Аннотация – В этом документе представлен анализ патента US11611582B2, описывающего компьютерный метод обнаружения фишинговых угроз. Анализ охватывает различные аспекты патента, включая его технические детали, потенциальные области применения и последствия для специалистов по кибербезопасности и других секторов промышленности.

Актуальность для развивающегося ландшафта DevSecOps подчёркивает вклад в более безопасные и эффективные жизненные циклы разработки программного обеспечения, поскольку патент предлагает методический подход к обнаружению фишинга, который может быть применён различными инструментами и сервисами для защиты пользователей и организаций от вредоносных действий в Интернете. Специалистам по кибербезопасности следует рассмотреть возможность включения таких методов в свои стратегии защиты для упреждения возникающих угроз.

I. Введение

Патент US20220232015A1 описывает способ динамического обнаружения фишинговых угроз с использованием заранее определённой статистической модели. Метод основан на машинном обучении и позволяет анализировать сетевые запросы в режиме реального времени и выявлять потенциальные попытки фишинга, чтобы проактивно защищать пользователей и системы от таких атак. Статистическая модель и набор признаков позволяют адаптироваться к новым фишинговым схемам.

II. Основная идея

Основная идея патента заключается в предоставлении масштабируемого и автоматизированного подхода к обнаружению попыток фишинга в режиме реального времени с использованием машинного обучения с целью упреждающей защиты пользователей от становления жертвами все более изощренных фишинговых атак. Динамический анализ атрибутов веб-запросов позволяет выявлять новые фишинговые сайты, которые могут отсутствовать в статических списках.

- Описывается реализованный метод динамического обнаружения фишинговых угроз с использованием заранее определённой статистической модели. Цель состоит в том, чтобы в режиме реального времени определить, представляет ли запрашиваемый сетевой ресурс потенциальную угрозу фишинга.
- При получении запроса на доступ к сетевому ресурсу извлекается набор признаков, связанных с запросом. Эти характеристики могут включать доменное имя, возраст домена, его регистратор, IP-адрес, географическое местоположение и т.д.
- Извлечённые характеристики загружаются в предварительно обученную модель, которая выдаёт оценку вероятности того, что запрошенный ресурс представляет собой фишинговую угрозу. Если оценка превышает заранее определённый порог, формируется оповещение.
- Модель обучается на наборах данных, содержащих известные фишинговые и нефишинговые примеры с периодическим обновлением данных.

III. Область применения

Конкретные детали реализации и точки интеграции будут варьироваться в зависимости от требований каждой отрасли и существующего технологического стека. Однако основные возможности динамического обнаружения фишинга с использованием машинного обучения могут быть адаптированы для обеспечения значительных преимуществ в области безопасности в широком спектре секторов, сталкивающихся с растущей угрозой фишинговых атак.

A. Телекоммуникации:

Телекоммуникационные компании могут интегрировать систему обнаружения фишинга в свою сетевую инфраструктуру для защиты клиентов от атак, осуществляемых с помощью SMS, MMS или других служб обмена сообщениями.

Возможности обнаружения в режиме реального времени помогут блокировать фишинговые ссылки до того, как они дойдут до конечных пользователей, снижая риск взлома учётной записи и кражи личных данных.

Телекоммуникационные провайдеры могут предлагать защиту от фишинга в качестве дополнительной услуги, позволяющей выделиться на рынке и завоевать доверие клиентов.

B. Информационные технологии:

ИТ-компании могут внедрять решение для обнаружения фишинга в рамках своих предложений по безопасности для клиентов, помогая защитить от атак, нацеленных на сотрудников и клиентов.

Поставщики услуг управляемой безопасности (MSSP) интегрируют эту технологию в свои службы мониторинга угроз и реагирования на инциденты, чтобы обнаруживать и блокировать попытки фишинга в режиме реального времени.

Поставщики программного обеспечения как услуги (SaaS) могут встроить функцию обнаружения фишинга в свои платформы для сканирования подозрительных URL-адресов и вложений, повышая безопасность своих приложений.

C. Финансовый сектор:

Финансовые учреждения могут использовать систему обнаружения фишинга для защиты своих клиентов от целенаправленных фишинговых атак, направленных на кражу учётных данных для входа в систему, номеров кредитных карт и других конфиденциальных финансовых данных.

Решение может быть интегрировано в платформы онлайн-банкинга, мобильные приложения и системы электронной почты для сканирования и выявления потенциальных попыток фишинга в режиме реального времени.

