ПОИСК СВЯЗЕЙ МЕЖДУ СУЩНОСТЯМИ В КРИМИНАЛИСТИЧЕСКОМ АНАЛИЗЕ ЦИФРОВЫХ ИСТОЧНИКОВ ДАННЫХ

Чугаева Т.В., студентка 661 группы кафедры системного программирования СПбГУ, chugaevatv@gmail.com, Губанов Ю. А., ст. преп. кафедры системного программирования СПбГУ, yuri.gubanov@gmail.com, Тимофеев Н.М., nikita.timofeev@gmail.com

Аннотация

В криминалистическом анализе цифровых устройств участвует большой объём различных данных, разобраться в которых вручную достаточно сложно. Эксперты-криминалисты пользуются специальным ПО, которое позволяет автоматизировать и ускорить анализ.

В данной работе представлена модель, позволяющая упростить одну из главных для экспертов задач: нахождение связей между подозреваемыми.

Введение

Анализ цифровых устройств при расследовании преступлений уже давно стал стандартной процедурой. Странные фразы в истории мгновенных сообщений, удалённые файлы, фотографии сомнительного содержания – всё это может пригодиться для поимки злоумышленника и доказательства его вины в суде [1].

Специальное ПО для криминалистического анализа данных, используемое криминалистами, не только ускоряет процесс нахождения артефактов, но и предоставляет полученную информацию в стандартизированном формате. Такое ПО также гарантирует, что во время анализа устройства данные не были изменены.

В анализе могут участвовать источники данных одного или нескольких лиц. Во втором случае эксперту может понадобиться исследовать взаимодействие группы лиц между собой, выявить наиболее активных участников, основные пути передачи информации и т.д. Однако, так как 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), эксперту будет непросто увидеть целостную картину взаимодействий.

Целью данной работы является создание модели для нахождения связей между людьми или группами лиц в криминалистическом анализе цифровых источников данных.

Исходные данные

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

Описание модели

В первую очередь конкретизируем понятие связи и определим, между кем и когда она может существовать.

В модели используется понятие «сущность» – это человек или группа лиц, представленные идентифицирующими данными (например, адрес электронной почты, номер телефона, имя учётной записи, псевдоним). Каждая сущность может содержать в себе несколько номеров телефонов, имён в социальных сетях и т.д.

Между двумя сущностями может существовать связь, если состоялся хотя бы один факт взаимодействия между ними:

1. звонок
2. голосовое сообщение
3. короткие текстовые сообщения
4. мгновенные сообщения
5. электронное письмо

Каждая связь имеет вес, характеризующий её значимость. Значение веса зависит от типа, количества и времени взаимодействий. Например, звонок считается более значимым типом взаимодействия, нежели электронное письмо. Использование телефона предполагает общение тет-а-тет и большую вовлечённость в диалог, тогда как у письма могут быть несколько получателей, и ответ на него может прийти через несколько часов или даже дней.

Особенности реализации

Реализация предложенной модели состоит из следующих шагов:

1. создание контактов для источников данных
2. выделение сущностей с помощью объединения контактов, принадлежащих одному человеку
3. вычисление весов для связей между сущностями

Создание контакта для источника данных

В рамках данной работы предполагается, что источник данных принадлежит одному человеку. Соответственно, все найденные номера телефонов, почтовые адреса и имена учётных записей относятся к этом же человеку. Например, при анализе ноутбука считается, что все найденные на устройстве учётные записи Skype, с которых был выполнен вход, принадлежат хозяину ноутбука.

На этапе анализа для каждого источника данных создаётся контакт с именем источника данных и дополнительной информацией. Например, при анализе резервной копии данных мобильного телефона создаётся контакт с именем устройства и номером телефона. Соответственно этот контакт будет родительским для остальных контактов, найденных в анализируемом источнике данных.

Выделение сущностей

Алгоритм выделения сущностей работает со списком контактов, найденных во время первичного анализа цифрового источника данных.

