6ZD-05

攻撃抑止効果を狙ったアクティブ・ディフェンス実現についての一考察

戸川來星　寺田真敏
東京電機大学

1. はじめに

警察庁は，サイバー攻撃に関わる悪性のアクセス，サイバー犯罪の検挙数が増加していることなどを受けて，サイバー空間における脅威は極めて深刻な状況が続いていると報告している[1]．このようなサイバー攻撃が止まない理由の 1 つとして，攻撃者は優位，被攻撃者は劣位という非対称性の問題がある．非対称性への対策として，従来からある受動的な対策だけではなく，アクティブ・ディフェンスといった積極的防御や攻撃を対策に活かすオフェンシブ・セキュリティというアプローチが導入され始めている．

本研究は，偽装環境を使って攻撃者を惑わす，攻撃者の情報を収集して先んじて対処する，などのアクティブ・ディフェンスというアプローチをサイバー攻撃対策に活用していくことを目的としている．本稿では，アクティブ・ディフェンスについて関連研究や関連文献を調査し，結果をもとにアクティブ・ディフェンスのアプローチを整理した．さらに，分類整理のなかで Web サイトへの攻撃抑止効果を狙ったプロトタイピングを通して，アクティブ・ディフェンスの実現方法について考察した結果を報告する．

2. 関連研究

2.1 解決したい課題

アクティブ・ディフェンスは元々軍事用語として，アメリカ国防総省が「限定的な攻撃や反撃を用いて，敵対する相手の地域や陣地を無力化すること」[2]と定義している．このため，アクティブ・ディフェンスに反撃・逆襲の意味を含む場合がある．その一方，サイバー空間における規範についても議論されており，国際連合の下に設置された IGF (Internet Governance Forum) は，2018 年にサイバー空間の信頼性と安全性のための Paris Call を発表し[3]，このなかでサイバー傭兵や非国家主体による攻撃を容認しないものとした．

本研究では，非対称性への対策としてアクティブ・ディフェンスを活用していくためには，取り組まれている研究分野を明らかにすることを通して，アクティブ・ディフェンスが取り扱う範囲と，その実現性を示すことが必要あると考え，課題とした．

2.2 アクティブ・ディフェンス

本節では，アクティブ・ディフェンスとして取り組まれている研究分野を整理する．整理にあたっては，情報戦に対抗するためにアメリカ軍によって定義されたインフォメーション・オペレーションズ[4]の区分を用いた．表 1 にインフォメーション・オペレーションズを元に調査した関連研究 31 件の分類結果を示す．

(1) **Degrade (低下)**

Degrade (低下) では，システム内部の攻撃対象となりうる箇所に対して，その場所を動的に変更することで，攻撃者に対抗する MTD(Moving Target Defense)という技術分野における提案が多く見られた．攻撃者の標的を見失わせたり，攻撃者のリソースを大幅に消費させたり，攻撃の不確実性を高めたりすることで，防御側が劣位という非対称性の是正を狙っている．文献[5]では，リバースプロキシの IP アドレスを動的に変更することで，攻撃者に攻撃の的を絞らせず，攻撃者のリソースを大幅に消費させるアプローチを，深層強化学習ベースで提案し，その効果を検証している．

表 1: アクティブ・ディフェンスの研究分野

区分	関連研究(重複有)
Destroy (破壊)	0
Disrupt (中断)	0
Degrade (低下)	5[5]
Deny (拒否)	0
Deceive (欺瞞)	10[6]
Exploit (搾取)	0
Influence (影響)	1[7]
Protection (防御)	11[5][6]
Detection (検知)	22[6][8]
Restoration (回復)	6[8]
Response (対応)	20[5][6][8]

