Анализ больших данных в ИБ предприятий. Перспективы развития

En **Big Data Analysis in the Information Security of Enterprises. Development Prospects**

N. I. Kasperskaya,
President
InfoWatch
Natalya.Kaspersky@infowatch.com

Big Data is a hot topic in high technology. At the same time, this topic does not correspond very often to the field of information security (IS), although initially the term "big data" was created to solve IS problems.

Today, behavioral analysis systems implemented to prevent risks and threats to information security of enterprises are a very promising direction. In the near future such systems will be widely demanded in companies from different verticals. These products will allow enterprises to move from the paradigm of responding to information security threats to early detection and prediction of these threats, to the prevention of information security risks. The technical basis for such solutions is now the technology of analyzing threats to the information security of enterprises.

Keywords: random antennas, confidential information security, simulation interference, module-repeater

Большие данные (Big Data) – актуальная тема в области высоких технологий. При этом с областью обеспечения информационной безопасности (ИБ) данная тематика соотносится не слишком часто, хотя изначально сам термин « большие данные» возник для решения проблем ИБ.

Сегодня системы поведенческого анализа, внедряемые с целью предотвращения рисков и угроз информационной безопасности предприятий, – весьма перспективное направление. В самое ближайшее время такие системы будут широко востребованы в компаниях из различных вертикалей. Эти продукты позволят предприятиям перейти от парадигмы реагирования на угрозы информационной безопасности к раннему обнаружению и прогнозированию данных угроз, к предотвращению ИБ-рисков. Технической основой для такого рода решений сейчас выступают технологии анализа угроз информационной безопасности предприятий.

DLP (Data Leakage Prevention), HBAS (Human Behavior Analytical System), SIEM (security information and event management), information, restricted access information, information security incident, negative consequences, threat, risk, information security risk, event, social communication, information security event

Наталья Ивановна Касперская,
президент
ГК InfoWatch
Natalya.Kaspersky@infowatch.com

Общие положения

Определение больших данных

О больших данных (БД) много говорится последние 8–10 лет, но достаточно мало – применительно к области информационной безопасности (ИБ). Хотя, на мой взгляд, само понятие «большие данные» изначально возникло для того, чтобы помочь в решении проблем ИБ.

Давайте определимся с термином – что мы будем понимать под большими данными для реализации целей настоящего исследования. Итак, большие данные – это не просто данные, которых много, это цифровые данные, обладающие некоторыми свойствами.

Определение 1. *Во-первых, большие данные слишком объемны для ручного просмотра и анализа, то есть они действительно большие. Во-вторых, на основе анализа этих данных можно делать выводы, которые невозможно сделать на менее масштабной выборке. В-третьих, эти данные собраны из различных источников (ин-*

тернет-трафик, смартфоны, геопозиционирование, звонки, почта, видео и т. д.) и имеют различную природу (представлены в виде фото-, видео-, аудиоматериалов, текстов и т. д.). Как правило (хотя и не обязательно), большие данные касаются людей. В-четвертых, стоит отметить, что эти данные имеют временную компоненту, то есть они зависят от времени, могут рассматриваться ретроспективно, по ним можно делать прогнозы на будущее и т. д.

Многие источники, например, исследовательская компания IDC, при определении понятия Big Data говорят о четырех *V*: Volume, Variety, Velocity и Value (объеме, разнообразии, быстроте и ценности) [1]. IDC дает следующее определение: «Big Data – это технологии и архитектуры нового поколения для экономичного извлечения ценности из разноформатных данных большого объема путем их быстрого захвата, обработки и анализа» [2]. Это определение в целом не противоречит данному выше.

Я неслучайно начинаю с определения БД. Дело в том, что до сих пор определение это не устоялось, и в литературе встречаются различные дефиниции. В данной статье я буду придерживаться определения 1.

Большие данные в разных сферах

Анализ больших данных используется во многих областях, в частности, в сфере маркетинга, в розничной торговле, интернет-рекламе, поисковых системах, социальных сетях и телекоммуникациях. Во всех этих случаях основной целью анализа БД является анализ поведения потребителей или покупателей для предложения им наиболее подходящей рекламы, то есть конечной целью служит зарабатывание средств или увеличение продаж.

Другим довольно известной областью применения БД является их анализ спецслужбами. Конечной целью здесь является либо обнаружение преступных элементов (например, террористов), либо анализ массового поведения людей в сети для выявления оппозиционных настроений и организаций, упреждение забастовок и бунтов, предвидения результатов волеизъявления, моделирования группового поведения [3].

Второй пример – использования БД в целях разведки или в рамках других мероприятий, связанных с обеспечение госбезопасности. Казалось бы, логичным шагом будет перенос такой практики с государственного на бизнес-уровень. Однако этого в должной мере не происходит. В настоящий момент, к сожалению, анализ больших данных мало применяется в области корпоративной информационной безопасности, несмотря на то, что именно в корпорации цифровые данные локализованы и наиболее удобны для проведения всестороннего изучения.

У бизнес-структур есть огромное преимущество, например, перед социальными сетями, с точки зрения массового сбора информации, поскольку теоретически они имеют для этого все возможности, обладая доступом к таким ресурсам, как:

- переписка сотрудников по электронной почте;
- их активность в Интернете, осуществляемая из корпоративной сети;
- информация на рабочих компьютерах сотрудников, которая, согласно законодательству, принадлежит компании;
- все кадровые данные (персональные данные, трудовая книжка, анкеты и т. п.) сотрудников;
- информация об используемых сотрудниками электронных устройствах.

