コンピュータセキュリティ　7－1
（2000．1．21）

UNIXワークステーションにおける不正利用者の判別方法

中國　真教† 堂薗　浩††
原　重臣†† 野口　義夫††

†佐賀大学大学院工学系研究科
††佐賀大学理工学部

本研究では，UNIXワークステーションをある利用者が利用する際，その利用者のキーボード入力の特徴を捉えることにより不正利用者の判別を行う．キーボード入力の癖は，利用者それぞれに特徴があり，例えば，入力ミスの多いキーやキーボード入力の速度，実行するUNIXコマンドの種類やオプションの使い方など，キーボードの操作やUNIXの操作の熟練度の違いによりキーボード入力の特徴が異なる．このような判断基準を設け，利用者が実際にどのような操作を行ったのかをキーボードから直接情報を取得し，利用者の特徴や癖をコンピュータが学習することにより不正利用者の判別を試みる方法について実験し考察を行う．

A Method to Distinguish an Injustice User in UNIX Workstation

Masanori Nakakuni,† Hiroshi Douzono,††
Shigeomi Hara †† and Yoshio Noguchi††

†Master Cource of Science and Engineering, Saga University

††Faculty of Science and Engineering, Saga University

When a user uses a work station, the distinction of the injustice user will be possible by reading the characteristics of the user's keyboard input and learning. For example, the user will be dentified by the habits: the UNIX command that it is carried out, the keys which are mistyped, and the speed of the keyboard inputs in the session. User's characteristics can be read from input of keyboard constancy and learned by computer, and the distinction of the automatic injustice user will be possible. We made some experiments of this method and the results are reported in this paper.

1. はじめに

近年，インターネットの急速な広がりと，数多くのネットワーククラッキングツール[1]の流通のために，ネットワーククラッキングの発生が増加している．これはインターネットの利用において深刻な問題[2][3][4]となっており，ネットワーククラッキング対策はコンピュータネットワークの管理者にとって重要な課題の一つとなっている．近頃のクラッキングツールはGUIベースのものが多く，コンピュー

タの初心者でも操作が簡単で，クラッキングに関する多少の知識があれば，比較的容易にクラッキングを行うことが出来る．そのためコンピュータの初心者でさえクラッカーになることができる．クラッキングツールの種類は様々であるが，例えば，ネットワーク上を流れるパケットの中身を盗み見るツールは，何者かがこのようなツールを利用することによって，ネットワーク利用者の個人データを盗み見られることがある．このように盗聴された個人情報には，インターネット上のネットワークサーバにアクセスするためのIDとパスワードが含まれる場合があり，そのような情報が盗聴され，悪用される場合がある．

一般にUNIXを搭載したコンピュータではユーザ名とパスワードを用いて正規利用者であることを認証する方式をとっているが，利用を許されていない人物が正規利用者のパスワードを盗み取るなど，何らかの方法を用いてコンピュータを不正に利用する場合がある．これらの不正利用を阻止するための現在の主なコンピュータセキュリティ対策はネットワークを経由したコンピュータへの不正アクセスを阻止するものが多いが，不正アクセスが成功した場合の対策はあまり講じられていない．コンピュータへの不正アクセスが成功した後の不正利用者の発見方法はコンピュータ内部に蓄積する利用状況や利用時間などの履歴をそのコンピュータの管理者が手動で読み取り，不審な記録を発見するという方法が従来の主な手法である．このようにコンピュータに蓄積する記録を管理者が調べるだけでは正規利用者であるのか不正利用者であるのかの判別が困難な場合があり，不正利用者の発見が遅れてしまうことがある．また，記録の解析は管理者にとって重労働であるので，コンピュータが自動で不正利用者を判別することが望まれる．

本研究では，UNIXマシンの利用者が実際にどのような操作を行ったのかをキーボードから直接情報を取得し，利用者の特徴や癖をコンピュータが学習することにより不正利用者を判別する実験を行い，その結果について考察を行う．

2. 利用者の特徴

利用者には個人ごとに様々な特徴がある．それは，まず，利用者それぞれのUNIXの用途の違いにより利用者それぞれに特徴が現れる．UNIXの利用者の特徴として，次のような利用者の例を挙げる．

