Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.31, NO.4, Aug. 2021 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2021.31.4.715

다크웹 오프체인 데이터를 이용한 다계층 비트코인 클러스터링 기법*

이 진 희,[1†] 김 민 재,[1] 허 준 범[2‡]
[1,2]고려대학교 (대학원생, 교수)

Multi-Layer Bitcoin Clustering through Off-Chain Data of Darkweb*

Jin-hee Lee,[1†] Min-jae Kim,[1] Junbeom Hur[2‡]
[1,2]Korea University (Graduate student, Professor)

요 약

비트코인은 분산되고 투명하며 강력한 암호화를 통해 데이터 수정이 불가능한 암호화폐 중 하나이다. 그러나 익명성으로 인해 다크웹 등에서 불법 거래를 위한 지불 수단으로 사용되기도 한다. 이 문제를 해결하기 위해 비트코인 트랜잭션의 특성을 기반으로 하는 클러스터링 기법이 제안되었으나 기존 휴리스틱 기법에서는 여전히 클러스터링 되지 못하고 있는 경우가 존재한다. 이러한 거짓 부정을 줄이기 위해 비트코인 트랜잭션의 특성뿐만 아니라 오프체인 데이터를 이용한 휴리스틱을 제안한다. 우리는 오프체인 데이터를 수집하고 활용하기 위해 Silk Road 4의 리뷰 데이터를 분석하여 리뷰 데이터의 31.68%가 실제 비트코인 트랜잭션으로 매치시킬 수 있음을 발견했고 수집된 데이터에 대해 91.7%까지 거짓 부정을 줄일 수 있었다.

ABSTRACT

Bitcoin is one of the cryptocurrencies, which is decentralized and transparent. However, due to its anonymity, it is currently being used for the purpose of transferring funds for illegal transactions in darknet markets. To solve this problem, clustering heuristic based on the characteristics of a Bitcoin transaction has been proposed. However, we found that the previous heuristis suffer from high false negative rates. In this study, we propose a novel heuristic for bitcoin clustering using off-chain data. Specifically, we collected and analyzed user review data from Silk Road 4 as off-chain data. As a result, 31.68% of the review data matched the actual Bitcoin transaction, and false negatives were reduced by 91.7% in the proposed method.

Keywords: Bitcoin clustering, De-anonymization, Darkweb

I. 서 론

2008년 비트코인의[1] 출현 이후, 이더리움[2], 모네로[3], 제트캐시[4] 등 수많은 블록체인 기반 암호화폐가 개발되었다. 블록체인은 분산되고 투명하며 강력한 암호화를 통해 데이터 수정이 불가능하기에 비트코인을 통한 거래는 익명성을 제공하는 거래 자산으로 사용가능하다. 그러나 Foley에[5] 따

Received(06. 16. 2021), Modified(07. 07. 2021), Accepted(07. 14. 2021)
* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2019-0-01697, 블록체인 플랫폼 보안취약점 자동분석 기술 개발)(No.2019-0-00533, 컴퓨터 프로세서의 구조적 보안 취약점 검증 및 공격 탐지대응)과 ICT명품인재양성 사업의 연구결과로 수행되었음(IITP-2021-0-01819). 또한 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2021R1A6A1A13044830).
† 주저자, akrso93@korea.ac.kr
‡ 교신저자, jbhur@korea.ac.kr(Corresponding author)

르면 익명성을 악용해 비트코인 거래량의 25%가 불법 거래에 사용되고 있다고 보고되었다. 비트코인을 사용한 불법거래를 추적하려면 비트코인 전체 네트워크에서 자금의 흐름을 분석할 수 있어야 한다. 따라서, 비트코인 트랜잭션을 정확하게 추적할 수 있는 기술이 필요한 상황이다.

불법거래를 이용하려는 사람들은 다크웹을 이용한다. 다크웹은 일반적으로 사용되는 서페이스 웹보다 더 강력한 보안이 적용된 웹 페이지이다. 실제로 서페이스 웹은 전체 웹의 단 4% 정도를 차지하며, 나머지 96%는 딥웹이다. 그 중에서도 다크웹은 일반적인 웹 브라우저가 아닌 Tor 브라우저를 통해 접속 가능하며, 어니언 라우팅 기술을 통해 자신의 IP주소와 서버 이름을 암호화한다. 이는 다크웹에서 어떠한 행동을 하더라도 IP주소와 서버 이름이 노출되지 않으며 Silk Road 4와 같은 다크넷 마켓에서 다양한 불법 거래가 이루어지는 결과를 가져온다. 실제로 사람들은 Tor 브라우저를 통해 Silk Road 4와[8] 같은 불법 마켓에 접속하여 다양한 불법 활동에 참여하고 있다.

이러한 불법 거래를 추적하고 실제 소유자를 식별하기 위해 비트코인 클러스터링 휴리스틱[9][10][11]이 제안되었다. 클러스터링이란 블록체인 네트워크에 기록된 비트코인 트랜잭션의 특성을 분석하여 동일한 지갑에 속한 비트코인 주소들을 하나로 묶는 작업을 말한다. 그러나 비트코인 트랜잭션을 통해 얻을 수 있는 정보는 제한적이며 실제로 같은 지갑에 포함되어 있지만, 정보의 부족으로 인해 하나의 지갑으로 클러스터링하지 못하는 거짓 부정이 존재한다.

이러한 한계를 극복하기 위해 이 논문에서는 비트코인 트랜잭션 데이터인 온-체인 데이터 뿐만 아니라 웹 페이지에서 얻을 수 있는 오프-체인 데이터를 이용하고자 한다. 이를 위해 2019년 11월부터 2020년 9월까지 실제 다크넷 마켓에 접속하여 활용 가능한 데이터를 직접 수집했으며 거래 패턴을 분석했다. 다크넷 마켓은 마약과 같은 불법 거래에서 구매자와 판매자 사이에 거래의 흐름을 찾을 수 없도록 중간에서 모든 거래에 관여하고 있음을 발견했다. 중간에서 모든 거래에 관여하는 다크넷 마켓의 주소들을 식별함으로서 클러스터링에 사용 가능한 오프-체인 데이터를 발견했다. 또한, 수집된 데이터에서 비트코인 값과 거래 날짜 및 시간이 포함된 리뷰데이터를 발견했으며 비트코인 트랜잭션과 연결 지을 수 있음을 알 수 있었다. 결과적으로, 온-체인 데이터와 오프-체인 데이터를 이용해 비트코인 클러스터링 정확도를 향상시킬 수 있는 다계층 휴리스틱을 제안한다.

연구 결과를 평가하기 위해 다계층 휴리스틱 알고리즘을 구현하고 Silk Road 4에서 수집한 데이터에 한하여 정확도 및 거짓 부정 감소율을 측정한다. 실험 결과에 따르면 Silk Road 4의 불법 거래 정보 중 31.68%가 실제 비트코인 거래와 일치한다. 또한, 발견된 비트코인 주소는 동일한 클러스터에 중복되는 경우를 제외하고 133개의 클러스터를 형성한다. 133개의 클러스터에서 거짓 긍정을 유발할 수 있는 11개의 클러스터를 제외한 122개의 클러스터를 Silk Road 4 클러스터로 결합한다. 이는 수집된 데이터셋에서 91.7%의 거짓 부정을 제거하는 효과를 보인다.

II. 배경 지식

2.1 다크넷 마켓

다크넷 마켓은 마약, 위조 지폐, 개인정보, 해킹 서비스 및 기타 불법 물품과 관련된 거래를 위해 설계된 다크웹 페이지다. 아마존[12], 알리익스프레스[13]와 같은 온라인 거래 마켓과 마찬가지로 판매자는 다크넷 마켓에 자신이 판매하려는 불법 상품을 게시하고 구매자는 다크넷 마켓에 게시된 불법 상품에 대해 구매 요청을 제시할 수 있다. 불법 상품에 대한 지불 수단으로 비트코인을 사용하게 되며 다크넷 마켓에서는 비트코인이 구매자로부터 판매자에게 도달하기까지 중간에서 자금을 관리한다.

