Исследование методов деанонимизации биткоин-кошельков

Антонов Илья Михайлович

Научный руководитель: Подлесных Дмитрий Артурович, каф. ИВМ МФТИ

Московский физико-технический институт

15 апреля 2025 г.

Модель UTXO в Bitcoin

UTXO (Unspent Transaction Output):

- Выход транзакции (TXO) - базовая единица
- Транзакция в блокчейне - набор входных и выходных выходов транзакций
- Биткоин существует в форме выходов транзакций

Структура UTXO и транзакций

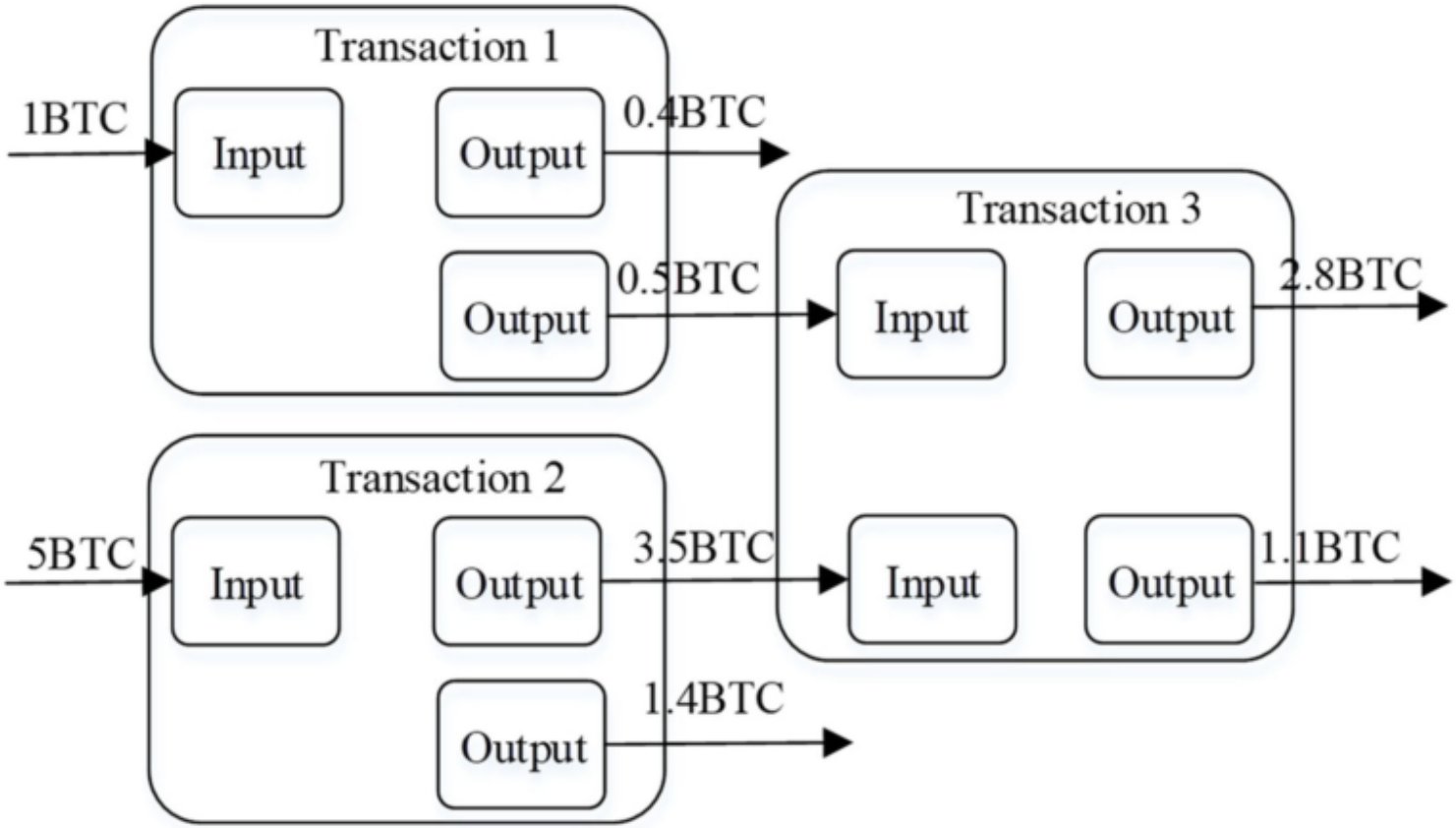

Рис.: Схема модели UTXO: входы ссылаются на предыдущие UTXO, создаются новые выходы

