コマンド入力連鎖に見られる個人の特徴抽出

6 B − 3

白井 治彦　　西野 順二　　小高 知宏　　小倉 久和
(福井大学工学部)

1　はじめに

コンピュータ・ネットワークが発達した現在, 正当なユーザになりすましシステムに被害を与えるコンピュータ犯罪は, 深刻な社会問題である。現在はこのような対策として, パスワードによるユーザの認証システムが主流である。しかし, パスワード破りによる犯罪事例も数多く報告されている。この場合にも, ユーザが使用するコマンド列に何らかの個人特徴が見つかれば, その特徴を基にシステムが絶えず監視を続けることで, たとえパスワードが破られても使用者が正当なユーザかどうかの識別が出来る。我々はこれまでに, コマンド列における個人特徴の着目点とその成果を報告してきた [1] [2]。今回は, あるコマンドに続いて次に入力されるであろうコマンドの確率に注目した。以下では, 個人の過去の入力履歴情報を基にコマンド系列の遷移確率表を構成し, 入力されつつあるコマンド列の遷移状況と対比することで, 本人かどうかの判別を行う手法を提案する。

2　コマンド入力連鎖

2.1　本手法の原理

ユーザがコンピュータと対話的に作業を行う時, そのコマンド系列は何らかの個人の特徴が含まれていると考えられる。今回用いた手法では, まず, ユーザのあるコマンド入力に続いて次にどのようなコマンドが入力されるかを絶えず監視する。そして, 入力されつつあるコマンド列の遷移状況が, 以前に入力されたコマンド列を基に予測される遷移状況と似ているかを調べることで本人かどうかを見分ける。例えば図1のようなコマンド列の遷移状況の時, 以前にこのユーザが図1のコマンド遷移がどの程度の確率で生じたかを調べる。この確率は, ユーザ毎に過去の入力履歴情報に基づいてあらかじめ計算しておく。これら一連の入力について確率の平均値を計算することで, ある閾値をもって本人かどうかの判断をする。

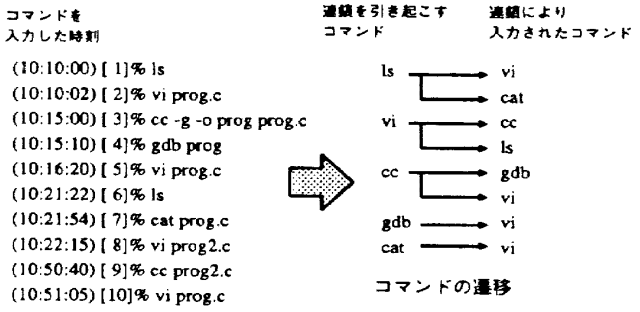


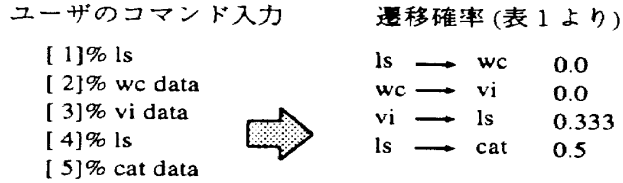
図1　UNIX コマンドとその遷移状況の例

2.2　個人認証のためのシステム構成

本認証システムは, 2つのステップから構成する。第一のステップでは, 個々のユーザが行う計算機への入力コマンド列をあらかじめ採取・蓄積しておき, 各ユーザのコマンド遷移の特性を計算する。ここで, 入力コマンド列を格納したファイル (事前学習用ログファイル) を用いてコマンド遷移の確率を計算し, 遷移確率を収めたデータファイル (遷移確率表) を各ユーザ毎に構成する。ログファイルを基に生成された遷移確率表の例を表1に示す。

表1　図1の例より生成された遷移確率表

連鎖を引き起こすコマンド ＼ 連鎖により入力されたコマンド	ls	vi	cc	gdb	cat
ls	0.0	0.5	0.0	0.0	0.5
vi	0.333	0.0	0.666	0.0	0.0
cc	0.0	0.5	0.0	0.5	0.0
gdb	0.0	1.0	0.0	0.0	0.0
cat	0.0	1.0	0.0	0.0	0.0



遷移確率 ＝ (0.0 + 0.0 + 0.333 + 0.5) / 4 = 0.2083

図2　遷移確率表を用いて遷移確率を求める例

第二のステップでは, 入力ログファイルと遷移確率表を用いて, ある一連の実用時入力動作が正当なユーザのものであるかどうかを判別する。それぞれのコマンド連鎖について, 第一のステップで作られた遷移確率表より遷移確率を求め, その平均を計算する。このと

An intrusion detection technique using characteristics of command chaines in interactive computer environment.
Haruhiko Shirai, Junji Nishino, Tomohiro Odaka, and Hisakazu Ogura.
Faculty of Engineering, Fukui University
3-9-1 Bunkyo,Fukui 910,Japan