다크넷 마켓은 회원가입 및 거래 절차에서 판매자와 구매자의 익명성을 보호하기 위해 여러 매커니즘을 사용한다. 먼저 회원가입 단계에서 이메일, 전화번호, 거주지 등의 개인정보를 입력하지 않으며, 간단하게 CAPTCHA 인증 후, 아이디, 비밀번호, PIN 코드만 입력하면 완료된다. 그 이후, 자신의 PGP키를 설정할 수 있으며, 다른 유저와의 통신에서 PGP키를 사용한 암호 메시지를 통해 개인정보의 노출을 방지한다. Aldridge[14]의 조사에 따르면, 판매자가 PGP키를 지원하게 되면서 불법 거래 시장의 탄력성이 증가한다고 했다. 다음으로 거래 절차에서, 다크넷 마켓은 구매자와 판매자의 거래 흐름을 알 수 없도록 일회성 비트코인을 제공한다.

때로는 익명성을 강화하여 비트코인 추적을 어렵

게 하는 Cryptmixier[15], BitMix[16] 및 TumbleBit[17]과 같은 코인 믹싱 서비스를 사용한다. 믹싱 서비스는 비트코인의 추적을 어렵게 하도록 입력된 코인을 다수의 코인으로 여러 차례에 걸쳐 나누는 서비스이다.

또한, 다크넷 마켓은 에스크로 방식의 거래를 사용한다. 에스크로 스타일 거래는 모든 거래를 마켓이 중간에서 관리하므로 구매자와 판매자 사이의 직접적인 거래가 연결되지 않는다. 에스크로 방식 거래는 Fig.1과 같은 순서이다. 전체 과정에서 가장 먼저 판매자는 다크넷 마켓에 불법 품목을 게시한다. 이렇게 게시된 불법 품목은 다크넷 마켓을 이용하는 모든 유저가 볼 수 있다. 구매하려는 유저가 원하는 품목을 구매 요청하면, 마켓에서는 구매자에게 일회용으로 발급된 비트코인 주소와 거래에 지불해야할 비트코인 금액을 표시한다. 구매자가 해당 비트코인 주소로 정확한 금액을 입금하면 판매자는 구매자에게 물품을 배송하게 된다. 그 이후, 구매자가 물품을 정상적으로 수령하게 되면 구매자가 확인 후 거래가 종료된다. 마지막으로, 거래가 종료된 후, 14일 이내에 거래에 대한 리뷰를 남길 수 있으며 구매자 ID, 거래 품목, 비트코인 금액, 배송 완료 시간이 기록된다. 이는 판매자의 신뢰를 평가하는 지표로 이용될 수 있다. 만약 기간 내에 입력하지 않은 경우, No Feedback이라는 문구와 함께 자동으로 등록된다.

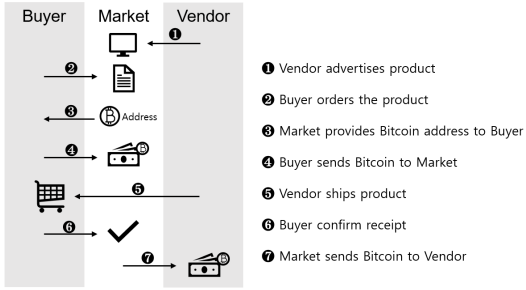


Fig. 1. Escrow Trading System

2.2 비트코인 주소

비트코인 주소는 공개 키 혹은 공개 키를 나열한 스크립트를 해시한 뒤 인코딩[18]하여 생성된다. 해시 대상(공개키 혹은 스크립트) 및 인코딩 방법[19][20](Base58 또는 Bech32)에 따라 비트코인 주소 형식은 PublicKeyHash, ScriptHash, WitnessPublicKeyHash, WitnessScriptHash 4가지로 구분된다.

PublicKeyHash는 SHA-256 및 RIPEMD-160을 연속으로 사용하여 공개 키를 해싱 한 후 Base58 인코딩으로 생성 된 주소이다. 비트코인에서 가장 일반적인 주소 유형이며 '1'로 시작하는 특성이 있다.

ScriptHash[19]는 SHA-256, RIPEMD-160을 통해 여러 공개 키를 나열하는 스크립트를 해싱하고 Base58로 인코딩하여 생성 된 비트 코인 주소이다. m-of-n 다중 서명 트랜잭션[20]을 지원하므로 소유자를 인증하려면 m개의 서명과 n개의 키 쌍이 필요하다. 모든 ScriptHash주소는 '3'으로 시작한다.

WitnessPublicKeyHash[21][23]는 SHA-256 및 RIPEMD-160을 사용하여 공개 키를 해싱 한 후 Bech32 인코딩에 의해 생성 된 주소이다. 이 주소 유형은 'bc1'로 시작하는 특징이 있다.

WitnessScriptHash[21][23]은 SHA-256, RIPEMD-160을 통해 여러 공개 키를 나열하는 스크립트를 해싱하고 Bech32에서 인코딩하여 생성 된 비트코인 주소이다. 이 주소 유형도 'bc1'로 시작하는 특징이 있다.

거래에 사용 된 주소 유형에 따라 비트 코인 트랜잭션은 다음 두 가지 유형 중 하나로 분류된다. 하나는 Legacy 트랜잭션이고 다른 하나는 SegWit 트랜잭션[21][22]이다. Legacy 트랜잭션에는 서명 정보가 포함되어 있으므로 트랜잭션 크기가 크다. 반면 SegWit 트랜잭션에서는 서명 정보가 거래와 분리되어 거래 규모가 줄어든다. 트랜잭션 크기는 트랜잭션 생성시 발생하는 수수료와 연관되어 있으므로 트랜잭션 크기가 줄어들면 트랜잭션 수수료가 줄어드는 장점이 존재한다.

Legacy 트랜잭션의 경우 PublicKeyHash 및 ScriptHash 주소가 출력 주소로 사용된다. PublicKeyHash 및 ScriptHash 주소로 생성된 트랜잭션을 각각 P2PKH (Pay to Public Key Hash) 및 P2SH (Pay to Script Hash)라고한다. 반면에, WitnessPublicKeyHash와 WitnessScriptHash는 SegWit 트랜잭션에 사용되며, 각 유형으로 생성된 트랜잭션을 각각 P2WPKH (Pay to Witness Public Key Hash) 및 P2WSH (Pay to Witness Script Hash)라고한다. 또한, ScriptHash 주소에 WitnessPublicKeyHash 및 Witnes

sScriptHash 주소를 중첩하여 SegWit 트랜잭션에 사용할 수 있다. 각 케이스의 거래를 P2SH에서는 P2WPKH, P2SH에서는 P2WSH라고 한다.

2.3 비트코인 트랜잭션과 클러스터링

2.3.1 비트코인 트랜잭션

비트코인 트랜잭션은 비트코인 주소 사이에 비트코인 값을 주고받는 작업을 말한다. 또한, 비트코인 트랜잭션에 대한 모든 기록은 블록체인에 공개적으로 등록된다.

비트코인 주소는 공개키의 해시값으로 표현되는 무작위 문자열이며 은행의 계좌번호와 같은 역할을 한다. 비트코인 주소를 생성하기 위한 공개키의 쌍인 개인키는 트랜잭션을 생성할 때 서명에 사용되며 주소의 소유권을 증명한다. 비트코인 지갑은 저장된 공개키, 개인키 쌍과 함께 공개키에서 파생된 주소들을 포함한다.

비트코인 트랜잭션은 하나 이상의 입력 주소와 출력 주소로 구성된다. 사용자가 비트코인을 보낼 때, 자신의 지갑에 포함된 비트코인 주소에서 입력 주소를 선택할 수 있으며, 출력 값으로 출력 주소와 비트코인 값을 선택하여 트랜잭션을 생성할 수 있다. 트랜잭션은 다수의 출력을 포함할 수 있으며 사용자는 하나의 트랜잭션으로 다수의 사용자에게 비트코인을 보낼 수 있다.

Fig.2은 비트코인 트랜잭션의 예시이다. 각각의 트랜잭션은 해시값을 가지고 있으며 트랜잭션의 ID와 같은 역할을 한다. 게다가, 트랜잭션에는 입력 주소와 출력 주소가 포함된다. 좌측에 기록된 문자열이 입력 주소이며, 우측에 기록된 문자열이 출력 주소이다.

