특집 01

암호화폐 거래 추적 기술의 이해

김태일 (코어시큐리티)

목 차

1. 서 론

암호화폐는 익명성(anonymity) 또는 가명성(pseudonymity)을 가진다는 특성으로 인해 원래 의도와는 다르게 범죄 집단의 자금세탁 및 자금 추적 회피 수단으로 널리 사용되어왔다. 또한 물리적 접근없이 해킹 등을 통해 암호화폐의 탈취가 가능해 암호화폐는 그 자체로 범죄의 대상이 되고 있는 실정이다. 이에 세계 주요국에서는 암호화폐와 관련된 여러 불법 행위를 추적하기위해 KYC(Know Your Customer), KYT(Know Your Transaction), 트래블 룰과 같은 규제들을 도입하고 있다. 이런 규제들이 실효성을 가지기 위해서는 암호 화폐 거래 추적 기술이 필수적이다. 이에 미국,영국을 비롯한 주요 선진국에서는 암호화폐의 거래 추적 기술에 대한 연구가 활발히 진행 중이며 체이널리시스, TRM과 같은 암호화폐 거래 추적 및 규제 지원 전문 기업들도 속속 등장하고 있다. 아쉽게도 국내의 경우 아직 관련 연구가 활발하지는 않지만 관심의 정도는 날로 증가하고 있다. 이에 본 고에서는 암호화폐 거래 추적 주요 기법과 이슈들을 개괄적으로 살펴 암호화폐 거래 추적 기술에 대한 이해를 돕고자 한다.

2. 암호화폐 거래 추적 개요

암호화폐 거래 추적 활동은 크게 온체인(on-chain) 분석과 오프체인(off-chain) 분석으로 나눌 수 있다. 온체인 분석은 블록체인 네트워크 상에서 이루어지는 통신 내용과 거래 내역을 분석하는 활동으로 블록체인 원장 분석, 블록체인 노드 간 통신 내용 분석 등 패시브(passive)한 분석 방법과 더스팅(dusting)처럼 실제 추적 대상에 자금을 송금한 후 발생하는 거래를 추적하는 액티브(active)한 분석 방법을 포함한다. 한편 오프체인 분석은 블록체인 네트워크 외 다른 소스로부터 수집한 데이터를 분석하여 거래 추적 등에 필요한 정보를 획득하는 방법으로 주로 OSINT, HUMINT 등의 활동이 포함된다.

암호화폐 거래 추적의 활동의 핵심은 식별, 어드레스 클러스터링, 자금흐름 추적이다. 이 중 식별은 온체인 분석을 통해 신원 파악에 필요한 기

초 정보를 수집하고 오프체인 분석을 통해 신원을 특정하는 방식으로 이루어지며, 자금 흐름 분석과 어드레스 클러스터링은 주로 블록체인 원장 분석과 블록체인 노드 간 통신 내용 분석을 통해 이루어진다.

3. 식별(identification)

암호화폐 추적에서 식별은 거래자의 신원을 특정하는 과정이다. 블록체인 원장에는 거래자의 신원을 확인할 수 있는 아무런 정보도 없으므로 원장 분석 만으로는 추적대상의 신원을 특정할 수 없다. 거래자의 신원을 특정하기 위해서 가장 널리 사용되는 정보는 KYC 규제에 따라 수집된 실명 인증 및 신원 정보인데 이러한 정보를 보유하고 있는 대표적인 곳은 규제를 준수하고 있는 거래소이다. 이러한 이유로 많은 거래 추적에서 조사자는 추적 대상과 거래소와의 연결 고리를 찾는데 집중한다. 예를 들어 추적 대상과 직접적/간접적 거래 관계에 있는 주소들 중 거래소로 직접 송금이 이루어졌거나 거래소로부터 입금을 받은 주소를 찾은 후, 거래소가 보유한 거래 정보와 신원 정보를 활용하여 해당 주소로부터 자금을 수신하거나 해당 주소로 자금을 보낸 거래소 내 사용자의 신원을 1차적으로 확인할 수 있다. 이 후 특정된 사용자를 대상으로 한 조사를 통해 추적 대상의 신원을 식별할 수 있게 된다.

