抑止力型トラスト：災害報告検証者の信頼性向上のための仕組みの検討

SCIS 2020 2020 Symposium on Cryptography and Information Security
Kochi, Japan, Jan. 28 – 31, 2020
The Institute of Electronics, Information and Communication Engineers

抑止力型トラスト：災害報告検証者の信頼性向上のための仕組みの検討
Deterrence-Based Trust: A Study for Improving Reliability of Disaster Report Verifiers

北川 沢水[1,] 向平 浩貴[1] 上原 航汰[1] 大木 哲史[1]
小泉 佑揮[2] 河辺 義信[3] 長谷川 亨[2] 西垣 正勝[4]
Takumi KITAGAWA[1] Koki MUKAIHIRA[1] Kota UEHARA[1] Tetsushi OHKI[1]
Yuki KOIZUMI[2] Yoshinobu KAWABE[3] Toru HASEGAWA[2] Masakatsu NISHIGAKI[4]

あらまし ソーシャルメディアが災害時情報伝達に有用であることが認識されつつある．一方で，ソーシャルメディア上に氾濫する情報は玉石混交であり，雑多な情報の中から信頼性の高い必要な情報を取捨選択することは容易ではない．災害時には，いち早く正確な情報を収集し，迅速な対応をしなければならない．この課題に対し，我々はこれまでに，2 層化したボランティアを活用したクラウドソーシングによる情報クレンジングを提案している．本方式では，身元確認に応じた一部の登録ボランティアに，他の匿名ボランティアの目付役を依頼する．ソーシャルメディアに流れる無数の情報の信憑性を多数の匿名ボランティアが確認し，匿名ボランティアの作業結果の正当性を少数の登録ボランティアが確認するという 2 層構造を設けることにより，ボランティアに科される身元確認を抑えながら相応の情報クレンジングを実現することを狙っている．本稿では，このうち，身元確認情報の種類と虚偽報告発信に対する抑止力の間の関係性を探る．200 名（有効回答は 66 名）規模のオンラインアンケートによる調査を通じて，プライバシー懸念と抑止力効果の両面から身元確認の際に用いる本人情報としては何が好適であるか分析する．

キーワード 災害時通信，トラスト，生体情報

1. はじめに

近年，自然災害の甚大化にともない，人々の災害に対する意識が向上しつつある．災害時においては，「黄金の 72 時間」と呼ばれる災害初期に，消防，警察などのレスキュー組織や救急隊員が被災者や被災状況に関する最新の情報を収集し，被災者を迅速に救出することが重要である．しかし，複数のレスキュー組織間での情報共有の難しさ，119 番通報などの高信頼な通信インフラの輻輳や障害により，迅速で最適な救助活動の実現は非常に困難な課題となっている．

一方で，電話網と比較してインターネットは耐障害性が高いことが知られており，Twitter などのソーシャルメディアを用いた被災情報の伝達が有用である事が認識されつつある．ソーシャルメディアは全ての参加者から各個のリアルタイムな情報を集約できるため，地方公共団体においても災害時対応のためのソーシャルメディア活用率は年々増加している[1]. 2019 年 10 月に発生した台風 19 号による豪雨によって被害を受けた長野県は，Twitter を活用し独自に救助要請を収集することで約 50 件の救助に繋げることができた[2]．しかし，ソーシャルメディアを 119 番通報の代替として用いるには様々な課題が存在する．その 1 つにソーシャルメディア上の情報の信頼性の問題が存在する．一般に，他のメディアと比べてインターネット上の情報の信頼度は低いといわれている[3]．ソーシャルメディアにおいても雑多なユーザと情報が氾濫しており，その中から信頼性の高い情報のみを選択することは困難である．

以上のことから，災害時にソーシャルメディア上の情報の信頼性を保証する仕組みが必要となる．

[1] 静岡大学大学院総合科学技術研究科, Graduate School of Integrated Science and Technology, Shizuoka University
[2] 大阪大学大学院情報科学研究科, Graduate School of Information Science and Technology, Osaka University
[3] 愛知工業大学情報科学部, Department of Information Science, Aichi Institute of Technology
[4] 静岡大学創造科学技術大学院, Graduate School of Science and Technology, Shizuoka University

2. 2 層情報クレンジングシステム

2.1. コンセプト

災害時，いち早く正確な情報を収集し迅速な対応を行うための手段として，我々は先行研究にて，ソーシャルメディア上に流れる情報の信憑性の確認を災害現場に集まったボランティアに依頼する方法を提案している[4]．このようなクラウドソーシング型の情報クレンジングが効果的に機能するためには，虚偽報告を行うボランティア（クラウドワーカ）は居ないという前提が必要である．しかし，ボランティアが匿名である場合などには，この前提が常に満たされるとは限らない．ボランティアの身元確認を徹底することができれば，身元が割れていることが虚偽報告を行うにあたっての抑止力になり得る．しかし，災害時に全ボランティアに身元確認を強制するようなことは，確認する側にとってもされる側にとっても手間や心的負担が過大となるため非現実的である．以上の問題に対し，我々は，クラウドソーシング型の情報クレンジングを 2 層化するという方式を提案している[5]．本方式では，身元確認に応じた一部のボランティア（登録ボランティア）に対し，他のボランティア（匿名ボランティア）の報告の正誤検証を依頼する．ソーシャルメディアに流れる無数の情報の信憑性を多数の匿名ボランティアが確認し，匿名ボランティアの作業結果の正当性を少数の登録ボランティアが確認するという 2 層構造を設けることにより，ボランティアに課される身元確認を抑えつつ，情報クレンジングを行う機構を実現する．

