Алекс («Сэнди») Пентленд

Как защитить большие данные от самих себя

Три совета, призванные уберечь конфиденциальную информацию от всевидящего ока государства

Об Агентстве национальной безопасности на протяжении первых десятилетий его существования было мало что слышно. Главная задача АНБ — шпионить за Советским Союзом. Враг был четко обозначен и хорошо различим. Арсенал спецсредств в те времена использовался довольно ограниченный: прослушивающие устройства для телефонов, самолеты-разведчики, скрытые микрофоны — вот, вроде, и все.

Однако после событий 11 сентября 2001 г. ситуация изменилась. Теперь главным врагом АНБ стала разветвленная сеть террористов-одиночек. Отныне любой гражданин любой страны становился потенциальной мишенью для негласного наблюдения со стороны АНБ. По мере появления новых цифровых технологий передачи информации стала меняться и сущность шпионажа. Мир наводнили различные гаджеты и девайсы, способные выходить в Интернет. И тут оказалось, что традиционные методы работы и инструменты из арсенала Агентства национальной безопасности безнадежно устарели.

Пытаясь справиться с новыми реалиями, АНБ решило использовать следующий принцип: «Собирай все, что можешь!» Бывший директор агентства Кит Александер (Keith Alexander) как-то однажды высказал аналогичную мысль: по его мнению, чтобы найти иголку в стоге сена, нужно взять сразу весь стог. Вот тогда-то АНБ и кинулось записывать телефонные разговоры практически всех американцев и собирать информацию чуть ли не обо всех гражданах других государств. В скором времени дошло до того, что в АНБ каждые два часа накапливались колоссальные залежи информации, сравнимые по объему разве что с данными статистического ведомства США.

Illustration by Bill Mayer

ОБ АВТОРЕ

Алекс («Сэнди») Пентленд (Alex "Sandy" Pentland) — глава Лаборатории человеческого развития при Массачусетском технологическом институте; один из руководителей программы Всемирного экономического форума по исследованиям больших данных и персональной информации; его самая свежая книга, «Социальная физика» (*Social Physics*), вышла в начале этого года.

Единственным местом, пригодным для хранения этого гигантского информационного «стога сена», оставалась традиционная инфраструктура АНБ — его защищенные от постороннего глаза «закрома». Объемы собранных данных были поистине фантастическими — именно в этом-то и был их минус, поскольку персональная информация почти о каждом жителе нашей планеты вдруг стала доступна любому аналитику АНБ, стоило ему только захотеть с ней ознакомиться. Получился парадокс: шпионское ведомство США стало еще более беззащитным, чем когда-либо, поскольку оказалось не в силах предотвращать утечки информации, а причина этого крылась как раз в гигантских объемах собранных данных. В результате Эдварду Сноудену, которого возмутил размах скрытой деятельности АНБ, удалось-таки скачать тысячи секретных файлов с сервера, расположенного на Гавайях, а потом сбежать в Гонконг и передать секретные документы представителям СМИ.

Государство и бизнес всегда используют в своей деятельности информацию о гражданах, например данные статистического ведомства, — в этом нет ничего нового. Другое дело, когда ведомство, отвечающее за сбор разведывательной информации, закачивает данные на скрытые от посторонних глаз сервера и практически бесконтрольно ими распоряжается: ничего подобного раньше не происходило. Поэтому вовсе не удивительно, что сделанные Сноуденом разоблачения вызвали бурю негодования в обществе.

До сих пор, когда речь заходила о шпионской деятельности АНБ, основное внимание уделялось главным образом моральным и политическим аспектам проблемы, в то время как технические вопросы фактически оставались за кадром. Выяснилось, что не только государственная политика по сбору и обработке больших данных была неадекватной, но и сам процесс ее формирования постоянно запаздывал; но такого быть не должно, ведь государство обязано идти в ногу с научно-техническим прогрессом. Словом, простых решений нет. И все же в настоящей статье мы предлагаем некоторые меры, которые смогут хоть как-то исправить ситуацию.

