1W-05

攻撃コンテンツの不可視化特徴及びURL文字列特徴を用いたDrive-by download攻撃検知手法の提案

髙田航太[†1]　吉田豊[†1]　稲村浩[†1]　中村嘉隆[†1]
公立はこだて未来大学[†3]

1. はじめに*

ドライブバイダウンロード攻撃(Drive-by Download Attack：以下 DbD 攻撃)とは，攻撃コードの仕組まれた悪性サイトにユーザがアクセスするだけでマルウェアに感染させられてしまう攻撃である．近年被害が増加しているランサムウェアにも，この攻撃方法を用いるものがある．スキルの低い攻撃者でもこの手法を用いた攻撃を容易に行えるよう幇助するエクスプロイトキットが存在する．これはユーザの OS もしくはソフトウェアの脆弱性を自動的に判断し攻撃を仕掛ける犯罪ツールである．これまでに DbD 攻撃に対してリダイレクト解析や HTTP ヘッダに着目した検知手法など様々な検知手法が提案されてきた．本研究では DbD 攻撃の入り口サイトの構造を解析し，その構造的特徴と攻撃に用いられる URL の文字列特徴から，従来の検知手法では困難であった検知を可能にするシステムの実現を目的とする．これにより，正規サイトが改ざんされている場合でも検知することが可能になり，従来のシグネチャーベースの検知手法では難しかった未知の悪性サイトの検知が可能になる．

2. 関連研究

2.1 攻撃コンテンツの不可視化に着目した研究

DbD 攻撃の検知方法として，Web コンテンツに着目した荻野ら[1]の研究がある．攻撃コンテンツは気付かれにくいように不可視化されている場合がある．これはリダイレクトの元に用いられる iframe コンテンツ等が透過処理されていたり，画面の描画領域外にコンテンツが配置されている場合である．攻撃コンテンツに不可視化という特徴があれば，シグネチャーベースの検知では難しかった未知の攻撃にも対応できる．

2.2 エクスプロイトキットで用いられる URL 文字列特徴に着目した研究

DbD 攻撃にはエクスプロイトキットが頻繁に用いられている．エクスプロイトキットが使用された攻撃に用いられる URL には，エクスプロイトキット毎に特徴があることが知られている．佐藤ら[2]は攻撃に用いられている URL の文字列特徴から，使用されたエクスプロイトキットを特定し攻撃を検知する手法を提案している．エクスプロイトキットが用いられている攻撃には高い精度で検知が可能であるが，エクスプロイトキットの新種や亜種への対応が困難である問題点がある．

3. 提案手法

本研究では攻撃コンテンツの不可視化だけでなく，攻撃に用いられる URL の文字列特徴によって未知の入り口サイトにも対処可能な，DbD 攻撃の入り口サイトの判別を利用者の手元で可能にする．

3.1 公開データセットを用いた入口サイトの構造的な特徴及び，URL 文字列特徴の発見

公開データセットから入り口サイトとみられるHTMLファイルソースを解析し，攻撃元となったコンテンツの座標や大きさなど構造的特徴をまとめた．関連研究[1]では不可視化という特徴について取り上げて研究を進め，検知率は約80%であった．これは不可視化という特徴だけに注目した結果である．関連研究[2]によってエクスプロイトキット毎に用いられるURLに特徴が存在することが示されているが，攻撃コンテンツにもエクスプロイトキット毎に特徴が見られる．図1に攻撃コンテンツのエクスプロイトキット毎の不可視化特徴の例を示す．

	width	height	top	z-index	opacity
NEUTRINO系	300~320	300~310	-1050~-1180	-	-
Rig系(静的)	250~270	250~270	-1000~-1150	-	-
Rig系(動的)	5~20	5~19	-	-	-
NEBULA系	1	1	-500	-	-
ANGLER系(iframe)	50	50	-	-1	0
ANGLER系(embed)	30~50	30~50	-	-1	0

図1：EK毎の攻撃コンテンツの不可視化特徴の例

*Detection of Drive by Download Attacks Detection based on Features of Malicious Contents' Visibilies and URLs' Strings
†1 KOTA TAKADA, YUTAKA YOSHIDA, HIROSHI INAMURA and YOSHITAKA NAKAMURA, Future University Hakodate.

