Летом 2013 года мир был взбудоражен сенсационной новостью: Эдвард Сноуден, сотрудник Агентства национальной безопасности (АНБ), организовал утечку засекреченной документации, подробно описывающей, как правительство США использует информационные технологии для слежки за потенциальными террористами. Сквозь эту брешь к нам хлынули сведения о том, что секретные службы собирают миллионы телефонных звонков, электронных писем, фотографий и видеороликов, получая их от Google, Facebook, Microsoft и других гигантов в сфере коммуникаций. Но что же делают потом с этой информацией агентства типа АНБ?

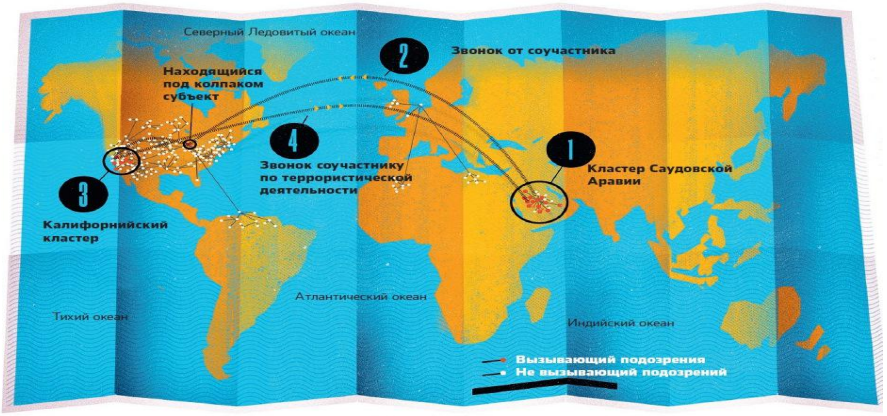
Какие объемы данных мы с вами производим? Согласно недавним исследованиям, проведенным компанией IBM, человечество порождает ежедневно 2,5 квинтиллиона байтов информации. (Если эти байты представить как уложенные плашмя плотно друг к другу монетки, то они бы покрыли весь земной шар в пять слоев.) В эту сумму включается записанная информация — фотографии, видеоролики, сообщения в социальных сетях, текстовые файлы, записи телефонных разговоров, финансовые отчеты и результаты научных экспериментов. Сюда же отнесены и те данные, которые существуют лишь несколько секунд — такие, как содержание телефонных разговоров или чатов по скайпу.

Сбор данных службами безопасности опирается на базовый тезис, что всю их массу можно проанализировать так, чтобы с ее помощью выявить связи между разными людьми. Разбираясь в этих связях, можно найти зацепки для следственных действий.

Главный принцип в обработке данных — снабжение каждого фрагмента меткой, и на основе этих метаданных компьютерные алгоритмы смогут выявлять интересующие службу безопасности связи. Метаданные — это данные, описывающие другие данные. Таковы, к примеру, имена и размеры файлов на вашем компьютере. В цифровом мире этикетка, наклеенная на фрагмент данных, будет называться меткой. Снабжение данных меткой — это обязательный первый шаг в их обработке, поскольку именно метка позволяет аналитику (или его программе) классифицировать и организовывать имеющуюся информацию для ее дальнейшей обработки и анализа. Метки позволяют совершать манипуляции с фрагментами данных, не вникая в их содержание. Это очень важный юридический момент в работе службы безопасности, поскольку закон США не позволяет вскрывать переписку граждан США, равно как и иностранцев, пребывающих в стране на законных основаниях, без соответствующего ордера.

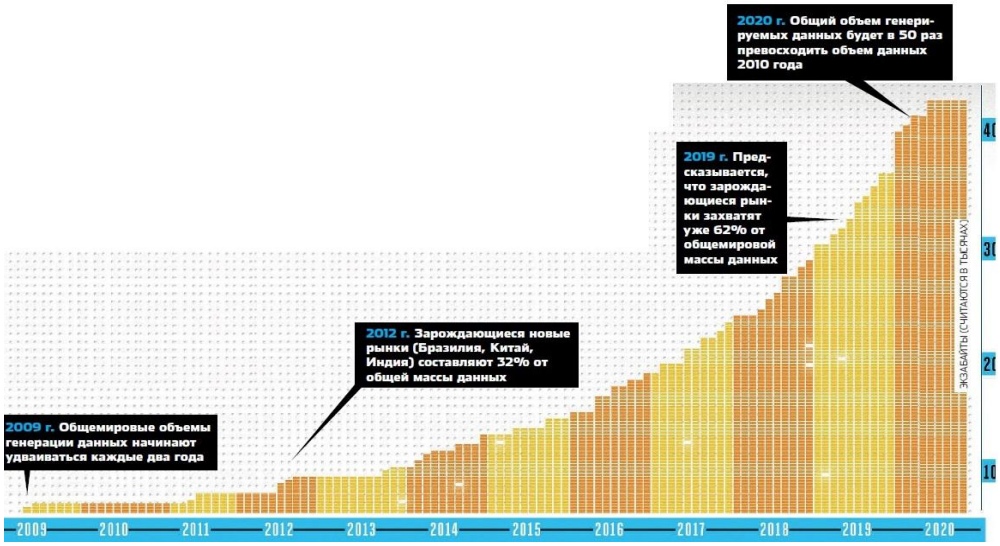
Компания IDC, занимающаяся анализом данных, сообщает, что только 3% всей информации, циркулирующей в компьютерном мире, при создании сопровождается присвоением какой-либо метки. Поэтому АНБ использует особую, весьма сложную программу, которая на всю собранную информацию «наклеивает» соответствующие метки. Они представляют собой основу для любой системы, устанавливающей связи между различными видами данных — например, между видеофайлами, документами и записями телефонных разговоров. Скажем, система обработки данных может привлечь внимание следствия к подозреваемому, который вывешивает в сети террористическую пропаганду, посещает сайты, где описана технология изготовления самодельных взрывных устройств, и вдобавок покупает скороварку. (Эта схема соответствует поведению братьев Царнаевых, которых обвиняют в теракте на Бостонском марафоне.) Подобная тактика основывается на предположении, что террористов отличают специфические профили данных, хотя многие эксперты ставят это предположение под сомнение.