(2) **Deceive (欺瞞)**

Deceive (欺瞞) では，攻撃者を偽装環境であるハニーポットを活用する手法の提案が多く見られた．リダイレクトした偽装環境を本物と思わせ，攻撃者のリソースを消費させたり，攻撃者を混乱させたりすることを狙っている．文献[6]では，電力供給の IoT デバイスにおいて，攻撃者を偽装環境であるハニーネットにリダイレクトして，攻撃プロセスを遅らせるとともに，攻撃行動を記録することで，攻撃に対抗するシステムを提案している．

(3) **Influence (影響)**

Influence (影響) に区分した文献[7]では，冷戦期の核戦略と結びついて発展してきた抑止が，サイバー攻撃対策の 1 つとして注目されていることから，米国でのサイバー抑止政策を分析した結果について報告している．

(4) **Restoration (回復)**

Restoration (回復) では，攻撃行動を記録して回復を目指したり，攻撃にリアルタイムに対応しながら回復を試みたりする提案などが見られた．文献[8]では，システムが処理しきれないほどの大量のリクエストを送信することでサービスの停止を狙う DoS 攻撃に対して，ゲーム理論を活用した対処を検討，および提案し，その効果を検証している．

3. 抑止を想定したアクティブ・ディフェンス

本章では，関連研究での分類整理に基づき，サイバー攻撃対策の 1 つとして注目されている抑止について，Web サイトへの適用検討とプロトタイピングについて述べる．なお, 表 1 において, 相手に行動を思いとどまらせる抑止は，敵対者の意思決定に影響を与え，こちらに有利な行動を取らせる Influence (影響) に区分した．

3.1 Web サイトにおけるサイバー抑止(Cyber Deterrence)

抑止は，「敵の意思決定に決定的な影響を与える手段によって，敵が悪意ある行動を取らないよう説得するもの」[9]，「受け入れがたい対抗措置を取るという信頼性のある威嚇

Considerations on the Active Defense Approach for Effect of Cyber Attack Deterrence
RAIKI TOGAWA and MASATO TERADA, Tokyo Denki University

か，ある行動を取るコストが予期される利得を上回るとの考えによって，その行動を防止すること」[10]と定義されており，サイバー攻撃対策の 1 つとして検討されている．抑止を実現するための要件については，文献[7]で帰属(attribution)，伝達(signaling)，信頼性(credibility)の 3 つが示されている．

Web サイトを対象に抑止を想定したアクティブ・ディフェンスの実現例としては，アクセス者に対するダイアログによる警告[11]や，不正アクセス監視での発信元 IP アドレスの開示[12]などがある．本稿では，アクティブ・ディフェンスをサイバー攻撃対策に活用していくため，これまでの手法を一歩進め，Web サイトで発生している不正アクセスの状況を Web サイト自身が開示する手法を検討し，実現のための要件を帰属，伝達，信頼性の観点からまとめた．

手法：

攻撃者に対しては攻撃を検知していることを示し，周辺利用者に対しては攻撃の情報を広報することで抑止を図る．

要件 1：帰属

攻撃者を追跡できる情報を公開する．

要件 2：伝達

Web サイトを攻撃した場合，その攻撃の情報がリアルタイムで公開され，周辺利用者に知られることになると攻撃者が理解できるようにする．

要件 3：信頼性

他の攻撃者による攻撃の情報が公開されている様子を確認できるようにすることで，こちらが実際に情報の公開を履行する意思，および攻撃を検知する能力を有していることを攻撃者が信じられるようにする．

3.2 プロトタイプ

本節では，要件に基づき実装したプロトタイプについて述べる．

(1) 抑止実現のための方針

Web サイトが受けた攻撃の情報を自ら公開することで[要件 2，要件 3]，攻撃者の意思決定に対して，攻撃が検知されていること，攻撃者の情報や攻撃の手法[要件 1]が周辺利用者に知られること，分析されること[要件 2]を通して影響を与え，攻撃を抑止する．

