

特別賞

暗号方式を用いた  
傷つかない SNS の提案

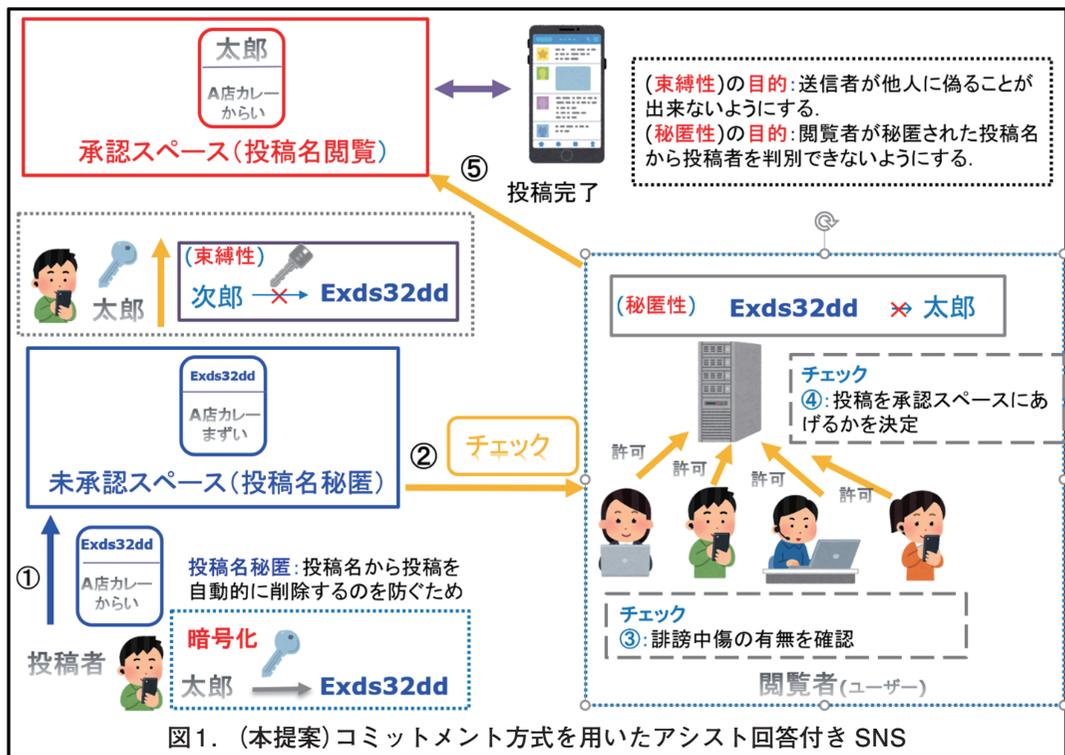
大阪大学大学院 工学研究科 電気電子情報通信工学専攻  
博士後期課程 2 年

宮地 秀至

# 1. 緒言

近年、コロナウィルスの影響により、人と人が対面で出会う機会が減少した。これにより、人との繋がりや情報を入手する機会としてSNS(ソーシャルネットワークサービス)を利用する人々が増加した[8]。SNSでは他のユーザーの情報を簡単に入手でき、自分の情報を容易に他のユーザーに伝達可能であるため、現在の社会の重要なコミュニケーションツールとなっている。

SNSで生活が便利になる一方で、SNSにおける誹謗中傷の問題が深刻になっている[9]。例えば、2020年にSNSの誹謗中傷が原因で芸能人が自殺に追い込まれた[1]。2021年7月2日には、あるSNSユーザーがいじめの加害者であるという誤った情報を多くのユーザーから誹謗中傷される問題が生じた[2]、一方で誹謗中傷について調査した文献によると、『誹謗中傷を行った』と答えたユーザーは全体の1.5%程度しか存在しない[3]。このことから誹謗中傷を行っている本人は意図せず他のユーザーを傷つけていると考えられる。このようにSNS上では、他のユーザーの気持ちを考えて投稿することが難しいため、結果としてSNSを使用するユーザー誰もが被害者かつ加害者になる可能性があることが既存のSNSの問題点である。