2 層情報クレンジングのイメージが図 1 である．図中，VA，IC，FR はそれぞれ Volunteer Authority，Incident Commander，First Responder の略称である．2 層情報クレンジングは以下の 3 段階の手順によって実行される．各手順についての詳細は 2.2 節，2.3 節，2.4 節で述べる．

1. 情報サマライズ：ソーシャルメディアメッセージのカテゴライズ
2. 1 層目の情報クレンジング：Event Report に対するレピュテーション
3. 2 層目の情報クレンジング：Volunteer に対するレピュテーション

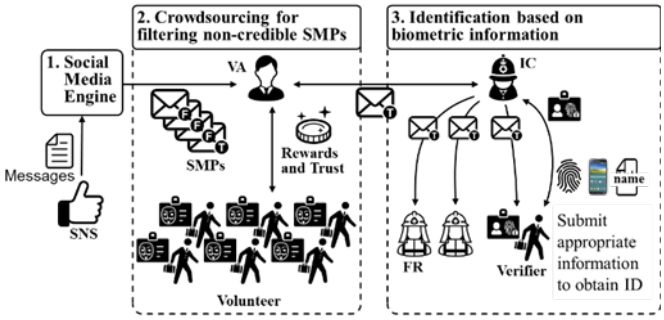

図 1　2 層情報クレンジングのイメージ図

2.2. ソーシャルメディアメッセージのカテゴライズ

1 段目では，ソーシャルメディアエンジンを用いてソーシャルメディア上に発信された情報の中から被災地に関する情報を自動抽出する．ソーシャルメディアエンジンは，ソーシャルメディアメッセージをパースし，テキストマイニング技術によって，ソーシャルメディアメッセージを時間，場所，内容に応じて分類してアイテム化する．Twitter のリツイートによって拡散した情報など，同一内容のメッセージについては縮約され，独立したアイテムごとに個々の「Event Report」としてまとめられる．ソーシャルメディアに流れる情報は玉石混合であるため，この段階の Event Report には誤った情報も含まれている．そこで，Event Report を 2 段目の Volunteer Authority に送信し，Volunteer に各 Event Report の真偽の確認を依頼する．

2.3. Event Report に対するレピュテーション

2 段目では，Volunteer 間での多数決による情報クレンジングを行う．Volunteer Authority は，1 段目のソーシャルメディアエンジンから届いた Event Report をクラウドソーシングシステムに投入し，個々の Event Report の真偽の確認をジョブとして公開する．被災地に到着した Volunteer は，クラウドソーシングシステムにアクセスすることによってジョブ（Event Report）の一覧を確認する．その中から，自分の状況あるいは目的に合致した Event Report（例えば，現在の自分の所在地近辺の Event Report や，報奨金額が高い Event Report）を選び，その真偽確認を自らのジョブとして受注する．Volunteer は直接現地に赴き，Event Report の内容の真偽を確認し，その結果をクラウドソーシングシステムに報告する．1 つの Event Report に対して 3 名以上の Volunteer が割り当てられ，過半数の Volunteer から「真」の報告が返された Event Report だけが，3 段目の Incident Commander に送られる．

一般的にソーシャルメディアに流れるメッセージは多量のため，ソーシャルメディアエンジンから出力される Event Report も相応の数となることが予想される．莫大な量の Event Report の真偽を判断していくには，Volunteer（クラウドワーカ）を十分に確保できる仕組みを考える必要がある．そこで，Volunteer には匿名で参加してもらい，正しい報告を行った際に報酬（以降「コイン」と表記する）を受け取ることができるようにする．Volunteer のインセンティブを更に高めるために，多くのコインを獲得した Volunteer には政府から褒章等の特典が与えられるようにしても良いだろう．

Volunteer の匿名性から，現地に赴いての確認を行わずに報告だけを行い，不正にコインを獲得しようとする者が現れる可能性がある．そのため，Event Report の真偽の判断は Volunteer 間の多数決により行うこととする．また，Volunteer

が受け取るコインは擬似通貨であり，災害活動中は現金通貨に換金する事はできない．災害活動中に虚偽報告を行わなかった Volunteer のみが災害終息後に現金通貨に換金できる．

2.4. Volunteer に対するレピュテーション

3 段目では，身元確認を済ませた登録ボランティア（以降「Verifier」と表記する）による情報クレンジングを行う．Incident Commander は，2 段目の Volunteer Authority から届いた Event Report を Verifier に伝える．Verifier は Event Report に基づいて災害現場に赴き，Event Report の真偽を確認する．万一，Event Report の内容に虚偽が認められた場合には，Verifier はそれを Incident Commander に報告する．この「偽」の報告は，Incident Commander を通じて Volunteer Authority にも届けられる．

