

---

2021年 2月 12日  
修士論文発表会

# 取引履歴の特徴量に基づく Bitcoinアドレスの識別リスクの評価

松本 寛輝

菊池研究室

# 背景

## ■ Bitcoinの匿名性

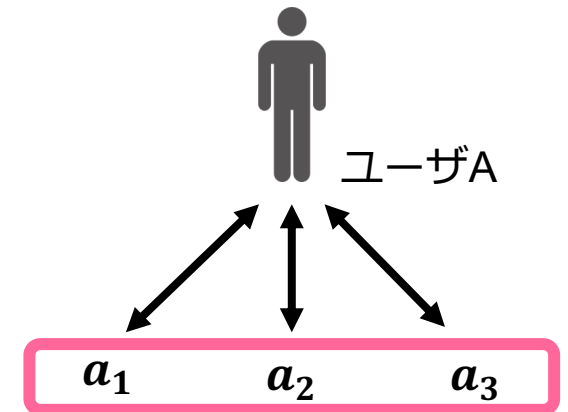
- Bitcoinアドレス(ランダムな文字列)に基づく
- アドレスとユーザが繋がる情報はない
- ユーザを特定できない = 匿名性が高い

## ■ 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Aが管理している
  - ※ ユーザAの情報(氏名など)は特定できない



Bitcoinアドレス



Bitcoinアドレスの識別

# Bitcoinの取引例

- 例) ユーザAがユーザBに5BTC送金

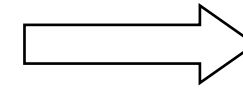
所有者	アドレス	送金前の残高	送金後の残高
ユーザA	$a_1$	3 BTC	1 BTC
	$a_2$	3 BTC	0 BTC
ユーザB	$b_1$	0 BTC	5 BTC

$a_1$ は取引のおつりを  
受け取っている

ユーザA



5 BTC



ユーザB

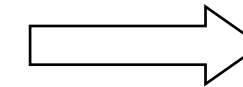


取引情報

$a_1$  (3 BTC)

$a_2$  (3 BTC)

**Input**



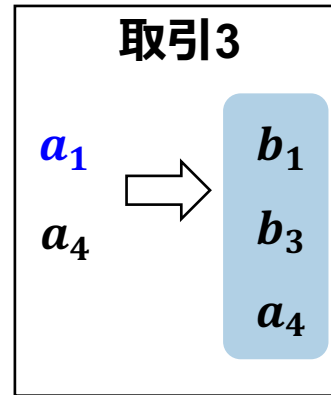
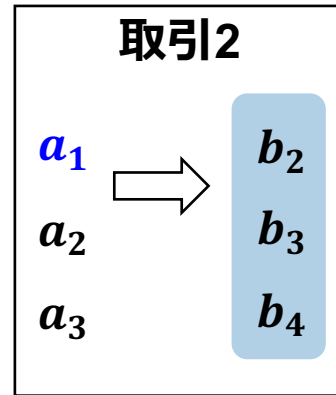
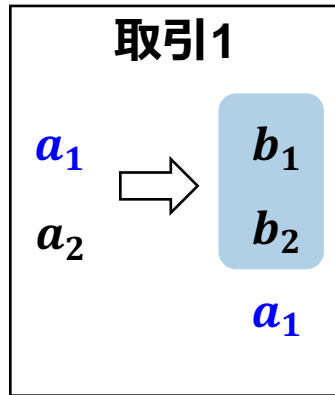
$b_1$  (5 BTC)

$a_1$  (1 BTC)

**Output**

# Bitcoinアドレスの識別手法(先行研究)

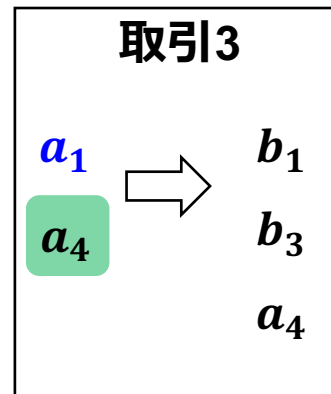
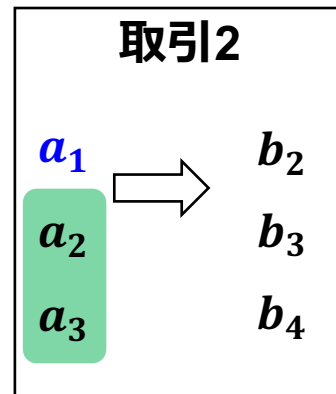
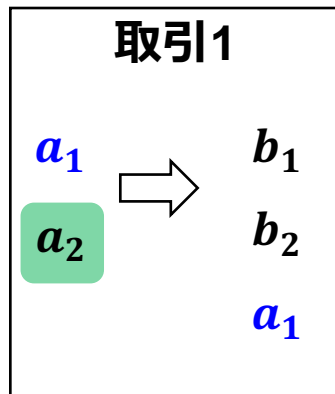
## ■ 永田ら[2018]による宛先アドレスを用いたアドレス識別



$a_1$ の宛先アドレス集合 $S(a_1)$

$$S(a_1) = \{b_1, b_2, b_3, b_4, a_4\}$$

## ■ Meiklejohnら[2013]による入力アドレスを用いたアドレス識別



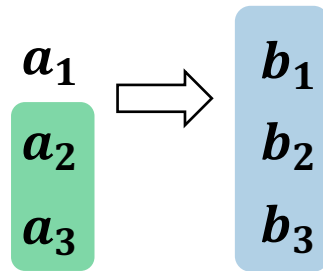
$a_1$ の入力アドレス集合 $I(a_1)$

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

# 先行研究[永田,2018]の課題と提案手法

- 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( $a_1$ )はユーザが自身で選択できるため識別精度に影響を与える
- 「送金元アドレス」と「出カアドレス」を用いたアドレス識別を提案
  - 自身のアドレス( $a_1$ )に対して送金を行うアドレスは指定できない

