РОССИЙСКАЯ ФЕДЕРАЦИЯ

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(19) **RU** (11) **2 580 027**(13) **C1**

(51) МПК
G06F 21/55 (2013.01)
G06F 21/60 (2013.01)

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2014141809/08, 17.10.2014**

(24) Дата начала отсчета срока действия патента:
17.10.2014

Приоритет(ы):
(22) Дата подачи заявки: **17.10.2014**

(45) Опубликовано: **10.04.2016** Бюл. № **10**

(56) Список документов, цитированных в отчете о поиске: **US 2012/0158626 A1, 21.06.2012. US 2013/0031628 A1, 31.01.2013. WO 2007/070838 A2, 21.06.2007. US 2010/0281536 A1, 04.11.2010. RU 2446459 C1, 27.03.2012.**

Адрес для переписки:
125212, Москва, Ленинградское ш., 39а, стр. 3, АО Лаборатория Касперского, Управление по интеллектуальной собственности, Надежде Васильевне Кащенко

(72) Автор(ы):
Кошелев Максим Глебович (RU)

(73) Патентообладатель(и):
Закрытое акционерное общество "Лаборатория Касперского" (RU)

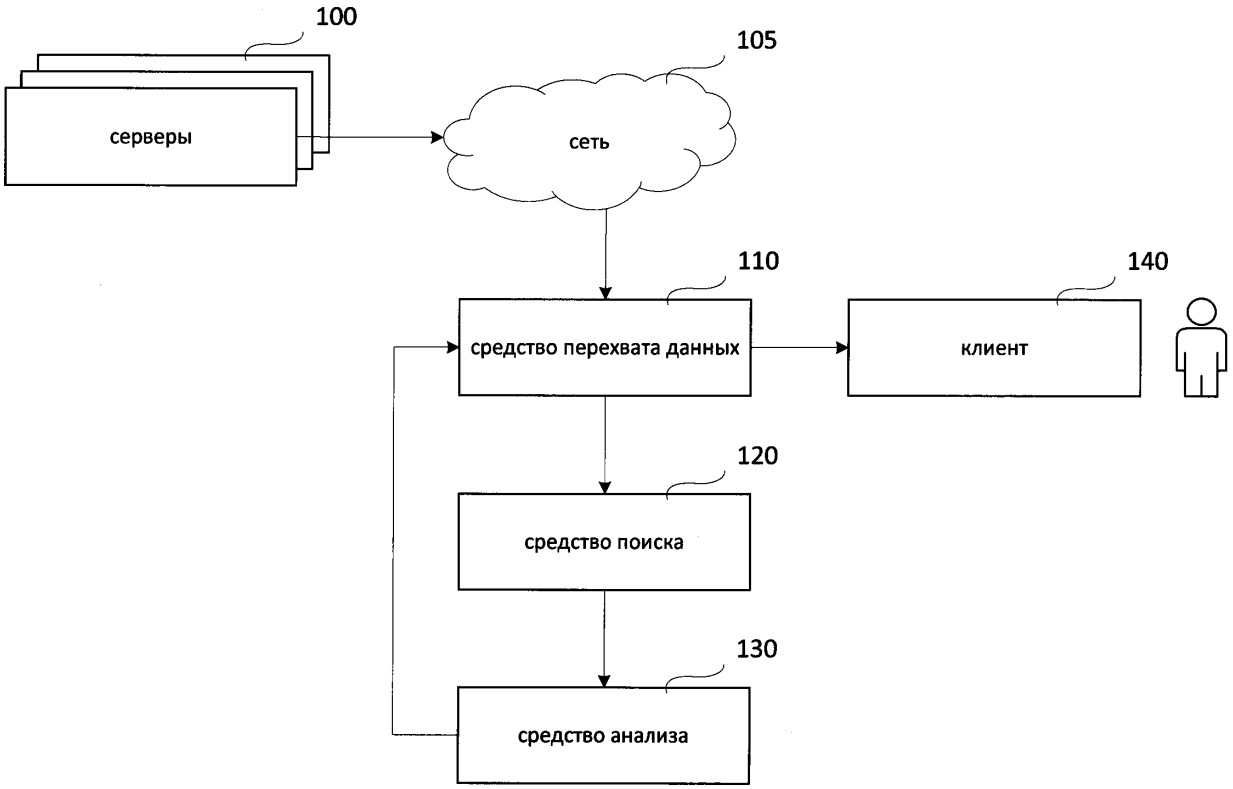
(54) СИСТЕМА И СПОСОБ ФОРМИРОВАНИЯ ПРАВИЛ ПОИСКА ДАННЫХ, ИСПОЛЬЗУЕМЫХ ДЛЯ ФИШИНГА

(57) Реферат:

Изобретение относится к информационной безопасности. Технический результат заключается в повышении надежности обнаружения фишинга в полученных пользователем данных. Система формирования правил поиска данных, используемых для фишинга, содержит средство перехвата; средство категоризации для определения категории перехваченных данных, где категориями выступают: текст, гиперссылка, мультимедийные данные, сценарий, приложение flash, java апплет; и передачи данных, разделенных на категории, средству анализа данных; средство анализа данных для поиска признака фишинга по данным, разделенным на категории; вычисления параметров найденных признаков фишинга, где параметрами признаков фишинга выступают: весовые коэффициенты для признака фишинга и категории, флаги обнаружения признака фишинга и категории, количество признаков фишинга по категориям; и передачи параметров найденных признаков фишинга средству формирования правил; средство формирования правил для настройки логических связей между найденными признаками фишинга на основании вычисленных параметров признаков фишинга; формирования на основании настроенных логических связей между найденными признаками фишинга правила поиска данных, используемых для фишинга. 2 н.п. ф-лы, 3 ил.

RU 2580027 C1

RU 2580027 C1



Фиг. 1

RU 2580027 C1

RU 2580027 C1

RUSSIAN FEDERATION

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(19) **RU** (11) **2 580 027**(13) **C1**

(51) Int. Cl.
G06F 21/55 (2013.01)
G06F 21/60 (2013.01)

(12) ABSTRACT OF INVENTION

(21)(22) Application: **2014141809/08, 17.10.2014**

(24) Effective date for property rights:
17.10.2014

Priority:
(22) Date of filing: **17.10.2014**

(45) Date of publication: **10.04.2016** Bull. № **10**

Mail address:
125212, Moskva, Leningradskoe sh., 39a, str. 3, AO Laboratorija Kasperskogo, Upravlenie po intellektualnoj sobstvennosti, Nadezhde Vasilevne Kashchenko

(72) Inventor(s):
Koshelev Maksim Glebovich (RU)

(73) Proprietor(s):
Zakrytoe aktsionernoe obshchestvo "Laboratorija Kasperskogo" (RU)

(54) SYSTEM AND METHOD OF GENERATING RULES FOR SEARCHING DATA USED FOR PHISHING

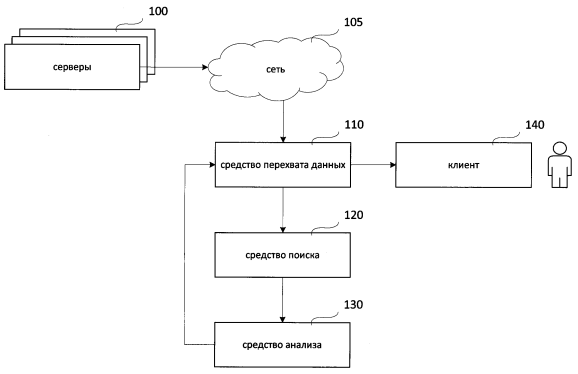
(57) Abstract:

FIELD: information technology.