Таким образом, компании обладают огромным объемом данных о сотрудниках, но, как правило, используют эти данные разрозненно, по департаментам и подразделениям, а не для анализа и извлечения общих выводов, которые могут быть полезны для предотвращения злоупотреблений и неприятных корпоративных инцидентов. При этом, очевидно, что бизнес мог бы использовать эти данные таким же образом, как их используют спецслужбы.

У компаний, собирающихся использовать БД для внутреннего использования, есть существенное отличие от компаний-рекламодателей в целях, которые они преследуют. Последние выясняют общие предпочтения пользователей и мало интересуются конкретной личностью. Например, «рекламщики» знают, что женщины 20 лет любят высокие каблуки, и предлагают туфли на высоких каблуках всем женщинам этого возраста. При этом конкретная женщина их не интересует. Для компании же информация о каждом отдельно взятом сотруднике может быть очень ценной, особенно если речь идет о топ-менеджере или о специалисте, имеющем доступ к конфиденциальной корпоративной информации. Эти кадры очень ценны для компании, и если они допускают нарушения, то это представляет громадные риски для предприятия.

Человек – центр системы анализа больших данных предприятия

Вообще говоря, человек является основным источником рисков для компании, таких как халатность и разгильдяйство, уход к конкурентам или переманивание других сотрудников, хищения и растраты, утечки важной информации и использование инсайда, сговор с поставщиками, коммерческий подкуп, хищения и много других. Подобные поступки недобросовестных людей на предприятии ведут к финансовым потерям, оттоку клиентов, снижению репутации компании, серьезным юридическим последствиям и пр.

При этом выявить «плохого» сотрудника крайне сложно до тех пор, пока его преступление не становится очевидным, то есть когда предприятию уже нанесен урон. Характерным в этом смысле является пример сотрудника одной промышленной компании, который явился на работу с пистолетом и, узнав об увольнении, расстрелял пятерых [4]. Если бы руководство заранее обратило внимание на странное поведение своего сотрудника, вероятно, трагедии удалось бы избежать.

А такая возможность была, поскольку преступник ранее имел судимость. Будь у предприятия система, анализирующая различные каналы коммуникации сотрудников, в том числе социальные сети, она перехватила бы агрессивные послания данного субъекта в сети. Совокупность этих данных с допускае-

мыми им странными высказываниями в адрес ответственных лиц могло помочь руководству предпринять какие-либо упреждающие действия: отправить на лечение, обратиться в полицию и т. д.

То есть, теоретически, угрозу неадекватного поведения человека можно обнаружить на ранней стадии, выявив в его действиях некоторые отклонения до того, как этот человек совершил какой-нибудь противоправный поступок в отношении компании, коллег и т. д.

Задачей данной статьи является попытка предложить подход по выявлению рисков предприятия, связанных с поведением людей. Но прежде, чем перейти к этому, хочу дать краткую справку по состоянию дел в области информационной безопасности предприятий на текущий момент.

Текущая ситуация в области информационной безопасности предприятий

Изменение ситуации с цифровыми данными

Пока информационные системы компаний оставались ограниченными по своим возможностям, а циркулирующие в компьютерной сети предприятия объемы данных росли медленно, легкомысленное отношение к анализу потоков информации было оправданным. Однако в последнее десятилетие ситуация стала резко меняться по нескольким направлениям:

1) огромный рост объемов корпоративной информации: по оценке IDC, к 2025 году объем информации в мире вырастет в 10 раз по сравнению с 2016 годом [5];

2) резкий рост объема неструктурированной информации: поставщики услуг сообщают, что такие данные растут на 60–80 % ежегодно [6];

3) появление новых источников данных (например, открытых данных в области здравоохранения и данных, полученных от устройств Интернета вещей);

4) резкое повышение количества рисков и угроз для современных компаний (детально об этом см. подраздел «Ситуация с рисками»);

5) появление практически во всех развитых странах законов, связанных с защитой данных и серьезной ответственностью за нарушения оных;

6) все более выраженное желание компаний предотвращать риски, а не реагировать на них постфактум.

Пункты 1 и 2, по сути, говорят о том, что корпоративные сети постепенно становятся типичной системой больших данных. Проверим эту гипотезу, сравнив ее с определением 1.

Данные корпоративной сети сегодня являются довольно разрозненными (*Variety*) – они включают тексты, коды программ, cookies от посещений сайтов, данные с датчиков анализа состояния ИТ-сетей, данные видеонаблюдения, данные пропускных систем на территорию организации (СКУД) и пр. Причем число источников данных постоянно растет, что делает последние еще более разрозненными. Данные накапливаются и хранятся в больших объемах (*Volume*). Все они имеют временну́ю компоненту (*Velocity*) и, как правило, связаны с людьми. И, конечно, все эти данные обладают большой ценностью для владельцев и управленцев корпораций и предприятий (*Value*). По этой причине использование общепринятых подходов к анализу больших данных применительно к корпоративным ИТ-системам является совершенно оправданным.

Существующие аналитические системы

Если посмотреть на имеющиеся системы корпоративной информационной безопасности, то можно заметить, что все они ориентированы на защиту от определенного, довольно узкого класса угроз: антивирусы – на защиту от вирусов, системы защиты от уязвимостей – на защиту от уязвимостей, системы шифрования – на защиту информации от несанкционированного доступа и т. д. При этом ни одна из этих систем не способна предсказывать и предотвращать риски в целом, а также не занимается анализом поведения людей.