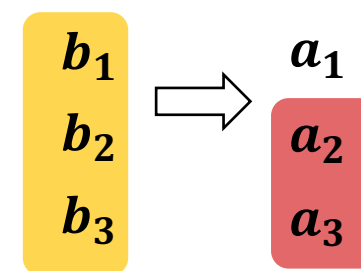
既存手法：アドレス $a_1$ が送金



入カアドレス  
集合 $I(a_1)$

宛先アドレス  
集合 $S(a_1)$

提案手法：アドレス $a_1$ が受け取り

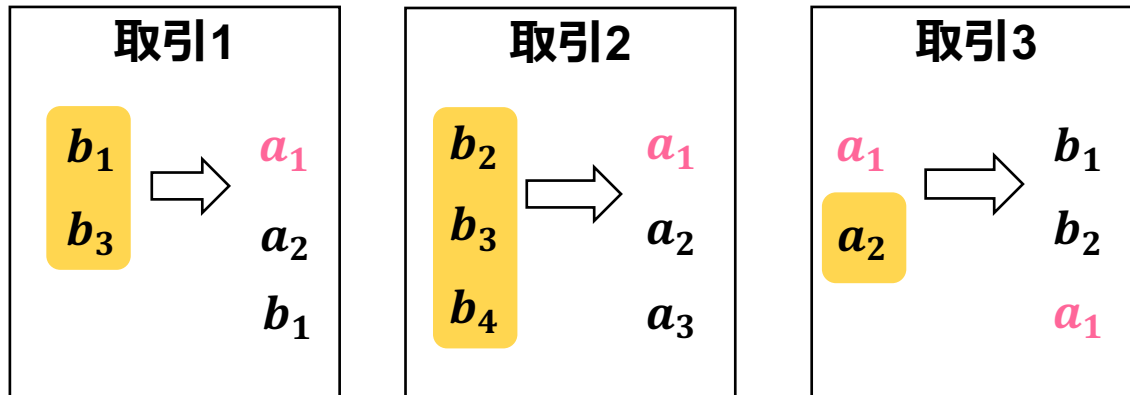


送金元アドレス  
集合 $R(a_1)$

出カアドレス  
集合 $O(a_1)$

# 受け取りに着目したの識別手法(提案手法)

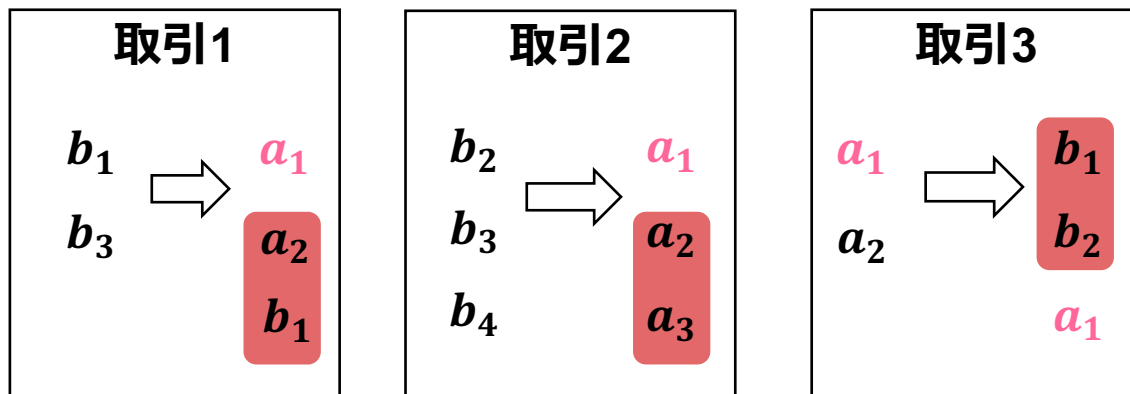
## ■ 提案手法1. 送金元アドレスを用いたアドレス識別



$a_1$ の送金元アドレス集合 $R(a_1)$

$$R(a_1) = \{b_1, b_2, b_3, b_4, a_2\}$$

## ■ 提案手法2. 出力アドレスを用いたアドレス識別

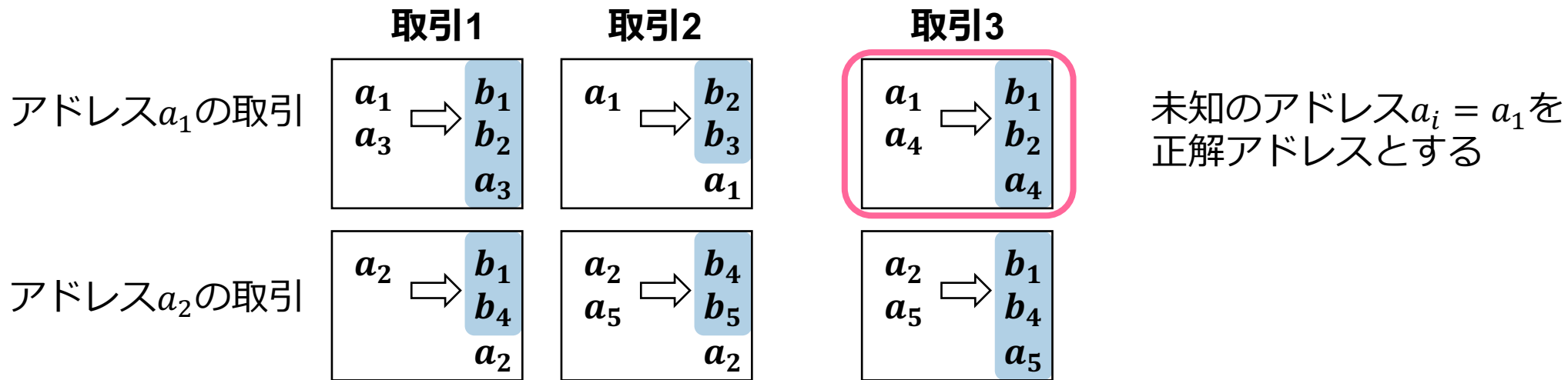


$a_1$ の入力アドレス集合 $O(a_1)$

$$O(a_1) = \{b_1, b_2, a_2, a_3\}$$

# アドレス識別のアルゴリズム

- 宛先集合 $S$ を用いた未知のアドレス $a_i$ の識別手法 (集合 $R, I, O$ も同様)



アドレス	学習アドレス集合 $S(a)$ (取引1,2)	評価アドレス集合 $S'(a_i)$ (取引3)	Jaccard係数 $\frac{ S(a) \cap S'(a_i) }{ S(a) \cup S'(a_i) }$
$a_1$	$\{b_1, b_2, b_3, a_3\}$	$\{b_1, b_2, a_4\}$	$\frac{ \{b_1, b_2\} }{ \{b_1, b_2, b_3, a_3, a_4\} } = 0.40$
$a_2$	$\{b_1, b_4, b_5\}$		$\frac{ \{b_1\} }{ \{b_1, b_2, b_4, b_5, a_4\} } = 0.20$

Jaccard係数の  
値が大きい  
 $a_1 = a_i$ と予測  
→ 正解

# 研究目的と実験概要

---

## ■ 研究目的

- 4つの集合のうち、識別精度が高いアドレス識別手法を明らかにする
  - » 識別精度を正確に評価することが困難

## ■ 実験概要

- 実験1. 取引回数に基づく識別
  - » 10年間で継続して使用されたアドレスの識別リスクを定量的に示す
- 実験2. 利用目的に基づく識別
  - » 5種類のアドレスを収集し、最も識別精度が高い利用目的を明らかにする