Благодаря активному обнаружению и блокированию фишинговых угроз финансовые компании снизят потери от мошенничества, сохранить доверие клиентов и соблюдать нормативные требования по защите данных.

D. Здравоохранение:

Организации здравоохранения могут использовать технологию обнаружения фишинга для защиты конфиденциальных данных пациентов и предотвращения фишинговых атак, которые могут поставить под угрозу конфиденциальность, целостность и доступность систем здравоохранения.

Решение может быть развёрнуто для мониторинга сообщений электронной почты, порталов пациентов и других цифровых каналов на предмет признаков попыток фишинга, нацеленных на медицинский персонал или пациентов.

Обнаруживая и блокируя фишинговые угрозы, поставщики медицинских услуг могут снизят риск утечки данных, защитить конфиденциальность пациентов и обеспечить непрерывность критически важных медуслуг.

E. Электронная коммерция:

Интернет-магазины могут интегрировать возможности обнаружения фишинга в свои платформы электронной коммерции, чтобы защитить клиентов от фишинговых атак, которые приведут к захвату аккаунта, мошенническим транзакциям и краже личных данных.

Обнаружение в режиме реального времени может помочь идентифицировать и блокировать попытки фишинга, отправляемые с помощью поддельных электронных писем с подтверждением заказа, запросов на верификацию учётной записи или в службу поддержки клиентов.

Активно противодействуя фишинговым угрозам, компании электронной коммерции смогут поддерживать доверие клиентов, сокращать возвратные платежи и потери от мошенничества, а также защищать репутацию бренда.

IV. Предлагаемое решение

Ключевыми аспектами являются извлечение релевантных артефактов из веб-запросов, использование обученной статистической модели для оценки запросов, обновление модели с течением времени и формирование предупреждений, когда оценка превышает пороговое значение. Ключевыми компонентами являются:

Извлечение признаков:

- При получении запроса на доступ к сетевому ресурсу извлекается набор признаков, связанных с запросом.
- Эти признаки могут включать полное доменное имя (FQDN), возраст домена, регистратора домена, IP-адрес, географическое местоположение и т.д.
- Извлечение признаков позволяет представить ключевые атрибуты веб-запроса, которые могут указать, является ли он потенциальной попыткой фишинга.

Статистическая модель:

- Извлечённые признаки вводятся в предварительно обученную статистическую модель, которая выдаёт оценку вероятности.
- Модель обучается с использованием методов машинного обучения на наборах данных, содержащих известные фишинговые и нефишинговые примеры.
- Могут использоваться различные модели ML, такие как логистическая регрессия, деревья принятия решений, нейронные сети и т.д.
- Модель распознает шаблоны и комбинации значений признаков указывающие на фишинг.

Обучение и обновление модели:

- Статистическая модель изначально обучается на помеченном наборе данных перед развёртыванием.
- Набор периодически обновляется с использованием новых обучающих данных, чтобы адаптироваться к развивающимся моделям фишинга.
- Обновление модели позволяет ей распознавать новые методы фишинга и сохранять точность с течением времени.

Установление порогового значения и формирование предупреждений:

- Результатом работы модели является оценка вероятности того, что веб-запрос является попыткой фишинга.
- Если оценка превышает заранее определённый порог, генерируется оповещение.
- Пороговое значение может быть скорректировано для настройки чувствительности системы на основе желаемого соотношения частоты ложноположительных и ложноотрицательных срабатываний.

- Могут быть предприняты защитные действия, такие как блокировка веб-запроса при срабатывании оповещения.

V. Технологический процесс

Технологический процесс охватывает полный жизненный цикл предлагаемой системы обнаружения фишинга, от первоначального сбора данных и разработки модели до развёртывания в режиме реального времени, формирования предупреждений и непрерывного обновления модели. Ключевыми этапами являются извлечение признаков, обучение и оценка модели, оценка сетевых запросов в режиме реального времени, формирование предупреждений и ответов, а также периодическая переподготовка модели для адаптации к меняющимся тактикам фишинга.

Сбор и предварительная обработка данных:

- Сбор набор данных известных фишинговых и легитимных запросов к сетевым ресурсам.
- Предварительная обработка необработанных данных запроса для извлечения соответствующих признаков, таких как URL, возраст домена, регистратор, IP-адрес, географическое местоположение и т.д.
- Разметка запроса на фишинговый или безвредный.

Извлечение признаков:

- Определение набора отличительных признаков, которые могут отличить попытки фишинга от легнтимных запросов, на основе знаний предметной области и предварительных исследований.
- Реализация логики извлечения объектов для анализа соответствующих атрибутов из предварительно обработанных данных запроса.
- Преобразование извлечённых значений признаков в подходящий формат (например, числовые векторы) для ввода в модель машинного обучения.

