「マルチメディア，分散，協調とモバイル
(DICOMO2022)シンポジウム」 令和4年7月

アカウントにアクセスするユーザ群の信頼性推定技術に関する検討

大森芳彦[1]　　山下高生[1]

概要： インターネット上で e コマースなどが人々の生活に浸透していく中で，オンラインサービスの不正利用や、ネットワークやサーバへの攻撃による脅威が発生している．このような脅威によるリスクを低減するためには，オンラインサービスのアカウントにアクセスしてきたユーザについて，どの程度の信頼性で確認できているかを推定することで，個々のユーザのアクセス毎にどのような脅威があるのかを把握できることが必要である．さらに，一定期間にアカウントにアクセスしてきたユーザ群のなかで，攻撃の意図を持つユーザがどの程度の割合を占めているかを推定することで，ネットワークやオンラインサービスへの脅威の全体像を把握できることも有効であると考えられる．

本稿では，一定期間にアカウントにアクセスしてきたユーザ群の信頼性について，ネットワークの運用時に観測可能なユーザ群の振る舞いの情報のみを用いて，ユーザ群に占める本人の割合として推定する技術を提案する．提案技術については，ユーザ群に占める本人の割合の推定値の精度を，マルチエージェントシミュレーションにより評価する．その結果，マルチエージェントシミュレーションで得られた本人の割合の推定値は実際の値と近似でき，提案技術は高い精度で推定値を算出できることが分かった．また，提案技術は，従来のような通常認証に成功したユーザを本人とみなした場合では，本人であるユーザエージェント数の割合が 0.9 のときで，通常認証に成功して本人とみなしたアクセス数のなかに，非本人が約 0.39 の割合で含まれていることを推定できた．そのため，ネットワークやオンラインサービスの運用において，通常認証に成功した非本人の割合の増減によって，通常認証の強度を含めたセキュリティ対処の強度を制御することに有効であると考えられることを示した．

A Study on the Method of Scoring Users' Reliability in Authentication

YOSHIHIKO OMORI[1]　　TAKAO YAMASHITA[1]

1. はじめに

インターネット上で e コマース,SNS(Social networking service),インターネットバンキングなどのオンラインサービスが人々の生活に浸透していく中で，フィッシング詐欺，マルウェア攻撃,ソーシャルエンジニアリングなどの方法により，オンラインサービスの不正利用や，ネットワークやサーバへの攻撃による脅威が発生している．このような脅威によるリスクを低減するためには，オンラインサービスのアカウントにアクセスしてきたユーザについて，どの程度の信頼性で確認できているかを推定することで，個々のユーザのアクセス毎にどのような脅威があるのかを把握できることが必要である．さらに，一定期間にアカウントにアクセスしてきたユーザ群のなかで，攻撃の意図を持つユーザがどの程度の割合を占めているかを推定することで，ネットワークやオンラインサービスへの脅威の全体像を把握できることも有効であると考えられる．

このような個々のユーザやユーザ群の信頼性の程度を推定することができれば，ネットワークやオンラインサービスの運用時の脅威の状況にもとづいて，セキュリティの対処の強度を制御することが可能となる．セキュリティの対処の例としては，アカウントにアクセスしてきたユーザが，そのアカウントに紐づいたユーザであることを十分に信頼できない場合には，2 段階認証などの追加の認証を行うことが考えられる．また，一定期間にアカウントにアクセスしてきたユーザ群の信頼性が低下している場合には，アカウントにアクセスするユーザ群全体に対してユーザ認証の強度を高めたり，WAF(Web Application Firewall)でのトラヒックの制限を強めたりすることが考えられる．

ユーザの信頼性の程度を把握して，ユーザの利便性の低下を極力抑えつつ，オンラインサービスをセキュアに提供するための技術に適用できる概念として，UEBA(User Entity Behavior Analytics)が注目されている．UEBA は，個々のユーザの振る舞いに関する情報の収集，分析を行い，正常なユーザの振る舞いをモデル化することで，攻撃者などの異常な振る舞いを検出しようとする概念である．ユーザの振る舞いの情報としては，例えば，オンラインサービスへのアクセス時刻，位置情報，使用端末の種別，回線の種別，マウスやキーボードの操作などがある．UEBA の概念を具体化して，オンラインサービスのアカウントにアクセスしてきたユーザの信頼性を推定する場合，推定した結果はスコア化され，スコアが基準値よりも高い場合には，アカウントに紐づくユーザとは異なる第三者がアクセスしているリスクが高いと判断する．

1 NTT ネットワークサービスシステム研究所

このように，UEBA は，個々のユーザがどの程度の信頼性があるのかをユーザ毎にスコア化することで，ユーザ毎にセキュリティの対処の強度を制御することが可能である．しかし，その一方で，一定期間にアカウントにアクセスしてきたユーザ群からもたらされるネットワークやオンラインサービスへの脅威の全体像については，UEBA はユーザ毎に振る舞いをモデル化していることから，ユーザのスコアをそのまま用いて把握することは困難であると考えられる．

UEBA の他に，ユーザの信頼性の程度を把握するための技術として，ユーザからの攻撃の有無の確率や，ユーザからの攻撃によってもたらされる，ネットワークにとって危険なインシデントの発生確率を推定する方法がいくつか提案されている[1][2][3]．これらの方法では，個々のユーザの振る舞いなどから攻撃の有無について確率で推定したり，ユーザからの攻撃によってネットワークの機能異常が発生する条件付確率などからインシデントの発生確率を推定したりしている．このような確率は，個々のユーザの振る舞いや，ユーザからの攻撃ごとに推定しているため，個々のユーザがどの程度の信頼性があるのかを推定したり，ユーザからの攻撃ごとにリスクを推定したりするために利用することが可能と考えられる．しかし，その一方で，一定期間にアクセスしてきたユーザ群の信頼性や，ユーザ群からの攻撃全体にもとづいたリスクについては推定していないことから，これらの方法については，ネットワークやオンラインサービスへの脅威の全体像の把握にそのまま用いることは困難であると考えられる．

以上を踏まえて，本稿では，ネットワークやオンラインサービスへの脅威の全体像を把握することに着目し，一定期間にアクセスしてきたユーザ群の信頼性の推定を可能にする技術について提案，評価を行う．ユーザ群の信頼性については，攻撃の意図を持つユーザは他人のアカウントにアクセスするものと想定し，一定期間にアカウントにアクセスしてきたユーザ群のなかで，それらのアカウントに紐づいたユーザ(以下，本人と呼ぶ)がどの程度の割合を占めているかを扱う．また，ユーザ群の信頼性の推定には，UEBA や上述した方法[1][2][3]と同様に，個々のユーザの振る舞いに関する情報を用いて，信頼性をスコア化することとする．