Для создания сущности необходимо найти все контакты, которые предположительно принадлежат одному человеку или группе лиц. Для этого проводится сравнение некоторых характеристик контактов (в том виде, в котором они были извлечены): имя учётной записи, адрес электронной почты, номер телефона, имя и фамилия (проверяются вместе), ник. Если совпадает хотя бы одна характеристика, контакты будут объединены в одну сущность.

При сравнении контакта источника данных и его вложенного контакта, они также объединяются в одну сущность. В дальнейшем к этой сущности могут быть добавлены другие контакты.

В итоге каждая сущность представляет собой имя (имя первого добавленного контакта) и список всех её контактов.

Выделение весов для связей между сущностями

Исходные данные, полученные после анализа источника данных, могут содержать список контактов и взаимодействий между ними: звонки, текстовые сообщения, письма и т.п. Как было сказано выше, связь между двумя сущностями существует, если между ними (а точнее между их контактами) было взаимодействие. Определим вес связи, для того чтобы эксперт мог исследовать наиболее значимые отношения между сущностями. Соответственно перед тем как вычислить веса для связей между сущностями необходимо получить веса для связей между их контактами.

Определим вес направленной связи W_oriented для пары контактов (C1, C2) как среднее арифметическое трёх параметров:

- тип связи: звонок, голосовое сообщение, короткое текстовое сообщение, мгновенное сообщение, письмо (указаны в порядке убывания значимости)
- процент количества взаимодействий C1 с C2
- процент времени взаимодействия C1 с C2.

Для устранения случаев незначительных взаимодействий между контактами, добавлены минимальные пороговые значения для общего количества и общего времени взаимодействий (контакта C1).

Соответственно, вес ненаправленной связи W_undirected между контактами C1 и C2 будем считать средним арифметическим весов W_oriented (C1, C2) и W_oriented (C2, C1).

Пусть есть две сущности E1 с контактами C1..Cn и E2 с контактами C1..Cm, где n и m – количества контактов сущностей E1 и E2 соответственно. Тогда вес связи между E1 и E2 будет равен среднему арифметическому весов W_undirected(Ci, Cj), где i1..n и j1..m. У весов для связей между сущностями также существует минимальное пороговое значение.

В рамках данной работы были выбраны именно такие веса для построения неориентированного графа, однако ясно, что аналогичным образом можно получать веса и для направленных связей между сущностями.

Результаты

Результатом предложенной модели является неориентированный граф, в котором вершины и рёбра представляют контакты и связи,

соответственно. В рамках работы было проведено тестирование предложенной модели на данных мобильных телефонов, почтовых ящиков и различных систем мгновенного обмена сообщениями. Пример графа, полученного при анализе нескольких профилей Skype, представлен на Рис. 1.

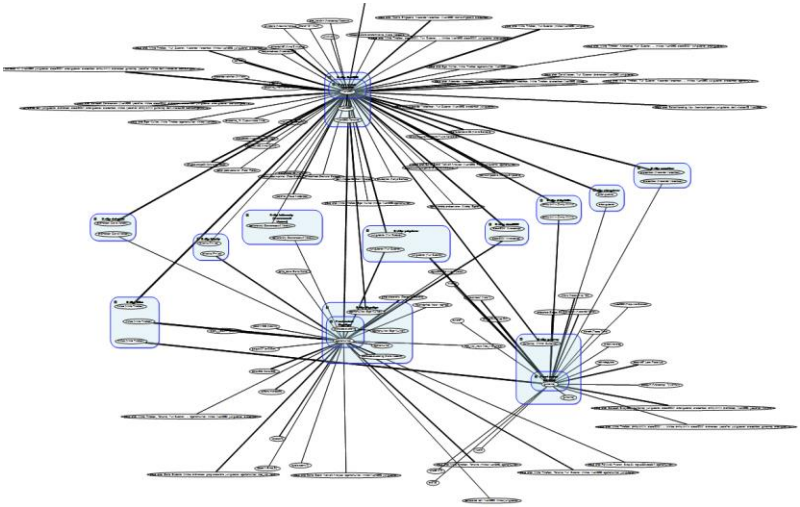

Рисунок 1: Пример графа связей

Контакты, принадлежащие одной сущности или одному контакту источника данных, объединены (см. Рис. 2). Можно увидеть взаимодействия как между сущностями, так и между отдельными контактами.