Обучение модели:

- Выбор алгоритма машинного обучения для задачи классификации фишинга (например, случайный лес, SVM, нейронные сети).
- Разделение наборов данных на обучающие и тестирующие подмножества.
- Обучение выбранной модели на шаблонах, которые сопоставляют входные признаки с фишинговыми / вредоносными метками.
- Настройка параметров модели с помощью таких методов, как перекрёстная проверка, для оптимизации производительности.

Оценка модели:

- Оценка производительности обученной модели на длительном тестировании.
- Расчёт оценочных показателей, например точности, прецизионности, F1-параметра и т.д.
- Анализ производительности модели для оценки её эффективность в обнаружении попыток фишинга и определения области для улучшения.

Развёртывание модели:

- Интеграция обученной модели обнаружения фишинга в систему оперативного сетевого мониторинга.
- Извлечение признаков (характеристик) из входящих сетевых запросов в режиме реального времени.
- Применение модели к признакам каждого запроса, чтобы получить оценку вероятности фишинга.
- Сравнение полученного значения с заданным пороговым, чтобы классифицировать запрос как фишинговый или доброкачественный.

Формирование оповещений и реагирование на них:

- Если показатель фишинг-запроса превышает пороговое значение, формируется оповещение с соответствующими сведениями, такими как URL, IP источника, оценка риска и т.д.
- Доставка оповещения группам безопасности по соответствующим каналам, таким как электронная почта, SMS, интеграция SIEM и т.д.
- Запуск автоматических ответных действий в зависимости от серьёзности предупреждения, такие как блокировка запроса или помещение связанного сетевого трафика на карантин.
- Проведение ручного расследования и исправления высокоприоритетных событий ИБ-аналитиками.

Обновление модели:

- Накопление и сбор новых примеров фишинговых и вредоносных запросов в рабочей среде.
- Переобучение модели обнаружения фишинга на обновлённом наборе данных.
- Оценка производительности обновлённой модели и развёртывание для замены существующей модели в случае выявления повышенной точности.
- Отслеживание прогнозов модели с течением времени для обнаружения отклонения от значений или снижение производительности, которые могут потребовать дальнейших обновлений.

VI. Извлечение признаков

Извлечение признаков является ключевым этапом процесса обнаружения фишинга. Оно включает в себя идентификацию и выбор соответствующих характеристик или атрибутов из необработанных данных запроса сетевого ресурса. Извлечённые характеристики, такие как полное доменное имя, возраст домена, регистратор, IP-адрес и местоположение, служат входными данными для

статистической модели динамической оценки риска фишинга.

Цель состоит в том, чтобы преобразовать данные запроса в набор информативных характеристик, которые могут быть введены в статистическую модель для определения того, является ли запрос потенциально вредоносным.

- **Полное доменное имя (FQDN)**: полное доменное имя запрашиваемого ресурса, которое включает имя хоста, поддомен (при наличии), домен второго уровня и домен верхнего уровня (TLD). Например, "mail.example.com" — это полное доменное имя, где "mail" — это имя хоста, "example" — это домен второго уровня, а ".com" — это TLD.
- **Возраст домена**: относится к тому, как давно было зарегистрировано доменное имя. Недавно зарегистрированные домены с большей вероятностью будут связаны с попытками фишинга.
- **Регистратор домена**: организация, через которую было зарегистрировано доменное имя. Определённые регистраторы могут чаще использоваться фишинговыми сайтами.
- **IP-адрес**: числовая метка, присвоенная серверу, на котором размещён запрашиваемый ресурс.
- **Географическое местоположение**: физическое местоположение сервера на основе его IP-адреса. Запросы, исходящие из неожиданных географических регионов, могут указывать на более высокий риск фишинга.

Извлечение этих составных-признаков позволяет представить ключевые элементы запроса в структурированном формате, который может быть проанализирован с помощью статистической модели. Значения признаков преобразуются и нормализуются, чтобы сделать их пригодными для ввода в алгоритм машинного обучения.

Дополнительные признаки могут также быть извлечены в зависимости от конкретной реализации. Процесс извлечения признаков, по сути, преобразует необработанные данные запроса в вектор релевантных атрибутов, которые кратко отражают информацию, необходимую для оценки риска фишинга.

Путём тщательной разработки и выбора признаков можно оптимизировать точность и эффективность последующей модели обнаружения фишинга. Выделенные признаки предназначены для сбора шаблонов и сигналов, которые отличают легитимные запросы от попыток фишинга на основе домена, сервера и характеристик запроса.

