SCIS 2016 2016 Symposium on Cryptography and Information Security
Kumamoto, Japan, Jan. 19- 22, 2016
The Institute of Electronics, Information and Communication Engineers

人間行動を用いた自律進化型防御システムの提案
Proposal of the Autonomous Evolution of Defense System using the Human Behavior

仲小路 博史*† 藤井 康広* 磯部 義明* 重本 倫宏* 鬼頭 哲郎*
林 直樹* 川口 信隆* 下間 直樹* 菊池 浩明†

Hirofumi Nakakoji Yasuhiro Fujii Yoshiaki Isobe Tomohiro Shigemoto Tetsuro Kito
Naoki Hayashi Nobutaka Kawaguchi Naoki Shimotsuma Hiroaki Kikuchi

あらまし 高度化・巧妙化するサイバー攻撃による被害が年々増加しており，アンチウィルスやファイヤウォール等の入口対策のみで標的型攻撃の被害を防止することは困難になってきている．著者らはマルウェアの侵入および感染を前提として，プロキシに認証を追加することによりマルウェアによる外部サーバとの通信を制御して実害の発生を抑えるとともに，認証を追加する条件を日々のセキュリティの運用を通して自律的に最適化する自律進化型防御システム（AED）の検討を進めている．本稿では，CAPTCHA 認証を用いて不審な URL への接続を制御するリスクベースプロキシ制御技術と，ユーザの認証履歴を機械学習することで認証の追加条件を最適化する認証条件最適化技術とを提案し，一部機能の実装および評価結果を報告する．

キーワード マルウェア，プロキシ，CAPTCHA 認証，機械学習，AED

1 はじめに

近年のサイバー攻撃の目的は，自己顕示欲の誇示から金銭搾取や政治的活動，諜報活動に変化している．これに伴い，犯行に加担する主体も単独から組織あるいは水平分業化した連合体へと変化してきている．また，攻撃に用いられる手法もゼロデイ攻撃や水飲み場攻撃等に巧妙化しており，金融機関や政府機関，制御システム等の重要インフラを狙った標的型攻撃が多発している．このような状況もあって，マルウェアの侵入を全て検知あるいは防止することは不可能[1][2]であるため，マルウェアの侵入，感染を前提とした多層的な防御が重要といわれている[3]．このような巧妙化，組織化する攻撃に対し，単独組織による対策も限界にきている．そのため，組織間でインテリジェンス（脅威情報や IT 機器の脆弱性情報およびそれらに関する分析や対処支援情報）を共有して攻撃に備える集団防御の概念が浸透してきた．集団防御を実現するべく，ISAC[4][5]のような公益法人が業界毎にインテリジェンスの共有を進めてリスクの軽減を図っている．また公益法人だけではなく，FireEye[6]，Threat Connect[7]等の民間企業がインテリジェンス共有サービスを開始している．しかし，インテリジェンス共有の仕組みは整いつつあるもののインテリジェンスの活用が進んでいないのが実態である．例えば，2015 年 6 月に日本年金機構が標的型攻撃を受けて 125 万件もの個人情報が漏えいしたが[8]，これを端緒として短期間に同様のマルウェアによって東京商工会議所や早稲田大学等，計 44 もの組織の情報漏えい被害が発生した[9]．この標的型攻撃では EMDIVI [10]と呼ばれるマルウェアが用いられていたが，そのマルウェアの特性や対処方法等のインテリジェンスが適切に共有されて且つ対策に迅速に活用されていれば，これらの被害の発生は抑えられた．これら 44 件の漏えい事故は小規模であったためあまり話題に上がらなかったが，大規模漏えい事故が頻発する蓋然性は極めて高いといえる．

著者らはこのような状況に鑑み，共有されたインテリジェンスを活用することでサイバー攻撃に対して集団防御を実現する自律進化型防御システム（AED：Autonomous Evolution of Defense）の研究を進めている．本技術は

* 株式会社日立製作所，神奈川県横浜市戸塚区吉田町 292 番地，Hitachi, Ltd., 292, Yoshida-cho, Totsuka-ku Yokohama-shi, Kanagawa
† 明治大学，東京都中野区中野 4-21-1，Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo

信頼関係のない他の組織から共有された不確実なインテリジェンスであっても，本来業務への悪影響を最小限に抑えつつ，対策に活用できるようにするものである．これにより，共有されたインテリジェンスに基づくシームレスな対策を実現し，EMDIVI で二の舞を演じた同様の被害の発生を未然に防ぐ．

