2020 / 10 / 27

CSS2020@オンライン
2B5-2 : ブロックチェーンプラットフォーム

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]　菊池 浩明[2]

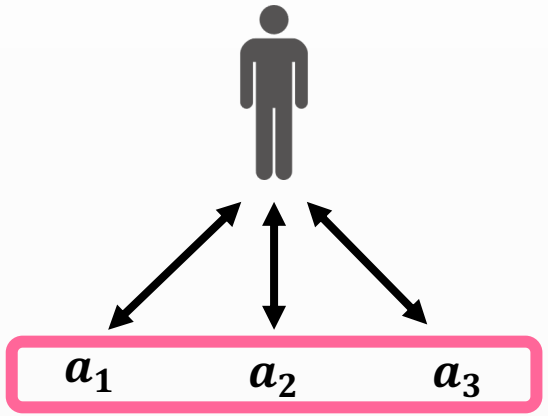
1. 明治大学大学院先端数理科学研究科
2. 明治大学総合数理学部

背景

- Bitcoinの匿名性
  - Bitcoinアドレス(ランダムな文字列)に基づく
  - アドレスとユーザが繋がる情報はない
  - ユーザを特定できない ＝ 匿名性が高い

Bitcoinアドレス

- Bitcoinアドレスの識別
  - 同一ユーザが管理しているアドレスの特定
    - » 例) アドレス a_1, a_2, a_3 は同一のユーザが管理している
      ※ ユーザの情報(氏名など)は特定できない



Bitcoinアドレスの識別

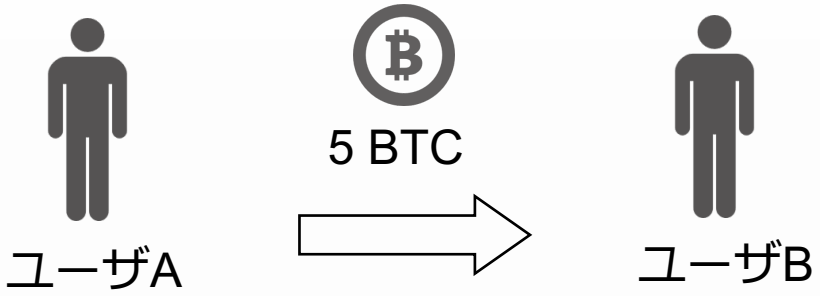
Bitcoinの取引構造

- 例) ユーザAがユーザBに5BTC送金

送金前のアドレス

- ユーザAのアドレス
  $a_1 = 3$ (BTC)
  $a_2 = 3$ (BTC)

- ユーザBのアドレス
  $b_1 = 0$ (BTC)

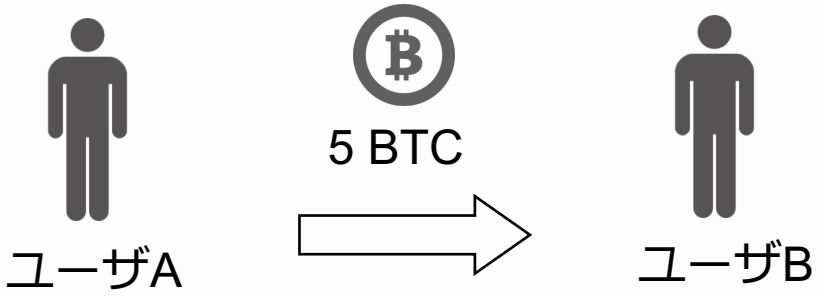


$\boldsymbol{Tx}$(取引)

Input	Output
$\boldsymbol{a_1}$ (3BTC)	$\boldsymbol{b_1}$
$\boldsymbol{a_2}$ (3BTC)	$\boldsymbol{a_1}$

Bitcoinの取引構造

- 例) ユーザAがユーザBに5BTC送金



送金前のアドレス

- ユーザAのアドレス
  $a_1 = 3$ (BTC)
  $a_2 = 3$ (BTC)
- ユーザBのアドレス
  $b_1 = 0$ (BTC)

$\boldsymbol{Tx}$(取引)