Системы, контролирующие информационную безопасность предприятий, можно разделить на пять основных групп с точки зрения подхода к анализу данных:

1) системы анализа элементов инфраструктуры – сканеры уязвимостей, патч-менеджмент (управление обновлениями программного обеспечения), инвентаризационные системы, системы сбора и управления инцидентами SIEM (*security information and event management*);

2) системы перехвата и анализа содержания для предотвращения утечек конфиденциальной информации – DLP (*data leakage prevention systems*), системы блокировки спама (антиспам) и фильтрация трафика (web-фильтры);

3) системы анализа конкретных угроз и защиты от них – антивирусы, межсетевые экраны, системы защиты от DDoS-атак и пр.;

4) системы видеонаблюдения и анализа видеопотока;

5) прочие, в том числе системы шифрования и контроля доступа к сети (NAC – *network access control*, IDS – *intrusion detection system*), контроля доступа в офис (СКУД) и др.

Каждая из подобных систем выдает собственный вердикт, касающийся ситуации с угрозами или состоянием сети. Значительная часть угроз блокируется автоматически (вирусы, атаки типа отказа в обслуживании и т. д.). Некоторые системы могут работать как в автоматическом режиме, так и в режиме рекомендательном, выдавая определенные назначения для дальнейшей обработки инцидентов системным администратором или офицером безопасности (например, так работают все DLP-системы). Некоторые системы требуют отдельного изучения потоков событий в ручном режиме (например, системы видеонаблюдения).

В идеале, конечно, офицеру безопасности хотелось бы получать некоторый интегрированный отчет о состоянии ИБ в целом на предприятии. И не просто отчет, а документ, включающий рекомендации по возможным действиям. Как ответ на эту потребность несколько лет назад стали появляться надстройки над аналитическими системами ИБ, которые занимаются интегрированным анализом инцидентов, происходящих в системе предприятия. Пер-

вой такой ласточкой стали SIEM-системы, предназначенные для анализа информации, поступающей от IDS, антивирусов, межсетевых экранов, ПАК для ИБ, маршрутизаторов и т. д. с последующим выявлением отклонений от норм согласно заложенным в систему критериям [7]. В основе работы SIEM лежат чистые математика и статистика.

По сути, системы типа SIEM – это первый шаг на пути полноценной аналитики больших данных в корпоративной ИБ. В то же время у них есть два существенных недостатка:

- SIEM – это система анализа свершившихся рисков, но не предотвращения таковых;
- SIEM анализирует только технические компоненты ИБ, тогда как человеческий фактор ими не учитывается, и сотрудник как носитель основных рисков предприятия остается за рамками этой картины.

Ситуация с рисками

Современному предприятию приходится сталкиваться со значительно бульшим числом рисков, чем это было 10–15 лет назад. Помимо роста числа атак, коих ежеквартально фиксируется уже сотни миллионов [8], постоянно растет их сложность. Также в последние годы получили широкое распространение новые виды угроз: майнеры, вирусы-шифровальщики, атаки на промышленные системы и др.

Любая новая технология, как правило, несет в себе риски, а разработчики средств защиты всегда идут в арьергарде у новых технологий. То есть сначала появляется технология, а потом экспериментальным путем пользователи узнают, что помимо удобств она таит в себе те или иные проблемы и опасности.

В этом смысле предприятия оказываются совершенно беспомощными перед постоянно нарастающей лавиной информационных и других технологических угроз, поэтому они весьма заинтересованы в получении решения по предотвращению и комплексному анализу угроз и рисков.

Эта тенденция в ИБ проявилась также около 10 лет тому назад. По крайней мере, некоторые эксперты рынка еще в 2009 году [9] отмечали начало смены парадигмы информационной безопасности – переход от задач чисто информационной безопасности к задачам по анализу рисков. При этом до сих пор решения по анализу рисков остаются в области ограниченных консалтинговых услуг либо не очень серьезных наработок. И никакие системы, имеющиеся в настоящий момент на рынке, не ставят перед собой комплексных задач по анализу рисков на основе изучения поведенческих аспектов самого слабого звена любой системы безопасности – человека.

Предваряя рассмотрение практических аспектов поднятой проблематики, подведем промежуточный итог изложением тезисов первой части статьи.

Анализ больших данных используется во многих сферах нашей жизни, но плохо применяется в интересах корпоративной безопасности.

Основным источником рисков любой системы является человек.

Все существующие системы корпоративной безопасности нацелены либо на борьбу с определенными типами угроз, либо на защиту определенных объектов. Пока не создано систем, которые бы могли предсказывать риски на основе исследования поведенческих аспектов человека.

Построение ИБ-системы, основанной на поведенческом анализе людей

Ограничения будущей модели

Очевидно, что весь комплекс рисков оценить и предсказать невозможно. Однако в отношении некоторые из них это вполне реально. Попытаемся представить модель, на основе которой можно будет построить эффективную систему анализа рисков предприятия, связанных с ИТ. Для этого сделаем некоторые предположения и внесем ограничения для данной системы:

- для данной модели не будем рассматривать риски, связанные с:
  - природными катастрофами, войнами и другими форс-мажорными событиями;
  - техническими сбоями и проблемами ИТ-систем;
  - общеизвестными типами угроз (компьютерными вирусами, атаками типа «отказ в обслуживании» и пр.), для защиты от которых разработаны специальные средства;
  - экономическими проблемами и рисками бизнеса;
- будем рассматривать риски предприятия, связанные со следующими бизнес-субъектами:
  - внешними поставщиками;
  - подрядчиками;
  - конкурентами;
  - сотрудниками;
  - руководителями.

Разработка собственно модели

Далее займемся построением модели анализа рисков с учетом упомянутых выше ограничений. Назовем нашу модель ***Human Behavior Analytical System*** (**HBAS**), то есть система поведенческого анализа человека.

Стандартная модель любой обработки данных состоит из трех компонентов:

Сбор данных => Анализ данных => Принятие решения.