本稿では 2 章で AED の概念および関連研究について紹介する．3 章で AED の実装および評価結果について述べ，4 章でまとめる．

2　自律進化型防御システムの提案

2.1　研究の背景と課題

近年の標的型攻撃はサイバーキルチェーン[11]と呼ばれる「偵察（Reconnaissance)」「武器化（Weaponization)」「配送（Delivery)」「エクスプロイト」(Exploitation)「インストール（Install)」「遠隔操作（Command & Control)」「目的実行（Actions on Objectives)」の 7 つのステップに整理できると考えられており，最終ステップである目的実行までのいずれかのステップで防御を成功させることによって実質的な被害を抑えられる．前述した日本年金機構を端緒とする 44 件の連続的な漏えい事件では，サイバーキルチェーンの初～中期段階として位置づけられる高度なマルウェアの侵入，感染といった「偵察」から「インストール」までは防げなかったとしても，被害が発覚してすぐにインテリジェンスを共有および活用できていれば「遠隔操作」(内部感染拡大や情報詮索）や「目的実行」(個人情報のアップロード）のステップで情報漏えいを防げた可能性が高い．例えば，マルウェアを解析したり，プロキシログ等から得られたりする情報漏えい先等のインテリジェンスを迅速に共有して対策に活用できていればこのような被害拡大は防げた．このような目前の脅威の存在に気付きながらも，自組織の防衛に有用であるはずのインテリジェンスを活用した対策に踏み込めない原因として，図 1 に示すセキュリティ対策の 3 段階阻害要因があると考える．

・阻害要因 1：真実把握の困難性
・阻害要因 2：対策方法が不明
・阻害要因 3：対策による既存業務への悪影響の懸念

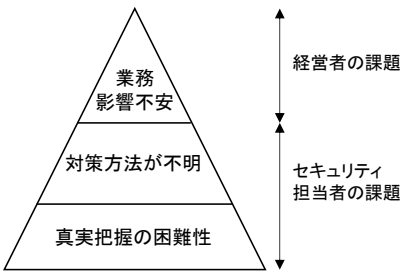

図 1　セキュリティ対策の３段階阻害要因

阻害要因 1 は現在起きている脅威が何か，どのようなリスクが自組織に内在するのかを把握できない問題に起因する．阻害要因 2 は自組織のリスクを把握できたとしても，専門性の欠如等からリスクを軽減する方法がわからない問題に起因する．阻害要因 3 はリスク軽減方法がわかったとしても，対策によるリスク軽減効果や本来業務への悪影響の事前把握が難しく意思決定が遅れてしまう問題に起因する．阻害要因 1，2 はサイバー攻撃に起因するリスク排除を優先するセキュリティ担当者の抱える課題であり，阻害要因 3 は事業利益の最大化を優先する経営者の課題である．このように立場の異なる人間の合意形成の難しさが迅速な対策を困難にしている．

著者らは阻害要因 1 を解決するために，標的型攻撃等で悪用されるマルウェアを自動的に解析し，マルウェア感染による影響を把握する多種環境マルウェア動的解析システム（M3AS）[12]を開発してきた．前述した日本年金機構の情報漏えい問題で悪用された EMDIVI もそうであるように，近年のマルウェアの 82.8%はマルウェア感染後にインターネットに接続して新たなマルウェアのダウンロードや，遠隔操作者（C&C サーバ）との通信，機密情報の窃取のためのアップロード通信等，外部サーバとの通信が発生することが知られている．つまりマルウェアに感染したとしても，上記の外部サーバとの通信を検知，遮断することによって実害の発生リスクを軽減できる．情報処理通信機構はインターネットへの出口に設置したプロキシのユーザ認証機能を有効化することでマルウェア感染による被害拡大を抑止できるとし，同設定を推奨している[13]．また，マルウェアのアクセス先をブラックリストとして定義し，プロキシや FW 等で遮断とすることも阻害要因 2 の解決策となる．ブラックリストとしては Spamhaus[16]や MalwareDomainList[17]等，リストそのものが提供されているものから，FireEye EX[18]や上記 M3AS 等のサンドボックス解析手法を用いてマルウェアから動的に作成されるものがある．このように著者らは M3AS の研究開発を通じて，セキュリティ担当者の抱える課題（阻害要因 1，2）の解決を図ってきた．

しかし，プロキシに対応したマルウェアのうち 8.7%がプロキシ認証を突破すると報告[14]されており，プロキシ認証が万全とはいえない状況となってきている．また，マルウェアの実行環境がインターネットと通信可能かを判断するためにマルウェア実行初期に正規なサーバに対して疎通確認を行う場合もあり，マルウェアのアクセス先の全てが不正なサーバとはいえない状況であった．これらに対してプロキシ認証に加えて新たな認証行動を強いたり，解析の結果得られたアクセス先を一様にブロックしたりする対策は，情報システムの利便性の低下や正規なサーバへのアクセスの遮断等，業務への悪影響に繋がるため問題視されていた．これが経営者の抱える課題（阻害要因 3）の代表例である．