VII. Статистическая модель

Статистическая модель является ключевым элементом динамической системы обнаружения фишинга. Он принимает извлечённые характеристики запроса сетевого ресурса в качестве входных данных и выводит оценку вероятности, указывающую на вероятность того, что запрошенный ресурс представляет собой угрозу фишинга.

Тип модели: предлагается использовать ML-методы для обучения статистической модели, в частности, упоминается метод случайного леса как одна из возможных реализаций. Метод случайного леса — это метод коллективного обучения, который создаёт несколько деревьев решений и выводит класс, который является способом вывода классов отдельными деревьями. Он известен своей способностью хорошо обобщать новые данные.

Входные данные модели: входными данными для модели является набор признаков, извлечённых из запроса сетевого ресурса, таких как полное доменное имя, возраст домена, регистратор, IP-адрес, географическое местоположение и т.д. Эти объекты преобразуются и нормализуются в подходящий формат (например, вектор объектов) перед загрузкой в модель.

Выходные данные модели: выходные данные модели представляют собой оценку вероятности от 0 до 1, которая предполагает вероятность того, что запрошенный ресурс является попыткой фишинга. Если оценка превышает заранее установленный порог (например, 0,8), ресурс классифицируется как потенциальная угроза фишинга.

Обучение модели: Статистическая модель обучается на наборе данных, содержащем примеры известных фишинговых и нефишинговых (доброкачественных) сетевых ресурсов. Модель учится распознавать шаблоны и комбинации значений признаков, указывающих на фишинг. Алгоритм случайного леса корректирует параметры модели, чтобы свести к минимуму ошибки неправильной классификации.

Оценка модели: производительность обученной модели оценивается с использованием таких показателей, как точность, прецизионность, оценка F1 и т.д., в отдельном наборе тестов. Это помогает понять, насколько хорошо модель обобщается на невидимые данные, и направляет выбор модели и настройку параметров.

Обновление модели: для адаптации к меняющимся тактикам фишинга статистическую модель можно периодически переподготовлять с использованием новых помеченных данных. Это позволяет модели изучать новые шаблоны и сохранять свою точность с течением времени по мере изменения характеристик попыток фишинга.

Статистическая модель представляет собой классификатор машинного обучения, лежащий в основе динамической системы обнаружения фишинга. Она обучена прогнозировать вероятность того, что сетевой ресурс представляет собой фишинговую угрозу, на основе его выделенных характеристик. Архитектура модели, процедура обучения и стратегия обновления разработаны таким образом, чтобы обеспечить точное, адаптивное выявление попыток фишинга в режиме реального времени.

Использование статистического подхода, основанного на данных, позволяет системе извлекать сложные закономерности из исторических данных о фишинге и

обобщать их для обнаружения новых, ранее невиданных попыток. Это обеспечивает более динамичную и надёжную защиту по сравнению со статическими методами, основанными на правилах.

VIII. Обучение и обновление модели

Обучение и обновление модели относятся к процессам первоначального построения статистической модели на основе обучающего набора данных и последующего её уточнения с течением времени по мере поступления новых данных. Это важнейшая часть конвейера машинного обучения, которая позволяет системе обнаружения фишинга адаптироваться и поддерживать точность перед лицом возникающих угроз.

Начальное обучение модели:

- Перед развёртыванием статистическая модель (например, классификатор случайного леса) обучается на помеченном наборе данных, содержащем примеры известных фишинговых и вредоносных запросов к сетевым ресурсам.
- Каждый обучающий пример состоит из извлечённых признаков (полное доменное имя, возраст домена, регистратор, IP, местоположение и т.д.) и соответствующего ярлыка (фишинговый или вредоносный).
- Во время обучения модель учится распознавать шаблоны и комбинации значений признаков, которые отличают попытки фишинга от легитимных запросов.
- Параметры модели оптимизированы для минимизации ошибок прогнозирования в обучающих данных.

Периодическое обновление модели:

- Подчёркивается важность периодического обновления модели новыми помеченными данными для адаптации к меняющимся тактикам фишинга.
- По мере появления новых типов фишинговых атак характеристики их запросов со временем могут меняться.
- Обновление модели позволяет изучить эти новые шаблоны, сохраняя при этом информацию о ранее замеченных признаках фишинга.
- Частоту обновлений модели можно регулировать в зависимости от объёма и скорости сбора новых фишинговых данных.

