言語圏ごとのパスワード生成・管理の傾向比較

森　啓華[1]　シュウ　インゴウ[1]　森　達哉[1]

概要：パスワードを対象とした研究は英語をベースとした研究事例が多く，非英語圏におけるパスワードの研究事例はきわめて少ない．本研究の目的は，言語圏の違いがパスワードの生成や管理方法にいかなる影響を与えるかを理解することである．そのためのアプローチとして，クラウドソーシングサービスを用い，中国語圏，日本語圏，英語圏の代表的な例として中国，日本，英国のユーザに対してオンラインサーベイを行い，パスワードに関する教育歴や知識，パスワード生成方法，管理方法を調査した．各言語圏のユーザにアプローチするために，中国語圏ユーザには Sojump を，日本語圏ユーザには Lancers を，英語圏ユーザには Amazon Mechanical Turk をそれぞれ用いた．サーベイの結果，各国において 80%以上のユーザがパスワードマネージャーを使わず，自分自身でパスワードを生成していること，パスワードを考える方法やベースにする単語の選び方に言語圏ごとの特徴的な差異があることを明らかにした．またパスワードの再利用率やパスワードマネージャーの利用率についても国ごとに差があることを示した．オンラインサーベイで得られた知見を確かめるために，過去に漏洩した大規模なパスワードコーパスの解析を行い，ユーザスタディの結果と比較した．この結果，中国のユーザは日付を，英国のユーザは動物関連の単語を好むという点においてユーザスタディと一致した．

キーワード：パスワード，セキュリティ，ユーザスタディ，言語圏

1. はじめに

パスワードは人間の記憶を使った認証方式であり，生体認証や多要素認証等で用いられる外部端末が不要である．パスワードによる認証は高い利便性を提供する一方で，様々な欠陥があることが知られている．しかしながらパスワード認証は様々なシステムで幅広く採用されており，今後も長期に渡って使われていくことが予想される．

パスワードが抱える技術的な問題は，覚えやすさと安全性の両立が難しい点にあり，このトレードオフは人間が持つ能力の限界に起因している．すなわちブルートフォース攻撃，辞書攻撃，クレデンシャルスタッフィング攻撃等のパスワードに対する攻撃への耐性を高めるためには，サイト毎に複雑なパスワードを使いわけることが必要であるが，そのような条件は人間の記憶の限界を超えてしまう．人間の能力限界に起因する問題を解決する手段として，パスワードマネージャーによるパスワードの生成・管理が推奨されている．

パスワードマネージャーはユーザビリティとセキュリティを両立する有望な技術であるが，その普及は進んでいない．後述する我々のオンラインサーベイの結果によると，パスワードマネージャーを使ってパスワードを生成しているユーザの割合は中国が 10.3%，日本が 2.71%，英国が 16.4%であり，高くないことがわかる．したがって，今日大多数のユーザは自らが考えたパスワードを使っている．

ユーザが生成するパスワードには一定の規則があり，多くの場合で予測可能であることが知られている．例えば人の名前，特定の日付，歌詞，映画，小説，テレビなどから得られた単語をパスワードに設定する傾向や，キーボードの配置パターンに基づくパスワード (“zxcvbn” など) が使われる [1]．これらの既存研究では言語として英語が使われることを前提とした解析が行われている．パスワードとして利用可能な文字列は ASCII の部分集合で構成されることが主であるが，ユーザが使っている言語は様々である．したがって，英語を前提とした解析から得られる知見が他の言語圏にも適合するかは定かではない．

非英語圏を対象としたパスワードに関する研究事例は少ない．Zhigong Li ら [2] は中国のウェブサイトから漏洩したパスワードデータを解析し，中国のユーザは数字を好むこと，またピンインを使用するユーザが存在することを示した．また，中国のユーザの傾向を考慮し，ピンインを辞書に含めることで，パスワード推測の成功確率が高まることを示した．森ら [3] は様々なサイトから漏洩したパスワー

[1] 早稲田大学

ドのうち，中国，日本，英国に属すると考えられるユーザが生成したパスワードを解析した．その結果，各国のユーザがパスワードに使う文字種や．日付の年月日に関連する数字の使い方等に関して，文化の差異が存在することを示した．愛ら [4] は漏洩パスワードデータとユーザスタディを組み合わせて日本人が付けるパスワードを解析した．その結果，日本語パスワードは他言語のパスワードと比較して，人名を利用したパスワードや単純な言葉の繰り返し表現が使われることが多いことを示した．

本研究は言語圏の違いがパスワードの生成や管理方法にいかなる影響を与えるかを明らかにすることを目的として，3 つの異なる言語圏として中国語圏，日本語圏，英語圏を対象としたパスワード解析を行う．本研究が既存研究と比較してユニークな点は，3 つの言語圏の解析結果を直接比較すること，ユーザスタディにおいては中国，日本，英国の各国のユーザをそれぞれの国でよく使われるクラウドソーシングプラットフォームを使うことで，各言語圏で典型的であると考えられるユーザにアプローチすることを目指したこと，およびパスワード生成のみならず，パスワード管理方法に着目する点にある．

