Title	暗号通貨Bitcoinにおいて犯罪に使用される取引の検知手法に関する研究
Sub Title	Research on fraud detection in Bitcoin based on transaction analysis
Author	豊田, 健太郎(Toyoda, Kentaro)
Publisher	
Publication year	2018
Jtitle	科学研究費補助金研究成果報告書 (2017.)
JaLC DOI	
Abstract	Bitcoinは, 信頼できる中央機関の存在を必要としない金融取引システムであり, 少額の取引手数料での送金, 透明性の高い寄付および出資などの手段として注目を浴びている。 しかしながら, Bitcoinは匿名性があり, 投資詐欺などに悪用されることが問題となっている。 そこで本研究では, 投資詐欺に関連したBitcoinアドレスをWebから収集する手法を提案し, Bitcoinの取引履歴を機械学習を用いて解析する手法を提案し, さらに, 犯罪に用いられる取引をリアルタイムに検知するシステムを構築した。結果, 918個のBitcoinアドレスを発見し, 偽陽性率を3.8%に抑え, 88%の精度で検知可能なことを示した。 Bitcoin is one of the most successful decentralized cryptocurrencies to date. However, it has been reported that it can be used for investment scams, which are referred to as HYIP (High Yield Investment Programs). So far, no schemes has been proposed to detect HYIP operators' Bitcoin addresses, although it is useful from the security forensics aspect. We have proposed a novel scheme to identify HYIP operators' Bitcoin addresses by analyzing transactions history. We collected 918 HYIP operators' Bitcoin addresses from the Internet and analyzed the characteristics of transactions where the collected Bitcoin addresses are involved. Based on this analysis, we proposed a machine learning technique to classify given Bitcoin addresses into HYIP operators ones or not. By evaluating the classification performance , our best scheme achieves that 88% of HYIP addresses are correctly classified, while maintaining false positive rate less than 3.8%. We also built a web application for practice.
Notes	研究種目 : 研究活動スタート支援 研究期間 : 2016 ~ 2017 課題番号 : 16H07168 研究分野 : 情報工学
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KAKEN_16H07168seika

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　５　月　２８　日現在

機関番号：　３２６１２
研究種目：　研究活動スタート支援
研究期間：　2016～2017
課題番号：　１６Ｈ０７１６８

研究課題名（和文）暗号通貨Bitcoinにおいて犯罪に使用される取引の検知手法に関する研究

研究課題名（英文）Research on Fraud Detection in Bitcoin Based on Transaction Analysis

研究代表者

豊田　健太郎（Toyoda, Kentaroh）

慶應義塾大学・理工学部（矢上）・特任助教

研究者番号：６０７２３４７６

交付決定額（研究期間全体）：（直接経費）　　2,300,000円

研究成果の概要（和文）：Bitcoinは，信頼できる中央機関の存在を必要としない金融取引システムであり，少額の取引手数料での送金，透明性の高い寄付および出資などの手段として注目を浴びている．
しかしながら，Bitcoinは匿名性があり，投資詐欺などに悪用されることが問題となっている．
そこで本研究では，投資詐欺に関連したBitcoinアドレスをWebから収集する手法を提案し，Bitcoinの取引履歴を機械学習を用いて解析する手法を提案し，さらに，犯罪に用いられる取引をリアルタイムに検知するシステムを構築した．結果，918個のBitcoinアドレスを発見し，偽陽性率を3.8%に抑え，88%の精度で検知可能なことを示した．

研究成果の概要（英文）：Bitcoin is one of the most successful decentralized cryptocurrencies to date. However, it has been reported that it can be used for investment scams, which are referred to as HYIP (High Yield Investment Programs). So far, no schemes has been proposed to detect HYIP operators' Bitcoin addresses, although it is useful from the security forensics aspect. We have proposed a novel scheme to identify HYIP operators' Bitcoin addresses by analyzing transactions history. We collected 918 HYIP operators' Bitcoin addresses from the Internet and analyzed the characteristics of transactions where the collected Bitcoin addresses are involved. Based on this analysis, we proposed a machine learning technique to classify given Bitcoin addresses into HYIP operators ones or not. By evaluating the classification performance , our best scheme achieves that 88% of HYIP addresses are correctly classified, while maintaining false positive rate less than 3.8 %. We also built a web application for practice.