本稿で提案する AED は上記阻害要因 3 を解決し，インテリジェンスを活用した集団防御を実現することにある．

2.2　関連研究

前述したようにマルウェアに感染したクライアントによる社外サーバへの情報漏えいを防止する一つの手法として，社外サーバへアクセスする際に CAPTCHA 認証を求める方式が提案されている[19]．CAPTCHA は機械と人とを判別する逆チューリングテストであり，マルウェアのようなプログラム（機械）では CAPTCHA を解読することができない特性を利用する．CAPTCHA 認証を行うことで，クライアントに感染したマルウェアがブラウザを乗っ取り，社外サイトへアクセスすることを防ぎつつ人間による意図的なアクセスを許可することができる．しかし，CAPTCHA 認証は人間にとっても認証困難であることが多いため，外部サイトへのアクセスの度に CAPTCHA 認証を行っていては日々の業務の妨げとなる．

悪性な社外サイトへのアクセスを防ぎつつも，安全性の高いサイトへのアクセス時には CAPTCHA 認証を省略することで業務への影響を軽減する手法の一つに，ブラックリストとホワイトリスト，これら 2 つのリスト以外に一定の基準を満たした不審な外部サイトはグレーリストへ振り分けておき，グレーリストへのアクセスに対してのみ CAPTCHA 認証を行う方法が提案されている[15]．しかし，本方式は不審な外部サイトを決定付ける基準が静的であるため日々進化する脅威に追従するのは難しい．

セキュリティ担当者と経営者の合意形成の困難さを解決するために，セキュリティを高めた対策にすべきか業務効率等の利便性を重視した対策にすべきかを，セキュリティと利便性との間のトレードオフを評価して対策を選定する方式が提案されている[20]．上記対策の選定手法ではトレードオフを前提としてそれらのバランスを数理的に解決するものであり，リスクとコスト低減のいずれかの妥協が必要となる．

2.3　自律進化型防御システム

本節では，前述した問題を解決しリスク低減と利便性とを両立するために，2 つの技術を提案する．1 つは，正当性や信頼性の欠如するインテリジェンス，ここでは不審な URL リスト（グレーリスト）を活用して，業務へ与える悪影響を最小限に抑えつつ悪性サイトへの接続リスクをも低減し，阻害要因 3「対策による既存業務への悪影響の懸念」を解決するリスクベースプロキシ制御技術である．もう 1 つは，グレーリストから不審な URL ルールを動的に生成し，リストに無い未知の URL への追加認証をも実現することで，さらなるリスク低減を実現する認証条件最適化技術である．

2.4　リスクベースプロキシ制御

グレーリストに登録された URL]へのアクセスに対し，マルウェアでは解決困難な認証手段（本稿でも CAPTCHA 認証を用いるが，他の手段でも構わない）を追加するこ

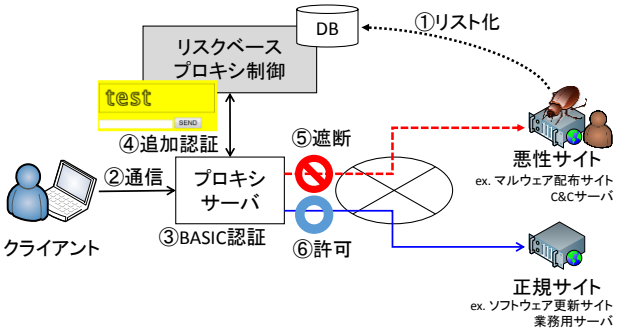

図 2　リスクベースプロキシ制御の概要

とで，ユーザの本来の業務の可用性を損なわず，標的型攻撃で多用されている遠隔操作型マルウェアに対策可能なリスクベースプロキシ制御機能のアーキテクチャを提案する．本機能の概要を図 2 に示す．

プロキシはユーザ認証機能（BASIC 認証や LDAP 認証，AD 認証連携等）をサポートしており，多くの組織で外部サーバへアクセスするユーザのアクセス履歴を記録するとともに，パスワードを知らないマルウェアのアクセスをブロックすることで遠隔操作型のマルウェアのリスクを低減していた．しかしながら前述したように近年様々な方法でプロキシ認証を突破するマルウェアが出現してきている．具体的にはブラウザにキャッシュされている認証情報を詐取するものや，認証済みのブラウザプロセスに悪性なコードをインジェクションしたりするマルウェアが確認されている．これらの最新の脅威に対抗するため，M3AS 等の解析結果に基づき悪性サイトをグレーリストに管理①しておき，クライアントの通信②に対して既存のユーザ認証③に加えて，新たな認証（例えば CAPTCHA 認証）を追加④する．この認証により，マルウェアからの外部アクセスを排除⑤し，且つ業務の可用性を保持⑥することが可能となる．

