Аннотация – Анализ документа "AntiPhishStack: модель многоуровневого обобщения на основе LSTM для оптимизированного обнаружения фишинговых URL", будет охватывать различные аспекты, включая методологию, результаты и последствия для кибербезопасности. В частности, будет рассмотрен подход документа к использованию сетей с долгой краткосрочной памятью (LSTM) в рамках многоуровневой структуры обобщения для обнаружения фишинговых URL-адресов. Будет изучена эффективность модели, стратегии её оптимизации и её производительность по сравнению с существующими методами.

В ходе анализа также будут рассмотрены практические применения модели, способы её интеграции в существующие меры кибербезопасности и её потенциальное влияние на сокращение числа фишинговых атак. Подчёркнута актуальность документа для специалистов по кибербезопасности, ИТ-специалистов и заинтересованных сторон в различных отраслях, а также важность передовых методов обнаружения фишинга в современном цифровом ландшафте.

Это изложение послужит ценным ресурсом для экспертов по кибербезопасности, ИТ-специалистов и других лиц, интересующихся последними разработками в области обнаружения и предотвращения фишинга.

I. Введение

В документе под названием "Модель многоуровневого обобщения на основе LSTM для оптимизации фишинга" обсуждается растущая зависимость от революционных онлайновых веб-сервисов, что привело к повышенным рискам безопасности и постоянным проблемам, создаваемым фишинговыми атаками.

Фишинг, вводящий в заблуждение метод социальной и технической инженерии, представляет серьёзную угрозу безопасности в Интернете, направленный на незаконное получение идентификационных данных пользователей, их личного счета и банковских учётных данных. Это основная проблема преступной деятельности, когда атакующие преследуют такие цели, как продажа украденных личных данных, извлечение наличных, использование уязвимостей или получение финансовой выгоды.

Исследование направлено на улучшение обнаружения фишинга с помощью AntiPhishStack, работающего без предварительного знания особенностей фишинга. Модель использует возможности сетей долгой краткосрочной памяти (LSTM), типа рекуррентной нейронной сети, которая способна изучать зависимость порядка в задачах прогнозирования последовательности. Он симметрично использует изучение URL-адресов и функций TF-IDF на уровне символов, повышая его способность бороться с возникающими фишинговыми угрозами.

II. Методология и значимость исследования

В документе представлена новая модель обнаружения фишинговых сайтов. Важность этого исследования заключается в совершенствовании методов обнаружения фишинга, в частности, за счёт внедрения обобщённой двухфазной стековой модели, названной AntiPhishStack.

Эта модель предназначена для обнаружения фишинговых сайтов, не требуя предварительного знания особенностей, специфичных для фишинга, что является значительным улучшением по сравнению с традиционными системами обнаружения, которые полагаются на машинное обучение и ручные функции.

Это исследование вносит вклад в продолжающийся дискурс о симметрии и асимметрии в информационной безопасности и предоставляет перспективное решение для повышения сетевой безопасности перед лицом развивающихся киберугроз.

Источник данных, использованный в исследовании, включает два контрольных набора, содержащих доброкачественные и фишинговые или вредоносные URL-адреса. Эти наборы данных используются для экспериментальной проверки модели. В документе наборы данных обозначены как DS1 и DS2, причём DS1 включает доброкачественные сайты Яндекса и фишинговые сайты PhishTank, а DS2 состоит из доброкачественных сайтов из common-crawl, базы данных Alexa и фишинговых сайтов из PhishTank.

III. Ключевые компоненты

Антифиш-стековая модель работает в два этапа (обобщённая модель двухфазного стека):

- **Этап I**: модель симметрично запоминает URL-адреса и функции TF-IDF на уровне символов. Эти функции обучаются на базовом классификаторе машинного обучения, использующем K-кратную перекрёстную проверку для надёжного прогнозирования среднего значения.
- **Этап II**: для динамической компиляции используется двухуровневая многоуровневая сеть LSTM с пятью адаптивными оптимизаторами,

обеспечивающими превосходное прогнозирование этих функций.

- Кроме того, симметричные прогнозы на обоих этапах оптимизированы и интегрированы для обучения мета-классификатора XGBoost, что способствует получению окончательного надёжного прогноза.