本研究の知見は以下の通りである．

- パスワードを単語ベースで考える場合，中国・日本のユーザは個人情報，英国のユーザは一般的な名詞を好む傾向があることを明らかにした．
- 各国で 8 割以上のユーザはパスワード生成器を使わず，自らパスワードを考えていることを明らかにした．
- パスワード管理方法は各国で異なり，英国で最もパスワードマネージャー使用率が高いことを示した．
- ユーザサーベイで見られた傾向の一部は実際の漏洩データでも観測できることを示した．

本論文の構成は以下の通りである．2 章では本研究の背景として漏洩パスワードとパスワード攻撃について述べる．3 章ではユーザスタディに関して被験者募集，サーベイ内容，調査結果を説明する．続いて 4 章では漏洩パスワード解析の手法と結果を示す．5 章では本研究の制約，今後の課題について論じ，6 章では関連研究をまとめる．最後に 7 章で本研究の結論を述べる．

2. 研究背景

2.1 漏洩パスワード

ウェブサービスを攻撃する目的の一つに機密情報を抜き取ることが挙げられる．機密情報の一つとしてパスワードがある．攻撃者はウェブサービスの利用者のパスワードを手に入れると，売ってお金儲けをしたり，手に入れたパスワードを用いて次の攻撃を仕掛けたりする．

漏洩したパスワードの売買は主に裏サイトなどで行われるが，インターネット上に無料で公開されることもある．公開されたパスワードは悪用されるだけでなく，研究目的やセキュリティ目的で利用されることもある．例えば Troy Hunt は漏洩したパスワードとメールアドレスのペアを集め，ユーザが自分のパスワードが漏洩していないかを確認することができるサービスを運用している [5].

2.2 パスワード攻撃

パスワード攻撃は大きく二種類に分けられる．オフライン攻撃とオンライン攻撃である．オフライン攻撃とはパスワードハッシュに対する攻撃で，一方向性関数でハッシュ化されたパスワードハッシュを，平文パスワードに戻すことを目的とする．オンライン攻撃とは稼働しているウェブサービスに対する攻撃で，ユーザ ID とパスワードを入力することで不正アクセスをすることを目的とする．

オフライン攻撃とオンライン攻撃のどちらにおいても，攻撃者はパスワードを推測する必要がある．推測する手法はいくつかあり，よく知られているものにブルートフォース攻撃と辞書攻撃がある．ブルートフォース攻撃はアルファベットや数字を順に並べて試行する手法で，確実だが効率が悪い．一方辞書攻撃は可能性のあるパスワードの辞書を作成し，順に試行する方法である．攻撃の成功率は辞書の精度に依存する．辞書を生成する時に，ユーザの個人情報を考慮したり，生成規則を工夫したりすることで，辞書の精度を高めることができる．

近年話題に上がっている攻撃にリスト型アカウントハッキングがある．これは辞書攻撃の一種で，過去に漏洩したメールアドレスとパスワードのペアを攻撃用の辞書として用いることでウェブサービスにログインを試みる攻撃である．2016 年には Ameba にて，2018 年には niconico にてこの攻撃が観測されている [6][7].

3. ユーザスタディ

3.1 手法

3.1.1 被験者募集

中国，日本，英国の被験者を募集するために，Sojump[8], Lancers[9], Amazon Mechanical Turk[10] を用いた．Sojump は中国語で，Lancers は日本語で，Amazon Mechanical Turk は英語で運用されているクラウドソーシングサービスである．各国の典型的であるとされるユーザを募集するために対応する言語のサービスを用いた．

中国の被験者は 525 人，日本の被験者は 529 人，英国の被験者は 165 人であった．本調査では母国語が各言語のユーザを各国のユーザと定めた．各被験者には国に応じて 0.4-1.0 ドルの報酬を支払った．

3.1.2 サーベイ内容

被験者に対して 5 つの項目からなるサーベイを行った．インフォームド・コンセント，デモグラフィック，パスワードに関する教育歴や知識，パスワード生成方法，パスワード管理方法の 5 つである．すべての質問に答えるのに

表 1　デモグラフィック

	性別 (人)	年齢 –19 / 20–29 / 30–39 / 40–49 / 50–59 / 60–
中国	F: 270 M: 205 O: 1	1.47 / 36.8 / 51.3 / 7.77 / 2.31 / 0.42 (%)
日本	F: 217 M: 298 O: 1	2.52 / 22.1 / 33.9 / 29.5 / 9.11 / 2.91 (%)
英国	F: 32 M: 83 O: 1	11.2 / 88.8 / 0.00 / 0.00 / 0.00 / 0.00 (%)

表 2　使用デバイス (複数可)

	中国	日本	英国
PC(%)	92.4	94.2	96.6
スマートフォン (%)	98.3	68.8	93.1
タブレット (%)	47.1	18.6	46.6

表 3　コンピュータサイエンスを専攻していたか・しているか

	中国	日本	英国
はい (%)	17.6	9.88	18.1
いいえ (%)	81.9	90.1	81.9
答えたくない (%)	0.42	0.0	0.0

10-15 分かかるよう設計した．サーベイの初めに，本調査の目的と回答の使用方法を明記した．続いてデモグラフィックス，パスワードに関する知識 (安全なパスワードの作り方，管理方法を知っているか？など) を尋ねた．最後に実際に使っているパスワード生成方法，管理方法について尋ねた．