SUBSTANCE: invention relates to information security. System for generation of rules on searching data used for phishing has interception device; categorization device for determining category of intercepted data, where categories are: text, hyperlink, multimedia data, script, flash application, java applet; and transmission of data divided into category, data analyser; data analyser to search sign of data phishing divided into categories; calculation of parameters of the found phishing signs, where parameters of phishing signs feature: weight coefficients for phishing signs and categories, flags of detection phishing signs and categories, number of phishing signs by categories; and transmission of parameters of found phishing signs to rules generating system; rules generating system for the adjustment of logical connections between found phishing signs based on calculated parameters of phishing signs; formation of rules for searching data used for phishing based on set logical connections between found features phishing signs.

EFFECT: technical result consists in improving reliability of phishing detection in data obtained by the user.

2 cl, 3 dwg



Фиг. 1

RU 2580027 C1

RU 2580027 C1

Область техники

Изобретение относится к системам и способам поиска данных, используемых для фишинга.

Уровень техники

Бурное развитие интернет-технологий в последнее десятилетие, появление большого числа устройств, передающих данные через интернет (таких как персональные компьютеры, ноутбуки, планшеты, мобильные телефоны и т.д.), а также простота и удобство их эксплуатации привели к тому, что огромное число людей в своих повседневных делах стало пользоваться интернетом, будь то получение информации, работа с банковскими счетами, осуществление покупок, чтение почты, посещение социальных сетей, развлечения и т.д. Часто при работе в интернете (например, при покупке товаров, переводе денег, регистрации и т.д.) пользователям приходится передавать на внешние сервера свою конфиденциальную информацию (такую как номера кредитных карт и банковских счетов, пароли к учетным записям и т.д.), ту самую информацию, от надежности которой зависит финансовая безопасность пользователей.

Огромное число пользователей, использующих интернет, привело к увеличению активности мошенников, получающих с помощью разнообразных техник и методов доступ к конфиденциальным данным пользователей с целью их кражи для дальнейшего использования в собственных целях. Одним из самых популярных методов является так называемый фишинг (англ. phishing), т.е. получение доступа к конфиденциальной информации пользователя с помощью проведения рассылок писем от имени популярных брендов, личных сообщений внутри различных сервисов (например, внутри социальных сетей), а также создания и регистрации в поисковых сервисах сайтов, выдающих себя за легальные сайты банков, интернет-магазинов, социальных сетей и т.д. В письме или сообщении, посылаемом мошенниками пользователям, часто содержатся ссылки на вредоносные сайты, внешне неотличимые от настоящих, или на сайты, с которых будет осуществлен переход на вредоносные сайты. После того, как пользователь попадает на поддельную страницу, мошенники, используя различные приемы социальной инженерии, пытаются побудить пользователя ввести свою конфиденциальную информацию, которую он использует для доступа к определенному сайту, что позволяет мошенникам получить доступ к аккаунтам и банковским счетам. Кроме однократного раскрытия своей конфиденциальной информации, пользователь рискует получить с такого сайта-подделки одно из вредоносных приложений, осуществляющих регулярный сбор и передачу мошенникам информации с компьютера жертвы.

Для борьбы с описанным выше способом мошенничества применяются технологии, направленные на выявление фишинговых сообщений (например, в электронной почте), а также поддельных сайтов. Для этого используются базы доверенных и недоверенных адресов сайтов, шаблоны фраз из фишинговых сообщений и т.д. При обнаружении факта присутствия такого подозрительного объекта, пользователь информируется о потенциальной опасности.

Например, в публикации US 20130086677 описана технология определения фишинга по анализу гиперссылок, который включает в себя как поиск по базам недоверенных гиперссылок, так и определение схожести исследуемых гиперссылок с недоверенными. Недостаток данного метода заключается в том, что зачастую гиперссылки в html-коде не указаны в явном виде, адрес, на который указывает гиперссылка, определяется только в момент осуществления перехода. Кроме того, наличие недоверенной гиперссылки само по себе не является фактом наличия фишинга в исследованных

данных, а определение схожести без учета других данных (например, лексического анализа текста, в котором встречаются гиперссылки) ведет к высокому уровню ложного обнаружения.

В другой публикации US 20090006532 описана технология динамической защиты от фишинга на основе обработки данных, собранных с нескольких пользовательских устройств. И хотя такая технология способна повысить уровень обнаружения фишинга, она обладает существенным недостатком по времени реагирования на возникшую угрозу.

Еще в одной публикации US 8776196 описана технология автоматического обнаружения фишинга на основе установленных методов. Недостаток данной технологии заключается в том, что она не справляется с новыми методами фишинга или с методами, построенными так, чтобы бороться со стандартными способами обнаружения фишинга.

Хотя описанные выше методы работы эффективны при столкновении с уже известными угрозами, ссылками и страницами, которые были изучены специалистами по информационной безопасности и признаны потенциально опасными, они гораздо хуже справляются с новыми угрозами. Кроме того, из-за ручного или полуавтоматического анализа время реакции между появлением нового фишингового сайта или сообщения и его обнаружением и последующей блокировкой на компьютере пользователя может быть довольно значительным, что является достаточным фактором для того, чтобы большое количество пользователей успели стать жертвами мошенников.

Настоящее изобретение позволяет более эффективно решить задачу обнаружения фишинга на веб-ресурсах и электронной почте.

Раскрытие изобретения

Изобретение предназначено для формирования правил поиска данных, используемых для фишинга.

Технический результат настоящего изобретения заключается в повышении надежности обнаружения фишинга в полученных пользователем данных за счет выполнения поиска данных, используемых для фишинга в полученных пользователем данных на основе сформированных правил поиска.

Технический результат настоящего изобретения достигается путем использования системы формирования правил поиска данных, используемых для фишинга, которая содержит: средство перехвата, предназначенное для перехвата данных, передаваемых от сервера клиенту и передачи данных средству категоризации, средство категоризации, предназначенное для определения, по меньшей мере, одной категории перехваченных данных, и передачи данных, разделенных на категории, средству анализа данных, средство анализа данных, предназначенное для поиска по данным, разделенным на категории, признаков фишинга, вычисления параметров признаков фишинга и передачи их средству формирования правил, средство формирования правил, предназначенное для формирования по вычисленным параметрам признаков фишинга, по меньшей мере, одного правила поиска данных, используемых для фишинга, по меньшей мере, в одном из следующих случаев: по меньшей мере, один из параметров признаков фишинга превышает заданный порог обнаружения фишинга, совокупная величина, по меньшей мере, двух параметров признаков фишинга превышает заданный порог обнаружения фишинга.

