Поиск связей между сущностями в криминалистическом анализе источников данных

Чугаева Татьяна Васильевна

Научный руководитель:
д.ф.-м.н., проф. Терехов Андрей Николаевич
Рецензент:
архитектор ООО “Белкасофт” Тимофеев Никита Михайлович

Введение

- Эксперты используют специальное ПО для криминалистического анализа данных
- Извлекаемые данные разнообразны и их количество может быть довольно объёмным
- Например, при анализе жёсткого диска могут найтись десятки и тысяч электронных писем и миллионы сообщений

Цель работы

Создание модели для нахождения связей между людьми или группами лиц в криминалистическом анализе цифровых источников данных

Задачи

- Создать модель для нахождения связей между людьми или группами лиц в криминалистическом анализе источников данных
- Реализовать модель: полученный граф связей должен использоваться в дальнейшем для выделения сообществ
- Выполнить апробацию модели
- Внедрить модель в коммерческий продукт цифровой криминалистики

Обзор существующих решений

- Мобильный криминалист
- Forensic Toolkit
- UFED Link Analysis
- Nuix
- IBM i2 Analyst’s Notebook

Исходные данные

- Анализ источника данных может привести к обнаружению различных артефактов: сообщения, письма, звонки, изображения, данные реестра и т.д.
- Среди них может найтись список контактов, полученных из истории мгновенных сообщений, адресной книги мобильного телефона

Описание модели

- Сущность – это человек или группа лиц, представленные идентифицирующими данными, такими как адрес электронной почты, номер телефона, имя учётной записи, псевдоним и т.п.
- Каждая сущность может содержать в себе несколько данных одного типа (например, два адреса электорнной почты)

Описание модели

- Взаимодействие между сущностями - звонок, голосовое сообщение, короткое текстовое сообщение, мгновенное сообщение или электронное письмо
- Между двумя сущностями существует связь, если состоялся хотя бы один факт взаимодействия между ними

Описание модели

- Вес связи между сущностями характеризует её значимость и зависит от
  - типа взаимодействия
  - количества взаимодействий
  - продолжительности взаимодействий
- Например, звонок является более значимым типом взаимодействия, чем электронное письмо

Создание контакта для источника данных

- Для каждого источника данных создаётся контакт
- На этапе анализа известны учётные записи и адреса электронной почты, в которые был выполнен вход с анализируемого устройства
- Предполагается, что такие контакты принадлежат владельца устройства

Создание сущностей

- Необходимо найти все контакты, которые предположительно принадлежат одному человеку или группе лиц
- Для этого проводится сравнение контактов по некоторым характеристиками: имя учётной записи, адрес электронной почты, номер телефона, фамилия и имя, псевдоним

Вычисление весов для связей

Вес для связи между сущностями считается на основе весов между их контактами

Вес связи между контактами – среднее арифметическое параметров:

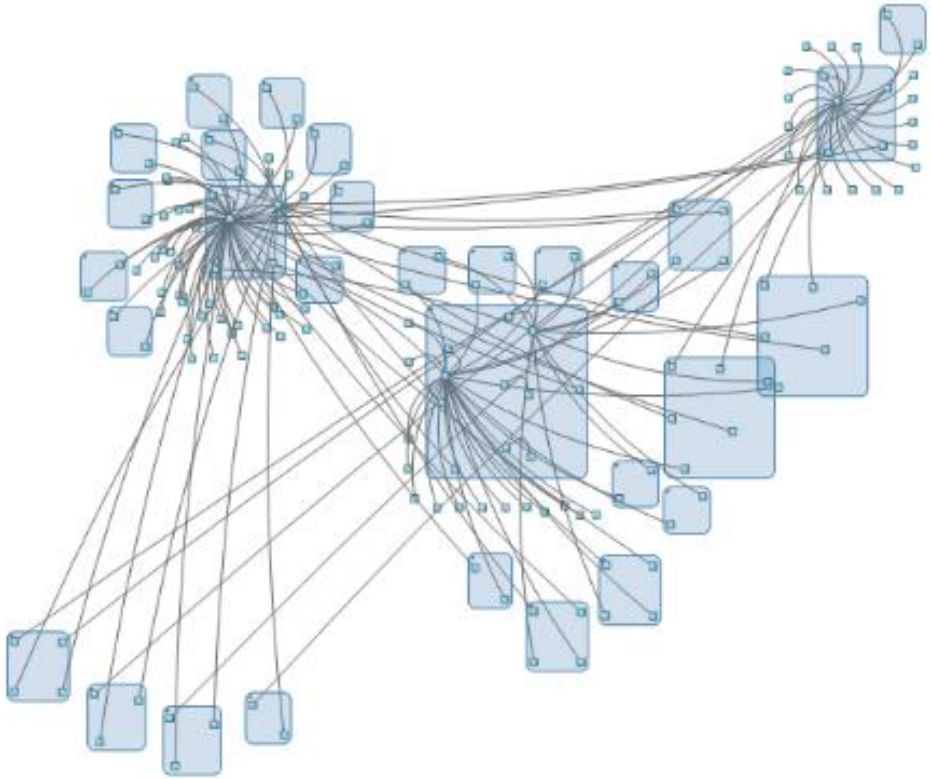
- тип связи: звонок, голосовое сообщение, короткое текстовое сообщение, мгновенное сообщение, письмо (указаны в порядке убывания значимости)
- процент количества взаимодействий
- процент времени взаимодействия