3.2　結果

中国から 525 回答，日本から 561 回答，英国から 143 回答が得られた．そのうち回答に矛盾がなく有効なものは中国，日本，英国でそれぞれ 476，516，116 回答であった．被験者のデモグラフィックを表 1 に示す．中国・日本の被験者の年齢層は多様であったが，英国の被験者は 30 歳未満に限られた．

ユーザが日常的に利用しているデバイスを表 2 に示す．3 ヵ国において 90%以上のユーザが PC を利用していた．スマートフォンの利用率は中国・英国で 90%以上だったが，日本では 69%であった．この理由として日本では他国と比べてフィーチャー・フォンの利用率が高いことが考えられる [11].

3.2.1　パスワードに関する教育歴・知識

ユーザのパスワードに関する教育歴を表 3，表 4，表 5 に示す．表 4 はパスワード生成や管理に関する指導を受けた人の割合を示している．日本で 62%のユーザが指導を受けていて，これは中国・英国と比べて高くなっている．一方パスワードが漏洩した場合のリスクについて説明を受けているユーザは中国 30%，日本 35%，英国 44%と英国が最も高く，中国と英国の間に有意差が見られた．国によってどのように注意喚起するかの方法が異なると言える．

3.2.2　パスワード生成

ユーザにどのようにパスワードを生成しているか尋ねた．結果を図 1 に示す．3 ヵ国において自分でパスワードを考

表 4　パスワードの生成・管理方法に関する指導 (複数可)

	中国	日本	英国
学校で受けた (%)	12.4	28.1	18.1
職場で受けた (%)	15.1	46.5	26.7
受けていない (%)	79.4	38.0	62.1

表 5　パスワードが漏洩するリスクに関する説明 (複数可)

	中国	日本	英国
学校で受けた (%)	15.3	15.1	19.0
職場で受けた (%)	22.3	26.0	31.0
受けていない (%)	70.4	64.7	56.0

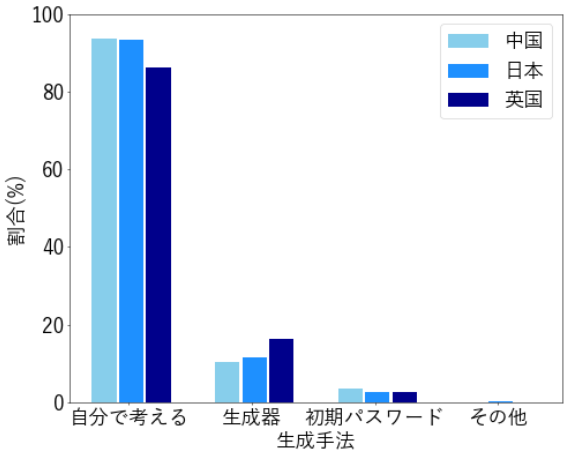

図 1　パスワード生成方法 (複数可)

表 6　パスワード生成器で使う文字種類 (複数可)

	記号	数字	大文字	小文字
中国 (%)	51.0	79.6	51.0	79.6
日本 (%)	57.4	92.6	81.5	88.9
英国 (%)	94.4	100	100	94.4

えている割合が高く 80%を超えている．この方法で生成されたパスワードはもっとも推測されやすく危険である．パスワードの安全性に関する教育が生成方法に影響を与えていないことがわかる．パスワード生成器の使用率は 3 ヵ国において低かったが，その中では英国が 16.4%ともっとも高い割合であった．中国と日本のパスワード生成方法は類似していたが，英国の方法は異なっていた．これらの違いに統計的有意差が見られた．

パスワード生成器を使っているユーザに対して，どのような文字種類からパスワードを生成しているか尋ねた．結果を表 6 に示す．安全なパスワードを生成するためには全ての文字種類を使用することが最も効果的だが，それがなされていたのは英国だけであった．

次に，パスワードを自分でつくると答えたユーザに対して思考過程に関する 2 つの質問をした．

- パスワードを作る時どの言語を使っているか
- どのような思考過程でパスワードを作っているか

表 7 から英国のユーザは主に英語を使ってパスワード

表 7　使用言語 (複数可)

	中国	日本	英国
中国語 (%)	72.6	0.0	0.0
日本語 (%)	0.0	70.6	0.0
英語 (%)	61.2	60.7	93.6
その他 (%)	1.57	6.93	12.8

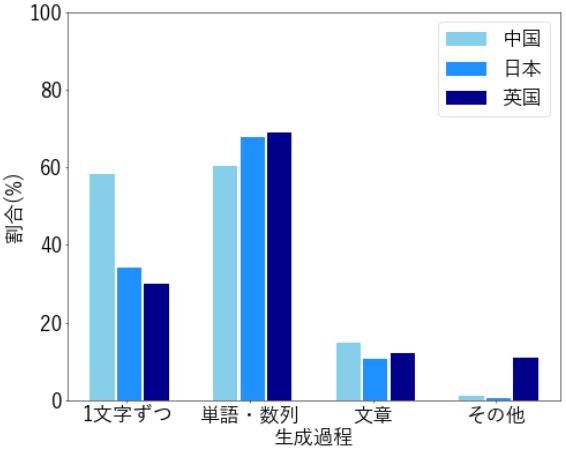

図 2　パスワードの生成過程 (複数可)