2 段目では，Volunteer 間での多数決による情報クレンジングが行われている．しかし，Volunteer は匿名であるため，結託あるいはシビル攻撃によって多数決の結果を操作する余地が残る．そこで，3 段目の Verifier による情報クレンジングで，Volunteer の信頼性の検証を行う．Verifier の報告は Incident Commander が取りまとめ，報酬と Volunteer の管理を行う Volunteer Authority に伝えられる．Volunteer Authority は，一度でも虚偽の報告を行った Volunteer は信頼できないものとし，当該 Volunteer からの他の報告と当該 Volunteer のコインを抹消する．

1 段目のソーシャルメディアエンジンによって，ソーシャルメディア上のメッセージが集約され，更に 2 段目の情報クレンジングによって，不正なメッセージが一旦ふるい落とされる．よって，ある程度少人数の Verifier がいれば，3 段目の情報クレンジングを行うことができると考える．このように，メッセージの信頼性と人の信頼性を 2 層で確認する情報クレンジングを行うことで，ソーシャルメディア上に存在する大量の雑多な情報の中から，信頼性の高い情報を効率的に抽出可能となることが期待される．

2.5. シナリオ

2 層情報クレンジングの具体的なシナリオは以下の通りである．

1. 災害現場でローカルコインを発行する．
2. 災害現場でクラウドソーシングシステムを稼働させる．
3. ソーシャルメディアに流れるすべてのメッセージをパースし，当該の災害に関連するメッセージを自動抽出するソーシャルメディアエンジンを運用する．エンジンによって，メッセージを時間，場所，内容等からカテゴライズすることでアイテム化し，Event Report を生成する．
4. エンジンは，各 Event Report からそのメッセージの内容の真偽を確認させるジョブを生成してクラウドソーシングシステムに投入する．また，各ジョブには，そのジョブ毎に報酬金額を設定する．
5. 災害現場に到着した Volunteer は，クラウドソーシングシステムにアクセスする．クラウドソーシングシステムにジョブが次々と投入されるが，それぞれのジョブが多数の Volunteer の中の誰かによって実行される．
6. Volunteer は，自身の状況に応じたジョブを請け負い，そのメッセージの内容の真偽を実際に確かめに行く．
7. ジョブを遂行した Volunteer は，その結果をクラウドソーシングシステムに入力する．Volunteer には報奨金額分のコインが与えられる．
8. 多数決により正しいと判断されたジョブに該当する Event Report だけが，Incident Commander に送られる．
9. Incident Commander は Event Report の内容に基づいて，当該 Event Report を担当する Verifier を決める．
10. Verifier は現地へ赴いて Event Report の真偽を確認し，Incident Commander に結果を報告する．
11. Incident Commander が Verifier の報告を Volunteer Authority に通知する．Volunteer Authority はそれを記録する．
12. 災害終息後，Volunteer はコインを現金通貨に換金可能となる．虚偽の報告を行ったことが発覚した Volunteer は換金対象外となる．
13. 多くのコインを獲得した Volunteer には，政府から褒章等の特典が与えられる．

2.6. 考察

前節手順 5 において，各 Event Report の報酬金額を適切に設定することにより，Volunteer 対して効果的にジョブを発行できる．例えば，隔離された状況に置かれている被災者からの救難メッセージのような緊急度の高い情報や，多くの人々に拡散されているような影響力の大きい情報に対しては，ソーシャルメディアエンジンが自動的に高い報酬金額を設定してジョブを発行することによって，Volunteer が優先してそのメッセージの真偽を確かめに行くことが期待される．

Volunteer には，Event Report の真偽確認以外のジョブを発注することもできる．例えば，災害によってネットワークの切断が発覚した際には，「P2P 通信パケットのフォワーダとして，自身のスマートフォンの P2P 通信機能を ON にしたまま，通信途絶エリアに滞在する」というジョブを，クラウドソーシングシステムを通じて Volunteer に発行することができる．

3 段目の Verifier による情報クレンジングは，Volunteer が不正な Event Report を「真」と偽る攻撃を発見するための手段となるが，Volunteer が正しい Event Report を揉み消す

攻撃に対しては有効ではない．心無い Volunteer が，結託あるいはシビル攻撃によって，「この Event Report は偽である」という報告をクラウドソーシングシステムに多数送信した場合，当該 Event Report は多数決の結果，ブロックされ，Incident Commander には通知されず，Verifier による情報クレンジングの対象からも外れてしまう．また，Volunteer による自作自演行為（Volunteer 自身が放火する→ソーシャルメディアに火事が発生していると投稿する→その虚偽投稿に関する Event Report のジョブが発行される→自分自身でそのジョブを請け負ってコインを獲得する）が行われる可能性がある．この 2 点については改善が必要である．

3. 抑止力型トラスト

Volunteer の信頼性を検証する役割を担う Verifier は，2 層情報クレンジングにおける「信頼の最後の砦」である．このため，「Verifier は嘘をつかない」という状況を作り出すことが重要となる．そこで，本研究では，Verifier の信頼性を向上させる仕組みとして，Verifier に本人情報を登録させることを本人確認および虚偽情報の発信抑止のための手段として活用する方法を検討する．ここで，本人情報とは，「その情報のみで本人を特定できる情報」とする．例として，個人情報，生体情報，携帯電話番号が挙げられる．