本稿の構成について述べる．2 節では，オンラインにおけるユーザの信頼性を推定する従来の技術について説明する．3 節では，2 節で説明した技術について，ネットワークやオンラインサービスへの脅威の全体像の推定に適用する場合の課題について述べる．4 節では，3 節で述べた課題を解決するユーザ群の信頼性推定技術について提案する．5 節では，提案方式の評価を行う．最後に，6 節で本稿のまとめを述べる．

2. 従来の技術

本節では，オンラインにおけるユーザの信頼性を推定する従来の技術として，UEBA の機器の一例および，インシデント予測とリスク評価の提案方式について説明する．説明にあたっては，ユーザの振る舞いに関する情報の使われ方や，ユーザの信頼性のスコア化の方法の観点を中心に行う．

2.1 UEBA を用いたユーザの信頼性推定[4]

UEBA を用いた機器におけるユーザの信頼性推定の一例について，図 1 を用いて説明する．図 1 で，ユーザは，Web サービスや社内システム，その他クラウドサービスなどのアカウントにシングルサインオン(以下，SSO と呼ぶ)によりログインして，これらのサービスを利用する．SSO システムは，ユーザの振る舞いに関する情報を収集して，UEBA 機能にてユーザの本人性のスコアを算出する．ユーザの振る舞いに関する情報としては、1 節で述べたようなユーザのアカウントへのアクセス時刻，使用している端末やデバイスの種別，アクセス時の位置情報など，ユーザの特徴を示す情報がある．

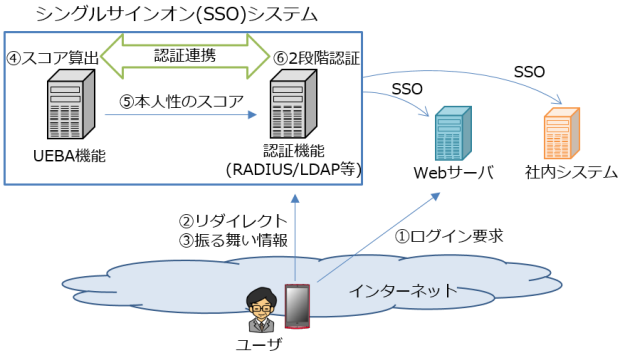

図1 UEBAを用いたユーザの信頼性の推定

ユーザが Web サービスを利用する場合，ユーザは Web サーバにアクセスしてログインを要求する(図 1①)．Web サーバは，ユーザを認証するために，SSO システムにリダイレクトをする(同②)．

SSO システムの UEBA 機能では，ユーザがアカウントにアクセスする度に，そのアカウントに紐づけてユーザの振る舞い情報を収集，保管する(同③)．ユーザがアカウントにアクセスすると，ユーザの振る舞い情報ごとに，過去に同じアカウントにアクセスした時のユーザの振る舞い情報と比較して，その差分をスコア化する(同④)．例えば，過去にアクセスした時の位置情報が国内のみであったのに対し，今回アクセスした際の位置情報が海外の場合，今回アクセスしてきたユーザは，本人性が低く，リスクが高いと推定して，スコアを高くする．その他のユーザの振る舞い情報についてもそれぞれ同様にスコアを算出する．これらのスコアは，UEBA 機能の独自の基準にもとづいて，過去の振る舞いとの差分の大きさによって加算される．このように

ユーザの振る舞い情報ごとに算出されたスコアは，図 2 に示す通り積み上げられて，合算された値がアクセスしてきたユーザの本人性のスコアとなり，ユーザの信頼性の推定値となる．

認証機能では，UEBA 機能と連携して，アクセスしてきたユーザの本人性のスコアにもとづいて，あらかじめ保守者が設定した閾値を超えている場合には，2 段階認証などの厳格な認証を実施した後に，アカウントへのアクセスを許可する(同⑤,⑥)．閾値を超えていない場合には、デフォルトの認証方法でユーザを認証する。

以上のように, UEBA を用いたユーザの信頼性推定では，ユーザがオンラインサービスのアカウントにアクセスする際に，過去の振る舞いとの差分を観測し，ユーザの信頼性として本人性をスコア化して，認証方法に反映している．

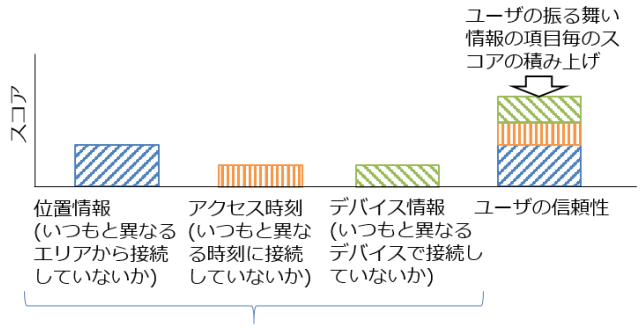

図2 ユーザの信頼性のスコア化

2.2 インシデント予測とリスク評価の提案方式[3]

産業用制御システムを対象にしたサイバーセキュリティのリスク評価の方式がいくつか提案されている．これらの方式では, ネットワークへの攻撃によってもたらされる，ネットワークにとって危険なインシデントの発生確率を算出することでリスクの評価をしており，ユーザの信頼性を確率でスコア化することにも適用可能と考えられる．

この方式におけるサイバーセキュリティのリスク評価には，2 つの方向性がある．一つは，複数の攻撃間の関係と潜在的な攻撃の予測に関するものである．もう一つの方向性は，インシデント間の因果関係にもとづいたリスク評価で, これらのインシデントの発生を予測するものである．

図 4 に，Zhang ら[3]が提案しているリスク評価モデルを示す．このリスク評価モデルには，IDS(Intrusion Detection System)から収集される攻撃のエビデンスと，産業用制御システムの管理装置から収集される機能異常やインシデント発生のアノマリのエビデンスが入力される．まず，攻撃のエビデンスとアノマリのエビデンスが収集されて，各モデルのベイジアンネットワークに入力されると，将来的に起こりうる全てのインシデントについての各々の発生確率が，これらのマルチモデル型のベイジアンネットワークによって算出される．次に，各インシデントについて，それらが発生したときの影響度によってカテゴリ分けと，影響度の数値化を行う．最後に，インシデントの発生確率と，数値化されたインシデントの影響度の積によって，サイバーセキュリティのリスク評価を行う．

