СПб ГУТ)))

Разработка системы мониторинга браузера для предотвращения фишинговых атак

Студент: ПЕТРОВ Павел Евгеньевич

Группа: ИКБ-83

Научный руководитель:
К.т.н., доцент каф. ЗСС, ИЗРАИЛОВ Константин Евгеньевич

Актуальность темы исследования и постановка научно-технической задачи

Цель: Повышение эффективности применения инструментария для обнаружения фишинга и предупреждения пользователя

Объект: Действия пользователя в интернете

Предмет: Предотвращение потери персональных/корпоративных данных

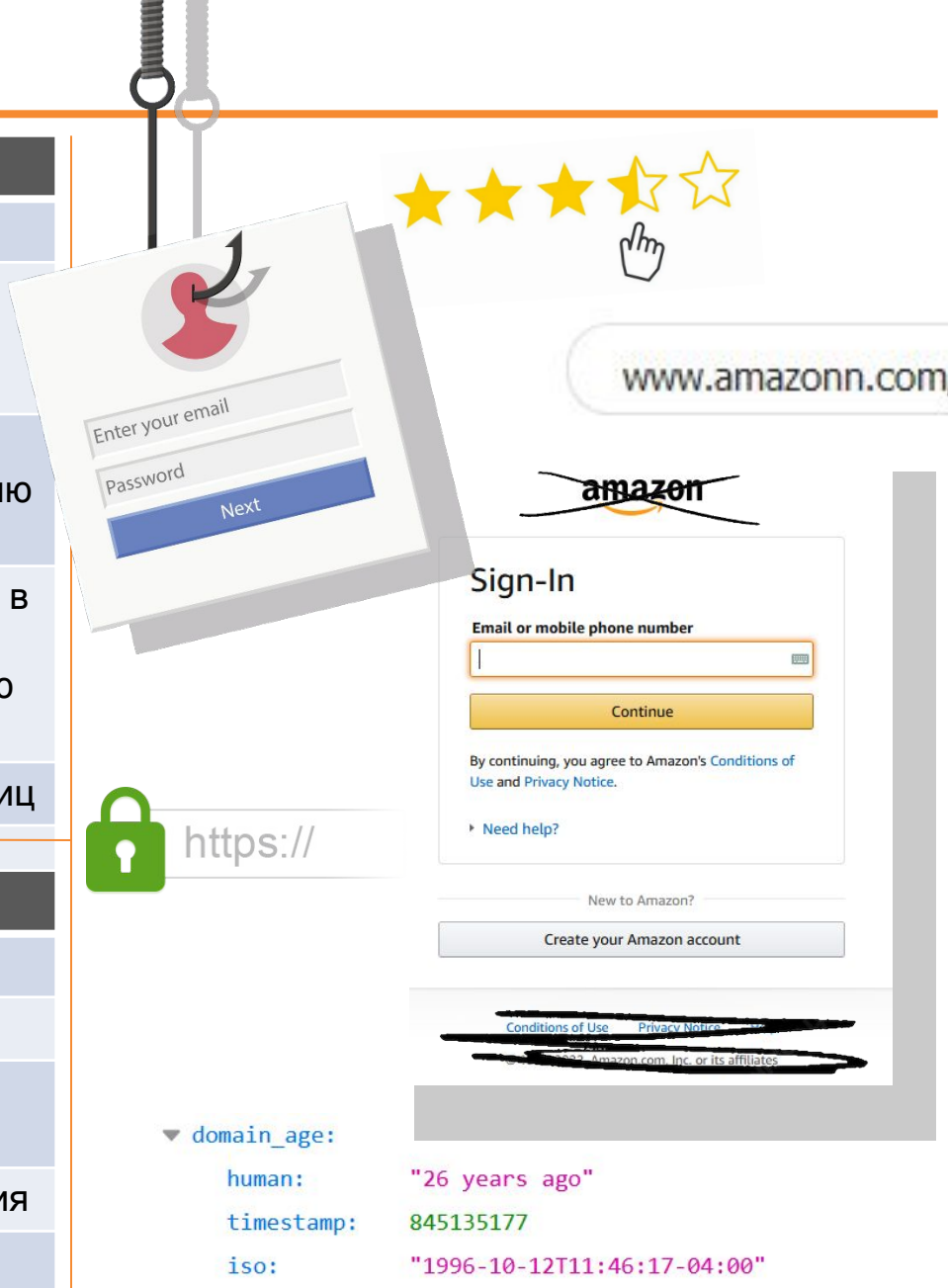
Задачи:

- Анализ фишинговых сайтов
- Исследование существующих методов выявления фишинговых страниц
- Создание собственного метода
- Разработка ПО для выявления потенциальных угроз
- Проведение экспериментов на прототипе

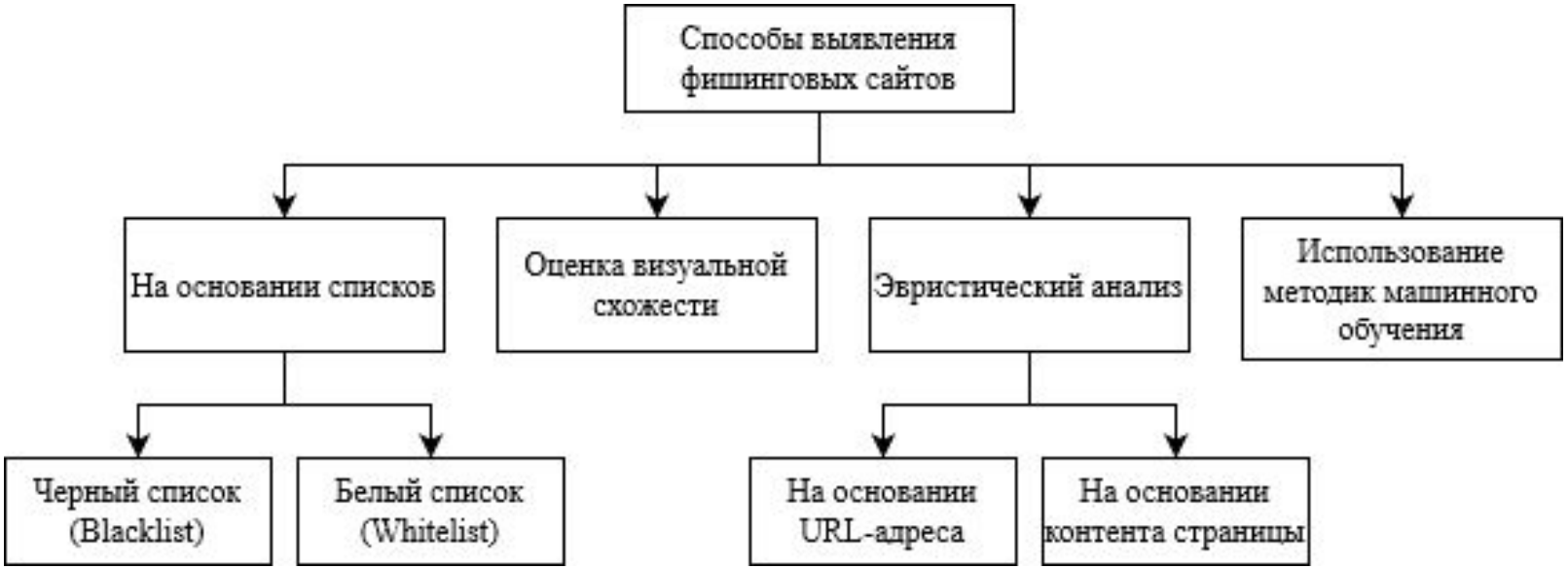
Анализ фишинговых сайтов

Признаки фишинговых сайтов
Отсутствие SSL сертификата
Некорректное написание URL-адреса, пытающегося походить на легитимный через малозаметное изменение букв и символов
Отсутствие стандартных компонентов популярных страниц, колонтитулов, полей меню и заголовков
Наличие веб-формы для заполнения, которая, в большинстве случаев, просит ввести персональные данные или конфиденциальную информацию
Рейтинг сайта в интернете среди других страниц
Дата, время и адрес регистрации домена