# 実験1. 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日から2019年11月18日の10年分の取引
- 取引回数が最低2回以上のアドレスを使用

## ■ 実験手順

1. 取引回数を  $n$  と定義
  - 21回送金と受け取り取引を行なった ( $n = 21$ )
  - 多くのアドレスの取引回数をまとめて評価する時は10刻みなどに量子化 ( $n = 21, 22, \dots, 30 \rightarrow n = 30$ )
2. 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
3. 既存手法(集合  $S, I$ ) と提案手法(集合  $R, O$ ) を用いてアドレスを識別

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

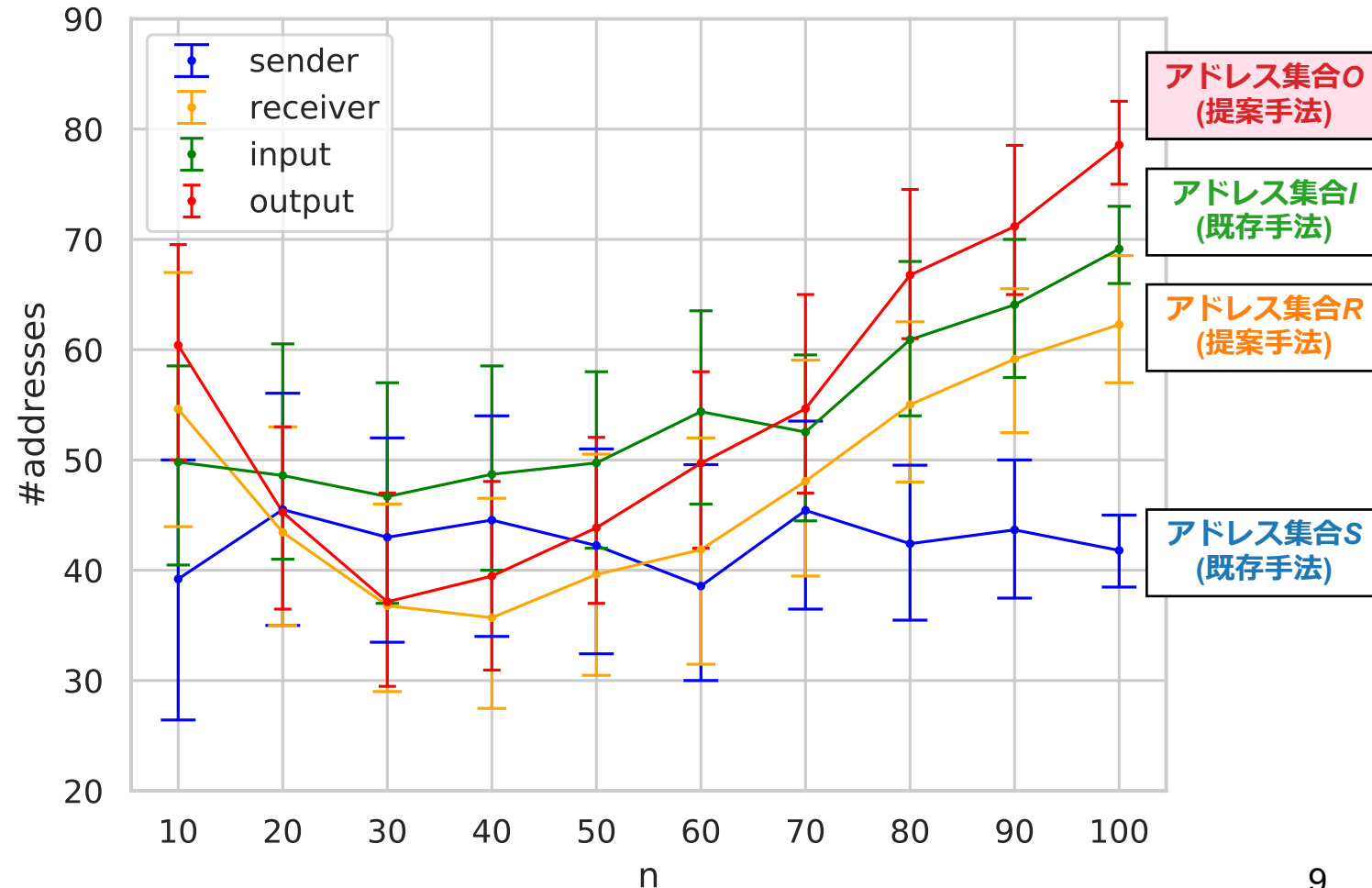
# 実験1.結果 取引回数と識別精度の推移

## ■ 実験結果

- エラーバーは95%の信頼区間を示す

## ■ 識別精度

- **集合0の識別精度が最も高い**  
( $n = 100$  のとき, 78.6)
- 集合0の最も低い識別精度  
( $n = 30$  のとき, 37.1)
  - » 集合R, I, Oは $n = 30$ の前後で識別精度が最も悪くなる



# 実験1.提案手法の精度が高いことの仮説検定

## ■ 仮説検定

- 提案方式 $R, O$ の識別精度が既存手法 $S, I$ よりも高いことを確かめる
- 平均値の $t$ 検定を実施
- 帰無仮説：2つのアドレス集合の識別率は一致する
- $n = 100$ における4つの方式の精度を比較

## ■ 検定結果

- 集合 $O$ を用いた提案手法の識別精度は既存手法 $S, I$ よりも統計的に有意な水準である

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

# 実験2. アドレスの利用目的に基づく識別

## ■ アドレス数

- Bitcoinの利用目的のうち, 代表的な5つのサービス(交換所等)を対象
- 2019年4月1日から9月30日の半年間で取引をおこなったアドレスを収集
- 取引回数が最低2回以上のアドレスを使用

## ■ 実験手順

1. 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
2. 既存手法(集合 $S, I$ )と提案手法(集合 $R, O$ )を用いてアドレスを識別

利用目的	概要	アドレス数	サンプリング数
Bitcointalk (BBS)	登録を行っているユーザのアドレス	844	30
Bitcoin ATM	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	106	30
Dark web	違法商品・サービスを提供, 利用しているアドレス	49	30
Exchange	交換所を利用しているユーザのアドレス	274	30
Mining Pool	マイニング報酬を受け取るアドレス	85	30
合計	-	1,358	150

# 実験2.結果 利用目的ごとの識別率

## ■ 実験結果

- Dark webで利用されていたアドレスの識別数が最も高い(平均 **22.3**)
  - › Dark webはサイトの運営期間が短く、アドレスの利用期間も短いと考えられる
- アドレス集合Iは総合的な識別精度が最も高い(合計 **84個**)
  - › 利用目的別にみると**集合S**や**集合O**の識別精度が上回る

識別手法	集合	利用目的					合計
		BBS	ATM	Dark web	Exchange	Mining Pool	
永田ら	S	12.8	16.6	23.9	4.1	17.8	75
提案手法 1.	R	17.2	3.6	22.2	14.7	5.0	63
Meiklejohnら	I	17.6	16.5	22.3	12.6	15.0	84
提案手法 2.	O	19.5	4.3	20.9	22.6	5.0	72
	平均	16.8	10.2	22.3	13.5	10.7	