研究分野：　情報工学

キーワード：　Bitcoin　データ解析　取引詐欺

１．研究開始当初の背景

Bitcoin は，信頼できる中央機関の存在を必要としない金融取引システムであり，少額の取引手数料での送金，透明性の高い寄付および出資などの手段として注目を浴びている．Bitcoin では，全ての取引は誰でも確認可能な公開台帳（ブロックチェーン）に記録され，利用者がそれらの取引を検証することにより実現される．しかしながら，Bitcoin は無数に作成可能な公開鍵を用いて取引を行うため匿名性があり，麻薬組織の金銭の授受，マネーロンダリング，金融詐欺行為などに悪用されることが問題となっている．

２．研究の目的

そこで本研究では，公開されている Bitcoin のブロックチェーンを統計的機械学習を用いて解析し，犯罪に起因する取引の特徴量の発見，さらにそれらの特徴量を活かし，犯罪に用いられる取引の特徴を明らかにすることで犯罪に用いられる取引をリアルタイムに検知するシステムを構築する．(1) ポンジ・スキーム，(2) 入金済みの商品を発送しない詐欺行為，(3) 口座管理サービスによる顧客の口座から預金を不正に取り出す窃盗行為，(4) 実通貨との取引所における詐欺行為の 4 つを対象とし，それぞれの犯罪に特徴的な取引量，取引形態，取引時間，取引手数料，および取引の軌跡といった情報を明らかにする．さらに発見した特徴量をどのように活用すれば犯罪に使用されうる取引，もしくはアカウントを検知できるかといったアルゴリズムにまとめる．最後に実際のブロックチェーンに対してアルゴリズムを適用した検知システムを提供する．

３．研究の方法

(1)　平成 28 年度は，公開されている Bitcoin のブロックチェーンを統計的機械学習を用いて解析し，犯罪に起因する取引の特徴量を明らかにすることを目標とする．そのために，まず本研究費で購入予定のサーバに Ubuntu 16.04 をインストールし，Bitcoin の公式サイトにおいて提供されている bitcoind を導入し，ブロックチェーンをダウンロードする．次にダウンロードしたブロックチェーンを解析するプログラムの記述を行う．本プログラムは C++言語で記述予定であり，既に公開されているブロックチェーンをデータとして処理可能な形式に変換するツール blockparser (https://github.com/znort987/blockparser) を参考にする．可能であればブロックチェーン全体を読み込み，各取引を処理できる形式に変換する．

(2)　そして犯罪に起因する取引の特徴量候補を抽出する過程に移る．まず，高収益投資プログラム（HYIP: High Yielding Investment Programs)と呼ばれる比較的古典的な投資詐欺に注目した．そこで Bitcoin における HYIP に関連した取引履歴から，取引形態，取引時間，取引手数料，および取引の軌跡といった情報を抽出する．ここで，実際の HYIP に用いられた取引を特定するために，実際に被害にあったユーザなどが Bitcoin に関するフォーラムの (bitcointalk.org)および不正なサービスのリスト(www.badbitcoin.org)等に告発しているためこれらの情報を利用する．それらの取引に使用されているアカウントの前後の資金の流れを追うことで，各犯罪に共通した特徴量を発見する．具体的には，取引量，取引形態，取引時間，取引手数料，および取引の軌跡といった情報を利用する．これらの情報のうち，取引量，時間，手数料は統計量であるため，分類器である SVM (Support Vector Machine)および RF (Random Forests)をはじめとする教師あり学習法を用いて学習を行い既知の犯罪の取引を分類できるかを R 言語を用いて検証する．