A. URL-особенности

- **Структура URL-адресов**: в документе подчёркивается, что злоумышленники часто создают фишинговые URL-адреса, которые кажутся пользователям законными. Они используют тактику блокирования URL-адресов, чтобы обманом заставить пользователей раскрыть личную информацию.
- **Легкие функции**: исследование направлено на обнаружение фишинговых веб-сайтов с использованием облегчённых функций, в частности системы маркеров URL с весовым коэффициентом, которые позволяют быстро обнаруживать их без доступа к содержимому веб-сайта.
- **Вычисление веса**: приводится формула для вычисления весat W_i for i-th d неопределённого слова в URL-адресе, которая используется для присвоения значения веса каждому URL-адресу для прогнозирования фишинга.
- **Компоненты URL**: описываются компоненты URL-адреса, включая протокол, IP-адрес хоста или местоположение ресурса, основные домены, домены верхнего уровня (TLD), номер порта, путь и необязательные поля, такие как запрос.
- **Индикаторы фишинга**: несколько дополнительных признаков идентифицируются как индикаторы фишинга, такие как использование IP-адреса вместо доменного имени, наличие символа "@", символа "//", префиксов и суффиксов доменных имён, разделённых знаком "-", и использование нескольких поддоменов.
- **HTTPS и возраст сертификата**: отмечается, что большинство законных сайтов используют HTTPS, и возраст сертификата имеет решающее значение. Требуется сертификат, заслуживающий доверия.
- **Favicon**: favicon может использоваться для перенаправления клиентов на сомнительные сайты, когда он находится во внешнем пространстве.
- **Анализ вспомогательных функций**: в документе представлен анализ вспомогательных функций, таких как IP-адрес, символ "@", символ "//", префиксы и суффиксы доменных имён, HTTPS и значок, объясняющий, как эти функции можно использовать для идентификации фишинговых веб-сайтов

B. Символьные особенности

- **TF-IDF**: используется термин, обратный частоте документа (TF-IDF) на уровне символов, чтобы определить относительную важность символов в URL-адресах по всему корпусу анализируемых URL-адресов.
- **Расчёт TF-IDF**: оценка TF-IDF состоит из двух частей: частоты использования термина (TF), которая представляет собой нормированное количество терминов в документе, и обратной частоты использования документа (IDF), которая состоит из логарифмов отношения общего количества документов к количеству документов, содержащих термин.
- **Уровни TF-IDF**: упоминается, что векторы TF-IDF могут генерироваться на разных уровнях, таких как уровень слова, уровень символа и уровень n-граммы, причём уровень символа особенно важен для данного исследования.
- **Ограничения TF-IDF**: хотя TF-IDF полезен для извлечения важных ключевых слов, у него есть ограничения, такие как невозможность извлечения терминов с орфографическими ошибками, что может быть проблематичным, поскольку URL-адреса могут содержать бессмысленные слова.
- **Символьный TF-IDF**: чтобы устранить ограничения TF-IDF для URL-адресов, которые могут содержать орфографические ошибки или бессмысленные слова, в исследовании используется подход TF-IDF на уровне символов с максимальным количеством функций 5000.
- **Естественное изучение функций**: модель обрабатывает строки URL как последовательности символов, которые считаются естественными функциями, не требующими предварительного знания функций для эффективного изучения моделью.
- **Обобщение стека для извлечения объектов**: модель использует обобщение стека для извлечения локальных объектов URL из последовательностей символов, а для окончательного прогнозирования разработан метаклассификатор.
- **Преимущества подхода**: подход позволяет предлагаемой модели обучаться на последовательностях символов URL как естественных признаках, что упрощает процесс обучения и потенциально улучшает способность модели обнаруживать фишинговые URL-адреса без предварительного знания особенностей

C. Модель обобщения стека

- **Двухфазный подход**: модель разделена на две фазы. На этапе I используются классификаторы машинного обучения для генерации среднего прогноза, в то время как на этапе II используется двухуровневая стековая обобщённая модель на

основе LSTM, оптимизированная для наилучшего прогнозирования при обнаружении фишинговых сайтов.

- **Интеграция прогнозов**: средний прогноз из фазы I объединяется с основным прогнозом из фазы II. Затем для получения окончательного прогноза используется метаклассификатор, в частности XGBoost.
- **Метод обобщения стека**: в модели используется обобщение стека, методология коллективного обучения, которая объединяет различные алгоритмы машинного обучения и модели глубокого обучения для повышения эффективности обнаружения.
- **Model Flow**: включает в себя сбор наборов данных, разделение их на обучающие и тестовые наборы, построение этапов модели обобщения стека и объединение прогнозов для получения окончательного.
- **Важность функции**: модель подчёркивает важность функций TF-IDF на уровне URL и символов, которые используются симметрично для обнаружения фишинговых веб-страниц.
- **Существенные преимущества**: модель обладает рядом преимуществ, включая независимость от предварительного знания функций, высокую способность к обобщению и независимость от экспертов по кибербезопасности и сторонних сервисов.
- **Улучшенное обнаружение фишинга**: модель предназначена для интеллектуального выявления новых фишинговых URL-адресов, ранее не идентифицированных как мошеннические, демонстрируя надёжную работу на контрольных наборах данных.