В другом частном случае реализации системы в качестве сервера может выступать, по меньшей мере: сайт, почтовый сервер, служба мгновенного обмена сообщениями, служба операционной системы.

Еще в одном частном случае реализации системы в качестве клиента может выступать, по меньшей мере: браузер, почтовый клиент, клиент службы мгновенного обмена сообщениями, средство просмотра электронных документов.

В другом частном случае реализации системы категории данных могут быть определены, по меньшей мере, как: текст, гиперссылки, мультимедийные данные, сценарии, приложения flash, java апплеты.

Еще в одном частном случае реализации системы средство анализа вычисляет параметры признаков фишинга, по меньшей мере, одним из следующих способов: через выставление весовых коэффициентов для найденных признаков, через выставление весовых коэффициентов для выделенных категорий, через выставление флагов для найденных признаков, через выставление флагов для найденных категорий.

В другом частном случае реализации системы правила поиска данных, используемых для фишинга, формируют, по меньшей мере, одним из следующих способов: через сравнение суммарного вычисленного весового коэффициента с заданным в качестве критерия для данных, используемых для фишинга, через сравнение максимального вычисленного весового коэффициента с заданным в качестве критерия для данных, используемых для фишинга, через сравнение количества категорированных данных, обладающих весовым коэффициентом выше определенного с заданным в качестве критерия для данных, используемых для фишинга.

Технический результат настоящего изобретения достигается путем использования способа формирования правил поиска данных, используемых для фишинга, по которому: перехватывают данные, передаваемые от сервера клиенту, категорируют перехваченные данные, выполняют поиск признаков фишинга по категорированным данным, вычисляют параметры найденных признаков фишинга, определяют необходимость формирования правил поиска данных, используемых для фишинга, по меньшей мере, в одном из следующих случаев: по меньшей мере, один из параметров признаков фишинга превышает заданный порог обнаружения фишинга, совокупная величина, по меньшей мере, двух параметров признаков фишинга превышает заданный порог обнаружения фишинга, после чего формируют по вычисленным параметрам признаков фишинга, по меньшей мере, одно правило поиска данных, используемых для фишинга.

В другом частном случае реализации способа в качестве сервера может выступать, по меньшей мере: сайт, почтовый сервер, служба мгновенного обмена сообщениями, служба операционной системы.

Еще в одном частном случае реализации способа в качестве клиента может выступать, по меньшей мере: браузер, почтовый клиент, клиент службы мгновенного обмена сообщениями, средство просмотра электронных документов.

В другом частном случае реализации способа категории данных могут быть определены, по меньшей мере, как: текст, гиперссылки, мультимедийные данные, сценарии, приложения flash, java апплеты.

Еще в одном частном случае реализации способа параметры признаков фишинга вычисляют, по меньшей мере, одним из следующих способов: через выставление весовых коэффициентов для найденных признаков, через выставление весовых коэффициентов для выделенных категорий, через выставление флагов для найденных признаков, через выставление флагов для найденных категорий.

В другом частном случае реализации способа правила данных, используемых для фишинга, формируют, по меньшей мере, одним из следующих способов: через сравнение суммарного вычисленного весового коэффициента с заданным в качестве критерия для данных, используемых для фишинга, через сравнение максимального вычисленного

весового коэффициента с заданным в качестве критерия для данных, используемых для фишинга, через сравнение количества категорированных данных, обладающих весовым коэффициентом выше определенного с заданным в качестве критерия для данных, используемых для фишинга.

Еще в одном частном случае реализации способа совокупная величина параметров признаков фишинга, вычисляется, по меньшей мере, одним из следующих способов: через вычисление суммы параметров признаков фишинга, через вычисление суммы признаков фишинга, через вычисление меры центральной тенденции параметров признаков фишинга, через вычисление показателя рассеивания параметров признаков фишинга относительно их математического ожидания.

Краткое описание чертежей

Фиг. 1 представляет структурную схему системы формирования правил поиска данных, используемых для фишинга в данных, передаваемых от сервера клиенту.

Фиг. 2 представляет структурную схему способа формирования правил поиска данных, используемых для фишинга в данных, передаваемых от сервера клиенту.

Фиг. 3 представляет пример компьютерной системы общего назначения, персональный компьютер или сервер.

Хотя изобретение может иметь различные модификации и альтернативные формы, характерные признаки, показанные в качестве примера на чертежах, будут описаны подробно. Следует понимать, однако, что цель описания заключается не в ограничении изобретения конкретным его воплощением. Наоборот, целью описания является охват всех изменений, модификаций, входящих в рамки данного изобретения, как это определено приложенной формулой.

Описание вариантов осуществления изобретения

Объекты и признаки настоящего изобретения, способы для достижения этих объектов и признаков станут очевидными посредством отсылки к примерным вариантам осуществления. Однако настоящее изобретение не ограничивается примерными вариантами осуществления, раскрытыми ниже, оно может воплощаться в различных видах. Сущность, приведенная в описании, является ничем иным, как конкретными деталями, необходимыми для помощи специалисту в области техники в исчерпывающем понимании изобретения, и настоящее изобретение определяется в объеме приложенной формулы.

Фиг. 1 представляет структурную схему системы формирования правил поиска данных, используемых для фишинга в данных, передаваемых от сервера клиенту.

Система формирования правил поиска данных, передаваемых от сервера клиенту на наличие данных, используемых для фишинга, состоит из серверов 100, сети 105, средства перехвата данных 110, средства поиска 120, средства анализа 130 и клиента 140.

Серверы 100 (представленные, по меньшей мере, одним сервером) предназначены для обмена через сеть 105 данными с клиентом 140 и предоставления ему доступа к определенным ресурсам или услугам. В качестве серверов могут выступать:

- сайты (например, сайт банка или интернет-магазина);
- файловые серверы;
- службы мгновенного обмена сообщениями (например, Skype);
- службы операционной системы (например, службы загрузки файлов).

Сеть 105 предназначена для осуществления обмена данными между серверами 100 и клиентом 140. В качестве сети могут выступать:

- локальная сеть;

- глобальная сеть Интернет.

Средство перехвата данных 110 предназначено для перехвата данных, передаваемых от сервера 100 клиенту 140, категоризации перехваченных данных, передачи категорированных данных средству поиска 120 и последующего получения параметров данных, используемых для фишинга, от средства анализа 130, модификации перехваченных данных с учетом полученных параметров, и передачи модифицированных данных клиенту 140. В качестве перехватываемых данных могут выступать:

- html-код сайтов и т.д.;
- flash-приложения и java-апплеты, загружаемые с сайтов и т.д.;
- мультимедийные данные (например, изображения, содержащиеся на сайтах или скриншоты работы клиента);
- электронные документы (документы Microsoft Office, PDF и т.д.);
- сообщения электронной почты.

