information Security Testing and Assessment (2008)
https://csrc.nist.gov/pubs/sp/800/115/final
(2) Nakanishi, F., et al.：Automated Attack Path Planning and Validation (A2P2V), Black Hat USA 2021 Arsenal (2021)
(3) 独立行政法人 情報処理推進機構 セキュリティセンター：共通脆弱性評価システムCVSS v3概説 (2022)
https://www.ipa.go.jp/security/vuln/CVSSv3.html
(4) MITRE：ATT&CK
https://attack.mitre.org
(5) RAPID7：metasploit
https://www.metasploit.com
(6) NMAP.ORG：NMAP Free Security Scanner
https://nmap.org
(7) tenable：Tenable Nessus
https://jp.tenable.com/products/nessus
(8) 木藤圭亮，ほか：ペネトレーションテスト自動化に向けたサイバー攻撃手段の定量的評価法の提案，2022年暗号と情報セキュリティシンポジウム(SCIS2022), 2B3-2 (2022)