Непрерывное обучение:

- Некоторые архитектуры машинного обучения, такие как онлайн-обучение или инкрементное обучение, специально разработаны для поддержки непрерывного обновления модели по мере поступления новых данных.
- Вместо переподготовки всего совокупного набора данных эти методы постепенно корректируют параметры модели на основе мини-пакетов новых примеров.
- Непрерывное обучение помогает снизить вычислительную нагрузку, связанную с повторным переобучением, и позволяет быстрее адаптироваться к новым угрозам.

Управление данными:

- Эффективное обновление модели требует тщательного управления обучающими данными с течением времени.
- Набор данных необходимо дополнить новыми примерами фишинга и вредоносными примерами при сохранении баланса между классами.
- Такие методы, как активное обучение, используют для стратегического выбора наиболее информативные примеры для маркировки, оптимизируя использование человеческих усилий по аннотированию.

Оценка и мониторинг:

- После каждого обновления переподготовленную модель следует оценивать с помощью отдельного набора тестов, чтобы убедиться в её производительности и отсутствии ухудшений
- Постоянный мониторинг прогнозов модели в процессе производства также важен для выявления отклонений от концепции или ошибок, которые могут потребовать дальнейших обновлений.

Обучение и обновление модели необходимы для долгосрочной эффективности системы обнаружения фишинга. Начальный процесс обучения формирует базовые знания модели, в то время как периодические обновления позволяют ей со временем адаптироваться к новым моделям фишинга, а непрерывное обучение, активный отбор данных и мониторинг производительности, помогают оптимизировать процесс обновления и поддерживать точность модели перед лицом возникающих угроз.

IX. Установление порогового значения и формирования предупреждений

Установление порогового значения и формирование предупреждений относятся к процессу принятия решения о том, следует ли классифицировать данный запрос сетевого ресурса как попытку фишинга, на основе оценки вероятности, выводимой статистической моделью, и выдачи соответствующего предупреждения, если решение положительное. Это шаг, который преобразует прогнозы модели в фактические решения по обеспечению безопасности и уведомления.

Порог оценки вероятности:

- Статистическая модель выводит оценку от 0 до 1 для каждого запроса сетевого ресурса, указывающую на предполагаемую вероятность того, что это попытка фишинга.

- Для принятия окончательного решения о классификации используется предварительно заданное пороговое значение (например, 0,8).
- Если оценка превышает пороговое значение, запрос классифицируется как потенциальная угроза фишинга; иначе - считается безопасным.

Выбор порогового значения:

- Выбор порогового значения предполагает компромисс между ложноположительными результатами (легитимные запросы ошибочно классифицируются как фишинговые) и ложноотрицательными результатами (попытки фишинга ошибочно классифицируются как безвредные).
- Более высокий порог снижает количество ложных срабатываний, но может пропустить некоторые реальные попытки фишинга. Более низкий порог улавливает больше фишинговых запросов, но также помечает больше безопасных запросов.
- Оптимальный порог определяется на основе конкретных требований безопасности и относительной стоимости ложных срабатываний и ложноотрицательных результатов в контексте развёртывания.

Формирование оповещений:

- Когда оценка запроса превышает пороговое значение фишинга, формируется предупреждение, указывающее на потенциальную угрозу.
- Оповещение может включать соответствующие сведения о запросе, такие как запрошенный URL-адрес, IP-адрес источника, соответствующий показатель вероятности и т.д.
- Оповещения доставляются по различным каналам, таким как журналы консоли, уведомления по электронной почте, SMS-сообщения, системы управления инцидентами и событиями безопасности (SIEM) и т.д.

Проверка и фильтрация оповещений:

- Чтобы уменьшить количество ложных срабатываний, сформированные оповещения могут проходить дополнительные этапы проверки перед их эскалацией.
- Это может включать сравнение сведений о предупреждении со списками разрешений известных безопасных ресурсов, проверку на наличие потока предупреждений из того же источника или применение других эвристических фильтров.
- Ручная проверка подмножества предупреждений аналитиками безопасности может помочь со временем настроить пороговые значения и правила проверки.

Действия по реагированию на предупреждение:

- В зависимости от серьёзности и достоверности классификации фишинга предупреждения могут инициировать различные ответные действия.
- Предупреждения с меньшей степени серьёзности регистрируются для последующего анализа, в то время как предупреждения с большей степени серьёзности вызывают немедленную блокировку запроса ресурса и помещение связанного сетевого трафика на карантин.
- Автоматические реакции дополнены действиями по расследованию и исправлению, выполняемыми вручную, на основе сведений о предупреждении.