В качестве средства перехвата данных могут выступать:

- proxy-сервер (например, proxy-сервер для перехвата http трафика);
- драйвер, установленный на компьютер пользователя (например, дисковый драйвер для перехвата обращений к файлам).

В качестве категорий, по которым категорируются перехваченные данные, могут выступать:

- гиперссылки;
- мультимедийные данные;
- сценарии (например, Javascript или VBA);
- тексты;
- java апплеты;
- flash приложения.

Например, один из компонентов антивирусного продукта, отвечающий за противодействия нежелательной рекламе, работая в качестве proxy-сервера, производит категоризацию html-кода перехваченной страницы, выделяя текст, содержащийся на странице, ссылки на изображения и flash-приложения, присутствующие на странице, гиперссылки и т.д.

В качестве методов модификации перехваченных данных могут выступать:

- замена гиперссылок, ведущих на сайты, осуществляющие мошенническую деятельность, на предупреждения;
- деактивации на страницах элементов управления, инициирующих отправку персональных данных пользователя;
- добавление информации, содержащей предупреждения;
- частичное или полное удаление перехваченной информации.

Например, один из компонентов антивирусного продукта, отвечающий за противодействие фишингу, при обнаружении на странице, посещаемой пользователем, недоверенной гиперссылки, может заменить ее html-код текстовой вставкой таким образом, чтобы в браузере, которому будут переданы соответствующие данные, вместо недоверенной гиперссылки было отображено предупреждение о блокировке потенциальной угрозы (к примеру, код CitiBankнакод[Ссылка заблокирована]).

Средство поиска 120 предназначено для поиска признаков фишинга в категорированных данных, вычисления параметров для найденных признаков фишинга и передачи вычисленных параметров, а также категорированных данных средству анализа 130. В качестве методов поиска признаков фишинга в зависимости от признаков

могут выступать:

- определение размера категорированного теста;
- поиск по тексту лексем из словаря, содержащего лексемы, составляющие сообщения о фишинге, и подсчет их количества;
- получение параметров сайтов (IP-адрес, владелец и т.д.) на которые указывают категорированные гиперссылки;
- сравнение по похожести категорированных мультимедийных данных с недоверенными или доверенными мультимедийными данными.

В качестве признаков фишинга могут выступать следующие признаки:

- размер текста меньше заданного порога (например, менее 20 слов) на странице сайта или в письме;
- недоверенная ссылка;
- неизвестная ссылка на изображения с известного сайта организации;
- кол-во лексем в тексте больше заданного порога (например, больше 20% слов и фраз текста удовлетворяют лексемам из словаря, где словарь содержит лексемы, составляющие сообщения о фишинге);
- изображение доверенного логотипа совместно с недоверенной или неизвестной ссылкой;
- размер текста больше заданного порога (например, более 1000 слов) при отсутствии изображений и ссылок;
- несколько больших изображений без текста;
- домен, не принадлежащий организации;
- ссылка на изображение не указывает на домен организации;
- комбинации указанных выше признаков.

В качестве вычисленных параметров признаков фишинга могут выступать:

- весовые коэффициенты для признаков фишинга;
- весовые коэффициенты для категорий;
- флаги обнаружения признаков фишинга;
- флаги обнаружения категорий;
- количество признаков фишинга по категориям.

В качестве примера вычисления параметров фишинга может служить выставление одним из компонентов антивирусного продукта, отвечающего за противодействие фишингу, весового коэффициента для неизвестной ссылки, содержащейся на исследуемом сайте (например, от 0.0, соответствующего доверенной ссылке до 1.0, соответствующего недоверенной ссылке). Для определения весового коэффициента проверяется наличие ссылки в списке доверенных ссылок (если ссылка найдена, выставляется весовой коэффициент равный 0.0, и расчеты прекращаются), наличие ссылки в списке недоверенных ссылок (если ссылка найдена, выставляется весовой коэффициент равный 1.0 и расчеты прекращаются), соответствие одной из лексем из словаря лексем недоверенных ссылок (если лексема найдена, искомый весовой коэффициент увеличивается на 0.25), наличие владельца сайта в списке недоверенных владельцев (если владелец найден, искомый весовой коэффициент увеличивается на 0.15) и т.д.

Средство анализа данных 130 предназначено для вынесения решения о создании правил поиска данных, используемых для фишинга, создания правил определения наличия фишинга в перехваченных данных на основе вычисленных параметров, вычисления с помощью созданных правил параметров данных, используемых для фишинга в категорированных данных и передачи вычисленных параметров средству перехвата данных 110. В качестве параметров данных, используемых для фишинга,

могут выступать:

- списки недоверенных гиперссылок;
- перечисление мультимедийных данных, содержащих фишинг;
- позицию и размер текстового блока, содержащего информацию, определенную как фишинг;
- участки html-кода, ответственные за передачу данных пользователя серверу.

В качестве способов вынесения решения о возможности создания правил поиска могут выступать:

- по меньшей мере, один из параметров признаков фишинга превышает заданный порог обнаружения фишинга;
- совокупная величина, по меньшей мере, двух параметров признаков фишинга превышает заданный порог обнаружения фишинга.

Совокупная величина параметров признаков фишинга может вычисляться одним из следующих способов:

- через вычисление суммы параметров признаков фишинга;
- через вычисление суммы признаков фишинга;
- через вычисление меры центральной тенденции параметров признаков фишинга;
- через вычисление показателя рассеивания параметров признаков фишинга относительно их математического ожидания.

Клиент 140 предназначен для получения данных от сервера 100, обработки полученных данных и отображения пользователю результата этой обработки. В качестве клиента 140 могут выступать:

- браузер;
- почтовый клиент;
- клиенты службы мгновенного обмена сообщениями (например, Skype, QIP);
- средства просмотра электронных документов.

Примером анализа данных, передаваемых от сервера клиенту на наличие данных, используемых для фишинга, может служить следующая ситуация: пользователь, используя браузер, с помощью поискового сервиса нашел сайт продаж интересующего его товара. После перехода по предоставленному адресу он попадает на страницу оформления заказа, где среди прочего требуется указать реквизиты своей кредитной карты. Сайт является фишинговым и создан мошенниками для воровства данных по кредитным картам.

После того как пользователь с помощью браузера 140 зашел на фишинговый сайт 100, информация в виде html-кода была передана от последнего браузеру 140. На компьютере пользователя данные с сайта 100 были перехвачены средством перехвата данных 110.

Средство перехвата данных 110 перехватывает представленные в виде html-кода данные, предназначенные для браузера 140, категорирует перехваченный html-код, для чего вначале производит эмуляцию html-кода, которая заключается в:

- исполнении сценариев, содержащихся в html-коде (например, сценариев на языке JavaScript, расшифровывающих данные, содержащихся в html коде, таких как гиперссылки, тексты и т.д.);
- загрузке мультимедийных данных, вызываемых из html-кода (например, изображения логотипов банка, под который маскируется сайт);
- загрузке flash-приложений, java-апплетов и т.п., вызываемых из html-кода;
- эмуляции действий пользователя (например, ввод текста в элементы ввода и нажатие кнопок с целью определения последующих действий);

- отрисовке страницы в буфер памяти или файл для последующего использования вместе с другими мультимедийными данными для поиска признаков фишинга.