一方で、デジタルデータに紐づくプライバシーの保護が重要な課題となっている。例えば、コロナの感染者と接触したときに知らせてくれるシステムではゼロ知識証明が使用されている[10]。よって本研究では、暗号方式を用いてプライバシーを保護しつつ誹謗中傷を削減可能なSNSを提案する。

本提案では、コミットメント方式を用いたアシスト回答付き（閲覧者回答付き）SNSを提案する、本提案 SNS では、以下の3つの特徴を持つ。コミットメント方式を用いた本提案 SNS を図1にて示す。

1. 2つの投稿スペース（未承認スペース、承認スペース）に情報を投稿。
2. 投稿者（送信者）は、未承認スペースにて投稿名を暗号化して投稿。承認スペースでは、復号した投稿名と情報を投稿。
3. 閲覧者（受信者）は、未承認スペースにて投稿名を暗号化した状態でコメントを投稿者に送信可能。

## 2. 定義

### 2.1 コミットメント方式

本提案では、投稿者及び閲覧者のプライバシー保護となりすまし防止のためにコミットメント方式を用いる。ここでは、コミットメント方式の定義について記述する。

コミットメント方式  $com$  は送信者  $S$  と受信者  $R$  により構成され、コミットメントフェーズとデコミットメントフェーズの2つのフェーズで実行される暗号方式である。

本稿で提案するコミットメント方式は束縛性と秘匿性の2つの安全性を持つ。まず、鍵生成アルゴリズムにより、共通パラメータ  $pp$  を作成する。コミットメントフェーズは送信者  $S$  がメッセージ値  $a$  からコミットメント値  $c$  とデコミットメント値  $d$  を生成し、受信者  $R$  にコミットメント値  $c$  を送るフェーズである。デコミットメントフェーズでは  $S$  が  $R$  にコミットメント値  $c$  に対するデコミットメント値  $d$  を送り、 $R$  は受け取った  $d$  からコミットメント値  $c$  を生成できるかどうかを検証する。もし、コミットメント値  $c$  が生成できたならば、 $S$  は  $a$  を出力し、そうでない場合は、 $\perp$  を出力する。

コミットメント方式は以下の3つのアルゴリズムから構成される。

- ・鍵生成： $Keygen \rightarrow pp$
- ・コミットメントフェーズ： $S(pp, a \in \mathbb{M}) \rightarrow (c, d)$
- ・デコミットメントフェーズ： $R(pp, a \in \mathbb{M}, c, d) \rightarrow a \text{ or } \perp$

### 2.2 コミットメント方式における束縛性：

$\mathcal{A}$  を敵対者、 $pp$  を共通パラメータ、 $d, d'$  をデコミットメント値、 $c$  を秘匿値、 $R$  を入力  $(d, PP)$  から  $c$  を出力するアルゴリズムとする。以下の確率が成立するとき、コミットメント方式は束縛性を満たすという。

$$\Pr \left[ \begin{array}{l} Keygen \rightarrow pp \\ \mathcal{A}(pp) \rightarrow (d, d', c): s.t. d \neq d' \wedge \\ R(pp, d) = R(pp, d') = c \end{array} \right] < \varepsilon(k)$$

### 2.3 コミットメント方式における秘匿性：

$\mathcal{A}$  を敵対者、 $pp$  を公開パラメータ、 $c_1, c_2$  を秘匿値とする。以下の確率が成立するとき、コミットメント方式は秘匿性を満たすという。

$$|\Pr_{c_1}[\mathcal{A}(pp, c_1) = 1] - \Pr_{c_2}[\mathcal{A}(pp, c_2) = 1]| < \varepsilon(k)$$

### 3. 着想に至った経緯

この章では、Twitter や Instagram などの既存の SNS における誹謗中傷に対する機能を説明し、その機能における問題点を記述する。

#### 【誹謗中傷の投稿に対する既存の SNS における機能】

- ・ Twitter：誹謗中傷であると認識した投稿に対して Twitter にて報告できる。
- ・ Instagram：投稿に対するコメントや DM(ダイレクトメッセージ)の機能を非表示にする [5]。