(2) 機能面の特徴

要件に基づき実装したプロトタイプの特徴は，次の通りである．

- Web サイトのアクセスログから，攻撃日時，攻撃元 IP アドレス，攻撃手法の 3 点を攻撃情報として抽出し，公開する．[要件 1，要件 2，要件 3](図 1)
- 攻撃情報の抽出には，既存ツールの iLogScanner[13]を活用する．[要件 3]
- 利用者が条件を指定することにより，攻撃日時，攻撃元 IP アドレス，攻撃手法などでグルーピングして情報を取得できるようにする．[要件 1，要件 2]

図 1: 攻撃情報を公開する Web サイト

3.3 考察

(1) 帰属

Web サイトが受けた攻撃情報を公開するため，より現実感や強制力をもって攻撃元に働きかけられるようになる可能性がある．特に，踏み台の追跡調査においては，踏み台として悪用されていることの認知や，さらに上流へとつなげていくことを通して，攻撃者の特定に寄与できることが考えられる．

(2) 伝達

攻撃情報を公開する Web サイトで，攻撃検知した件数や，踏み台として悪用されていることの認知につながった件数を数字やレポートなどで示したり，グラフなどを用いて可視化したりすることで，情報が効果的に伝わりやすくなると考える．

(3) 信頼性

攻撃情報の抽出に既存ツールの iLogScanner を活用していると開示することで，実装の透明性，抽出手法の信頼性などを，より高められることが見込める．

4. おわりに

本稿では，アクティブ・ディフェンスをサイバー攻撃対策に活用していくことを念頭に置きながら，サイバー空間におけるアクティブ・ディフェンスのアプローチを整理した．さらに，Web サイトへの攻撃情報を公開することで攻撃を抑止することを意図したプロトタイプ開発をおこなった．

プロトタイプについては，今後実環境で運用することを通して，有効性などを検証するとともに，アクティブ・ディフェンスを活用した，サイバー攻撃における攻撃者と被攻撃者の非対称性の是正に向けた対策を検討していきたい．

参考文献

[1] 警察庁，令和４年上半期におけるサイバー空間をめぐる脅威の情勢等について，https://www.npa.go.jp/publications/statistics/cybersecurity/data/R04_kami_cyber_jousei.pdf.
[2] DOD Dictionary of Military and Associated Terms, https://irp.fas.org/doddir/dod/dictionary.pdf
[3] Paris Call, https://pariscall.international/en/
[4] Information Operations: Doctrine, Tactics, Techniques, and Procedures, https://irp.fas.org/doddir/army/fm3-13-2003.pdf
[5] Xinzhong Chai; Yasen Wang; Chuanxu Yan; Yuan Zhao; Wenlong Chen; Xiaolei Wang, DQ-MOTAG: Deep Reinforcement Learning-based Moving Target Defense Against DDoS Attacks, https://ieeexplore.ieee.org/document/9172847
[6] Feng Li; Kunsan Zhang; Shuting Chen; Huiting Yang; Bing Wang, Research on Key Technologies of Active Defense for Distribution Internet of Things Service Security, https://ieeexplore.ieee.org/document/9277037
[7] 栗田 真広，サイバー攻撃に対する「抑止」の現状，https://dl.ndl.go.jp/view/download/digidepo_9104304_po_20140210.pdf?contentNo=1
[8] Chengwei Wu; Ligang Wu; Jianxing Liu; Zhong-Ping Jiang, Active Defense-Based Resilient Sliding Mode Control Under Denial-of-Service Attacks, https://ieeexplore.ieee.org/document/8716684
[9] U.S. Strategic Cyber Deterrence Options, https://centaur.reading.ac.uk/79976/1/22839264_Jasper_thesis.pdf
[10] Joint Chiefs of Staff, Joint Publication 3-0: Joint Operation, 22 October 2018, https://irp.fas.org/doddir/dod/jp3_0.pdf
[11] DOD SAFE, https://safe.apps.mil/
[12] Internet Storm Center, https://isc.sans.edu/data/
[13] IPA，ウェブサイトの攻撃兆候検出ツール iLogScanner, https://www.ipa.go.jp/security/vuln/iLogScanner/index.html.