情報処理学会論文誌 Vol.62 No.9 1487–1498 (Sep. 2021)

Bitcoin取引履歴の特徴量に基づくアドレス識別リスクの評価

松本 寛輝[1,a)] 菊池 浩明[2,b)]

受付日 2020年11月30日, 採録日 2021年6月7日

概要：Bitcoinでは，プライバシ保護の観点からアドレスの使用を1度限りとし，ユーザは自分のアドレスを明かさないようにすることが推奨されている．しかしながら，取引に利用されているアドレスの中には再利用されているものも多く，送金を行った取引情報から特徴量を学習することで同一ユーザが所有するアドレスの識別リスクを分析した研究成果が報告されている．本研究では，Bitcoinの受け取り取引に着目した新たな2つのアドレス識別手法を提案し，識別精度を明らかにすることを目的とする．Bitcoinの取引情報を用いて識別実験を実施し，アドレスの取引回数と利用目的の観点による精度の変化を評価する．実験結果より，提案手法のアドレス識別精度は既存手法を利用した識別精度よりも統計的に有意な水準で高いことが示された．

キーワード：Bitcoin，暗号通貨

Study on Risk of Bitcoin Address to be Identified from the Features of Transaction History

Hiroki Matsumoto[1,a)] Hiroaki Kikuchi[2,b)]

Received: November 30, 2020, Accepted: June 7, 2021

Abstract: Crypto-currency such as Bitcoin suggests a use of one-time address for the privacy enhancement. However, addresses were often reused, several studies reported that learning transaction history has a risk of address identification. Our study purposes two new methods that focus on features related output address. Our study aims to evaluate risk to be identified by the proposed methods with respect to on the number of transactions per address and the kinds of address usages. In this paper, our experiment shows that our method identifies addresses more accurately than the conventional method in statistically significant level.

Keywords: Bitcoin, crypto-currency

1. はじめに

Bitcoin [1] を代表とする暗号資産は高い匿名性を持つとされ，国を超えた送金や投資目的など様々な用途で利用されている．しかしながら，Bitcoinが持つ匿名性はBitcoinアドレスが持つ仮名のランダム性に基づくものであり，取引履歴の統計情報からのアドレスの識別やユーザ居住地などの属性情報が推定されるリスクが知られている [2], [3], [4]．この問題に対してオープンソースプロジェクトのBitcoin.orgではプライバシ保護の観点からアドレスの使用を1度限りとし，ユーザは自分のアドレスを明かさないようにすることを推奨している [5]．ユーザは取引ごとに新たなアドレスを作成し，利用することで匿名性を高めることが期待できる．

その一方で，用途によってはアドレスが長期間にわたり繰り返し利用されることがある．代表的な例として，自身のアドレスへ寄付を受け付ける目的で掲示板やSNSなどに公開する場合やMining pool事業者などが営利用アドレスを意図的に使い回す場合などがある．同じアドレスを一定回数繰り返し使用することで同一ユーザの所有するアドレスが識別されるリスクは否定できない．

1 明治大学大学院先端数理科学研究科
Graduate School of Advanced Mathematical Sciences, Meiji University, Nakano, Tokyo 164–8525, Japan
2 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164–8525, Japan
a) cs192026@meiji.ac.jp
b) kikn@meiji.ac.jp

本稿の初稿は2020年10月のコンピュータセキュリティシンポジウム2020（CSS2020）で報告された．

現に，Meiklejohn らは，取引の *Input* に並列に指定された複数アドレスが同一ユーザによることを指摘し，そのリスクを分析している [6]．また，永田らは，取引の宛先を学習することで，同一ユーザの所有するアドレスであるかを識別できることを示している [8]．

しかしながら，取引履歴から学習されるユーザの特徴はそれらに限らない．本研究では，取引の *Input* でなく *Output* に指定されたアドレスにも，ユーザを特定する重要な特徴があることを新たに主張する．加えて，取引の宛先だけでなく，送信元の情報の特異性にも着目し，新たな識別方式を 2 つ提案する．

一方，識別精度を正確に評価するのは困難である．精度を左右する要素として，そのアドレスに関わる取引の頻度や，交換所などのサービス事業者によるものか，単なるエンドユーザのものかといった利用目的など，多くの条件を考慮する必要があるためである．そこで，掲示板，ATM，交換所，マイニングプール，Darkweb サービスの 5 つの代表的なアドレスを取り上げ，それらによる識別率の変化を明らかにする．

本稿の貢献は次のとおりである．

- 新たに 2 つの識別方法を提案する．
- 10 年間のアドレスデータセットを用いて，長期間継続して利用されているアドレスの有する識別リスクを定量化する．
- 5 種類の利用目的を考慮したアドレスの識別精度を明らかにする．

本稿の構成は以下のようになっている．2 章では Bitcoin アドレスの識別と取引構造を定義し，既存手法を説明する．3 章では本稿で提案する新たな識別手法とアドレスの利用目的について述べる．4 章で 4 つの手法を用いたアドレス識別実験を実施する．5 章では実験結果に関する考察と既存手法と提案手法の精度を比較した検定などの評価を行い，最後に 6 章で本稿の結論をまとめる．

2. 基本定義

2.1 Bitcoin アドレスの識別の定義

Bitcoin アドレスの管理とは，ユーザがアドレスに対応する秘密鍵を保有し，資産を自由に移動する権限を有することと定める．1 ユーザは複数のアドレスを管理することができる．

Bitcoin アドレスの識別問題とは，あるユーザが管理している複数のアドレスを与えて，そのユーザの管理する他のアドレスを他人のアドレスから識別する問題である．

たとえば，**図 1** のアドレスの識別例を考えよう．ここではユーザ A が a_1，a_2，a_3 のアドレスを管理している．a_1 が与えられたとき，対象全アドレスの集合から a_2 や a_3 を選べれば識別が成功したと考える．Bitcoin アドレスが識別されても，必ずしもアドレスの所有者が特定されるわけ

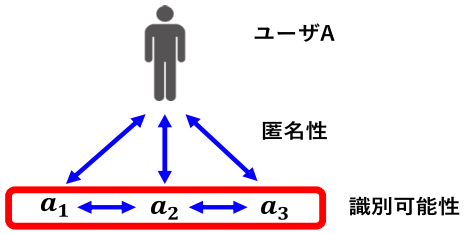


図 1 Bitcoin アドレスの識別定義
Fig. 1 Definition of address identification.

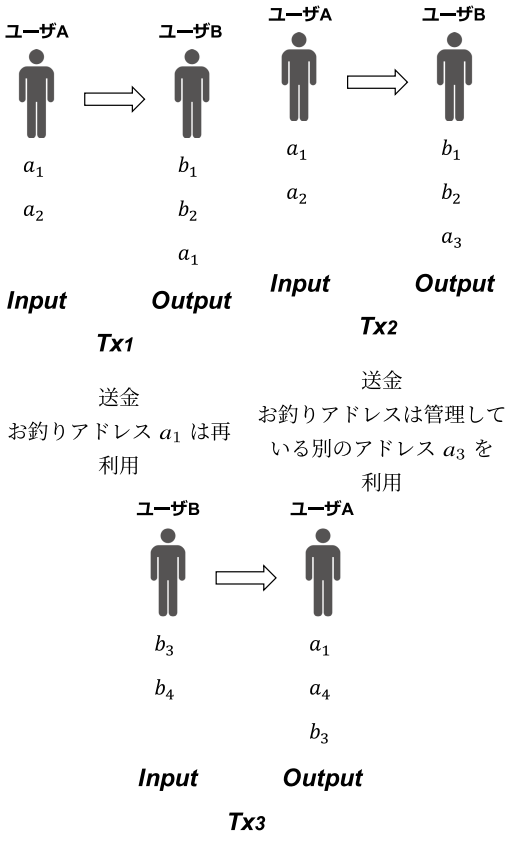


図 2 Bitcoin の送金，受け取りを行う取引構造
Fig. 2 Examples of transactions of Bitcoins.