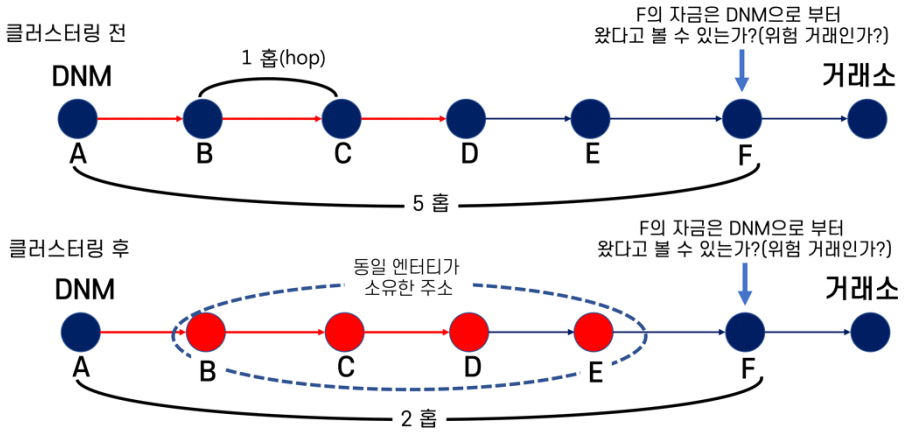
식별이 개인만을 대상으로 하는 것은 아니다. 거래소, DNM 등과 같은 서비스 역시 식별의 대상이 된다. 서비스 식별은 개인의 신원을 특정하는 것과는 다른 방법으로 이루어진다. 개인처럼 KYC 관련 정보가 존재하지 않기 때문이다. 서비스 식별은 실제 서비스로의 입출금을 반복하면서 서비스에서 사용하는 주소들을 수집한 후 후술할 어드레스 클러스터링 기법을 활용하여 더 많은 주소들을 확보하기도 하고, OSINT를 통해 공개된 정보를 수집하고 분석하여 식별을 하기도 한다.

4. 어드레스 클러스터링(address clustering)

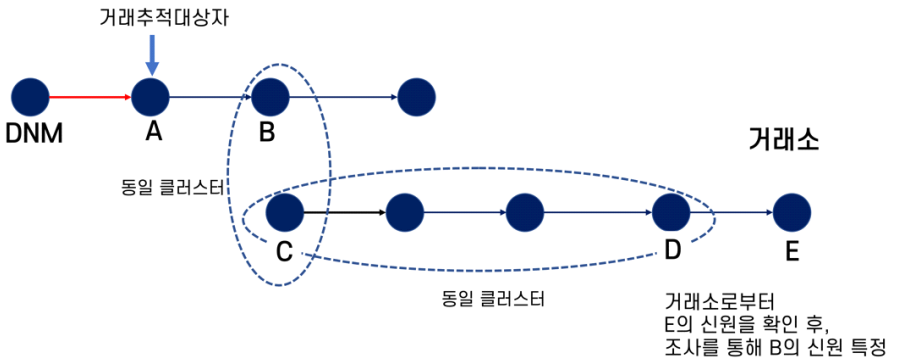
4.1 어드레스 클러스터링 개요 및 필요성

암호화폐 거래 추적의 맥락에서 클러스터링은 동일한 엔터티(entity)가 소유한 암호화폐 주소들을 식별하고 이들을 하나의 그룹으로 묶어내는 분석 활동을 뜻한다. UTXO 모델을 사용하고 있는 비트코인의 경우 프라이버시 보장을 위해 주소를 재사용하지 않도록 권고하고 있으며 지갑 소프트웨어들은 이러한 권고에 따라 사용자들이 주소를 재사용하지 않을 수 있도록 돕고 있다. 이러한 프랙티스로 인해 개인의 경우 일반적으로 수십 개에서 수백 개 정도의 주소를 사용하게 되며 서비스의 경우에도 적게는 수천 개에서 많게는 수백만개 이상의 주소를 사용하고 있다. 암호화폐 거래 추적의 궁극적인 목적이 행위자와 행위 사실을 밝혀내기 위한 것임을 감안하면 효과적인 분석을 위해 동일 엔터티가 소유하고 있는 주소들을 하나의 그룹으로 묶는 것이 필수적이라는 사실을 직관적으로 이해할 수 있다. 암호화폐 거래 추적에 있어 클러스터링이 중요한 이유를 몇 가지 살펴보면 다음과 같다.