АНБ собирает метаданные по телефонным переговорам. Эти метаданные позволяют выявить террористов, не вникая в содержание самих переговоров. Среди миллионов звонков можно нащупать определенные паттерны, как это иллюстрирует приведенный на фото сценарий. 1. Звонок из Саудовской Аравии от известной организации, поддерживающей терроризм, адресованный в кластер возможных сообщников. 2. Звонок от организации, известной своей террористической деятельностью, адресованный гражданину США, привлекшему внимание Агентства национальной безопасности. 3. Метаданные по телефонным переговорам, которые ведет подозреваемая личность, формируют кластер сообщников в Калифорнии. 4. Детализация телефонных разговоров показывает, что один из сообщников в Калифорнии связывается с кем-то в кластере Саудовской Аравии. NSA привлекает внимание ФБР к этой связи и получает право на прослушку этой линии.



АНБ — крупный заказчик программного обеспечения, позволяющего работать с большими базами данных. Одна из таких программ носит имя Accumulo. Прямых доказательств, что ее используют для слежки в системах международной связи, нет, а создавалась она именно для снабжения метками миллиардов разрозненных фрагментов данных. Это «секретное оружие» службы безопасности, созданное средствами программирования Google, написано открытым кодом. В нынешнем году компания Sqrrl выпустила эту программу на рынок и надеется, что ею заинтересуются в сфере здравоохранения и финансов для работы с огромными массивами рабочих данных.

АНБ имеет право перлюстрировать международные каналы связи и собирает огромнейшие объемы данных. Это триллионы фрагментов различных сообщений, которые люди пишут по всему свету. Агентство не занимается охотой на преступников, террористов или шпионов, которых выявляют с помощью его работы, а просто сливает полученную информацию другим правительственным службам — Пентагону, ФБР и ЦРУ. Далее работа ведется по такой схеме. Сначала один из 11 судей секретного суда FISA (Foreign Intelligence Surveillance) принимает от государственного агентства запрос на разрешение переработать определенные данные, полученные АНБ. Получив разрешение (а с этим, как правило, проблем не бывает), запрос сначала переадресуют в отдел ФБР по контролю за электронными средствами связи (ECSU). Этот ход должен обеспечить юридическую корректность — агенты ФБР проверяют запрос и подтверждают, что объектом слежки не являются граждане США. ECSU переадресовывает такой же запрос в отдел ФБР по методам перехвата данных. Там получают информацию с интернет-серверов и передают ее в АНБ, чтобы там ее пропустили через свои программы переработки данных. (Многие компании, работающие в сфере коммуникаций, отрицают тот факт, что их серверы открыты для доступа со стороны АНБ. Федеральные чиновники, напротив, сообщают о фактах такого сотрудничества.) И наконец, АНБ передает соответствующую информацию в то правительственное агентство, от которого поступил запрос.



Что же замышляет АНБ?

Неприятности у АНБ начались с того момента, когда Сноуден открыл всему миру факт, что правительство США собирает метаданные по телефонным переговорам всех клиентов оператора Verizon, причем в их число входят и миллионы американцев. В ответ на запрос ФБР судья FISA Роджер Уилсон издал постановление, обязывающее компанию Verizon передавать в ФБР детализацию всех телефонных разговоров. В АНБ подобную практику называют «системой раннего предупреждения», которая позволяет обнаруживать террористическую деятельность.

Не успело общество переварить сведения о метаданных, как Сноуден обрушил на него рассказ о еще одном направлении в работе АНБ, имеющем обозначение US-984XN. Каждая поисковая платформа, каждый источник сырой разведывательной информации получает свое обозначение — SIGAD (Signals Intelligence Activity Designator, «указатель разведдеятельности») и кодовое имя. Служба SIGAD US-984XN известна нам по чаще упоминаемому кодовому имени — PRISM. Система PRISM — это сбор цифровых фотографий, где-то хранящихся и куда-то пересылаемых файлов, электронных писем, чатов, видеороликов и видеопереговоров. Эта информация изымается у девяти ведущих интернет-компаний. В правительстве США утверждают, что именно эти мероприятия помогли схватить Халида Уаззани, натурализованного гражданина США, которого ФБР обвиняет в планах взорвать Нью-Йоркскую фондовую биржу.

Схемы, обнародованные Сноуденом, показывают, что АНБ, помимо всего прочего, использует в своей деятельности средства слежки, работающие в режиме реального времени. Аналитики агентства могут получать оповещения о подключении пользователя к сервису или отсылке письма, а также о входе в тот или иной чат.

В июле Сноуден опубликовал сверхсекретный доклад, в котором описано программное обеспечение, позволяющее просматривать сотни различных баз данных. Сноуден утверждает, что эти программы позволяют аналитику самого низшего уровня бесконтрольно вмешиваться в чужие процессы обмена информацией. В докладе приведены примеры: «Мой клиент говорит по-немецки, но находится в Пакистане. Как мне его найти?» или «Мой клиент для поиска своих целей использует GoogleMaps. Можно ли воспользоваться этой информацией для определения его email-адреса?» Описанная программа позволяет, задав один такой вопрос, одновременно провести поиск по 700 серверам, раскиданным по всему свету.