#### 3.1 誹謗中傷の投稿に対する Twitter の問題点

Twitter では、自身が誹謗中傷であると認識した投稿を直接 Twitter にて報告可能であるが、『投稿の削除が実行されない場合は弁護士に相談ください』という web サイトを多くみかける [4]。これは Twitter の運営側が行う誹謗中傷の投稿の削除が追い付いていないからである。つまりこの報告機能だけでは、誹謗中傷の報告が多いために深刻な誹謗中傷の投稿を見逃す可能性がある。ゆえに深刻な誹謗中傷を防ぐためには報告機能だけでは解決できないのである。

#### 3.2 誹謗中傷の投稿に対する Instagram の問題点

Instagram は写真や動画をベースにした SNS である。Instagram で問題になる投稿として、『これをすれば二重になれる』や『これをやるだけで鼻が高くなる』といった情報の信頼性が不明な投稿である。これらの投稿のコメント欄では、『この方法を行っても効果がなかった』や『この方法は正しくないと言われている』というコメントが投稿されている。投稿のコメントの内容が正しいかどうかは不明だが、これらのコメントは『この投稿は正しくない可能性もある』とユーザーへの警鐘を与える。しかし、コメント欄が閉鎖された場合、信頼性のない投稿が行われた場合でも削除されず残り、その情報が拡散される可能性がある。つまりコメント欄を閉鎖する機能では、信頼性の欠ける情報の拡散を防ぐことはできないのである。

#### 3.3 本研究の提案と本研究の意義

誹謗中傷は各個人によって捉え方が異なるので、深刻な誹謗中傷がある投稿の早急な削除ができなくなる。また、元々の投稿自体に誹謗中傷がある場合、投稿に対するコメント欄やダイレクトメッセージの制限は問題の解決策にはならない。

本研究では、投稿者や閲覧者のプライバシーを保護し、誹謗中傷の情報や信頼のない情報を自然に削除可能な SNS を提案する。

本申請内容は、既存の SNS に存在する誹謗中傷を削減するという社会的意義を持つ。また本提案 SNS では、量子コンピュータが実現されても安全なプライバシー保護を実現する。これにより量子コンピュータが実現されても、安心かつ傷つかない SNS の提供を可能である。

## 4. 本提案の SNS

本章では、本提案  $Com\_SNS$  について説明する。まず4.1章にて本研究で使用する用語とアシスト回答について記述する。4.2章では、 $Com\_SNS$  の構成手順を記述し、4.3章において本提案 SNS の新規性について記述する。さらに4.4章にて、構成した  $Com\_SNS$  がコミットメント方式の束縛性と秘匿性を満たすことで、他の悪意のあるユーザーからでも安全に構成可能なことを示す。

### 4.1 本研究で使用する用語

- ・  $pp$  : 共通パラメータで全ユーザーに与えられる
- ・  $sk$  : 秘密鍵
- ・  $pk$  : 公開鍵
- ・  $M$  : 投稿内容
- ・  $r$  : 乱数値
- ・  $k$  : セキュリティパラメータ
- ・  $\varepsilon(k)$  : 無視可能関数
- ・ ユーザー : 本提案の SNS を使用している人
- ・ 送信者 : 投稿を行うユーザー
- ・ 受信者 : 投稿をチェックするユーザー
- ・ 未承認スペース : 送信者によって投稿された投稿を最初に置くスペース
- ・ 承認スペース : チェックされた投稿が置かれるスペース
- ・ アシスト回答 : 未承認スペースにて投稿を確認した後、『誹謗中傷あり』『誹謗中傷なし』の2つの選択肢から投稿内容の結果を記録。

### 4.2 コミットメント方式を応用した本提案 SNS の構成方法

本章では、本提案  $Com\_SNS$  を用いた投稿における手順について記述する。

[本提案 SNS への投稿における手順]

投稿者および閲覧者 :

0. 投稿名を登録する。

投稿者 :