- (그림 1)의 예에서 볼 수 있는 것처럼 거래의 위험성이나 연관성을 판단하기 위해서는 거래의 위험성이나 연관성을 판단하기 위해 홉수는 주된 판단 기준 중 하나이며 클러스터링 결과에 따라 엔터티 간의 홉수가 달라질 수 있다. <그림 1>의 상단 그림에서 거래소에 자금을 송금하는 F는 DNM(DarkNet Market)으로부터 자금을 간접적으로 수신했으나 DNM과의 홉수가 5이므로 DNM과의 연관성이 적은 것으로 판단할 수 있다. 하지만 (그림 1)의 하단 그림에



(그림 1) 클러스터링과 위험거래 식별



(그림 2) 클러스터링과 거래자 신원 특정

서처럼 B,C,D,E가 동일 엔터티가 소유한 주소로 밝혀졌다면 DNM과 F의 홉수는 2가 되어 DNM과 F가 강한 연관성을 가지고 있다고 판단할 수 있다. 따라서 거래소의 입장에서 F의 자금은 위험도가 높은 DNM으로부터 펀딩된 것으로 판단하여 F가 거래소에 입금하는 거래를 위험 거래로 식별하게 된다.

- (그림 2)에서는 클러스터링이 추적 대상의 신원 식별 기회를 더 많이 제공할 수 있음을 보여준다. 그림의 예를 보면 클러스터링을 수행하지 않은 경우 거래 추적 대상자인 A의 신원을 확인하기 어려우나 클러스터링을 통해 B,C,D 등의 주소가 동일 엔터티에 속한 주소라는 사실을 알아낸 경우 거래소로부터 E의 신원을 확인한 후 조사를 통해 B의 신원을 특정할 수 있는 기회를 가지게 된다.

4.2 어드레스 클러스터링 기법

현재 가장 널리 사용되는 어드레스 클러스터링 기법들은 체인지 주소 분석, 주소 특성 분석, 지갑 특성 분석, 클러스터 프로파일링 등을 포함한다. 이 중 체인지 주소 분석은 비트코인처럼 UTXO 거래 모델을 사용하는 암호화폐에만 적용된다. 지면 관계 상 모든 기법들을 상세히 설명하기 어려우나 각 각의 기법에 적용되는 주된 추론들을 소개하면 다음과 같다.

4.2.1 체인지 어드레스 식별

UTXO 거래 모델을 사용하는 암호화폐에는 체인지(change, 잔돈)라는 개념이 존재한다. 체인지 주소는 일정 금액을 지급한 후 남는 잔액을 돌려받을 주소로 거래 아웃풋에 체인지가 존재한다면 이 체인지는 주소는 인풋에 사용된 주소와 동일 지갑에 속한 주소로 판단할 수 있다. 거래 아웃풋으로 출력된 UTXO들 중에서 체인지를 식별하는데 사용되는 주된 추론들은 다음과 같다.

(추론 1) 아웃풋 중 하나는 새롭게 생성된 주소이고 다른 아웃풋들은 이전에도 사용되었던 주소들이라면 새롭게 생성된 주소가 체인지 주소일 확률이 높다.

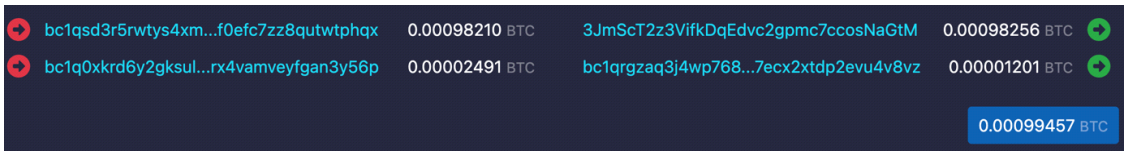
이 추론은 지갑의 특성을 관찰한 결과에 기반을 두고 있다. 오늘날 암호화폐 지갑들은 사용자가 별도로 지정하지 않는 이상 매 거래마다 체인지 주소를 새롭게 생성하는 경향을 가지고 있다. 따라서 아웃풋으로 출력된 UTXO를 잠그는데 사용된 주소들 중 이전 거래에서는 관찰되지 않은 새로운 주소가 존재한다면 지갑이 새롭게 생성한 체인지 주소일 가능성이 높다고 볼 수 있다.