ではないことに注意が必要である．アドレス a_1，a_2，a_3 の情報から管理しているユーザの名前 A などが特定されることではない．

2.2 Bitcoin アドレスの取引

Bitcoin アドレスに関する取引の例を**図 2** に示す．Bitcoin の取引 Tx 中に，送金を行うアドレスは *Input*，Bitcoin を受け取るアドレスは *Output* に指定されている．図 2 ではユーザ A が管理しているアドレス a_1，a_2 を用いてユーザ B が管理しているアドレス b_1，b_2 へ送金を行っている．このとき，ユーザ A は送金を行った際のお釣りを受け取るため *Output* に自身のアドレス a_1 を指定している．図 2 の Tx_1 では，送金時に使用したアドレス a_1 を再度用いてお釣りを受け取っているが，Tx_2 のように，ユー

ザ A が自分が管理する別のアドレス a_3 でお釣りを受け取ることもある．Tx_3 では，ユーザ A が管理しているアドレス a_1，a_4 に対してユーザ B が管理しているアドレス b_3，b_4 から送金を行っている．このとき，ユーザ B は送金を行った際のお釣りを受け取るため *Output* に自身のアドレス b_3 を指定している．

2.3 関連研究

Bitcoin アドレスの匿名性に関して次の研究 [6], [8] がある．加えて，Bitcoin アドレスや管理ユーザのプライバシに焦点を当てた研究報告についても述べる．

2.3.1 Meiklejohn らの入力アドレスを用いた識別

Meiklejohn らは取引の入力アドレスを用いたアドレスのグループ化手法を提案している [6]．

Bitcoin では，不正な送金操作への対策として楕円曲線を使用した ECDSA デジタル署名アルゴリズムを採用している．この署名は，トランザクションの入力アドレス（公開鍵）に対応する秘密鍵を用いて作成され，採掘者（マイナ）によって検証が行われる．Meiklejohn らはこの仕組みを利用し，Bitcoin の送金元アドレス（入力アドレス）が 1 つのトランザクションに複数指定されている場合，すべての入力アドレス a の秘密鍵を同一のユーザ（管理者）によって管理されていると推測した．たとえば，図 2 の取引情報からは，アドレス a_1，a_2 が取引 Tx_1，Tx_2 の *Input*（入力アドレス）に指定されているため，a_1，a_2 は同一のユーザによって管理されていることが推測される．トランザクションの署名には，a_1 と a_2 の両方の秘密鍵が必要なことから，推定精度は高いと考えられる．この仕組みは Bitcoin 取引の分析方法としても利用されている [9]．

Kappos らは暗号資産 Zcash [*1] における識別実験を実施しており，Bitcoin 以外の暗号資産にも同様の識別手法が有効であることを報告している [7]．

2.3.2 永田らの送金先アドレス集合を用いた識別

永田らは取引の送金先アドレス集合を用いた識別手法を提案している [8]．

この手法は，アドレスの取引頻度と送金先履歴の宛先アドレスを用いて，過去に行った取引履歴からアドレスを識別する．永田らの方式は，ユーザごとに取引を行う固有の相手が決まっているので，取引の宛先アドレス情報からユーザの追跡が可能になる，という仮説に基づいている．この仮説では，不特定多数のユーザに対して送金を行う機会が少ないという統計的な性質を根拠としている．したがって，Meiklejohn らの入力アドレスを利用した推定と比較して，推定精度は低い．また，永田らはアドレスの取引数はアドレスの識別に影響を与えない，と主張している．

永田らによる宛先アドレスを用いたアドレス識別手法で用いる送金時に利用するアドレスはユーザによって任意に変更することが可能である．したがって，自身が管理していないアドレスを意図的に用いることでアドレス識別を回避できる．

2.3.3 プライバシに関する研究

Bitcoin のアドレスやユーザの属性を推定する複数の研究が行われている．Dupont らはアドレスの取引時刻に着目し，取引の時刻分布からユーザが居住している地域のタイムゾーンを推定する方式を提案している [2]．井垣らはアドレスの平均取引時間分布を用いることで，最大で 77%のアドレスのタイムゾーンを推定可能と報告している [3]．我々は交換所を利用しているユーザのタイムゾーンに関する分析を行っている [4]．

アドレスの利用目的を推定する研究では，Harlev らが著名なサービス事業者のアドレスを用いてアドレスの利用目的を推定し，77%の正解率で識別可能であることを報告している [12]．我々は Bitcoin の利用者をユーザとサービス事業者の 2 つの観点から，それぞれの利用方法に違いが生じることを報告している [13]．

Bitocin の取引構造を分析し，資金の流れを追跡する研究が行われている．Ron らはブロックチェーン上に記録されたすべての取引を分析し，複数のアドレスに共通した特異な取引構造があることを報告している [9]．Ron らが行った取引パターンの分析に対して，取引の特徴量を学習させないためにミキシングサービスが利用されることがある．廣澤らは実際にミキシングサービスを利用した追跡の困難性について報告している [10]．Garba らは web サイト上に公開されているアドレスを収集し，Bitcoin の支払い時における中間者攻撃のリスクについて考察している [14]．Huang らはランサムウェアの支払いに使用された Bitcoin アドレスに注目し，ランサムウェアの被害者フォーラムで報告されたアドレスを収集し，資金の流れを追跡している [15]．坂間らは Bitcoin の取引時に使用するデジタル署名を分析し，利用者が使用しているウォレットが原因でアドレスの秘密鍵が漏洩する危険性について考察している [11]．

3. 提案方式

3.1 アドレス集合の定義

アドレス a_1 の**宛先アドレス集合** $S(a_1)$ は，a_1 から期間内に 1 度でも送金を行ったアドレスの集合である．宛先アドレス集合は先行研究で永田らが定義した送金先アドレス集合と同一である．アドレス a_1 の**入力アドレス集合** $I(a_1)$ は，a_1 から送金を行った際に同時に *Input* フィールドに指定されたアドレスの集合である．入力アドレス集合は Meiklejohn らが定義した入力アドレスの集合と同一である．アドレス a_1 の**送金元アドレス集合** $R(a_1)$ は，a_1 に対して期間内に 1 度でも送金を行ったアドレスの集合とする．アドレス a_1 の**出力アドレス集合** $O(a_1)$ は，a_1 に対し

[*1] Zcash: Privacy-protecting digital currency (https://z.cash/ja/)

表 1 a_1 についての 4 つのアドレス集合
Table 1 Definitions of four sets of addresses for a_1.