Требования к разрабатываемой системе
Модульный принцип построения
Централизованное управление системой
Совместное применение автоматического и ручного контроля
Использование принципов машинного обучения
Журналирование



Синтез метода выявления фишинговых сайтов



Название базы данных	Количество URL-адресов	Постоянное обновление базы	Наличие бесплатного API	Бесплатный доступ к базе адресов
OpenPhish	14,7 млн	+	-	-
PhishTank	7,2 млн	+	+	+
PhishStats	4,1 млн	+	+	-
Common Crawl	Ок. 700 тыс.	-	-	+
Phishing.Database	Ок. 630 тыс.	-	-	+

В разрабатываемом прототипе применяются два варианта списков:

- Корпоративные списки;
- Списки на основе базы данных адресов.

Синтез метода выявления фишинговых сайтов

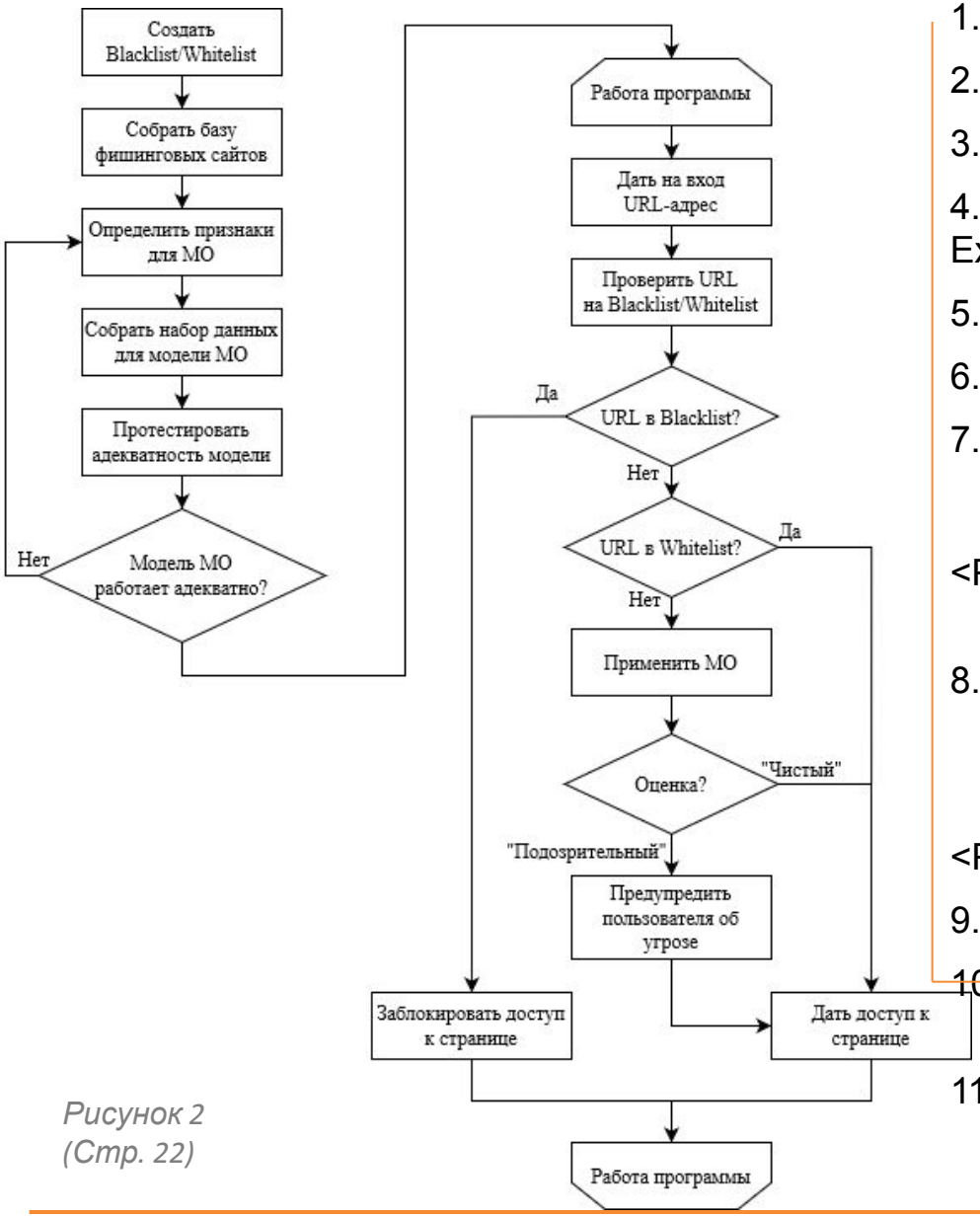


Рисунок 2
(Стр. 22)

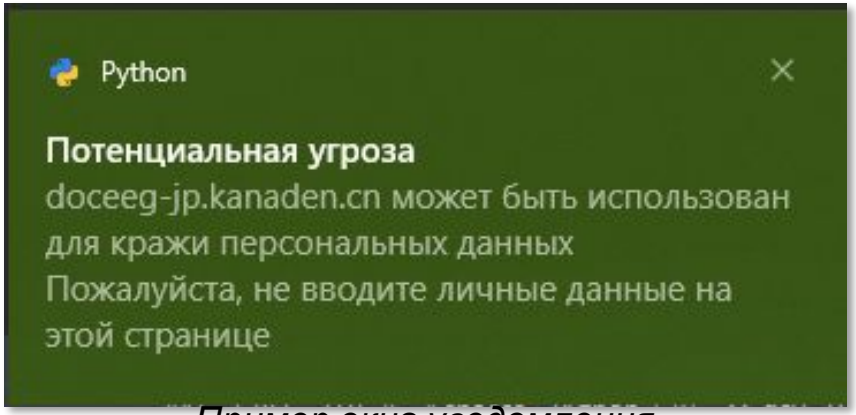
1. <Blacklist, Whitelist> = Analysis<Company_Politics>
2. <DB> = Comparison<OpenDB_List>
3. <Features> = SelectFeatures<Expert, Data>
4. <DataSet> = Create_DataSet<Features, Expert_Decision>
5. <Model> = Train<Classificator, DataSet>
6. <Adequate> = IsModelGood<Model, TestData>
7. If <URL> in <Blacklist>: Goto <Шаг 11> => <Block>
   Else If <URL> in <Whitelist>: Goto <Шаг 9> => <Pass>