災害現場には，警察署や消防署などからも多くの署員や隊員が First Responder として参加している．そこで，First Responder をトラストアンカとして機能させ，First Responder がボランティアの身元確認を目視によって行った場合に，First Responder からボランティアに PKI 秘密鍵と PKI 公開鍵が発行される仕組みを運用する．身元確認を済ませ，PKI 秘密鍵と PKI 公開鍵を所持した時点で，ボランティアは Verifier としてメッセージを発信する資格を得る（図 2，図 3）．Verifier の本人情報は，First Responder の公開鍵によって暗号化された形で登録されており，プライバシーに関する懸念が最小となるように配慮されている．

Verifier の登録を抑止力の起点とした情報送信は以下の手順で動作する（図 4）．

① ボランティアはルート CA の公開鍵PK_{root}を信頼している．

② First Responder は，ルート CA から秘密鍵 SK_{fr}，公開鍵 PK_{fr}，公開鍵証明書 $Sig_{SKroot}(PK_{fr})$を発行してもらっている．ここで，$Sig_x(Y)$は，鍵 X によるデータ Y の署名を表す．

③ ボランティアは，災害現場において First Responder に接触する．

④ First Responder はボランティアの本人情報 ID_{ve} を受け取り（例えば，First Responder が所持するスマート端末でボランティアの顔写真を撮影する），これを確認する．

⑤ First Responder はボランティアに秘密鍵 SK_{ve}，公開鍵 PK_{ve}，公開鍵証明書 $Sig_{SKfr}(PK_{ve})$を発行し，SK_{ve}，PK_{ve}，$Sig_{SKfr}(PK_{ve})$，$Sig_{SKroot}(PK_{fr})$をボランティアに渡す．

⑥ First Responder は ID_{ve} を暗号化し，$\{Sig_{SKfr}(PK_{ve}), Enc_{PKfr}(ID_{ve})\}$の形でアーカイブする．ここで，$Enc_x(Y)$は，鍵 X によるデータ Y の暗号化を表す．

⑦ ⑤によって，ボランティアは Verifier としてのメッセージ送信が可能となる．その際の通信内容 M は，$\{M, Sig_{SKve}(M), Sig_{SKfr}(PK_{ve}), Sig_{SKroot}(PK_{fr})\}$のデータ形式で送受信される．

⑧ 万一，ある verifier からの情報 M が虚偽であったことが発覚した場合，First Responder は $Enc_{PKfr}(ID_{ve})$を復号し，ID_{ve} を警察に届け出る．警察は，ID_{ve} を手掛かりにして，虚偽情報の発信者を捜索して特定する．

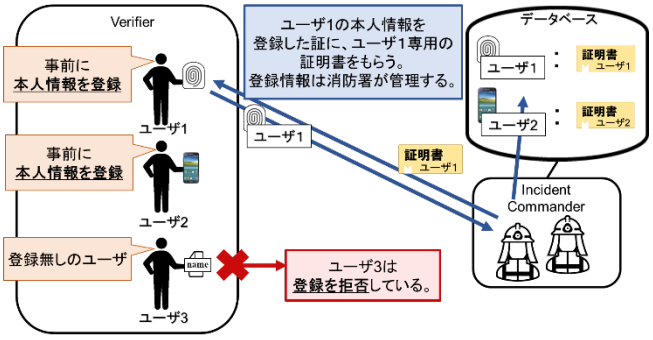

図 2　Verifier の登録

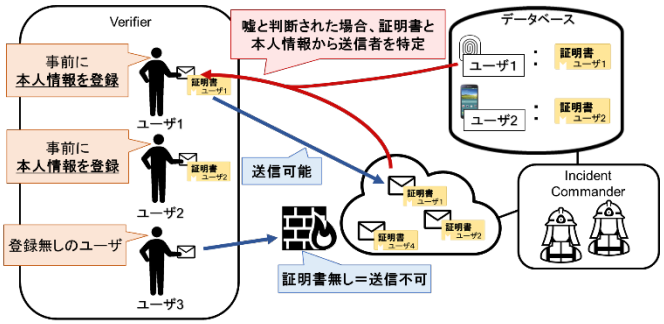

図 3　Verifier からの情報送信

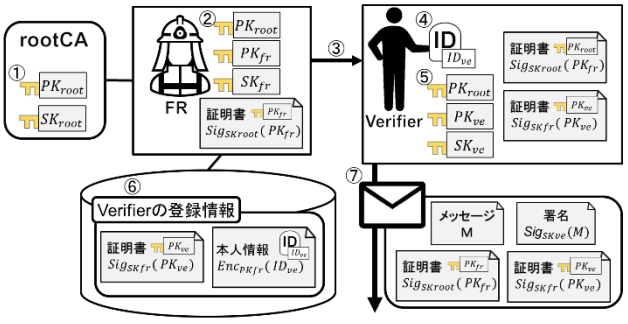

図 4　登録手続き

近年，生体認証が普及し，生体情報から個人を追跡・特定できるということが一般に知られるようになった. さらに，生体情報に関するプライバシー保護の重要性が強く認識されつつある．これは，多くのユーザが“生体情報が漏れてしまうと，個人を特定されるかもしれない”という危機感を有していることの表れといえる．上記の方法は，ユーザが生体情報に対して抱くこの危機感を，シビル攻撃や虚偽情報発信に対する抑止力として利用し，ソーシャルメディアに情報を投入するユーザの信頼性を高めることを狙っている．