アドレス集合	定義	図 2 中での例	識別方式
宛先アドレス集合 $S(a_1)$	a_1 から送金するアドレスの集合	$\{a_3, b_1, b_2\}$	永田ら [8]
送金元アドレス集合 $R(a_1)$	a_1 に送金を行うアドレスの集合	$\{b_3, b_4, a_2\}$	本研究
入力アドレス集合 $I(a_1)$	a_1 の送金時に同時に *Input* に指定されるアドレスの集合	$\{a_2\}$	Meiklejohn ら [6]
出力アドレス集合 $O(a_1)$	a_1 の受け取り時に同時に *Output* に指定されるアドレスの集合	$\{a_4, b_1, b_2, b_3\}$	本研究

て送金が行われた際に，取引の *Output* フィールドに指定されたアドレスの集合とする．

これらのアドレス集合を**表 1** に整理する．4 つのアドレス集合には識別を行うアドレス a_1 を含めないことに注意されたい．

3.2 Jaccard 係数を用いた識別

アドレス識別の評価には Jaccard 係数を用いた集合の類似度を利用する．Jaccard 係数とは，ある集合 A と集合 B について，$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ で定められる類似度である．

Jaccard 係数を用いたアドレス識別手法を宛先アドレス集合 S を利用して，説明する．**図 3** はアドレス集合 S（特徴量）と取引の例を示す．

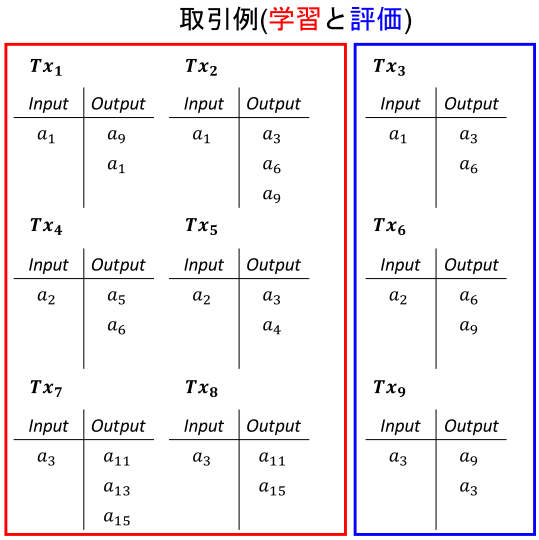
識別対象のアドレス a_1，a_2，a_3 から送金された取引 $Tx_1, Tx_2, \ldots, Tx_9$ について，3 つの期間における宛先アドレス集合を S_1，S_2，S_3 とする．学習アドレスを $L = S_1 \cup S_2$ と評価アドレスを S_3 とする．

ここである未知の宛先集合 $S_3(a_i) = \{a_3, a_6\}$ が，a_1，a_2，a_3 のどのアドレスから送信されたかを識別したい．そこで，$S_3(a_i)$ と 3 つの学習データ La_1，La_2，La_3 の Jaccard 係数を求めると，以下のようになる．

$$J(La_1, S_3(a_i)) = \frac{|\{a_3, a_6\}|}{|\{a_3, a_6, a_9\}|} = 0.67$$
$$> J(La_2, S_3(a_i)) = \frac{|\{a_3, a_6\}|}{|\{a_3, a_4, a_5, a_6\}|} = 0.50$$
$$> J(La_3, S_3(a_i)) = \frac{\phi}{|\{a_3, a_6, a_{11}, a_{13}, a_{15}\}|} = 0$$

他のアドレスと比較し $J(La_1, S_3(a_i))$ が Jaccard 係数の値が最も高いため，$S_3(a_i)$ は $a_i = a_1$ から送信されたと推測する．アドレス a_1 の評価アドレスは $S_3(a_1) = \{a_3, a_6\}$ であることから，この例ではアドレス識別に成功している．本評価では，Jaccard 係数が最も高い値を持つアドレスが複数存在する場合は，アドレス識別が失敗したと見なす．

アドレスの集合のような離散値に対する類似尺度として Dice 係数，Simpson 係数 [18] を利用する方法があげられる．これらの中で，Jaccard を用いる利点として，Jaccard 係数が持つロバスト性がある．Jaccard 係数を利用した類似度指標は外乱に強く，比較する特徴量の大きさが一様でなくても適切に類似度を与える．Bitcoin のトランザクション には，1 度の取引で 100 個以上の宛先や受け取りアドレスが利用されることがある．したがって，識別対象のアドレスは取引回数や 1 回あたりの取引アドレス数に応じて，特徴量の分散が大きいため，類似度を比較する際に Jaccard 係数を用いることが有効である．



アドレス	宛先アドレス集合 S		
	S_1	S_2	S_3
a_1	$\{a_9\}$	$\{a_3, a_6, a_9\}$	$\{a_3, a_6\}$
a_2	$\{a_5, a_6\}$	$\{a_3, a_4\}$	$\{a_6, a_9\}$
a_3	$\{a_{11}, a_{13}, a_{15}\}$	$\{a_{11}, a_{15}\}$	$\{a_9\}$

アドレス	学習アドレス集合 $L = S_1 \cup S_2$	評価アドレス集合 S_3
a_1	$\{a_3, a_6, a_9\}$	$\{a_3, a_6\}$
a_2	$\{a_3, a_4, a_5, a_6\}$	$\{a_6, a_9\}$
a_3	$\{a_{11}, a_{13}, a_{15}\}$	$\{a_9\}$

図 3 宛先アドレス集合 S と Jaccard 係数を用いた識別手法
Fig. 3 Identification of a set of destination addresses using Jaccard coefficient.

類似尺度の違いによる識別率への影響を確認するために，Dice 係数と Simpson 係数を利用して図 3 に示した取引のアドレス識別を行った場合，**表 2** に示す結果が得られる．Simpson 係数の結果からは a_1，a_2 のどちらが正解アドレスか予測することができない．Dice 係数の結果は a_1 と a_2 の差が 0.13 となり，Jaccard 係数の 0.17 よりも小さく識別力が低い．いずれの手法と比べても Jaccard 係数を用い

表 2 3 種類の類似尺度を用いた識別例
Table 2 Example of discrimination using three types of similarity measures.

2 つの集合	Jaccard	Dice	Simpson
La_1，$S_3(a_i)$	0.67	0.80	1.00
La_2，$S_3(a_i)$	0.50	0.67	1.00
La_3，$S_3(a_i)$	0	0	0

入力: $a_1, \ldots, a_m$ の取引 Tx の集合
未知アドレス x の送金元アドレス集合 $R(x)$
出力: x の推定アドレス $a_x \in \{a_1, \ldots, a_m\}$
1: $a_1, \ldots, a_m$ の送金元アドレス集合 $R(a_1), \ldots, R(a_m)$ を求める．
2: 未知のアドレス x について，類似度最大のアドレス a_{x^*} を求める．
すなわち，$a_{x^*} = \underset{a \in \{a_1, \ldots, a_m\}}{\arg\max} J(R(x), R(a))$
3: a_{x^*} を出力する．

図 4 提案手法 1 送金元アドレス集合 $R(a)$
Fig. 4 Method 1. A set of recipient addresses $R(a)$.