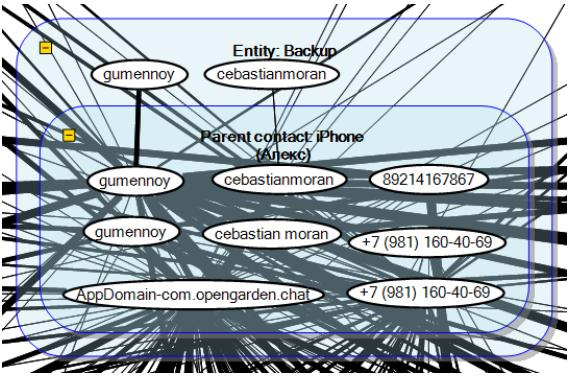

Рисунок 2: Пример сущности, содержащей различные контакты

При визуализации графа было решено устанавливать толщину ребра в зависимости от веса для наглядности и не отображать само значение веса (см. Рис. 3).

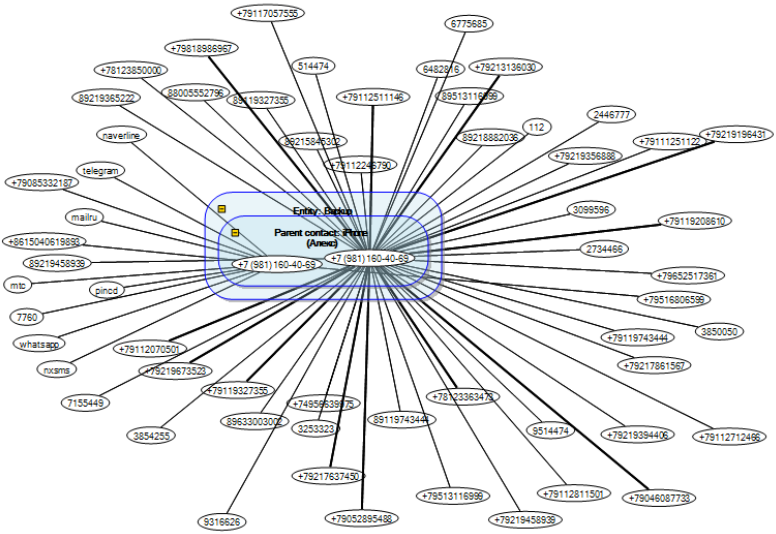

Рисунок 3: Толщина рёбер в графе зависит от веса: самые тонкие рёбра имеют наименьший вес

Представленная модель была реализована и интегрирована в современный продукт компьютерной криминалистики Belkasoft Evidence Center [2], разрабатываемый компанией Belkasoft. Также веса связей между сущностями использовались в дальнейшем для выделения сообществ (групп тесно связанных между собой сущностей) [3].

Заключение

В работе была предложена модель для нахождения связей между людьми или группами лиц в криминалистическом анализе цифровых источников данных. Описаны особенности реализации: выделение сущностей, вычисление весов для связей между ними. Также приведены результаты тестирования модели на различных источниках данных.

Литература

1. Yuri Gubanov Retrieving Digital Evidence: Methods, Techniques and Issues — 2012. https://articles.forensicfocus.com/2012/07/11/retrieving-digital-evidence-methods-techniques-and-issues/ [дата просмотра: 10.04.2016].
2. Belkasoft Evidence Center 2016. https://belkasoft.com/ec [дата просмотра: 10.04.2016].
3. Куликов Е.К., Губанов Ю. А., Тимофеев Н.М. Выделение сущностей в криминалистическом анализе источников данных// Материалы межвузовской научной конференции по проблемам информатики «СПИСОК-2016». — 2016.