$\boldsymbol{a_1}$

$\boldsymbol{a_2}$

Input

$\boldsymbol{b_1}$ **(5BTC)**

$\boldsymbol{a_1}$ **(1BTC)**

Output

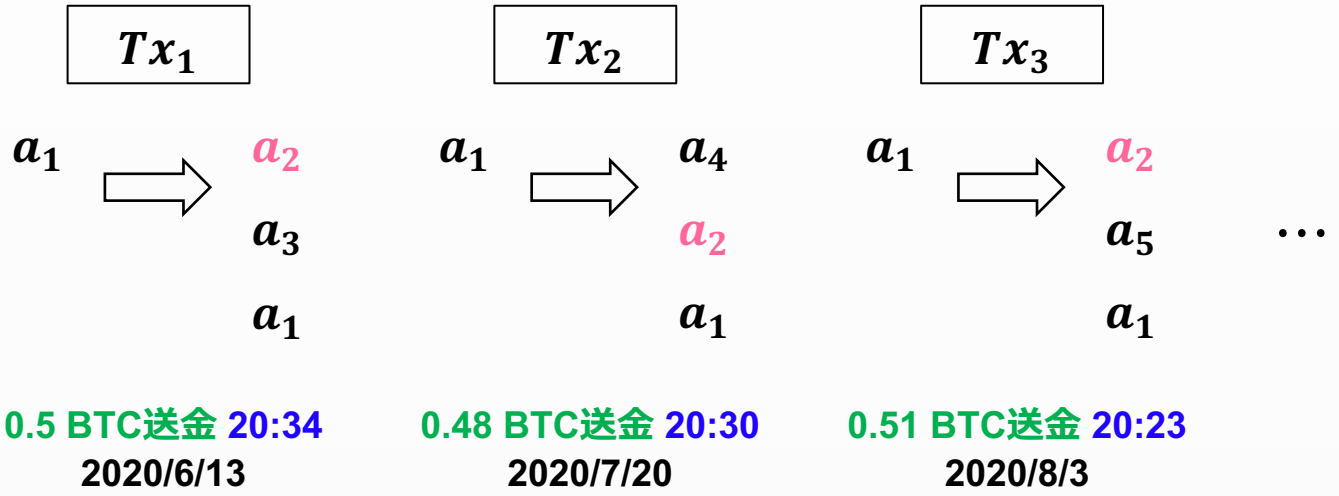
送金後のアドレス

- ユーザAのアドレス
  $\boldsymbol{a_1 = 1}$ **(BTC)**
  $a_2 = 0$ (BTC)
- ユーザBのアドレス
  $\boldsymbol{b_1 = 5}$ **(BTC)**

a_1は**取引のおつり**を
受け取っている

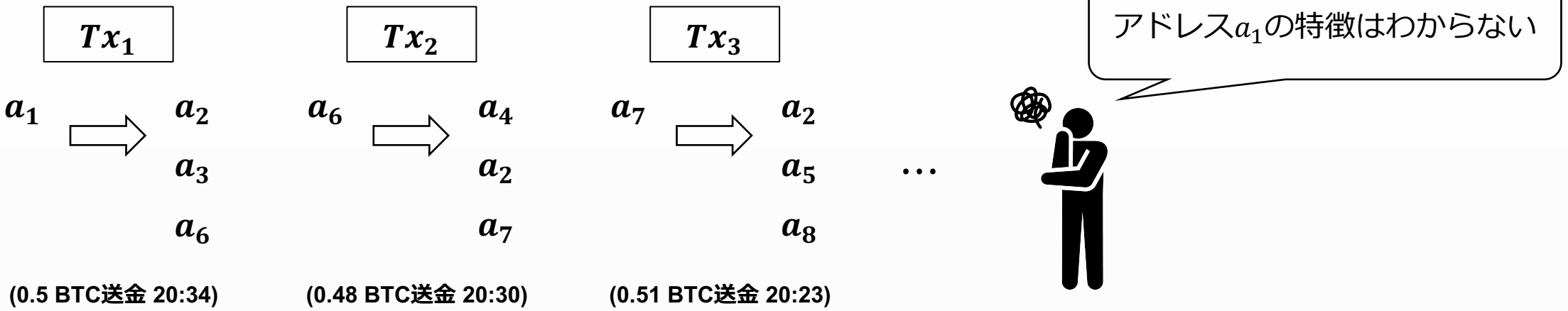
Bitcoinアドレスの識別リスク

- Bitcoinアドレスの識別リスク
  - 取引(Tx)の回数が増加 = アドレスの特徴量が学習される
  - 例) 金額, 時間帯, 送金を行う相手(宛先アドレス)