Поскольку мы пытаемся построить систему анализа рисков ИБ, то стандартную трехзвенную модель для нашего случая можно уточнить следующим образом:

Сбор данных => Анализ данных и выявление проблемных событий (инцидентов) => Обработка инцидентов (расследование, устранение) => Корректировка модели.

Рассмотрим далее каждый компонент отдельно.

Сбор данных для анализа

Для того чтобы анализировать любые данные, в том числе большие, их надо сначала собрать.

Как уже отмечалось ранее, процесс сбора данных в единой корпоративной системе значительно упрощается по сравнению с разрозненными системами, поскольку значительный объем необходимой информации сосредоточен внутри корпорации либо доступен ее работникам.

Какие данные могут потребоваться для получения представления

Таблица 1. ***Системы для сбора данных в корпоративной сети***

№ п/п	Тип данных	Система, которая может обеспечить сбор данных
1	*Данные о действиях сотрудников:*	
	а) посещение интернет-сайтов	Web-filter
	б) компьютерные игры, соцсети и другие «убийцы» рабочего времени	Web-filter, DLP, системы контроля рабочего времени
	в) работа с компьютерными файлами (открытие/закрытие, копирование, внесение изменений)	Системы контроля конечных устройств (*Endpoint Security*)
2	*Данные о коммуникациях сотрудников внутри организации и за ее пределами:*	
	а) анализ электронной почты, чатов, мессенджеров и пр.[1]	DLP
	б) круг общения сотрудника внутри предприятия (так называемый граф связей)	DLP
3	*Данные о публичной жизни сотрудника вне предприятия:*	
	а) мнения и высказывания сотрудников, сделанные ими на различных ресурсах и в соцмедиа.	Системы автоматического мониторинга соцмедиа (SMM)
	б) фотографии, аудио- и видеоматериалы, размещенные сотрудником в социальных сетях	- SMM; - Системы распознавания фото- и видеоизображений
	в) круг общения сотрудника в Сети	- Построение графа связей; - SMM

о возможных рисках предприятия, связанных с неправильным или опасным поведением людей?

1. Данные о действиях сотрудников:
- посещение интернет-сайтов;
- компьютерные игры, социальные сети и другие «убийцы» рабочего времени;
- работа с компьютерными файлами (открытие/закрытие, копирование, внесение изменений).

2. Данные о коммуникациях сотрудников внутри организации и за ее пределами:
- анализ электронной почты, чатов, мессенджеров и пр.;
- круг общения сотрудника внутри предприятия (так называемый граф связей);
- круг общения, связанный с поставщиками, подрядчиками и другими партнерами организации, а также с конкурентами (граф внешних связей).

3. Данные о публичной жизни сотрудника вне предприятия:
- мнения и высказывания сотрудников, сделанные ими на различных ресурсах и в соцмедиа;
- фотографии, аудио- и видеоматериалы, размещенные сотрудником в социальных сетях;
- круг общения сотрудника в Сети.

Перечисленные выше данные обладают всеми свойствами больших данных (см. определение 1), поэтому и методики их обработки будут аналогичными.

Какими способами можно собрать указанные выше данные?

Существует два основных способа сбора интересующих нас данных о персонале:
- *организационный* – люди сами доставляют информацию о предполагаемых инцидентах ответственным за безопасность;
- *автоматический* – сбор информации осуществляется с помощью различных автоматических систем.

Будем считать, что организационным способом можно собрать любые из вышеперечисленных данных. Приведем конкретные системы сбора информации о людях, сохраняя нумерацию, приведенную выше (табл. 1).

По сути, речь идет об установке «датчиков» на все каналы коммуникаций сотрудников. Сбор данных предполагает отслеживание коммуникаций, а также действий сотрудников предприятия, в режиме реального времени. В настоящее время не существует решений, которые бы позволяли собирать данные из различных систем воедино и проводить на их основе контентный анализ.

Подход HBAS аналогичен подходу решений типа SIEM, собирающим данные с разных источников в корпоративной сети и делающим выводы в бинарной логике: «инцидент/не инцидент» [10]. Только в отличие от SIEM, следящего в основном за техническими характеристиками системы, такими как аномалии в трафике, срабатывание антивируса или другого датчика, HBAS будет делать выводы, анализируя содержание коммуникаций и действий людей.

Рассуждая теоретически, было бы интересно объединить системы контентного анализа HBAS с системой анализа инцидентов SIEM, чтобы получить полную и разностороннюю картину инцидентов, происходящих на предприятии, но это не является задачей данной работы.

Анализ полученных данных, принятие решения

Для начала попробуем построить модель системы анализа контента или коммуникаций, то есть представим, как будет работать HBAS. Объединяя различные датчики, мы сумели собрать большое количество данных со всей компании. Вопрос: что теперь делать с этим массивом информации?

Анализируя потоки собранных данных, можно выявлять происходящие в компании события.

Определение 2. *Назовем* ***событием*** *некое действие сотрудника или группы сотрудников, которое потенциально может нести в себе угрозу безопасности. Если же событие нарушит какое-либо правило политики безопасности предприятия, то оно будет называться* ***инцидентом****.*

Нужно распределить все полученные события по степени их риска.

[1] *В данной статье рассматриваются только легальные способы получения информации, поэтому такой способ, как, например, перехват разговора по мобильному телефону, здесь не упоминается.*

Для этого нам надо построить **Разборщик событий**. Определим основные типы возможных активностей пользователей информационных систем компании (сотрудников), которые связаны с содержанием переписки или других каналов коммуникаций. Предлагаем выделить четыре основных типа таких активностей и один дополнительный:

- накопление, хранение и распространение информации ограниченного доступа;
- социальное общение пользователей;
- поиск в сети Интернет (поиск по ключевым словам);
- посещение web-сайтов;
- (возможно – нарушение системы контроля доступа в помещение).