Установление порогового значения и формирование предупреждений устраняют разрыв между вероятностными прогнозами модели обнаружения фишинга и детерминированными решениями и действиями в области безопасности, необходимыми для защиты пользователей и систем. Выбирая соответствующие пороговые значения, формируя информативные предупреждения и запуская пропорциональные ответные действия, этот компонент вводит в действие разведданные, собранные моделью, для обеспечения эффективной защиты от фишинга.

X. Преимущества,-недостатки и значимость предлагаемого решения

Этот патент иллюстрирует важную эволюцию от реактивного обнаружения фишинга на основе сигнатур к более динамичному адаптивному подходу, основанному на статистическом моделировании.

Патент представляет автоматизированный подход, основанный на данных, для обнаружения попыток фишинга в режиме реального времени путём изучения обобщённых шаблонов вместо использования статических правил. Динамический характер позволяет адаптироваться к развивающимся методам фишинга. Формирование вероятностных оценок риска позволяет определять приоритетность наиболее подозрительных случаев.

Описывая гибкий конвейер машинного обучения с извлечением признаков, обновлением модели и формированием предупреждений, патент обеспечивает основу для создания более эффективных антифишинговых систем. Предлагаемый метод может значительно улучшить способность организации активно выявлять и блокировать фишинговые угрозы до того, как они станут жертвами пользователей. Однако для его внедрения и обслуживания требуется значительный сбор данных и инженерные усилия.

Статистическая модель обучается на исторических примерах фишинга и неопасных примерах для изучения закономерностей, которые различают эти два класса. Его можно периодически переподготовлять на основе новых данных, чтобы адаптироваться к меняющимся тактикам фишинга.

Основные преимущества:

- Обеспечивает упреждающее обнаружение попыток фишинга в режиме реального времени, включая новые, невиданные ранее атаки, путём анализа шаблонов в атрибутах URL / домена
- Предоставляет оценку вероятности, позволяющую определить приоритетность наиболее опасных угроз
- Адаптируется к меняющимся тактикам фишинга с течением времени посредством периодической переподготовки модели
- Формирует информативные оповещения с ключевыми запросами для расследования группами безопасности
- Позволяет настраивать чувствительность обнаружения путем настройки порога оповещения

Недостатки:

- Требуются значительные исторические и достоверные данные о фишинге для начального обучения модели
- Необходим постоянный сбор помеченных данных для переподготовки и обновления модели с течением времени
- Могут отсутствовать некоторые новые шаблоны фишинга, не отражённые в обучающих данных
- Извлечение эффективного набора признаков требует тщательного проектирования и экспертных знаний в предметной области
- Может генерировать ложноположительные результаты, которые требуют дополнительной фильтрации / проверки

A. *Извлечение признаков*

Извлечение признаков является важным шагом, который позволяет создавать эффективные модели ML для обнаружения фишинга путём представления данных запроса в информативном формате. Однако для разработки и поддержания надёжного набора признаков требуются значительные знания и усилия. Сочетание ручной разработки признаков с автоматическим обучением представлению может помочь устранить некоторые из этих недостатков и создать более мощные гибридные модели обнаружения.

1) *Преимущества:*
- Предоставляет ключевые характеристики запросов сетевых ресурсов в структурированном формате, подходящем для анализа с помощью моделей машинного обучения. Извлечение соответствующих признаков имеет решающее значение для построения точных моделей обнаружения фишинга.
- Фиксирует дискриминационные шаблоны, которые отличают попытки фишинга от легитимных запросов. Тщательно спроектированные признаки могут обеспечивать надёжные сигналы для классификации.
- Уменьшает размерность необработанных данных запроса, делая их обработку более эффективной с точки зрения вычислений. Работать с компактным набором информационных признаков быстрее, чем анализировать полное содержимое запроса.
- Специалисты по извлечению признаков из предметной области используют свои знания для создания высокоэффективных признаков для конкретной задачи обнаружения фишинга. Ручная разработка признаков, основанная на опыте, может дать очень эффективные наборы признаков.
- Извлечённые признаки можно комбинировать с автоматически изучаемыми признаками в процессе глубокого обучения для создания мощных гибридных моделей. Это позволяет получить максимум пользы как от ручного проектирования объектов, так и от обучения представлению.