プライバシーの保護と課題

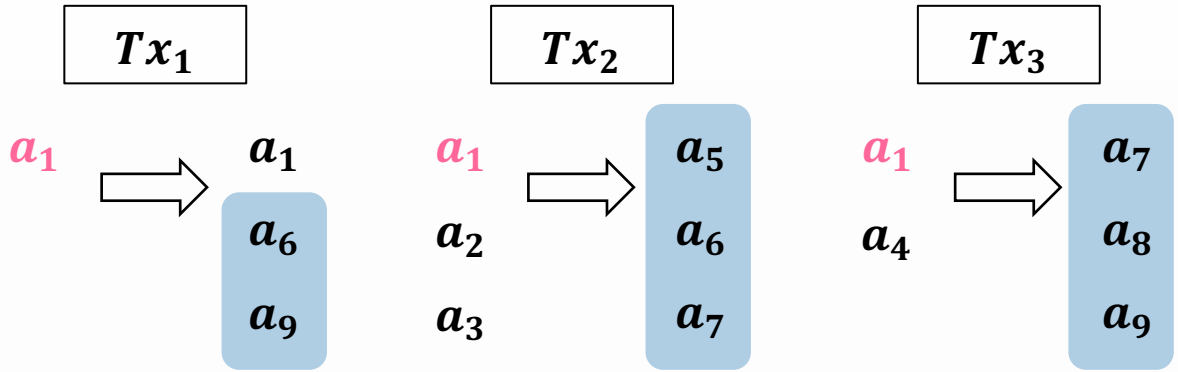
- Bitcoinアドレスとプライバシーの保護
  - 開発コミュニティ Bitcoin.org は「**一度利用したアドレスを再び利用しない**」ことを推奨
  - アドレスa_1は取引毎に新しいアドレス(a_6, a_7, a_8)を作成・利用する



- Bitcoinアドレスを長期期間・繰り返し利用する
  - SNSへの公開(寄付), 一部営利目的で利用されるアドレス, 等

先行研究 Bitcoinアドレスの識別

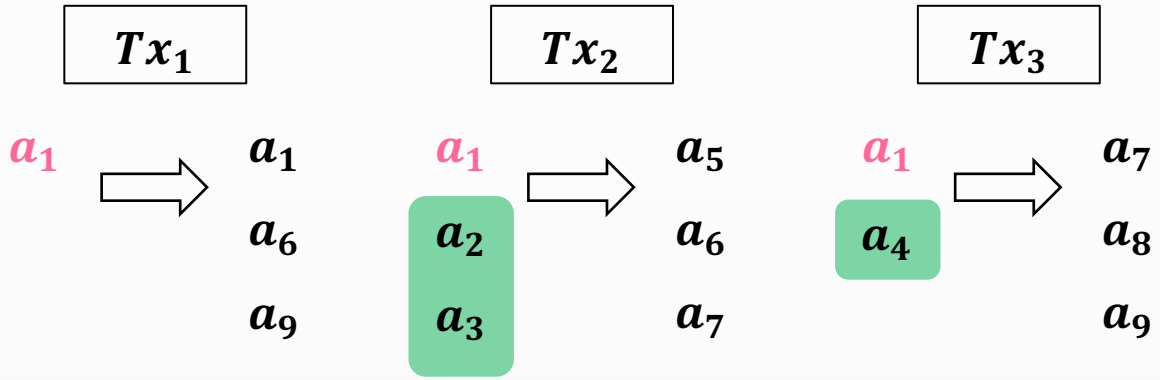
- **永田ら[2018]による「送金先アドレス」(宛先アドレス)を用いたアドレス識別**