[利用者1] エディタやLaTeXを使い，Mailメールを読む程度．使用するUNIXコマンドは`ls`や`mkdir`のような基本的なもの使用する．

[利用者2] UNIXを計算機サーバとして，主にコンパイラを使用する．また，ソースファイルの編集のためにエディタを使用する．

[利用者3] メールの読み書き，各種ファイルの作成を全てUNIX上で行う．また，UNIXコマンドやUNIXシェルを巧みに操る．

[利用者4] UNIXの利用者でもあり，システム管理者でもあるため，一般ユーザが使用するコマンドから，システム管理に関係するコマンドまで，様々なコマンドを実行する．

[利用者5] UNIXに大変興味を持ち，好奇心旺盛で様々なUNIXコマンドの実行を試みる．

その他には，キーボード操作の馴れ，キー入力の間違いが多いキーなどに利用者の特徴が現れる．

3. 利用者の特徴の学習

不正利用者を判別するためには利用者のキーボード入力をコンピュータが読み取り，その利用者の特徴や癖をコンピュータが学習し評価することによって判定を行う．利用者のキーボード入力から不正利用者の判定までのおおまかな流れを図1に示す．

コンピュータが利用者から得た情報を学習するためには，それらの情報を何らかの方法で評価し，その評価の度合を数値化する必要がある．そこで，不正利用者の判別に有効であると考えられる判定項目をいくつか設け，それらの判定項目に該当する情報を利用者のキーボード入力から取得する．不正利用者の判定を行うためには，それぞれの判定項目での評価を数値化し，まず，初期の学習の段階では利用者のキーボード入力の特徴や癖を学習し，ある程度の学習を行ったところで不正利用者であるかの判定を行う．

いくつかの判定項目における結果から総合的な評価を行い，正規利用者の過去の操作履歴と利用者の現在のセッションでの操作履歴を比較することによって正規利用者であるか不正利用者であるかを判別する．ここでのセッションとは，図2のようにログインからログアウトまでに行う一連の作業を指す．

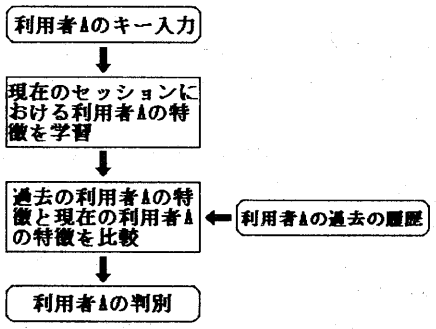


図1 不正利用を判定するまでの流れ

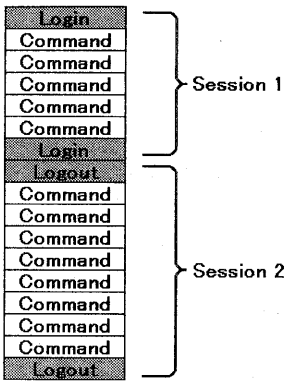


図2 セッション

4. 不正利用者を判別するための判定項目

不正利用者を判別するために判定項目をいくつか設ける．その項目は以下の通りである．

(1) キー入力を行った回数
(2) 入力した文字の修正回数
(3) 入力した文字の種類
（UNIX コマンドと共に使われる特殊記号）
(4) キー入力の速度
(5) 実行したコマンドとその種類
(6) アクセスしたファイルとその種類
(7) 移動先のディレクトリ

ここでは，これらの7つの判定項目を設けることにする．学習時にはこれらの判定項目を更に細分化した判定項目を作成し学習を行う．例えば，判定項目 (2) を細分化する場合,「一文字のみを削除し修正した文字とその回数」や「BS(バックスペース) キーの使用回数」などの項目に細分化する．

そして，利用者のキーボード入力や UNIX の操作に関する熟練度の変化なども考慮しながら，コンピュータが利用者のキーボード入力を読み取り，それを学習する．

5. 判定項目ごとの評価の数値化

不正利用者の判定を行うためには，それぞれの判定項目での評価を数値化しなければならないが，その数値化の方法は，それぞれの判定項目で多少異なる．ここでは評価の度合いを数値化する例として，「入力した文字の修正回数」と「実行したコマンドとその種類」を挙げて説明する．これらの説明のために，以下では次の記号を用いる．

i: セッションの番号
j: セッションにおける判定項目の番号
k: セッションにおける判定項目を細分化した評価項目の番号
N_i: セッションの回数
N_{jk}: 判定項目 j における細分化された判定項目の個数
w_{jk}: 判定項目 jk における重み
X_{ijk}:セッション i における判定項目 jk の評価値
Y_{ij}: セッション i における判定項目 j の評価値
Z_i: セッション i におけるキー入力を行った回数
U_i: セッション i において入力したコマンドの個数

