РОССИЙСКАЯ ФЕДЕРАЦИЯ

(19) **RU** (11) **2 634 178**(13) **C1**

(51) МПК
G06F 21/56 (2013.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2016139470, 10.10.2016**

(24) Дата начала отсчета срока действия патента:
10.10.2016

Дата регистрации:
24.10.2017

Приоритет(ы):
(22) Дата подачи заявки: **10.10.2016**

(45) Опубликовано: **24.10.2017** Бюл. № **30**

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

(72) Автор(ы):
Крюков Андрей Владимирович (RU),
Лискин Александр Викторович (RU),
Иванов Антон Михайлович (RU)

(73) Патентообладатель(и):
Акционерное общество "Лаборатория Касперского" (RU)

(56) Список документов, цитированных в отчете о поиске: **US 7093239 B1, 15.08.2006. US 8607335 B1, 10.12.2013. WO 2008/054732 A2, 08.05.2008. US 2011/0016530 A1, 20.01.2011. US 2007/0266436 A1, 15.11.2007. RU 2491623 C1, 27.08.2013. DANIEL RENTZ "Microsoft Compound Document File Format", опубл. 31.01.2012 на 25 страницах [найдено 22.06.2017], найдено в Интернет по адресу URL:** (см. прод.)

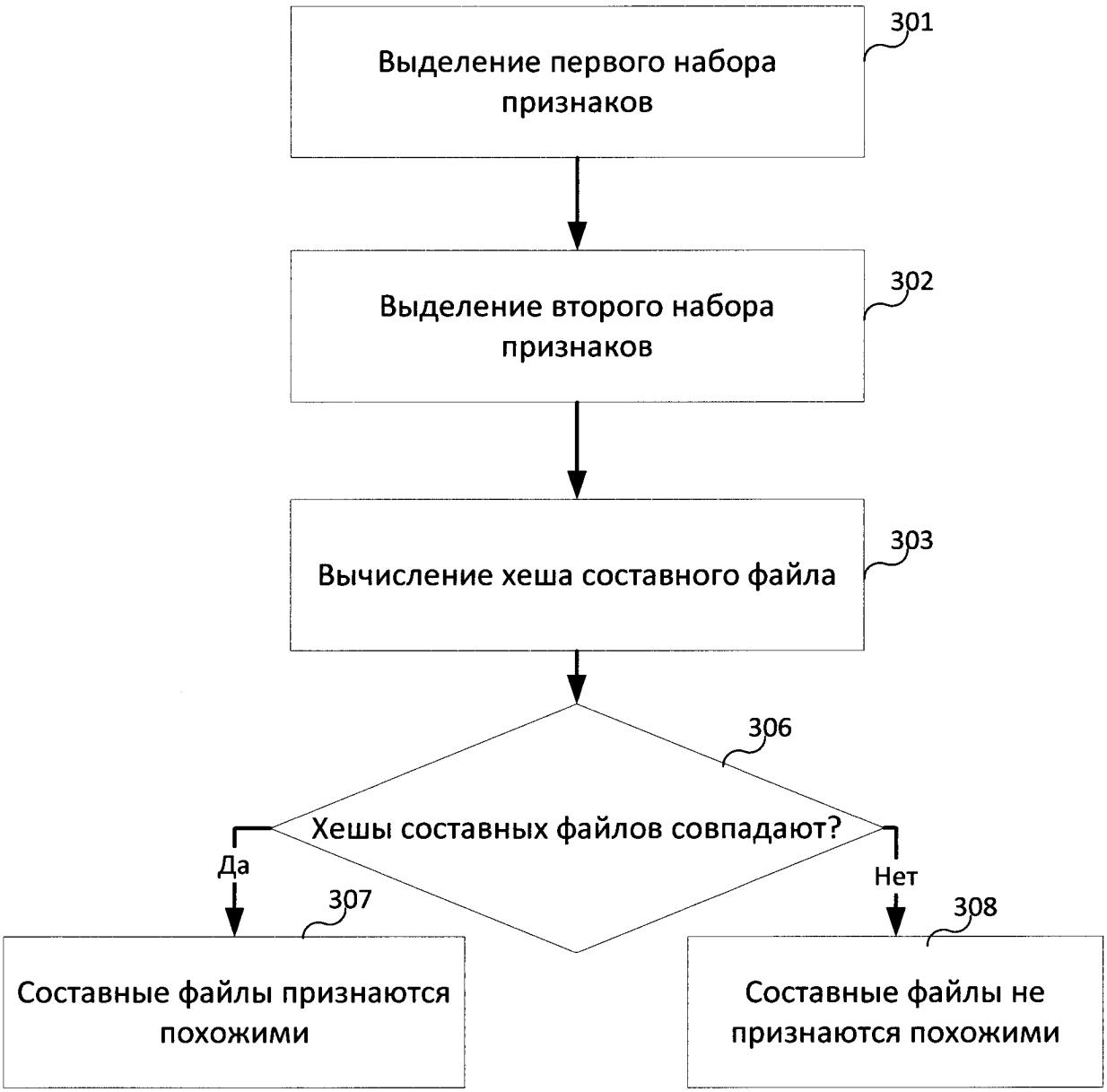
(54) Способ обнаружения вредоносных составных файлов