アドレスa_1の
宛先アドレス集合S

$S(a_1) = \{a_5, a_6, a_7, a_8, a_9\}$

- ***Meiklejohn*ら[2013]による「入力アドレス」を用いたアドレス識別**



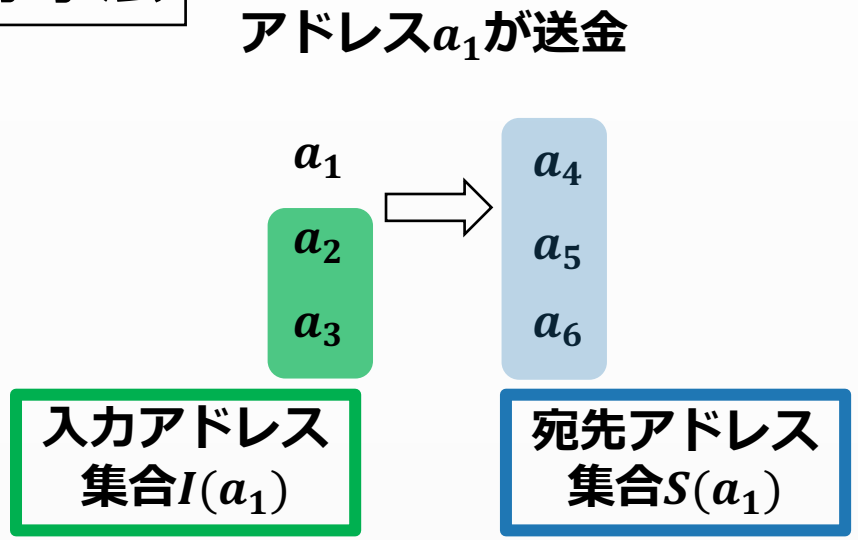
アドレスa_1の
入力アドレス集合I

$I(a_1) = \{a_2, a_3, a_4\}$

先行研究[永田ら, 2018]の課題

- 識別対象のアドレスが取引の送金側(*Input*フィールド)である場合のみ想定
  - 送金に使用するアドレスは**ユーザが自身で選択できる**ため識別率に影響を与える

既存手法

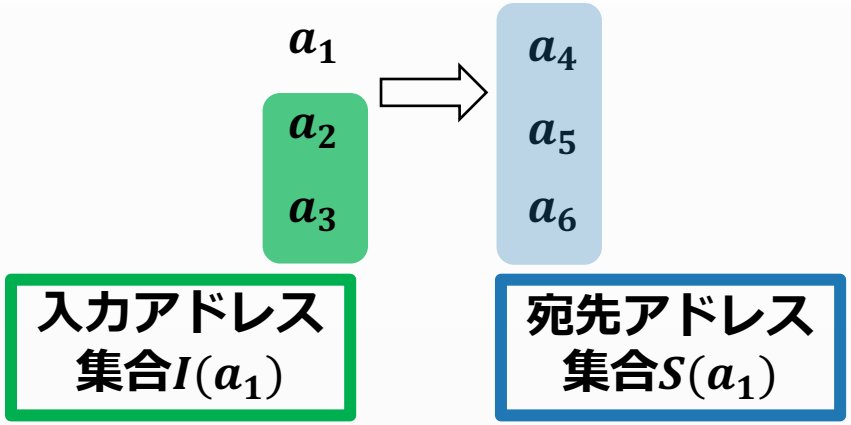


提案方式

- **「送金元アドレス」と「出力アドレス」を用いたアドレス識別を提案**
  - 識別対象のアドレスが取引の受け取り側(*Output*フィールド)にある
  - 自身のアドレスに対して送金を行うアドレスは指定できない