Для соблюдения однозначности условимся, что каждое событие, детектируемое системой, может относиться только к одной из вышеперечисленных активностей.

Далее необходимо разработать **Классификатор инцидентов**.

Для создания классификатора можно просто перечислить все возможные инциденты, но целесообразнее создать некоторые их группы, в которые/из которых можно добавлять/убирать события. По опыту внедрений систем InfoWatch, на предприятиях обычно выделяются следующие группы инцидентов[2]:

1) Утечка информации ограниченного доступа;

2) Информация, порочащая репутацию компании и ее ключевых лиц;

3) Экономические преступления;

4) Кража и/или уничтожение материальных активов;

5) Угрозы физическим лицам (сотрудникам компании);

6) Нелегальный контент;

7) Проблемы с трудовыми отношениями;

8) Развлечения во время работы.

Далее будем называть группы событий 1–8 ***инцидентными группами***.

Теперь необходим оценить каждую из этих инцидентных групп по ее значимости с точки зрения возможных последствий (потерь) для компании. Чем больше возможные потери, тем данное событие критичнее (важнее). Под возможными потерями будем понимать следующие прямые или косвенные потери предприятия:

а) прямые финансовые потери;

б) упущенная выгода;

в) негативное влияние на репутацию компании;

г) штрафы и другие санкции регуляторов;

д) легальные преследования со стороны третьих лиц;

е) неблагонадежное/неадекватное поведение сотрудников;

ж) ущерб сотрудникам (физическим лицам);

з) снижение мотивации и эффективности персонала, угроза увольнения;

и) возможности последующего влияния на принимающих решения лиц.

Теперь нам необходимо оценить инцидентные группы с точки зрения их опасности для организации. Для этого каждой инцидентной группе мы присвоим определенный приоритет – от единицы (опасность максимальна) до четырех (опасность невелика). Приведем описание приоритетов:

1 – очень высокая степень опасности;

2 – высокая степень опасности;

3 – средняя степень опасности;

4 – низкая степень опасности.

Методика присвоения приоритетов конкретным инцидентным группам может быть различной, но, как правило, подобные приоритеты расставляются руководителями службы безопасности на основе их опыта, исходя из задач предприятия и направлений, заданных руководством. В данной работе автор прибегает к своему опыту для расстановки приоритетов инцидентным группам 1–8.

На основании заданных приоритетов для каждой инцидентной группы составим **Матрицу пересечений** (табл. 2). Для этого выберем возможные действия пользователя в сети (колонка «Активность»). Этим активностям сопоставим конкретные инциденты или то, что на языке информационной безопасности называется событием (см. определение 2). Инциденты (события) выбраны на основе многолетнего опыта работы компании InfoWatch с корпоративными клиентами.

В графе «Срабатывание политики» зададим условие, при котором система должна реагировать на инцидент. Это может быть единичное событие или конкретные числовые значения, при которых событие считается инцидентом безопасности.

Далее каждой активности и каждому событию сопоставим приоритет, и в следующей колонке выберем возможные последствия («+», если последствие будет иметь место, и «–», если не будет). Ниже приведены таблицы анализа инцидентов для случая утечки информации ограниченного доступа (см. табл. 2) и для случая экономических преступлений (табл. 3).

Подобные таблицы составляются по всем остальным инцидентным группам.

На основе табл. 2 и 3 строится *система анализа* или *решающий модуль*, способный автоматически обнаруживать и фиксировать инциденты информационной безопасности на основе как совокупности инцидентных факторов, так и единичных событий.

Решающий модуль должен работать **Поверх** систем, из которых он будет получать данные: DLP, SMM, Web-filter и др. Сутью решающего модуля будет:

Извлечение данных => Анализ извлеченных данных => Принятие решения.

Условно архитектуру подобного решения можно изобразить следующим образом (см. рисунок).

Описанная выше система является ни чем иным как интеллектуальным категоризатором всей проходящей через компанию и исходящей из нее информации, дополненным анализатором поведения и связей.

[2] *Данные группы приводятся для примера, хотя в целом они соответствуют основным категориям рисков, которые, согласно практическому опыту автора, имеются у предприятий.*

Таблица 2. ***Матрица пересечений инцидентов и событий***

Группа	Активность	Событие	Срабатывание политики	Приоритет	Возможные последствия								
					а	б	в	г	д	е	ж	з	и
Утечка информации ограниченного доступа	Накопление, хранение и распространение информации ограниченного доступа	Выявление информации ограниченного доступа посредством DLP-системы	Единичное событие	1	+	+	+	±	±	–	–	–	–
	Накопление, хранение и распространение информации ограниченного доступа	Накопление информации ограниченного доступа на рабочей станции сотрудника (выкачивание файлов из сетевых ресурсов на локальный диск), к которой он имеет права доступа	1. Более *N* МБ в течение X дней 2. Более *N* файлов в течение X дней 3. Более *X* % от «нормальной» активности, установленной с помощью «слепка»[3]	2	±	±	±	±	±	–	–	–	–
	Накопление, хранение и распространение информации ограниченного доступа	Наличие на рабочей станции сотрудника информации ограниченного доступа, к которой он не имеет прав доступа	Единичное событие	2	±	±	±	±	±	–	–	–	–
	Социальное общение	Прямой договор о краже, купле-продаже информации ограниченного доступа, обсуждение уголовного законодательства	Единичное событие	1	+	+	+	±	±	–	–	–	–
	Поиск в сети Интернет	Поиск информации о купле-продаже информации ограниченного доступа (базы персональных данных, номера кредитных карт, пр.)	Единичное событие	1	+	+	+	±	±	–	–	–	–
	Поиск в сети Интернет	Поиск информации по уголовному законодательству в части кражи информации ограниченного доступа и нарушения политик информационной безопасности компании	Единичное событие	2	±	±	±	±	±	–	–	–	–
	Социальное общение	Обсуждение возможностей по сокрытию (стеганография) и шифрованию информации, обходу и/или отключению средств защиты (особенно DLP-систем)	Единичное событие	2	±	±	±	±	±	–	–	–	–
	Поиск в сети Интернет	Поиск информации по сокрытию (стеганография) и шифрованию информации, обходу и/или отключению средств защиты (особенно DLP-систем)	Единичное событие	2	±	±	±	±	±	–	–	–	–