   Else: Goto <Шаг 8>
8. <Page_Features> = AnalyzePage<Page>
   <Result> = CheckPage<Model, Page_Features>
   If <Result> == «Good»: Goto <Шаг 9> => <Pass>
   Else If <Result> == «Bad»: Goto <Шаг 10> => <Pass> + <Notify>
9. <Pass>
10. <Notify> = ShowNotification<URL, Text>
    <Pass>
11. <Block>

Работа программы завершается при закрытии окна браузера

Разработка архитектуры средства выявления фишинговых сайтов

Модульный слой архитектуры

- Модуль инициализации программы;
- Модуль оценки Blacklist/Whitelist;
- Модуль сбора информации о странице;
- Модуль оценки страницы на основе МО;
- Модуль уведомлений;



Пример окна уведомления

Информационный слой архитектуры

- Модуль работы с журналом

Входные данные	Blacklist/Whitelist	Набор данных для обучения модели классификатора МО	База данных фишинговых сайтов	Список ключевых слов для определения параметров, отправляемых на оценку классификатору
Промежуточные данные	Набор данных, который собирается в процессе работы программы и отправляется заранее обученной модели для итоговой оценки проверяемой страницы			
Выходные данные	Результат оценки страницы моделью	Окно ввода пользовательского решения (экспертный режим работы)	Уведомление о потенциальной угрозе	