を生成していることがわかる．他にもアラビア語，ノルウェー語，タミル語などを使っているユーザが存在した．それに対して中国・日本では英語と母国語を両方使っている傾向があった．この違いを利用することで，パスワードに対して辞書攻撃を仕掛ける時に，効果的な辞書の作成ができるだろう．また図 2 の結果からパスワードの生成方法にも 3 ヵ国で明らかな差があることがわかる．単語・数列をベースにパスワードを考える方法は 3 ヵ国において一般的に使われており，中国で 60%，日本で 68%，英国で 69%のユーザがこの方法を用いていた．一方，1 文字ずつランダムに選ぶ方法は中国でのみ一般的で，中国 58%，日本 34%，英国 30%であった．

単語・数列ベースでパスワードを考えているユーザに対して，以下の 2 つの質問をした．

- どのような単語・数列をベースにしているか
- ベースの単語・数列をどのようにパスワードに変換しているか

表 8 は単語・数列の種類とそれをベースにパスワードを生成しているユーザの割合である．英国では音楽関連や動物関連の一般的な単語が好まれていたのに対し，中国・日本では個人情報 (誕生日，名前など) が使われる傾向があった．この結果にも統計的有意差が見られた．ベースの単語をパスワードに変換する方法にも 3 ヵ国で有意差が見られた (図 3)．日本では単語をつなぎ合わせるだけのユーザが 52.6%と最も高く，続いて単語を混ぜ合わせるユーザが 30.8%，そのまま使うユーザが 17.2%であった．それに対し，中国・英国ではリートという手法 (a を@，l を 1 のように置き換える手法) を使うユーザがそれぞれ 32.0%，

表 8　ベースに使う単語・数列 (複数可)

	中国 (%)	日本 (%)	英国 (%)
個人情報			
名前	**43.5**	**23.1**	2.90
苗字	**37.5**	**16.6**	0.0
ニックネーム	**36.1**	**24.3**	13.0
誕生日	**43.9**	**17.2**	5.80
電話番号	**29.4**	1.85	7.25
クレジット番号	6.69	1.54	0.0
好きな人	**19.0**	6.77	7.25
記念日	**38.3**	12.6	10.1
一般的な単語			
有名人	**19.3**	13.2	7.25
サイト名	9.29	**11.4**	1.45
地名	11.5	8.92	11.6
愛に関する単語	**19.0**	0.615	2.90
音楽に関する単語	14.5	9.85	**20.3**
スポーツに関する単語	8.92	5.23	5.80
動物に関する単語	10.4	7.08	**15.9**

43.5%と最も多かった．

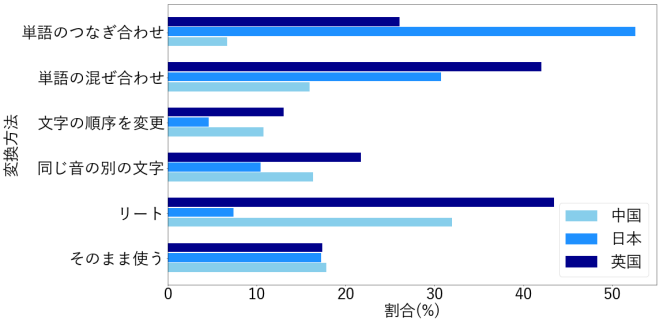

図 3　単語・数列の変換方法 (複数可)

3.2.3　パスワード管理

パスワード管理方法に関して，以下の 5 つの質問を尋ねた．

- どのようにパスワードを保管しているか
- パスワード管理ソフトを使ったことがあるか
- パスワードで保護されるアカウントをいくつ持っているか
- ユニークなパスワードをいくつ持っているか
- パスワードを再利用しているか

ユーザのパスワードの保管方法は表 9 に示す通りであった．中国・英国では過半数のユーザがパスワードを全て覚えているのに対し，日本では 23%であった．英国ではパスワード生成器を使っているユーザが 16%と他国と比べて多かったが，全てのパスワードを覚えているユーザが過半数を占める結果となった．人が覚えられるパスワードは推測されやすい可能性が高く危険である．セキュリティ意識が高いユーザとそうでないユーザの習慣に大きな差があると考えられる．また中国・日本ではノートや日記などの紙媒

表 9　パスワード保管方法 (複数可)

	中国	日本	英国
全て覚える (%)	**55.5**	23.4	**53.4**
一部覚える (%)	17.6	34.7	14.7
紙媒体に記録する (%)	**44.5**	**51.7**	19.8
電子媒体に記録する (%)	23.3	29.5	8.62
ブラウザに保存する (%)	15.1	15.5	13.8
パスワード管理ソフトを使う (%)	3.57	8.91	**21.6**
その他 (%)	0.210	0.388	0.0