ただし，同一の verifier が，ある場所で ve1 と名乗って First Responder fr1 と接触した後に，別の場所で ve2 と名乗って First Responder fr2 と接触した場合，
暗号化をほどくことなく $\{Sig_{SKfr1}(PK_{ve1}), Enc_{PKfr1}(ID_{ve1})\}$ と $\{Sig_{SKfr2}(PK_{ve2}), Enc_{PKfr2}(ID_{ve2})\}$ の突合を検査することはできない．このため本方式は，シビル攻撃に対する耐性は完全ではない．これについては今後の課題である．

本研究では，抑止力を通じて情報の信頼性が高まる仕組みを「抑止力型トラスト」と定義する．抑止力型トラストの利用によって,“嘘が明るみになると発信者が自分であると特定されてしまう”という心理が働き，情報発信者が嘘をつきにくくなる．この結果，本人情報が添付されたメッセージが虚偽情報である可能性が低くなることが期待され，Verifier の信頼性を高めることにつながると考えられる．

4. 抑止型トラストの効果に対するアンケート調査

4.1. 調査項目

提案方式と抑止力型トラストの有効性を確認するため，アンケート調査を行う．本研究では,「ソーシャルメディア上に虚偽情報が流れるのを防ぐにはどうすれば良いか？」という目的に対し,「生体情報を用いた抑止力型トラストを利用することにより虚偽情報の発信を抑制できる」という仮説を立てた．この仮説を明らかにするために以下の調査項目を,アンケート調査によって明らかにする．

RQ. 1 どの本人情報を用いれば，虚偽情報送信の抑止力が高まるのか？

RQ. 2 嘘の程度によって虚偽情報送信の抑止力の効果は変化するか？

RQ. 3 プライバシーを晒す程度と虚偽情報送信の抑止力効果は同等か？

ここで，②の嘘の程度についてだが，嘘には事実を偽装する嘘や誇張する嘘など様々な種類が存在する．嘘の種類によって，ユーザが嘘を発信しやすくなり，抑止力が働きにくくなることが考えられる．本来であれば，嘘を分類できる指標を用いて,それに応じた質問を作成するべきである．しかし,当該指標に関する既存研究の有無を調査した結果,今回のアンケートに適応可能なものが得られなかった．そこで,今回は,「自分が得られる利益」「他人が被る不利益」の 2 軸で嘘を分類し，嘘の分類表である図 5 を作成した．図 5 の分類を基軸とし,「自分が得られる利益」と「他人が被る不利益」が等しい嘘として,「救助」「配給」「いたずら」の 3 つをアンケートに用いることとした．

嘘の分類		他人が被る不利益		
		大	小	無し
自分が得られる利益	大	救助		
	小		配給	
	無し			いたずら

図 5　嘘の分類

4.2. アンケートの質問構成

前節で述べた調査項目を明らかにするため，後述の①~⑦の質問を作成し，アンケートを実施した．番号は，実際に実施したアンケートでの質問順である．また，今回の調査に用いる生体情報と個人情報は「最低限個人を特定できる情報」かつ「スマートフォンで入力可能な情報」を条件に,「顔」「指紋」「声紋」「基本 4 個人情報（氏名＋住所＋生年月日＋性別)」「氏名＋住所」「携帯電話番号」「運転免許証」の 7 つとした．

① 生体認証の利用に関する質問

生体認証をよく利用している人ほど，生体情報が個人を特定できる情報であることを知っているため，プライバシーをさらす程度や抑止力の程度に差が出るのではないか,と考えた. そこで, どのような生体認証を利用しているか,また，その生体認証を利用し始めてどれくらい経つのか,といった質問を行う．

② IMC

今回のアンケートは, 心理尺度による「測定」ではなく,質問文を独自に作成する「調査」である．このような調査が適切に行われるためには，回答者が質問項目の意味を正確に理解し，回答者が考えている回答を，5 件法などの指定された回答形式に合致させ回答させる必要がある．しかし，このような場合，Satisfice 問題が発生し，オンライン調査では大きな影響を与える[10]．Satisfice 問題の中でも，特に強い Satisfice の対策が必須とされている．強い Satisfice とは,「調査項目の内容を理解するための認知コストを払わず，誰でも選択可能な選択肢を選んだり，当てずっぽうに選んだりすること」である．この Satisfice 問題の対策として，事前にスクリーニングを行う方法が広く検討されおり[11][12], その 1 つが, IMC（Instructional manipulation check）である[13]．IMC とは，リッカート尺度や複数選択式の設

問に対して，回答者にあえて「正しく答えないように」求めることで，Satisfice の有無を確認する方法である．IMC は，和文での構成も提案されている[10].