비트코인의 초창기부터 지금까지의 모든 트랜잭션은 공개된 장부이며 Blockchain.com과 같은 블록체인 검색 페이지에서 확인할 수 있다. 따라서 비트코인 트랜잭션의 흐름은 네트워크의 모든 비트코인 사용자에 의해 추적될 수 있다. 더 나아가 동일한 지갑에 속한 비트코인 주소들을 찾아낼 수 있다면 비트코인 거래를 추적하는 것도 가능하다. 이는 비트코인을 이용한 불법 거래를 추적하기에 유용할 것이다. 따라서 비트코인 주소를 클러스터링 하는 것은 실제 비트코인 소유자의 거래를 분석하기 위해 가장 중요한 단계이다. 그러나 비트코인 트랜잭션은 흐름을 알 수 없도록 변경 주소와 같은 매커니즘이 적용되어 있으며 믹싱 서비스도 이용하기 때문에 비트코인 클러스터링은 어려운 문제로 남아있다.

2.3.2 비트코인 클러스터링

비트코인 클러스터링은 동일한 지갑에 속한 주소를 찾는 작업이다. 일반적으로 클러스터링은 클러스터링의 대상의 시드 주소로 시작된다. 시드 주소는 지갑에 포함된 하나 이상의 주소이며 환전소나 믹싱 서비스와 같은 업체에서 직접 공개한 경우와 유저들이 거래를 통해 발견하고 식별된 경우를 포함한다. 이러한 시드 주소로 시작하여 동일한 지갑에 속한 다른 주소가 클러스터링 알고리즘을 통해 함께 수집되어 점점 더 큰 클러스터를 생성한다. 시드 주소가 알려진 정보를 통해 라벨을 지정할 수 있는 경우 전체 클러스터는 Fig.3과 같이 동일한 이름으로 레이블될 수 있다. 정확한 클러스터링이 된다면 클러스터 간 트랜잭션 흐름을 보다 투명하게 분석하는 것에 도움이 된다.

그러나 현재까지 연구된 비트코인 클러스터링 기법들이 항상 정확한 결과를 생성하지 않는다. 대부분 ground-truth 정보가 부족하기 때문이다. 현재까지 비트코인 트랜잭션 데이터만 활용한 클러스터링 휴리스틱이 제안되었다. 지금까지 연구된 기존 비트

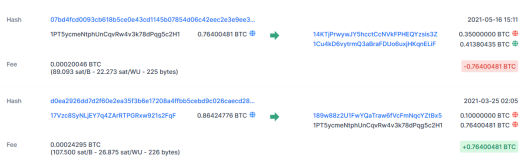


Fig. 2. Bitcoin Transaction

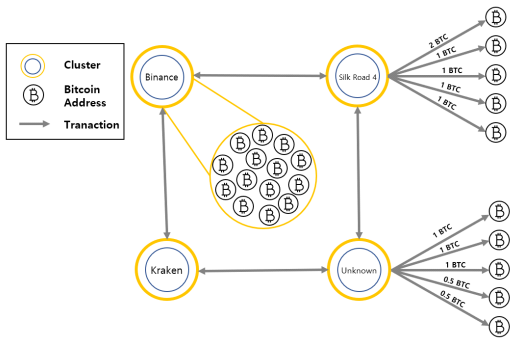


Fig. 3. Bitcoin Clustering Overview

코인 클러스터링 휴리스틱은 다음과 같다.

Multi-input heuristic[9][11] : 한 트랜잭션에서 두 개 이상의 입력 주소가 하나의 출력으로 비트코인을 보내는 경우 입력 주소는 모두 동일한 지갑으로 클러스터링 가능하다.

초기 비트코인 시스템에서는 출력의 수와 관계없이 동일한 사용자에 의해 제어되었지만, 여러 사용자가 하나의 트랜잭션에 참여할 수 있는 Coinjoin[26]의 출현으로 출력이 하나인 경우에만 이 휴리스틱이 가능하게 수정되었다.

Change address heuristic[9][11] : 비트코인 트랜잭션을 생성하면서 Shadow address라고 불리는 일회성 변경 주소가 출력 주소 중 하나로 생성되고 입력 주소와 동일한 사용자가 소유한다. Shadow address는 전송하려는 비트코인 값에서 남는 값을 자신에게 전송하는 주소이다. 그러나, 출력 주소 중 어떠한 주소가 Shadow address이며 동일한 지갑에 속할지 구별하는 것은 어려우며 다음 설명할 2가지 휴리스틱은 이를 식별하기 위한 휴리스틱이다.

Consumer heuristic[10] : 소비자 지갑에서 이루어진 트랜잭션은 항상 2개 이하의 출력 주소를 가진다. 따라서 1개 혹은 0개의 변경 주소를 포함한다. 출력 주소 중 라벨이 기록된 특정 클러스터의 주소가 포함 된 경우 나머지 출력 주소가 변경 주소(즉, Shadow address)가 된다.

Optimal change heuristic[10] : 지갑 소프트웨어는 트랜잭션 생성 시 불필요한 입력 주소를 선택하지 않는다는 가정을 기반으로 한다. 기본적으로 트랜잭션이 유효하려면 입력 값의 합계가 출력 값의 합계보다 커야한다. 따라서 최소 입력 값이 변경 주소의 값보다 작으면 생략 가능한 주소가 선택된 경우이므로 변경 주소의 값은 입력 주소의 값의 최소값보다 작아야한다. 트랜잭션의 출력이 최소 입력보다 작은 경우 변경 주소일 가능성이 크다.

Consumer heuristic 및 *Optimal change heuristic*은 클러스터링을 위해 *Change address heuristic*을 보완하는 아이디어다. 게다가, 지갑 소프트웨어에서 자동적으로 선택되는 경우만을 포함한다. 불행하게도, 다크웹 사용자는 추적을 피하기 위해 일반적인 패턴을 사용하지 않는 경향이 있으므로 다크웹 환경에서는 적용 가능성이 적다.

2.4 기존 휴리스틱의 문제점

이전 섹션에서 서술한 4가지 휴리스틱은 온-체인 데이터(즉, 비트코인 트랜잭션 데이터)를 통해 비트코인 전체 네트워크에 적용 가능한 휴리스틱이다. 그러나, 온-체인 데이터만을 사용한 비트코인 클러스터링은 실제로는 동일한 지갑에 포함되어 있지만 하나의 클러스터로 묶지 못하는 경우가 존재한다. 온-체인 데이터 이외에 오프-체인 데이터(예, 웹 데이터)를 이용한다면 비트코인 주소들 사이에 연결점을 가능성이 존재한다. 따라서, 오프-체인 데이터를 수집하고 분석하여 비트코인 클러스터링 정확도를 향상시킬 수 있는 연구가 필요하다.

2.5 클러스터링 효과

비트코인 클러스터링의 효과는 불법 거래를 추적하고 거래 규모를 추정하는데 이용될 수 있다. Fig. 4에서 볼 수 있듯이, 불법 거래에 사용된 비트코인 주소가 수집되면, Chainalysis, BlockSci와 같은 블록체인 분석 툴을 통해 범죄 규모 및 실제 소유주를 추적할 수 있다. 이 과정에서 블록체인 분석 툴은 현재까지 연구된 비트코인 클러스터링 휴리스틱을 비트코인 네트워크에 적용시켜 수많은 클러스터를 생성하고 저장한다. 정확한 클러스터링을 통해 더 정확한 규모 추정 및 추적이 필요한 만큼 비트코인 클러스터링 휴리스틱은 더 많이 연구되어야 한다.

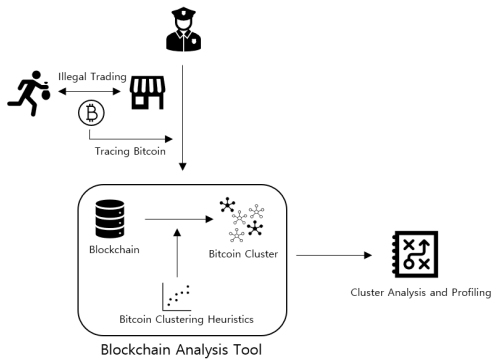


Fig. 4. Bitcoin Clustering Flowchart

III. 다크넷 마켓 거래 분석

오프-체인 데이터를 수집하고 분석하기 위해 불법

거래가 활발하게 이루어지며 비트코인을 사용한 거래가 활성화된 다크넷 마켓에 대해 심층 분석했다. 그 중에서도 거래가 가장 활발한 Silk Road 3.1과 Silk Road 4에서 데이터를 수집했으며, 비트코인 트랜잭션과 연결 지을 수 있는 결과를 얻었다.