表 10　パスワード管理ソフトに関して

	中国	日本	英国
使っている (%)	3.57	8.91	21.6
以前使っていた (%)	10.9	3.49	4.31
使ってみたい (%)	**46.6**	16.7	17.2
使うつもりはない (%)	13.2	**44.4**	31.0
知らなかった (%)	25.6	26.6	23.3
その他 (%)	0.0	0.0	2.59

表 11　ユーザのアカウント保持数

	1	2–5	6–10	11–20	21–50	51–100	101–
中国 (%)	1.47	21.4	**42.4**	26.9	6.93	0.84	0.0
日本 (%)	0.39	18.2	**34.7**	25.6	17.2	3.10	0.78
英国 (%)	1.72	2.59	**35.3**	21.6	28.4	6.90	3.45

体に記録するユーザの割合がそれぞれ 45%，52%，電子媒体に記録する割合がそれぞれ 23%，30%と英国と比べて多かった．パスワード管理ソフトの利用率は英国で最も高く 21%を超えていた．パスワード管理ソフトの使用経験に関するアンケート結果を表 10 に示す．中国のユーザで最も多かったのはパスワード管理ソフトを使ってみたいというユーザで 47%を占めた．日本のユーザで最も多かったのはパスワード管理ソフトを使うつもりはないというユーザで 44%を占めた．中国のユーザはパスワード管理ソフトの使用に前向きだが日本のユーザは後ろ向きであった．後ろ向きである理由として “パスワード管理ソフトの故障や安全性が心配だから”，“必要がないから”，“面倒だから” などが挙げられた．

ユーザにパスワードで保護されているアカウント保持数とユニークなパスワード数を尋ねた．結果を表 11，表 12 に示す．3 カ国において 6–10 個のアカウントを持っているユーザが最も多かった．英国ではユニークなパスワード数も 6–10 個保持するユーザが多かったが，中国・日本では 2–5 個の割合が高かった．英国に比べて中国・日本はパスワード再利用率が高いと考えられる．次に，どのようなアカウントで同じパスワードを使うかを尋ねることで，ユーザのパスワード再利用に関する調査をした．結果を表 13 に示す．これは表 11，表 12 の結果と一致している．

3.2.4　パスワードに関する教育歴とパスワード習慣

パスワードに関する教育 (生成・管理方法の指導，漏洩リスクに関する説明) とパスワードの扱い方に関連がある

表 12　ユーザの保持するユニークなパスワード数

	1	2–5	6–10	11–20	21–50	51–100	101–
中国 (%)	9.03	**53.2**	26.5	7.35	3.57	0.42	0.0
日本 (%)	3.68	**42.2**	24.6	16.1	11.2	1.94	0.19
英国 (%)	1.72	31.9	**32.8**	13.8	15.5	2.59	1.72

表 13　パスワード再利用

	中国	日本	英国
している (%)	75.2	64.5	42.2
していない (%)	24.8	35.1	53.4
その他 (%)	0.0	0.39	4.31

かどうかの調査をした．教育を受けたユーザと受けていないユーザの間で，パスワード生成器やパスワード管理ソフトの使用率，パスワードの再利用率に有意差があるかを確認した．結果を表 14 に示す．オッズ比とカイ二乗検定を用いて教育がパスワード生成器，パスワード管理ソフトの利用またはパスワード再利用に影響を与えているかどうかを評価した．中国では教育によってパスワード生成器やパスワード管理ソフトの利用を促せていることがわかる．これは中国ではパスワード管理ソフトに関心があるユーザが多く (表 10)，教育によって後押しすることが容易であるからだと考える．また日本・英国ではパスワード生成器を用いることでパスワード再利用を抑制できていたが，中国ではそうではなかった．一方で，中国・日本ではパスワード管理ソフトを使うことでパスワードの再利用を抑制することができていたが，英国ではそうではなかった．日本，英国においてパスワードに関する教育がパスワードマネージャーの利用やパスワード再利用の抑制に繋がっていなかった．“パスワードマネージャーを利用すること” と “パスワード再利用をしないこと” は安全にパスワード認証を行うための習慣の一部である．これらの習慣を促すために，セキュリティ意識向上を図る方法を考え直す必要がある．

4.　漏洩パスワード解析

3 章の結果から得られた知見を元に，漏洩パスワードを解析を行った．

4.1　データ収集

本調査では，複数のウェブサイトから漏洩したデータの寄せ集めである Exploit.in[12]，中国のウェブサイトから漏洩した 7k7k2000w，500W_16610 を使用した．詳細を表 15 に示す．Exploit.in のデータセットは複数のウェブサイトからの漏洩データであるが，各データがどこのウェブサイトから漏洩したかはわからない．各データセットはメールアドレスとパスワードのペアを含む．それぞれのデータセットを合わせ，以下の処理を行った．