る手法がアドレスを識別する目的に適していると考える．

3.3 提案方式 1：取引の送金元アドレスを用いた識別

本手法では，識別対象のアドレスが Bitcoin を受け取る際の取引に着目する．これは自身のアドレスに対して送金を行うアドレスは指定することができない，という仮説に基づいている．

本手法では，対象アドレス a に向けて送金をした取引があるアドレス，すなわち，送金元アドレス集合 $R(a)$ を用いて Jaccard 係数に基づいて識別する．提案方式を**図 4** に示す．

永田ら [8] による宛先アドレスを特徴量としたアドレス識別では，送金時に指定するアドレスをユーザが任意に変更可能である．それゆえ，1 度送金したアドレスに対して，再度送金を行う際に自身が管理する別の撹乱用のアドレスを推定させることでアドレスの識別を回避することができる．一方，自身のアドレスが受け取り（*Output* フィールド）に指定される取引では，送金元のアドレスは取引相手が決めることになる．ユーザの制御ができない取引相手が管理するアドレスを特徴量とすることで，精度識別が高くなると考えた．

3.4 提案方式 2：取引の出力アドレスを用いた識別

本手法では，識別対象のアドレスに対して送金が行われる際に，同時に Bitcoin を受け取るアドレスに着目した．提案方式 1 における $R(a)$ の代わりに，出力アドレス $O(a)$ を用いる方式を提案方式 2 とする．提案方式を**図 5** に示す．

提案手法 2 では，受け取りを行う取引の *Output* フィールドのアドレスを特徴量とする．Bitcoin の取引が 1 対 1 のユーザ間で行われる場合，*Output* フィールドに指定された複数のアドレスはすべて送金相手のアドレスと考えることができる．したがって，*Output* フィールドに共起するアドレスを特徴量として用いた識別が有効であると考えた．

入力: $a_1, \ldots, a_m$ の取引 Tx の集合
未知アドレス x の出力アドレス集合 $O(x)$
出力: x の推定アドレス $a_x \in \{a_1, \ldots, a_m\}$
1: $a_1, \ldots, a_m$ の出力アドレス集合 $O(a_1), \ldots, O(a_m)$ を求める．
2: 未知のアドレス x について，類似度最大のアドレス a_{x^*} を求める．
すなわち，$a_{x^*} = \underset{a \in \{a_1, \ldots, a_m\}}{\arg\max} J(O(x), O(a))$
3: a_{x^*} を出力する．

図 5 提案手法 2 出力アドレス集合 $O(a)$
Fig. 5 Method 2. A set of output addresses $O(a)$.

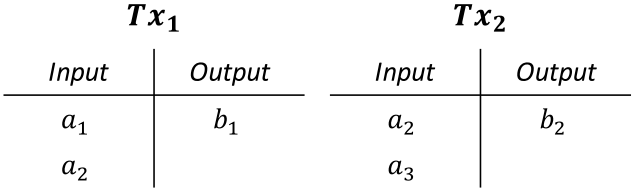


図 6 Meiklejohn らの手法 [6] と入力アドレス集合 $I(a)$ による識別の違い
Fig. 6 Differences in identification by Meiklejohn et al. [6] and $I(a)$.

3.5 Meiklejohn らの手法 [6] を応用した識別

本稿では，Meiklejohn らの手法 [6] を応用し，入力アドレスを特徴量とした Jaccard 係数に基づくアドレス識別を実施する．Meiklejohn らは，入力アドレスが複数ある取引に着目し，ブロックチェーンの署名の仕組みを利用して所有者が同じアドレスのグループ化を行っている．アドレス集合 I を用いたアドレス識別とは，Meiklejohn らが着目した入力アドレスを特徴量とした手法である．本手法では，3.2 節で示した $S(a)$ の代わりに，入力アドレス集合 $I(a)$ を Jaccard 係数の引数に用いて識別率を求める．

次に，これらの手法の違いを**図 6** に示した取引の例を利用し，説明する．Meiklejohn らの手法では，Tx_1 の *Input* フィールドに指定された a_1，a_2 が同じ管理者であると推定する．同様に，Tx_2 の a_2，a_3 が同じ管理者であると推定する．また，Tx_1，Tx_2 ではそれぞれ a_2 が共通して利用されていることから，a_1，a_2，a_3 は同じ管理者であると推定する．このように，取引の *Input* フィールドに指定されたアドレス間の関係を収束するまで推移的に拡大し，管理者の推定を行う．

一方，$I(a)$ を特徴量とした方式では，すべての異なる 2 アドレス間で Jaccard 係数を算出するため，推移性が必ずしも成り立つわけではない．たとえば，アドレス a_1，a_2，a_3 の入力アドレス集合 $I(a_1) = \{a_2\}$，$I(a_2) = \{a_1, a_3\}$，$I(a_3) = \{a_2\}$ を用いた識別手法は以下のようになる．

$$J(I(a_1), I(a_2)) = 0$$
$$J(I(a_1), I(a_3)) = 1$$
$$J(I(a_2), I(a_3)) = 0$$

この例では，アドレスの管理者が $I(a_1)$，$I(a_3)$ と $I(a_2)$ の 2 人であると推定される．したがって，Meiklejohn らの手法と集合 $I(a)$ による識別は厳密には異なる手法であることに注意が必要である．

4. 実験

4.1 実験目的

2 つの提案方式の識別精度を明らかにすることを目的とする．精度は次の条件に大きく依存すると考えられる．

(1) アドレスあたりの取引回数 n

(2) Bitcoin の利用目的

ここで，n は識別対象のアドレスが送金，受け取りを行った取引回数とする．たとえば，(1) については，アドレス a_1 が 21 回から 30 回の送金，受け取りをともなう取引を行ったとき，アドレス a_1 の取引回数は $n = 30$ となる．(2) については，交換所やマイニングプール業者のように同じアドレスで繰り返し送金，受け取りを必要とする場合と投資目的のエンドユーザとでは，取引の振舞いが大きく変わることを想定している．

そこで，これらの条件を変化させて，先行研究の 2 方式（入力アドレス集合 I [6] と宛先アドレス集合 S [8]）と提案方式を比較するために，次の実験を行う．

実験 1　取引回数による識別精度（(1) の評価）

実験 2　利用目的による識別精度（(2) の評価）

(1) アドレスあたりの取引回数と (2) Bitcoin の利用目的がアドレス識別精度に大きく依存すると考えられる理由にアドレス集合に含まれる情報量の差がある．(1) では，取引回数が増加するとアドレス集合に含まれる情報も多くなり，精度が上がると考えられる．(2) では，投資目的のため頻繁に取引が行われる交換所と自身のアドレスへ寄付を受け付ける目的でアドレスを公開している Bitcointalk では取引の特徴が著しく異なるため，アドレス集合に含まれる情報に差が生じて精度が向上すると考えられる．

4.2 アドレスの利用目的

(1) Bitcointalk [*2] は暗号資産に関する情報を交換する掲示板サイトである．Bitcointalk では，アカウントを登録しているユーザがプロフィールページに自身のアドレスを公開していることがある．これは，ユーザが自身への寄付を受け取ることが目的と考えられる．

(2) Bitcoin ATM [*3] は Bitcoin を預貯金することができるオフラインのサービスである．ユーザは Bitcoin アドレスの公開鍵情報（QR コード）を ATM に入力し，入金したい金額を現金で入れることで Bitcoin ATM のアドレスからユーザのアドレスへ送金が行われる．我々の研究報告 [16] において収集したカナダに設置された ATM のアドレス，ATM を利用しているユーザのアドレスを用いる．

(3) Darkweb は匿名通信路 Tor ネットワークである．Darkweb を利用する際には特殊なブラウザを用いることで送信元を匿名のままでアクセスする．Darkweb のウェブページ上で掲載しているプロモーション用のアドレスや違法商品（クレジットカード番号など）を取り扱うアドレスを収集する．

