4X-08

ユーザのコマンド履歴による Adaboost を利用した認証手法の改善への試み

中田　明秀†　　小高　知宏†　　白井　治彦‡　　黒岩　丈介†

†福井大学大学院工学研究科　　‡福井大学工学部

1　はじめに

私たちが一般的に用いる侵入検知手法としてパスワードや所有物を利用した認証手法が挙げられる. しかし, これらの手法は, 基本的にログイン時にしか用いられておらず, 認証に必要となる情報が盗まれてしまった場合, 無防備となってしまう可能性が高い.

そこで, 本研究ではログイン中のユーザの挙動に注目した認証手法として,Linux 等の端末で入力したユーザのコマンド履歴を用いて認証を行う手法について取り上げた. ここで用いたコマンド履歴であるが, 本研究では一般的に公開されている Schonlau[1] のコマンド履歴のデータを用いて, 様々な比較手法を用いて認証評価を行うことが出来る.

しかし, この認証手法では, 様々な手法により解析が行える半面, 一つ一つの比較手法ではあまり精度の良いユーザの認証を行うことが出来ないという問題点がある. そこで,Adaboost[2] という機械学習アルゴリズムを用いて, 複数の比較手法の結果を組み合わせることで, どの程度認証の精度を改善することができるのかを研究の目的とした.

2　コマンド履歴を用いた認証手法

ここで, コマンド履歴を用いた認証について, どのようにして行うのかについて示す. この認証手法ではまず, 正当なユーザが入力したコマンド履歴の情報について, あらかじめ学習を行う. 次に, あとから入力されたものとを比較して認証を行い, 図 1 のようにして行われる.

学習モデルの構築では, 正当なユーザが入力したコマンド履歴の情報を用いて学習を行う. ここでは, ユーザが入力したコマンドの種類や出現頻度, コマンドの前後関係といったものを学習させる. これを学習モデルと定義し, のちに構築される検査モデルとの比較に用いられる.

学習モデルによる十分な学習が終わったら, 次に検査モデルの構築が行われる. こちらも先ほどと同様にし

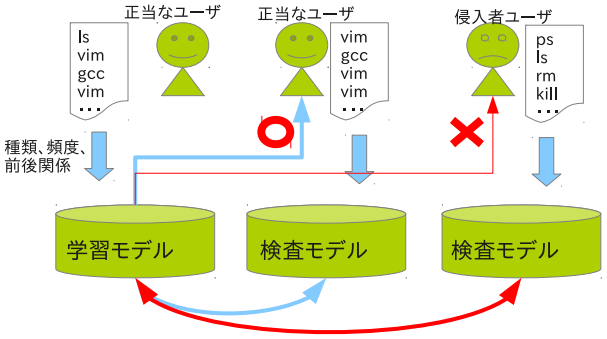

図 1: コマンド履歴を用いた認証手法

て, ユーザが入力したコマンド履歴の情報を元にモデルの構築を行う. しかし, 先ほどとは異なり, ここで構築されるモデルには侵入者ユーザが入力したものについても行うことが出来る. ここで構築したモデルを検査モデルと呼ぶことにする.

ここで, 先ほど構築した学習モデルと検査モデルとを比較して, そのユーザが同一のユーザであるかどうかの比較が行われる. 比較には, 後述する比較手法に基づいて評価を行い, その差異がどの程度のものであったかによって, 検査モデルが正当なユーザか侵入者ユーザかどうかを判断して認証が行われる.

3　Adaboost による学習アルゴリズム

ここで,Adaboost と呼ばれる機械学習アルゴリズムを用いて, 本研究で用いた比較手法の認証の精度の向上を図ったのかについて簡単に示す. Adaboost では, 与えられた比較手法により, その結果が正しいかどうかを判断できる弱識別器を複数用意して行われる. このうち, どの識別器の精度が最も良かったかを重みという概念で評価し, 繰り返し学習を行うことによって, より正確な結果を導き出せる強識別器を生成する.