③　プライバシーに関する質問

各プライバシー情報について，その情報を他人に提示する機会があった場合，どの程度のプライバシーを晒していると感じるかを調査するため，それぞれの情報に対して，「どのくらいプライバシーを晒していると思うか」と言う質問を行う．回答には 5 件法を用い，「1：全くそう思わない」「2：あまりそう思わない」「3：どちらとも言えない」「4：まあそう思う」「5：そう思う」とた．本回答の結果をプライバシースコアとする．

④　メッセージ送信に関する質問

抑止力型トラストの有効性を調査するため，被験者には，抑止力型トラストが実装された緊急時通信網で，いたずら・配給・救助の 3 つの嘘を発信しようとしているユーザをそれぞれイメージしてもらい，「嘘のメッセージの送信時に，各情報を添付した場合でも，メッセージをそのまま送信すると思うか」を③と同様 5 件法で回答させる． また，当該質問は添付型と事前登録型の 2 種類の設問文を作成し，回答時に被験者をランダムに添付型回答群と事前登録型回答群に振り分けることとした．

ここでは，「嘘のメッセージを送信すると思うか」という質問をしているため，“値が小さいほど抑止力が高い”ということに注意されたい．この回答結果を抑止力スコアとする．

⑤　生体認証に関する質問

生体認証の知識の差によって，抑止力に差が出る可能性を考え，文献[14]の生体認証の脆弱性に関する記述を参考に，生体認証の知識を問う質問を作成した．その際，複数ある脆弱性のうち，今回は生体認証特有の課題に絞り，利用した．またここでは，前述した Satisfice 問題の対策の一つとして，回答の一貫性を利用したフィルタリングを行うこととした．そのため，次の例の通り，いくつかの脆弱性に関しては，1 つの脆弱性につき 2 つの質問を作成し，回答の一貫性を確認できるようにした．

例）脆弱性「複製：物理的に生体情報を複製できる」

質問①：生体情報は物理的な方法で複製出来ない

質問②：生体情報の成功な偽造物を作ることができる

これらの質問に，Yes/No で回答してもらい，その脆弱性に関する知識の有無を調べる．質問①で No と答えた（つまり正解した）人は，質問②でも Yes と答える（つまり正解する）はずである．不正解のときも同様になるため，ここで回答に一貫性が持てない場合は，文章を読み飛ばしている可能性が高いと判断できる．また，具体的な質問になりすぎると，正しい回答を悟らせてしまう可能性もある．その 2 点に注意し，実験実施者内（著者ら 3 名以上）で推敲を重ね，質問文を作成した．最終的に，17 つの脆弱性から 31 問の質問を作成し，その内一貫性のとれた質問は 11 ペアとなった．さらに，「この質問には必ず右（左）を選択してください」というダミー質問を 2 つ追加した．

⑥　SNS の利用に関する質問

本提案がソーシャルメディアに関するものであることから，日常的な SNS の利用に影響を受けるのではないか，と考えた．そこで，文献[15]を参考に，本研究の分析に利用できそうなものを抜粋し，アンケートの質問に取り入れた．

⑦　基本情報

被験者の基本情報を問う項目は「デモグラフィック項目」と呼ばれ，分析には極めて重要である一方で，プライバシーや個人情報に関するデリケートな質問である[16]．そこで，今回アンケートで質問する項目については，分析に必要な情報のみに絞っている．また，警戒心を抱かせないため，アンケートの最後に聞くよう配慮を行った．

以上の質問からアンケートシステム Limesurvey [7]を利用しアンケートを作成，被験者については，クラウドソーシングサービスである Lancers [8]を利用して募集を行った．また，被験者を募集する際に，「生体認証を利用したことがある人」を調査対象にしていることを明記した．

4.3. 結果

被験者数は，20～60 代の 200 名（男性 127 名，女性 73 名）となった．前述したとおり，回答時に被験者をランダムに添付型回答群と事前登録型回答群に振り分けを行っている．その結果，添付型回答群 86 名，事前登録型回答群 114 名となった．この被験者について，まず IMC でのフィルタリングを行う．IMC の設問指示に従い，回答せずに次に進んだ人たちを遵守群，指示を読まずに回答をしてしまった人たちを違反群とする．その結果，遵守群 121 名，違反群 79 名となった．次に，遵守群 121 名に対して，回答の一貫性を用いたフィルタリングを行う．前節の⑤で記した通り，全 11 ペアの質問を利用し，一貫しなかったペアを誤答数としてカウントし，集計した結果を図 6 に示す．本来であれば，誤答数 0 問の被験者を分析で利用すべきだが，8 名では適切な分析ができないため，今回は誤答数が 0～2 問であれば許容範囲とした．以上より，最終的な分析対象は 66 名となった．以降の結果では，当該 66 名の結果をま

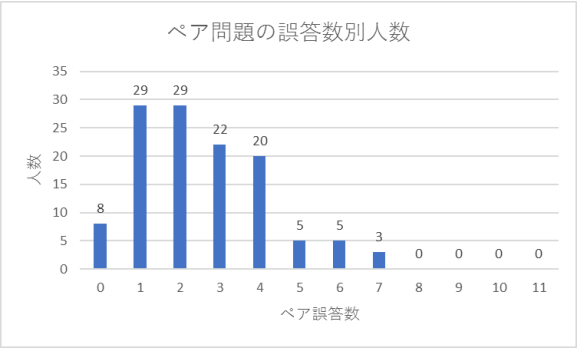

図 6　ペア問題の誤答数別人数

.