(4) Exchange（交換所）はユーザの所有する Bitcoin を現金と交換するサービスである．Bitcoin ATM とは異なり，ユーザは交換所へ登録を行うことでオンライン上で Bitcoin の売買が可能となる．主要な交換所のアドレスは WalletExploer [*4] で公開されている．

(5) Mining Pool は多数のマイナが協力し Bitcoin の取引情報をまとめたブロックに対して取引の検証を行い，報酬を得るための仕組みである．報酬を得るためには膨大な計算資源が必要とされており，個人がマイニング報酬を受け取ることは難しいと考えられている．マイニング報酬を受け取った Mining Pool が管理しているアドレスを対象とする．

4.3 データ収集

本研究で取得したアドレスと取引の数を**表 3** に示す．5 種類の利用目的に基づく Bitcoin アドレスの取引記録は Blockchain Explorer [*5] の API を用いて収集した．本実験では，収集したアドレスのうち 2 回以上取引を行っていたアドレスを利用し，アドレスの収集期間を 10 年間（観測期間 $D = 10$）と半年間（$D = 0.5$）について評価する．

観測期間 $D = 10$ は，Bitcoin の取引が開始された 2009 年からアドレスを収集した 2019 年までの全期間である．しかし，5 つの利用目的を $D = 10$ で比較できない．たとえば，Bitcoin ATM に登録されたアドレスは機器が製造された時期や設置された時期によって取引数や期間が異なる

表 3 収集したアドレスデータ
Table 3 Summary of address data.

利用目的	アドレス数	取引数	収集期間 D
Bitcointalk BBS	44,067	3,139,677	2009/1/4 – 2019/11/18
Bitcointalk BBS	1,968	28,832	
Bitcoin ATM	404	26,843	
Darkweb	82	35,048	2019/4/1–9/30
Exchange	680	33,252	
Mining Pool	96	24,449	

[*2] Bitcointalk (https://bitcointalk.org/)
[*3] Coin ATM Radar Bitcoin ATM Map (https://coinatmradar.com/)
[*4] WalletExplorer.com (https://www.walletexplorer.com/)
[*5] Blockchain Explorer (https://www.blockchain.com/ja/explorer)

表 4 長期間継続して利用された Bitcointalk アドレスと取引回数
Table 4 Number of addresses used for a long period in Bitcointalk.

取引回数 n	アドレス数	サンプリング数
10	12,493	100
20	4,948	100
30	2,535	100
40	1,408	100
50	842	100
60	499	100
70	335	100
80	211	100
90	153	100
100	117	100
合計	23,541	1,000

からである．そこで，すべての利用目的について識別に利用するアドレスの取引が行われていた共通の期間である，2019 年 4 月から 9 月までの半年間を対象としている．

4.4 実験 1. 取引回数に基づくアドレス識別

$D = 10$ となる長期間継続して利用されているアドレスの識別手法を以下に述べる．

(i) 表 3 に示した約 10 年間分の Bitcointalk アドレス 44,067 個を対象とする．
(ii) 44,067 個の Bitcointalk アドレスのうち取引回数が 2 回以上，100 回以下となる 23,541 個のアドレスを識別対象のアドレスとして使用する．取引回数とサンプリング数を**表 4** に示す．
(iii) 識別に使用するアドレスとして，取引回数ごとに 100 個のアドレスを 100 回層別サンプリングする．
(iv) サンプリングされたアドレスについて，取引回数の比率を保って取引データを 7 対 3 に分割し，7 割の特徴量を学習アドレス集合に，3 割に含まれる特徴量を評価アドレス集合とする[*6]．
(v) 提案手法と従来手法を用いてアドレスを識別する．

4.5 実験 2. 利用目的に基づくアドレス識別

$D = 0.5$ となる 5 種類の利用目的に対するアドレスの識別手法を以下に述べる．

(i) 表 3 に示した半年間（$D = 0.5$）の 5 種類の利用目的（Bitcointalk, Bitcoin ATM, Darkweb, Exchange, Mining Pool）に分類された計 3,230 個のアドレスを対象とする．
(ii) 3,230 個のアドレスのうち取引回数が 2 回以上となる 1,358 個のアドレスを識別対象のアドレスとして使用する．取引回数を**表 5** に示す．

[*6] たとえば，取引回数が 60 回（$n = 60$）のアドレスでは，学習アドレス集合と評価アドレス集合を構成する取引数の割合は 42 : 18 となる．

表 5 5 種類の利用目的の識別アドレス数
Table 5 Number of addresses for five Bitcoin usages.

利用目的	アドレス数	サンプリング数
Bitcointalk BBS	844	30
Bitcoin ATM	106	30
Dark web	49	30
Exchange	274	30
Mining Pool	85	30
合計	1,358	150

(iii) 識別に使用するアドレスは表 5 の 5 種類の利用目的ごとに 30 個のアドレスを 100 回層別サンプリングする．
(iv) サンプリングされたアドレスについて，取引回数の比率を保って取引データを 7 対 3 に分割し，取引データに含まれる 7 割の特徴量を学習アドレス集合に，3 割に含まれる特徴量を評価アドレス集合にする．
(v) 提案手法と従来手法を用いてアドレスを識別する．

本実験では，識別対象のアドレスを利用目的ごとに独立して識別率を評価する．たとえば，Exchange アドレスの評価では，Exchange として収集されたアドレス 274 個から 30 個のアドレスをサンプリングし，アドレスの識別率を評価する．Bitcointalk と ATM など，異なる利用目的のアドレスを混在して識別率を評価していないことに注意が必要である．

4.6 実験結果 1.

識別に成功した平均アドレス個数の結果を**表 6** と**図 7** に示す．表 6 では 100 回の試行のうち，識別に成功したアドレス数の平均値を示す．図 7 では 4 つのアドレス集合の識別精度について，95%信頼区間を表すエラーバーとともに示す．

4 つのアドレス集合のうち，出力アドレス集合 O の 547 個が最も識別に成功したアドレス数が多い．この集合 O の Jaccard 係数の分布を**図 8** に示す．$n = 10$ のとき，Jaccard 係数は 0.15 から 0.20 の値に分布しており，分散が大きい．この分散は取引回数の増加にともない，小さくなっている．

取引回数 $n = 40$ のアドレスが最も識別率が低く，$n = 100$ のアドレスが最も高い．3 つのアドレス集合 R, I, O では n が 40 回を超えるとアドレス識別率が増加し，出力アドレス集合 O では最大で 2.1 倍まで識別率が増加している．これは，永田らの報告による取引回数とアドレスの識別率には相関がない結果と異なっている．

そこで，より精査して見るために，表 6 に，4 つのアドレス集合の取引数と識別に成功したアドレス数の標準偏差を示す．集合 S は取引数に依存することなく識別率が推移しているように見えたが，識別したアドレス数の標準偏差は 2.2 で安定していた．集合 R, I, O の標準偏差の値がいずれも 7 以上の値であり，集合 S と比較して 3 倍以上の変動があった．

表 6 取引回数と平均アドレス識別個数
Table 6 Successful mean identified addresses for number of transactions.

アドレス集合	取引回数 n										合計	平均	標準偏差
	10	20	30	40	50	60	70	80	90	100			
S	39.2	45.5	43.0	44.6	42.2	38.6	45.4	42.4	43.7	41.8	426	42.6	2.2
R	54.6	43.4	36.8	35.7	39.6	41.9	48.1	55.0	59.2	62.3	477	47.7	9.1
I	49.8	48.6	46.7	48.7	49.7	54.4	52.5	60.9	64.1	69.1	545	54.5	7.2
O	60.4	45.3	37.1	39.5	43.9	49.7	54.7	66.8	71.2	78.6	547	54.7	13.4
平均	51.0	45.7	40.9	42.1	43.9	46.1	50.2	56.3	59.5	62.9	-	-	-

表 7 5 種類の利用目的の平均アドレス識別個数
Table 7 Successful mean of identified addresses for five Bitcoin usages.