3.1 Silk Road 3.1

지난 10년 동안 다양한 연구[27-33]에서 Silk Road 1과 2의 거래 특성을 분석했다. 또한, Silk Road 1과 2의 클러스터는 식별되었으며 누구에게나 공개되어 있다. 그러나 Silk Road 3.1과 4와 같은 최신 다크넷 마켓을 분석하고 거래 패턴을 조사하려는 시도는 없었다. 따라서 최신 다크넷 마켓의 비트코인 클러스터와 거래 흐름을 알 수 없는 현황이다.

먼저, Silk Road 3.1에서는 에스크로 방식의 거래를 이용한다. Fig. 1의 3단계에서는 거래를 위한 비트코인 주소를 반복해서 제공하고 있었다. 이 주소는 Silk Road 3.1에서 제공한 비트코인 주소였기 때문에 이 주소들을 모두 수집하고 기존 휴리스틱을 통해 클러스터링되는지 확인할 필요가 있었다. Fig. 2의 2단계와 3단계를 반복하여 다수의 거래를 시도하면서 비트코인 주소를 수집했다. 2019년 11월 25일부터 12월 16일까지 직접 수작업으로 절차를 반복하여 75개의 비트코인 주소를 수집했으며, 중복 수집된 경우를 제거하고 결과적으로 48개의 고유 주소를 찾았다. *Multi-input heuristic*과 *Change address heuristic*을 통해 48개의 비트코인 주소에 대해 클러스터링 결과를 확인하기 위해 오픈 소스 클러스터링 툴인 BlockSci를 통해 분석했다.

그 결과, 48개 주소 중 18개가 하나의 클러스터로 그룹화되었으며, 나머지 30개도 다른 하나의 클러스터로 그룹화되었다. 두 개의 클러스터에는 기존 휴리스틱을 통해 각각 57개, 72개의 주소를 포함하고 있었다. Silk Road 3.1이 2개의 클러스터로 분할되어 있음을 발견했으며, 이는 기존 휴리스틱이 실제 환경에서 거짓 부정을 초래하고 있음을 증명한다. 동일한 웹에서 제공하는 주소라는 점을 이용해 129개의 비트코인 주소를 하나의 지갑으로 클러스터링 가능했다.

그 후, 129개의 비트코인 주소에 기록된 트랜잭션을 분석하여 다른 오프-체인 정보를 얻으려 했다. 그 결과 Silk Road 3.1의 2가지 패턴을 발견했다. 먼저, Silk Road 3.1에서 제공하는 비트코인 주소는 1-of-2 다중 서명을 지원하는 ScriptHash이다. 수집된 48개의 주소는 모두 '3'으로 시작하는 주소였으며 이러한 주소는 ScriptHash 형식의 주소이다. 두 번째로, 거래에 사용된 주소들은 구매자로부터 거래에 대한 코인을 입금 받으면 곧바로 다른 비트코인 주소로 전달하는 패턴을 가지고 있었으며, Shadow address 없이 단 1개의 출력 주소만을 가지는 트랜잭션이다.

Silk Road 3.1을 조사하던 중 2019년 12월 17일에 서버가 닫혔으며 다음 버전인 Silk Road 4는 2020년 4월 1일에 다시 열렸다. 버전이 다르지만 Silk Road 3.1에서 발견된 주소들과 Silk Road 4 사이에 연관이 있다는 증거를 관찰했다. 따라서 Silk Road 4의 거래를 조사하기로 했다.

3.2 Silk Road 4

Silk Road 4에서 비트코인 주소를 수집하면서 에스크로 시스템에서 비트코인을 제공하는 방식이 변경되었음을 발견했다. Silk Road 3.1에서는 129개의 비트코인 주소를 반복적으로 거래에 사용했다면, Silk Road 4에서는 한 번에 거래에 일회성 비트코인 주소를 새로 생성하여 제공하는 방식으로 변경되었다. 일회성 비트코인 주소를 수집하면서 '3'으로 시작되는 비트코인 주소만 제공되고 있음을 발견했다. 따라서 Silk Road 4에서 ScriptHash 기반 주소를 사용하고 있는 것을 알아냈다. 그러나, Silk Road 4의 비트코인 트랜잭션을 얻기 위해서는 다른 접근이 필요하다.

Silk Road 4의 비트코인 트랜잭션을 얻기 위해 웹 페이지의 구조와 내용에 대한 심층 분석을 수행했다. 그 중 Silk Road 4의 리뷰데이터가 비트코인 트랜잭션과 유사한 형태임을 발견했다. 리뷰데이터에는 사용자 ID, 품목명, 비트코인 값, 배송 날짜에 대한 정보가 포함되어 있다. 비트 코인 값과 배송 날짜는 비트코인 트랜잭션의 비트코인 값과 트랜잭션 생성시간과 연관될 수 있다. 따라서 2020년 4월 10일부터 9월 30일까지 총 606개의 Silk Road 4 리뷰데이터를 수집했다.

리뷰데이터에 기록된 비트코인 값은 불법 품목 구매자가 실제로 Silk Road 4에 지불한 값이다. 따라서 블록체인에는 출력과 동일한 가치를 가진 비트코인 트랜잭션이 존재해야 한다. 비트코인 값은 소수 8번째 자리까지 표현되며 날짜 범위에 따라 같은 값의

거래를 찾을 가능성이 높아진다. 따라서 정확한 범위 내에서 동일한 비트코인 값을 찾아내기 위해 리뷰 데이터에 기록된 정보를 활용하여 트랜잭션의 검색 범위를 좁혔다. 구체적으로 Silk Road 4의 판매글에 따르면 결제 후 품목이 구매자에게 도달하기 까지 2~4일이 소요되며 에스크로 시스템은 14일 이내에 판매자에게 비트코인을 전달한다. Silk Road 4의 리뷰데이터에서 배송 날짜는 "*n* days ago"로 표현된다. 따라서 Silk Road 4는 "*n* days ago"에 해당하는 날짜부터 14일 이내에 판매자에게 비트코인을 보내기 위한 트랜잭션을 생성했다.

3.3 Silk Road 3.1과 Silk Road 4 비교

Silk Road 4의 리뷰데이터와 비트코인 트랜잭션을 매칭시키기에 앞서, Silk Road 3.1에서 발견된 패턴을 적용 가능한지 비교 분석한다. 만약 같은 패턴이 발견된다면 정확도를 향상하는데 도움이 될 것이다.

ScriptHash 주소 : Silk Road 4 에스크로에서 제공하는 비트코인 주소가 Silk Road 3.1과 마찬가지로 ScriptHash인지 분석했다. Fig.2의 2단계와 3단계를 반복하며 제공된 비트코인 주소를 직접 수집했다. 결과적으로, Silk Road 4에서도 마찬가지로 '3'으로 시작하는 주소들만 제공했으므로 ScriptHash 주소를 사용하고 있음을 증명한다.

1-of-2 다중 서명 : Silk Road 4에서 여전히 1-of-2 다중 서명을 지원하는 ScriptHash를 사용하는지 조사했다. 비트코인 주소가 다중 서명을 사용하는지 확인하기 위해 해당 비트코인 주소를 포함하는 트랜잭션이 필요하다. 그러나 Silk Road 4에서 제공한 비트코인 주소는 트랜잭션 기록이 없기 때문에 리뷰데이터와 매치시켜 찾아낸 트랜잭션을 분석했다. 리뷰데이터에 기록된 배송일로부터 14일 이내에 발생한 트랜잭션이 정확하게 1개인 경우 리뷰데이터와 직접적으로 연관된 비트코인 트랜잭션일 확률이 높다. 실험 결과 Table 1에서 볼 수 있듯이 606개의 리뷰 데이터 중 31.68%가 정확히 하나의 트랜잭션과 일치함을 발견했다. 그중 10.48%는 1-of-2 다중 서명을 지원하는 ScriptHash를 사용하고 있고, 나머지 89.52 %는 WitnessPublicKeyHash를 중첩하는 ScriptHash를 사용하였다. 따라서 Silk Road 4는 Silk Road 3.1과 달리 다양한 서명 유형을 지원함을 발견하고, Silk Road 4 주소 식별을 위한 유일한 지표로 서명 유형을 사용하기 어렵다는 결론을 도출했다.