Adaboost では, 図 2 のようなアルゴリズムにより, 学習が行う. 初めに, 重みの初期値を D_1 として, 比較となる検査モデルについて, 重みの初期値を与える. 次に, もっとも精度が良いと判断された弱識別器 H_t を選択し, その誤り率 ϵ_t が 0.5 以下であれば, 信頼率 α_t を求める. そして, それぞれの検査モデルが正しく判別できたかどうかによって, 重みの更新が行われる. 具体的には, 誤って判断されたモデルについては重みを高くし, 正しく判断されたモデルに対しては重みが低くなるように更新

Improvement of the Intrusion Detection Method Based on User Command History Using the Adaboost Machine Learning Meta-algorithm
†Akihide NAKATA　†Tomohiro ODAKA　‡Haruhiko SHIRAI　†Jousuke KUROIWA
†Graduate School of Technology, University of Fukui Unified Graduate School
‡School of Technology,University of Fukui

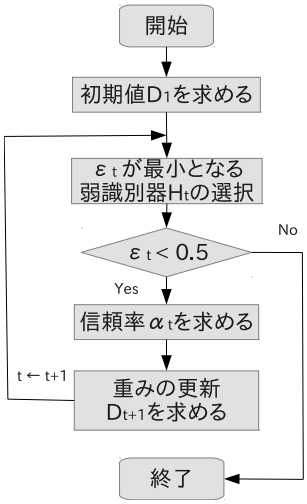

図 2: Adaboost による学習アルゴリズム

される.

こうして十分に学習を行い, 最終的に強識別器を生成することによって精度の高い認証を行うことが出来る.

4 実験

本研究では,Schonlau データにおける 50 人分のユーザのコマンド履歴を用いて, 実験を行った. Schonlau データには, 各ユーザごとにそれぞれ 5000 個のコマンドデータを持った学習データと 1 セッション 100 個のコマンドを基本単位として,100 つのセッションを持った検査データが構築されている.

そこで, 本研究では, 検査データに含まれる各セッションのデータについて正しくユーザを判別できるかどうかについて実験を行った. ここで, 学習データを学習モデル, 検査データの各セッションを検査モデルとして, モデルを比較することにより, それらのユーザモデルが同一ものであったかどうかについての比較を行った.

また, 比較には表 1 にある 6 つの手法を用いた. これらの違いは, コマンド履歴の情報を特徴化するのに用いた方法であり,1 つ 1 つのコマンドもしくは連続した 2 つのコマンドに注目するかというものである. また, ユーザのモデルを比較した時の差異を計る方法として, コマンドのヒット率,COS 類似度,TF-IDF を用いたが, こちらについて具体的な説明は当日のスライドに示す.

5 結果

ここで, 本研究で用いた各手法及び Adaboost による結果について示す. 図 3 は,FRR(他人受け入れ率) と FAR(本人拒否率) との関係を示したものである. このグラフから,method1～6 のそれぞれの比較手法よりも,Adaboost により学習を行ったものの方が FRR,FAR は遥かに減少していることが分かる. そのため, コマンド履歴を用いた認証手法においても, 複数の手法を組み合わせることで認証の精度を改善することが出来るといえる.

表 1: コマンド履歴の比較に用いる手法

method1	1 コマンドによるコマンドのヒット率
method2	1 コマンドによる COS 類似度
method3	1 コマンドによる TF-IDF
method4	2 コマンドの連鎖によるコマンドのヒット率
method5	2 コマンドの連鎖による COS 類似度
method6	2 コマンドの連鎖による TF-IDF

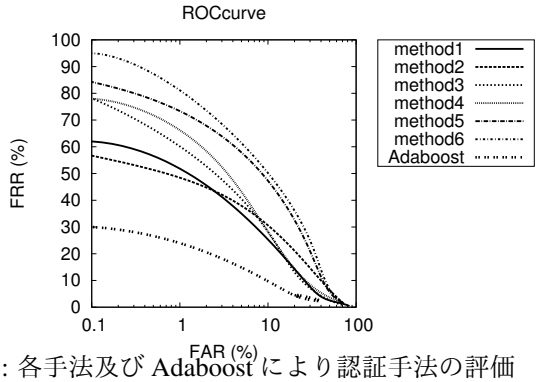

図 3: 各手法及び Adaboost により認証手法の評価

6 考察とまとめ

本研究では, コマンド履歴を用いた認証手法について, 複数の比較手法を用いることで認証精度の改善を計るためのものとして Adaboost を用いた. これにより, 最適と思われる FAR,FRR において, それぞれ 10%程度に抑えることができ, 比較手法を組み合わせる前よりも良くなるという結果が得られた. Adaboost による学習では, 多くの手法を組み合わせるほど精度が良くなるため, ここで述べたもの以外の手法とも組み合わせることが出来れば, さらに改善が可能ではないかと考えられる.

参考文献

[1] Matthias Schonlau,William DuMouchel,Wen-Hua Ju,Alan F.Karr,Martin Theus and Yehuda Vardi. ”Computer Intrusion:Detecting Masquerades”. Statistical Science 2001,Vol.16,No.1,1-17.

[2] 塚本　浩司, 颯々野　学”Adaboost と能動学習を用いたテキスト分類”. 自然言語処理 146-13 (2001.11.21).