き，遷移確率表に無いコマンド連鎖についてはその確率を0(ゼロ)とする。表1の遷移確率表を用いてコマンド連鎖の遷移確率を求める例を図2に示す。ここで求められた平均値は，本人であれば高く，他人であれば低くなる傾向がある。そこであらかじめ適当な下限閾値(たとえば0.20)を設定し，ある程度のコマンド連鎖について求めた値がその下限値を下回る場合，不当なユーザとして報告する。このような個人認証システムの構成図を図3に示す。

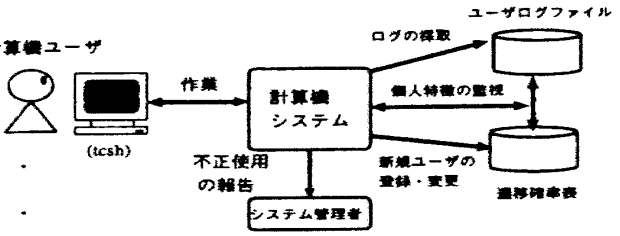


図3　個人認証のためのシステム構成図

3　実験結果

本手法の適応性を調べるため，いくつかの実験を行った。実験データは，いずれもUNIXシステム上のシェルコマンド(tcsh)より採取したものである。

3.1　実験1の設定とその考察

本研究室の学生4名の被験者(1～4)について本手法を適用した。実験に使用したログファイル採取量を表2に示す。ログデータ採取期間は約7ヵ月，作業内容はプログラム開発及び文書作成が中心で，被験者はいずれもUNIX使用歴が4～5年以上のある程度使い慣れたユーザである。

表2　実験1のためのログデータ採取量

単位：ステップ

	data1	data2	合計
被験者 1	9052	9053	18105
被験者 2	9765	9765	19530
被験者 3	8719	8720	17439
被験者 4	13384	13384	26768
合　計	40920	40922	81842

この実験では，まずログファイル採取量を二分割する。前半のデータ(**data1**)を第一ステップにおける入力ログファイルとして遷移確率表を作るために使用し，後半のデータ(**data2**)を第二ステップにおいて正当なユーザかどうかを調べるために遷移確率の平均を求めるデータとした。その結果が**表2**である。表中の数値は(例えば被験者1と被験者1の結果0.166546は被験者1の**data1**に依る遷移確率表を用いて，被験者1の**data2**について)遷移確率の平均値を計算した結果である。この結果より，確率平均の下限値を適切に設定してやれば，本人かどうかの判定が可能である。

表3　実験1の遷移確率の計算結果(平均値)

	被験者 1	被験者 2	被験者 3	被験者 4
被験者 1	0.166546	0.120222	0.098571	0.121772
被験者 2	0.15541	0.190057	0.133453	0.17444
被験者 3	0.137989	0.126164	0.175452	0.104019
被験者 4	0.156735	0.157354	0.096693	0.192803

3.2　実験2・3の設定

実験2では，本学科の2・3年の学部生約200名のデータ9ヵ月分を対象とした。今度の被験者は比較的UNIX使用歴が浅い初心者である。この実験のデータとして，**data1**には4月から9月までのログデータ全てを，**data2**には10・11・12月それぞれのログデータを用いた。

実験3では，学部生約100人(現在3年生)のデータ2年分を用い，**data1**には前年度のログデータ全てを，**data2**には今年度の一ヵ月毎のログデータとした。

これらの実験結果は，口頭で報告する。

4　まとめ

今回の報告で，ユーザの入力コマンド列を解析することで個人の特徴を抽出し，システムセキュリティの強化に役立てる方法を提案した。本手法は比較的システム実装が容易で，当然，従来のユーザ認証法(パスワード等)との併用も可能である。ただし，いくつかの問題点も残されている。第一にユーザの数が多くなればそれだけユーザ間の類似が増し，個人を特定するのが困難になる点である。第二に，今回の方法では，作業内容が限定されていれば個人の遷移確率も比較的高くなるが，多岐に渡ればそれだけ確率値を下げ，これも特定を困難にしてしまう。個人毎の遷移確率表を生成する最適な条件を見つけることや工夫により，この問題も解決できる。

参考文献

[1] 加藤友彦，高田光男，小高知宏，小倉久和: "対話的計算機環境におけるキーボード入力系列のモデル化と認証への応用"，電子情報通信学会誌(B),Vol.J78-A,No.9,pp.1251-1254.Sep(1995).

[2] 小高知宏，加藤友彦，高田光男，西野順二，小倉久和: "計算機利用者のシステム操作入力文字列に基づく認証手法の検討"，電子情報通信学会誌(B),Vol.J79-A,No.4,pp.1001-1004.Apr(1996).