この提案手法をユーザの信頼性推定に適用する場合，図 3 のマルチモデル型のアーキテクチャの中の攻撃モデルにおいて，オンラインサービスのアカウントに，そのアカウントに紐づくユーザ以外の第三者がアクセスするような攻撃を対象にする方法が考えられる．この場合，アカウントが乗っ取られることで起こり得るシステムの機能の異常動作，および機能の異常動作によってもたらされる，オンラインサービスにとって危険なインシデントの発生確率を算出する．

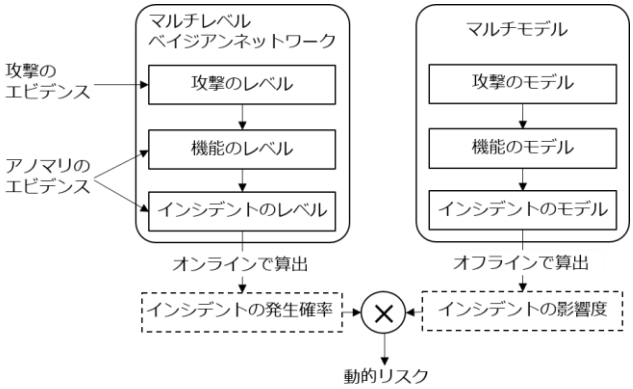

図3 リスク評価モデル[3]

3. 課題

2.1 節で説明した通り, UEBA を用いたユーザの信頼性推定では，ユーザがオンラインサービスのアカウントにアクセスする際の認証時において，ユーザの振る舞いにもとづいて，ユーザの本人性をスコア化することが可能である．しかし，その一方で，スコアは，UEBA 機能の独自基準にしたがって算出したものであり，異なる機器の UEBA 機能間のスコアの互換性はないと考えられる．また，同一機器において，複数のスコア間でユーザの本人性の相対的な比較はできるものの，それらのスコアに対応する各ユーザの本人性の程度を推定することは困難である．そのため，ネットワークやオンラインサービスの保守者は，攻撃者がもたらす脅威が発生した時の影響評価と対応判断に，機器毎に異なるノウハウの修得が必要になるという問題がある．この問題は，ネットワークやオンラインサービスへの脅威の全体像を把握するために，個々のユーザの本人性のスコアを収集，統計処理して，一定期間にオンラインサービスのアカウントにアクセスしてきたユーザ群の本人性の推定に適用しようとした場合にも該当する．

2.2 節で説明したインシデント予測とリスク評価の方式では，第三者によってアカウントが乗っ取られることでもたらされる，オンラインサービスにとって危険なインシデントの発生確率によって，ユーザの信頼性をスコア化することができると考えられる．この方式はベイジアンネット

ワークを用いてインシデントの発生確率を算出することから，アカウントが乗っ取られる確率や，アカウントが乗っ取られた場合にシステムの機能が異常動作する条件付き確率，およびシステムの機能の異常動作が発生した場合にインシデントが発生する条件付き確率などがあらかじめ知っていることが前提となっている．また，図 3 に示したマルチモデルで用いられる各パラメータの値についても既知であることが前提となっている．そのため，ネットワークやオンラインサービスを運用する際には，これらの確率やパラメータの値を入手しておく必要があるが，保守者にとって容易には収集することができない情報であることが想定される．この問題も，上述した問題と同様に，ネットワークやオンラインサービスへの脅威の全体像を把握するために，インシデントの発生確率を収集，統計処理して，一定期間にオンラインサービスのアカウントにアクセスしてきたユーザ群の信頼性の推定に適用しようとした場合にも該当する．

本稿では，ネットワークやオンラインサービスへの脅威の全体像として，ユーザ群の認証時における本人性をスコア化することとし，従来の技術を用いた場合の問題点を踏まえて次の 3 つの課題を設定する．

(T-1) 一定期間にアクセスしてきたユーザ群における本人の割合をスコア化する．

(T-2) ネットワークやオンラインサービスの運用時に観測，収集可能なユーザの振る舞いにもとづいて，ユーザの本人性をスコア化する．

(T-3) ユーザの振る舞い情報の収集には，ユーザの利便性の低下を抑止する．

4. 課題を解決する提案技術

本節では，3 節で設定した 3 つの課題を解決するためのユーザの信頼性評価技術について提案する．

4.1 課題解決へのアプローチ

オンラインサービスのアカウントにアクセスするユーザの本人性を算出するために，ユーザの認証時において図 4 に示す処理を行う．図 4 で，ユーザにアカウントの利用を許可するための手続きとして世の中で広く用いられているユーザ認証(以下，通常認証と呼ぶ)に加えて，新たな使用方法としての認証(以下，超強力認証と呼ぶ)を導入する．この超強力認証は，アカウントにアクセスしてきたユーザ群のなかから，ランダムにサンプリングして行う．

超強力認証は，アカウントにアクセスするユーザ群のなかで本人ではないユーザの割合を推定するために行うユーザ認証である．この超強力認証は，非本人がなりすましてユーザ認証に成功することを，推定値の精度に関するネットワークやオンラインサービスの運用条件の許容範囲にもとづく確率で，ほぼ抑止することを可能とする．そのため，超強力認証には，スマートカードなど，多要素かつ暗号学的に十分な強度を持つ認証方法を用いる．

超強力認証は，高いセキュリティを有する反面，そのトレードオフでユーザの利便性の相対的な低さのために，本人がユーザ認証に失敗することが想定される．そこで，本人が超強力認証に失敗することを極力抑えるために，超強力認証方法を複数併用し，アクセスするアカウントに紐づくユーザに関する振る舞い情報にもとづいて，本人であれば認証失敗することを極力抑えることができる超強力認証方法を選択することとする．

アクセスするアカウントに紐づくユーザに関する振る舞い情報の例として，次のような情報が考えられる．

(B-1) アクセスするアカウントに紐づくユーザが，IC カード機能を持つ社員証でオフィスのドアを開いて入室している(図 4).