# 結論

---

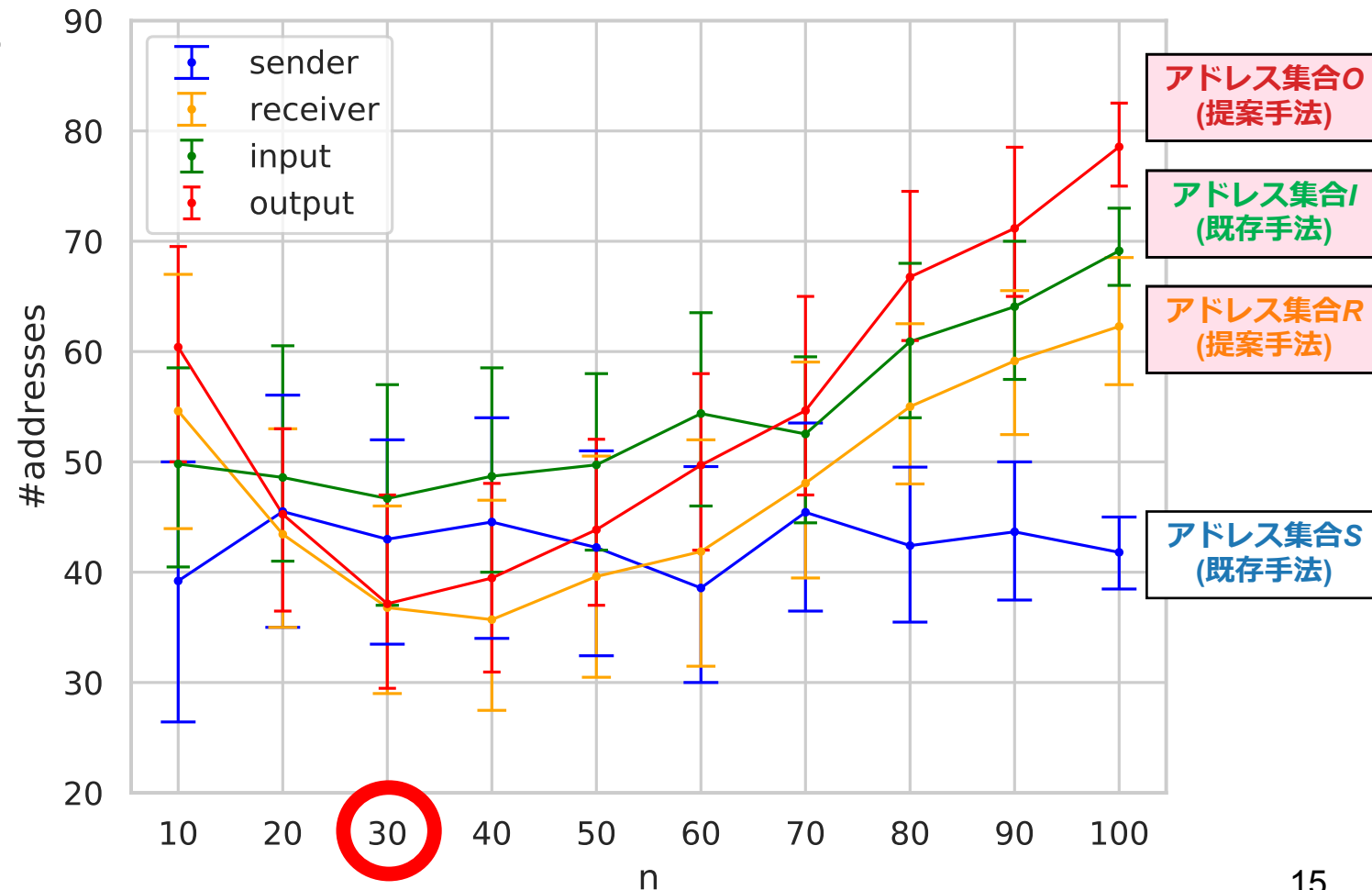
- 受け取り取引に注目した新たな2つの識別手法を提案
  - **提案手法の集合 $O$** を用いた識別精度は**既存手法の集合 $S, I$** よりも統計的に有意な水準で高い
- 実験結果
  - 実験1. 取引回数に基づく識別
    - » 集合 $O$ では,  $n = 30$ のとき, アドレス識別精度が**最も低い(37.1個)**
    - » 集合 $O$ では,  $n = 100$ のとき, アドレス識別精度が**最も高い(78.6個)**
  - 実験2. 利用目的に基づく識別
    - » 5種類のうち **Dark web** の識別精度が高い
    - » 集合 $I$ の精度は総合的に高い, 利用目的別では集合 $S, O$ の精度が高い

---

# 質疑応答スライド

# 実験1. $n = 30$ 付近で識別精度が低下

- $n = 30$ 付近で識別精度が低下
  - 調査1.アドレスが毎回更新される新しいウォレットが多い
    - » アドレスの取引開始時期の調査
  - 調査2.学習量が少ないため
    - » 学習アドレス集合の大きさを調査
  - 調査3.特定の利用目的のものが偏っている
    - » 収集した10年分のアドレスと利用目的を調査