После завершения эмуляции средство перехвата данных 110 непосредственно производит саму категоризацию проэмулированных данных и передает полученные категорированные данные (такие, как текст, гиперссылки, изображения и т.д.) средству поиска 120. Затем, после получения от средства анализа 130 параметров данных, используемых для фишинга (например, обнаруженные недоверенные гиперссылки), средство перехвата данных 110 производит модификацию перехваченного html-кода (например, заменяя обнаруженные недоверенные гиперссылки текстовыми предупреждениями о возможном фишинге), после чего передает модифицированный html-код браузеру 140.

Средство поиска 120 производит поиск по категорированным данным признаков фишинга (например, наличие гиперссылок недоверенных сайтов, фраз в текстах, часто используемых на фишинговых сайтах, несоответствие названия гиперссылок и адресов, на который они ведут, присутствие среди изображений логотипов известных компаний и т.д.), вычисление по найденным признакам фишинга их параметров (например, для текста это может быть весовой коэффициент, означающий вероятность принадлежности к фишингу, для гиперссылок - флаг принадлежности к недоверенным ссылкам и т.п.), после чего осуществляет передачу вычисленных параметров и категорированных данных средству анализа 130.

Средство анализа 130 на основе полученных параметров признаков фишинга создает правила определения наличия фишинга (например, правило по которому присутствие недоверенных ссылок и весового коэффициента текста, в котором встречаются недоверенные ссылки, превышающего 0.75 (где весовой коэффициент 0.0 означает доверенную ссылку, а 1.0 - недоверенную ссылку), считается фактом обнаружения фишинга), после чего применяет правила к категорированным данным и вычисляет по ним параметры данных, используемых для фишинга (например, какие гиперссылки являются недоверенными и какой текст указывает на фишинг), после чего передает вычисленные параметры средству перехвата данных 110.

Браузер 140, получив модифицированный html-код, обрабатывает его и отображает в своем окне. В результате пользователь, даже попав на сайт, использующий фишинг, ограничивается в возможностях по его использованию и тем самым защищается от потери своих конфиденциальных данных, таких как реквизиты банковской карты.

Другим примером анализа данных, передаваемых от сервера клиенту на наличие данных, используемых для фишинга, может служить ситуация, в которой пользователь, используя почтовый клиент, читает электронные письма. Одно из полученных писем отправлено мошенниками с целью заманить пользователя на специально созданный сайт.

После того как пользователь с помощью почтового клиента 140 открыл письмо, полученное с почтового сервера 100, информация в виде pdf-документа была передана последним почтовому клиенту 140. На компьютере пользователя данные с почтового сервера 100 были перехвачены средством перехвата данных 110.

Средство перехвата данных 110 перехватывает данные, предназначенные для почтового клиента 140, представленные в виде pdf-документа, производит их категоризацию и передает полученные категорированные данные (такие, как текст, гиперссылки, изображения и т.д.) средству поиска 120. Затем, после получения от средства анализа 130 параметров данных, используемых для фишинга (например, текст и изображения), средство перехвата данных 110 производит модификацию

перехваченного pdf-документа (например, дополняя pdf-документ текстовыми предупреждениями о возможном фишинге), после чего передает модифицированный pdf-документ почтовому клиенту 140.

Средство поиска 120 производит поиск по категорированным данным признаков фишинга, вычисление по найденным признакам фишинга их параметров, после чего осуществляет передачу вычисленных параметров и категорированных данных средству анализа 130.

Средство анализа 130 на основе полученных параметров признаков фишинга создает правила определения наличия фишинга (например, правило, по которому присутствие изображений из списка недоверенных совместно с текстом, весовой коэффициент которого превышает 0.9 (где весовой коэффициент 0.0 означает доверенное изображение, а весовой коэффициент 1.0 - недоверенное), считается фактом обнаружения фишинга), после чего применяет правила к категорированным данным и вычисляет по ним параметры данных, используемых для фишинга (например, какой блок данных в pdf-документе указывает на фишинг), после чего передает вычисленные параметры средству перехвата данных 110.

Почтовый клиент 140, получив модифицированный pdf-документ, обрабатывает его и отображает в своем окне. В результате пользователь при просмотре полученного письма видит предупреждение о возможном фишинге в его содержимом.

Фиг. 2 представляет структурную схему способа формирования правил поиска данных, используемых для фишинга в данных, передаваемых от сервера клиенту.

Способ формирования правил поиска данных состоит из следующих этапов: перехвата данных 210, категоризации данных 220, поиска признаков 230, вычисления параметров 240, создания правил 250, использования правил 260 и модификации данных 270.

На этапе 210 происходит перехват данных, передаваемых серверами 100 клиенту 140. В качестве перехватываемых данных могут выступать:

- html-код сайтов и т.д.;
- flash-приложения и java-апплеты, загружаемые с сайтов и т.д.;
- мультимедийные данные (например, изображения, содержащиеся на сайтах или скриншоты работы клиента);
- электронные документы (документы Microsoft Office, PDF и т.д.);
- сообщения электронной почты.

На этапе 220 перехваченные данные категорируются по критериям. В качестве категорий, по которым категорируются перехваченные данные, могут выступать:

- гиперссылки;
- мультимедийные данные;
- сценарии (например, Javascript или VBA);
- тексты;
- java апплеты;
- flash приложения.

На этапе 230 по категорированным данным производится поиск признаков фишинга. В качестве признаков фишинга могут выступать следующие признаки:

- размер текста меньше заданного порога (например, менее 20 слов) на странице сайта или в письме;
- недоверенная ссылка;
- неизвестная ссылка на изображения с известного сайта организации;
- кол-во лексем в тексте больше заданного (например, больше 20% слов и фраз текста удовлетворяют лексемам из словаря, где словарь содержит лексемы, составляющие

сообщения о фишинге);

- изображение доверенного логотипа совместно с недоверенной или неизвестной ссылкой;

- размер текста больше заданного порога (например, более 1000 слов) при отсутствии изображений и ссылок;

- несколько больших изображений без текста;

- домен, не принадлежащий организации;

- ссылка на изображение не указывает на домен организации;

- комбинации указанных выше признаков.

На этапе 240 вычисляются параметры найденных признаков фишинга. После чего на основе вычисленных параметров признаков фишинга выносится решение о возможности создания правил поиска данных, используемых для фишинга. В качестве вычисленных параметров признаков фишинга могут выступать:

- весовые коэффициенты для признаков фишинга;

- весовые коэффициенты для категорий;

- флаги обнаружения признаков фишинга;

- флаги обнаружения категорий;

- количество признаков фишинга по категориям.

В качестве способов вынесения решения о возможности создания правил поиска могут выступать:

- по меньшей мере, один из параметров признаков фишинга превышает заданный порог обнаружения фишинга;

- совокупная величина, по меньшей мере, двух параметров признаков фишинга превышает заданный порог обнаружения фишинга.