(추론 2) 아웃풋에 사용된 주소 중 하나가 인풋으로 사용된 주소들과 타입이 동일하고 아웃풋으로 사용된 나머지 주소들은 그렇지 않은 경우 주소 타입이 동일한 주소가 체인지 주소일 확률이 높다.

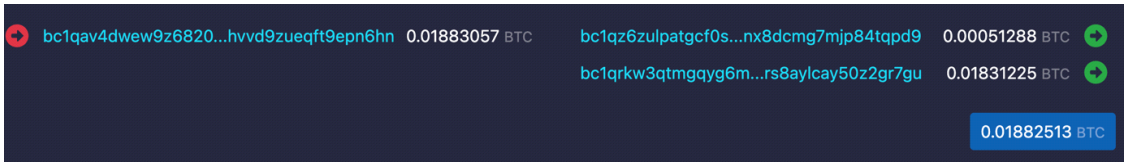
비트코인의 경우 기능과 목적에 따라 1, 3, bc1q, bc1p로 시작하는 4가지 유형의 주소가 사용된다. 이 추론은 지갑이 한가지 유형의 주소를 사용하는 경향이 강하다는 관찰에 기반하고 있다. (그림 3)의 예를 보면 인풋으로 사용된 주소들은 bc1q로 시작하는 주소들이며 아웃풋에는 각 각 3으로 시작하는 주소와 bc1q로 시작하는 주소가 관찰된다. 이 경우 인풋과 주소 타입이 같은 bc1qrgzaq...가 체인지 주소일 가능성이 높다. 이 추론은 어디까지나 현재 지갑이 한가지 타입의 주소를 사용하는 경향이 있다는 관찰에 기반한 것이므로 미래에도 이 추론이 반드시 유효하다고 보기는 어렵다. 잘 알려진 지갑 소프트웨어 중 일부는 지금도 프라이버시 강화를 목적으로 여러 유형의 주소를 동시에 사용할 수 있도록 설계되어 있다.

(추론 3) 코스펜딩인 거래의 경우 인풋 UTXO들의 액면 금액 총합보다 큰 값을 가지는 아웃풋 UTXO가 지출이고 그렇지 않은 쪽이 체인지일 가능성이 높다.

(그림 3)의 거래는 명백히 코스펜딩이라고 볼 수 있다. 이 거래는 0.00098256을 지출하기 위한 거래로 0.00098256보다 큰 액면 금액을 가지는 UTXO를 가지고 있지 않았기 때문에 두 개의 UTXO를 모아서 지출한 것으로 이해할 수 있다. 따라서 0.00001201은 체인지이다. 이 추론에서 유의할 점은 인풋 UTXO가 여러 개일 지라도 정확하게 동일한 주소들이 여러 개 인풋에 사용된다면 코스펜딩 추론을 적용할 수 없다는 사실이다. 또한 코스펜딩과 트랜잭션의 모습이 유사한 코인조인의 경우에도 당연히 이런 추론을 적용할 수 없다.



(그림 3) 주소 타입을 활용한 체인지 어드레스 식별



(그림 4) UTXO의 액면 금액을 활용한 체인지 어드레스 식별

(추론 4) 아웃풋으로 출력된 UTXO의 액면금액이 보다 단순한 쪽이 지출이며 그렇지 않은 쪽이 체인지일 가능성이 보다 높다.

이 추론은 지출 거래에서 지출하는 금액이 보통 사람이 입력하기 쉬운 값인 경우가 많다는 관찰에 기초한다. (그림 4)의 거래에서 0.01831225 보다는 0.00051288이 숫자의 관점에서 더 단순해 보이므로 지출일 가능성이 보다 높다고 볼 수 있다. 이 추론은 앞의 다른 추론들과는 달리 상대적으로 약한 추론이다.

4.2.2 주소의 특성을 활용한 클러스터링

(추론 6) 인풋에 사용된 주소가 다중서명 주소이고 아웃풋에 동일한 속성의 다중서명 주소가 존재하는 경우 두 주소는 서로 동일 엔터티에 의해서 관리되는 주소이다.