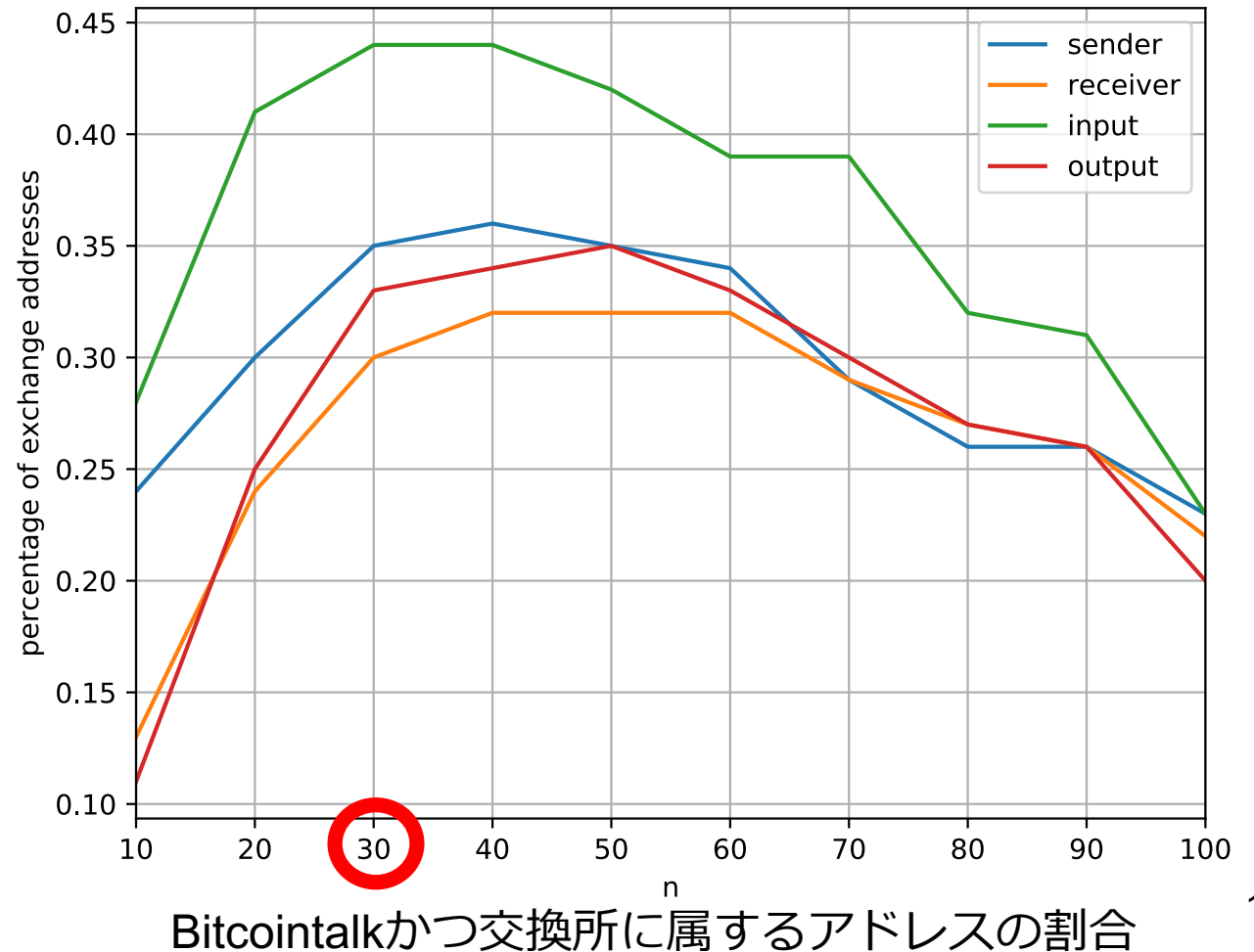
# 実験1. 識別精度が低下した原因

## ■ 原因「調査3. 特定の利用目的のものが偏っている」

- 収集したアドレスのうち  
交換所(Exchange) で利用されている  
アドレスを調査
- $n = 30$ の付近で識別精度とは  
逆の振る舞い(上に凸)となった

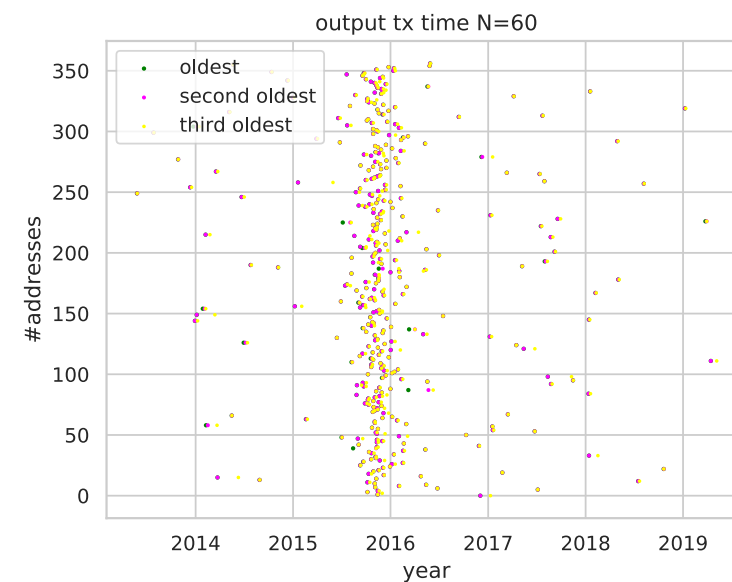
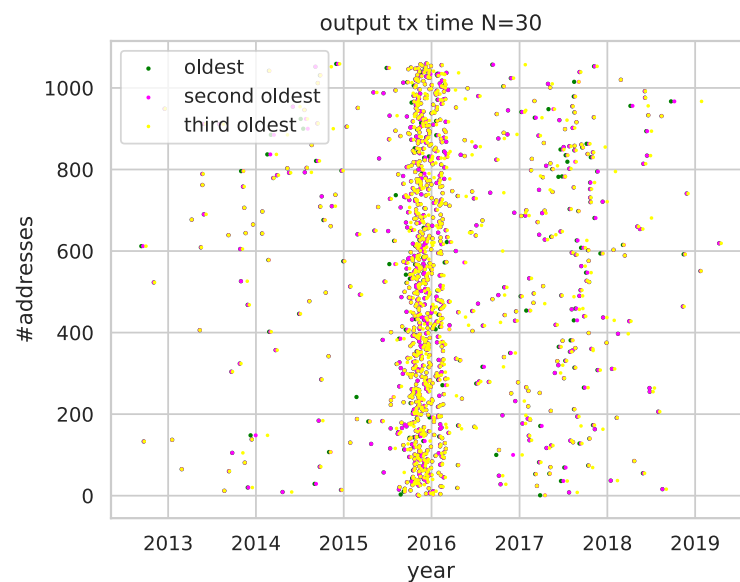
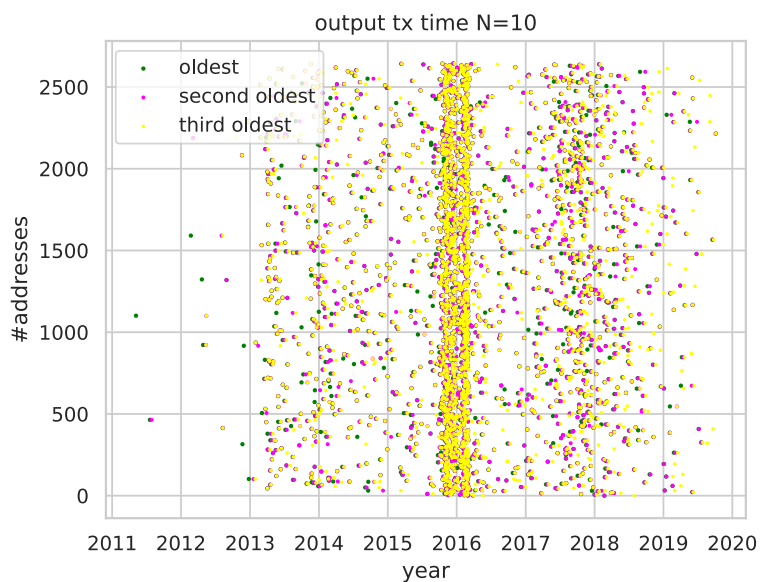
## ■ 交換所アドレスが含まれると 識別精度が低下する