続いてリスクベースプロキシ制御機能の構成を図 3 に示す．

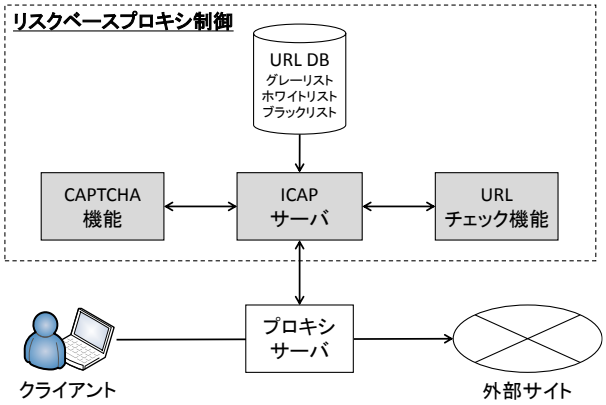

図 3　リスクベースプロキシ制御機能の構成

リスクベースプロキシ制御機能は，プロキシに機能拡張するための標準仕様 ICAP（Internet Content Adaptation Protocol）[21]を利用して，URL チェックや CAPTCHA 認証機能との連携を実現する．

リスクベースプロキシ制御機能は以下の 5 つの機能およびデータベースから構成され，クライアントのアクセス先のリスクに応じて，CAPTCHA 認証を追加する．各構成要素を以下に述べる．

A) プロキシサーバ
本機能は，ユーザによるインターネットアクセスを代理するとともに，ICAP コアと連携してユーザのアクセス先に応じた制御を行う．

B) ICAP サーバ
本機能はプロキシサーバに届いたクライアントからのリクエスト情報を ICAP に従ってプロキシサーバから受信し，URL チェック機能や CAPTCHA 機能と連携して，ユーザからのリクエストに対して認証の追加やアクセスの遮断といった制御を実施する．

C) URL チェック機能
本機能は URL データベースと連携して，ICAP サーバが制御しようとしている URL にどの程度のリスクがあるかを応答する．

D) CAPTCHA 機能
本機能は CAPTCHA 認証に必要な歪み画像および認証フォームの生成を行う．また ICAP サーバからの要求に応じてユーザからの CAPTCHA 追加認証応答の正当性を判定して返答する．

E) URL データベース（URLDB）
本データベースには，正当性や信頼性の欠如する不確実な URL 群をグレーリストとして，不正である確度の高い URL 群をブラックリストとして，安全性の高い URL 群をホワイトリストとしてそれぞれ管理する．

図 4　シーケンス図

上記機能によるリスクベースプロキシ制御機能の処理シーケンスを図 4 に示す．

1. クライアントがプロキシサーバに対してアクセス要求を送信．
2. プロキシサーバはクライアントからのアクセス要求に対し，ICAP サーバへアクセス可否の確認要求を送信．
3. ICAP サーバは URLDB と連携する URL チェック機能に対し，アクセス要求に含まれるアクセス先のリスクを確認．
4. URL チェック機能は URL（あるいはドメインや FQDN）のリスクを送信．

上記リスクがブラックの場合

5. ICAP サーバはアクセス拒否応答をプロキシサーバに送信．
6. プロキシサーバがアクセスを拒否するメッセージをクライアントに送信．

上記リスクがホワイトの場合

5. ICAP サーバはアクセス許可応答をプロキシサーバに送信．
6. プロキシサーバがアクセス先サーバへのアクセス要求中継を依頼．
7. アクセス先サーバがプロキシサーバに要求に対応するコンテンツを送信．
8. プロキシサーバがクライアントにコンテンツを送信．

上記リスクがグレーの場合

5. ICAP サーバは，アクセス先を CAPTCHA 機能に差し替えた応答をプロキシサーバに送信．
6. プロキシサーバは CAPTCHA 機能に対してアクセス要求中継を依頼．
7. CAPTCHA 機能は CAPTCHA 追加認証フォームをプロキシサーバに送信．
8. プロキシサーバが CAPTCHA 追加認証フォームをクライアントに送信．
9. クライアントが CAPTCHA 追加認証フォームに含まれる CAPTCHA 画像を取得するためのアクセス要求をプロキシサーバに送信．
10. プロキシサーバは ICAP サーバへアクセス可否の確認要求を送信．
11. ICAP サーバはアクセス先が CAPTCHA 機能であるためアクセス許可レスポンスをプロキシサーバに送信．
12. プロキシサーバは CAPTCHA 機能に対するアクセス要求中継を依頼．
13. CAPTCHA 機能は CAPTCHA 画像をプロキシサーバに送信．
14. プロキシサーバは CAPTCHA 画像を取得しクライアントに送信．
15. クライアントはユーザから入力された CAPTCHA 認証への回答（テキスト値）を含んだアクセス

要求をプロキシサーバに送信.