D. Эксперименты

Представлена экспериментальная проверка предложенной модели. Она была протестирована на двух контрольных наборах данных, которые включали доброкачественные и фишинговые или вредоносные URL-адреса.

- Модель продемонстрировала исключительную производительность при обнаружении фишинговых сайтов, достигнув точности 96,04%. Этот результат был заметно выше по сравнению с существующими исследованиями.
- Модель оценивалась с помощью различных матриц, включая кривую AUC-ROC, точность, отзыв, F1, среднюю абсолютную ошибку (MAE), среднеквадратичную ошибку (MSE) и точность.
- Сравнительный анализ с базовыми моделями и традиционными алгоритмами машинного обучения, такими как метод опорных векторов, дерево решений, наивный байесовский алгоритм, логистическая регрессия, метод K-ближайших соседей и последовательная минимальная оптимизация, выявил превосходную эффективность обнаружения фишинга в модели.
- Было установлено, что эта модель эффективна при выявлении новых фишинговых URL-адресов, которые ранее не были идентифицированы как мошеннические.
- Модель работает без предварительного знания особенностей фишинга, что является значительным преимуществом в достижении прогресса в области кибербезопасности

E. Оценка оптимизатора в LSTM

- **Производительность оптимизатора**: в статье оценивается производительность пяти различных адаптивных оптимизаторов: AdaDelta, Adam, RMSProp, AdaGard и SGD (Stochastic Gradient Descent), чтобы определить, какой из них лучше всего подходит для предлагаемой модели защиты от фишинга.
- **Эпохи и скорость обучения**: для реализации двухуровневого LSTM с разными оптимизаторами рассматривается разное количество эпох. Скорость обучения, важнейший параметр, настраивается для каждого оптимизатора, для контроля модели.
- **Точность, MSE и MAE**: в документе указаны точность, среднеквадратичная ошибка (MSE) и средняя абсолютная ошибка (MAE) для каждого оптимизатора с использованием модели обобщения стека на основе LSTM на двух наборах данных (DS1 и DS2).
- **Результаты для наборов данных**: оптимизатор AdaGard обеспечил высочайшую точность при минимальных значениях MSE и MAE в DS1, в то время как оптимизатор Adam достиг наивысшей точности в DS2.
- **Кривые точного воспроизведения**: кривые точного воспроизведения представлены для каждого набора функций, указывая на компромисс между точностью и повторным воспроизведением для различных оптимизаторов.
- **Выбор оптимизатора**: анализ показывает, что скорость обучения в значительной степени способствует успеху предлагаемой модели с адаптивными оптимизаторами. Оптимизатор Adam выделяется своей производительностью с определённой скоростью обучения при использовании двухуровневого LSTM со 100 эпохами.
- **Сравнительный анализ**: сравнивается средняя производительность оптимизаторов на DS1 и DS2, при этом DS2 показывает несколько лучшую точность.

- **Значимость оптимизаторов**: оценка оптимизаторов имеет решающее значение для точности модели, которая является ключевым компонентом машинного обучения и искусственного интеллекта, отвечающим за формирование модели для получения наиболее точных результатов из возможных

IV. Ключевые выводы

Конструкция модели позволяет эффективно идентифицировать новые фишинговые URL-адреса, ранее не идентифицированные как мошеннические, тем самым снижая вероятность ложноотрицательных результатов. Использование K-кратной перекрёстной проверки и двухуровневой сети LSTM помогает предотвратить переоснащение и улучшить способность модели правильно классифицировать фишинговые сайты, тем самым снижая вероятность ложных срабатываний.

