概述：

由於匿名性，比特幣被廣泛用於洗錢和違禁物品的交易。 但是由於每一筆比特幣交易歷史記錄，都保存在一個叫做「區塊鏈（Blockchain）」的公共記錄中，包括管理帳戶的資訊以及交易的數量。 所以比特幣的匿名性只是「偽匿名」，比特幣的交易仍然可以追溯到交易者本身。

但是由於區塊鏈非常龐大（目前為 80GB 量級），對其進行即時查詢、即時統計的過程，需要耗費較多的時間。 最近我將區塊鏈放置在微軟開發的分散式記憶體圖資料庫GraphEngine中，加快了即時查詢和即時統計的速度。

背景：

Blockchain 是記錄所有比特幣交易記錄的一個資料集合。 在blockchain的結構中，從內向外以此為：支付記錄(Transaction)，區塊(Block)，區塊鏈(Blockchain)。 區塊鏈由多個區塊拼接而成，每個區塊中保存著多條支付記錄，每條支付記錄中包含交易的詳細資訊。 區塊鏈示意圖如下：



支付記錄的結構如下：

|  |  |
| --- | --- |
| Time | 支付發生的時間 |
| Hash | 支付的64位hash值 |
| Inputs(可以有多個) | |  |  | | --- | --- | | Addr | 發起方位址 | | tx\_index | 發起方比特幣來源 | |
| outs(可以有多個) | 接收方地址清單 |
| amount | 支付金額 |

可以將支付記錄看做圖的一個節點，二發起方比特幣來源則是指向前一筆支付記錄的邊。

設計與實現：

首先下載所有的區塊，在下載的過程中，刪除無關內容，並轉化為容易反序列化的json格式。

然後逐行讀取支付記錄，將支付記錄存儲在 GraphEngine的存儲模型中，對於此程式設計的存儲模型如下：

然後進行可疑支付追蹤，追蹤包含兩個部分：**靜態追蹤與即時追蹤**。 靜態追蹤指的是根據一些公認的指標，對於所有的支付記錄預先進行可疑判別。 靜態追蹤可用於生成統計結果，預測可疑支付等操作。 即時追蹤指的是使用者輸入某個位址，以及自訂判別條件，系統就能給出此位址相關的可疑支付記錄。

在靜態追蹤中，對於比特幣的可疑支付判斷標準尚未確定，目前滿足以下任一條件則判別為可疑交易：1)交易金額大於50個比特幣; 2)在1秒之內多條支付記錄匯向同一個位址; 3)存在A->B->A的環形支付記錄。

                在動態追蹤中，使用者可以輸入如限制金額，交易間隔等參數進行搜索，介面如下：

結果：

目前下載的block容量為21GB，height為34萬（共40萬+），完成靜態追蹤的初始化耗時1小時。 動態查詢一個Address（局域網下）耗時為毫秒級別。

備註：

難度主要在於靜態處理，需要對千萬級別的節點進行預處理，而預處理的過程中往往會涉及大量與此節點相關的節點（平均達到100個），所以至少進行billion級別的遍歷。

對於動態處理，相比基於檔案系統的搜索，通過 GraphEngine，將資料放入記憶體中，能真正做到常數級別的搜索。