16. プロキシサーバは ICAP サーバへアクセス可否の確認リクエストを送信.
17. ICAP サーバは認証の正当性を CAPTCHA 機能に確認.
18. CAPTCHA 機能は正当性を判定し，ICAP サーバに結果を送信.
19. 正当性が確認できた場合，ICAP サーバはアクセス先を本来クライアントが要求していた URL に差し替えた返答をプロキシサーバに送信．正当性が確認できなかった場合は処理 5 に遷移.
20. プロキシサーバが本来のアクセス先へのアクセス要求中継を依頼.
21. アクセス先サーバがプロキシサーバにコンテンツを送信.
22. プロキシサーバがクライアントにコンテンツを送信.

上記の処理シーケンスによりアクセス先のリスクに応じて，遮断，アクセス許可，CAPTCHA 認証追加の制御を行う.

2.5 認証条件最適化

前節で述べたリスクベースプロキシ制御機能によって，例えグレーリストに誤って登録された業務上アクセスが必要な正規サイトへのアクセスも，ユーザの認証が成功すれば支障なくアクセスできるようになる．これにより悪性サイトに接続してしまうリスクを低減しつつ，業務への悪影響を抑え，阻害要因 3 の解決に寄与する．しかしながらグレーリストの拡充に伴い，誤って登録される正規サイトの数も増えてくることが想定され，その度にユーザに追加認証をさせることは，業務効率の低下につながり望ましくない．また，URL データベースに事前に登録されていない URL（未知の URL）へのアクセスに対しての防護策が存在しない.

そこで本節では，グレーリストに登録された不審サイトの安全性を，ユーザの認証結果に基づいて評価するリスト良質化機能を提案する．さらに危険性あるいは安全性の高いサイトに共通する属性を学習することにより，未知の URL に対するアクセス時でもリスクベースプロキシ制御機能で防護可能とさせる不審属性学習機能も提案する．本稿では防御システムの自律進化を支援するこれら 2 つの機能をまとめて認証条件最適化機能と呼ぶ.

2.5.1. リスト良質化

リスト良質化機能はグレーリストに含まれる URL に対して，クライアントによる認証結果を表 1 に示す 3 種類のステータスに分類して認証実績として付与する．さらに，その URL の認証実績の統計値に基づいて危険度の高い URL（ブラック），安全な URL（ホワイト）に分別する.

具体的には CAPTCHA 認証要求数に対し，認証成功となった数を認証成功数，認証失敗となった数を認証失敗数，

表 1 認証ステータス

ステータス	説明
認証成功	CAPTCHA の回答が正しく入力された状態
認証失敗	CAPTCHA が入力されが，回答が正しくなかった状態
認証無試行	CAPTCHA 認証に一定時間（例えば 10 秒等）回答が無かった状態

認証無試行となった数を認証無試行数として，それぞれの値を各 URL の認証実績値として定義する.

本稿では単純に，認証要求数に対する認証無試行数の割合が一定値以上あるいは一定数以上の認証実績値を持つ URL をブラックリストに追加する．また，認証要求数に対する認証成功数の割合が一定値以上あるいは一定数以上に達した認証実績値となった場合はホワイトリストに追加する．本機能により，認証を繰り返すことでグレーリストの URL がホワイトリストおよびブラックリストに機械的に振り分けられる.

2.5.2. 不審属性学習

不審属性学習機能の構成を図 5 に示す．不審属性学習機能は，プロキシサーバから得られるアクセス先 URL やグレーリストに格納された URL に関連する情報を調査して属性値として付与する属性付与機能と，機械学習を用いてルールを生成する学習機能，そして生成したルールを用いて認証を追加するか否かを判断する予測機能から構成される.

属性付与機能は表 2 に示す情報を，グレーリストに URL が追加されたタイミング，あるいはクライアントがアクセス先 URL への代理接続をプロキシに依頼したタイミングで，アクセス先URL を管理している WEB サーバや，DNS，GeoIP[22]等の外部サービスを用いて取得する．表中の#9 は，DNS ラウンドロビン機能を悪用した Fast-Flux 攻撃[21]の場合に不一致になる可能性が高いことを，また#10 や#11 は正規のサイトが HTTP と HTTPS の両方のコンテンツを同等にメンテナンスされている可能性が高いことを想定した属性である.