(3)　平成 29 年度は，より多くの HYIP のデータを収集するため，bitcointalk.org のトピックから自動的にスクレイピングする．多くの HYIP が投資を募るために bitcointalk.org において宣伝を行うことに着目し，該当するトピック内から HYIP に関連した Bitcoin アドレスを収集する．HYIP を宣伝するトピック内には，実際に投資もしくは還元の証拠を示すために Bitcoin トランザクションの ID が投稿されることが多いため，全投稿に含まれる Bitcoin トランザクションを抽出した上で，HYIP を運営する側の Bitcoin アドレスを抽出する．トランザクションには送金側と受領側の双方の Bitcoin アドレスが含まれるため，抽出されたトランザクションが HYIP の投資もしくは還元のいずれの証拠であるかを判別する必要がある．提案方式では，トランザクション ID と共に投稿される文章に対して単純なテキストマイニングを施すことでこれを解決する．

(4)　さらに，前年度の研究を継続および発展させ，得られた特徴量および取引の軌跡のパターンを基に検知アルゴリズムを構築し，ブロックチェーン上の犯罪に使用される取引を検知するシステムの構築まで行う．サーバマシン上に上述のアルゴリズムを実装し，システムの構築を行う．本システムは平成 28 年度に購入したサーバで収集したブロックチェーンを処理し，R 言語およびウェブアプリケーション開発パッケージである RShiny (https://shiny.rstudio.com) を使用したウェブアプリケーションとして稼動させる．

４．研究成果

(1) まず HYIP に関連した Bitcoin アドレスをマニュアルで収集した結果について示す．実際に HYIP の Bitcoin に関するフォーラムの (bitcointalk.org) トピックを読んでいくと，時折``payed''と書かれた

投稿とともに，その証拠のトランザクションの識別子が公開されていることがある．これは既に出資した出資者によって投稿されるケースが多い．その理由として，この出資者はその後に別の出資者から出資がなければ利息の配当を受けられないため，そのHYIPが信頼できることを示すためである．当然，そのような投稿の信憑性は不明であるが，いくつかの信頼できる手掛かりがあると考えられる．1つ目に，支払いを完了したことを示す投稿の日時がトランザクションの承認時間の直後（数分後)にあった場合である．2つ目に，いくつかのHYIPはPublic Noteと呼ばれるメモ書きをトランザクションに付加している場合である．上記の手法で，2013年から2016年に存在していた計48個のHYIPに関連したBitcoinアドレスを収集した．しかしながら，これらのアドレスだけでは各HYIPの特徴を捉えるには不十分である．そこで，非匿名化手法を用いて，収集した48個のBitcoinアドレスのそれぞれに対して，各エンティティが保持する他のBitcoinアドレスを抽出した．これらの中にはHYIPに直接利用されていないものも含まれている可能性があるが，それらの識別は困難であるため，ここでは非匿名化手法で抽出された全アドレスをHYIPに関連しているとみなす．その結果，収集した48個のHYIPのうち29%のHYIPは単一のBitcoinアドレス，38%のHYIPは10以上，23%のHYIPは100以上のBitcoinアドレスで運用されていることがわかった．

(2) 以上の結果より，HYIPに使用されているアドレスが含まれる一連のトランザクションを調査した．その結果，エンティティ毎に，トランザクションの送金・受領の方向および額の大きさ，時系列上での順序に傾向があることがわかり，それらを特徴量化することが有効であると考えられる．そこで，トランザクション毎にそのエンティティの収支を整数値でパターン化した上で，その整数値のシーケンスをテキストマイニングの分野で用いられるbigramとして扱い，パターンの遷移の頻度をTF-IDFにより特徴量化することを提案した．さらに，投資家は出資した後に利息を受領するというHYIPに特有な傾向があり，これを特徴量として算出する．これらの特徴量を元に，既に素性が明らかとなっているHYIPおよびnon-HYIPのアドレスを教師あり機械学習によって学習し，2クラスの分類器を生成した．図1にHYIPの割合を変化させた場合の分類器毎の識別精度の比較を示す．TPR（True Positive RateはHYIPのうち正しくHYIPとして識別した割合，FPR（False Positive Rate）はnon-HYIPを誤ってHYIPとして識別した割合を表す．さらにWith ACは，単一