- 各データを [メールアドレス]:[パスワード] というコロン区切りのデータに整形した．

表 14　パスワードに関する教育歴とパスワード習慣

	中国			日本			英国		
	odds 比	95%CI	p 値	odds 比	95%CI	p 値	odds 比	95%CI	p 値
生成・管理方法の指導を受けたかどうか									
パスワード生成器を利用	2.27	1.20–4.29	**<0.01**	0.95	0.55–1.66	0.87	1.59	0.59–4.30	0.35
パスワード管理ソフトを利用	6.02	2.23–16.3	**<0.01**	0.78	0.42–1.44	0.42	1.12	0.45–2.77	0.81
パスワード再利用	0.73	0.45–1.20	0.22	1.09	0.75–1.58	0.66	0.80	0.37–1.74	0.58
漏洩リスクの説明を受けたかどうか									
パスワード生成器を利用	1.92	1.05–3.51	0.03	0.86	0.48–1.53	0.60	1.18	0.44–3.16	0.74
パスワード管理ソフトを利用	3.58	1.33-9.60	**<0.01**	0.98	0.52–1.84	0.94	1.00	0.41–2.44	1.00
パスワード再利用	0.77	0.49–1.20	0.24	1.11	0.76–1.62	0.60	0.45	0.21–0.99	0.05
パスワード生成器を使っているかどうか									
パスワード再利用	1.52	0.72–3.24	0.27	0.44	0.26–0.77	**<0.01**	0.12	0.03–0.56	**<0.01**
パスワード管理ソフトを使っているかどうか									
パスワード再利用	0.17	0.06–0.46	**<0.01**	0.31	0.17–0.58	**<0.01**	0.28	0.09–0.81	0.02

- メールアドレスのトップレベルドメイン (cn,jp,uk,com,net,edu,org) によりデータを分類し，ユニークなデータセットにした．
- ドメイン (一定数以上のメールアドレスに使われているもの) を www.alexa.com/siteinfo で検索し，サイト訪問者の位置情報を確認した．訪問者の 90%以上が中国・日本・英国のユーザであれば，各国のデータセットとして採用した．
- パスワードが含まれていないデータを取り除いた．

処理の後，調査対象となったデータ数を表 16 に示す．なお今回使用したデータは森ら [3] が利用したものと同じものである．

4.2　解析手法

3.2.2 章の結果を元に，漏洩パスワードに音楽・動物関連の単語が出現する割合，日付の出現する割合を調査した．

音楽関連の単語として，billboard が毎週算出しているランキングからアーティスト名のみを利用した．各国の漏洩パスワードに対して，自国のアーティスト名が出現する割合を算出した．動物関連の単語として 31 種類の動物の中国語名 (ピンイン)・日本語名 (ローマ字)・英語名を利用した．3 言語を合わせて 93 単語のリストとし，各国の漏洩パスワードに出現するかどうか調査をした．アーティスト名も動物名も，大文字小文字の区別はしなかった．アーティスト名にスペースが含まれる場合は，スペースを削除して各単語を連結したものをアーティス名とした．単語リストの詳細を表 17，表 18 に示す．日付の出現として，YYYYMMDD・MMDDYYYY・DDMMYYYY・MMDD・DDMM の形式に当てはまるものを調査した．YYYY は 1900 から 2099 まで，MM は 01 から 12 まで，DD は 01 から 31 に制限をした．

音楽・動物関連の単語，日付形式の出現を調査する方法として，これらの単語・数列が含まれているパスワードの割合，これらと完全に一致するパスワードの割合を調べた．含まれているかどうかを確認する時，単純に単語・日付を含むかどうかを確認すると誤検出の可能性が高くなる．そこで単語の場合は対象の単語の両側がアルファベットでないこと，日付の場合は日付形式の両側が数字でないことを条件とした．またアーティスト名・動物名を含むかどうかの調査対象は 3 文字以上の単語のみに絞った．なお 3 文字以上の動物関連の単語は中国語が 24 単語，日本語が 31 単語，英語が 31 単語であった．

表 15　漏洩パスワードセット

データセット	データ数	サービス
Exploit.in	805,499,579	—
7k7k2000w	19,138,452	ゲーム
500W_16610	4,768,600	ポータルサイト

表 16　調査対象のデータ

国	Exploit.in	中国語のサイト
中国	1,402,370	4,487,396
日本	462,087	—
英国	388,350	—

表 17　使用した単語リスト

リスト名	URL
アーティスト (中国)	https://www.billboard.com/charts/china-v-chart
アーティスト (日本)	https://www.billboard.com/charts/japan-hot-100
アーティスト (英国)	https://www.billboard.com/charts/official-uk-songs
動物リスト (ピンイン)	http://www.lexisrex.com/Chinese-Vocabulary/Animals-pinyin
動物リスト (ローマ字)	http://www.lexisrex.com/Japanese-Vocabulary/Animals-romaji
動物リスト (英語)	http://www.lexisrex.com/Chinese-Vocabulary/Animals-pinyin
	http://www.lexisrex.com/Japanese-Vocabulary/Animals-romaji

表 18　音楽関連の単語リスト

チャート名	開始週	最終週	アーティスト数 (3 文字以上)
china-v-chart	2015-12-05	2018-08-04	702 (700)
japan-hot-100	2011-04-16	2018-08-04	3707 (3688)
official-uk-songs	2011-02-05	2018-08-04	717 (713)