2) *Недостатки:*
- Требуется значительный опыт работы в предметной области и ручные усилия для определения и внедрения эффективных признаков. Разработка хорошего набора признаков для обнаружения фишинга требует много времени и в значительной степени зависит от экспертных знаний.
- Разработанные признаки могут не отражать все релевантные шаблоны, особенно новые в развивающихся фишинговых атаках. Существует риск пропустить важные сигналы, о которых эксперты не подумали.
- Код извлечения признаков нуждается в регулярном обновлении, чтобы соответствовать изменениям в веб-технологиях и методах фишинга. Обслуживание конвейера признаков может быть постоянной инженерной нагрузкой.
- Извлечённые признаки специфичны для определённых типов фишинга, что ограничивает способность модели обобщаться на новые варианты атак. Чрезмерно специализированные признаки могут привести к хрупкости моделей.
- Использование исключительно признаков, разработанных вручную, может привести к снижению производительности по сравнению со сквозным глубоким обучением на необработанных данных. Для некоторых задач заученные представления могут превосходить созданные вручную признаки.

B. *Статистическая модель*

Статистические модели, особенно гибридные подходы, сочетающие инженерные признаки и глубокое обучение, предлагают мощные возможности для динамического и адаптивного обнаружения фишинга. Однако они также создают проблемы, связанные с качеством данных, проектированием признаков, сложностью вычислений и

устойчивостью к атакам противника. Эффективное развертывание требует тщательного устранения этих ограничений посредством постоянного сбора данных, обновления моделей и экспертного надзора.

1) *Преимущества:*
- Обеспечивает динамическое и адаптивное обнаружение фишинговых угроз путем изучения шаблонов из исторических данных. Статистическая модель может распознавать сложные комбинации признаков, указывающих на фишинг, помимо простых правил.
- Выводит показатель вероятности, который количественно определяет риск того, что запрос является попыткой фишинга. Это обеспечивает более детальную информацию, чем двоичная классификация, позволяя проводить детальную оценку рисков и расставлять приоритеты.
- Может обновляться с течением времени путем переподготовки на основе новых данных для адаптации к меняющимся тактикам фишинга. Прогностическая способность модели может сохраняться по мере того, как злоумышленники меняют свои методы.
- Подходит для обнаружения в режиме реального времени благодаря быстрому выводу результатов после обучения модели. Позволяет интегрировать в системы оперативного мониторинга и предотвращения.
- Гибридные модели, сочетающие ручную разработку признаков и глубокое обучение, показали более высокую точность обнаружения фишинга по сравнению с традиционными моделями ML. Использует сильные стороны как человеческого опыта, так и обучения на основе данных.

2) *Недостатки:*
- Для начального обучения требуется большой маркированный набор данных, получение которого может быть дорогостоящим и отнимать много времени. Наборы данных о фишинге должны постоянно обновляться, чтобы включать новые шаблоны атак.
- Производительность модели в значительной степени зависит от качества и репрезентативности обучающих данных. Предвзятые или неполные наборы данных могут привести к искаженным прогнозам и "слепым зонам".
- Разработка признаков по-прежнему играет решающую роль в построении эффективных моделей ML для обнаружения фишинга. Соответствующие признаки должны создаваться вручную, что требует значительного опыта в предметной области.
- Традиционные модели ML, такие как Random Forest, снижают производительность и не обнаруживают новые схемы фишинга, не замеченные во время обучения. Поддержание моделей в актуальном состоянии является постоянной задачей.
- Обучение моделей глубокого обучения может быть дорогостоящим с точки зрения вычислений и потребовать специализированного оборудования. Повышенная сложность также затрудняет интерпретацию и отладку моделей.
- Риск враждебных атак, при которых фишеры намеренно создают сообщения, чтобы избежать обнаружения моделью. Модели ML могут быть хрупкими и уязвимыми для манипуляций.

C. Обучение и обновление модели

Обучение и обновление модели необходимы для поддержания эффективности системы обнаружения фишинга по мере появления новых угроз. Однако этот процесс также создает операционные сложности, связанные со сбором данных, маркировкой, вычислительными ресурсами и управлением изменениями. Тщательная разработка схемы переподготовки, средств контроля качества данных и механизмов мониторинга имеет решающее значение для реализации преимуществ при одновременном смягчении недостатков.

1) *Преимущества:*
- Позволяет модели обнаружения фишинга адаптироваться к развивающимся угрозам, со временем изучая новые помеченные примеры. Периодическая переподготовка помогает модели распознавать новые схемы фишинга.
- Методы непрерывного обучения позволяют постепенно обновлять модель новыми данными, снижая вычислительные затраты по сравнению с полной переподготовкой, что обеспечивает более частое и эффективное обновление модели.
- Стратегии активного обучения могут оптимизировать отбор новых примеров для маркировки, сводя к минимуму ручные затраты на аннотирование. Это помогает управлять текущим процессом обработки данных.
- Регулярная оценка модели на новых наборах тестов гарантирует, что обновления действительно повышают производительность и не приводят к регрессиям. Мониторинг поведения модели в рабочей среде выявляет потенциальные проблемы на ранней стадии.
- Обновление модели разнообразным набором новых примеров фишинга и вредоносных программ повышает её надежность и универсальность для различных вариантов атак. Широкий набор обучающих программ помогает модели справляться с широким спектром угроз.