Иными словами, это система искусственного интеллекта, которая способна автоматически обнаруживать инциденты информационной безопасности.

Данная система будет состоять из следующих трех модулей.

1. *Перехватчики всей информации в режиме реального времени.* Сюда входят перехватчики систем DLP, web-filter, анализаторы соцмедиа и поисковых запросов[4].

2. *Аналитический модуль*, который собирает информацию, полученную с перехватчиков, строит по ним модели поведения и выдает их пользователям системы в виде решения. Пользователями могут быть риск-аналитики, офицеры безопасности и управляющие компаний.

3. *Система хранения, аккумулирующая все инциденты.* На основании данных из системы хранения можно провести ретроспективный анализ полученных данных: проанализировать всю информацию, полученную в прошлом, и сделать выводы о том, как ситуация может развиваться в будущем, тем самым предотвращая возможные инциденты.

Впоследствии система может дополняться другими перехватчиками или интегрироваться с системами SIEM. Это позволит строить более сложные модели поведения и предсказывать случаи, которые лежат на стыке человеческого поведения и технических проблем в инфраструктуре. Однако надо понимать, что чем больше информации поступает в блок анализа, тем более сложные и менее предсказуемые модели будут получаться на выходе.

Представленный выше подход является частным случаем реализа-

[3] *Слепок с базы ключевых слов в DLP-решениях – это файл, содержащий набор хэш-сумм ключевых слов, соответствующий набору слов в базе, на основе которой был сделан слепок.*

[4] *В принципе, в систему можно добавить любой другой перехватчик, а не только четыре означенных: например, перехват видеопотока или перехват системы контроля доступа в помещение. Для таких перехватчиков, однако, нужно будет разрабатывать отдельный аналитический модуль, который довольно сложен в реализации.*

Таблица 3. ***Анализ событий, нарушающих политики ИБ в области экономических преступлений***

Группа	Активность	Событие	Комментарий	Приоритет	Возможные последствия							
					а	б	в	г	д	е	ж	з	и
Экономические преступления	Социальное общение	Прямой договор об организации отката, завышения цен и других мошеннических действиях, обсуждение уголовного законодательства	Единичное событие	1	+	±	±	±	±	+	–	±	+
	Социальное общение	Обсуждение цен или других условий с конкурентами	Единичное событие	1	+	–	+	–	–	+		–	+
	Поиск в сети Интернет	Поиск информации по уголовному законодательству в части экономических преступлений	Единичное событие	2	±	±	±	±	±	±	–	±	–
	Социальное общение	Пересылка по любым каналам пустых бланков[5] с печатями	Единичное событие	2	+	±	±	±	±	+	–	±	+
	Социальное общение	Пересылка факсимиле руководителей или пустых бланков с подписями ключевых лиц	Единичное событие	2	+	±	±	±	±	+	–	±	+

ции подхода к анализу больших данных в целях обеспечения информационной безопасности предприятия.

Возможные проблемы создания и использования подобной системы

Для реализации описанной выше системы существует ряд довольно серьезных технических ограничений, а также некоторые возможные проблемы морального характера. Рассмотрим их подробнее.

Проблема 1. Объектом интереса и исследования системы HBAS является человек, причем человек, который предположительно является нарушителем. Понятно, что поведение человека, знающего, что за ним следят, отличается от его поведения в обычной ситуации. Любой нарушитель будет всячески пытаться обмануть систему.

С другой стороны, данная система совершает перехват различных событий и одновременно анализирует множество разнообразных параметров, что делает варианты ее обхода слишком трудоемкими и сложными. Рано или поздно проскользнет то, что не свойственно поведению коллег нарушителя – это и выдаст данного индивидуума. Полностью контролировать свое поведение способны разве что тренированные агенты спецслужб, против которых, скорее всего, система будет бессильна, однако вероятность наличия таковых в штате среднестатистической организации, прямо скажем, не слишком велика.

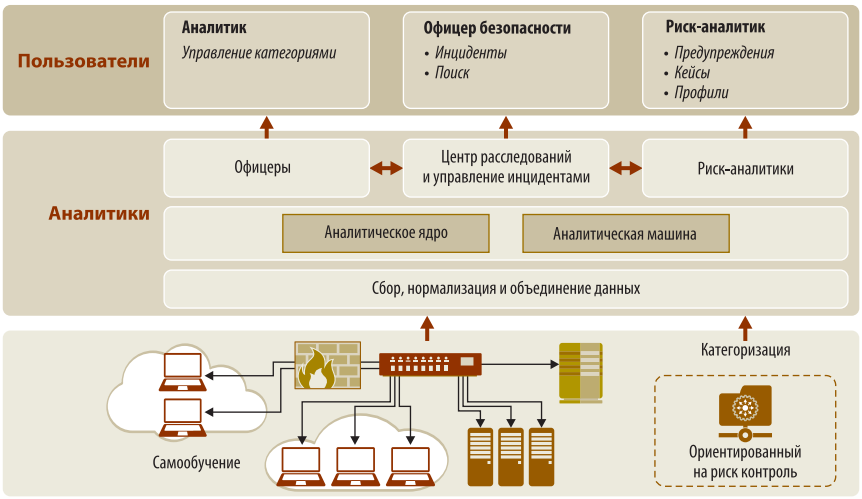

Рисунок. ***Схема работы решения HBAS***

Проблема 2. Подбор коэффициентов и определение инцидентов делается экспертным путем. Не существует доказательств, что результат подбора точно отражает систему угроз предприятия.