まず,「入力した文字の修正回数」(判定項目番号 $j=1$ とする) のセッション i における評価値 Y_{i1} は次のように定義する．

$$Y_{i1} = \sum_{k=1}^{N_{1k}} \frac{X_{i1k}}{Z_i} w_{1k} \tag{1}$$

次に,「入力するコマンドの種類」(判定項目番号 $j=2$ とする) のセッション i における評価値 Y_{i2} は次のように定義する．

$$Y_{i2} = \sum_{k=1}^{N_{2k}} \frac{X_{i2k}}{U_i} w_{2k} \tag{2}$$

このように，それぞれの判定項目 j での評価値を定義する．そこで，これらの判定項目での評価値からセッション i における評価値は次のように定義する．

$$J_i = \sum_{j=1}^{N_j} Y_{ij} \qquad (3)$$

利用者があるセッションで普段と違う振る舞いをし，キーボード入力に変化があれば，そのセッションの前後で評価値が増減する．

6. 実験

実験は，まず，被験者を3人選択し，それぞれ利用者A，B，Cとする．利用者それぞれのUNIXの使用経験は以下のとおりである．

[利用者A] UNIX使用暦約6年．UNIX上ではメールの読み書き，文章作成，プログラムの作成など様々な作業を行う．システム管理者も兼ねる．

[利用者B] UNIX使用暦約2年．UNIX上ではメールの読み書き，プログラムの作成をUNIX上で行う．

[利用者C] UNIX使用暦約2年．UNIX上ではメールの読み書き，文章作成，プログラムの作成をUNIX上で行う．

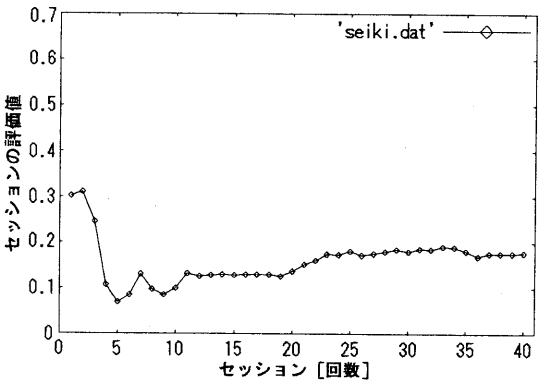
UNIXの操作において，いずれの利用者の場合も，1回のセッションでのコマンドの実行回数の制限は無いものとする．このような条件で，利用者Aが40回のセッションに渡り一人でキーボード入力を行った場合(実験1)と，正規利用者Bに成りすました不正利用者Cが存在することを想定し，ある回数のセッションまで利用者Bがキーボード入力を行った後に，利用者Cが利用者Bと交代してキーボード入力を行った場合(実験2)の2通りのキーボード入力の様子をコンピュータに学習させ，それぞれの場合についての評価を試みる．また，今回の実験で使用する評価項目は，4節で示した評価項目のうち「キー入力の速度」を除外した6つの評価項目を使用する．