2) *Недостатки:*
- Требуется постоянный поток новых помеченных как фишинговые и безвредные примеров для переобучения модели, что может быть сложным и

дорогостоящим для получения в масштабе. Маркировка новых обучающих примеров требует ручной работы экспертов предметной области и занимает много времени. Разработка эффективных рабочих процессов аннотирования и интерфейсов имеет решающее значение.

- Если распределение новых обучающих данных значительно отличается от исходных, в обновленной модели может наблюдаться снижение производительности или нестабильность.
- Частые обновления моделей могут быть дорогостоящими с точки зрения вычислений, особенно для больших моделей глубокого обучения. Методы инкрементного обучения помогают, но все еще могут требовать значительных ресурсов.
- Обновление модели изменяет её поведение, что негативно сказывается на последующих системах и рабочих процессах, основанных на её прогнозах.
- Существует риск того, что модель будет переоснащена недавним учебным примерам и потеряет производительность при использовании старых фишинговых шаблонов. Сбалансировать сочетание старых и новых данных во время переподготовки непросто.

D. Установление порогового значения и формирование предупреждений

Определение порога и формирование предупреждений играют решающую роль в реализации модели обнаружения фишинга путем преобразования её вероятностных результатов в конкретные действия по обеспечению безопасности. Однако этот процесс также сопряжен с проблемами, связанными с настройкой порога, управлением ложными срабатываниями и переутомлением при оповещении. Тщательный дизайн и постоянное совершенствование логики определения пороговых значений в сочетании с производительностью модели являются ключом к достижению баланса между снижением рисков и операционной эффективностью.

1) Преимущества:

- Позволяет преобразовать вероятностные выходные данные статистической модели в практические решения по обеспечению безопасности. Сравнивая показатель вероятности фишинга модели с заданным порогом, система может автоматически определить, следует ли помечать запрос как потенциальную угрозу.
- Предоставляет настраиваемый параметр (пороговое значение) для балансирования компромисса между ложноположительными и ложноотрицательными результатами. Настройка порогового значения позволяет контролировать чувствительность оповещений на основе их толерантности к риску и операционных ограничений.
- Позволяет генерировать информационные оповещения с соответствующими сведениями о подозрительном запросе, такими как URL-адрес, IP-адрес источника и соответствующая оценка риска. Эта контекстуальная информация помогает службам безопасности быстро выявлять и расследовать потенциальные случаи фишинга.
- Поддерживает гибкие каналы доставки оповещений, такие как журналы консоли, электронная почта, SMS или интеграция с системами информации о безопасности и управления событиями (SIEM).
- Позволяет реализовать дополнительную логику проверки и фильтры для дальнейшего уменьшения ложных срабатываний. Например, оповещения могут быть отключены для доменов, внесенных в белый список, или диапазонов IP-адресов, или если показатель достоверности модели ниже определенного уровня.

2) Недостатки:

- Выбор соответствующего порогового значения требует тщательной настройки и может быть связан с методом проб и ошибок. Слишком низкое значение порога приведёт к большому количеству ложных срабатываний, в то время как слишком высокое значение – к пропуску реальных попыток фишинга.
- Требуется регулярно корректировать оптимальный порог по причине изменения с течением времени характеристик фишинговых атак, что, в свою очередь, требует постоянного мониторинга и анализа производительности системы и меняющегося ландшафта угроз.
- Пороговое значение сводит обширную информацию, предоставляемую оценкой вероятности модели, к бинарному решению (оповещение или отсутствие оповещения). Это приведёт к потере нюансов и детализации при оценке риска пограничных случаев.
- Предупреждения, сформированные системой, по-прежнему могут требовать ручной проверки и расследования аналитиками безопасности. Хотя установление порогового значения помогает определить приоритетность наиболее подозрительных случаев, оно не устраняет полностью необходимость в суждениях и вмешательстве человека.
- Эффективность оповещений зависит от точности лежащей в основе статистической модели. Если прогнозы модели неправильно откалиброваны, даже хорошо настроенный порог может привести к неоптимальным результатам.