비트코인 주소 중에는 다중서명 속성을 가진 주소들이 존재한다. 다중서명 주소는 여러 엔터티가 UTXO의 소유권 증명에 참여해야 하는 경우 사용되는 특성으로 보통은 에스크로우 서비스, 자금의 안전한 관리 목적 등에 활용된다. 일반적으로 다중서명 주소는 개인보다는 서비스에서 주로 사용되며 동일 엔터티가 사용하는 다중서명 주소는 대체로 그 속성(예 다중서명 2/3, 다중서명 3/3)이 동일하다. (그림 5)의 예에서 인풋 주소는 다중서명 2/3(2 out of 3) 속성을 가지고 있으며 그림상에 명시되어 있지는 않으나 bc1qg74... 역시 조사해보면 다중서명 2/3 속성을 가지고 있어 같은 서비스 클러스터에 포함되는 주소임을 알 수 있다.

4.2.3 지갑의 특성을 활용한 클러스터링

비트코인의 경우 트랜잭션 속성 중 일부 예를들어 RBF(Replace By Fee), 버전, 잠금시간, 세그윗 사용 여부는 지갑마다 그 설정이 다를 수 있다. 따라서 이처럼 지갑의 설정 값을 엿볼 수 있는 트랜잭션 속성을 분석하여 동일 지갑에 속해 있는 주소를 식별할 수 있다.

(그림 5) 다중서명 주소 속성을 활용한 어드레스 클러스터링

5. 자금 흐름 분석

자금 흐름 분석은 블록체인 원장을 조사하여 자금을 송수신한 주소와 금액 그리고 시간을 식별하는 것을 주 목적으로 한다. 식별된 자금의 흐름은 보통 그래프로 표현된다. 자금을 송수신한 클러스터가 노드가 되며 자금 흐름의 방향과 자금의 크기는 노드와 노드를 잇는 엣지로 표현한다. 일반적이고 정상적인 거래에서 자금의 송수신 주소를 식별하는 것은 매우 직접적이고 쉬운 일이지만 자금 흐름 분석에는 몇 몇 어려움이 존재한다.

5.1 믹싱(mixing)

텀블링(thumbling)이라고도 불리는 믹싱은 자금이 어디로 전달되었는지를 감추는 것을 목적으로 한 자금흐름 난독화 기술이다. (그림 6)은 와사비월렛(Wasabi wallet) 등에서 사용되는 코인조인(coin join)을 활용한 믹싱의 예이다. (그림 6)의 예에서는 좌측 박스의 주소로부터 믹서로 입금된 UTXO가 좌측 주소로 전달되었다. 이는 디믹싱(demixing) 기술을 활용하여 분석 한 내용이기는 하지만 그림에서 볼 수 있듯이 믹싱된 자금은 어디로 전달되었는지 직관적으로 알기가 어렵고 분석 과정도 매우 복잡하다. 믹싱은 자금 세탁이나 거래 추적 회피에 널리 사용되며 거래 추적을 어렵게 하는 주된 방해 요소 중 하나이기 때문에 믹싱된 자금의 흐름을 추적하는 기술인 디믹싱은 자금 흐름 분석에서 가장 중요한 기술 중 하나라고 할 수 있다.

믹싱 서비스는 수탁형(custodial)과 비수탁형(non-custodial)으로 구분된다. 수탁형 믹싱 서비스는 자금을 믹싱하여 되돌려주는 서비스를 제공하는 업체나 업자에게 자금을 전달하고 지정한 시간과 주소로 믹싱된 자금을 돌려받는 형태로 이루어진다. 수탁형 믹싱 서비스들은 서비스들마다 고유한 알고리즘을 개발하여 사용하기 때문에 이를 분석하여 믹싱 서비스를 식별하고 디믹싱 알고리즘을 개발하려는 노력들이 오랫동안 계속되어 왔다.

비수탁형 믹싱 서비스는 대체로 믹싱을 중계하는 서버를 통해 다수의 사용자들이 모여 믹싱을 수행하는 형태로 이루어진다. 이 때 서버는 믹싱에 대한 어떠한 기록도 남기지 않으며 KYC를 요구하지도 않는다. 따라서 사용자들은 와사비월렛(wasabi wallet), 사무라이월렛(samourai wallet)과 같은 전용 클라이언트를 이용하여 아무런 규제 없이 자유롭게 믹싱을 수행할 수 있다.