Совокупная величина параметров признаков фишинга может вычисляться одним из следующих способов:

- через вычисление суммы параметров признаков фишинга;

- через вычисление суммы признаков фишинга;

- через вычисление меры центральной тенденции параметров признаков фишинга;

- через вычисление показателя рассеивания параметров признаков фишинга относительно их математического ожидания.

На этапе 250 на основе вычисленных параметров признаков фишинга создают правила поиска данных, используемых для фишинга. В случае когда правила создать не удалось, предпринимается попытка вернуться к этапу 220, чтобы повторно категорировать перехваченные данные и в конечном итоге сформировать новые правила (например, если в перехваченных данных присутствовали изображения, то изначально они попали в категорию мультимедийных данных, но впоследствии с помощью алгоритмов распознавания были категорированы как текст).

Правила формируются таким образом, чтобы настроить логические связи между найденными признаками фишинга на основании вычисленных параметров (например, изображение логотипа известной компании является фактором, повышающим весовой коэффициент неизвестной гиперссылки до уровня, при котором гиперссылка начинает считаться недоверенной).

Также при формировании правил могут использоваться уже существующие правила, полученные ранее, которые будут усложнять (увеличивать количество логических связей между признаками фишинга) или облегчать (соответственно, уменьшать) формируемое правило. Найденные признаки фишинга, а также вычисленные по ним параметры и построенные впоследствии правила с целью последующего использования

на новых категорированных данных могут сохраняться как локально на компьютере пользователя, так и удаленно в облаке (в этом случае они могут быть загружены и другими пользователями и использованы на других компьютерах).

Одним из способов построения правил является использование нейронной сети, которая может обучаться формированию новых более точных правил на основе параметров, полученных при обработке полученных ранее данных. В такой нейронной сети в качестве нейронов выступают алгоритмы (или, другими словами, процессы) расчета признаков фишинга и вычисления параметров признаков фишинга, обменивающиеся между собой категорированными данными и вычисленными параметрами.. Могут использоваться разные методы обучения, например, "обучение с учителем" (англ. Supervised learning), при котором часть нейронов (т.е. способы расчета признаков фишинга) и связей (т.е. какие признаки фишинга и параметры признаков принимают нейроны) определяет аналитик на основе обработанных ранее данных. В другом случае реализации может использоваться метод "обучение без учителя" (англ. Unsupervised learning), в котором нейроны перестраиваются в зависимости от ранее обработанных данных без участия в процессе анализа аналитика. В качестве результата работы описанная выше нейронная сеть может возвращать или факт использования обработанных данных для фишинга, или вероятность такого факта (от 0 до 1, где 0 означает доверенные данные, а 1 - недоверенные).

На этапе 260 происходит применение сформированных правил к категорированным данным, осуществляется поиск данных, используемых при фишинге, вычисление параметров найденных данных. В качестве параметров найденных данных могут использоваться:

- списки недоверенных гиперссылок;
- перечисление мультимедийных данных, содержащих фишинг;
- позицию и размер текстового блока, содержащего информацию, определенную как фишинг;
- участки html-кода, ответственные за передачу данных пользователя серверу.

На этапе 270 происходит модификация перехваченных данных на основании вычисленных параметров данных, используемых при фишинге, и передача модифицированных данных клиенту 140. В качестве модификаций перехваченных данных могут выступать:

- замена гиперссылок, ведущих на сайты, осуществляющие мошенническую деятельность, на предупреждения;
- деактивации на страницах элементов управления, инициирующих отправку персональных данных пользователя;
- добавление информации, содержащей предупреждения;
- частичное или полное удаление перехваченной информации.

Фиг. 3 представляет пример компьютерной системы общего назначения, персональный компьютер или сервер 20, содержащий центральный процессор 21, системную память 22 и системную шину 23, которая содержит разные системные компоненты, в том числе память, связанную с центральным процессором 21. Системная шина 23 реализована, как любая известная из уровня техники шинная структура, содержащая, в свою очередь, память шины или контроллер памяти шины, периферийную шину и локальную шину, которая способна взаимодействовать с любой другой шинной архитектурой. Системная память содержит постоянное запоминающее устройство (ПЗУ) 24, память с произвольным доступом (ОЗУ) 25. Основная система ввода/вывода (BIOS) 26, содержит основные процедуры, которые обеспечивают передачу информации

между элементами персонального компьютера 20, например, в момент загрузки операционной системы с использованием ПЗУ 24.

Персональный компьютер 20, в свою очередь, содержит жесткий диск 27 для чтения и записи данных, привод магнитных дисков 28 для чтения и записи на сменные магнитные диски 29 и оптический привод 30 для чтения и записи на сменные оптические диски 31, такие как CD-ROM, DVD-ROM и иные оптические носители информации. Жесткий диск 27, привод магнитных дисков 28, оптический привод 30 соединены с системной шиной 23 через интерфейс жесткого диска 32, интерфейс магнитных дисков 33 и интерфейс оптического привода 34 соответственно. Приводы и соответствующие компьютерные носители информации представляют собой энергонезависимые средства хранения компьютерных инструкций, структур данных, программных модулей и прочих данных персонального компьютера 20.

Настоящее описание раскрывает реализацию системы, которая использует жесткий диск 27, сменный магнитный диск 29 и сменный оптический диск 31, но следует понимать, что возможно применение иных типов компьютерных носителей информации 56, которые способны хранить данные в доступной для чтения компьютером форме (твердотельные накопители, флеш-карты памяти, цифровые диски, память с произвольным доступом (ОЗУ) и т.п.), которые подключены к системной шине 23 через контроллер 55.

Компьютер 20 имеет файловую систему 36, где хранится записанная операционная система 35, а также дополнительные программные приложения 37, другие программные модули 38 и данные программ 39. Пользователь имеет возможность вводить команды и информацию в персональный компьютер 20 посредством устройств ввода (клавиатуры 40, манипулятора «мышь» 42). Могут использоваться другие устройства ввода (не отображены): микрофон, джойстик, игровая консоль, сканер и т.п. Подобные устройства ввода по своему обычаю подключают к компьютерной системе 20 через последовательный порт 46, который, в свою очередь, подсоединен к системной шине, но могут быть подключены иным способом, например, при помощи параллельного порта, игрового порта или универсальной последовательной шины (USB). Монитор 47 или иной тип устройства отображения также подсоединен к системной шине 23 через интерфейс, такой как видеоадаптер 48. В дополнение к монитору 47, персональный компьютер может быть оснащен другими периферийными устройствами вывода (не отображены), например колонками, принтером и т.п.

Персональный компьютер 20 способен работать в сетевом окружении, при этом используется сетевое соединение с другим или несколькими удаленными компьютерами 49. Удаленный компьютер (или компьютеры) 49 являются такими же персональными компьютерами или серверами, которые имеют большинство или все упомянутые элементы, отмеченные ранее при описании существа персонального компьютера 20, представленного на Фиг. 3. В вычислительной сети могут присутствовать также и другие устройства, например маршрутизаторы, сетевые станции, пиринговые устройства или иные сетевые узлы.