Разработка архитектуры средства выявления фишинговых сайтов

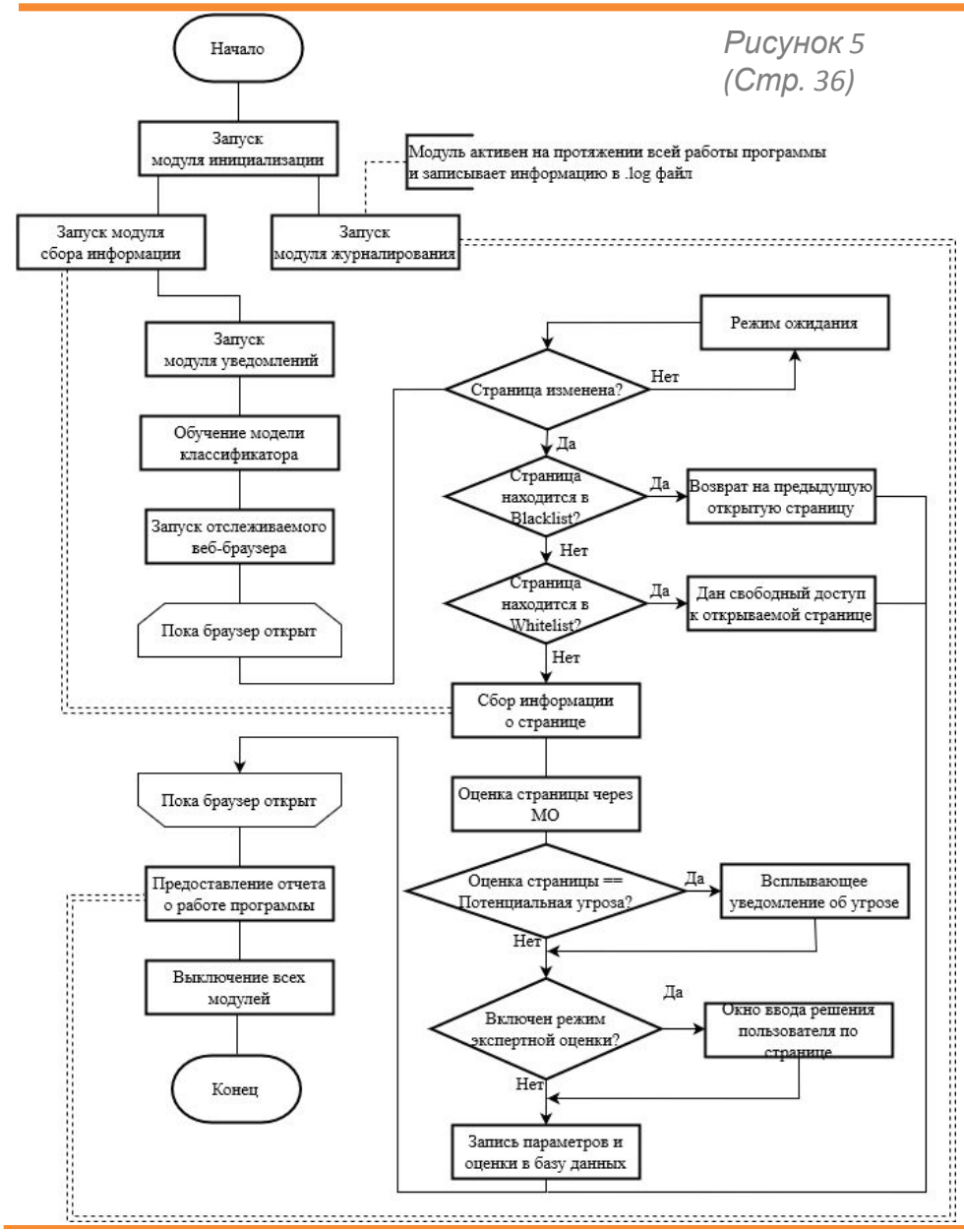


Рисунок 5 (Стр. 36)

Пунктирными линиями на изображении показан процесс взаимодействия модулей для передачи данных

Общая схема подразделяется на два связанных между собой этапа.

Первый этап относится к работе прототипа до и после непосредственной работы пользователя в окне браузера, можно считать его внешним этапом.