	アドレス集合	利用目的					合計	平均
		BBS	ATM	Darkweb	Exchange	Mining pool		
文献 [8]	S	12.8	16.6	23.9	4.1	17.8	75	15.0
提案方式 1	R	17.2	3.6	22.2	14.7	5.0	63	12.5
文献 [6]	I	17.6	16.5	22.3	12.6	15.0	84	16.8
提案方式 2	O	19.5	4.3	20.9	22.6	5.0	72	14.5
	平均	16.8	10.2	22.3	13.5	10.7	-	-

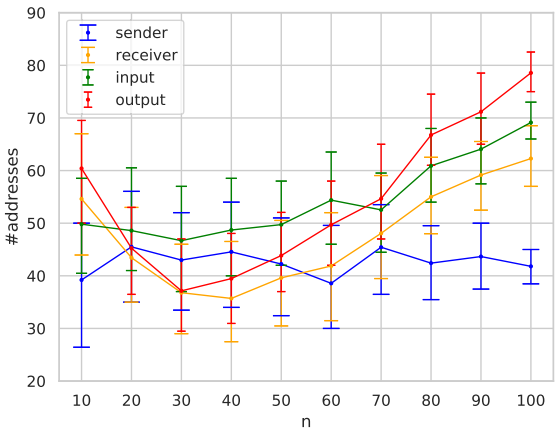


図 7 取引回数 n についての平均アドレス識別個数
Fig. 7 Successful mean of identification addresses with respect to numbers of transactions n.

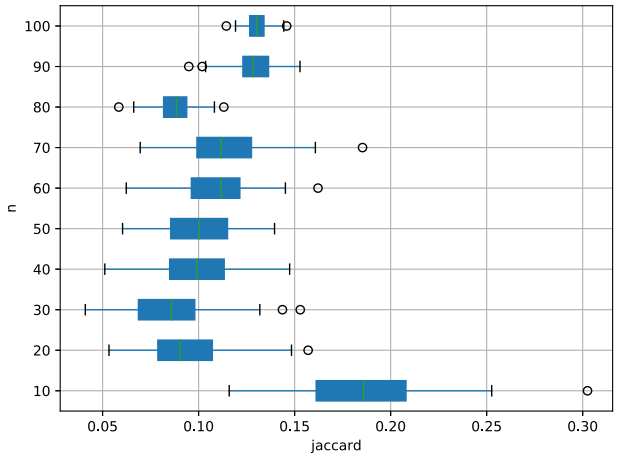


図 8 出力アドレス集合 O における取引回数についての Jaccard 係数の分布
Fig. 8 Distributions of Jaccard coefficient of a set of output addresses O for the number of transactions.

表 8 公開されているアドレス（ATM と Exchange）を除いた 3 種類の利用目的の平均アドレス識別個数
Table 8 Mean Successful identified addresses for three Bitcoin usages excluded known published addresses (ATM and Exchange).

	集合	3 種類		5 種類（表 7）	相対
		合計	平均	平均	誤差
文献 [8]	S	55	18.2	15.0	3.2
提案方式 1	R	44	14.8	12.5	2.3
文献 [6]	I	55	18.3	16.8	1.5
提案方式 2	O	45	15.1	14.5	0.6
	平均	-	16.6	14.7	1.9

4.7 実験結果 2.

5 種類の利用目的に基づくアドレス識別結果を**表 7** に示す．識別に成功したアドレス数が最も多い利用目的は Darkweb であり，4 つのアドレス集合を用いた識別結果の平均個数は 22.3 個であった．また，4 つのアドレス集合のうち入力アドレス集合 I は最も識別精度が高い．

次に，ATM や交換所などの通常推定する必要のない既知公開アドレスを取り除いた識別結果を**表 8** に示す．5 種類の利用目的で評価を行った表 7 と比較して，4 つの集合すべてにおいてアドレス識別数の平均値が増加した．

5. 評価と考察

5.1 $n = 30$ 付近での識別率の低下について

図 7 より，3 つのアドレス集合 R，I，O は $n = 30$ 付近において識別率のピークがあり，$n > 30$ では再び増加している．この非単調な識別率の原因として，次が考えられる．

(1) $n < 30$ のアドレスに，アドレスが毎回更新される新しいウォレットで使われる割合が多いため．

(2) $n < 30$ となるアドレスに，特定の利用目的のものが偏っているため．

そこで，(1) を調査するために，各取引数 n におけるアドレスの開始年度を調べた．仮説が正しければ，$n < 30$ における年度に偏りがみえるはずである．

図 9 に識別対象のアドレスに関する取引数と取引開始年の散布図を示す．残念ながら，開始年度における著しい偏りはなく，(1) の仮説が原因とは考えにくい．

そこで，(2) を調べるために，利用目的の中の交換所のアドレスに注目する．交換所のアドレスは取引パターンも固有で識別率も高いために，全体の識別率を支配していると考えたためである．4 つのアドレス集合で識別に成功したアドレスのうち，交換所のアドレスが含まれている割合を**図 10** に示す．4 つのアドレス集合では $n = 30$ から

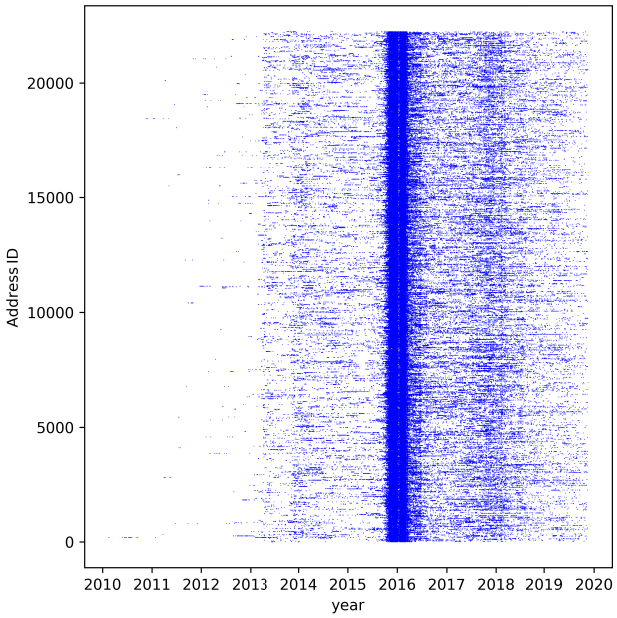


図 9 $D = 10$ 年の間におけるアドレス取引分布
Fig. 9 Distribution of transactions for $D = 10$ data.