Использование графа связей

- Граф связей – неориентированный взвешенный граф
- На основе его данных можно сделать выводы о структуре взаимодействий, поведении сущностей
- Граф связей используется в задаче выделения сообществ (групп тесно связанных вершин в графе)

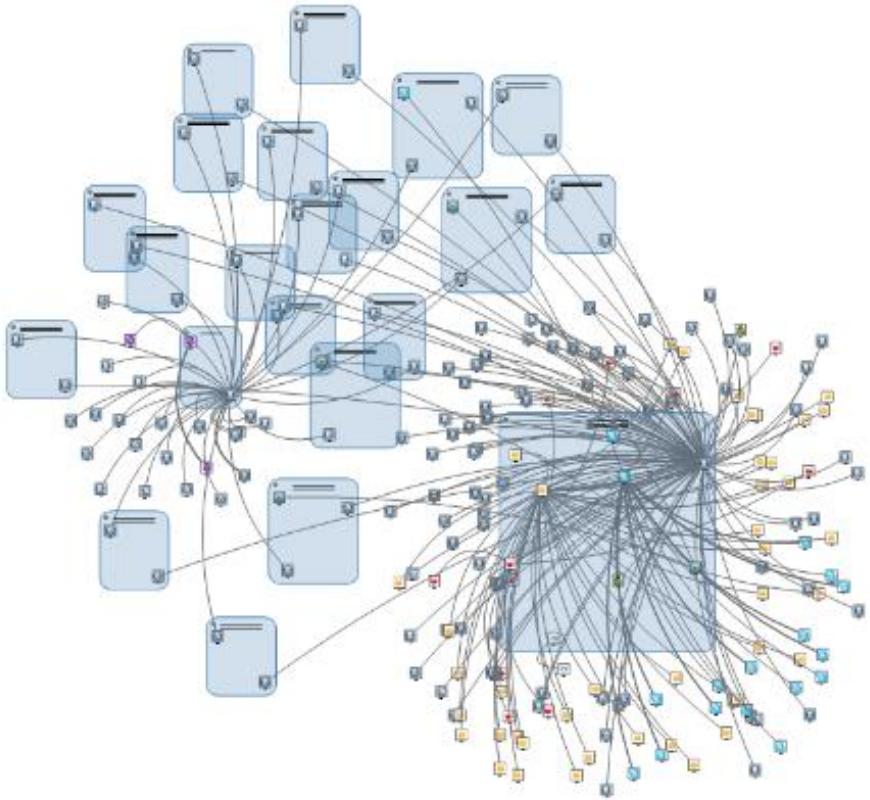
Апробация модели

- 3 учётные записи Skype (471 контакт)



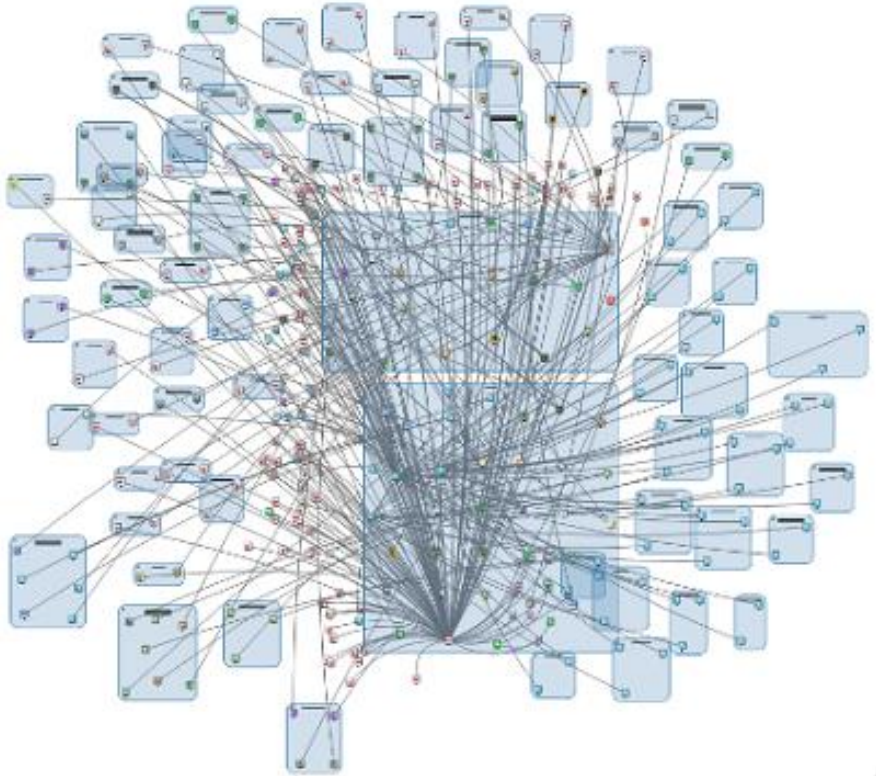
Апробация модели

- 2 резервные копии мобильных телефонов с операционной системой Android (336 контактов)



Апробация модели

- 2 резервные копии мобильных телефонов с операционной системой iOS (561 контакт)



Результаты

- Создана модель для нахождения связей между людьми или группами лиц в криминалистическом анализе источников данных
- Предложенная модель реализована, при этом полученный граф связей используется для выделения сообществ
- Выполнена апробация модели
- Модель внедрена в продукт Belkasoft Evidence Center