Второй – основной – этап работы программы относится к алгоритмам работы во время использования отслеживаемого браузера. В данном этапе происходит мониторинг открытых страниц, их оценка в Blacklist/Whitelist, работа алгоритмов машинного обучения и уведомление пользователя о потенциальных угрозах.

Разработка архитектуры средства выявления фишинговых сайтов

Выб… при… дл…	…ия	Описание
Phishing_	1}	«-1» – на странице нет формы для заполнения; «0» – есть форма заполнения, но сайт НЕ в списке фишинговых адре… «1» – …
Rank	1}	«-1» – … «0» – … «1» – …
Suspiciou	1}	«-1» – … «0» – … «1» – … верх…
Passwor	}	«0» – … «1» – …
Login	}	
Iframe	}	«0» – … «1» – …
IQS_Phis	}	«Fals… «True…
IQS_Susp	}	«Fals… «True» = 1 – Страница в списке подозрительных адресов.
IQS_Risk	1}	«-1» – Значение ранга от 0 до 33; «0» – Значение ранга от 34 до 66; «1» – Значение ранга от 67 до 100.
Result	{-1, 1}	«-1» – сайт чистый; «1» – сайт потенциально опасен.

```
0,0,-1,0,0,0,0,0,-1,-1
0,0,1,0,0,0,0,0,-1,-1
0,0,0,0,0,1,0,0,-1,-1
0,0,0,1,1,1,0,0,-1,-1
0,0,0,1,1,0,0,1,0,-1
0,-1,1,0,0,0,0,1,0,1
0,-1,1,0,0,0,0,1,1,1
0,-1,1,1,1,0,1,1,1,1
0,-1,0,0,0,0,1,1,1,1
0,-1,1,1,1,1,1,1,1,1
0,-1,1,0,0,1,0,1,1,1
0,-1,1,0,0,0,1,1,1,1
0,-1,1,0,0,1,1,1,1,1
0,0,0,0,1,1,0,0,-1,-1
0,0,-1,0,0,1,0,0,-1,-1
0,0,1,0,1,0,0,0,-1,-1
0,0,-1,1,1,1,0,0,-1,-1
0,-1,0,1,0,0,1,1,1,1
0,-1,1,1,0,0,1,1,1,1
0,1,-1,0,0,0,0,0,-1,-1
0,0,-1,0,1,0,0,0,-1,-1
0,-1,1,1,1,0,0,0,-1,-1
0,1,1,0,0,0,0,0,-1,-1
0,-1,1,0,0,0,0,0,0,-1
0,-1,1,1,1,0,0,0,0,-1
0,-1,1,1,1,0,0,1,1,-1
0,0,0,0,0,0,0,0,-1,-1
0,-1,0,0,1,1,0,1,0,-1
```

```
C:\Users\petru\AppData\Local\Programs\Python\

              precision    recall  f1-score

          -1       0.80      1.00      0.89
           1       1.00      0.50      0.67

    accuracy                           0.83
   macro avg       0.90      0.75      0.78
weighted avg       0.87      0.83      0.81


Accuracy Score: 83.0 %
```

Термин «IQS» означает, что параметры берутся через API с онлайн-ресурса для проверки сайтов «IPQualityScore.com»

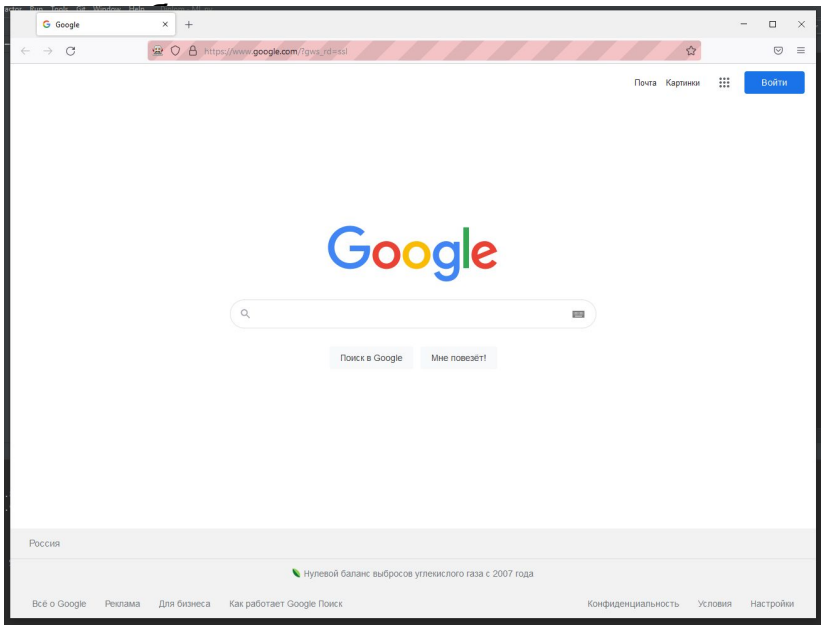
Разработка архитектуры средства выявления фишинговых сайтов