1. 投稿内容  $M$  を作成する。
2. 投稿者は秘密鍵  $sk$  と乱数  $r$  をランダムに選択し、 $S(sk, r, pp, M)$  からコミットメント値  $c$  を作成する。
3.  $c$  と投稿内容  $M$  を未承認スペースにて投稿する。このとき、投稿名がコミットメント値  $c$  として送信される。

閲覧者 :

4. 未承認スペースにおいて閲覧者が投稿内容  $M$  に誹謗中傷があるかどうかを確認し、承認スペースに載せてもよいかを判断する。
5. 投稿内容の結果を『誹謗中傷あり』『誹謗中傷なし』の2択から選択し、閲覧者回答として送信者に送信

#### 投稿者：

- 受信者から『誹謗中傷なし』と判断されると、承認スペースにて投稿内容  $M$  とアシスト回答が送信され、送信者にアシスト回答の結果が送信される。送信者はアシスト回答を受信した後、承認スペースにて  $(sk, r)$  を承認スペースへ送信する。ここで、送信者が  $(sk, r)$  を承認スペースにて送信しない場合、次の閲覧者のフェーズに進まないで、承認スペースにて投稿が表示されない。

#### 閲覧者：

- 閲覧者が  $(sk, r, M)$  からコミットメント値  $c$  を作成できるかどうかを検証することで、投稿名が正しいかどうかを確認する。一致した場合、 $(sk, r, M)$  がコミットメント値  $c$  と関連した値であることが判明できるので、承認スペースへの投稿を許可し、送信者の元に投稿が承認スペースにて配置されたことを通知する。このときアシスト回答の結果も投稿と共に表示される。
- $(sk, r, M)$  を用いた  $c$  との検証が一致しない場合、送信者に一致しないことを通知する。
- 承認スペースにて投稿名と投稿内容が明示された形式で表示される。

### 4.3 本提案 SNS の新規性 / 独創性と新たな機能

#### 【新規性】

- 誹謗中傷のある投稿が承認スペース上に拡散されないための閲覧者回答機能：

閲覧者回答付き SNS では送信者(投稿者)と受信者(閲覧者)の双方に対する効果を持つ。  
(閲覧者に対する効果) 閲覧者は承認スペースにて『誹謗中傷あり』、『誹謗中傷なし』の2つの選択肢から1つを選択し、閲覧者回答を記録する。

(投稿者に対する効果) 承認スペースのある投稿に対して閲覧者が回答した閲覧者回答に『誹謗中傷あり』という回答が一定数以上存在した場合、その投稿を承認スペースから削除し、投稿者にアシスト回答と削除された結果が送信される。ただし『誹謗中傷あり』というアシスト回答が基準以下の場合、4.1章の手順6のように投稿者にアシスト回答が送信され、承認スペースへの投稿が可能となる。

- アシスト回答を受信者の投稿名を秘匿した状態で送信者にコメントを送信可能 / コミットメント方式を用いて安全に設計

承認スペースにて受信者が誹謗中傷のある投稿を確認してアシスト回答を記録する際に、受信者の名前は秘匿された状態で承認スペースにて、アシスト回答が保存される。一方で投稿者は、受け取ったアシスト回答の結果に不満がある場合、閲覧者の名前を開示請求することが可能である。

投稿者から開示請求が行われるまで閲覧者の名前が開示されない仕組みをコミットメント方式の束縛性と秘匿性を用いて実現する。秘匿された閲覧者の名前を他のユーザーから識別できないことをコミットメント方式の秘匿性から実現し、閲覧者が名前を偽造できないことをコミットメント方式の束縛性から実現する。

