УДК 004.2

ОЦЕНКА ВЛИЯНИЯ МОДУЛЯ МАШИННОГО ОБУЧЕНИЯ НА ФУНКЦИОНАЛ DLP СИСТЕМЫ

Яруллин К.Л.
ФГБОУ ВО «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»
Научный руководитель – Кривоносова Н.В.
ФГБОУ ВО «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»

Доклад освещает тематику современных решений в области информационной безопасности (далее ИБ), а именно основные технологические особенности решений DLP. Освещаются сложности, связанные с принципиальным строением классической системы DLP, предлагается решение, основанное на применении технологии машинного обучения.

Распространенным средством обеспечения ИБ в организации с точки зрения защиты от внутренних угроз является система DLP. DLP система - это программный продукт, созданный для защиты организации от утечек информации (Data Leak Prevention - предотвращение утечек данных). Важно отметить что DLP система борется именно с утечками, связанными с деятельностью сотрудников компании и чаще всего не является универсальным решением для защиты и от внешних угроз. Предметом работы DLP системы являются данные, собираемые самой системой.
Данными в этом случае могут быть:
• информация о передаваемых сотрудниками файлах;
• время активности сотрудников;
• текст, содержащийся в электронных письмах, названия документов;
• акты записи информации на носители (чаще всего интересует запись или попытки записи на внешние носители).
Собранные данные хранятся и обрабатываются, затем уполномоченному сотруднику отдела ИБ предоставляется отчет по реакции системы. В зависимости от настроек этой реакцией может быть запись (например, изменение «рейтинга доверия сотруднику», предупреждение для сотрудника ИБ) или действие (например, блокировка пользователя). Соответственно действие системы способно повлиять на бизнес-процесс в организации, а запись только привлечь внимание сотрудника ИБ.
Вышеупомянутая зависимость характера работы DLP системы от настроек является важной особенностью, ведь именно настройки определяют будет ли DLP система работать эффективно или же в случае некорректной настройки будет мешать бизнес-процессам.
Настройка (программирование) реакций DLP системы лежит на уполномоченном сотруднике ИБ, так же как и анализ результатов работы DLP системы (отчеты, уведомления, действия).

В общем случае работа уполномоченного сотрудника ИБ с DLP системой представляет собой замкнутую систему:

1. DLP система собирает данные по параметрам, анализирует по параметрам, создает реакцию по параметрам;
2. Сотрудник ИБ анализирует результат работы DLP системы и вносит необходимые изменения в её параметры;
3. Возвращение к пункту 1.

Подобная нагрузка на уполномоченного сотрудника ИБ может становиться проблемой т.к.:

1. Требует специальных навыков сотрудника;
2. Требует анализа огромных объемов данных в больших сетях;
3. Создает сильную зависимость функционирования DLP системы от работы конкретного сотрудника/группы сотрудников ИБ.

Решением проблемы в современных системах DLP может стать внедрения модуля машинного обучения.

Машинное обучение является в общем случае распространенным методом избавления от сложных алгоритмов и инструкций за счет решения задачи на основании предыдущего опыта решения множества сходных задач. На данный момент является популярным решением, доказавшим свою эффективность во многих сферах.

DLP система в ходе функционирования предоставляет среду, благоприятную для машинного обучения, в частности - это большие объемы отфильтрованной информации, по которой нейросеть (здесь – логическая единица, представляющая собой систему способную к обучению) может обучаться, анализировать и реагировать на угрозы.

При внедрении модуля машинного обучения его работа с системой DLP выглядит следующим образом:

1. DLP система собирает данные по параметрам, анализирует по параметрам, создает реакцию по параметрам;
2. Модуль машинного обучения изучает предыдущий опыт и на его основании предлагает реакцию (например, изменение параметров DLP системы или уведомление сотрудника ИБ)

Результатом взаимодействия двух современных и актуальных систем становится:

• отсутствие необходимости постоянного участия сотрудника;

• способность модуля машинного обучения быстро обрабатывать огромные объемы данных, что позволяет находить новые паттерны, а значит новые потенциально опасные действия пользователей.