Способ решения данной проблемы – экспертным образом ставить произвольные коэффициенты, пробовать их в реальной системе на ретроспективных данных и корректировать по результатам. Возможно, потребуется несколько таких итераций, прежде чем система начнет нормально работать.

Проблема 3. Необходимость настройки системы под конкретную бизнес-модель конкретного предприятия. Вероятность того, что экспертным образом подобранные для одного предприятий коэффициенты не подойдут другому, довольно высока. Это значит, что систему придется адаптировать для различных индустрий.

Проблема 4. Системы искусственного интеллекта неизбежно вторгаются в частное поле человека, так как следят за его поступками. На Западе нарушение тайны частной жизни (*privacy*) является наказуемым деянием и преследуется по закону.

Конечно, говорить о *privacy*, когда у каждого в кармане есть смартфон, буквально «излучающий» различные данные, когда люди с удовольствием носят фитнес-браслеты, транслирующие всю информацию

[5] *Пустые бланки документов с печатями и подписями, обычно не попадают в перечень информации ограниченного распространения и поэтому отнесены в Активность «Социальное общение».*

о здоровье их владельца куда-то в облако, и когда на каждом углу стоит камера видеонаблюдения, несколько странно. Тем не менее, для предприятий вопрос доступа к событиям частной жизни можно решить путем подписания с сотрудником договора-согласия на то , что за ним будут наблюдать автоматические системы. По крайней мере, данный алгоритм работает, когда речь идет об имеющихся на рынке системах защиты от утечек, функционирование которых не лишено той же этической проблемы, так как они «читают» содержание переписки.

Системы типа HBAS на мировом рынке

Первой в мире публичной системой класса HBAS стала разработка компании IBM QRadar User Behaviour Analytics, представленная в 2013 году. IBM тогда объявила о создании системы анализа больших данных с целью проведения бизнес-разведки [11].

С тех пор решения подобного класса, который исследовательская компания Gartner назвала неблагозвучной для русского уха аббревиатурой UEBA (*User and Entity Behavioral Analytics* – поведенческая аналитика пользователей и организаций) [12] начали появляться, но мировой рынок подобных средств пока довольно мал – примерно 185 млн долл. по итогам 2017 года. Однако сегодня уже отмечается его стремительный рост и, по прогнозам аналитиков, в 2024 году объем превысит 2,4 млрд долл. [13].

На данный момент автору удалось насчитать более десятка компаний, которые заявляют, что разрабатывают решения подобного класса и готовы предложить их российским заказчикам. Правда, публичных проектов пока нет. Но в целом рынок уже созрел для появления спроса на системы UEBA и, скорее всего, пилотные внедрения уже прошли [14].

Интересно, что разные компании при этом используют разные подходы и алгоритмы для решения задачи. Например, IBM в своем решении QRadar UBA использует расширение SIEM-системы от IBM также с названием QRadar. Аналогичный путь выбрала калифорнийская компания Exabeam с продуктом Exabeam Advanced Analytics. За ними последовала и Hewlett Packard с решением HP ArcSight UBA (которое, после отделения от Hewlett Packard называется Micro Focus ArcSight UBA), чье решение данного класса тоже основано на известной SIEM-системе собственной разработки.

Другой путь – это переход от анализа больших данных с общими целями к использованию их в сфере информационной безопасности. Примерами такого подхода могут служить компании Splunk и Microsoft. Первая, которая традиционно фокусировалась на обработке и быстрой аналитике больших массивов текстовых данных, использовала эти знания для создания приложения для анализа поведения человека – продукта Splunk UBA. Аналогичный подход выбрала компания Microsoft, которая для входа в новый рынок купила за 200 млн долл. израильскую компанию Aorato [15], занимавшуюся разработками в области анализа больших данных и систем машинного обучения.

Третий подход – это построение решения по анализу поведения на основе сбора данных со своих клиентских агентов. К подобного рода решениям относятся продукт израильской компании ObserveIT ITM и – с некоторыми ограничениями – StaffCop Enterprise от российской компании «Атом Безопасность».

Наконец, последний способ реализации желаемой функциональности – это реализация его на основе DLP-системы. По этому пути идет компания InfoWatch, а также ряд других российских производителей, включая «МФИ Софт» и Solar Security.

Пример реализации подобной системы в продукте InfoWatch

Приведу пример подобного решения, разработанного компанией InfoWatch. Созданная нами система предназначена для выявления угроз возникновения финансовых, репутационных и кадровых потерь для компании. Под финансовыми потерями понимаются отток клиентов, потеря проектов, тендеров, утечки технологий, прямое вредительство со стороны сотрудников, штрафы, выплаты, административная ответственность для компании, мошенничество сотрудников. Под репутационными потерями понимаются высказывания, публикации и компрометирующие действия сотрудников, подрывающие авторитет компании. Под кадровыми потерям понимаются увольнения ценных сотрудников, низкая эффективность сотрудников (трата рабочего времени впустую, плохая атмосфера, конфликты в коллективе). Система позволяет обнаруживать недобросовестных сотрудников до того, как они нанесут ущерб компании.