表 1 アンケート結果と分析

item / Info	Privacy score	Deterrence score			Effect score			Traceability (Authentication accuracy)	Impersonation resistance
		Prank	Distribution	Rescue	Prank	Distribution	Rescue		
Non	-	1.76 (1.14)	1.98 (1.25)	1.78 (1.21)	-	-	-	0%	×
Face	4.17 (1.14)	4.24 (1.02)	3.68 (1.17)	3.05 (1.40)	0.07 (1.42)	-0.49 (1.65)	-1.12 (1.93)	(100 - EER)%	○
Fing	3.93 (1.15)	4.10 (1.11)	3.44 (1.34)	3.07 (1.31)	0.17 (1.16)	-4.89 (1.38)	-0.85 (1.56)	(100 - EER)%	○
voice	3.27 (1.23)	3.61 (1.18)	3.05 (1.22)	2.73 (1.18)	0.34 (1.54)	-0.22 (1.62)	-0.54 (1.67)	(100 - EER)%	○
PI4	4.80 (0.459)	4.63 (0.662)	3.90 (1.32)	3.39 (1.38)	-0.17 (0.738)	-0.90 (1.37)	-1.41 (1.40)	100%	△
PI2	4.61 (0.628)	4.51 (0.746)	3.80 (1.21)	3.20 (1.36)	-0.10 (0.86)	-0.80 (1.38)	-1.41 (1.53)	99%	△
DL	4.61 (0.771)	4.63 (0.733)	4.10 (1.20)	3.51 (1.43)	-0.02 (1.07)	-0.51 (1.47)	-1.10 (1.73)	100%	○
PN	4.02 (0.790)	4.27 (0.922)	3.54 (1.27)	2.88 (1.47)	0.24 (1.04)	-0.49 (1.52)	-1.15 (1.51)	100%	△

とめたものを記す． 66 名のうち，添付型回答群は 25 名，登録型回答群は 41 名であった．また，ダミー質問に対して誤った回答した被験者が 3 名居たが，IMC のフィルタリングで除外できていたため，結果に影響はない．また，調査に利用したプライバシー情報に関して，「なにもなし : non, 顔 : face, 指紋 : fing, 声紋 : voice, 氏名+住所+生年月日+性別 : PI4, 氏名+住所 : PI2, 運転免許証 : DL, 携帯電話番号 : PN」と表記する．プライバシースコアと事前登録型の抑止力スコアの平均と分散を表 1 の Privacy score と Deterrence Score の欄に示す．表中の値については「平均値（分散）」となっている．また，有効数字は 3 桁とする

4.4. 分析

先行研究では, 4.1 節の調査項目を明らかとするために，アンケートの結果の分析を行った[4]．その結果，本人情報の登録が，虚偽情報発信の抑止力になることが明らかとなった．また，その抑止力は嘘の程度が大きくなるにつれて弱まることも明らかとなった．

一方で，どの本人情報が抑止力型トラストの運用に最適なのかは，明らかとなっていない．そこで本稿では，「3 種類の本人情報（生体情報，個人情報，携帯電話番号）のうち， どの情報が抑止力型トラストに適しているのか？」を調査課題とし，アンケート結果をもとに分析・検討を進める．調査課題を達成するために検討しなければならない項目は以下の通りである．

① 「抑止力」と「プライバシーを晒す程度」の関係
② 各本人情報の追跡可能性について
③ 各本人情報のなりすまし耐性について

①については，先行研究[4]での分析を再掲する．先行研究では，プライバシーを晒す程度と虚偽情報発信の抑止力が同等であるか調査を行うため，プライバシースコアと抑止力スコアについて相関分析を行った．その結果，いたずらと配給に関しては, 低い負の相関があることが分かった．救助に関しては，相関無しである．プライバシーと抑止力の間には強い相関があると予想していたが，予想とは異なる結果となった．プライバシーだけが抑止力に関係するというわけではなく，嘘の深刻度にも抑止力は左右されるからだと考えられる．また，個人差が大きいことも原因の一つだと考えられる．

②と③については，抑止力として働くには，必ず本人を特定できるという必要性がある．そのため，抑止力型トラストに適している本人情報の判断要素として，追跡可能性は必要不可欠な項目である．また，他人の本人情報や偽装した本人情報を登録できてしまうと，これはアンケートでいえば non の項目に当てはまるものとなるため，全く抑止力は働かなくなると考えられる．従って，なりすまし可能性もまた必要不可欠な項目である．表 1 には，②および③についての分析結果も示す．

Effect Score は，「(抑止力スコア) － (プライバシースコア)」の値である．これにより，簡易的にプライバシーを晒す程度と，抑止力のバランスを比較することができる．例えば，いたずらに着目すると，voice の値が最も高いため，バランスの良い本人情報だと判断することができる．ただし，抑止力の大きさまでは考慮していないため，計算の改良が必要と思われる．

Traceability については，基本 4 情報は，それが揃えば確実に本人を特定できる情報とされているため，確実に追跡ができる一方で，氏名＋住所でも一見確実に追跡ができそうだか，登録システムによってはマンション名以降が自由記述になっている場合も考えられる．その場合，同一マンションに同姓同名の住人が 2 人以上居た場合，確実に追跡できるとは限らない．そのため，追跡可能性の高さとして