このエクスプロイトキット毎の不可視化特徴と用いられるURL文字列特徴を判定部の判定ベクトルに用いる．これにより攻撃コンテンツの不可視化という特徴の有無に限らず，エクスプロイトキットが用いられている攻撃の検知を可能にする．

3.2 Web ブラウザのプラグインとしての実装

提案システムは Web ブラウザのプラグインとして実装する．クライアントサイドで実装する理由は二つある．一つは検知した情報をサーバとやりとりすることなしに，クライアントサイドで動作させた方が高い応答性が期待されるからである．検知に必要なデータが逐次サーバとやりとりされ，接続対地ごとの確認において通信による遅延が加わるための応答性への悪影響が見込まれる．もう一つの理由として，利用者の実際の Web アクセス行動に即した悪性サイトの検知が行える点である．

悪性サイトの発見を行うためにクローラによる自動巡回がおこなわれている．しかし膨大な Web ページ数に対してクローラが巡回できる範囲は限定的であるという問題があり，そのため効率の良いクローリングについての研究も行われている[3]．クライアントサイドに実装することで，利用者の実際の Web アクセス行動に即した検知が可能である．またクローラによる巡回範囲に限定された事前検知結果に依らず，本手法を用いれば任意の対象サイトについての検証が可能である．

提案システムが入り口サイトを検知するまでの概要を図 2 に示す．

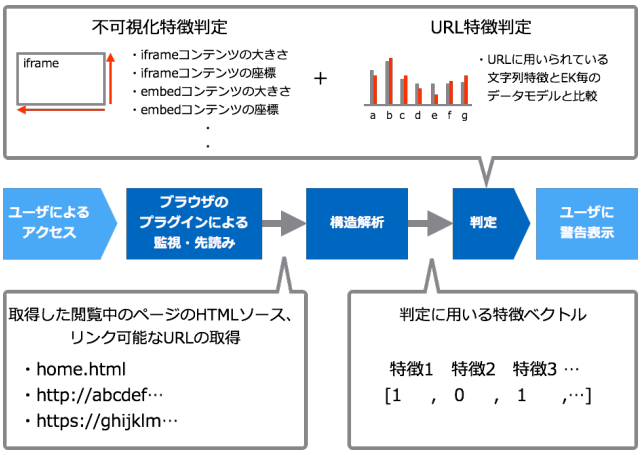

図 2 ：提案するシステムの概要

ユーザがあるサイトのページにアクセスすると，提案手法を実装したブラウザのプラグインが，そのページからリンクにより遷移可能なページの HTML ソースなどを先回りして入手する．入手したソースの構造を解析し，用いられているタグなどから特徴ベクトルを生成する．また用いられている URL についても，文字列特徴などによってエクスプロイトキットが用いられている可能性が高い場合，どのエクスプロイトキットに該当するか判断する．この特徴ベクトルを用いて判定を行う．ユーザが，悪性コンテンツを含むと判定されたページにアクセスを試みるとポップアップで警告が表示される．あらかじめ悪性データセットを用いて入り口サイトの特徴と URL のパターンからモデルを作成しておき，判定に用いる．

DbD 攻撃に用いられる攻撃コンテンツがすべて不可視化されているとは限らない．本手法では，攻撃コンテンツが不可視化されていなくても，URL 文字列特徴によって検知することができた．

4. おわりに

本稿では，Drive-by Download 攻撃の攻撃検知に対して，エクスプロイトキット毎の攻撃コンテンツの不可視化と，用いられる URL の文字列特徴に基づいた検知手法を提案した．攻撃コンテンツの不可視化特徴と用いられる URL パターンに着目することで，ブラックリスト方式などのシグネチャーベースでの検知が難しかった攻撃に対しても効果が見込める．

今後の課題として，不可視化特徴と URL 文字列特徴パターンだけでなく，Web コンテンツの他の要素も検知に用いられないか検討したい．また JavaScript などによって動的に攻撃コンテンツが生成されるものに関しても検知できるよう改良を加えたい．

参考文献

[1] 荻野貴大，高田哲司：不可視 Web コンテンツ特徴に基づく Drive-by Download 攻撃の検知と調査支援ツールの提案，研究報告マルチメディア通信と分散処理（DPS），Vol.2017-DPS-170，No.23，pp.1-8(2017)．

[2] 佐藤祐磨，中村嘉隆，高橋修：エクスプロイトキットで利用される文字列特徴を用いた悪性 URL 検知手法の提案，情報処理学会研究報告，Vol.2016-DPS-166，No.25，(2016)．

[3] 千葉大紀，森達哉，後藤滋樹：悪性 Web サイト探索のための効率的な巡回順序の決定方，コンピュータセキュリティシンポジウム 2012(CSS2012)論文集，Vol.2012，No.3，pp.805-812，(2012)．