Проблема кластеризации

- **Цель**: идентификация скриптов, принадлежащих одному пользователю/сущности
- **Задача**: построение кластеризации скриптов, где кластеры представляют конечные сущности
- **Метод**: разработка эвристик на основе микроструктуры транзакций
  - Фокус на механике работы с выходами (TXO) и адресами
  - Выявление групп скриптов-кандидатов на объединение

Основные эвристики кластеризации

Эвристика общего входа (Common Input Heuristic)

- Адреса, используемые как входы в одной транзакции, принадлежат одному владельцу
- Основана на необходимости подписывать транзакцию приватными ключами
- **Условие объединения:** Если в транзакции Δ участвуют два или более разных входных скрипта ($n_{in}(\Delta) \geq 2$), все входные скрипты объединяются в один кластер

Основные эвристики кластеризации

Эвристика адреса сдачи (Change Address Heuristic)

- Новый адрес среди выходов, вероятно, адрес сдачи того же владельца
- Основана на практике создания нового адреса для сдачи в большинстве кошельков
- **Условие объединения:** Если транзакция имеет 1) один вход, 2) два выхода, 3) один из выходов новый (не использовался), 4) второй выход был ранее использован, тогда входной адрес и новый выходной адрес объединяются в один кластер

Дополнительные эвристики

- **Эвристика оптимальной сдачи**
  - **Условие объединения:** В транзакции с одним входом и двумя выходами, если один выход имеет "круглое"значение $(v - round(v) < \epsilon)$, а другой нестандартное, то входной адрес и адрес нестандартного выхода объединяются в один кластер

Дополнительные эвристики

Эвристика цепочек очистки (Peeling Chain)

- **Условие объединения:** В последовательности транзакций, где каждая имеет 1 вход и 2 выхода, причем один выход используется как вход в следующей транзакции, все адреса входов и адреса "переходящих"выходов объединяются в единый кластер

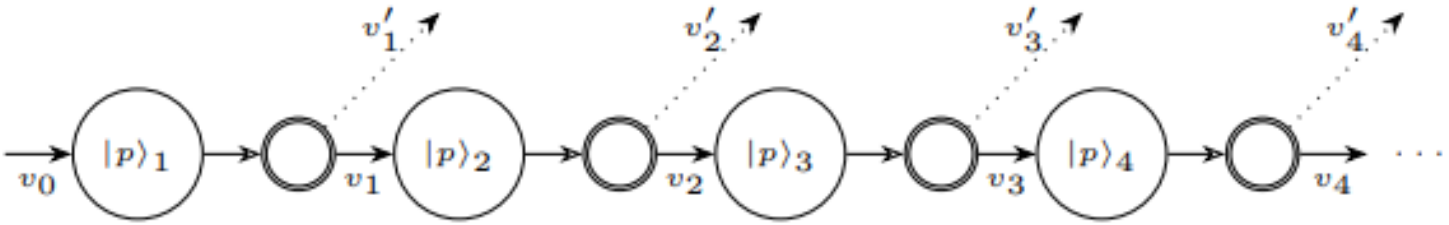

Рис.: Схема цепочки

Методология исследования

- **Используемые данные**:
  - Датасет из исследования (e Elmougy и Liu (ACM SIGKDD '23) - Elliptic++
  - Предварительно размеченные транзакции для оценки точности
- **Процесс анализа**:
  - Последовательное применение эвристик
  - Оценка эффективности каждой эвристики по отдельности

Результаты исследования

- **Исходное количество адресов**: 824,223
- **Результаты кластеризации**:
  - Общий вход: 48 кластеров (сокращение на 99.99%)
  - Адрес сдачи: 76,914 кластеров (сокращение на 90.65%)
  - Повторное использование: 22,996 кластеров (сокращение на 97.21%)
  - Оптимальная сдача: 40 кластеров (сокращение на 99.99%)
  - Цепочка очистки: 130,085 кластеров (сокращение на 84.19%)

Дальнейшие планы исследования

- **Графовый анализ кластеризованных адресов:**
- **Расширение набора эвристик:**

Источники данных и литература

- **Основной источник данных:**
  - Elmougy, Y., Liu, L. (2023). "Demystifying Fraudulent Transactions and Illicit Nodes in the Bitcoin Network for Financial Forensics."In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23). https://doi.org/10.1145/3580305.3599803
- **Литература по эвристикам:**
  - Meiklejohn, S., et al. (2013). "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names."
  - Ron, D., Shamir, A. (2013). "Quantitative Analysis of the Full Bitcoin Transaction Graph."
  - Harrigan, M., Fretter, C. (2016). "The Unreasonable Effectiveness of Address Clustering."