は，PI4 には及ばないと考えられる．運転免許証と電話番号は，本人に結び付けられているものであるため，確実に追跡可能である．生体情報には，他人を誤って本人と認識してしまう他人受入率と，本人を誤って拒否してしまう本人拒否率がある．通常，生体認証においては，他人受入率と本人拒否率が等しくなる認証閾値における等価エラー率（EER）を認証精度と考える．そこで，生体情報の追跡可能性を（100－EER）%によって表す．

Impersonation Resistance については，本方式において本人情報登録時には First Responder が対面で Verifier の身元確認を行うという前提がある．そのため，生体情報の偽装は難しいものとなるため，なりすまし耐性は十分に確保されていると考えられる．運転免許証にも顔写真が着いているため，同様の理由から，なりすまし耐性は十分であると考えられる．一方で，個人情報を確認するためには他の資料（写真付きの身分証明証）を必要とするため，本人確認に少なからず手間が掛かる．また，他の資料を用意しない場合には，なりすましが容易となる．そのため，なりすまし耐性は相対的に低いと考えられる．携帯電話番号は，Verifier が詐取した他人のスマートフォンを所持している場合には，その真偽を確認できない．そのため，個人情報と同様，なりすまし耐性は低いと考えられる．

5. まとめ

本稿では，災害時 SNS の情報の信頼性を確保するため，2 層情報クレンジングを提案し，そのうち Verifier の信頼性を確保する手法として，「抑止力型トラスト」を運用することを目指し，アンケートにより，その有用性を検証した．その結果，本人情報が虚偽情報発信の抑止力として働くことが分かり，その抑止力は嘘の程度によって変化することが分かった．本稿では先行研究を踏まえ，抑止力型トラストに最適な本人情報を選択するため，新たに分析事項を示し，追加の分析を行ったが，十分であるとは言えない．実装に向け，どのような情報が抑止力型トラストに向いているのか，検討を行っていく．

謝辞　本研究は NICT 受託研究課題 193 による．

参考文献

[1] 内官房情報通信技術（IT）総合戦略室 : 災害対応における SNS 活用ガイドブック，入手先〈https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/pdf/h2903guidebook.pdf〉(2018).

[2] NHK NEWS WEB : 長野県台風 19 号でツイッターの救助要請収集 約 50 件救助に，入手先〈https://www3.nhk.or.jp/news/html/20191110/k10012171761000.html〉(2019).

[3] 総省情報通信政策研究所：平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書，入手先〈http://ww.soumu.go.jp/menu_news/s-news/01iicp01_02000073.html〉(2018).

[4] 北川沢水ら : 生体情報を用いた抑止力型トラスト : 災害時通信の信頼性向上のための仕組みの検討，コンピュータセキュリティシンポジウム 2019 論文集，p84-91 (2019).

[5] Mohammad Jahanian, Toru Hasegawa, Yoshinobu Kawabe, Yuki Koizumi, Amr Magdy, Masakatsu Nishigaki, Tetsushi Ohki and K. K. Ramakrishnan: DiReCT: Disaster Response Coordination with Trusted Volunteers, Proceedings of 2019 International Conference on Information and Communication Technologies for Disaster Management, 2019

[6] Rousseau D. M., Sitkin, S. B., Burt, R. S., and Camerer, C.: Not so different after all: A cross-discipline view of trust, Academy of management review, Vol23, No.3, pp.393–404 (1998).

[7] LimeSurvey : LimeSurvey: the online survey tool - open source surveys（オンライン）, 入手先〈https://www.limesurvey.org/〉(2019).

[8] ランサーズ株式会社：Lancers（オンライン），入手先〈https://www.lancers.jp/〉(2019).

[9] 国立研究開発法人情報通信研究機構 :「スマートコミュニティアを支える高信頼ネットワーク構成技術の研究開発」に対する提案書 (2017).

[10] 三浦麻子，小林哲郎：オンライン調査モニタの Satisfice に関する実験的研究，社会心理学研究，Vol.31，No.1，pp.1-12 (2015).

[11] Berinsky, A. J., Margolis, M. F., and Sances, M. W.: Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. American Journal of Political Science, Vol.l58, pp.739–753 (2014).

[12] Chandler, J., Mueller, P., and Paolacci, G.: Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. Behavior Research Methods, Vol.46, pp.112–130 (2014).

[13] Oppenheimer, D. M., Meyvis, T., and Davidenko, N.: Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, Vol.45, pp.867–872 (2009).

[14] 日立製作所システム開発研究所:「バイオメトリクスセキュリティ評価基準の開発」2003 年度報告書（抜粋），入手先〈https://www.jaisa.or.jp/action/group/bio/pdfs/0715_02.pdf〉（2018）

［15］消費者庁 : インターネット消費者トラブルに関する総合的な調査研究報告書「ＳＮＳに関するアンケート結果」，入手先〈http://www.caa.go.jp/policies/policy/consumer_policy/caution/internet/pdf/adjustments_index_1_170111_0002.pdf〉(2018)

[16] 鈴木淳子：質問紙デザインの技法[第 2 版]，ナカニシヤ出版（2016）