디믹싱은 믹서식별과 믹싱된 자금 흐름 추적의

(그림 6) 코인조인을 이용한 믹싱과 디믹싱 예

2단계로 이루어진다. 유형과 관련없이 믹서들마다 트랜잭션에서 드러나는 패턴이 존재하기 때문에 이를 시그너처로 믹서들을 식별할 수 있다. 예를 들어 (그림 6)의 예는 와사비월렛을 이용한 믹싱으로 볼 수 있다. 와사비월렛을 이용한 믹싱 트랜잭션의 특징은 인풋과 아웃풋이 매우 많고 인풋과 아웃풋 주소가 모두 bc1q로 시작하며 아웃풋에 어노니머티 셋(anonymity set)이 다수 관찰되는 특징을 가진다. 어노니머티 셋은 믹싱을 통해 생성된 동일한 액면가를 가지는 UTXO들로 액면가의 상이함을 이용하여 자금의 흐름을 추정하는 것을 방지하기 위한 목적을 가지고있다.

믹싱된 자금 흐름 추적은 크게 두 가지 방법이 적용된다. 그 중 하나는 믹싱에 사용된 알고리즘을 파훼하는 것이고 또 다른 하나는 주소 재사용에 따른 실수를 이용하는 것이다. (그림 6)에서 박스에 있는 두 개 주소들은 모두 재사용된 주소들이며 다른 트랜잭션들로부터 동일한 엔터티가 소유하고 있는 주소임을 알 수 있는 것들이다.

5.2 체인호핑(chain hopping)

체인호핑은 비트코인(BTC)을 이더(ETH)로 교환하는 것처럼 특정 암호화폐를 다른 종류의 암호화폐로 교환하는 거래이다. 체인호핑은 주로 중앙화 거래소, 탈중앙화 거래소(DEX), P2P 거래소, 크로스체인 브릿지(cross chain bridge) 등을 통해서 이루어진다. 중앙화 거래소를 통해서 체인호핑을 하는 경우에는 거래소가 보유한 데이터베이스를 통해서 자금 흐름을 추적할 수 있으며, 탈중앙화 거래소나 P2P 거래소, 크로스체인 브릿지를 를 통해 체인호핑이 이루어지는 경우 원장을 분석하여 각 서비스들의 거래 패턴을 식별하고 이를 분석하여 자금의 흐름을 추적할 수 있다. (그림 7)은 탈중화 거래소 중 하나인 유니스왑(Uniswap)을 이용한 체인호핑의 예이다. 이 거래를 통해 0x9Ce950...은 유니스왑을 통해 1.59 이더를 3,788 폴카시티로 교환했다.

5.3 필 체인(peel chain)

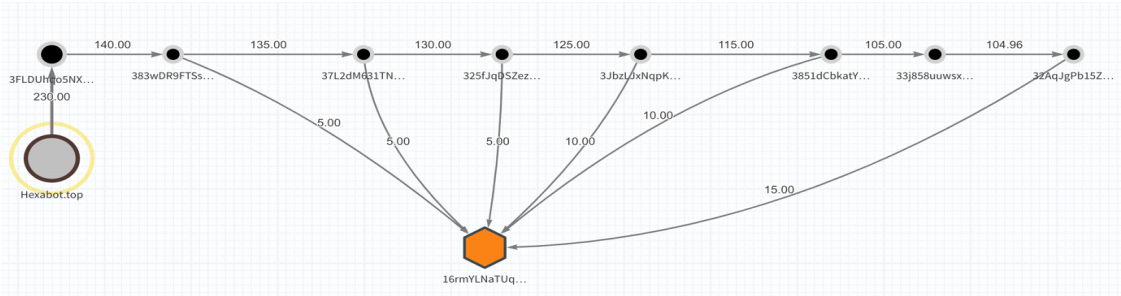
(그림 8)은 스트럭처링 또는 분할거래로도 불리는 필체인의 예를 보여준다. 필체인은 자금을 오랜 시간에 걸쳐 조금씩 나누어 특정 주소로 이동시키는 거래 방식으로 한번에 큰 규모의 자금을 이동시켜 주목을 끄는 일을 피하기 위해 사용된다. 비트코인의 경우 필체인은 작은 금액을 특정 주소로 보내고 대부분의 금액을 체인지 주소로 보내는 것을 반복하는 형태로 이루어진다. 이 때 작은 금액들은 필(peel)이라고하고 체인지 주소들의 연결된 형태를 체인(chain)이 된다. 필체인인은 언급한 바와 같이 오랜 시간에 걸쳐 자금을 이동시키는 특성을 가진다. 조사 당시 이미 필체인이 완성된 형태라면 자금의 흐름을 추적하는 것이 어렵지 않을 수 있으나 필체인이 진행 중이라면 다음 체인이 형성될 때까지 지속적인 모니터링이 필요하다. 또한 필체인