(B-2) アクセスするアカウントに紐づくユーザが，直近にモバイル端末からアカウントにアクセスしていた．

(B-1)の場合，超強力認証として，社員証を用いた方法を選択することで，本人であれば認証失敗することをほぼ抑止できると考えられる．また，(B-2)の場合であれば，アカウントにアクセスするために，あらかじめ登録済みのモバイル端末へのショートメッセージを用いた超強力認証方法を選択することが考えられる．

このようなユーザの振る舞いに関する情報の収集には，ユーザの端末とオンラインサービスのサーバ間のチャネルの他にも，(B-1)の場合のように，入退室管理のサーバと連携して，アウトオブバンドからも収集する．

以上の通常認証と超強力認証を用いて，一定期間にオンラインサービスのアカウントにアクセスしてきたユーザ群に占める本人の割合を算出する．

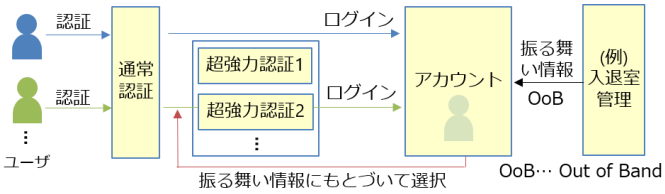

図4 ユーザの本人性を算出するための処理

4.2 提案技術の前提

通常認証と超強力認証を用いて，一定期間にオンラインサービスのアカウントにアクセスしてきたユーザ群に占める本人の割合を算出するためには，本人と非本人ごとに，通常認証と超強力認証の成否の関係を明らかにする必要がある．本節では，そのための前提について説明する．

世の中で広く使われている通常認証は，一般的に，セキュリティとユーザの利便性のトレードオフを最適化するように，その認証強度を選択して採用されている．そのため，通常認証に成功したユーザの中には，非本人も含まれることが想定される．また，本人であっても，ユーザ認証時の環境によっては，通常認証に失敗することが想定される．

超強力認証については，前述した通り，非本人が認証成功することをほぼ抑止する．

以上を踏まえて，本人と非本人における，通常認証と超強力認証の成否に関する前提を以下に述べる．また，これらの前提を踏まえた，通常認証と超強力認証の成否の関係を図 5 に示す．

(前提 1) 選択した超強力認証方法では，本人性の推定値の精度に関するネットワークやオンラインサービスの運用条件にもとづいて，許容される確率で非本人は認証失敗し，本人は認証成功する．

(前提 2) 一定期間にオンラインサービスのアカウントにアクセスしてきた全ユーザ(母集団)から，十分な数のランダムにサンプリングしたユーザ群において，次の割合は，母集団でのそれらの割合と近似できる．

(Ⅰ) 通常認証に成功する本人（図 6 $a_1 \fallingdotseq a_2$）
(Ⅱ) 通常認証に失敗する本人（図 6 $c_1 \fallingdotseq c_2$）
(Ⅲ) 通常認証に失敗する非本人（図 6 $b_1 \fallingdotseq b_2$）
(Ⅳ) 通常認証に失敗する非本人（図 6 $d_1 \fallingdotseq d_2$）

前提 1 については，前述した通り，複数の超強力認証方法のなかから，アクセスするアカウントに紐づくユーザの振る舞い情報から最適な超強力認証方法を選択することで，実際のインターネット上でも成り立つものとする．

前提 2 については，超強力認証を行うユーザのサンプリングの数を増やすことで，標本における前提 2 の各割合の期待値は，それぞれ母平均に近づいていく．そのため，十分な数のサンプリングをすることで，標本における前提 2 の各割合は，母集団でのそれらの割合と近似的に等しくなると仮定する．

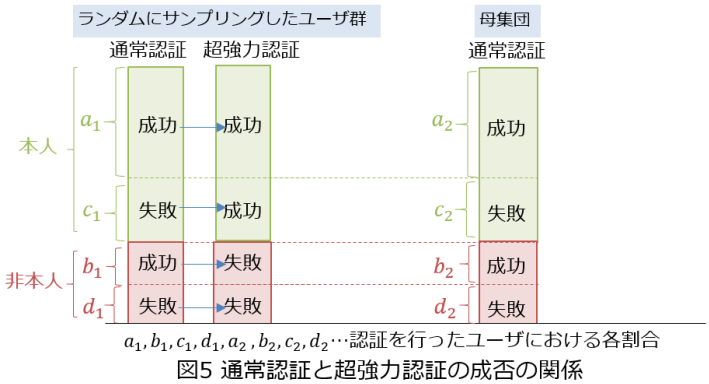

図5 通常認証と超強力認証の成否の関係

4.3 提案技術の方式

本稿で提案するユーザの本人性のスコア化の方式について，図 4 と図 6 を用いて説明する．

図 4 で，ユーザがオンラインサービスのアカウントにアクセスする場合，ユーザは，まず通常認証を実施する．その後，ランダムにサンプリングされたユーザは超強力認証を実施する．ここで，図 6 に示す通り，一定期間にアクセスしてきたユーザ全体(母集団)のなかで，通常認証に成功した本人の割合をa_2，通常認証に失敗した本人の割合をc_2，通常認証に成功した非本人の割合をb_2，通常認証に失敗した非本人の割合をd_2とする．そうすると，母集団のなかで本人が占める割合をスコア化するためには，a_2，b_2，c_2，d_2を求めればよい．非本人のうち，通常認証に成功した割合をβ_2，母集団のうち，非本人の割合をγ_2とすると，次の式が成り立つ．

$$\beta_2 = \frac{b_2}{b_2 + d_2} \tag{1}$$

$$\gamma_2 = \frac{b_2 + d_2}{a_2 + b_2 + c_2 + d_2} \tag{2}$$

また，母集団のうち，通常認証に成功した割合を$N_2^{(a+b)}$，通常認証に失敗した割合を$N_2^{(c+d)}$とすると，次の式が成り立つ．

$$N_2^{(a+b)} = a_2 + b_2 \tag{3}$$

$$N_2^{(c+d)} = c_2 + d_2 \tag{4}$$

これらの$N_2^{(a+b)}$と$N_2^{(c+d)}$については，通常認証の成否を観測することで求めることができる．

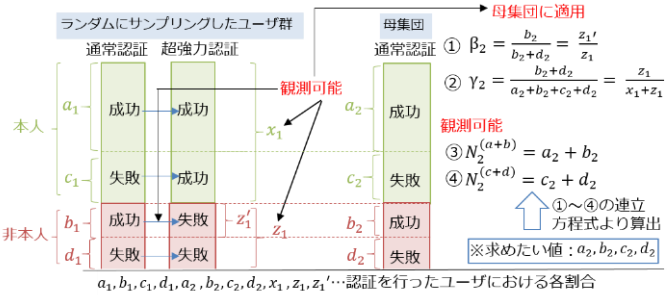

図6 本人であるユーザの割合の算出方法

超強力認証を行ったユーザのなかで，通常認証に成功した本人の割合をa_1，通常認証に失敗した本人の割合をc_1，通常認証に成功した非本人の割合をb_1，通常認証に失敗した非本人の割合をd_1とする．また，超強力認証に成功した割合をx_1，超強力認証に失敗した割合をz_1，通常認証に成功し，かつ超強力認証に失敗した割合をz_1'とする．そうすると，x_1，z_1，z_1'は，a_2，b_2，c_2，d_2，および 4.2 節の前提を用いて次の式で近似できる．