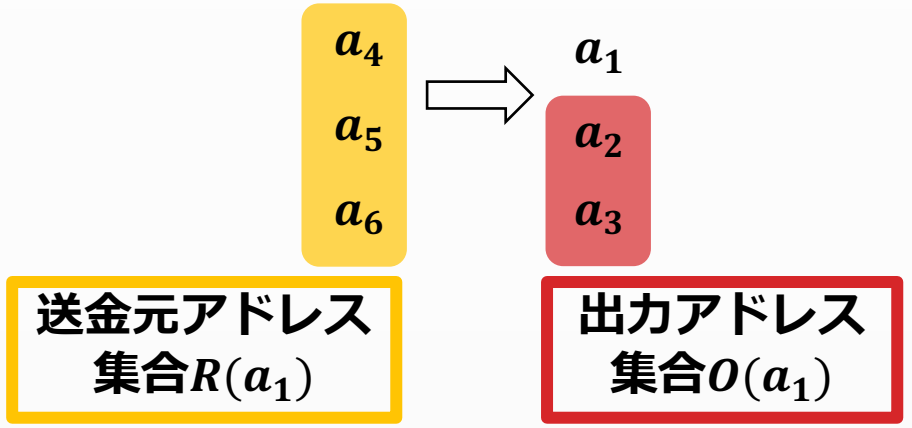
既存手法

アドレスa_1が送金



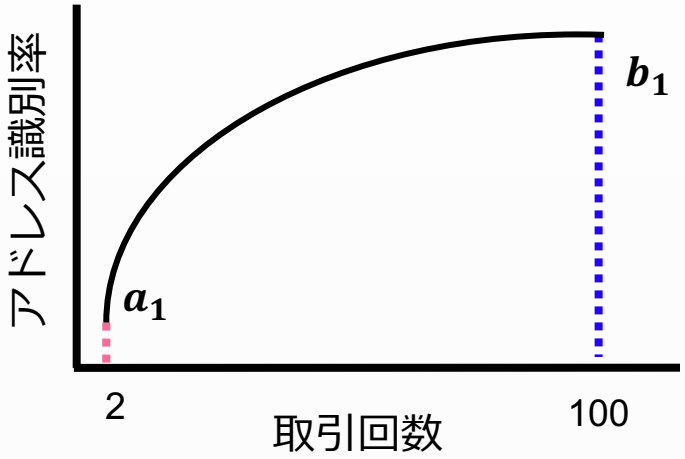
提案手法

アドレスa_1が受け取り



研究目的と研究課題

- 4つの識別方式のうち最も識別精度が高い手法はどれか？
  - **取引回数が与えるアドレス識別率への影響**を明らかにする
    - » 長期間, 繰り返し利用されたアドレスは識別されやすいのか？



2回取引したアドレスa_1
100回取引したアドレスb_1

アドレスの識別率
$a_1 < b_1$

  - **アドレス利用目的が与える識別率への影響**を明らかにする
    - » アドレスの利用目的(交換所, マイニング等)で識別率は変化するのか？

実験1 アドレス識別：取引回数に基づく識別

- アドレス数
  - 暗号資産の情報交換サイト*Bitcointalk*より収集
  - 2009年1月4日～2019年11月18日の約10年分の取引

- 識別アドレスの選択
  - 取引回数nを基準に層別サンプリング
    - » 100個のアドレスを抽出
  - 100個のアドレスのうち識別に成功した個数を求める

取引回数 n	アドレス数
2 - 10	12,493
11 - 20	4,948
21 - 30	2,535
31 - 40	1,408
41 - 50	842
51 - 60	499
61 - 70	335
71 - 80	211
81 - 90	153
91 - 100	117
合計	23,541

実験1 アドレス識別：実験手法

- アドレスの識別実験手法
  i. 各アドレス集合を7対3となるように分割する
  (学習アドレス集合を7, 評価アドレス集合を3と定義)