Базовое тестирование прототипа

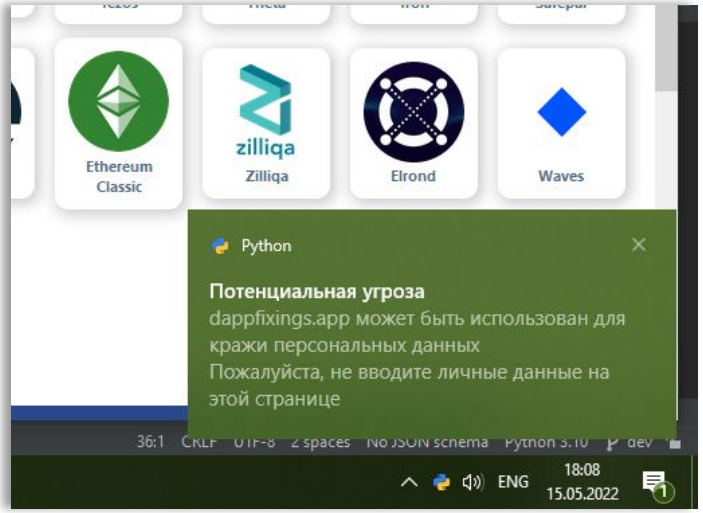
1. Запуск программы и модулей

```
C:\Users\petru\AppData\Local\Programs\Python\Python310\python.exe "E:/PYTHON SMTH/Diplom/scanner.py"
ML instance is ready
Notifier online
Scanner online
```

2. Запуск окна веб



3. Работа модуля уведомлений



4. Работа модуля журнали

Messages	Count
INFO	15
WARNING	5
CRITICAL	0

Visited pages:
www.google.com
www.jetbrains.com
rutube.ru
www.researchgate.net
phishtank.org
dappfixings.app

По окончании работы мы также получаем вывод информации в log-файл

Эксперимент по обнаружению фишинговых сайтов

Пошаговая схема эксперимента:

1. Собрать пул адресов (50% - фишинговые страницы; 50% - чистые страницы; в общем случае от 10 до 20 адресов) и листы доступа;
2. Запустить программу;
3. Дождаться запуска отслеживаемого окна веб-браузера;
4. Зайти на URL-адрес, находящийся в Blacklist;
5. Зайти на URL-адрес, находящийся в Whitelist;
6. Открыть ранее собранный пул и поочередно заходить на каждый из адресов;
7. Проверить получаемую программой оценку по открытой странице;
8. Проверить работу уведомлений (при срабатывании);
9. Закрыть веб-браузер, проверить правильность выключения прототипа.

Исходные данные эксперимента

Blacklist: «YouTube.com», «Wikipedia.org»;

Whitelist: «Ok.ru», «Vk.com».

Пул URL-адресов для эксперимента:

URL-адрес	Экспертная оценка сайта
Google.com	Чистый
www.sut.ru	Чистый
https://www.twitch.tv/	Чистый
github.com	Чистый
web.telegram.org	Чистый
http://mail.ru/	Чистый
avito.ru	Чистый
http://www.airsoftclub.ru	Чистый
www.aeocsen-aesonm.tkfixuh.ne.pw	Фишинговый
http://metamasf.cc	Фишинговый
https://bitpandac.net/	Фишинговый
sec-tool.net	Фишинговый
http://solanart.ltd	Фишинговый
www.paxfulport.com	Фишинговый
www.pockchain.net/	Фишинговый
www2.aenoeuon.icu	Фишинговый

Эксперимент по обнаружению фишинговых сайтов

Ход эксперимента

Шаг 6. Открыть ранее собранный пул и поочередно заходить на каждый из адресов

...аммой оценку по откр...