ОСНОВНЫЕ ПОЛОЖЕНИЯ

- Государство и бизнес всегда хотели получать конфиденциальную информацию о частной жизни граждан. Но смогут ли они это делать и впредь, не нарушая неприкосновенности частной жизни?
- Предлагаются следующие основные принципы: большие данные распределяются по локальным базам, которые подконтрольны отдельным организациям, а не АНБ или прочим госструктурам; все, кто осуществляет операции с персональной информацией (в том числе простые граждане), обязательно должны использовать криптографическую защиту.
- В эпоху цифровых технологий традиционные методы и подходы могут не сработать. Поэтому новаторство, эксперимент и открытость при работе с большими данными — единственные условия поиска эффективных решений.

ШАГ 1

ЛИКВИДАЦИЯ «СТОГА СЕНА»

И все-таки Александер был неправ относительно метода поиска иголки в стоге сена. Вовсе не нужно брать весь стог, достаточно проверить любую из его частей. Нет необходимости хранить гигантские объемы данных в «одной корзине» — это рискованно и для шпионов, и для тех, за кем они наблюдают. При таком положении вещей возрастает вероятность утечек информации, содержащей государственные секреты, а над обычными гражданами нависает угроза нарушения неприкосновенности частной жизни.

Разоблачения Сноудена продемонстрировали следующее: если информация находится в руках у государства, то степень ее централизации возрастает. Агентству национальной безопасности и другим госструктурам не нужно хранить все информационные ресурсы непременно в своих электронных «закромах»; вместо этого контроль над базами данных и шифрование надо передать той организации, которая их создала. Различные виды данных должны храниться отдельно друг от друга: скажем, финансовая информация — в одной базе, а сведения о состоянии здоровья гражданина — в другой, причем данные о физических лицах следует выделить

в особую категорию. В любом случае при наличии юридических оснований АНБ или любое другое ведомство всегда сможет получить доступ к любому из сегментов информационного «стога сена»; другое дело, что информация больше не будет храниться в одном-единственном кластере серверов.

Итак, необходима децентрализация баз данных, для чего их надо перестать «складировать» в едином центре; пусть телекоммуникационные и интернет-компании хранят всю информацию у себя, на своих серверах. В то же время нет никакой необходимости ликвидировать хранилища данных АНБ: и архивы, и программное обеспечение быстро устаревают.

Агентство национальной безопасности вряд ли добровольно согласится свернуть свою деятельность по сбору информации; следовательно, его надо к этому подтолкнуть при помощи специальных законодательных актов, включая директивы исполнительной власти. В результате выиграет само же АНБ, руководство которого, кажется, это хорошо понимает. Так, выступая на форуме по безопасности, состоявшемся летом прошлого года в Колорадо под патронажем Института Аспена, Эштон Картер (Ashton B. Carter), занимавший на тот момент пост заместителя министра обороны США, отозвался о причинах крупных неприятностей АНБ следующим образом: «Этот провал (т.е. утечка, организованная Сноуденом) возник из-за того, что использовались два подхода, которые необходимо пересмотреть <…>. Огромный массив информации был сосредоточен в одном месте. И это было ошибкой. И второе: у вас работал человек, которому были даны очень широкие полномочия, позволявшие получать доступ к той информации и пересылать ее. Такие ошибки больше повторять нельзя». Если хранить данные в разных «корзинах», используя криптографическую защиту, это не только предотвратит хищение информации такими как Сноуден, но и защитит ее от кибератак. Почему? Потому что в таком случае любому пользователю будет доступен не весь массив данных, а лишь отдельный его сегмент. Даже авторитарные государства должны быть заинтересованы в децентрализации, рассредоточении данных, ведь в противном случае эти режимы вероятнее всего ожидает печальная участь.