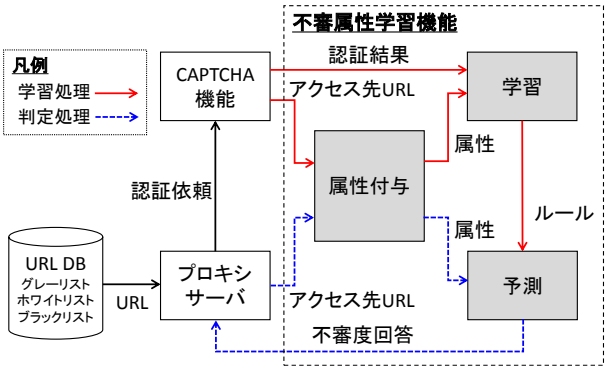

図 5 不審属性学習機能の構成

表 2　URL へ付与する属性

#	属性	取得先	説明
1	HTTP ステータスコード	アクセス先 URL	HTTP アクセスした際のステータスコード
2	HTTP コンテンツサイズ	アクセス先 URL	HTTP アクセスした際のコンテンツサイズ
3	HTTPS ステータスコード	アクセス先 URL	HTTPS アクセスした際のステータスコード
4	HTTPS コンテンツサイズ	アクセス先 URL	HTTPS アクセスした際のコンテンツサイズ
5	DNS A レコード	DNS サーバ	FQDN から正引きした IP アドレス
6	DNS 逆引き	DNS サーバ	IP アドレスから逆引きした FQDN
7	カントリーコード	GeoIP サービス	IP アドレスから所在国を推定した国／地域
8	AS 番号	GeoIP サービス	IP アドレスが属する AS の番号
9	逆引き一致	－	上記#5 と#6 の整合性
10	ステータスコード一致	－	上記#1 と#3 の整合性
11	コンテンツサイズ差	－	上記#2 と#4 の差

学習機能はプロキシの BASIC 認証（図 2-③）を試行した認証成功や認証失敗，あるいは CAPTCHA 認証（図 2-④）を試行した URL への認証成功，認証失敗，認証無試行結果を目的変数として，また付与した属性を説明変数として決定木学習アルゴリズムによりルールを生成する．さらに予測機能は，クライアントがグレーリストに登録されていない URL へアクセスする際に CAPTCHA 認証を表示させるか否かを上記ルールに基づき予測する．

本学習機能を定期的に実行することで変化を再学習させることができ，日々進化する脅威に追従することができるようになる．

3　自律進化型防御システムの実装と評価

本章では AED の主要機能であるリスクベースプロキシ制御機能と認証条件最適化機能の実装を行い，評価を行った．以降に評価の目的と結果を示す．

3.1　評価の目的

2.2 節で述べたようにリスクベースプロキシ制御機能および認証条件最適化機能におけるリスト良質化機能は既存研究に近い方式であることから，本稿では主に追加認証による利便性低下と不審属性学習機能による未知の URL への効果を中心に評価する．

評価 1

追加認証によるユーザビリティへの影響を評価するため，グレーリストに正規サイトが数多く誤登録されている状況を再現した上でリスクベースプロキシ制御機能を実際のユーザに利用してもらい，コンテンツ表示画面への影響や認証頻度等を検証する．

評価 2

グレーリストに登録されていない未知の URL に対して追加認証が適切に表示できるか評価するために，不審属性学習機能の評価を行う．本機能はクライアントによるアクセスの度に属性付与機能が呼び出されることから，本機能がボトルネックになる可能性がある．このため本機能の処理性能を評価する．加えて本機能により属性を学習させて得られたルールを用いて，未知の URL の予測精度を評価する．なお評価では URL として PATH を含まない FQDN を用いる．

3.2　評価結果

評価 1 では，意図的にグレーリストに正規なサイトを追加し，その結果得られる認証ステータスを集計するとともに，コンテンツ表示画面への影響を検証した．今回用意したグレーリストは，被験者が過去にアクセスした URL をアクセス数の多い順にソートし，その下位 30%（33,990 個）をグレーリストに追加して作成した．

被験者 10 名に対して約 11 日間実験した結果を表 3 に示す．グレーリストへの一致率は約 13.6%で，追加認証は 651 回実施された．この時の認証成功率は 27.3%となった．一方で，認証無試行が 274 件と，認証要求全体の 42.1%を占めた．認証無試行となったケースの多くは，アクセス先 URL の HTML ファイルが外部サーバに格納された CSS ファイルを外部参照していて，且つその外部サーバがグレーリストに登録されている場合であった．このため，グレーリストに誤登録されたサーバから CSS ファイル等を読み込む WEB サイトでは，図 6 のようなレイアウトの乱れが生じた．これは，CSS ファイルや外部データファイル等がグレーリストに含まれるサーバに登録され CAPTCHA 認証の追加対象となっていたとしても，ユーザに対して認証画面を表示する手段が無いため，結果的に認証無試行となることに起因する．この問題はグレーリストへの誤登録が多い場合に顕在化するため，グレーリストの精緻化やホワイトリストの充実化が重要となる．

表 3　リスクベースプロキシ制御機能の統計

項目	数	割合
処理リクエスト総数	92,101	100.00%
グレーリスト一致数	12,479	13.55%
認証要求数	651	100.00%
認証成功数	178	27.34%
認証失敗数	199	30.57%
認証無試行	274	42.09%