(強み) アシスト回答を束縛性と秘匿性を用いて実現することにより、悪意のある閲覧者が故意に『誹謗中傷あり』と回答する行動に対する抑止力となる。

- 承認スペースにおける投稿の確認を人間が実行する点：

(強み) AI 単独の判断ではなく、受信者も情報を判断することで、一見誹謗中傷のように思えるユーモアのある投稿の削除を回避する機能をもつ。

## 【独創性】

1. ユーザーによる不正を暗号技術により防ぐことができる点(投稿名を暗号化する(秘匿する))。  
(強み)投稿の内容をユーザー(閲覧者)が未承認スペースにて確認するが、投稿名は秘匿した状態で未承認スペースに保管される。これは閲覧するユーザーが投稿内容を確認せず、投稿名のみから『この投稿は誹謗中傷のある投稿であるので削除すべきである』と実行するユーザーの行動を防ぐためである。よって、ユーザーの悪意のある行動を防ぐことができることが強みである。
2. ユーザーの不正が理論的に回避可能(束縛性、秘匿性の達成)未承認スペースでは以下の不正が行われる可能性が存在する。
  - ・ 閲覧者(投稿を確認するユーザー)による不正：(不正方法)未承認スペースにおいて、秘匿されている投稿名を識別しようとする攻撃が存在する。この攻撃を防ぐことで**秘匿性を満たす**という。
  - ・ 投稿者(投稿を投稿するユーザー)による不正：(不正方法)未承認スペースにて閲覧者から、投稿を承認スペースにてアップすることを許可された後、投稿者が自分の投稿名を承認スペースに送信するが、ここで投稿者の投稿名と違う名前で同じ暗号値が生成できてしまうことで他の人になり済ませてしまう。投稿者が他の投稿名になりすましができないことを、**束縛性を満たす**という。(強み)本 SNS の各ユーザーが不正できないシステム(束縛性と秘匿性を満たすことの証明と検証)の実現により、本提案方式を他のプラットフォームへと応用可能な点。

### 4.4 本提案 Com\_SNS の安全性証明に必要な準備

本提案 SNS の構成方式である本提案 Com\_SNS において、未承認スペースでは投稿名をコミットメント値として秘匿された状態で保管する。投稿者または閲覧者が悪意のあるユーザーであった場合に、以下の攻撃が想定される。

#### ・ [受信者が悪意のあるユーザーの場合]

閲覧者が悪意のあるユーザーの場合、未承認スペースにて閲覧者がコミットメント値のみから投稿名を解読することを試みる。

#### ・ [送信者が悪意のあるユーザーの場合]

投稿者が悪意のあるユーザーの場合、承認スペースに投稿が移行されるときに、未承認スペースに用いた投稿名  $sk$  と異なる投稿名を投稿することを試みる。

以上の2つの攻撃を防ぐために本提案 Com\_SNS に必要な定義を定義する。

#### 定義(送信値偽造不可能性)：

$\mathcal{A}$ を敵対者、 $pp$ を公開パラメータ、 $sk, sk' \in \mathbb{M}$ ,  $r, r' \in$ 乱数、 $c$ をコミットメント値、 $R$ を入力( $pp, sk, r$ )から $c$ を出力するアルゴリズムとする。以下の確率が成立するとき、本方式は計算量的送信値偽造不可能性を満たすという。

$$\Pr \left[ \begin{array}{l} \text{Keygen} \rightarrow pp \\ \mathcal{A}(pp) \rightarrow (sk, r, c), (sk', r', c) : s.t. sk \neq sk' \wedge r \neq r' \wedge \\ R(pp, sk, r) = R(pp, sk', r') = c \end{array} \right] < \varepsilon(k)$$

定義 (秘匿値識別不可能性) :

$\mathcal{A}$  を敵対者、 $pp$  を公開パラメータ、 $c_1, c_2$  を秘匿値とする。以下の確率が成立するとき、本方式は秘匿値偽造不可能性を満たすという。

$$|\Pr_{c_1}[\mathcal{A}(pp, c_1) = 1] - \Pr_{c_2}[\mathcal{A}(pp, c_2) = 1]| < \varepsilon(k)$$

#### 4.5 本提案 $Com\_SNS$ の安全性証明

本章では、コミットメント方式の束縛性と秘匿性が成立することで、本提案  $Com\_SNS$  が送信値偽造不可能性と秘匿値識別不可能性を満たすことを証明する。

定理 1.