Table 1. Transaction analysis results

Transaction match rate	31.68%
Signature Type	1-of-2 multi-signature : 10.48%
	WitnessPublicKeyHash : 89.52%
Output Address	1-output : 80.73%
	more than 2 output : 19.27%

출력이 1개인 트랜잭션 : Silk Road 4 트랜잭션이 Silk Road 3.1과 같이 1개의 출력 주소를 가지는 트랜잭션을 나타내는지 확인했다. 분석 결과는 Table 1에 요약되어 있다. 실험 결과, 매치된 31.68% 주소와 관련된 트랜잭션 중 80.73%는 단 1개의 출력 주소 만 가지고 있음을 발견했다. 나머지 트랜잭션은 두 개 이상의 출력 주소를 가지고 있다. 따라서, Silk Road 3.1의 '1 출력 트랜잭션'에 관해서는 Silk Road 4에 적용할 수 없다.

결과적으로, 결정론적인 패턴은 ScriptHash 주소 형식뿐이다. Silk Road 3.1과 비교하여 세분화된 클러스터링에 대해 훨씬 더 제한된 조건이다.

3.4 다크넷 마켓 확장

리뷰 데이터 활용이 다른 다크넷 시장 분석에도 적용될 수 있는지 확인하기 위해 AlphaBay, Dream Market, Valhalla, Apollon, Agartha 시장의 웹 페이지를 추가로 조사했다. 그 결과, 과거에 AlphaBay, Dream Market, Valhalla 시장의 거래가 많았었지만 현재 운영되지 않는다. 따라서 현재 운영되는 다크넷 시장 인 Apollon[34] 및 Agartha[35]에 초점을 맞추고 이러한 시장에서 사용 가능한 리뷰 데이터를 분석했다. 각 시장에서 제공하는 리뷰 데이터의 비교 결과는 Table 2에 나타내었다.

기본적으로 사용자 ID, 품목명, 배송일은 다른 마

Table 2. Review data format by market

Darknet market	Silk Road 3.1	Silk Road 4	Apollon Market	Agartha Market
User ID	O	O	O	O
Item name	O	O	O	O
Price format	BTC	BTC	USD	X
Shipped date	O	O	O	O

켓에서도 동일하게 제공한다. 그러나 각 마켓에서 제공하는 화폐의 값이 다르다. 특히 BTC 값이 제공되는 Silk Road 3.1과 4와는 다르게 Apollon 마켓은 달러로 기록되며 Agartha 마켓은 가격에 대한 정보를 제공하지 않는다. 따라서 Agartha의 경우 시장의 웹 데이터만으로는 이와 관련된 숨겨진 거래를 식별하고 클러스터링하는 것이 거의 불가능하다. 반면 Apollon은 당시 환율을 정확하게 기록하여 USD 및 BTC 값을 정확하게 변환 할 수 있다. 따라서 변환 된 BTC 값을 기반으로 분석 및 클러스터링 접근 방식을 Apollon과 Silk Road 3.1 및 4에 적용 할 수 있다.

IV. 클러스터링 휴리스틱

이 섹션에서 기존 휴리스틱과 함께 사용가능한 다계층 휴리스틱을 제안한다. 다계층 휴리스틱은 블록체인의 트랜잭션 데이터뿐만 아니라 어플리케이션 데이터를 활용하여 *Multi-input heuristic*과 *Change address heuristic*과 같은 기존 클러스터링 휴리스틱을 보완하고 개선할 수 있다.

4.1 매치 주소

V는 리뷰 데이터에 기록된 비트코인 값(BTC)이고 D는 항목의 배송 날짜이다. (V,D)는 단일 리뷰 데이터 세트를 나타내고 $P(V,D)$는 리뷰 데이터 (V,D)가 게시되는 웹 페이지이다. A와 TX를 각각 비트코인 주소와 트랜잭션이다. 마지막으로 세 가지 함수를 정의한다. $ITX(A)$는 비트코인 주소 A가 입력 중 하나로 나타난 트랜잭션 TXs를 반환하는 함수이다. $OTX(A)$는 비트코인 주소 A가 출력 중 하나로 나타난 트랜잭션 TXs를 반환하는 함수이다. $Time(TX)$은 TX 트랜잭션이 비트코인 네트워크에 기록 된 날짜를 반환하는 함수이다. $|TXs|$ 는 TXs 트랜잭션 수이다.

정의 1

1) *ITX(A)*에서 A의 비트코인 값은 리뷰데이터에 기록된 비트코인 값과 동일하다.
2) 구매자가 Silk Road 4에 비트코인을 송금하는 트랜잭션을 $OTX(A)$로 정의하며 트랜잭션의 출력 주소에 A가 포함된다. $OTX(A)$는 1개 존재한다. $|OTX(A)| = 1$
3) Silk Road 4가 판매자에게 비트코인을 송금하는 트랜잭션을 $ITX(A)$로 정의하며 트랜잭션의 입력 주소에 A가 포함된다. $ITX(A)$ 는 1개 존재한다. $|ITX(A)| = 1$
3) 리뷰데이터에 기록된 배송 날짜 및 시간을 D라고 할 때, $Time(OTX(A)) \rightarrow D \rightarrow Time(X(A)) \rightarrow D+14$ 조건을 만족한다.

첫 번째 조건은 리뷰데이터에 기록된 비트코인 값이 비트코인 트랜잭션의 비트코인 값과 정확히 일치하는 경우만 검색해서 단 1개만 존재하면 매치되는 것을 설명한다. 두 번째 조건은 A가 Silk Road 4 에스크로에서 구매자에게 제공 한 비트코인 주소이며 구매자-에스크로(Silk Road 4) 거래에 대해 하나의 $ITX(A)$가 있어야 함을 의미한다. 또한 세 번째 조건은 에스크로-판매자 거래에 대한 $OTX(A)$ 1 개가 존재해야 한다. 네 번째 조건은 (1) 리뷰 데이터에 해당하는 실제 결제에 대한 구매자-에스크로 거래(특히 리뷰 데이터와 동일한 BTC 값을 포함하는 거래)가 배송 날짜 이전에 존재해야 함을 의미한다. (2) 검토 데이터에 해당하는 실제 결제에 대한 에스크로-판매자 거래는 리뷰데이터의 배송 날짜 이후 14 일 사이에 존재해야 한다.

결과적으로 A가 Silk Road 4의 리뷰 데이터와 일치하는 주소 인 경우 해당 트랜잭션에 실제로 사용되는 비트 코인 주소이다.

4.2 다계층 휴리스틱

i에 대해 C_i를 *Multi-input heuristic* 및 *Change address heuristic*을 사용하여 생성 된 비트코인 클러스터라고 한다. 그러면 다계층 휴리스틱은 다음과 같이 정의한다.

다계층 휴리스틱

C_1 및 C_2에 각각 MA_A및 MA_B가 포함되고 MA_A및 MA_B가 동일한 $P(V,D)$에서 온 경우 C_1 및 C_2는 동일한 사용자가 제어한다.

다계층 휴리스틱의 중요한 특징은 웹 페이지에서 얻은 오프 체인 데이터와 비트코인 트랜잭션 데이터를 활용해 기존 클러스터 간의 숨겨진 관계를 찾는 것이다. 기존 클러스터링 휴리스틱과 동시에 사용 가능하며 기존에 존재했던 거짓 부정을 제거하면서 자연스럽게 보완된다.

Ermilov[11]는 이미 제공된 태그 정보를 활용하는 비트코인 클러스터링 방법을 제안했다. 이런 정보들은 웹 포럼 또는 사용자 프로필을 통해 공개된다. 이러한 태그 정보가 주어지면 사용자는 알려진 클러스터 정보를 쉽게 제거하여 변경 주소를 찾을 수 있다. 반면에 이 논문에서의 비트 코인 클러스터링 알고리즘은 애플리케이션 계층의 오프-체인 데이터(특히, 다크넷 마켓의 리뷰 데이터)를 활용하며, 기존 클러스터와의 연관성은 이전에 공개 된 적이 없다. 안타깝게도 Silk Road 4를 분석 한 것처럼 이러한 사전 지식은 현재의 다크 넷 시장에서 더 이상 얻을 수 없다. 이 논문에서의 클러스터링 알고리즘은 이러한 실제적인 한계를 극복하고 클러스터 간의 숨겨진 관계를 찾고 추가로 연결함으로써 비트 코인 클러스터링 정확도를 높인다. 따라서 제안 된 휴리스틱 알고리즘은 비트 코인 거래를 현실 세계에서보다 투명하고 추적 가능하게 만드는 데 도움이 될 것이다.