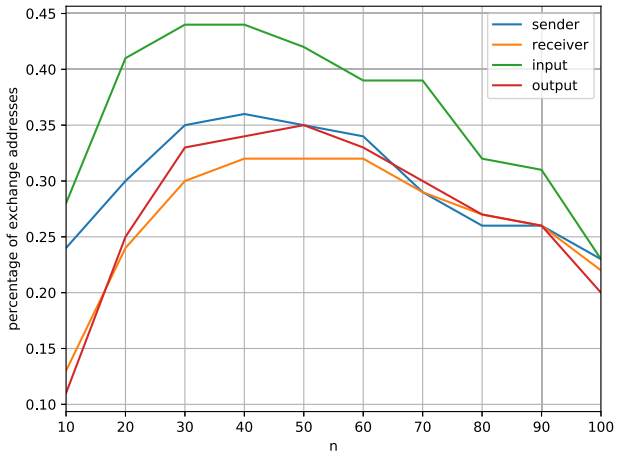


図 10 識別成功アドレス中の交換所アドレスの割合
Fig. 10 The fraction of exchange addresses in the identified addresses.

$n = 50$ 付近で交換所アドレスの割合が最も高い値を示している．図 7 と図 10 を比較すると $n = 30$ 付近でピークとなること，$n = 10$ から $n = 100$ までの割合が類似していることから，交換所のアドレスが識別率に影響を与えていると結論づける．

5.2 Darkweb アドレスが最も識別率が高い原因について

表 7 の結果より，最も識別率が高い利用目的は Darkweb で利用されたアドレスであった．この要因の 1 つに Darkweb の運用形態が原因であると考えられる．Darkweb では違法商品の売買など法的に問題のあるサービスで利用されているので，頻繁にサイトの公開と閉鎖が繰り返されており，取引に利用された Bitcoin アドレスの寿命も短い．収集した Darkweb のアドレスは Tor ネットワーク上のサイト上に掲載されている．違法サービスを取り扱い，アドレスが閲覧可能なページ上に掲載されていたことから，アドレスの追跡リスクを考慮せずに短期間に取引を行い，使い捨てていたと考える．また，表 3 に示した 5 種類の利用目的の中でも，Darkweb のアドレス数は最も少なく，取引数が最も多いため，アドレスあたりの特徴量が多い．したがって，これらが識別率に影響を与えていると考える．

5.3 提案手法の精度が高いことの仮説検定

提案方式 R，O のアドレス識別精度が従来手法 S，I より高いことを確かめるために平均値の t 検定 [17] を行う．検定の結果を**表 9** に示す．検定には $n = 100$ における 4 つの方式の精度を比較している．$n = 100$ の識別率を検定したのは提案手法 R，O と従来手法 I において最も識別率が高い値であったためである．

$n = 100$ 以外については，$n = 10$ から 100 までの識別結果に対して平均値の t 検定の p 値を**表 10** に示す．提案方式 1（送金元アドレス R）は先行研究 S に対して $n = 60$ 以上のすべてについて高精度で識別し，その差は十分に大きく，統計的に有意であった．提案方式 2（出力アドレス O）は先行研究 S と先行研究 I の両方に対して $n = 70$ 以上で統計的に有意な差で優れていた．また，提案手法 1，2 は，n が大きい場合だけでなく $n = 10$ の場合にも高精度で，統計的に有意であった．

表 9 4 種類のアドレス集合と平均値の検定結果（$n = 100$）
Table 9 t-test results for four address sets ($n = 100$).

		平均値の差	統計量 t	p 値	
提案方式 1	R, S	20.5	57.0	2.2×10^{-16}	***
	R, I	−6.8	-	-	
	R, O	−16.3	-	-	
提案方式 2	O, S	36.8	135.1	2.2×10^{-16}	***
	O, I	9.5	34.9	2.2×10^{-16}	***
	O, R	16.3	45.5	2.2×10^{-16}	***

***：$p < 0.05$

表 10 提案手法 R, O の平均値の検定結果（$n = 10$ から $n = 100$）
Table 10 t-test results for a sets of R and O (between $n = 10$ to $n = 100$).

n	p 値					
	R, S	R, I	R, O	O, S	O, I	O, R
10	*	*	-	*	*	*
20	1.0	-	1.0	0.6	-	3.6×10^{-3}
30	-	-	0.7	-	-	0.3
40	-	-	-	-	-	*
50	-	-	-	5.5×10^{-3}	-	*
60	*	-	-	*	-	*
70	*	-	-	*	*	*
80	*	-	-	*	*	*
90	*	-	-	*	*	*
100	*	-	-	*	*	*

$*: p < 1.0 \times 10^{-3}$

以上により，提案方式 2（出力アドレス O）は先行研究 S, I のいずれに対しても識別精度が高く，その差は十分に大きく，2 つのアドレス集合の識別率に差はないという帰無仮説の p 値が 0.05 未満であり，統計的に有意であることが示された．

5.4 目的と手法の関係

表 7 より，Bitcointalk（BBS），Exchange については，提案方式のアドレス集合 O は高い識別率を示した．これらのアドレスは，4.3 節で示した web ページ上に公開されている．そのため，集合 O を用いたアドレス識別は識別率が高くなると考える．送金取引を特徴量とする集合 S や集合 I では，ユーザが送金に利用するアドレスを自身で選択しているという特徴がある．それゆえ，アドレスが公開されていることを考慮し，同じアドレスを再利用した取引を意図的に避けているために識別率が下がったと考えられる．

永田ら [8] によって提案されたアドレス集合 S は ATM，Darkweb，Mining pool で最も高い識別率を示した．これらのアドレスは，独立した web ページなどで公開されている．ユーザは取引相手に対してアドレスを再利用し，自身のアドレスが識別されるリスクを重要視していないと考えられる．

一般的に，ユーザがアドレスの情報を公開している場合，集合 O を用いた識別率は高くなり，アドレスの情報を公開していない場合，集合 S を用いた識別率は高くなると考えられる．

5.5 ATM と Mining pool の識別率が低い原因

表 7 より，ATM，Mining pool について，提案手法の集合 R，O は既存手法の集合 S，I と比較して低い識別率を示した．

ATM に関するアドレスには，**図 11** に示す 3 種類のアドレス a，b，c がある．本実験では，ATM 機器のアドレス b とユーザのアドレス c のみを収集している．ATM のアドレスは，事業者によって取引されている．そのため，ATM 取引の送金元アドレスやお釣りを受け取る出力アドレス（*Output* フィールド）には，特定の ATM 業者のアドレス a が指定される．したがって，a や b は提案方式で識別しやすい．一方，ATM を利用したユーザのアドレス c はユーザごとに異なり，多様な利用をされるので識別しにくい．したがって，アドレス c の特徴が ATM アドレス全体の識別率を低下させていると考えられる．

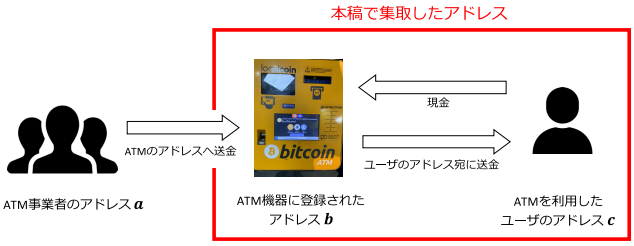


図 11 Bitcoin ATM を利用した取引例
Fig. 11 Examples of transactions of Bitcoin ATM.

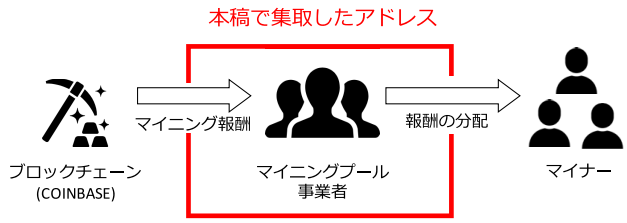


図 12 マイニング報酬の取引例
Fig. 12 Examples of transactions with mining rewards.