4.3 解析結果

漏洩パスワードを解析した結果が表 19 である．

アーティスト名が含まれている割合は，日本が最も高く 1.72%であった．これは調査対象としたアーティスト数が中国・英国は 700 ほどであったのに対して日本は 3,700 ほどであったからだと考える (表 18)．また表 8 から推測すると中国よりは英国でアーティスト名が使われていそうだが，そうではなかった．これは英国のアーティスト名が長いことが理由の一つだと考える．今回用いたアーティスト名の中で，8 文字以下のものは中国で 56%，日本で 26%，英国で 18%で，数はそれぞれ 392，967，126 であった．パスワードの長さは 8 または 6 の可能性が高く [3]，8 文字長を超えるアーティスト名は使われにくいだろう．

動物名が含まれているパスワードの割合は英国が最も高く 0.46%であった．これはユーザスタディと一致している．一方中国・日本に関して，ユーザスタディでは中国のユーザの方が動物名を利用する傾向があったが，解析結果では日本のパスワードに含まれている割合の方が高かった．本調査では 3 文字以上の動物名に絞ったため，中国語名が減ってしまった．これが中国のパスワードの割合が低くなった原因の一つであると考える．

日付の出現に関して，日付のみからなるパスワードは中国で最も多く見られ，続いて日本，英国とユーザスタディの調査結果と一致した (表 8 の誕生日，記念日)．一方日付を含むパスワードの割合は日本が中国を上回り，ユーザスタディと漏洩データ解析の結果で差が見られた．アンケートに答える時と実際にパスワードを生成する時で，思考が異なる可能性や記憶が曖昧な可能性，誕生日・記念日以外の日付を用いている可能性が考えられる．

表 19 漏洩パスワード解析

	中国		日本		英国	
	含む	一致	含む	一致	含む	一致
アーティスト名 (%)	0.94	0.08	1.72	0.30	0.17	0.06
動物名 (%)	0.17	0.01	0.26	0.03	0.46	0.17
日付 (%)	7.38	4.41	11.7	2.91	2.02	0.17

5. 議論

5.1 制約と課題

本調査に参加した英国のユーザは 165 人，そのうち有効回答が 116 件と中国・日本と比較して少数であった．またその回答は全て 30 歳未満のユーザのものであった．本調査で得られた傾向が英国のユーザ全般のものであるとは言い難い．より普遍的な調査をするためには，Amazon Mechanical Turk ではなく英国のユーザ向きのクラウドソーシングサービス (Prolific [13] など) を利用する必要がある．

本研究ではオンラインサーベイを元に漏洩パスワード内の音楽・動物関連の単語を解析した．しかし調査に使用した単語数が少ない．音楽関連の単語としてアーティスト名のみ，動物関連の単語として 31 種類の動物名のみを使用した．音楽関連の単語として曲名，歌詞，楽器名，音楽ジャンル名などを利用している可能性や，動物関連の単語として今回利用した 31 種類の動物名以外を利用している可能性もある．

また，ユーザが自己申告する内容と実際のパスワード生成の習慣は異なる可能性がある．幅広い分野の単語コーパスを用いて大規模に調査を行うことで，オンラインサーベイでは見られなかった傾向も見られると考える．

5.2 研究倫理

オンラインサーベイはパスワード生成に用いる単語や生成規則，管理方法など機密情報に関わる内容であった．そのためサーベイのはじめにインフォームド・コンセントを得た．具体的には，調査への参加は自由意志によるものでいつでも辞退できること，アンケート結果は研究目的でのみ使用されること，結果が公表される場合は被験者のプライバシーが保全されることを明記し，同意を得た．

本研究の漏洩パスワード解析で利用したデータセットはメールアドレスとパスワードのペアで構成される．このデータは限られた研修者のみがアクセスできる環境で安全に保管をし，研究の目的のみで使用した．

6. 関連研究

6.1 漏洩パスワード解析

Zhigong Li ら [2] は中国のウェブサイトから漏洩したパスワードを解析し，数字を含むパスワードが多かったこと，ピンインを使ったパスワードも存在したことを示した．そして，これらの傾向を考慮して攻撃の辞書にピンインを含めることで，パスワード推測の成功確率が高まることを示した．

Yue Li ら [14] は `www.12306.cn` から漏洩したデータを用いてパスワードへの個人情報の使用を調査した．この漏洩データは 130,000 以上のパスワードと個人情報のペアで構成され，個人情報にはメールアドレス，携帯番号，政府発行の ID などが含まれる．60%のパスワードがなんらかの個人情報 (特に誕生日やアカウント名) を含んでいることを示した．

また漏洩パスワードを独自で解析し，使われる頻度が高いパスワードのランキングを発表している団体や [15], [16]，解析して得られたパスワードの傾向をブログで公開している個人も存在する [17]．

6.2 パスワード生成に関するユーザスタディ

Shannon Riley ら [18] はユーザのパスワード生成，管理方法の調査を行い，最善だと思っている方法を使用しない

ユーザの存在を明らかにした．

Blase Ur ら [19] はパスワード生成方法に関するインタビューを行った．インタビューの中で被験者に 3 つのウェブサイト (銀行・メール・情報サイト) のパスワードを生成するよう指示をし，思考過程を調査した．安全なパスワードとはなにかを誤解しているユーザの存在を示しただけでなく，アカウントの重要度の認識に一般ユーザとセキュリティ組織の間で差が見られたことを明らかにした．