V. 실험 결과

이 섹션에서, 다계층 휴리스틱을 통한 실험 결과 및 정확도 측정에 대한 평가를 설명한다.

5.1 실험 세팅

리뷰 데이터와 비트 코인 클러스터링에 대해 일치하는 주소를 찾는 알고리즘을 구현하기 위해 Intel Xeon E5-2620 3.0GHz 프로세서와 256GB RAM이 장착 된 시스템을 사용했다. Bitcoin Core를 사용하여 전체 Bitcoin 블록체인 데이터를 다운로드했다. 또한, 블록체인 분석을 위해 BlockSci를 사용했다. 이는 특별한 트랜잭션 (예 : CoinJoin)을 식별하고 기본 클러스터링을 기반으로 주소를 서로 연결하는 것과 같은 유용한 분석 도구 라이브러리를 포함하는 오픈 소스 소프트웨어 플랫폼이다.

각 리뷰 데이터에 대해 일치하는 비트 코인 주소를 자동으로 찾기 위해 먼저 수집 한 606 개의 원시 리뷰 데이터를 (V,D) 튜플 형식으로 표현한다. 여기서 V는 리뷰 데이터에 있는 항목의 비트 코인 값이고 D는 섹션 4.1에 설명 된대로 항목의 배송 날짜이다. 사전 처리 된 (V,D) 튜플을 기반으로 일치하는 주소를 찾기위한 비트 코인 클러스터링 진행률은 각각 섹션 5.2 및 5.3에 설명되어 있다.

5.2 매치 주소 서치

매치 주소를 찾기 위해 먼저 검색 공간을 D에서 D + 14 사이의 기간으로 설정한다. 검색 공간에서 BlockSci를 사용하여 정의 1의 모든 조건을 충족하는 일치하는 주소 MAs를 찾는다. 또한 섹션 3.3에서 설명한 것처럼 Silk Road 4에서 제공하는 주소 유형은 ScriptHash이므로 MAs는 ScriptHash 기반 주소이다.

위의 모든 조건을 충족하는 후보가 단 하나만 존재하는 경우 주어진 (V,D)에 대해 일치하는 주소 MA로 간주된다. 섹션 3.3의 측정에 따르면 Silk Road 4 트랜잭션의 89.52%는 WitnessPublicKeyHash 서명 방법을 사용하고 있으며 80.73%는 1개의 출력 트랜잭션이다. Silk Road 4 거래의 대부분이 이러한 패턴을 따르더라도 무시할 수 없는 반대 사례가 있기 때문에 이러한 정보는 Silk Road 4의 고유한 특성에 대한 구체적인 지표로 사용될 수 없다. 따라서 일치하는 주소를 찾기 위해 이러한 서명 패턴을 활용하지 않고 정의 1의 조건과 주소 유형만 확인한다.

그 결과, Table 3에서 볼 수 있듯이 606개의 리뷰데이터에 대해 31.68%(즉, 192개의 비트코인 주소)를 찾을 수 있다. 나머지 68.32%는 리뷰 데이터를 입력 주소로 V와 동일한 BTC로 두 개 이상의 비트 코인 거래가 발견된 경우를 포함한다. 둘 이상의 거래가 발견될 경우, 식별을 위한 추가 조건이나 정보 없이는 Silk Road 4와 관련된 정확한 거래와 허위 거래를 구별할 수 없다.

또한, WitnessPublicKeyHash 서명과 1개의 출력 트랜잭션 패턴을 일치하는 주소를 찾기위한 조

Table 3. Matched Address Rate

Condition	ScriptHash	ScriptHash & WitnessPublicKeyHash & 1-output transaction
Rate	31.68%	35.31%

건으로 포함하면 일치율이 31.68%에서 35.31%로 증가한다. 그러나 추가 3.63 %에는 거짓 긍정이 포함되어 있을 가능성이 높다. 더 많은 조건을 추가하는 것이 Silk Road 4에 대해 더 많은 주소를 찾는데 도움이 될 수 있음을 나타내더라도 조건이 정확하지 않은 경우 거짓 긍정을 늘릴 수 있기 때문에 정확한 조건만 사용해야한다.

5.3 클러스터링

이전 섹션에서 발견한 MA들은 Silk Road 4에서 구매자에게 불법 거래를 위해 제공한 Silk Road 4가 제어하는 비트코인 주소이다. 따라서 MA는 하나의 클러스터로 병합되어야 한다. 특히 제안된 다계층 휴리스틱과 같이 이 주소들을 포함하는 클러스터들은 하나의 클러스터로 병합된다.

2020년 5월 14일부터 10월 8일까지 시간의 흐름에 따른 클러스터 분포를 측정하여 다계층 휴리스틱의 효과를 평가한다. 먼저 192개의 MA에서 중복된 클러스터를 제거하여 133개의 고유 클러스터를 분류했다. 133개의 고유 클러스터 각각 크기는 Table 4와 같다. n는 클러스터의 크기를 나타내며 각 분포에 따른 개수를 나타낸다. 우리는 이 클러스터들은 다계층 휴리스틱을 통해 클러스터링하기 전에 이전 연구[9][11]에 따라 거짓 긍정을 줄이기 위해 현저하게 크기가 큰 클러스터를 제거했다. 따라서, 크기가 10^4이상인 클러스터를 제거하기로 했으며, 총 11개의 클러스터가 제외되었다. 이 11개의 클러스터 중 2개는 이미 Binance 및 Luno.com으로 태그가 되어있었으며, 나머지 9개의 클러스터는 알려지지 않은 클러스터이다.

Fig.5.는 다계층 휴리스틱을 통해 시간이 지남에 따라 나머지 122개의 클러스터가 하나의 클러스터(즉, Silk Road 4 클러스터)로 병합되는 모습을 보여준다.

Table 4. Count by cluster size

Cluster Size(n)	Count
$n < 10^1$	80
$10^1 \leq n < 10^2$	22
$10^2 \leq n < 10^3$	14
$10^3 \leq n < 10^4$	6
$10^4 \leq n$	11

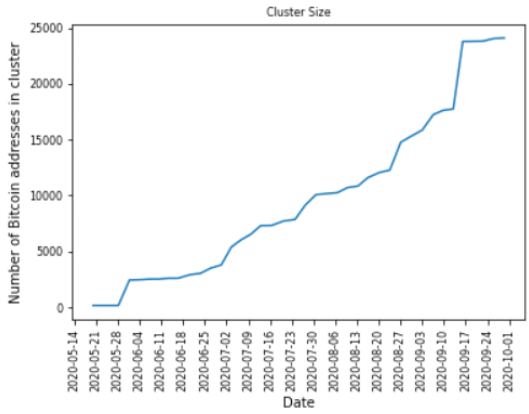


Fig. 5. Changes of clusters over time

마지막으로, 다계층 휴리스틱을 통한 거짓 부정의 감소를 측정한다. Table 5에서 볼 수 있듯이 우리는 MA를 통해 얻은 133개의 클러스터 중 알려지지 않은 클러스터 131개를 분류했다. 그리고 131개의 알려지지 않은 클러스터에서 122개의 Silk Road 4 클러스터를 식별했으며, 이를 하나의 클러스터로 병합하여 91.7%의 위음성을 줄일 수 있었다.

Table 5. Clustering Reslt of Multi-layer Heuristic

Cluster ID	Number of Clusters		
	Matched Address	Befor Multi-layer Heuristic	After Multi-layer Heuristic
Silk Road 4	0	122	1
Binance	1	1	1
Luno.com	1	1	1
Unkwown	131	9	9

5.4 한계 및 논의

이 연구에서 리뷰 데이터의 BTC가 주소 A에 대해 ITX(A)에 포함된 하나의 입력 주소의 BTC와 정확히 일치하는 경우만 고려했다. ITX(A)에 포함된 다수의 주소 합이 리뷰 데이터의 BTC와 동일한 경우도 매치 주소의 경우이다. 비트코인의 모든 트랜잭션에서 이러한 조합을 찾는 것이 이론적으로는 가능하지만 실제로는 어렵다.