Mining pool のアドレスは，Bitcoin のマイニング報酬を受け取った Mining pool 事業者のアドレスである．**図 12** にマイニング報酬を受け取る取引例を示す．マイニング報酬を受け取る取引では，新たに生成されたブロックを示す固有の値 *COINBASE* が入力アドレスに指定される．そのため，報酬を受け取るために使用されている Mining pool 事業者間でアドレスの特徴量に差が出ない．また，異なる Mining pool 事業者でも報酬を受け取る取引では，送金元が *COINBASE* に統一されてしまうため識別が難しい．したがって，マイニング報酬を受け取る取引の特徴が Mining pool アドレスの識別率を低下させていると考えられる．

以上の取引の特性が，提案手法の集合 R と集合 O における ATM と Mining pool の識別率を低下させている要因であると結論づける．また，取引の特性はブロックチェーンや ATM サービスの仕組みに影響されるため，受け取り取引を特徴量とする集合 R や集合 O では識別率を高くすることが難しい．識別率を向上させる工夫としては送金取引を特徴量とする集合 S や集合 I を併用するなど，利用目的に応じて識別手法を選択することが有効である．

6. 結論

本稿ではBitcoinアドレスを識別する新たに2つのアドレス集合 R，O を特徴量とする方式を提案した．10年間のアドレス用いたアドレス識別実験により，アドレス集合 O を用いた提案手法の識別結果が最も精度が高いことを示した．これは，既存手法のアドレス集合 S，I と比較して，統計的に有意である．5種類の利用目的を用いたアドレス識別実験ではDarkwebで利用されているアドレスの識別率が最も高いことを示した．これらの精度の違いが，各応用における取引の特徴から生じることを考察した．

今後は本稿で使用した5種類の利用目的に含まれる属性情報の有無がBitcoinアドレスの識別にどれだけ影響を与えるかについて検討することを課題とする．

参考文献

[1] Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008).

[2] Dupont, J. and Squicciarini, A.C.: Toward De-Anonymizing Bitcoin by Mapping Users Location, *Proc. 5th ACM Conference on Data and Application Security and Privacy* (*CODASPY '15*), pp.139–141 (2015).

[3] 井垣秀星，永田倖大，菊池浩明：平均取引時間分布の相関を用いたBitcoinユーザのタイムゾーンの推定，情報処理学会第81回全国大会，pp.481–482 (2019).

[4] 山崎孝順，草野蘭之介，松本寛輝，井垣秀星，菊池浩明：取引件数の時間分布の相関を用いたBitcoin取引所のユーザの属性推定，情報処理学会第82回全国大会，pp.407–408 (2020).

[5] Bitcoin.org：プライバシーの保護，入手先 ⟨https://bitcoin.org/ja/protect-your-privacy⟩（参照 2020-6-26）.

[6] Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M. and Savage, S.: A fistful of bitcoins: Characterizing payments among men with no names, *Proc. 2013 Conference on Internet Measurement Conference* (*IMC '13*), pp.127–140, ACM (2013).

[7] Kappos, G., Yousaf, H., Maller, M. and Meiklejohn, S.: An Empirical Analysis of Anonymity in Zcash, *Proc. 27th USENIX Security Symposium* (*USENIX Security '18*), pp.463–477 (2018).

[8] 永田倖大，菊池浩明：Bitcoinアドレスの送金先集合に基づく匿名性の評価，情報処理学会第80回コンピュータセキュリティ研究発表会（CSEC-80），pp.1–6 (2018).

[9] Ron, D. and Shamir, A.: Quantitative Analysis of the Full Bitcoin Transaction Graph, *Financial Cryptography and Data Security* (*FC 2013*), Lecture Notes in Computer Science, Vol.7859, pp.6–24, Springer, Berlin, Heidelberg (2013).

[10] 廣澤龍典，上原哲太郎：ビットコインのミキシングにおける資金移動の分析，情報処理学会第81回コンピュータセキュリティ・第41回インターネットと運用技術合同研究発表会，pp.1–8 (2018).

[11] 坂間潤一郎，金岡　晃：ビットコインにおけるデジタル署名の乱数分析，情報処理学会第87回コンピュータセキュリティ研究発表会，pp.1–5 (2019).

[12] Harlev, M.A., Yin, H.S., Langenheldt, K.C., Mukkamala, R.R. and Vatrapu, R.: Breaking Bad: De-Anonymising Entity Types on the Bitcoin Blockchain Using Supervised Machine Learning, *Proc. 51st Hawaii International Conference on System Sciences* (*HICSS 2018*), pp.3497–3506 (2018).

[13] 松本寛輝，井垣秀星，菊池浩明：Bitcoinサービス業者と利用者アドレスの種類の推定と評価，情報処理学会第182回マルチメディア通信と分散処理・第88回コンピュータセキュリティ合同研究発表会（CSEC-88），pp.1–7 (2020).

[14] Garba, A., Guan, Z., Li, A. and Chen, Z.: Analysis of Man-In-The-Middle of Attack on Bitcoin Address, *Proc. 15th International Joint Conference on e-Business and Telecommunications* (*ICETE 2018*), pp.388–395 (2018).

[15] Huang, D.Y. et al.: Tracking Ransomware End-to-end, *2018 IEEE Symposium on Security and Privacy* (*SP*), pp.618–631, DOI: 10.1109/SP.2018.00047 (2018).

[16] 井垣秀星，松本寛輝，菊池浩明：カナダにおけるBitcoin ATMの利用者調査，情報処理学会第82回全国大会，pp.411–412 (2020).

[17] 平均値の検定, 入手先 ⟨http://www.aoni.waseda.jp/abek/document/t-test.html⟩（参照 2020-11-28）.

[18] 集合の類似度（Jaccard係数, Dice係数, Simpson係数），入手先 ⟨https://mieruca-ai.com/ai/jaccard_dice_simpson/⟩（参照 2021-03-27）.

松本 寛輝 （正会員）

2019年東京電機大学理工学部情報システムデザイン学系卒業．2021年明治大学大学院先端数理科学研究科博士前期課程修了．現在，IT企業に所属．

菊池 浩明 （正会員）

1988年明治大学工学部電子通信工学科卒業．1990年同大学大学院博士前期課程修了．1994年同博士（工学）．1990年（株）富士通研究所入社．1994年東海大学工学部電気工学科助手．1995年同専任講師．1999年同助教授，2006年同情報理工学部情報メディア学科教授．1997年カーネギーメロン大学計算機科学学部客員研究員．2013年明治大学総合数理学部先端メディアサイエンス学科専任教授．2016年同大学大学院先端数理科学研究科長．2018年一般社団法人JPCERTコーディネーションセンター（JPCERT/CC）代表理事．WIDEプロジェクト暗号メールシステムFJPEMの開発，認証実用化実験協議会（ICAT），IPA独創情報技術育成事業等に従事．暗号プロトコル，ネットワークセキュリティ，ファジィ論理，プライバシ保護データマイニング等に興味を持つ．1990年日本ファジィ学会奨励賞，1993年情報処理学会奨励賞，1996年SCIS論文賞，2010年度，2017年度情報処理学会JIP Outstanding Paper Award，2013年IEEE AINA Best Paper Award，2014年情報セキュリティ文化賞．電子情報通信学会，日本知能情報ファジィ学会，IEEE，ACM各会員．本会フェロー．