図 6　表示不具合の例（左：適用前，右：適用後）

表 4　属性付与機能実装サーバのスペック

項目	スペック
CPU	Intel Core i7-2600
Memory	2GB
NIC	1Gbps(光ネクストビジネス)
OS	Ubuntu 14.04

評価 2 では本来，実際のマルウェア感染端末による CAPTCHA 認証（図 2-④）試行データを用いることが望ましいが，前述の評価では感染事象が発生しなかったこと，グレーリスト一致数（URL ユニーク）680 種，CAPTCHA 認証失敗/無試行数（URL ユニーク）が 74 種と，学習および予測評価に十分な量でなかったことから，プロキシの BASIC 認証（図 2-③）試行データを用いて評価することとした．BASIC 認証はブラウザ起動直後に一度しか認証しないため CAPTCHA 認証のように URL 毎に認証実績が付与されない．しかしマルウェア等のプログラムが BASIC 認証に失敗する URL は CAPTCHA 認証にも失敗すると考えられるため，認証失敗実績の傾向は類似すると考える．

属性付与機能を実装したサーバのスペックを表 4 に示す．このサーバで属性付与した結果，1 つの URL あたり 0.22 秒の処理時間を要した．

次に独自に入手した実際の約 2500 万件の BASIC 認証ログを URL 単位に集計して「BASIC 認証に全て失敗」と「BASIC 認証を一度でも成功」の 2 種に分類した URL を目的変数として，さらに表 2 に示した属性を説明変数として生成した 100,000 件の学習用データを決定木学習アルゴリズムで学習させてルールを生成した．さらに，予測精度の評価用 URL を別途 50,000 件用意して，上記ルールと属性から BASIC 認証の成否を予測させて精度を測定した．結果，606 のルールが生成され，表 5 に示す予測精度となった．

以上の結果より，既存の認証結果を学習することで，未知の URL に対しても 99%近い精度で認証の成否を予測できることを確認した．前述したように BASIC 認証ではブラウザが一度認証に成功した後に他の URL へ遷移した際に，再度に BASIC 認証を加えることはできない．一方で CAPTCHA 認証は URL 単位で認証を追加できる．このため，上記予測結果を用いて CAPTCHA 認証を追加／非表示

表 5 予測精度

項目	数	割合
全体	50,000	100.00%
認証成功（ホワイト）	49,174	98.35%
認証失敗（ブラック）	826	1.65%
予測成功	49,411	98.82%
ホワイト予測成功	48,983	99.61%
ブラック予測成功	428	51.82%
予測失敗	589	1.18%
ホワイト予測失敗	191	0.39%
ブラック予測失敗	398	48.18%

することによって，BASIC 認証後の URL データベースに存在しない不審な URL(BASIC 認証を追加できなかった可能性の高い URL)へのアクセスを CAPTCHA で遮断(約 52%)することができる．また本来業務で必要な正規サイトを誤予測してしまった場合（約 0.4%）には，ユーザが CAPTCHA 認証を意識的に入力することによって，業務を阻害することなく意図した WEB サイトへアクセスすることができる．

一方で，本来 BASIC 認証で遮断されるはずの未知の URL の約 48%に対しては本機能の検知漏れにより CAPTCHA 認証が表示されることなくアクセスできてしまう．このため本項目の精度改善が今後の課題となる．

4　考察とまとめ

本稿では，普及するインテリジェンスが不確実な状態であった場合でも，業務に悪影響を与えることなく対策に活かすことが可能な AED を提案した．

本システムでは，マルウェア動的解析システムやインテリジェンスサービスベンダが提供する不審な URL 情報をもとにプロキシに追加認証を加えることで，一様にアクセスを遮断するのではなく，ユーザの意思や認証行為を明確に示させることで，セキュリティ対策の 3 段階阻害要因の最終段階である「対策による既存業務への悪影響の懸念」を解決する．また認証結果と，その URL に関係する属性情報とを決定木学習アルゴリズムを用いて学習させることで，未知の URL に対して認証結果を 99%近い精度で予測できるルールの生成に成功した．本ルールに基づく予測機能を CAPTCHA 認証の追加判断条件に利用することで，52%程度の精度で不審な URL へのアクセス時に認証を追加できる見込みを得た．

一方で，不審な URL へのアクセスを 48%程度の確率で CAPTCHA 認証を経ずに許可しまう点は今後の課題であるが，URL 単位に認証実績が付与された CAPTCHA 認証の試行データを学習することにより，学習精度および予測精度が高まると考える．今後はリスクベースプロキシ制御機能の実証範囲を拡大し，上記 URL 単位の試行データを取得して再度評価したい．

AED は，不確実な情報でも自動対策に活用できる点で

実用性が高い．また，運用を継続することでURLリストやルールが自律的に進化するため，ネットワーク効果や運用コスト低減が期待できる．例えば複数の組織にそれぞれ導入したAED間でグレーリストや認証ステータスを共有することにより，特にばらまき型メール攻撃[24]のような組織を跨る攻撃に対して集団防御の効果を発揮することができると考える．

本稿中で使われているシステム・製品・サービス名は，一般に各社の商標または登録商標です．

参考文献

[1] Guardian News and Media Limited or its affiliated companies：Antivirus software is dead, says security expert at Symantec, 入手先< http://www.theguardian.com/technology/2014/may/06/antivirus-software-fails-catch-attacks-security-expert-symantec>(参照 2015-12-7) .