Каким же образом распределение данных по многочисленным узлам сети помогает обеспечивать конфиденциальность информации? Дело в том, что подобная архитектура позволяет отслеживать связи между базами данных и пользователями, ведь при проведении любой операции по обработке данных (поиск по базе, статистическая обработка массивов и т.д.) возникают свои характерные особенности выстраивания связей — свой, так сказать, рисунок, характерный для каждого конкретного взаимодействия с базами данных. Эти особенности взаимодействия — метаданные о метаданных — позволяют выявлять общие закономерности обмена конфиденциальной информацией. Как это понимать?

Для ясности приведем такую аналогию. Если нам, скажем, известны структура компании и особенности формирования связей между всеми ее подразделениями (т.е. мы знаем, как обычно должна перемещаться почта между подразделениями), тогда нам будет известна и информация о ежедневных потоках почты (исключая ее содержание) между сотрудниками компании. И вот если, например, оператор, который ведет базу данных, содержащую конфиденциальную информацию о состоянии здоровья работников, зафиксировал внезапный всплеск обращений к этой базе со стороны, например, финансового отдела компании, то он вправе задать вопрос, чем обусловлен этот всплеск. Точно такой подход можно использовать и в отношении операций с большими данными: если во время этих операций генерируются метаданные о метаданных, то появляется возможность контроля над потоками информации. В таком случае информацию смогут отслеживать телекоммуникационные компании, а за деятельностью АНБ теперь сможет наблюдать не только пресса, но и общественные организации. Благодаря использованию метаданных о метаданных граждане смогут поступать с АНБ точно так же, как эта организация привыкла вести себя по отношению ко всем остальным.

ШАГ 2

УКРЕПИТЬ КАНАЛЫ ПЕРЕДАЧИ ДАННЫХ

Итак, децентрализация, рассредоточение данных АНБ — это всего лишь один из способов обеспечить конфиденциальность в мире, перенасыщенном информацией. Для предотвращения утечек не менее важно обеспечить с помощью шифрования безопасные передачу и хранение данных, что особенно актуально на фоне активизации глобальной киберпреступности и возникновения угрозы кибервойн.

Все, кто использует персональные данные, будь то госструктуры, частные предприятия или отдельные лица, должны соблюдать несколько основных правил безопасности. Во-первых, внешний обмен данными должен осуществляться только между теми информационными системами, которые придерживаются сходных стандартов безопасности. Для того чтобы отслеживать, откуда информация поступает и куда передается, каждой операции должна соответствовать надежная цепочка используемых идентификаторов и прав доступа. При передаче информации любого вида нужно всегда проводить мониторинг метаданных и проверку, подобно тому как осуществляется верификация кредитных карт для предотвращения мошенничества.

И здесь вполне подойдет так называемая сеть доверия — компьютерная сеть, которая осуществляет верификацию пользователя и выносит

решение, имеет ли он право доступа к информации, какие операции ему разрешается осуществлять и как поступить в случае, если пользователь нарушил правила доступа. В сетях доверия сохраняется вся информация о верификации пользователей и о правах доступа, благодаря чему данный вид сетей поддается контролю и мы всегда можем проверить, насколько соблюдаются соглашения об использовании данных.

Со временем сети доверия доказали свою надежность и безопасность. Наиболее известная из них — Общество всемирных межбанковских финансовых телекоммуникаций *(SWIFT)*; около 10 тыс. банков и других организаций используют ее для денежных переводов. Насколько мы знаем, эту систему еще никому не удалось взломать, что и стало ее отличительной особенностью. Когда-то известного грабителя Вилли Саттона (Willie Sutton) спросили, почему он грабил банки, на что он якобы по-простецки ответил: «Потому что там лежат деньги». Но времена изменились: в наше время деньги находятся в системе *SWIFT*, внутри которой ежедневно перемещаются триллионы долларов. Благодаря мониторингу метаданных, автоматизированному контролю и принципу солидарной ответственности система *SWIFT* остается мошенникам не по зубам, а клиенты всегда могут быть уверены, что денежный перевод обязательно дойдет до адресата.