Сетевые соединения могут образовывать локальную вычислительную сеть (LAN) 50 и глобальную вычислительную сеть (WAN). Такие сети применяются в корпоративных компьютерных сетях, внутренних сетях компаний и, как правило, имеют доступ к сети Интернет. В LAN- или WAN-сетях персональный компьютер 20 подключен к локальной сети 50 через сетевой адаптер или сетевой интерфейс 51. При использовании сетей персональный компьютер 20 может использовать модем 54 или иные средства обеспечения связи с глобальной вычислительной сетью, такой как Интернет. Модем

54, который является внутренним или внешним устройством, подключен к системной шине 23 посредством последовательного порта 46. Следует уточнить, что сетевые соединения являются лишь примерными и не обязаны отображать точную конфигурацию сети, т.е. в действительности существуют иные способы установления соединения техническими средствами связи одного компьютера с другим.

В заключение следует отметить, что приведенные в описании сведения являются примерами, которые не ограничивают объем настоящего изобретения, определенного формулой.

Формула изобретения

1. Система формирования правил поиска данных, используемых для фишинга, которая содержит:

а) средство перехвата, предназначенное для:

• перехвата данных, передаваемых от сервера клиенту, где в качестве перехваченных данных выступает:

- html-код сайта,
- flash-приложение,
- java-апплет,
- мультимедийные данные,
- электронный документ;

• и передачи данных средству категоризации;

б) упомянутое средство категоризации, предназначенное для:

• определения по меньшей мере одной категории перехваченных данных, где в качестве по меньшей мере одной категории выступают следующие категории:

- текст,
- гиперссылка,
- мультимедийные данные,
- сценарий,
- приложение flash,
- java-апплет;

• и передачи данных, разделенных на категории, средству анализа данных;

в) упомянутое средство анализа данных, предназначенное для:

• поиска по данным, разделенным на категории, по меньшей мере одного признака фишинга;

• вычисления параметров найденных признаков фишинга, где в качестве вычисленных параметров признаков фишинга выступают, по меньшей мере:

- весовой коэффициент для признака фишинга,
- весовой коэффициент для категории,
- флаг обнаружения признака фишинга,
- флаг обнаружения категории,
- количество признаков фишинга по категориям;

• определения на основании параметров найденных признаков фишинга необходимости формирования по меньшей мере одного правила поиска данных, используемых для фишинга, по меньшей мере, в случае, когда:

- по меньшей мере один из параметров признаков фишинга превышает заданный порог обнаружения фишинга,
- совокупная величина по меньшей мере двух параметров признаков фишинга превышает заданный порог обнаружения фишинга;

• и передачи параметров найденных признаков фишинга средству формирования правил;

г) упомянутое средство формирования правил, предназначенное для:

• настройки логических связей между найденными признаками фишинга на основании вычисленных параметров признаков фишинга;

• формирования на основании настроенных логических связей между найденными признаками фишинга по меньшей мере одного правила поиска данных, используемых для фишинга.

2. Способ формирования правил поиска данных, используемых для фишинга, по которому:

а) перехватывают данные, передаваемые от сервера клиенту, где в качестве перехваченных данных выступает:

• html-код сайта,

• flash-приложение,

• java-апплет,

• мультимедийные данные,

• электронный документ;

б) определяют по меньшей мере одну категорию перехваченных данных, где в качестве по меньшей мере одной категории выступают следующие категории:

• текст,

• гиперссылка,

• мультимедийные данные,

• сценарий,

• приложение flash,

• java-апплет;

в) выполняют поиск по меньшей мере одного признака фишинга по данным, разделенным на категории;

г) вычисляют параметры найденных признаков фишинга, где в качестве вычисленных параметров признаков фишинга выступают, по меньшей мере:

• весовой коэффициент для признака фишинга,

• весовой коэффициент для категории,

• флаг обнаружения признака фишинга,

• флаг обнаружения категории,

• количество признаков фишинга по категориям;

д) определяют необходимость формирования по меньшей мере одного правила поиска данных, используемых для фишинга, по меньшей мере, в случае, когда:

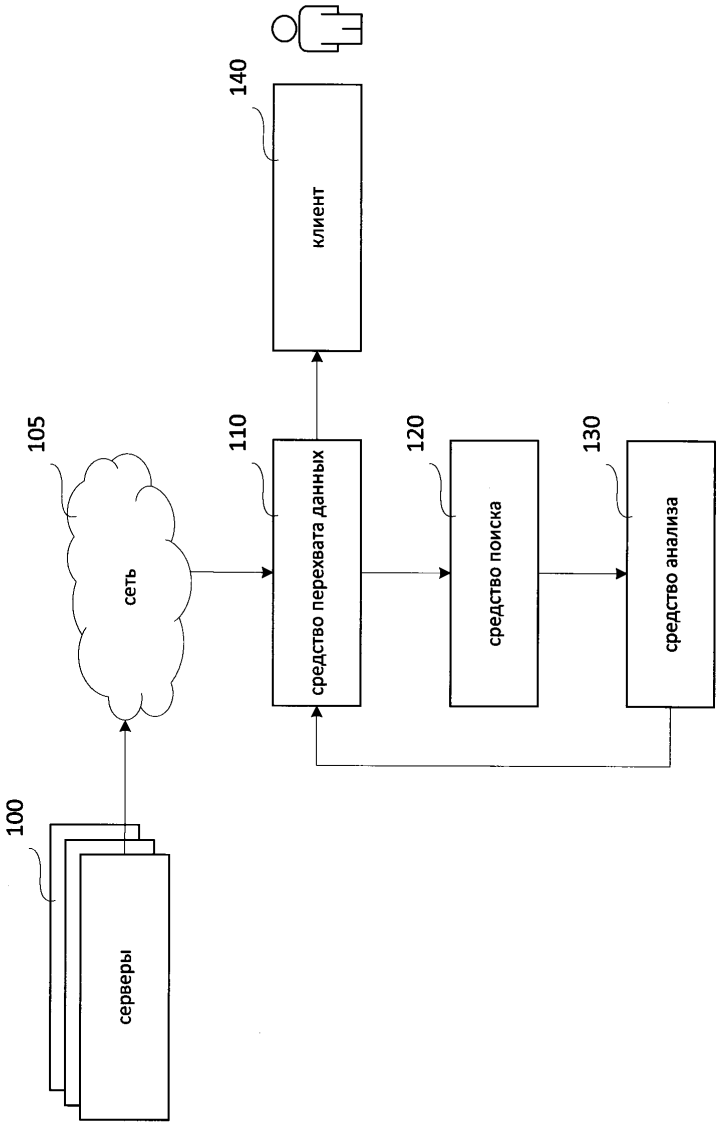
• по меньшей мере один из параметров признаков фишинга превышает заданный порог обнаружения фишинга;

• совокупная величина по меньшей мере двух параметров признаков фишинга превышает заданный порог обнаружения фишинга.