7. まとめ

本研究はまず中国・英国・日本のユーザにパスワード生成方法，管理方法に関するアンケートを取った．各国の言語で運用されているクラウドソーシングサービスを用いることで，言語圏ごとのパスワード習慣の差異に着目することを可能にした．結果として 3 ヵ国において主なパスワード生成方法は自分で作る，特に単語や数列ベースでパスワードを作ることであることを明らかにした．パスワード生成器は英国で最も普及していて，利用率は 16%であった．パスワードを単語・数列ベースで考えるユーザの中で，中国・日本では個人情報が，英国では一般的な単語 (音楽，動物関連など) がベースに選ばれる傾向があることを明らかにした．特に中国のユーザは誕生日，記念日などの数字を好む傾向があった．パスワード管理方法に関して，パスワード管理ソフトは英国で最も普及していた．中国ではパスワード管理ソフトに前向きな意見を持っているユーザが 40%を超えていたが，日本では後向きな意見を持っているユーザが 40%を超えていた．

ユーザサーベイを元に漏洩パスワードを解析することで，サーベイ結果と漏洩データが一致している部分もあるが一致していない部分もあることがわかった．アンケート回答時と実際のパスワード生成時では目的が異なるため，重要視する点や思考過程に差が出ると考えられる．アンケート結果を参考に漏洩データをより深く解析する必要ことで，実際のユーザの習慣を調査できるだろう．

ユーザのパスワード習慣の言語圏ごとの違いを理解することで，ユーザに最適な方法で安全なパスワード生成・管理を促すことができる．例えばパスワード生成時の警告文やブラックリストに含めるパスワード，パスワード管理ツールの機能や促進方法などを対象の国ごとに変更することが考えられる．

参考文献

[1] Ur, B.: Supporting Password-Security Decisions with Data, PhD Thesis (2016).

[2] Li, Z., Han, W. and Xu, W.: A Large-Scale Empirical Analysis of Chinese Web Passwords, *Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014.*, pp. 559–574 (2014).

[3] 森　啓華，森　達哉：文化の差異がパスワード生成に与える影響の考察 (2018).

[4] 愛野乃子，金岡　晃：日本人が付けるパスワードの特性調査と他国データとの比較，研究報告セキュリティ心理学とトラスト（SPT），pp. 1–5 (2018).

[5] Hunt, T.: ';–have i been pwned?, `https://haveibeenpwned.com/`.

[6] Ameba 広報担当：「Ameba」への不正ログインに関するご報告とパスワード再設定のお願い，`https://www.cyberagent.co.jp/news/detail/id=11977` (2016).

[7] ニコニコインフォ：他社流出パスワードを用いた不正ログインについて (2018/05)，`http://blog.nicovideo.jp/niconews/73053.html` (2018).

[8] 長沙冉星信息科技有限公司：Sojump, `https://www.wjx.cn/` (2018).

[9] Lancers.inc: Lancers, `https://www.lancers.jp/` (2018).

[10] Amazon.com: Amazon Mechanical Turk, `https://www.mturk.com/` (2018).

[11] デロイト トーマツコンサルティング：世界モバイル利用動向調査 2017，`https://www2.deloitte.com/jp/ja/pages/about-deloitte/articles/news-releases/nr20171214.html` (2017).

[12] Hunt, T.: Password reuse, credential stuffing and another billion records in Have I been pwned, `https://www.troyhunt.com/password-reuse-credential-stuffing-and-another-1-billion-records-in-have-i-been-pwned/` (2017).

[13] Prolific: Prolific, `https://prolific.ac/`.

[14] Li, Y., Wang, H. and Sun, K.: Personal Information in Passwords and Its Security Implications, *IEEE Trans. Information Forensics and Security*, Vol. 12, No. 10, pp. 2320–2333 (2017).

[15] splashdata: WORST PASSWORDS OF 2017 Top 100, `https://s13639.pcdn.co/wp-content/uploads/2017/12/Top-100-Worst-Passwords-of-2017a.pdf` (2017).

[16] keeper: The Most Common Passwords of 2016, `https://keepersecurity.com/public/Most-Common-Passwords-of-2016-Keeper-Security-Study.pdf` (2016).

[17] : あなたのパスワードは本当に安全ですか？ 300 万件のサンプル中の弱いパスワードの規則性を見てみよう，`http://36kr.com/p/5038663.html` (2015).

[18] Riley, S.: Password Security: What Users Know and What They Actually Do, *Usability News, Volume 8, Issue 1* (2006).

[19] Ur, B., Noma, F., Bees, J., Segreti, S. M., Shay, R., Bauer, L., Christin, N. and Cranor, L. F.: "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab, *Eleventh Symposium On Usable Privacy and Security, SOUPS 2015, Ottawa, Canada, July 22-24, 2015.*, pp. 123–140 (2015).