В свое время сети доверия были дорогостоящими и сложными в эксплуатации. Однако благодаря глобальному снижению стоимости вычислительных мощностей сети доверия стали вполне по карману небольшим организациям и даже физическим лицам. Возглавляемая мной исследовательская группа из Массачусетского технологического института вместе с Институтом систем информационного проектирования участвовала в создании открытого хранилища персональных данных *(PDS)*, а именно коммерческой версии системы данного типа. Идея, лежащая в основе программного обеспечения, которое мы сейчас обкатываем совместно с различными отраслевыми и государственными партнерами, состоит в том, чтобы как можно шире распространить принципы обеспечения информационной безопасности по подобию системы *SWIFT* с тем, чтобы компании, местные органы власти и частные лица смогли спокойно обмениваться конфиденциальной информацией — любой, в том числе медицинской и финансовой. Власти некоторых штатов США тоже начинают присматриваться к нашим разработкам, чтобы потом использовать их во внутренних и внешних сервисах анализа данных.

Благодаря более широкому распространению сетей доверия вести обмен данными юридическим и физическим лицам стало намного безопаснее. В результате стало легче внедрять безопасные распределенные системы хранения данных, которые помогают снизить угрозы, связанные с незаконным доступом к большим данным, принадлежащим организациям, и конфиденциальной информации простых граждан.

ШАГ

3

ПРОДОЛЖАЙТЕ ЭКСПЕРИМЕНТИРОВАТЬ

А теперь сделаем последний и самый главный вывод: проблема безопасности информации окончательно не решена. Признаем это. Понятно, что технологии развиваются, изменяются, а вслед за ними должны меняться и структуры, осуществляющие надзорные функции. Наступившая цифровая эра представляет собой, по сути, нечто совершенно новое в человеческой истории. В наше время уже нельзя полагаться только на традиционные методы и подходы — необходимо постоянно предлагать новые идеи и проверять их на практике.

Давление, оказанное на США со стороны других государств, отдельных граждан и высокотехнологичных компаний, уже побудило Белый дом несколько ограничить деятельность АНБ. Скажем, в попытке восстановить пошатнувшееся было доверие некоторые компании уже пытаются в судебном порядке получить информацию о запросах, поступавших от АНБ, — т.е. те самые метаданные о метаданных. А в мае Палата представителей уже приняла Закон о свободе. И хотя многие защитники принципа неприкосновенности частной жизни сочли его слабоватым, данный правовой акт ограничивает сбор данных, так сказать, оптом, т.е. всех подряд; кроме того, в законе сказано о необходимости поддерживать некоторую степень прозрачности в процессе сбора информации. (Законопроект находится на рассмотрении в Сенате.)

Перечисленные в статье шаги ведут нас в правильном направлении. Однако любые действия, предпринимаемые в настоящее время, следует рассматривать в качестве временного решения глобальной проблемы. Технологии постоянно развиваются и непрерывно изменяются, поэтому инновационная активность государства тоже не должна отставать от научно-технического прогресса. В конечном счете мы обязаны постоянно экспериментировать; необходимо понемногу продвигаться вперед, браться за новые проекты, чтобы понять, что работает, и отделить от того, что не работает, — вот главная наша задача. ■

Перевод: И.В. Ногаев

ДОПОЛНИТЕЛЬНЫЕ ИСТОЧНИКИ

■ Пентленд А. В цифровом обществе // ВМН, № 1, 2014.

■ Personal Data: The Emergence of a New Asset Class. World Economic Forum, January 2011. www.weforum.org/reports/personal-data-emergence-new-asset-class

■ Social Physics: How Good Ideas Spread — The Lessons from a New Science. Alex Pentland. Penguin Press, 2014.