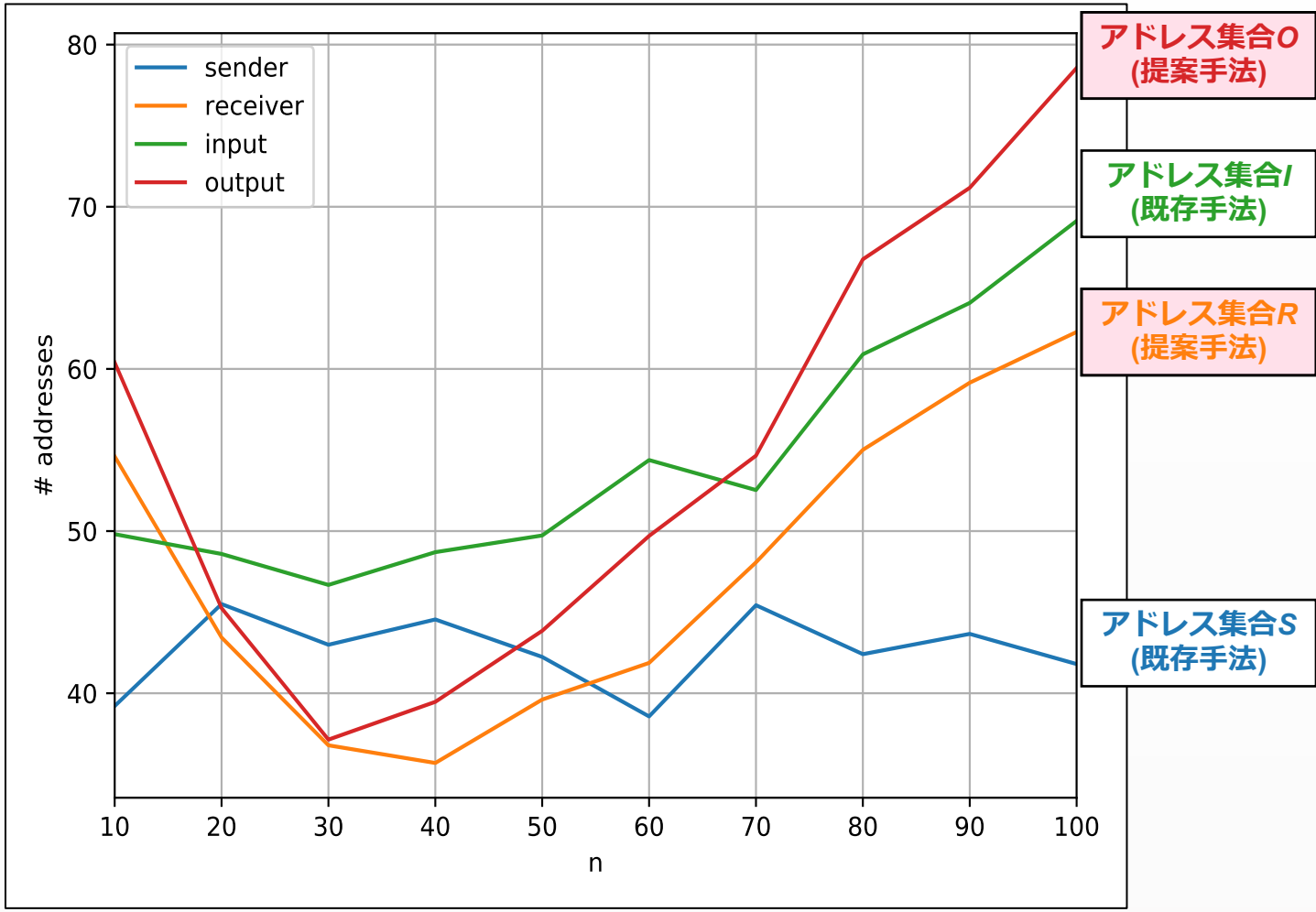
  ii. 100個のアドレス集合間の距離を*Jaccard*係数を用いて求める

  iii. *Jaccard*係数の値が最も高い集合のアドレス一致していた場合,
  「識別に成功した」と定義する

実験1 結果：取引回数とアドレス識別率の推移

- 取引回数nとアドレス識別率
  - $n = 40$　識別率が最も低い(35.7)
  - $n = 100$　識別率が最も高い(78.6)

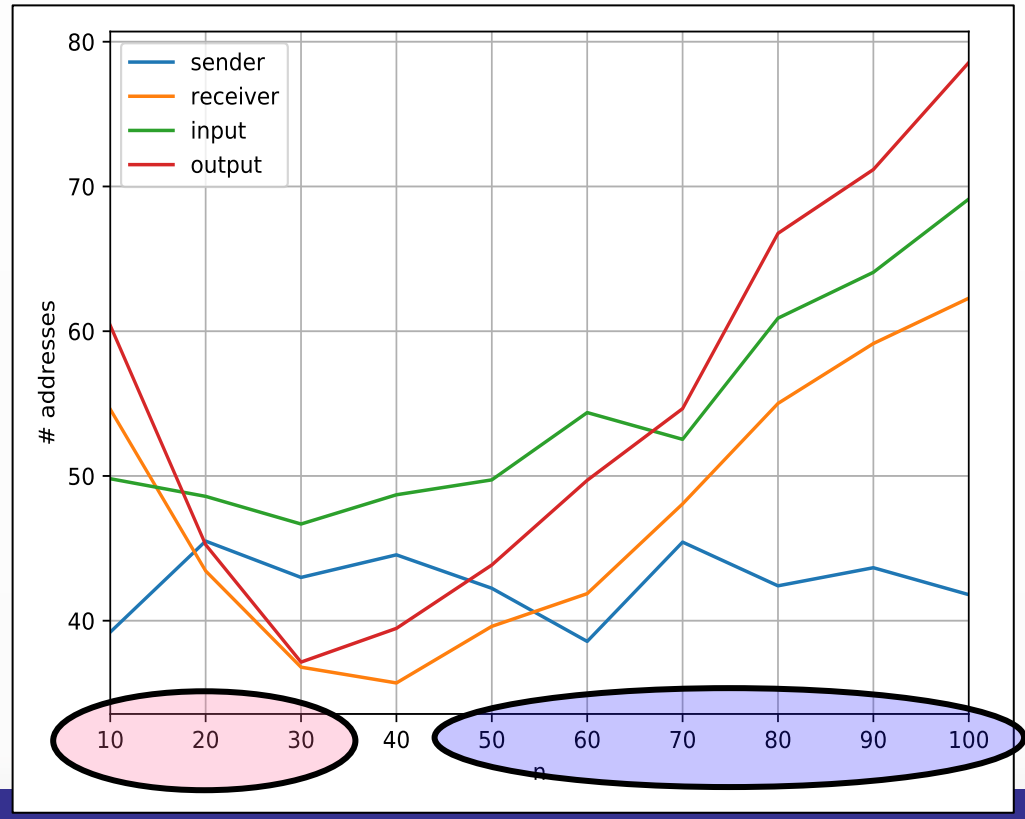
- アドレス識別率と識別手法
  - 出力アドレス集合O(提案手法)が最も識別率が高い(平均:54.7)
  - 宛先アドレス集合S(既存手法)が最も識別率が低い(平均:42.6)