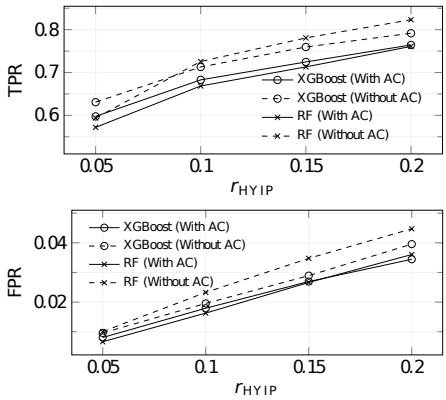

図 1. HYIPの割合を変化させた場合のTPR（True Positive Rate)とFPR (False Positive Rate).

のHYIPのBitcoinアドレスだけでなく，匿名化手法を用いて同一のユーザが保有する全てのBitcoinアドレスを用いて特徴量を計算した場合である．この図からわかるように，分類器によらず，HYIPの割合が大きい程，TPRは改善するが，同時にFPRが劣化することがわかった．本成果は，著名な国際会議IEEE GLOBECOM 2017に採録され，さらに仮想通貨に関する国際会議Decentralized 2017にてパネルディスカッションに招待された．(学会発表①，学会発表②)

(3) 上記の結果では，HYIPのBitcoinアドレスを48個収集したが，より多くの数を収集する必要があった．そこで，これまで手作業で行われていたHYIPに関連したBitcoinアドレスの収集をWebページ上から自動的にスクレイピングする手法を提案した．具体的には，多くのHYIPが投資を募るためにBitcoinに関する最もポピュラな掲示板であるBitcoinTalkにおいて宣伝を行うことに着目し，該当するトピック内からHYIPに関連したBitcoinアドレスを収集する．HYIPを宣伝するトピック内には，図2に示すように，実際に投資もしくは還元の証拠を示すためにBitcoinトランザクションのIDが投稿されることが多いため，全投稿に含まれるBitcoinトランザクションを抽出した上で，HYIPを運営する側のBitcoinアドレスを抽出する．トランザクションには送

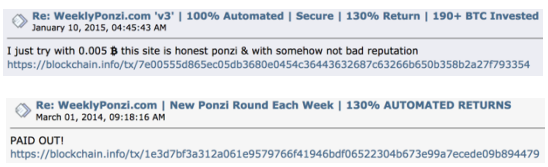

図 2 BitcoinTalkにおいてBitcoinトランザクションのIDの付加された実際の投資もしくは還元の証拠の例.

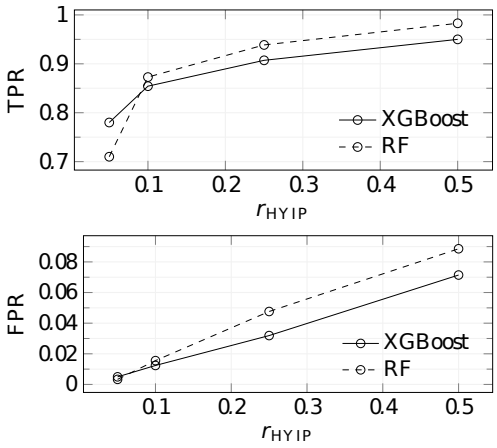

図 3 HYIP の割合を変化させた場合の TPR および FPR.

金側と受領側の双方の Bitcoin アドレスが含まれるため，抽出されたトランザクションが HYIP の投資もしくは還元のいずれの証拠であるかを判別する必要がある．提案方式では，トランザクション ID と共に投稿される文章に対して単純なテキストマイニングを施すことでこれを解決する．実際に BitcoinTalk 内の HYIP に関するトピック群に対して提案方式を適用し，計 918 個の Bitcoin アドレスを収集した．この成果は，既存研究では数十程度を大きく上回り，Bitcoin における HYIP の解明に大きく貢献した．