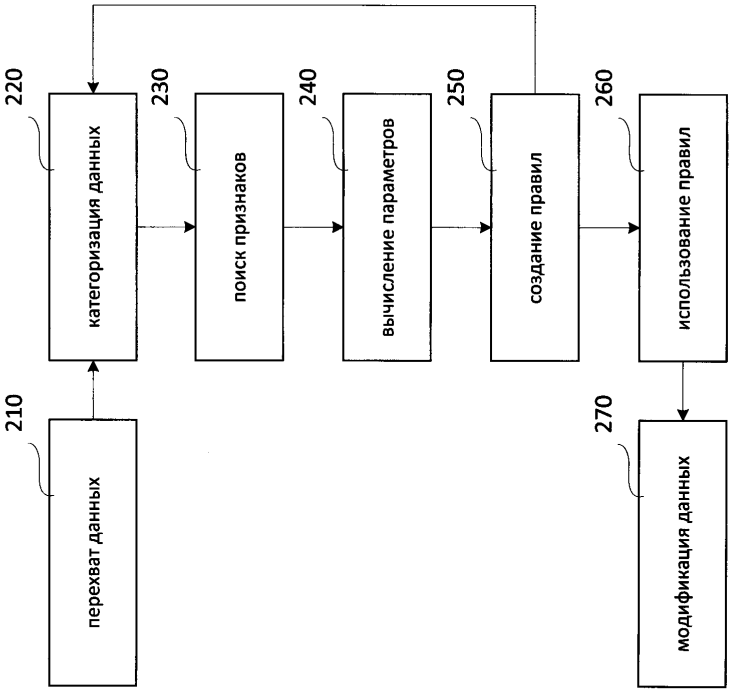
е) настраивают логические связи между найденными признаками фишинга на основании вычисленных параметров признаков фишинга;

ж) выполняют этапы в)-е) до тех пор, пока логические связи между найденными признаками фишинга на основании вычисленных параметров признаков фишинга не будут настроены;

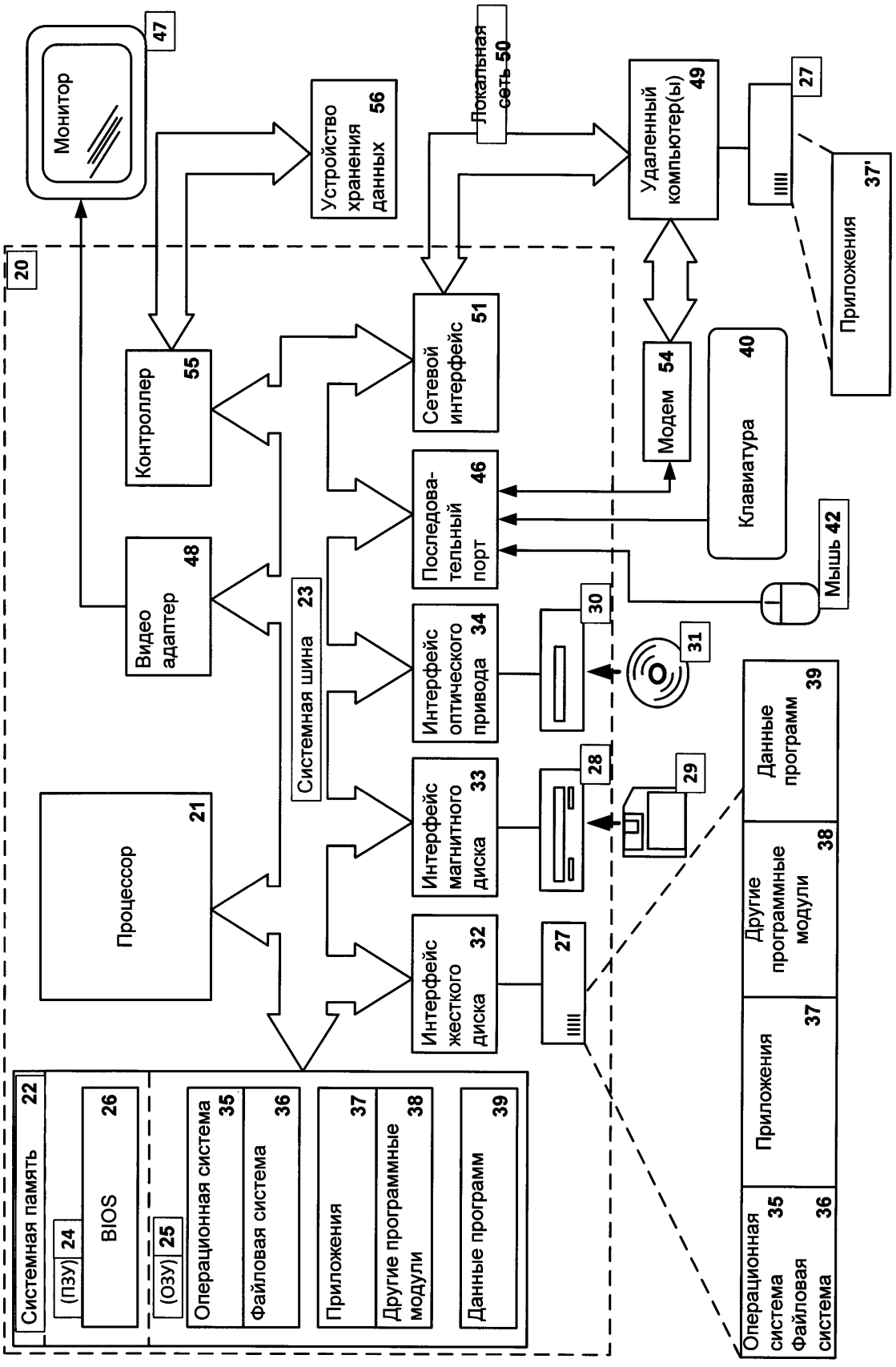
з) формируют на основании настроенных логических связей между найденными признаками фишинга по меньшей мере одно правило поиска данных, используемых для фишинга.


140
клиент
105
110
120
130
сеть
средство перехвата данных
средство поиска
средство анализа
100
серверы

Фиг. 1


210
перехват данных
220
категоризация данных
230
поиск признаков
240
вычисление параметров
250
создание правил
260
использование правил
270
модификация данных

Фиг. 2


Монитор 47
Устройство хранения данных 56
Локальная сеть 50
Удаленный компьютер(ы) 49
27
Приложения 37'
20
Контроллер 55
Сетевой интерфейс 51
Модем 54
Клавиатура 40
Видео адаптер 48
Последовательный порт 46
Мышь 42
Системная шина 23
Интерфейс оптического привода 34
30
31
Процессор 21
Интерфейс магнитного диска 33
28
29
Данные программ 39
Другие программные модули 38
Интерфейс жесткого диска 32
27
Приложения 37
Операционная система 35
Файловая система 36
Системная память 22
(ПЗУ) 24
BIOS 26
(ОЗУ) 25
Операционная система 35
Файловая система 36
Приложения 37
Другие программные модули 38
Данные программ 39

Фиг. 3