Transfers

Zero amount transfers are hidden.

Amount	Asset	Sending Cluster	Sending Address	Receiving Cluster	Receiving Address
1.59	ETH	0x9Ce950A0c52B2425…	0x9Ce950A0c52B24254C207b…	Uniswap V2	0x7a250d5630B4cF539739dF2…
1.59	ETH	Uniswap V2	0x7a250d5630B4cF539739dF…	Wrapped Ether Cont…	0xC02aaA39b223FE8D0A0e5C4…
1.59	WETH	Uniswap V2	0x7a250d5630B4cF539739dF…	Uniswap V2: POLC 0…	0xb5Bd39Fc0C57d902271Ba6C…
3,788.…	POLC	0xb5Bd39Fc0C57d902…	0xb5Bd39Fc0C57d902271Ba6C5cd8…	0x9Ce950A0c52B2425…	0x9Ce950A0c52B24254C207bB1485…

(그림 7) 유니스왑을 이용한 체인호핑 예



(그림 8) 필 체인 예

거래를 반복하여 작은 필체인들로 구성된 거대한 복합 필체인을 형성하는 방식으로 거래 추적을 복잡하게 만드는 사례도 빈번하게 관찰된다.

5.4 장외거래(OTC, Over the Counter)

장외거래는 규제를 준수하는 정상적인 거래소를 통하지 않고 비공개된 마켓에서 대량의 암호화폐를 사거나 파는 거래로 보통은 에이전시나 브로커에게 거래를 위탁하는 형태로 이루어진다. 장외거래는 시장의 가격 변동을 유발하지 않고 대량의 암호화폐를 사거나 팔 수 있으며, 거래 과정에서 신원을 감출 수 있다는 이점을 가진다. 장외거래를 통한 암호화폐 교환이나 현금화 과정은 상황에 따라서 블록체인 원장에 기록되거나 기록되지 않을 수 있다. 예를들어 대면거래를 통해 하드웨어 지갑과 현금을 교환하는 경우라면 해당 거래 사실은 블록체인 원장에 남지 않겠지만, 에이전시에게 자금을 송금하고 현금 등을 돌려받는 경우라면 일부 사실이 블록체인 원장에 남게된다. 따라서 장외거래 추적은 블록체인 원장 분석 만으로는 어려우며 오프체인 분석이 수반되어야 한다.

6. 결 론

본 고에서는 암호화폐 거래 추적을 식별, 어드레스 클러스터링, 자금흐름 추적으로 나누어 주요 기법들과 이슈들을 개괄적으로 살펴보았다.

암호화폐 기술의 발전 방향 중 하나는 프라이버시 강화이다. 이러한 프라이버시 강화 노력은 선의를 가진 원래 목적과는 다르게 각종 범죄에 널리 악용되고 있다. 이러한 상황에서 암호화폐 거래 추적은 범죄 식별과 예방 등 시장의 신뢰를 확보하는데 기여할 수 있는 중요한 기술이다. 이에 대한 관심들이 더 많은 연구로 이어지길 바라 본다.

저 자 약 력

김 태 일

이메일 : brian.kim@coresec.co.kr

- 2011년~현재 코어시큐리티(주) 대표이사
- 2009년~2012년 경찰수사연수원 외래교수
- 2015년~2018년 극동대학교 겸임교수
- 2018년~현재 (사)한국포렌식협회 협력부회장
- 관심분야 : 사이버보안(침해사고조사,위협사냥,교육훈련 및 평가 자동화 기술), 디지털포렌식, 암호화폐 추적