- 修士論文中では, 交換所アドレスと  
Bitcointalkアドレスのみを用いた  
識別実験も実施



# 実験1. 識別精度が低下した原因(調査1)

- 調査1. アドレスが毎回更新される新しいウォレットが多い
  - アドレスの取引開始時期の調査
  - 取引回数 $n = 30$ 前後で大きな差はなし

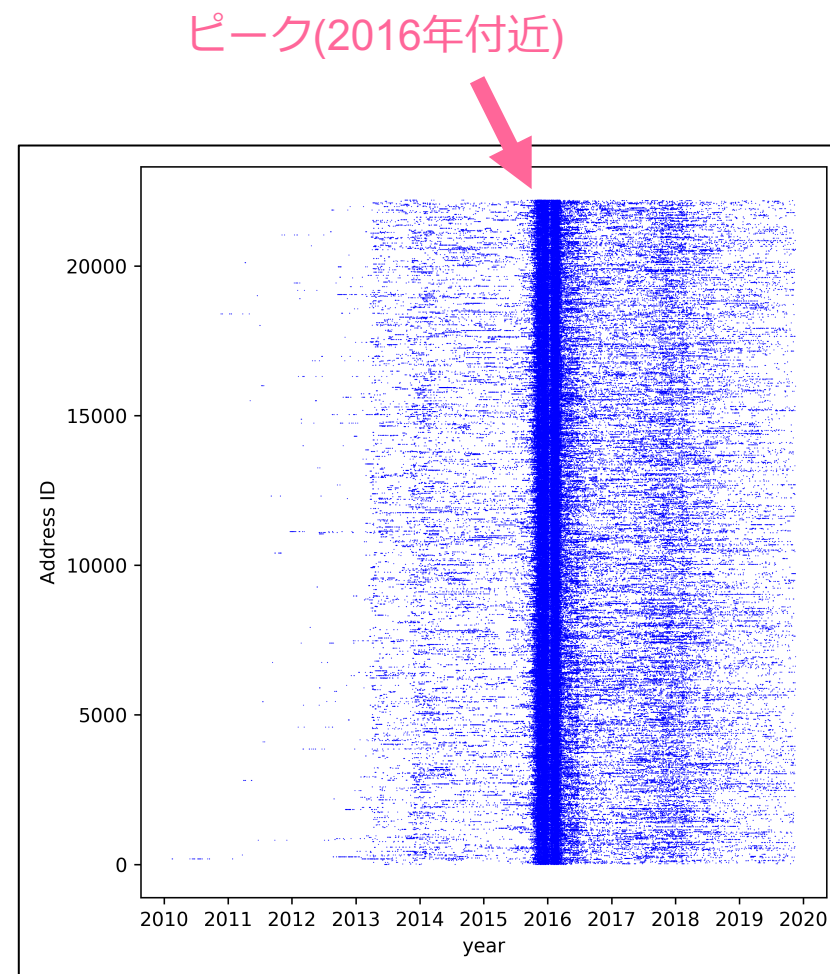


出カアドレス集合 $O$ について, 左から $n = 10, n = 30, n = 60$ における取引開始時期の散布図

# 参考：10年間の取引数の分布



Blockchain.com 1日あたりの確認済みトランザクション数  
(<https://www.blockchain.com/charts/n-transactions>)

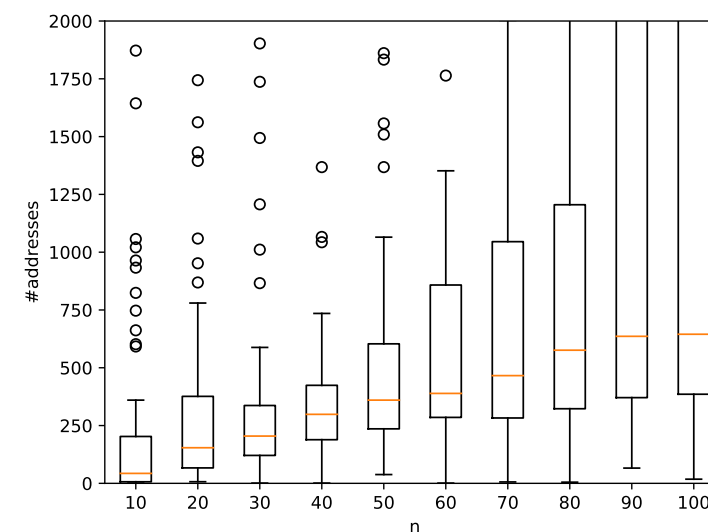
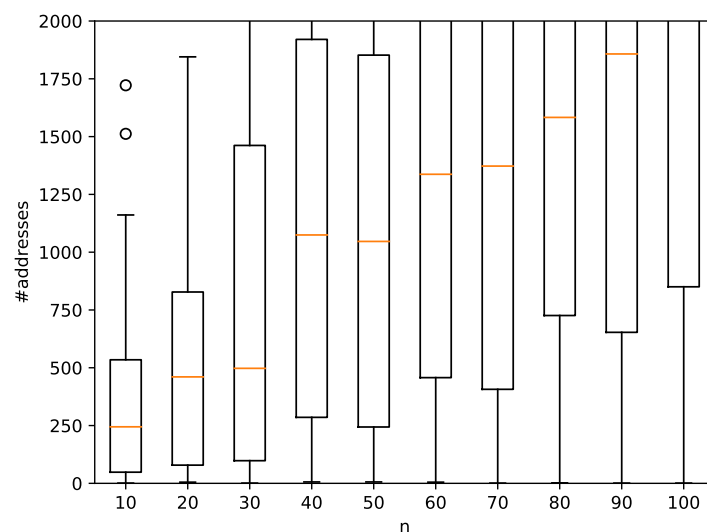
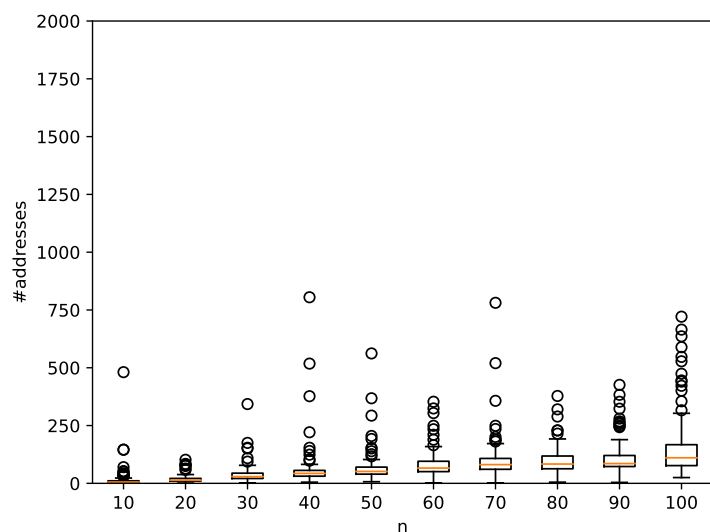


約10年間の取引分布(bitcontalkアドレス)