예를 들어, 실제로 현재 하루에 약 200 개의 블록이 생성되고 각 블록에는 평균 약 2,000 개의 트랜잭션이 포함된다. 즉, 하루에 400,000 개의 트랜잭션이 생성되므로 이러한 모든 트랜잭션의 입력 주소를 분석하고, BTC의 모든 다른 조합을 계산하고,

무차별 대입 접근 방식을 사용하여 주어진 BTC의 모든 검토 데이터에 대해 고유 한 조합을 찾는 것이 거의 불가능 하다. 게다가, 우리는 심지어 단일 비트코인 거래에도 1,000 개가 넘는 입력 주소와 같이 엄청난 수의 주소가 포함되어있는 것을 관찰했다. 이론과 실제 사이의 이러한 격차를 줄이고 문제를 극복할 수 있다면 더 많은 매치 주소가 발견 될 것임이 분명하여 제안 된 클러스터링 휴리스틱 알고리즘이 불법 사이의 더 심층적 인 관계를 밝히는 데 도움이 된다.

제안된 다계층 비트코인 클러스터링 휴리스틱이 숨겨진 관계를 발견하고 클러스터링 정확도를 향상시킬 수 있더라도 결과 클러스터는 여전히 실제 클러스터의 일부이다. Ground-truth 정보를 얻는 것은 사실상 불가능하기 때문에 수집된 데이터를 기반으로 클러스터링 알고리즘을 평가할 수 밖에 없다. Ground-truth 정보를 찾을 수 있는 한 가지 방법은 실제 불법 거래에 참여해 보는 것이다. 그러나 윤리적인 문제로 인해 더 이상의 정보를 얻을 수 없었고 주소 타입과 서명 패턴과 같이 공개적으로 접근 가능한 정보만 관찰한다. Ground-truth 정보를 사용한다면 정확도를 향상시킬 수 있지만, 실질적으로 어려운 일이며, 이는 모든 비트코인 클러스터링 기술이 근본적으로 직면하는 문제로 남아있다.

VI. 관련 연구

Meiklejohn[9]은 비트코인 트랜잭션 데이터를 기반으로 *Multi-input heuristic* 및 *Change address heuristic이라는 비트코인* 클러스터링을 위한 최초의 휴리스틱 알고리즘을 제안했다. 거래의 출력 주소 수가 참여 사용자 수와 동일한 Coinjoin[26] 의 출현으로, Coinjoin의 특성에 따라 *Multi-input heuristic*이 약간 수정되었다.

Nick[10]은 트랜잭션에서 주소 변경을 결정하는 데 도움이 되는 *Consumer heuristic* 및 *Optimal change heuristic*을 제안했다.

그 이후로 비트 코인 클러스터링을 위해 비트 코인 트랜잭션 데이터와 웹 데이터를 결합하는 연구가 제안되었다. Ermilov[11]는 웹에서 얻은 공개 정보를 활용하는 비트코인 클러스터링 방법을 제안했다. 이들은 이전에 공개된 정보를 활용한다. 따라서 공개된 정보가 없거나 제공되지 않는 환경에서는 사용할 수 없다.

익명 네트워크에서 숨겨진 서비스의 공격 환경과 구조를 이해하기 위해 많은 연구에서 Tor 트래픽 [36][37][38] 및 활동[39][40][41][42]을 분석했다. Biryeukov[43][44]은 Tor를 통해 호스팅되는 숨겨진 서비스를 분석 한 결과 불법 인신 매매를 위해 많은 숨겨진 서비스가 유지되고 있음을 발견했다. 최근 Van Wegberg[45]는 온라인 익명 시장에서 사이버 범죄의 상품화가 증가하는 것을 관찰했으며, 이는 야심 찬 범죄자의 진입 장벽을 낮추고 사이버 범죄의 증가를 촉진한다.

Ciancaglini[39]은 언어 및 항목과 같은 기능을 분류하여 Tor 숨겨진 서비스의 범죄 활동을 분석했다. 이 연구에서 그들은 사용자가 딥 웹에서 사이버 범죄 상품을 거래하는 방법을 분석하고 거래 패턴을 표면 웹의 거래 패턴과 비교했다. 그런 다음 그들은 이전 작업의 후속 작업[40]으로 다크 웹에서의 불법 거래를 분석했으며, 딥 웹의 많은 익명 네트워크가 사이버 및 물리적 영역 모두에서 범죄 활동을 숨기는 안전한 피난처가 되었음을 발견했다.

Soska 및 Christin[41]은 2013 년과 2015 년 사이에 16 개의 Tor 사이트에서 판매 된 제품 유형을 분석했다. 또한, 공급 업체 측면에서 보안 관행이 어떻게 진화했는지 그 결과 공급 업체가 PGP 키를 사용하여 통신을 암호화하여 범죄 활동을 숨길 가능성이 있음을 발견했다. Barratt[46]는 다크 웹 마약 시장, 특히 Silk Road를 조사하여 영국, 호주 및 미국의 온라인 불법 마약 시장에 대한 인지도를 조사하고 마약 구매자가 온라인 마약 시장을 얼마나 선호했는지 파악했다. 그들은 불법 마약 거래자뿐만 아니라 제네릭 의약품을 구매한 일반 구매자에게도 초점을 맞추고 온라인 다크넷 시장을 사용할 가능성이 높은 이유를 분석했다. 그 결과 마약 구매자에 대한 호소력이 국가별 억제력과 시장 특성에 의해 완화되었다고 주장하는 분석 결과를 제공했다. Foley는 다크 넷 시장에서 불법 활동의 규모를 추정하기 위해 몇 가지 기능을 제안했다. 이 연구에서 그들은 비트코인 사용자의 약 25%가 불법 활동에 연루되어 있으며, 연간 약 760 억 달러의 불법 활동이 비트코인과 관련되어 있다고 추정했다.

다크 웹에서 불법 거래 및 암호 화폐 거래량을 측정 한 연구도 있다. Christin[47]는 2011 년부터 2012 년까지 6 개월 동안 Silk Road 데이터를 수집하고 분석했다. 그들은 Silk Road 에서 판매

된 24,400 개의 개별 품목을 조사한 결과 Silk Road 가 불법 상품, 특히 불법 마약 거래에 압도적으로 사용된다는 사실을 발견했다. 또한, 이 연구는 Silk Road 에서 발생한 총 매출이 매월 120 만 달러 이상으로 평가되었으며, 이는 2012 년 Silk Road 운영자에 대한 수수료로 약 92,000 달러에 해당한다. Demant[48]은 Silk Road 2.0과 Agora를 분석했다. 그들은 2014 년부터 2015 년까지 웹 크롤러를 통해 데이터를 수집하고 거래 상품에 대한 수요를 조사했다. 분석에 따르면 대부분의 수익이 B2B 거래에서 발생했으며 해당 기간 동안 두 시장 모두에서 거래량이 감소한 것으로 나타났다. 그 결과, 다크 넷 시장은 유통 및 수익 측면에서 기존의 약품 시장과 유사하다는 것을 보여주었다.

VII. 결 론

이 연구에서는 비트코인 트랜잭션 계층과 어플리케이션 계층 정보를 모두 활용 한 다계층 비트코인 클러스터링 휴리스틱을 제안했다. 애플리케이션 계층에서 유용한 오프 체인 데이터를 얻기 위해 다크넷 시장, 특히 Silk Road 3.1 및 4에서 사용 가능한 데이터에 대한 포괄적인 분석을 수행하고 에스크로 방식 거래에서 사용되는 주소 및 서명 패턴과 같은 고유한 특성을 분석했다. 분석 결과, 우리가 수집 한 606개의 Silk Road 4 리뷰 데이터 중 약 31.68%가 실제 비트코인 거래와 일치함을 발견했으며, 이를 사용하여 식별 할 수 없는 122 개의 Silk Road 4 숨겨진 클러스터를 발견했다. 제안 된 휴리스틱은 기존 클러스터링 방법을 보완 할 수 있으며 이들의 거짓 부정 비율을 최대 91.7 %까지 크게 줄일 수 있다.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Decentralized Business Review, 21260, 2008.

[2] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," Ethereum project yellow paper 151. 2014: 1-32, 2014.

[3] S. Noether, "Ring SIgnature Confidential Transactions for Monero," IACR Cryptol. ePrint Arch.: 1098, 2015.

[4] D. Hopwood, S. Bowe, T. Hornby and N. Wilcox, "Zcash protocol specification." GitHub: San Francisco, CA, USA, 2016.

[5] S. Foley, J. R. Karlsen and T. J. Putniņš, "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?," The Review of Financial Studies 32.5: 1798-1853. 2019.