$$x_1 \fallingdotseq a_1 + c_1 \fallingdotseq a_2 + c_2 \tag{5}$$

$$z_1 \fallingdotseq b_1 + d_1 \fallingdotseq b_2 + d_2 \tag{6}$$

$$z_1' \fallingdotseq b_1 \fallingdotseq b_2 \tag{7}$$

式(1)と式(2)に式(5)，式(6)，式(7)を代入すると，β_2とγ_2は次の式で近似できる．

$$\beta_2 = \frac{b_2}{b_2 + d_2} \fallingdotseq \frac{b_1}{b_1 + d_1} \fallingdotseq \frac{z_1'}{z_1} \tag{8}$$

$$\gamma_2 = \frac{b_2 + d_2}{a_2 + b_2 + c_2 + d_2} \fallingdotseq \frac{b_1 + d_1}{a_1 + b_1 + c_1 + d_1} \fallingdotseq \frac{z_1}{x_1 + z_1} \tag{9}$$

x_1，z_1は超強力認証の成否を観測することで求めることができる．z_1'についても通常認証と超強力認証の成否を観測することで求めることができる．したがって，β_2とγ_2は，通常認証と超強力認証の成否を観測することで求めることができる．

以上より，$N_2^{(a+b)}$と$N_2^{(c+d)}$および，β_2とγ_2は，通常認証と超強力認証時に観測，収集することが可能であることから，式(3)，式(4)，式(8)，式(9)は，近似的に，a_2，b_2，c_2，d_2の4 元連立方程式と考えることができる．この連立方程式からa_2，b_2，c_2，d_2は，次のように求められる．

$$a_2 = N_2^{(a+b)} - \beta_2\gamma_2 \quad (10)$$

$$b_2 = \beta_2\gamma_2 \quad (11)$$

$$c_2 = N_2^{(c+d)} - (1-\beta_2)\gamma_2 \quad (12)$$

$$d_2 = (1-\beta_2)\gamma_2 \quad (13)$$

式(10)，式(11)，式(12)，式(13)から，次の各ユーザ群において，本人であるユーザの割合を求めることができる．

[一定期間に通常認証に成功したユーザ群]

本人である割合P_s：$P_s = \dfrac{a_2}{a_2+b_2} = \dfrac{N_2^{(a+b)} - \beta_2\gamma_2}{N_2^{(a+b)}}$ (14)

[一定期間に通常認証に失敗したユーザ群]

本人である割合P_f：$P_f = \dfrac{c_2}{c_2+d_2} = \dfrac{N_2^{(c+d)} - (1-\beta_2)\gamma_2}{N_2^{(c+d)}}$ (15)

[一定期間にアクセスしてきた全ユーザ]

本人である割合P_t：$P_t = \dfrac{a_2+c_2}{a_2+b_2+c_2+d_2} = 1-\gamma_2$ (16)

これらの式(14)，式(15)，式(16)を用いることにより，本人であるユーザの割合を推定して，ユーザの本人性をスコア化することができる．

以上より，提案技術は，ネットワークやオンラインサービスの運用時に観測，収集可能な通常認証および超強力認証の結果を用いて，一定期間にアクセスしてきたユーザ群における本人の割合をスコア化することができる．また，ユーザの振る舞いに関する情報にもとづいて，本人であれば認証成功する確率が十分高い，延いては相対的にユーザの利便性が高いと考えられる超強力認証をサンプリングしたユーザのみ実施していることから，3 節で設定した 3 つの課題(T-1)～(T-3)を解決することができる．

5. 評価

本節では，4 節で提案したユーザの本人性のスコア化の方式について評価を行う．

5.1 評価項目

提案技術では，一定期間にオンラインサービスのアカウントにアクセスしてきた全ユーザにおける通常認証の成否の情報，およびサンプリングしたユーザにおける超強力認証の成否の情報を用いて，統計学的な手法でユーザの本人性をスコア化している．そこで，通常認証と超強力認証を行うサンプル数，およびユーザ全体に占める本人の割合と，ユーザの本人性のスコアの精度の関係について評価を行う．

また，超強力認証については，4.2 節で説明した前提 1 より，本人は十分高い確率で認証成功することとしている．しかし，超強力認証は，ユーザの振る舞いにもとづいて認証失敗することを極力抑えることができる強力認証方法を選択して行うものの，実際のインターネット上では，ユーザの位置情報や利用端末などの利用環境によっては，本人であっても認証失敗するケースが考えられる．本人が超強力認証に失敗するケースが増えると，式(14)，式(15)，式(16)から求められるユーザの本人性のスコアが，実際の本人であるユーザの割合から乖離することから，スコアの精度に影響が出てくると考えられる．そこで，実際のインターネット上でのオンラインサービスの利用を踏まえて，超強力認証で本人が認証失敗する確率と，ユーザの本人性のスコアの精度の関係について評価を行う．

5.2 評価方法

ユーザの振る舞いは多種多様で複雑であることから，それらの振る舞いを評価に反映させるために，本稿では，マルチエージェントシミュレーションによる評価を行った．

マルチエージェントシミュレーションは，図 7 に示す通り，複数のエージェントがそれぞれ独自の行動モデルを持ち，周囲の状況にもとづいて，一定のルールに従って自律的に動作する．また，各エージェントが，他のエージェントや系と相互作用することで，多数の要素が絡み合った複雑系としての振る舞いをシミュレーションすることができる．

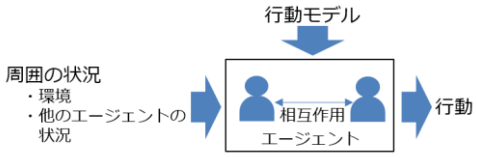

図7 エージェントの動作イメージ

本稿では，エージェントとして，大きく，ユーザ，認証システム，スコア算出システム，アウトオブバンド振る舞い情報システムを定義した．スコア算出システムは，ユーザ群に占める本人の割合を算出するエージェントである．アウトオブバンド振る舞い情報システムは，4.1 節で例示した入退室管理サーバからユーザの入退室の情報を収集するような，ユーザの端末とオンラインサービスのサーバ間のセッション以外からユーザの振る舞い情報を収集するエージェントである．

エージェント間の関係について図 8 を用いて説明する．ユーザエージェントには，本人であるユーザエージェントと非本人であるユーザエージェントを定義する．本人であるユーザエージェントは，ネットワークの利用おいて，①主に昼間に利用するユーザ，②主に夜間に利用するオフィス勤務のユーザ，③主に夜間に利用する在宅勤務のユーザの 3 種類のエージェントを定義し，それぞれの振る舞いを図 9 に示すような時間割に従うものとした．図 9 で，各ユーザのネットワーク(NW)の利用時間帯と位置情報については，ユーザの振る舞いに揺らぎを持たせるために，0.9 の確率で時間割にしたがった振る舞いをするものとし，0.1 の確率では補集合の事象となるようにした．例えば，主にネ