URL-адрес	Экспертная оценка сайта	Программная оценка сайта
Google.com	Чистый	Чистый
www.sut.ru	Чистый	Чистый
https://www.twitch.tv/	Чистый	Чистый
www.cyberforum.ru	Чистый	Фишинговый
web.telegram.org	Чистый	Чистый
http://mail.ru/	Чистый	Чистый
avito.ru	Чистый	Чистый
http://www.airsoftclub.ru	Чистый	Фишинговый
www.aeocsen-aesonm.tkfixuh.ne.pw	Фишинговый	Фишинговый
http://metamasf.cc	Фишинговый	Фишинговый
https://bitpandac.net/	Фишинговый	Чистый
sec-tool.net	Фишинговый	Фишинговый
http://solanart.ltd	Фишинговый	Фишинговый
www.paxfulport.com	Фишинговый	Фишинговый
www.pockchain.net/	Фишинговый	Фишинговый
www2.aenoeuon.icu	Фишинговый	Фишинговый

Матрица ошибок

TP = 7 **FP = 1**

FN = 2 **TN = 6**

		Экспертная оценка	
		Фишинговый	Чистый
Оценка модели	Фишинговый	7	1
	Чистый	2	6

Точность:

$$PR = \frac{TP}{TP + FP} = \frac{7}{7 + 1} = 0.875 => \mathbf{87.5}\ \%$$

Полнота:

$$RC = \frac{TP}{TP + FN} = \frac{7}{7 + 2} = 0.7(7) => \mathbf{77.8}\ \%$$

F-мера:

$$F = 2\frac{PR * RC}{PR + RC} = 2\frac{0.875 * 0.777}{0.875 + 0.777} = 0.823 => \mathbf{82.3}\ \%$$

Итоговая оценка качества:

$$AC = \frac{P}{N} = \frac{13}{16} = 0.8125 => \mathbf{81.25}\ \%$$

Эксперимент по обнаружению фишинговых сайтов

Сравнительный анализ прототипа с существующими решениями

Критерии эффективности (формульное представление):

Эфф = <Рез, Опер, РесЭк>, где «Эфф» – Эффективность; «Рез» – Результативность; «Опер» – Оперативность; «РесЭк» – Ресурсоэкономность.

Рез = <Рез_Кач, Рез_Колич>, где «Рез_Кач» - Качественная результативность (Критерии); «Рез_Колич» - Количественная результативность (% Точности).

Берем «Рез_Колич» в усредненном значении, далее рассчитаем «Рез_Кач»:

Рез_Кач = <Кр_1, ... Кр_8>, где «Кр_1», ... «Кр_8» - Ранее определенные критерии.

Эфф_Прототип > Эфф_Аналог(1..4)

1. Использование базы фишинговых сайтов;
2. Применение модели машинного обучения для принятия решения по проверяемой странице;
3. Гибкая настройка компонентов программы;
4. Возможность использования в корпоративных сетях;
5. Возможность применения дополнительной экспертной оценки со стороны пользователя;

Решение	Кр_1	Кр_2	Кр_3	Кр_4	Кр_5	Кр_6	Кр_7	Кр_8	Точность обнаружения угрозы
Фильтр Firefox	+	–	–	–	–	–	–	+	~ 73%
Фильтр IE	+	–	–	–	–	–	–	+	~ 70 %
FireEye EX	+	+	+	+	–	–	–	–	~ 87 %
Анти-фишинг в KIS	+	–	+	–	–	–	–	+	~ 91%
Прототип	+	+	+	+	+	+	+	+	~ 81.3 %

СПб ГУТ)))

Благодарю за внимание!

Вопросы?