(4) 上記で収集した HYIP の Bitcoin アドレスをどの程度の精度で識別できるかを明らかにした．また，精度向上のため，取引の頻度，一度送った Bitcoin アドレスからどの程度 Bitcoin が返金されるかといった HYIP の特徴の出易い特徴量を計 348 種類算出した．BitcoinTalk (https://blockchain.info/tags) よりギャンブル，寄付，取引所といった HYIP 以外の目的に使用される計 1523 の Bitcoin アドレスと Web スクレイピングにより収集した計 918 の HYIP の Bitcoin アドレスをサンプリングし，与えられた Bitcoin アドレスが HYIP のものであるかを二値分類する．TPR およびを評価項目とし，10-fold 交差検定により評価した．図 3 に HYIP の Bitcoin アドレスがデータセットに存在する割合に対する TPR および FPR を示す．この図より，いずれの分類器においても，HYIP の割合が大きくなるにつれ，TPR，FPR ともに増加し，r_{HYIP} = 0.25 の時，RF を用いた場合 TPR = 0.94，FPR = 0.047，XGBoost を用いた場合 TPR = 0.91，FPR = 0.031 となった．さらに r_{HYIP} で平均を取ると，RF を用いた場合 TPR = 0.88，FPR = 0.038，XGBoost を用いた場合 TPR = 0.87，FPR = 0.031 となった．ある程度の匿名性があると考えられている Bitcoin において，88%という高い精度でその素性を識別できることが示せたことは，仮想通貨を用いた犯罪のフォレンジックスの観点から極めて重要であると言える．

(5) 最後に上記の，平成 28 年度に購入したサーバで収集したブロックチェーンを処理し，R 言語およびウェブアプリケーション開発パッケージである RShiny (https://shiny.rstudio.com) を使用したウェブアプリケーションとして稼動させた．図 4 にそのキャプチャを載せる．本アプリケーションは，Bitcoin アドレスを入力すると，その Bitcoin アドレスの取引履歴を解析し，取引に関する特徴量を可視化し，さらにその Bitcoin アドレスが HYIP の運営に用いられたかの推定結果を示すことができる．本結果は技術展示会（慶應テクノモール 2017）にて公開した．

図 4 HYIP 検知ウェブアプリケーション．

(6) さらに，上記研究を進めるにあたり，ブロックチェーンの所有権が保存されるという性質が他のアプリケーションにも応用可能であることに気付いた．そこで，ブロックチェーンを用いた商品所有権管理システム（POMS: Product Ownership Management System）を提案した．本システムにより，商品の所有権を主張することを防止し，本物の所有者のみがその商品を所持することを保証する．この目的のために，分散管理型の仮想通貨である Bitcoin の考えを用いた．すなわち Bitcoin では残高の所有権証明を行うのに対し，提案システムでは商品の所有権証明を行う．システムが正しく稼動するために，いくつかの要件が存在する．例えば，正規の製造者のみが商品の所有権最初の所有者であり，かつその商品は自社のものでなければならない．これらの要件を考慮に入れ，ポストチェーンにおいても偽物検知が可能なように，ブロックチェーンを用いた商品所有権管理システムを提案する．まず初めにシステム全体に必要な要件をまとめた後に，サプライチェーンにおける各パーティならびに消費者が RFID の付加された商品の所有

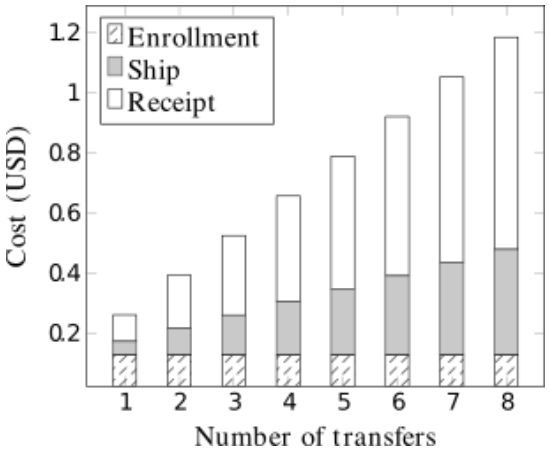

図 5 所有権の移行回数と製造者の負担するコスト.