ットワークを利用する時間帯では，0.9 の確率でネットワークを利用し，0.1 の確率で利用しないとした．位置情報についても，自宅か外出先かを確率で揺らぎを持たせた．非本人であるユーザエージェントについては，リバースブルートフォース攻撃やリスト攻撃などを行うことを想定し，突発的に一度に集中してオンラインサービスにアクセスすることとした．

認証システムエージェントは，各ユーザからのアクセス要求に対して，通常認証と超強力認証を行う．超強力認証はランダムにサンプリングしたユーザに行う．超強力認証は，スコア算出システムエージェントから通知された方法を選択して行う．

スコア算出システムエージェントは，認証システムエージェントから各認証結果を収集して，ユーザの本人性のスコアを算出する．また，アウトオブバンド振る舞い情報システムエージェントから，アクセスするアカウントに紐づいたユーザの振る舞い情報を収集して，本人であれば十分高い確率で認証成功する超強力認証方法を選択して，認証システムエージェントに通知する．

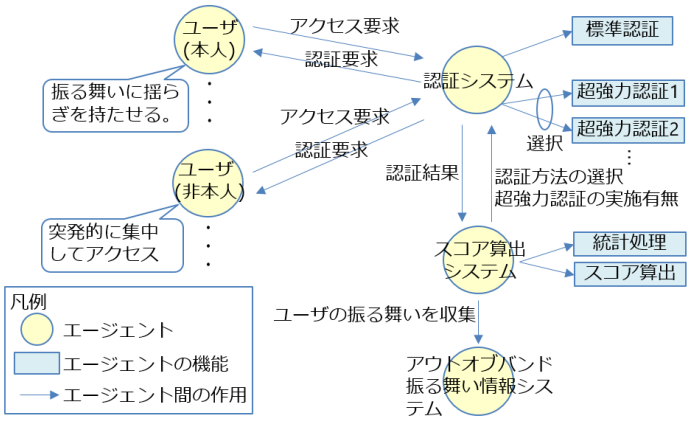

図8 エージェント間の関係

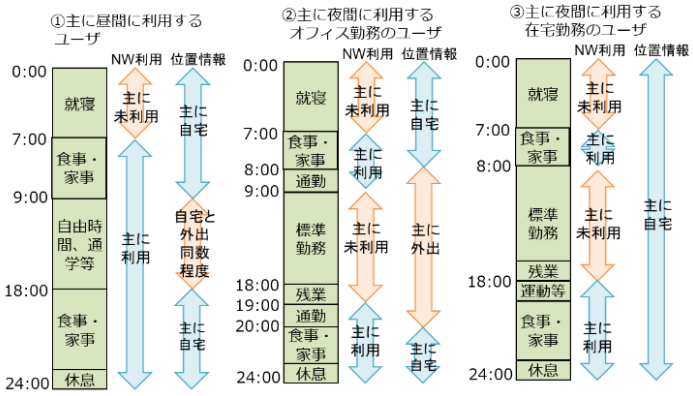

図9 本人であるユーザエージェントの種類ごとの時間割

5.3 評価条件

各エージェントの動作パラメータと，本稿で用いたパラメータ値を表 1 と表 2 に示す．

表 1 の No.4,5,6 における本人は，5.2 節で説明した，本人であるユーザエージェントの 3 種類にそれぞれ対応する．また，その値(割合)は，総務省統計局が発行している人口推計[5]および労働力調査[6]に記載の年齢別の人口と就業状態、および総務省が発行している 2021 年度の情報通信白書[7]に記載のインターネットの利用率と在宅勤務の普及状況のデータを用いて算出した．本人および非本人の各ユーザエージェントのアクセス頻度については，2021 年度の情報通信白書に記載されている 1 時間ごとのインターネットの利用率，および図 9 の時間割，および 5.2 節で説明した表 1 の No.7 にしたがって，1 時間ごとのアクセスする確率値として設定した．表 1 の No.8,9,10 における非本人が一度に行う集中アクセス数については，非本人の各ユーザエージェントがアクセスした 1 時間のアクセス数を，集中アクセスの都度，1 から 3600 回の範囲でランダムに設定し，その期待値を 1800 回とした．

表 2 の No.2 の非本人が通常認証に成功する確率については，例えば 4 桁の PIN(Personal Identification Number)をリトライも含めて 3 回ランダムに入力する場合などで，偶然認証成功することを想定した．また，表 2 の No.4,5 の本人が超強力認証に成功する確率については，今回は基本評価として，位置情報に依らず同一とした．

表1 ユーザエージェントのパラメータ

No.	ユーザエージェントの動作パラメータ	値
1	ユーザエージェント数	1億人
2	本人であるユーザエージェントの割合	0.9999
3	非本人であるユーザエージェントの割合	0.0001
4	本人のうち，主に昼間に利用するユーザエージェント(①)の割合	0.53
5	本人のうち，主に夜間に利用するオフィス勤務のユーザエージェント(②)の割合	0.38
6	本人のうち，主に夜間に利用する在宅勤務のユーザエージェント(③)の割合	0.09
7	本人のユーザエージェントが図10の時間割にしたがった振る舞いをする確率	0.9
8	非本人のユーザエージェントが一度に行う集中アクセス数の期待値	1800アクセス
9	非本人のユーザエージェントが一度に行う集中アクセス数の最大値	3600アクセス
10	非本人のユーザエージェントが一度に行う集中アクセス数の最小値	1アクセス

表2 認証システムエージェントのパラメータ

No.	認証システムエージェントの動作パラメータ	値
1	本人が通常認証に成功する確率	0.95
2	非本人が通常認証に成功する確率	0.003
3	ユーザに超強力認証を行う割合	0.1
4	本人が超強力認証に成功する確率(自宅にいる場合)	0.95
5	本人が超強力認証に成功する確率(自宅にいない場合)	0.95
6	非本人が超強力認証に成功する確率	0

5.4 評価結果

5.1 節で説明した評価項目ごとに，ユーザの本人性のスコアの精度の評価結果を示す．今回の評価では，ユーザ群全体のアクセス数における本人の割合と，通常認証に成功したアクセス数における本人の割合の推定値の精度を評価する．比較対象として，5.3 節で説明した評価条件におけるこれらの割合の実際の値と，従来の通常認証のみにおける認証成功したユーザの割合を用いる．

