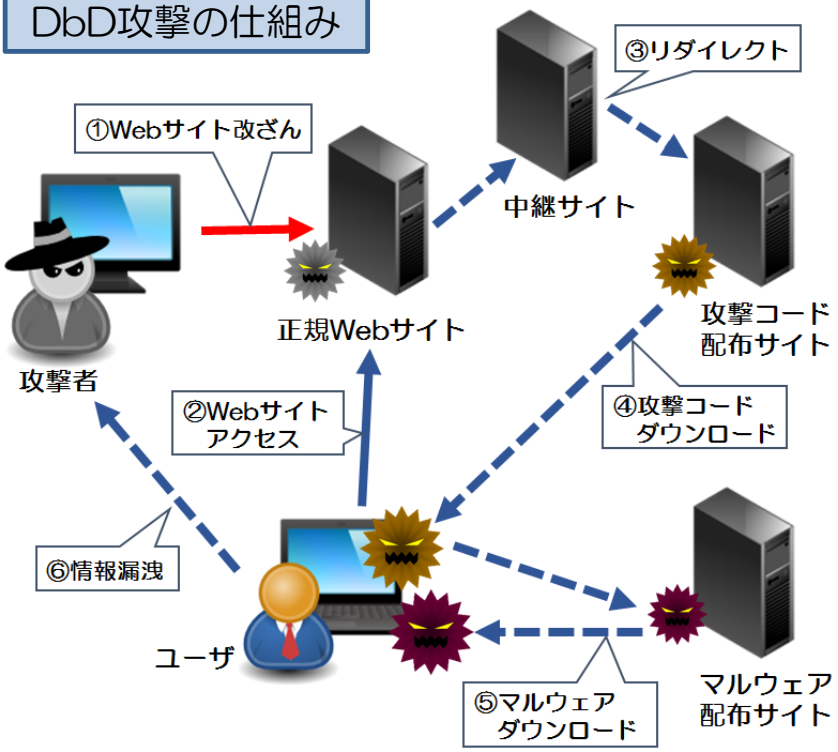
ヘッダフィールド名の出現傾向による Drive-by Download攻撃の検知手法

Detection of Drive-by Download Attack through usage pattern of HTTP header field names

榎本尚代・マネジメント分科会・情報セキュリティ大学院大学

Abstract ：
Drive-by Download is a method that uses the web pages to install malware automatically on your computer. It has become diverse and sophisticated. In addition, the attack tool called Exploit Kit used in the Drive-by Download can easily add attack codes, so vulnerabilities before the release of the patches are added in a second and tend to be exploited by attacks. Therefore, it is becoming more difficult to detect by the anti-virus products using pattern matching or by the access control using blacklist which had been mainline detection methods. In this research, we propose a method to distinguish malignant / benign communication using machine learning, characterized by the appearance frequency of HTTP header field names.

DbD攻撃の仕組み



提案手法

HTTPヘッダ情報を利用したDbD攻撃の検知手法が活発に研究されている。

↓

一つのヘッダ情報で悪性と判定するのは困難。複数のヘッダ情報を利用して検知精度の向上が課題である。

↓

DbD攻撃の一連の通信中に現れるヘッダ情報には、何か傾向があるのでは？

ヘッダフィールド名の出現傾向から悪性／良性通信の分類は可能か？

↓

DbD攻撃の一連の通信から抽出したヘッダフィールド名の出現回数を特徴量として、機械学習を用いて悪性通信と良性通信を分類する手法を提案。

実験

実験データ（アクセス先URL件数）
悪性通信データ：１３０６件　　良性通信データ：２９４件

アクセス先URL毎にどのヘッダフィールド名が何回出現したか集計。

↓

ヘッダフィールド名の出現回数を特徴量として、サポートベクターマシン（SVM）で精度評価

【集計方法 - 例】

ヘッダフィールド名：	ヘッダの値
Server:	Apache/2.2
Content-Length:	155
Content-Type:	text/javascript

ヘッダフィールド名で仕分けて出現回数を集計　　ヘッダの値は含まない

結果とまとめ

- SVMで精度評価をした結果、正解率は９６.６%。
  （再現率：９０.３%、F値：９０.４%）
- 本提案手法のソースコードは軽量。ユーザ端末やProxyで動作可能なため大掛かりなシステムの設置は不要。
- 機械学習を用いた検知手法は、データ件数を増やしたりチューニングすることで、判定精度向上の見込み。