権の移行および証明を可能とするプロトコルを示した．この提案プロトコルに基づき，ブロックチェーンを利用した分散アプリケーションプラットフォームである Ethereum 上に PoC (Proof-of-Concept)となるシステムを構築した．特性評価により，提案システムを用いて 1 商品あたりの所有権を移行・管理するコストを評価した．図 5 に所有権の移行回数と製造者が負担するコストを示す．この図からわかるように，所有権の移行回数が増える程，管理コストが大きくなることがわかる．さらに，所有権を 6 回程度移行する場合においては，1 米ドル以下で所有権を管理できることがわかった．これにより，高級嗜好品などの所有権管理を極めて低価格で管理できる可能性を示した．本研究成果は国際的に著名な論文誌である IEEE Access に採録された．(雑誌論文①)

(7) さらに，上記の所有権管理システムを R 言語およびウェブアプリケーション開発パッケージである RShiny (https://shiny.rstudio.com)を使用したウェブアプリケーションとして稼動させた．図 6 にそのキャプチャを載せる．本アプリケーションは，商品の ID (EPC: Electronic Product Code)を入力すると，その所有権の移行履歴を可視化できる．本結果は技術展示会（慶應テクノモール 2017)にて公開した．

図 6 ブロックチェーンを用いた所有権の移行履歴を可視化するアプリケーション.

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

①. Kentaroh Toyoda, P. Takis Mathiopoulos, Iwao Sasase, and Tomoaki Ohtsuki,``A Novel Blockchain-Based Product Ownership Management System (POMS) for Anti-Counterfeits in The Post Supply Chain,″ vol. 5, no. 1, pp. 17465-17477, DOI: 10.1109/ACCESS.2017.2720760, 2017. (査読有り)

〔学会発表〕（計 6 件）

①. Kentaroh Toyoda, Tomoaki Ohtsuki and P. Takis Mathiopoulos, ``Identification of High Yielding Investment Programs in Bitcoin via Transactions Pattern Analysis,'' in IEEE Global Communications Conference (GLOBECOM), Singapore, Dec. 2017.

②. Kentaroh Toyoda, ``Panel Discussion I: The Rise of Digital Cash: Blockchain Regulation, Compliance and Law Enforcement,'' in Decentralized, Limassol, Cyprus, Nov. 2017.

③. 豊田健太郎, P.T. Mathiopoulos,大槻知明, ``Web スクレイピングによって収集した高収益投資プログラムの運用に用いられた Bitcoin アドレスの識別,'' 電子情報通信学会 ソサイエティ大会, 2017 年 9 月.

④. 豊田健太郎, 大槻知明, P.T. Mathiopoulos, ``Web スクレイピングによる高収益投資プログラムに用いられた Bitcoin アドレスの収集およびその解析,'' 電子情報通信学会, 情報通信システムセキュリティ研究会 (CS), 福江, 長崎県, 2017 年 7 月.

⑤. 豊田健太郎, 大槻知明, P.T. Mathiopoulos, ``Bitcoin における高収益投資プログラムの取引抽出を目的とした特徴量抽出手法の検討,'' 電子情報通信学会, 情報通信システムセキュリティ研究会 (ICSS), 長崎, 長崎県, 2017 年 3 月

⑥. 豊田健太郎, P.T. Mathiopoulos, 笹瀬巌, 大槻知明, ``偽物商品流通防止に向けたブロックチェーンを利用した商品所有権管理システム,'' 情報処理学会, コンピュータセキュリティシンポジウム (CSS 2016), 秋田, 秋田県, 2016 年 10 月.

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等: なし

６．研究組織
(1)研究代表者
豊田 健太郎 (TOYODA, Kentaroh)
慶應義塾大学理工学部・助教
研究者番号：60723476

(2)研究分担者
(　　)

研究者番号：

(3)連携研究者
(　　)

研究者番号：

(4)研究協力者
(　　)