- **Разработка модели**: новый режим, внедрённый с помощью обобщённой модели двухфазного стека, предназначенной для эффективного обнаружения фишинговых сайтов.
- **Симметричное изучение URL-адресов и функций TF-IDF на уровне символов**: в модели симметричное изучение URL-адресов и функций TF-IDF на уровне символов. Это повышает способность модели бороться с возникающими фишинговыми угрозами.
- **Двухфазная работа**: на этапе I функции обучаются на базовом классификаторе машинного обучения с использованием K-кратной перекрёстной проверки для надёжного прогнозирования среднего значения. На этапе II используется двухуровневая многоуровневая сеть LSTM с пятью адаптивными оптимизаторами для динамической компиляции, обеспечивающими превосходное прогнозирование этих функций.
- **Интеграция прогнозов (Мета-классификатор XGBoost)**: симметричные прогнозы на обоих этапах оптимизированы и интегрированы для обучения мета-классификатора XGBoost, что способствует получению окончательного надёжного прогноза.
- **Независимость от предварительного знания функций, специфичных для фишинга**: модель работает без предварительного знания функций, специфичных для фишинга, что является значительным достижением в его обнаружении, которое демонстрирует сильную способность к обобщению и независимость от экспертов по кибербезопасности и сторонних сервисов.
- **Высокая производительность**: проверка (экспериментальная) на двух контрольных наборах данных, включающих «доброкачественные» и фишинговые или вредоносные URL-адреса, демонстрирует производительность модели, достигая заметной точности 96,04% по сравнению с существующими исследованиями
- **Независимость от экспертов по кибербезопасности и сторонних сервисов:** модель самостоятельно извлекает необходимые функции URL, устраняя зависимость от экспертов по кибербезопасности. Она также демонстрирует независимость от функций сторонних производителей, таких как рейтинг страницы или возраст домена
- **Независимость от предварительного знания функций**: подход, использованный в этой работе, рассматривает строки URL как последовательности символов, выступающие в качестве естественных функций, которые не требуют предварительного знания для эффективного изучения предлагаемой моделью
- **Повышение сетевой безопасности**: исследование добавляет ценности продолжающемуся обсуждению симметрии и асимметрии в информационной безопасности и предлагает перспективное решение для повышения сетевой безопасности перед лицом развивающихся киберугроз.

V. Преимущества и ограничения исследования

Для сравнения, традиционные фишинговые системы, основанные на машинном обучении и ручных функциях, борются с эволюционирующими тактиками. Другие модели, такие как модель CNN-LSTM и архитектура сквозного глубокого обучения, основанная на методах обработки естественного языка, показали ограничения в их обобщении тестовых данных и их зависимости от существующих знаний об обнаружении фишинга. Модель AntiPhishStack, напротив, демонстрирует высокую способность к обобщению и независимость от предыдущих знаний функций, что делает её надёжным и эффективным инструментом для обнаружения фишинга.

Преимущества исследования по сравнению с традиционными фишинговыми системами включают:

- **Независимость от предварительного знания функций**: AntiPhishStack не требует предварительного знания функций, специфичных для фишинга, что позволяет ему адаптироваться к новым и развивающимся тактикам более эффективно, чем традиционные системы, которые полагаются на предопределённые функции.
- **Независимость от экспертов по кибербезопасности и сторонних сервисов**: модель автономно извлекает необходимые функции URL, уменьшая зависимость от экспертов по кибербезопасности и сторонних сервисов, таких как рейтинг страницы или возраст домена, от которых могут зависеть традиционные системы.
- **Высокая точность**: Модель продемонстрировала исключительную производительность, достигнув

заметной точности 96,04% для контрольных наборов данных, что является значительным улучшением по сравнению с традиционными системами.

- **Адаптивность к развивающимся угрозам**: Конструкция модели позволяет ей извлекать уроки из обрабатываемых данных, что потенциально делает её более адаптируемой к постоянно меняющимся тактикам, используемым атакующими, в отличие от традиционных систем, которые могут требовать обновления вручную для сохранения эффективности.

Ограничения исследования включают:

- **Применение в реальном мире**: в документе не обсуждается производительность модели в реальных сценариях, где фишинговые тактики постоянно развиваются.
- **Производительность на других наборах данных**: производительность модели была проверена на двух контрольных наборах данных, но неясно, как она будет работать на других наборах или в других контекстах.
- **Зависимость от функций**: зависимость модели от функций TF-IDF на уровне URL и символов может ограничить её способность обнаруживать попытки фишинга, использующие другие тактики.
- **Вычислительные ресурсы**: в документе не обсуждаются вычислительные ресурсы, необходимые для реализации модели, что может быть потенциальным ограничением для некоторых пользователей.

Предлагаемая модель имеет ряд ограничений с точки зрения масштабируемости и производительности.

- Во-первых, зависимость модели от сетей долгой краткосрочной памяти (LSTM) может привести к неэффективности вычислений. Сети LSTM известны своими высокими требованиями к вычислениям и памяти, что может ограничивать масштабируемость модели при работе с большими наборами данных или в приложениях реального времени.
- Во-вторых, двухэтапный подход модели, который включает в себя обучение функций в базовом классификаторе машинного обучения, а затем использование двухуровневой многоуровневой сети на основе LSTM, может потребовать много времени и вычислительных ресурсов. Это потенциально может ограничить производительность модели в сценариях обнаружения фишинга в реальном времени.
- Наконец, хотя модель предназначена для работы без предварительного знания специфических функций фишинга, это также может быть ограничением. Модели может быть сложно точно обнаруживать новые или изощренные попытки фишинга, которые используют функции, не учтённые при обучении.

