Big Data:
риски, угрозы, защита

Александр ВЕСЕЛОВ,
ведущий инженер,
ЗАО «С-Терра СиЭсПи»

Андрей ШПАКОВ,
ведущий инженер,
ЗАО «С-Терра СиЭсПи»

С повсеместным распространением Интернета и других современных технологий объемы доступной для любого человека информации значительно увеличились. Ранее актуальными были вопросы «существует ли в принципе такая информация?» и «где ее можно найти?», сейчас основной вопрос – «как выделить из колоссального количества информации действительно нужную?». Мы неоднократно убеждаемся в справедливости так называемого закона Старджона: «90 процентов чего угодно — ерунда» («Ninety percent of everything is crap»). Неизменной во все времена остается ценность информации.

На сегодняшний день нет точного определения больших данных. В некоторых ситуациях под Big Data понимают информацию достаточно большого объема, хранящуюся в системе хранения данных (СХД), но назвать порог объема, при котором данные считаются «большими», никто не берется. На наш взгляд, наиболее точное определение звучит следующим образом: большие данные – это данные, которые невозможно обработать вручную, и их совместное рассмотрение показывает нам то, чего не видно по отдельности.

Одним из наиболее ярких примеров больших данных является информация, содержащаяся в социальных сетях. Снижение уровня анонимности приводит к увеличению объема персонализированной информации. Но данные конкретного профиля не так интересны, как выборка по сотням тысяч записей с анализом совпадений, связей между ними и т. д. Эта информация может использоваться и для проведения различных исследований (социальных, политических и т. п.), для помощи людям, повышения удобства использования сервисов пользователями, и для менее позитивных задач. На основе такой информации может предлагаться целевая реклама в виде баннеров, спам-рассылок и т. п. Производители навигационных систем могут пускать пользователей по «рекомендованным» маршрутам «с платными дорогами» без реальной необходимости. В более серьезных случаях трафик, полученный в процессе сбора данных геолокации, можно использовать для слежки за группами людей, определения размера этих групп, а также их активности. Новые «веяния» нашего законодательства, в частности обязательная идентификация в публичных Wi-Fi-сетях, будут снижать степень анонимности все сильнее и сильнее.

Как мы видим, большие данные (публичные и частные) можно использовать и в хороших, и в сомнительных целях. Защитить публичные данные от анализа невозможно в принципе, но каждый пользователь способен повысить свой уровень анонимности (если не хочет, чтобы информация о нем попала в упомянутые выше базы данных и исследования).

Со сведениями, находящимися в ограниченном доступе, ситуация иная – им необходима защита. Раскрытие таких данных и их последующий анализ могут привести к финансовым, репутационным, юридическим и другим проблемам. Как правило, информация подобного рода хранится в центре обработки данных (ЦОД), к тому же довольно часто ресурсы дублируются в резервном ЦОД. Тем самым обеспечивается катастрофоустойчивость информационной системы: при выходе из строя основного ЦОД данные остаются доступными в резервном.

Безопасность информации, которой оперирует ЦОД, должна обеспечиваться на всех стадиях ее использования. Организовать защиту больших данных при передаче между ЦОД так же важно, как

и защитить трафик при обращении пользователей к информации ЦОД. Но если в случае с трафиком пользователей говорить о больших данных не совсем уместно, то данные, которые передаются при взаимодействии между ЦОД, в полной мере соответствуют понятию «большие данные». Именно поэтому для репликации используются высокопроизводительные каналы связи (10 Гб/с и выше). Для защиты информации при передаче по таким каналам существуют специализированные канальные шифраторы, производимые как западными, так и российскими компаниями. Курс на импортозамещение, который декларируется в России сегодня, уже определяет, какому оборудованию стоит отдать предпочтение. К тому же многие данные (например, персональные) подлежат обязательной защите в соответствии с законодательством РФ, следовательно, не просто желательно, а необходимо использовать сертифицированные отечественные СКЗИ (средства криптографической защиты информации) высокой производительности. Не следует забывать и о влиянии характеристик среды передачи – «темной оптики» довольно большой длины. Защита такого канала традиционными криптографическими средствами представляет собой сверхдорогую задачу и применяется крайне редко. Средств защиты, учитывающих специфику защиты трафика СХД, на рынке не так много, но они уже есть (например, шлюзы C-Терра) и успешно используются для решения указанных задач.

Похожая ситуация возникает при миграции ЦОД. Ни для кого не секрет, что многие организации переводят свои ЦОД из региональных центров на территории с меньшими накладными расходами (на аренду, охлаждение, персонал и т. д.). Соответственно встает вопрос организации такой процедуры переезда, при котором деятельность ЦОД не будет остановлена ни на минуту. Ведь ЦОД нельзя остановить, погрузить в грузовики, перевести на новое место и запустить – к его ресурсам должен быть постоянный гарантированный защищенный доступ. Причем пользователи этих ресурсов чаще всего расположены в разных территориальных и часовых поясах. При переезде ЦОД неизбежен момент, когда часть ресурсов уже переехала в новое хранилище, а другая находится в старом. В таком случае используются решения по миграции от ведущих производителей СХД, но есть одна фундаментальная проблема – эти решения являются проприетарными и не совместимы друг с другом. Здесь помогут комплексы, функционирующие на более низких уровнях, в частности сетевом (L3) и канальном (L2).

Если использовать собственный ЦОД невозможно, многие компании переходят к хранению своих данных в инфраструктурном облаке (IaaS). При этом возникает вопрос сохранности данных и быстрой миграции к другому поставщику услуг. В подобных ситуациях целесообразно использовать виртуализированные решения по защите каналов связи. Они позволяют передать конфиденциальные данные между облаками и осуществить репликацию на локальное хранилище. Безусловно, при обработке информации, подлежащей обязательной защите в соответствии с законодательством РФ, помимо защиты канала связи необходимо учесть требования регуляторов к защите виртуальной среды. В зависимости от различных условий эти технические и/или организационные меры могут быть реализованы как пользователем, так и ее поставщиком.

Защитив взаимодействие между компонентами системы хранения, не нужно забывать о других аспектах защиты (например, пограничных межсетевых экранах, обеспечивающих защиту от сетевых атак, инспекцию трафика и т. д.).

> Организовать защиту больших данных при передаче между ЦОД так же важно, как и защитить трафик при обращении пользователей к информации ЦОД.

Как и практически любая информация, большие данные могут использоваться не только для пользы, но и во вред. Защитить публичные данные в целом от анализа невозможно, но каждый конкретный пользователь может повысить свой уровень анонимности и не участвовать в подобных исследованиях. Данные ограниченного доступа требуют комплексной защиты, особенно при передаче между территориально распределенными объектами. Если данные подлежат обязательной защите в соответствии с законодательством РФ (например, ПДн), то при построении системы защиты нужно учитывать требования регуляторов. ■