7. 実験結果

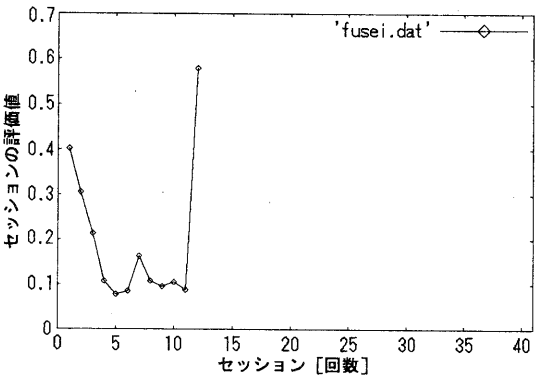
実験結果を図3に示す．まず，図3(a)は実験1において利用者Aの40回にわたるセッションごとのキーボード入力の様子を評価し，それを数値化したグラフである．このグラフでは1回目のセッションで利用者の特徴を学習した時は，コンピュータが利用者Aのキーボード入力の特徴を知らないので評価値は高くなっている．1〜5回目のセッションでは，セッション回数が増すに連れて評価値が減少し，利用者Aの特徴をコンピュータが学習している様子を示している．6〜11回目のセッションでは評価値に増減があるが，これは利用者Aの大まかな特徴を学習したことを示している．そして，12回目以降のセッションでは，それまでのセッションに比べて評価値の変化量が減少しているが，これは利用者の特徴の学習がほぼ完了したことを示している．また，19〜23回目のセッションでの評価値が徐々に増加し，それ以降のセッションでは再び評価値の変化量が減少しているが，これは，19〜23回目のセッションで利用者のキーボード入力に変化があったことを示し，24回目以降でコンピュータが利用者の変化を学習したことを示している．この19〜23回目のセッションの評価値の増加は，利用者Aがこれまでのセッションで実行したことの無いコマンドを実行したためにグラフに変化が現れた．

次に，図3(b)は実験2においてある回数のセッションまで正規利用者Bがキーボード入力を行った後に，不正利用者Cが正規利用者Bの代わりにキーボード入力を行った様子を評価し，それを数値化したグラフである．これは，12回目で利用者Bの代わりに利用者Cがキーボード入力を行った結果である．1〜11回目までのグラフの概形は実験1での11回目までのものと似ており，11回目で利用者Bの特徴の学習がほぼ完了しつつあるのだが，12回目のセッションでは評価値が急激に増加していることが分かる．これは1〜11回目のセッションでキーボード入力を行った利用者と12回目のセッションでキーボード入力を行った利用者が同一人物でないことを示している．つまり，不正利用者Cが正規利用者Bに成りすましてキーボード入力を行ったことを示している．このセッションで評価値が急激に増加した要因は以下のものだった．

(1) 正規利用者Bのセッションで使用回数が少なかったキーバインドを不正利用者Cは多用していた．
(2) 正規利用者Bのセッションで使用したことのないUNIXコマンドを不正利用者Cは多用していた．
(3) 正規利用者Bと不正利用者Cでは入力ミスの回数が多いキーが異なった．



(a) 実験1におけるキーボード入力の評価



(b) 実験2におけるキーボード入力の評価

図3 キーボード入力の評価値の変化

8. おわりに

本研究では，UNIXの利用者のキーボード入力から情報を取得し，利用者の特徴や癖をコンピュータが学習することにより不正利用者の判別を試みる方法について考察および実験を行った．今回の実験では評価項目の一つである「キー入力の速度」を評価項目に加えなかったが，その理由は，利用者が回線速度の遅いサイトを経由してUNIXマシンにログインした場合，キー入力の速度を評価することは困難だからである．「キー入力の速度」を評価項目として加える場合は，利用者のログイン元の所在を明確にし，UNIXマシンとログイン元の間の回線速度を加味する必要がある．当面は，「キー入力の速度」を評価項目から外し，本システムのUNIXマシンへの実装時にこの評価項目について検討を行いたい．また，利用者の特徴を学習するアルゴリズムと不正利用者の判別のための閾値の設定についても検討を行い，さらに精度の高いシステムを目標とする．

参考文献

[1] Security of the Internet, CERT. http://www.cert.org/encyc_article/tocencyc.html

[2] 通信白書11年度版 第1章 特集インターネット 第5節 課題と展望 1. 利用環境整備 (3) 不正アクセス，郵政省 (1998). http://www.mpt.go.jp/policyreports/japanese/papers/99wp/99wp-1-index.html

[3] 不正アクセスの動向，コンピュータ緊急対応センター (Oct.1999). http://www.jpcert.or.jp/nl/99-0004/99-0004-01.html

[4] 情報処理振興事業協会 セキュリティセンター. http://www.ipa.go.jp/SECURITY/index-j.html

[5] Simson Garfinkel, Gene Spafford 共著，山口英 監訳，UNIX セキュリティ，ASCII出版局 (1993).