5.4.1 通常認証と超強力認証を実施したアクセス数とユーザの本人性のスコアの精度の関係

図 10 に，超強力認証を実施したアクセス数とユーザの本人性のスコアの精度の評価結果を示す．通常認証を実施したアクセス数については，表 2 の No.3 より，超強力認証を実施したアクセス数の 10 倍になる．

図 10 の評価では，図 9 の時間割を約 6 日分実施したが，ユーザ群全体のアクセス数における本人の割合は，3 日程度で実際の値にほぼ近似した．また，通常認証に成功したアクセス数における本人の割合は，数時間程度で実際の値にほぼ近似した．このことから，数時間あるいは 3 日程度の範囲での非本人の割合の変動を観察できると考えられる．

また，従来の通常認証のみにおける認証成功したユーザの割合と実際の値との差は，約 0.04 であった．ユーザ群全体のアクセス数における本人の割合の推定値と実際の値との差は約 0.003 であり，従来の通常認証のみと比べて精度が向上している．

なお，図 10 では，ユーザ群全体のアクセス数における本人の割合が，推定値と実際の値ともに，表 1 の No.2 の本人であるユーザエージェントの割合よりも小さい．この割合の相違は，非本人の各ユーザエージェントがアクセスする場合には，一度に集中して平均 1800 回アクセスすることから，非本人であるユーザエージェントの割合よりも，非本人のアクセス数の割合が大きくなるためである．

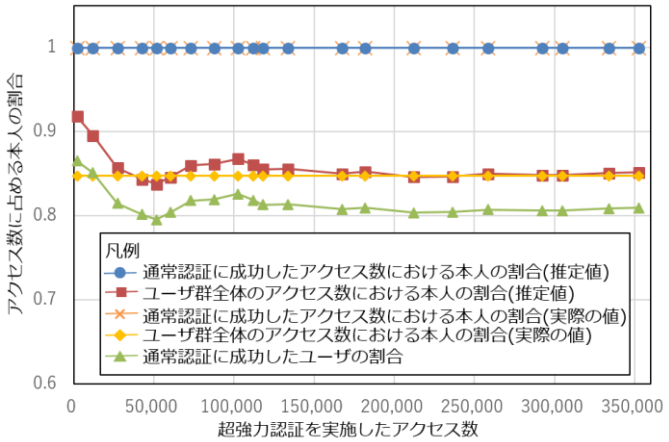

図10 超強力認証を実施したアクセス数とユーザの本人性のスコアの関係

5.4.2 ユーザ全体における本人の割合とユーザの本人性のスコアの精度の関係

図 11 に，ユーザ全体における本人の割合とユーザの本人性のスコアの精度の評価結果を示す．今回の評価は，全ユーザエージェント数のうち，本人であるユーザエージェント数の割合を 0.9 から 0.99 の範囲で行った．

図 11 より，ユーザ群全体のアクセス数における本人の割合，および通常認証に成功したアクセス数における本人の割合ともに，今回評価した本人であるユーザエージェント数の割合の範囲において，推定値と実際の値との差が 0.001 以下であり，推定値は実際の値とほぼ同じであった．

通常認証に成功したアクセス数における本人の割合の推定値については，本人であるユーザエージェント数の割合が 0.9 のときで約 0.61 であった．同様に，従来の通常認証のみにおける認証成功したアクセス数の割合は約 0.008 であった．従来のような通常認証のみにおける認証成功したユーザを本人とみなした場合では，通常認証に成功して本人とみなした約 0.8%のアクセス数のなかに，約 0.39 の割合で非本人が含まれていると推定することができる．そのため，提案技術は，ネットワークやオンラインサービスの運用において，通常認証に成功した非本人の割合の増減によって，通常認証の強度を含めたセキュリティ対処の強度を制御することに有効であると考えられる．

なお，図 11 では，ユーザ群全体のアクセス数における本人の割合，および通常認証に成功したアクセス数における本人の割合が，推定値と実際の値ともに，図 10 と比べると少ない．特に，ユーザ群全体のアクセス数における本人の割合は，推定値と実際の値ともに 0.06 以下である．これらの割合が少ないのは，図 10 と比べて，非本人であるユーザエージェントの割合が多く，その集中アクセス数の影響が大きいためである．

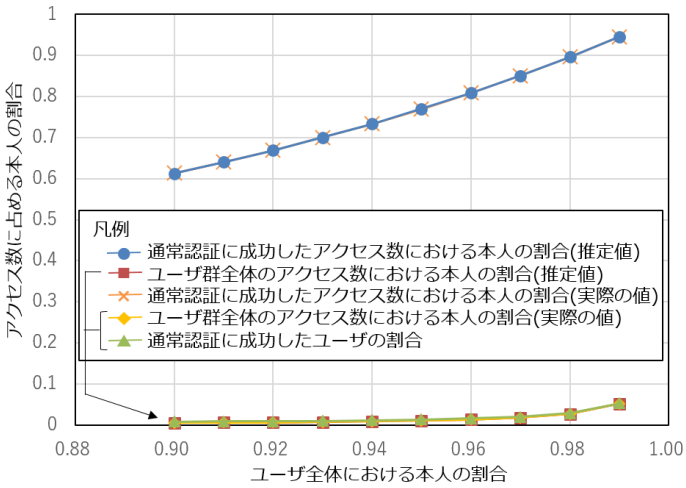

図11 全ユーザ数に占める本人の割合とユーザの本人性のスコアの関係

5.4.3 超強力認証に本人が認証失敗する確率とユーザの本人性のスコアの精度の関係

図 12 と図 13 に，超強力認証に本人が認証失敗する確率とユーザの本人性のスコアの精度の評価結果を示す．今回の評価は，超強力認証に本人が認証失敗する確率を 0 から 0.1 の範囲で行った．また，超強力認証に本人が認証失敗する確率が 0 のときのユーザの本人性のスコアが，実際にユーザが本人である割合となる．

図 12 より，ユーザ群全体のアクセス数における本人の割合，および通常認証に成功したアクセス数における本人の割合ともに，超強力認証に本人が認証失敗する確率が高くなるにしたがい，ユーザ群の本人の割合は小さく推定されて，精度が低下する．提案技術は，4.2 節で説明した前提 1 が成り立つことが必要であり，ネットワークやオンライ

ンサービスの運用において，ユーザ群の本人の割合の推定値の誤差の要件を踏まえて，本人が認証失敗する確率の許容範囲を満たす超強力認証方法を用いなければならない．例えば，図 13 で，ユーザ群全体のアクセス数における本人の割合の推定値の誤差を 0.01 程度の範囲にすることを想定した場合，超強力認証に本人が失敗する確率を 0.01 程度に抑える超強力認証方法が必要となる．このように，ユーザの振る舞い情報から最適な超強力認証を選択することが，提案技術のポイントになると考えられる．