VI. Значение для будущих исследований

- **Обобщение модели**: способность модели работать без предварительного знания особенностей фишинга предполагает, что будущие исследования могут быть направлены на разработку более обобщённых моделей, которые могут адаптироваться к различным типам киберугроз без обширной переподготовки.
- **Методы глубокого обучения**: успех модели на основе LSTM указывает на то, что методы глубокого обучения обладают значительным потенциалом в приложениях кибербезопасности. Будущие исследования могли бы дополнительно изучить интеграцию различных архитектур нейронных сетей и их эффективность в обнаружении угроз.
- **Извлечение признаков**: использование функций TF-IDF на уровне символов и анализа URL-адресов в модели демонстрирует важность извлечения признаков для обнаружения фишинга. Исследования могли бы быть сосредоточены на выявлении новых признаков и методов извлечения для повышения уровня обнаружения.
- **Стековое обобщение**: двухфазный подход, используемый в модели, которая объединяет классификаторы машинного обучения и сети LSTM, демонстрирует преимущества многоуровневого обобщения. В будущих исследованиях можно было бы изучить другие комбинации алгоритмов и моделей для повышения эффективности прогнозирования.
- **Эталонные наборы данных**: использование эталонных наборов данных для проверки модели подчёркивает необходимость всеобъемлющих и актуальных наборов данных в исследованиях кибербезопасности. Будущая работа может включать создание и поддержание наборов данных, отражающих последние тенденции в области угроз.

VII. Основной вклад в кибербезопасность

- **Независимость от предварительного знания функций**: способность модели извлекать информацию из строк URL в виде последовательностей символов без необходимости предварительного знания функций упрощает процесс обнаружения и делает его более адаптируемым к новым и неизвестным фишинговым атакам.
- **Высокая способность к обобщению**: использование в модели функций на основе символов URL для надёжного обобщения и точности проверки в сочетании с интеграцией многоуровневых функций в нейронной сети повышает её эффективность при обобщении различных фишинговых угроз.

- **Независимость от экспертов по кибербезопасности и сторонних сервисов**: благодаря автономному извлечению необходимых функций, URL модель снижает зависимость от экспертов по кибербезопасности и сторонних сервисов, что делает её самодостаточным инструментом для обнаружения фишинга.
- **Повышенная точность обнаружения**: экспериментальная проверка модели на контрольных наборах данных продемонстрировала исключительную производительность с заметной точностью 96,04%, что выше, чем в существующих исследованиях.
- **Вклад в симметрию в информационной безопасности**: исследование дополняет дискурс о симметрии и асимметрии в информационной безопасности, предоставляя модель, которая может симметрично изучать и обнаруживать фишинговые URL-адреса, тем самым повышая безопасность сети от возникающих киберугроз.

VIII. Предполагаемые направления будущих исследований

- **Улучшение способности к обобщению**: модель обладает сильной способностью к обобщению, используя функции на основе символов URL для надёжного обобщения и точности проверки. Будущие исследования могли бы быть сосредоточены на дальнейшем повышении этой способности, особенно в контексте развития тактики и методов фишинга.
- **Повышение независимости от экспертов по кибербезопасности и сторонних сервисов**: модель автономно извлекает необходимые функции URL, устраняя зависимость от экспертов по кибербезопасности и сторонних сервисов. В будущих исследованиях можно было бы изучить способы дальнейшего повышения этой независимости, возможно, за счёт разработки более сложных методов выделения признаков.
- **Оптимизация модели многоуровневого обобщения**: используется двухфазная модель многоуровневого обобщения, при этом на первом этапе генерируется прогноз среднего значения, а на втором этапе используется двухуровневая обобщённая модель стека на основе LSTM, оптимизированная для наилучшего прогнозирования при обнаружении фишинговых сайтов. Будущие исследования могли бы быть сосредоточены на оптимизации этой модели, возможно, с помощью различных алгоритмов или методов машинного обучения.
- **Повышение точности**: хотя модель продемонстрировала высокую точность обнаружения фишинговых сайтов, будущие исследования могут быть сосредоточены на способах дальнейшего повышения этой точности, особенно в контексте атак нулевого дня и других передовых методов фишинга.
- **Распространение модели на другие приложения кибербезопасности**: модель потенциально может быть адаптирована для других приложений кибербезопасности, помимо обнаружения фишинга.