実験1 結果：n=10～30のアドレス識別が低下

- 利用されているアドレスの属性が異なる
  - ウォレットアプリ：Bitcoinに興味がある人(ライトユーザ)
  - 交換所：投資家などBitcoin取引を頻繁に行う人(ヘビーユーザ)

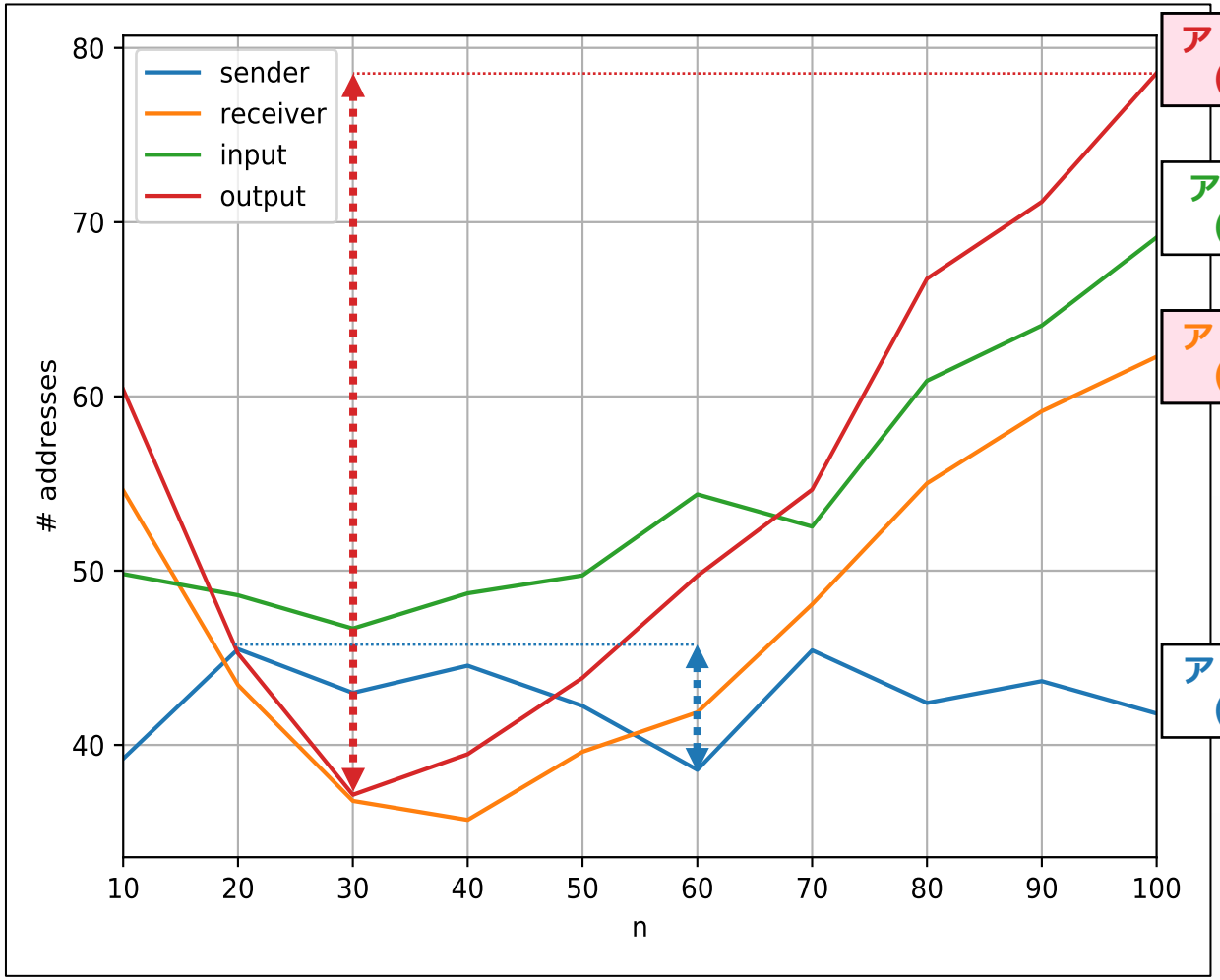
	ウォレットアプリ	交換所
アドレスの更新	○	×
長期間の利用	△	○
送金方法	自分で送金	複数のユーザの送金を まとめて実施
利用方法	スマホアプリ等で誰でも インストール可能	本人確認手続きなどを 行う必要あり
取引頻度	少ない	多い