[6] Google, "https://www.google.com/", 2021.3.24

[7] Naver, "https://www.naver.com/", 2021.3.24

[8] Silk Road 4, "http://silkroad7rn2puhj.onion/", 2021.3.24

[9] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names." Proceedings of the 2013 conference on Internet measurement conference, pp.127-140, 2013.

[10] J. D. Nick, "Data-driven de-anonymization in Bitcoin," Master's thesis, ETH-Zürich, 2015.

[11] D. Ermilov, M. Panov, and Y. Yanovich, "Automatic bitcoin address clustering," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 461-466, 2017

[12] Amazon, "https://www.amazon.com/", 2021.03.24

[13] Aliexpress, "https://www.aliexpress.com/", 2021.03.24

[14] J. Aldridge and R. Askew, "Delivery dilemmas: How drug cryptomarket users identify and seek to reduce their risk of detection by law enforcement," International Journal of Drug Policy, 41: 101 - 109, 2017.

[15] Cryptmixer, "https://cryptmixer.com/", 2021.03.24

[16] BitMix, “https://bitmix.biz/en” , 2021. 03.24
[17] E. Heilman, L. Alshenibr, F. Baldimtsi, A. Scafuro and S. Goldberg, “Tumblebit: An untrusted bitcoin-compatible anonymous payment hub,” Network and Distributed System Security Symposium, 2017.
[18] S. Delgado-Segura, C. Pérez-Sola, G. Navarro-Arribas and J. Herrera-Joancomartí, ″Analysis of the bitcoin utxo set,″ International Conference on Financial Cryptography and Data Security, Springer, Berlin, Heidelberg, pp.78-91, 2018.
[19] A. Gavin, “Pay to Script Hash,” Technical Report BIP-16, Bitcoin Improvement Proposal, 2012.
[20] K. Thomas, R. Jean-Pierre and V.Ruben, “Deterministic Pay-to-script-hash multi-signature addresses through public key sorting”, Technical Report BIP-67, Bitcoin Improvement Proposal, 2015.
[21] L. Eric, L. Johnson and W. Pieter, “Segregated Witness,” Technical Report BIP-141, Bitcoin Improvement Proposal, 2012.
[22] L. Johnson and W. Pieter, “Transaction Signature Verification for Version 0 Witness Program,” Technical Report BIP-143, Bitcoin Improvement Proposal, 2016.
[23] L. Johnson and W. Pieter, “Dealing with signature encoding malleability (consensus layer),” Technical Report BIP-146, Bitcoin Improvement Proposal, 2016.
[24] H. Kalodner, M. Möser, K. Lee, S. Goldfeder, M. Plattner, A. Chator and A. Narayanan, “BlockSci: Design and applications of a blockchain analysis platform,” 29th {USENIX} Security Symposium, pp. 2721-2738, 2020.
[25] Blockchain.com, “https://Blockchain.com”, 2021.03.21
[26] G. Maxwell, “CoinJoin: Bitcoin privacy for the real world,” https://bitcointalk.org/index.php/topic=279249.0, 2021.03.21
[27] N. Christin, ″Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace,″ Proceedings of the 22nd international conference on World Wide Web, pp.213-224, 2013.
[28] J. Van Buskirk, S. Naicker, R. B. Bruno, C. Breen and A. Roxburgh, “Drugs and the Internet,” 2016.
[29] M. Horton-Eddison, ″Updating Escrow: demystifying the CDM multisig process,″ 2017.
[30] A. ElBahrawy, L. Alessandretti, L. Rusnac, D. Goldsmith, A. Teytelboym and A. Baronchelli, “Collective dynamics of dark web marketplaces,” Scientific reports, 10(1), 1-8, 2020.
[31] W. Lacson and B. Jones. ″The 21st Century DarkNet Market: Lessons from the Fall of Silk Road.″ International Journal of Cyber Criminology, 10.1, 2016.
[32] M. Horton-Eddison and M. Di Cristofaro, ″Hard interventions and innovation in crypto-drug markets: The escrow example,″ Policy Brief, 11, 2017.
[33] V. Adewopo, et al. ″Plunge into the Underworld: A Survey on Emergence of Darknet,″ 2019 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, p.155-159, 2019.
[34] Apollon Market, “http://apollionih4ocqyd.onion”, 2021.02.19
[35] Agartha Market, “http://agarthaourmnyhq3.onion”, 2021.02.19
[36] S. J. Murdoch and G. Danezis, “Low-cost traffic analysis of Tor,” 2005 IEEE

Symposium on Security and Privacy, pp.183-195, 2005.

[37] P. Mittal, A. Khurshid, J. Juen, M. Caesar and N. Borisov, "Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting," Proceedings of the ACM conference on Computer and communications security, pp.215-226, 2011.

[38] A. Kwon, M. AlSabah, D. Lazar, M. Dacier and S. Devadas, "Circuit finger printing attacks: Passive deanonymization of tor hidden services," USENIX Security Symposium, pp.287-302, 2015.

[39] V. Ciancaglini, M. Balduzzi, M. Goncharov and R. McArdle, "Deepweb and Cybercrime," Trend micro report, 9: 5-6, 2013.

[40] V. Ciancaglini, M. Balduzzi, R. McArdle and M. Rösler, "Below the surface: Exploring the deep web," Trend Micro, pp.1-48, 2015.

[41] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," USENIX Security Symposium, pp.33-48, 2015.

[42] I. Sanchez-Rola, D. Balzarotti and I. Santos, "The onions have eyes: A comprehensive structure and privacy analysis of tor hidden services," Proceedings of the International Conference on World Wide Web, pp.1251-1260, 2017.

[43] A. Biryukov, I. Pustogarov and R. P. Weinmann, "Trawling for tor hidden services: Detection, measurement, deanonymization," 2013 IEEE Symposium on Security and Privacy, pp. 80-94, 2013.

[44] A. Biryukov, I. Pustogarov, F. Thill and R. P. Weinmann, "Content and popularity analysis of Tor hidden services," 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 188-193, 2014.

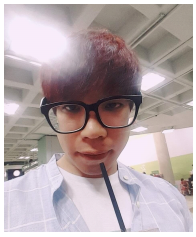
[45] R. Van Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. H. Ganan, B. Klievink and M. Van Eeten, "Plug and prey? measuring the commoditization of cybercrime via online anonymous markets," USENIX Security Symposium, pp.1009-1026, 2018.

[46] M. J. Barratt, J. A. Ferris and A. R. Winstock, "Use of Silk Road, the online drug marketplace in the United Kingdom, Australia and the United States," Addiction, vol 109, no.5, pp. 774-783, 2014.

[47] N. Christin, "Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace," Proceedings of the 22nd international conference on World Wide Web, pp. 213-224, 2013.

[48] J. Demant, R. Munksgaard and E. Houborg, "Personal use, social supply or redistribution? Cryptomarket demand on Silk Road 2 and Agora," Trends in Organized Crime, 21(1), pp. 42-61, 2018.

〈저자소개〉



이 진 희 (Jin-hee Lee) 학생회원
2018년 2월: 한성대학교 전자정보공학과 졸업
2019년 2월~현재: 고려대학교 컴퓨터학과 석사과정
〈관심분야〉 정보보호

김 민 재 (Min-jae Kim) 학생회원
2021년 2월: 고려대학교 컴퓨터학과 졸업
2021년 3월~현재: 고려대학교 컴퓨터학과 석박통합과정
〈관심분야〉 정보보호, 컴퓨터공학, 인공지능, 블록체인

허 준 범 (Junbeom Hur) 종신회원
2001년 2월: 고려대학교 컴퓨터학 학사
2005년 8월: 한국과학기술원 전산학 석사
2009년 8월: 한국과학기술원 전산학 박사
2009년 9월~2011년 8월: University of Illinois at Urbana-Champaign 박사후 연구원
2011년 9월~2015년 2월: 중앙대학교 컴퓨터공학부 조교수
2015년 3월~2016년 8월: 고려대학교 컴퓨터학과 조교수
2016년 9월~2021년 8월: 고려대학교 컴퓨터학과 부교수
2021년 9월~현재: 고려대학교 컴퓨터학과 교수
2021년 9월~2022년 8월: ETH Zurich 방문교수
〈관심분야〉 응용 암호, 네트워크 보안, 클라우드 보안, 시스템 취약점