$k$  をセキュリティパラメータ、 $pp$  を公開パラメータとし、 $sk$  を投稿者  $S$  の秘密鍵、 $r$  を乱数とし、 $c$  を未承認スペースにて送信したコミットメント値とする。コミットメント方式の束縛性が成立するとき、送信値偽造不可能性も成立する。

(証明)

$\mathcal{A}$  を本提案  $Com\_SNS$  の送信値偽造不可能性を破ることができる敵対者が存在することを仮定する。このとき、以下の式(1)が成立する。

$$\Pr \left[ \begin{array}{c} \text{Keygen} \rightarrow pp \\ \mathcal{A}(pp) \rightarrow (sk, r, c), (sk', r', c): s. t. sk \neq sk' \wedge r \neq r' \wedge \\ R(pp, sk, r) = R(pp, sk', r') = c \end{array} \right] > \varepsilon(k) \dots (1)$$

ここで、 $d = (sk, r)$ 、 $d' = (sk', r')$  とする。このとき、上記の式は

$$\Pr \left[ \begin{array}{c} \text{Keygen} \rightarrow pp \\ \mathcal{A}(pp) \rightarrow (d, c), (d', c): s. t. d \neq d' \wedge \\ R(pp, d) = R(pp, d') = c \end{array} \right] > \varepsilon(k) \dots (2)$$

と変形することができる。(2)の式はコミットメント方式の束縛性を破る式を表しており、本提案  $Com\_SNS$  の送信値偽造不可能性を破ることができる敵対者  $\mathcal{A}$  が存在する場合、コミットメント方式の束縛性をも破ることができる敵対者が存在する。ゆえに対偶より、コミットメント方式の束縛性が成立する場合、本提案  $Com\_SNS$  の秘匿値識別不可能性が成立する。

定理 2.

$k$  をセキュリティパラメータ、 $pp$  を公開パラメータとし、 $c_1, c_2$  を未承認スペースにて送信したコミットメント値とする。コミットメント方式の秘匿性が成立するとき、送信値偽造不可能性も成立する。

(証明)

$\mathcal{A}$  を本提案  $Com\_SNS$  の送信値偽造不可能性を破ることができる敵対者であると仮定する。このとき、以下の式(3)が成立する。

$$|\Pr_{c_1}[\mathcal{A}(pp, c_1) = 1] - \Pr_{c_2}[\mathcal{A}(pp, c_2) = 1]| < \varepsilon(k) \dots (3)$$

(3)の式はコミットメント方式の秘匿性の定義を破ることを表している。ゆえに、本提案 *Com\_SNS* の送信値偽造不可能性を破ることができる敵対者が存在すると仮定すると、コミットメント方式の秘匿性を破ることができる敵対者が存在する。対偶より、コミットメント方式の秘匿性が成立するとき、本提案 *Com* の送信値偽造不可能性が成立する。

## 5. 本提案 *Com\_SNS* の実用化

### 5.1 ある閲覧者が未承認スペースの誹謗中傷のない投稿に対して誹謗中傷ありとアシスト回答に記録する問題について

4章にて悪意のある閲覧者が故意に『誹謗中傷あり』と回答することを防ぐ抑止力の仕組みについて記述した。5.1章では、最初に故意に『誹謗中傷あり』と回答する悪意のあるユーザーを認識する条件について記述する。次に故意に『誹謗中傷あり』と回答する悪意のあるユーザーが存在する場合に、本提案 *Com\_SNS* で実行する仕組みについて記述する。

[故意に誹謗中傷ありと回答する悪意のあるユーザーを認識する条件]

明らかに悪意のある行動があると多くの投稿者や閲覧者から指摘がある場合

[誹謗中傷ありと回答する悪意のあるユーザーに対して実行すること]

閲覧者が投稿者から開示要求が行われたとしても、閲覧者が悪くない場合もあれば、投稿者が腹いせに開示要求を行う可能性もある。また、お互いの意見が届かなかっただけの可能性もある。本提案では、投稿者と閲覧者に対してどちらも悪意のある行動を実行することによるデメリットが存在するので、あえて悪意のある行動をとる確率は少ないと考えられる。

ゆえに本提案では、一定数以上の開示要求が行われる閲覧者に対しても未承認スペースにおける投稿の確認を行う権利を引き続き承認する。一方で、明らかに悪意のある行動があると多くの投稿者や閲覧者から指摘がある場合は、未承認スペースにおける投稿の確認する権利を剥奪する可能性がある。