通常認証に成功したアクセス数における本人の割合について，上述した通り，超強力認証に本人が認証失敗する確率が 0.1 のときで約 0.899 であった．同様に，従来の通常認証のみにおける認証成功したアクセス数の割合は約 0.8 であった．5.4.2 節と同じく，従来のような通常認証のみにおける認証成功したユーザを本人とみなした場合では，通常認証に成功して本人とみなした約 80%のアクセス数のなかに，約 0.101 の割合で非本人が含まれていると推定することができる．

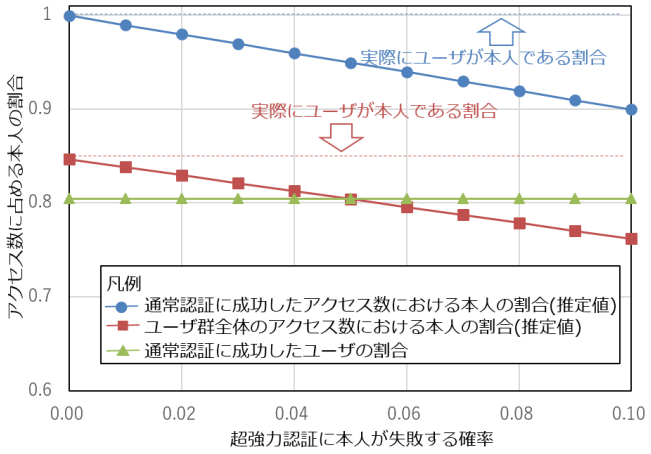

図12 超強力認証に本人が失敗する確率とユーザの本人性のスコアの関係

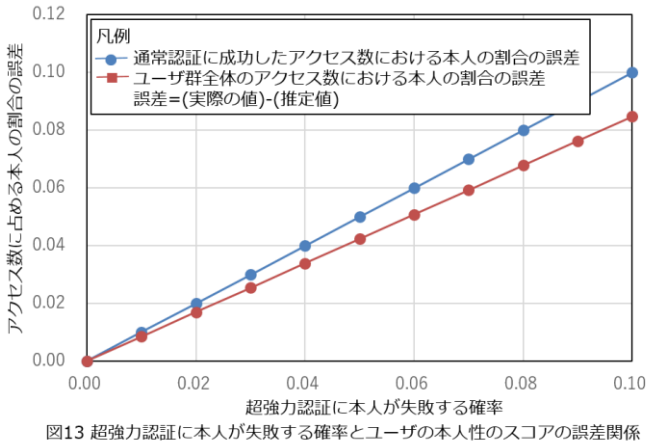

図13 超強力認証に本人が失敗する確率とユーザの本人性のスコアの誤差関係

6. まとめ

本稿では，インターネット上で e コマースなどのオンラインサービスを提供するにあたり，ネットワークやオンラインサービスへの脅威の全体像として，ユーザ群の信頼性を，一定期間にアカウントにアクセスしてきたユーザがアカウントに紐づくユーザと同一である割合(本人性)として推定，スコア化するための技術の提案，評価を行った．

提案技術では，ユーザ群の本人性を推定するにあたり，従来の UEBA の概念を適用した場合や，ネットワークにとって危険なインシデントの発生確率を適用した場合の問題点を踏まえて，3 節で述べた次の 3 つの課題を解決した．

(T-1) 一定期間にアクセスしてきたユーザ群における本人の割合をスコア化する．

(T-2) ネットワークやオンラインサービスの運用時に観測，収集可能なユーザの振る舞いにもとづいて，ユーザの本人性をスコア化する．

(T-3) ユーザの振る舞い情報の収集には，ユーザの利便性の低下を抑止する．

評価では，提案方式を用いて推定できるユーザの本人性として，ユーザ群全体のアクセス数における本人の割合，および通常認証に成功したアクセス数における本人の割合の推定値の精度について，マルチエージェントシミュレーションを行った．マルチエージェントシミュレーションでは，これらの割合の実際の値と，従来の通常認証のみにおける認証成功したユーザの割合を比較対象とし，シミュレーションで得られた推定値は実際の値とほぼ同じであった．また，従来のような通常認証のみにおける認証成功したユーザを本人とみなした場合では，本人であるユーザエージェント数の割合が 0.9 のときで，通常認証に成功して本人とみなした約 0.8%のアクセス数のなかに，非本人が約 0.39 の割合で含まれていると推定することができた．そのため，提案技術は，ネットワークやオンラインサービスの運用において，通常認証に成功した非本人の割合の増減によって，通常認証の強度を含めたセキュリティ対処の強度を制御することに有効であると考えられる．さらに，超強力認証に本人が認証失敗する確率が高くなるにしたがって，ユーザの本人の割合の推定値の精度は低下し，その確率が 0.1 の場合では，実際にユーザが本人である割合に比べて約 0.1 低くなった．したがって，超強力認証に本人が認証失敗する確率を低減させるために，ユーザの振る舞い情報から最適な超強力認証を選択することが，提案技術のポイントになると考えられる．

今後は，ユーザの振る舞い情報にもとづいて最適な超強力認証を選択することで，本人が認証失敗することを抑止できる方法の詳細検討と評価を進めていく．

参考文献

[1] Xie, Y. and Yu, S.Z.: A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Behaviors, IEEE/ACM Transaction on Networking, Vol.17, No.1, pp.54-65(2009).

[2] Bravo, S. and Mauricio, D.: DDoS Attack Detection Mechanism in the Application Layer Using User Features, Proc. IEEE International Conference on Information and Computer Technolo

gies, pp.97-100(2018).
[3] Zhang, Q., Zhou, C., Xiong, N., et al.: Multimodel-Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems, IEEE Transactions on Systems, Man, and Cybermetrics: Systems, vol.46, No.10, pp.1429-1444(2016)
[4] OneLogin: SmartFactor AuthenticationTM, OneLogin(online), available from https://www.onelogin.com/product/smatyfactor-authentication (accessed 2022-05-12)
[5] 総務省統計局：人口推計 令和３年２月報 (stat.go.jp), 総務省統計局(オンライン), https://www.stat.go.jp/data/jinsui/pdf/202102.pdf (参照 2022-05-12)
[6] 総務省統計局：労働力調査（基本集計）2021 年（令和３年）平均結果の要約，概要，統計表等 (stat.go.jp), 総務省統計局(オンライン), https://www.stat.go.jp/data/roudou/sokuhou/nen/ft/pdf/index1.pdf (参照 2022-05-12)
[7] 総務省：令和 3 年版 情報通信白書，総務省(オンライン)，https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/pdf/index.html（参照 2022-05-12）.