[2] Solutionary： 2014 NTT Group Global Threat Intelligence Report, Annual Threat Report, 入手先<http://www.solutionary.com/research/threat-reports/annual-threat-report/ntt-solutionary-global-threat-intelligence-report-2014/>(参照 2015-12-7).

[3] 情報通信処理機構,【注意喚起】ウイルス感染を想定したセキュリティ対策と運用管理を，入手先<https://www.ipa.go.jp/security/ciadr/vul/20150602-secop.html>（参照 2015-12-7）

[4] Telecom-ISAC JAPAN, 入手先< https://www.telecom-isac.jp/>（参照 2015-12-7）

[5] 一般社団法人金融ISAC, 入手先< http://www.f-isac.jp/>（参照 2015-12-7）

[6] FireEye：FireEye Threat Intelligence, 入手先< https://www.fireeye.jp/content/dam/fireeye-www/regional/ja_JP/products/pdfs/ds-threat-intelligence.pdf >（参照 2015-12-7）

[7] THREATCONNECT,INC. ：Enterprise Threat Intelligence Platform, 入手先<https://www.threatconnect.com/>（参照 2015-12-7）

[8] サイバーセキュリティ戦略本部,日本年金機構における個人情報流出事案に関する原因究明調査結果,入手先<http://www.nisc.go.jp/active/kihon/pdf/incident_report.pdf>（参照 2015-12-7）

[9] 久保 啓司, 標的型攻撃への対応，一般社団法人JPCERT コーディネーションセンターインシデントレスポンスグループ，入手先< https://www.jpcert.or.jp/present/2015/JNSAWG20150630-apt.pdf>(参照 2015-12-7)

[10] Symantec Corporation：Backdoor.Emdivi，入手先< https://www.symantec.com/security_response/writeup.jsp?docid=2014-101715-1341-99>（参照 2015-12-7）

[11] Eric M. Hutchins, et al.： Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains, ICIW2011, 2011.3

[12] 仲小路博史，重本倫宏，鬼頭哲郎他，“多種環境マルウェア動的解析システムの提案および評価”，情報処理学会論文誌，Vol.56，No.9，pp1730-1744，2015-9

[13] 情報通信処理機構,『高度標的型攻撃』対策に向けたシステム設計ガイド,入手先< https://www.ipa.go.jp/files/000046236.pdf>（参照 2015-12-7）

[14] 下間直樹，鬼頭哲郎，重本倫宏他，“マルチモーダルマルウェア解析システムを用いたプロキシアクセス型マルウェアの解析結果の考察”，コンピュータセキュリティ研究会，vol.2015-IOT-29，No.1，pp1-7，2015-5

[15] 角田朋，大鳥朋哉，藤井康広他，“グレーリストを用いたホワイトリストブラックリストの自動生成によるマルウェア感染検知方法の検討”，情報処理学会研究報告，Vol.66，No.16，2014.

[16] The Spamhaus Project，入手先<https://www.spamhaus.org/>（参照 2015-12-7）

[17] Malware Domain List，入手先< http://www.malwaredomainlist.com>（参照 2015-12-7）

[18] FireEye，Inc：E メール・セキュリティ，入手先<https://www.fireeye.jp/products/ex-email-security-products.html>（参照 2015-12-7）

[19] 土屋貴史，藤田真浩，高橋健太，加藤岳久，間形文彦，勅使河原可海，佐々木良一，西垣正勝. Man in the Browser 攻撃対策を実現する人間・サーバ間のセキュア通信プロトコル. 情報処理学会研究報告，Vol.69，No.22，2015.

[20] 芝口誠仁，稲場太郎，中山佑輝，岡田謙一. 仕事量及び利便性低下度に着目したセキュリティ対策選定手法. 情報処理学会研究報告，Vol.70，No.11，2009.

[21] Internet Engineering Task Force：Internet Content Adaptation Protocol (ICAP), 入手先<https://tools.ietf.org/html/rfc3507>(参照 2015-12-7)

[22] MaxMind，Inc：GeoIP2：業界をリードする IP 情報収集，入手先<https://www.maxmind.com/ja/geoip2-services-and-databases>（参照 2015-12-7）

[23] WIKIPEDIA：Fast flux，入手先 https://en.wikipedia.org/wiki/Fast_flux（参照 2015-12-7）

[24] 情報通信処理機構,【注意喚起】特定の組織からの注文連絡等を装ったばらまき型メールに注意,入手先< https://www.ipa.go.jp/security/topics/alert271009.html>（参照 2015-12-7）