実験1 結果：取引回数と識別率の相関関係

- 取引回数と識別率の相関関係
  - 10段階の取引回数のうち
    識別率が最も高い回数と低い回数の
    標準偏差を求めることで評価
  - 集合Oは相関関係が最も強い
  - 集合Sは相関関係が最も弱い

アドレス集合	標準偏差
S	2.2
R	9.1
I	7.2
O	13.4



実験2 アドレス識別：アドレスの利用目的

- 長期間のアドレス識別とアドレスの利用目的
  - 識別結果に偏り(Bitcointalkアドレス固有の特性)が生じる可能性
  - 代表的なアドレスの利用目的・サービス(交換所等)で使用されたアドレスに対してアドレス識別を行う

- 5つの代表的なアドレスの利用目的

利用目的	アドレス	考えられるユーザ層	アドレス識別率(予測)
BBS (Bitcointalk)	登録を行っているユーザのアドレス	暗号資産に関心が高いユーザ	△
ATM (BitcoinATM)	ATMに預貯金を行うユーザと ATMに登録されているアドレス	ATMが設置されている地域のユーザ	△
Darkweb	違法商品・サービスを 提供, 利用しているアドレス	身元を明かさず取引を行いたいユーザ	×
Exchange	交換所を利用しているユーザのアドレス	投資目的で利用しているユーザ	○
Mining Pool	マイニング報酬を受け取るアドレス	投資目的で利用しているユーザ	○

実験2 アドレス識別：5種類のアドレス識別

- アドレス数
  - 2019年4月1日～9月30日の半年間で取引をおこなったアドレスを使用
  - 利用方法別にサンプリング
    - » 25個のアドレスを抽出
- 実験手法
  - 前述の実験と同様の方法でアドレス識別実験を実施

利用方法	アドレス数
Bitcointalk BBS	844
Bitcoin ATM	106
Darkweb	49
Exchange	274
Mining Pool	85
合計	1,385

実験2 結果：5種類のアドレス利用方法と識別率

- 実験結果
  - Darkwebで利用されていたアドレスの識別率が最も高い(平均 0.74)

アドレス集合	利用方法 [%]					合計
	BBS	ATM	Darkweb	Exchange	Mining Pool	
S	0.43	0.55	0.80	0.14	0.59	75
R	0.57	0.12	0.74	0.49	0.17	63
I	0.59	0.55	0.74	0.42	0.50	84
O	0.65	0.14	0.70	0.75	0.17	72
平均	0.56	0.34	0.74	0.45	0.36	

結論

- 新たなアドレス集合を特徴量としたアドレス識別手法の提案
  - アドレス識別に影響が現れにくい送金元アドレス集合*R*, 出力アドレス集合*O*の提案

- 4つのアドレス集合を用いたアドレス識別実験を実施
  - 出力アドレス集合*O*(提案手法)はアドレス識別率が最も高い(54.7)
  - 4つの集合のうち3つの集合*R*,*I*,*O*において取引回数と識別率の相関関係が見られた
  - アドレスの利用目的別の識別率ではDarkwebの値が最も高い(0.74)

- 今後の課題
  - アドレスの利用目的を特徴量に含めた識別実験の実施