Работа программного продукта состоит из трех этапов: создания слепка нормального поведения, выявления аномалий поведения и анализа выявленных аномалий.

1. *Создание слепка нормального поведения.* Сразу после установки система начинает сбор информации по каждому сотруднику, формируя слепок, учитывающий более 50 различных факторов поведения (например, Иван Иванов делает в среднем 30 SQL-запросов в день, скачивает 2 мегабайта информации из сетевого хранилища, отправляет 20 писем в неделю и т. д.).

2. *Выявление аномального поведения.* После завершения процесса создания слепка система начинает вести 24-часовой мониторинг поведения каждого сотрудника. При выявлении отклонений от «слепка» система создает инцидент и определяет тип потерь, к которым этот инцидент может привести (например, финансовые потери). По совокупности инцидентов формируется рейтинг угрозы для каждого сотрудника. То есть система позволяет выявить опасных и ненадежных сотрудников.

3. *Анализ выявленных аномалий в поведении.* Используя инструменты визуализации и анализа поведения, офицер безопасности проводит расследование инцидентов по отдельным сотрудникам, чтобы сделать выводы о наличии реальной угрозы. После завершения расследования система формирует отчеты для представления руководству компании и принятия дальнейших решений.

Выводы

На взгляд автора, системы поведенческого анализа, внедряемые для предотвращения рисков и угроз информационной безопасности предприятий, – весьма перспективное направление. В ближайшее время они будут широко востребованы в компаниях из различных вертикалей.

Разработчики используют разные подходы к решению задачи. В нашей работе перечислены четыре таких подхода:

- расширение системы защиты от утечек (DLP) дополнительными каналами перехвата и построение над ними решающего модуля;
- расширение системы анализа событий и инцидентов (SIEM) возможностями по анализу поведенческих паттернов сотрудников;
- дополнение клавиатурных шпионов и других клиентских программ функционалом анализа больших данных;
- фокусировка систем анализа больших данных на узкие задачи информационной безопасности.

Однако какой бы путь ни был выбран, подобные продукты позволят предприятиям перейти от парадигмы реагирования на угрозы информационной безопасности к раннему обнаружению и прогнозированию данных угроз, к предотвращению рисков. Технической основой для такого рода решений сейчас выступают технологии анализа ИБ-угроз предприятий. ■

ЛИТЕРАТУРА

1. Аналитики IDC обещают, что «большие данные» дадут огромные прибыли [Электронный ресурс]. – Режим доступа: http://www.computerra.ru/81587/idc-bigdata/.

2. Беклемишев Андрей. Big Data и бизнес-аналитика [Электронный ресурс]. – Режим доступа: http://ca.idc.com/dwn/PRES_84767/02_effective_management_of_enterprise_data_andrew_beklemishev_idc.pdf.

3. Якорек Владислав. Big Data по-русски: на службе государства, рекламы и страховщиков [Электронный ресурс]. – Режим доступа: https://thisis.media/future/2505176/.

4. Сотрудник завода в США расстрелял пятерых коллег, узнав об увольнении [Электронный ресурс]. – Режим доступа: https://www.mk.ru/incident/2019/02/16/strelbana-zavode-illinoysa-uvolennyy-sotrudnik-ubil-pyat-kolleg.html.

5. Объем данных всего мира к 2025 году увеличится в 25 раз [Электронный ресурс]. – Режим доступа: https://aboutdata.ru/2017/04/27/volume-of-data-by-2025/.

6. Unfrustructured Data Growth Fueling Massive Migration to Object Storage [Электронный ресурс]. – Режим доступа: https://www.enterprisestorageforum.com/storage-management/unstructured-data-growth-fueling-massive-migration-to-object-storage.html.

7. Сапрыкина Анастасия. Обзор мирового и российского рынка SIEM-систем [Электронный ресурс]. – Режим доступа: https://www.anti-malware.ru/analytics/Market_Analysis/overview-global-and-russian-market-siem/.

8. Развитие информационных угроз в третьем квартале 2018 года. Статистика [Электронный ресурс]. – Режим доступа: https://securelist.ru/it-threat-evolution-q3-2018-statistics/92612/.

9. LETA-Group: Рынок информационной безопасности 2009 – начало эпохи compliance [Электронный ресурс]. – Режим доступа: http://www.letagroup.ru/common/files/LETA_research_2010.pdf.

10. Что такое SIEM и зачем они нужны? [Электронный ресурс]. – Режим доступа: http://www.is-systems.org/siem/about_siem/.

11. Джамак Питер. Бизнес-анализ больших данных [Электронный ресурс]. – Режим доступа: http://www.ibm.com/developerworks/ru/library/ba-big-data-bi/index.html.

12. Market Guide for User and Entity Behavior Analytics [Электронный ресурс]. – Режим доступа: https://www.gartner.com/doc/3872885/market-guide-user-entity-behavior/.

13. Global User and Entity Behavior Analytics (UEBA) Market Report [Электронный ресурс]. – Режим доступа: https://www.valuemarketresearch.com/report/user-and-entity-behavior-analytics-ueba-market/.

14. UEBA: что скрывается за новым трендом на рынке информационной безопасности [Электронный ресурс]. – Режим доступа: http://www.tadviser.ru/a/415979/.

15. TechCrunch, may 2014 [Электронный ресурс]. – Режим доступа: https://techcrunch.com/2014/11/13/microsoft-buys-israeli-hybrid-cloud-security-startup-aorato-in-200m-deal/.

16. Внутренние документы АО «ИнфоВотч». Описание концепции программного продукта InfoWatchPrediction.