# 実験1. 識別精度が低下した原因(調査2)

## ■ 調査2. 学習量が少ないため

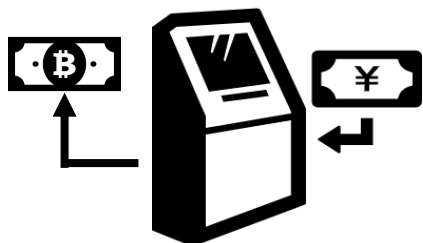
- 学習アドレス集合の大きさを調査
- 取引回数  $n = 30$  前後で大きな差はなし



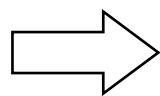
取引回数  $n$  について, 左から集合  $R, I, O$  における学習アドレス数の分布

# 提案手法R, Oで識別精度が低い利用目的

## ■ Bitcoin ATM



Bitcoinの購入



ATMのアドレスから送金

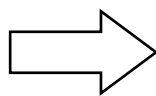


ユーザ

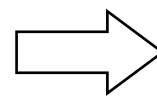
## ■ Mining pool(マイニングプール)



マイニング報酬



マイニングプール  
事業者



報酬の分配



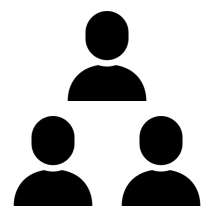
マイナー

ブロックチェーン  
(Bitcoin)

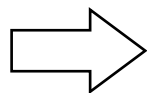
# 交換所アドレスが識別精度を低下させる原因

## ■ 交換所アドレスの特殊な取引

- 登録ユーザのアドレスに保管されているBitcoinを自社のアドレスへ送金する
  - » 分散されたBitcoinを一元に管理する
  - » 一定期間ごとに実施
- 右図では、100個のユーザのアドレス(赤枠)から自社のアドレス(青枠)に送金を行う取引



交換所登録ユーザ



交換所アドレス

chainFlyer

アドレスを検索 / トランザクションハッシュ / ブロックハッシュ

トランザクション

62a0b59372f4c7be314752f278f4f3f64c78ffcc1fad53fd601f45e6f2a855ba

75429 確認

受取日時 2019/09/14 13:45:49 JST

サイズ 29,952 bytes  
29,952 vbytes

ウェイト 119,808

送信額 1.60510349 ₿

手数料 0.00039353 ₿  
1.314 satoshis/byte  
1.314 satoshis/vbyte  
0.328 satoshis/weight unit

ブロックの高さ 594781

Input (100)

3KJEbGTyw6Cz7g4JH7K6w76KLJumFK1y8G	0.00001496 ₿
34AkHZuqQaLaChmEQThMApyAgFUUjJE	0.0000385 ₿
3PwcdcJzyQTsHmYuhWn3rgKWkuaP1LJ6d5	0.0000445 ₿
35RPobSwcD1EFVy6qmhct328TDPToPp3L	0.00011415 ₿
3Bv4i5EzZdKVSsKrH2gaF3L84MPkDubtp1	0.00014 ₿
35RPobSwcD1EFVy6qmhct328TDPToPp3L	0.000276 ₿
3LbxUAsGn6QsuMHXMcGodFoRD4rWVVSauti	0.0005 ₿

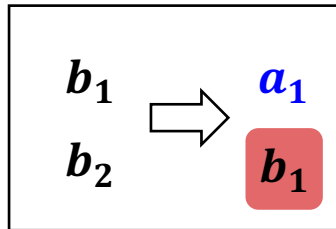
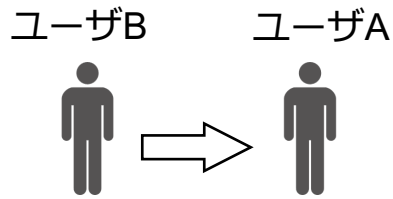
Output (1)

bc1qy30guv6m5ez7n10ayro8u23w3k5s8vg3elmxd zlh8a3skupyqn2lp5w	1.60510349 ₿
---	--------------

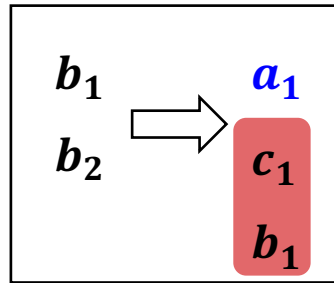
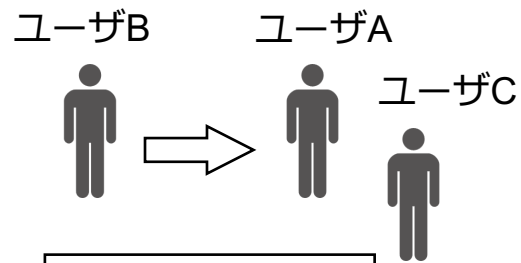
交換所アドレスの特殊な取引

# アドレス識別対策

- 集合 $O$ を利用したアドレス識別手法への対策
  - 受け取りアドレスを1つ定めて再利用する



1対1の取引例

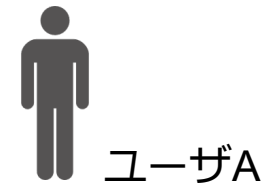


1対多の取引例

$a_1$ の入カアドレス集合 $O(a_1)$

$$O(a_1) = \{b_1, c_1\}$$

→ ユーザAが管理する  
他のアドレス( $a_2, a_3$ など)は識別できない



Bitcoinアドレスの識別

# 収集したアドレスの背景

## ■ アドレスの再利用

- 開発コミュニティ\*では「一度利用したアドレスを再び利用しない」ことを推奨
  - » 実験1で収集したアドレス**23,541個は2回以上取引を行っている**
- 前提：アドレスは再利用される

## ■ 収集したアドレスの例

- Bitcointalkでプロフィールページ\*\*にあるアドレス
  - » プロフィールページに書かれているアドレスは頻繁に更新されないはず
- 送金する際に青枠のアドレスを利用：多？少？
- 受け取る際に青枠のアドレスを利用：多

Summary - FlightyPouch	Picture/Text
<b>Name:</b> FlightyPouch <b>Posts:</b> 3378 <b>Activity:</b> 1232 <b>Merit:</b> 287 <b>Position:</b> Sr. Member <b>Date Registered:</b> October 11, 2016, 02:15:03 PM <b>Last Active:</b> Today at 12:37:36 AM	
<b>ICQ:</b> <b>AIM:</b> <b>MSN:</b> <b>YIM:</b> <b>Email:</b> hidden <b>Website:</b> <b>Current Status:</b> <input type="checkbox"/> Offline <b>Bitcoin address:</b> <span style="border: 2px solid blue; padding: 2px;">3PyrwHe7oDdk739x78n1sUVdnadJh4fmSb</span>	
<b>Gender:</b> <b>Age:</b> N/A <b>Location:</b> 0x6B3A0003A273A8bCF061cD3a611277Bec8810EDb <b>Local Time:</b> February 14, 2020, 07:34:55 AM	
<b>Signature:</b> Fast 1% Dice <input type="checkbox"/> Rakeback <input type="checkbox"/> <b>YOLOdice.com</b> <input type="checkbox"/> Competitions <input type="checkbox"/> Exchange <b>BTC LTC ETH DOGE</b>	
<b>Additional Information:</b> Show the last posts of this person. Show the last topics started by this person. Show general statistics for this member.	