### 5.2 本提案 *Com\_SNS* が実用化/商品化された際の効果

誹謗中傷を削減する仕組みとしてAIを使う方法[11]やTwitterのように運営者が直接削除する方法が存在している。しかし、AIや運営者が直接削除するシステムではユーザーに『自分の主観で投稿を判別しているのでは?』と感じさせてしまい、誹謗中傷の投稿は削除できても、誹謗中傷を行う風土自体を変えることができない。

本提案 SNS では、多くの人々が幅広く投稿を確認することで、『誹謗中傷を行わないように気を付けて発言する』という風土を実現することを目標とする。

本提案 SNS が実用化されることによって、誹謗中傷を投稿しにくいシステムの開発が注目されるようになり、安全に快適にシステムを使用するユーザーが増加する効果が期待される。

## 6. まとめ

本研究では、誹謗中傷を削減する *Com\_SNS* をプライバシーを保護できるようにコミットメント方式を用いて提案した。本提案 *Com\_SNS* では、2つのスペースと1つの閲覧者回答と、閲覧者回答における選択肢を誹謗中傷あり、誹謗中傷なしの2つに分けることで、悪意

のある投稿者や閲覧者に対しても本提案 Com\_SNS が実際に機能できることを可能にした。

また本提案は誹謗中傷を削減できる SNS であるが、この方式を組織的に利用することで、組織内相談窓口システムというようなデジタルコミュニケーションツールとしても応用可能である。

## 謝 辞

本研究を進めるにあたり、ご指導を頂きました宮地充子教授(大阪大学大学院工学研究科)に謹んで御礼申し上げます。また、本研究の一部は文部科学省「Society5.0に対応した高度技術人材育成事業成長分野を支える情報技術人材の育成拠点の形成(enPiT)」、文部科学省の平成30年度「Society 5.0実現化研究拠点支援事業」、さらにJSPS 科研費 JP21H03443、JP21K11751の助成、次世代研究者挑戦的研究プログラム JPMJSP2138の支援を受けています。

## 参考文献

1. <https://www.bbc.com/japanese/52787007>
2. <https://www.asahi.com/articles/ASP7233X9P72UTIL003.html>
3. ネット炎上の実態と政策的対応の考察—実証分析から見る社会的影響と名誉毀損罪・制限的本人確認制度・インターネットリテラシー教育の在り方—, [https://www.soumu.go.jp/iicp/chousakenkyu/data/research/icp\\_review/11/11-4yamaguchi2015.pdf](https://www.soumu.go.jp/iicp/chousakenkyu/data/research/icp_review/11/11-4yamaguchi2015.pdf)
4. <https://itbengo-pro.com/columns/23/>
5. <https://nlab.itmedia.co.jp/nl/articles/2108/12/news146.html#:~:text=Instagram%E3%81%AF8%E6%9C%8811,%E3%82%84%E3%82%8A%E3%81%A8%E3%82%8A%E3%82%92%E9%98%B2%E3%81%92%E3%82%8B%E3%81%A8%E3%81%84%E3%81%86%E3%82%82%E3%81%AE%E3%80%82>
6. <https://itbengo-pro.com/columns/23/>
7. <https://www.biz.ne.jp/matome/2002865/>
8. 新型コロナウイルス感染拡大前後で SNS 利用増加は 34% - 新しい生活様式における消費者の SNS 利用実態調査結果(アライドアーキテクト株式会社調べ)  
<https://www.profuture.co.jp/mk/column/11058>
9. コロナ禍で深刻化する差別コメント、誹謗中傷。ネット炎上の病理と防止策とは？  
<https://newswitch.jp/p/25392>
10. Joseph K. Liu et al. Privacy-Preserving COVID-19 Contact Tracing App: A Zero-Knowledge Proof Approach (2021), ISPEC2021
11. 「ヤフコメ」AI で誹謗中傷を非表示「利便性と安心」どう両立  
<https://www.sankei.com/article/20211030-NKKXXGSQYJNRNJKRZINBRX36P4/>