\* Bitcoin.org プライバシーの保護 (<https://bitcoin.org/ja/protect-your-privacy>)

\*\* bitcointalk プロフィールページ (<https://bitcointalk.org/index.php?action=profile;u=907855>)



# 実験2. アドレスの利用目的について

## ■ アドレスの利用目的

- Bitcoinの利用目的のうち, 代表的なサービス(交換所等)を対象

## ■ 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	識別精度(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	△
Bitcoin ATM	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	△
Dark web	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	○

# アドレス識別と社会への影響

## ■ アドレス識別の影響

- アドレスの管理者(ユーザ)の追跡や属性推定に利用

## ■ Bitcoinの市場の動き

- テスラ社がBitcoinへ投資
  - » 将来的にはBitcoinでの販売も検討
- 特定の商品を購入したアドレスの追跡リスク
  - » 交換所などから個人情報が漏洩する可能性
- Dark web(闇市場)の追跡
  - » Zcash(より匿名性の高い暗号資産)で追跡を行った研究もある

coindesk JAPAN

NEWS ▾ CHART ▾ CRYPTO / EQUITY / FX ▾ FEATURE ▾ EVENTS ▾ Q

月々約8,300円で有能な秘書を持つ方法——時間の質にこだわるエグゼクティブのBlack Card 【Sponsored】

今すぐ取引したい! ビットコイン 購入まで最短10分! GMOクリック証券

### テスラ、ビットコインに15億ドルを投資——史上初の4万7000ドル突破【更新】

2021年 2月 9日 17:25 · 2021年 2月 9日 17:25 更新

🐦 f B!

Daniel Palmer

Data

Bitcoin 5,069,452 円

米電気自動車 (EV) 大手のテスラが、暗号資産のビットコイン (BTC) に15億ドル (約1580億円) を投資していたことがわかった。

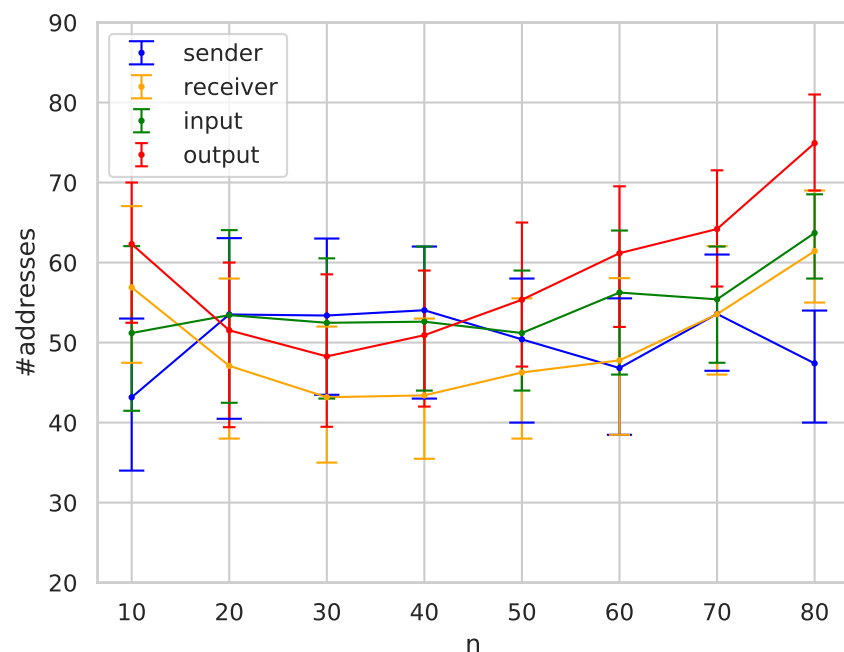
テスラが米証券取引委員会 (SEC) に提出した2020年度の年次報告書によると、同社は投資に対する方針を1月に更新。それに基づき15億ドルをビットコインに投資したと述べた。

テスラは今後、状況に応じてデジタル資産の購入と保有を検討していくと、同報告書に記した。また、テスラは同社のプロダクトに対する支払いにおいて、ビットコインの利用を受け入れる準備を検討していることも明らかにした。受け入れ次期は「近い将来」としている。

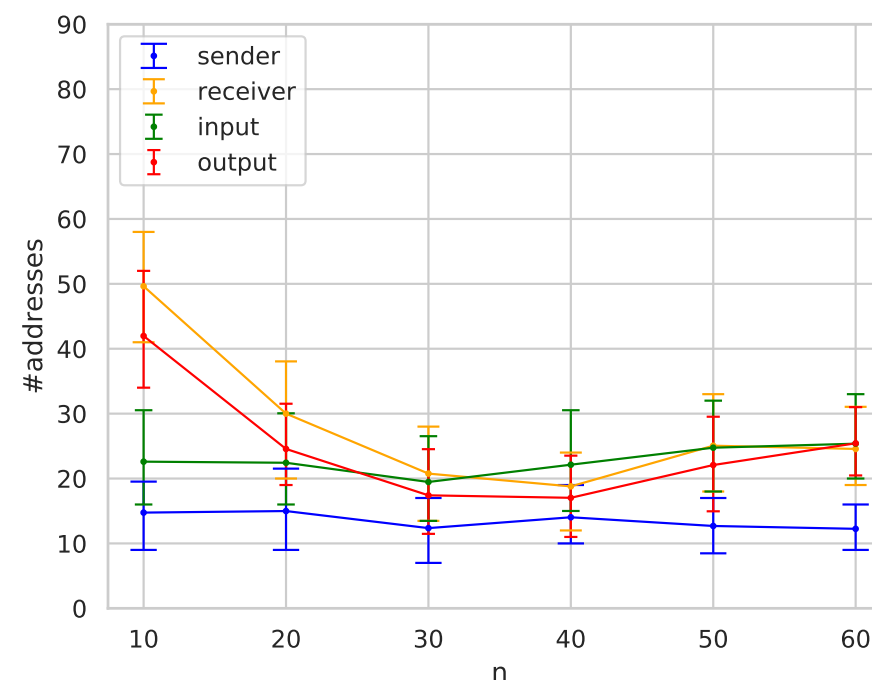
# 取引回数と利用目的の識別精度

## ■ 10年間のアドレスのBitcointalkと交換所アドレスの識別

□ エラーバーは95%の信頼区間を示す



取引回数とBitcointalkの識別精度



取引回数とExchangeの識別精度