(57) Реферат:

Изобретение относится к области защиты вычислительных устройств, а именно к способам обнаружения вредоносных составных файлов. Технический результат заключается в обеспечении защиты вычислительного устройства от вредоносных программ за счет обнаружения составного вредоносного файла. Способ обнаружения вредоносных составных файлов, в котором определяют при помощи средства вычисления, является ли файл составным; выделяют при помощи средства вычисления первый набор признаков из заголовка составного файла, если ранее на этапе было определено, что файл является составным; выделяют при помощи средства вычисления второй набор признаков из по меньшей мере одной директории составного файла, если ранее на этапе было определено, что файл является составным; вычисляют при помощи средства вычисления хеш составного файла с использованием по меньшей мере первого и второго набора признаков; признают при помощи средства сравнения составной файл вредоносным, если вычисленный хеш составного файла совпадает с хешем вредоносного составного файла; при этом хеш вредоносного составного файла хранится в базе данных хешей. 4 з.п. ф-лы, 4 ил.

RU 2634178 C1

RU 2634178 C1



Фиг. 3

(56) (продолжение):
https://web.archive.org/web/20120131060104/ http://www.openoffice.org/sc/compdocfileformat.pdf.

RU 2634178 C1

RU 2634178 C1

RUSSIAN FEDERATION

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(19) **RU** (11) **2 634 178**(13) **C1**

(51) Int. Cl.
G06F 21/56 (2013.01)

(12) ABSTRACT OF INVENTION

(21)(22) Application: **2016139470, 10.10.2016**

(24) Effective date for property rights:
10.10.2016

Registration date:
24.10.2017

Priority:
(22) Date of filing: **10.10.2016**

(45) Date of publication: **24.10.2017** Bull. № **30**

Mail address:
125212, Moskva, Leningradskoe sh., 39a, str. 3, AO "Laboratoriya Kasperskogo", Upravlenie po intellektualnoj sobstvennosti, Nadezhda Vasilevna Kashchenko

(72) Inventor(s):
Kryukov Andrej Vladimirovich (RU),
Liskin Aleksandr Viktorovich (RU),
Ivanov Anton Mikhajlovich (RU)

(73) Proprietor(s):
Aktsionernoe obshchestvo "Laboratoriya Kasperskogo" (RU)

RU 2634178 C1

(54) METHOD OF DETECTING HARMFUL COMPOSITE FILES

(57) Abstract:

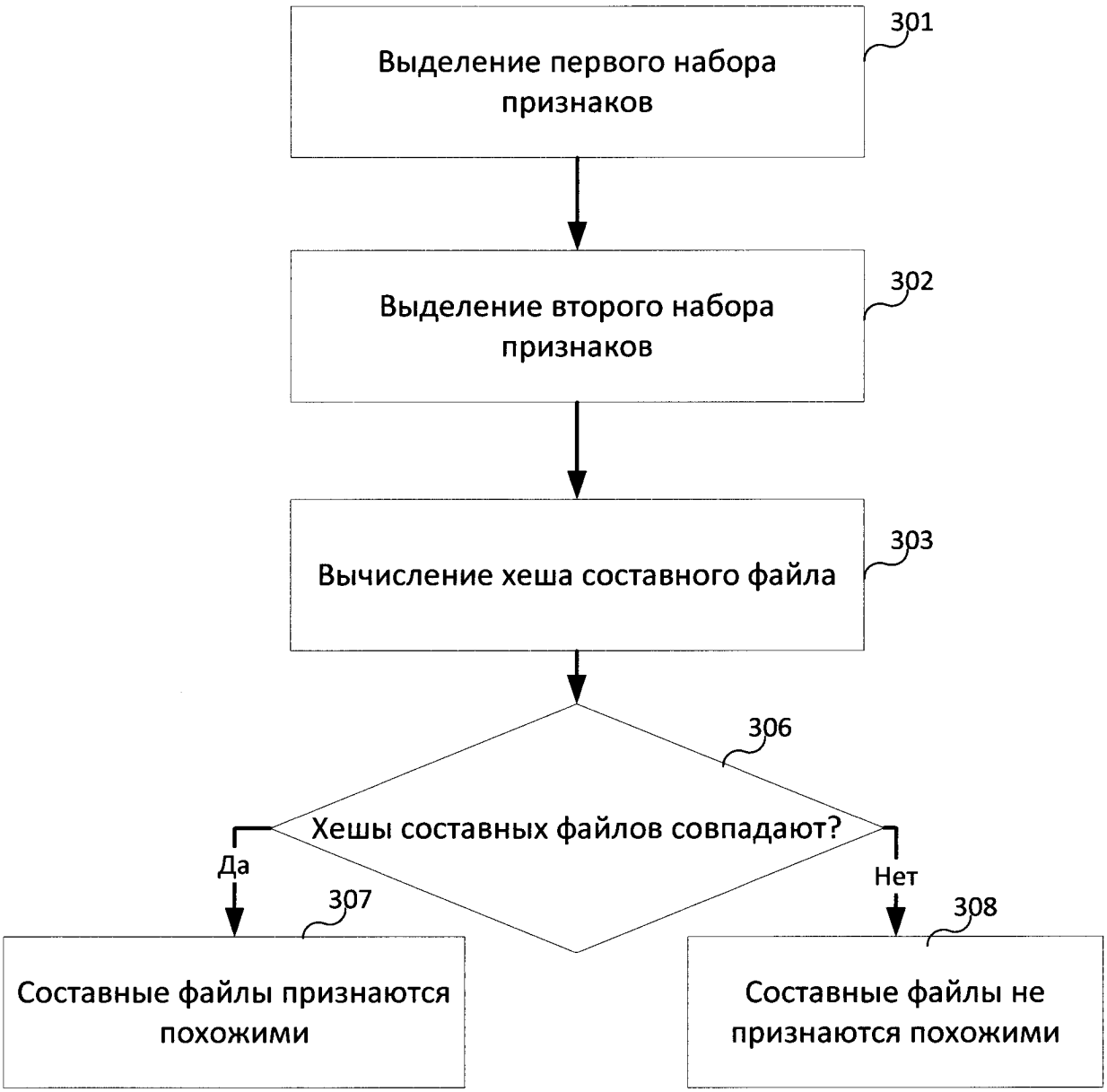
FIELD: information technology.

SUBSTANCE: determining by means of the calculation tool whether the file is composite; extracting the first set of attributes from the header of the composite file with the help of the calculation tool, if previously at the stage it was determined that the file is composite; selecting the second set of attributes from at least one directory of the composite file by the calculation tool if earlier at the stage it was determined that the file is composite; calculating using calculation tool a hash of composite file using at least the first and second set of features; recognizing by means of a comparison tool a composite file is malicious if the computed hash of the composite file is the same as the hash of malicious composite file; wherein the hash of a malicious composite file is stored in a hash database.

EFFECT: protecting the computing device from malicious programs by detecting a malicious composite file.

5 cl, 4 dwg

RU 2634178 C1



Фиг. 3

RU 2634178 C1

RU 2634178 C1

Область техники

Изобретение относится к области защиты вычислительных устройств, а именно к способам обнаружения вредоносных составных файлов.

Уровень техники

С каждым днем количество вредоносных приложений, распространяемых в сети Интернет, становится все больше и больше. Для защиты вычислительных устройств от вредоносных приложений часто используются антивирусные решения, которые, используя один или несколько методов обнаружения (детектирования), например сигнатурный или эвристический анализ, обнаруживают вредоносные приложения, которые, например, загружены из сети Интернет.

Но у методов детектирования также есть ограничения и недостатки: эвристический анализ применим не ко всем типам файлов, а сигнатурный анализ может быть не эффективен при обнаружении полиморфных вредоносных приложений - приложений, выполняющих одни и те же команды, но отличающиеся содержимым соответствующих файлов приложений. Часто такие полиморфные вредоносные приложения (а именно файлы таких приложений) создаются автоматизированным путем (например, генерируются автоматически): создатель вредоносного приложения, как правило, использует специальные среды разработки, которые могут из одного исходного кода вредоносного приложения скомпилировать огромное количество вредоносных файлов, которые будут иметь разное тело файла (содержимое файла), но запущенные из таких файлов приложения будут вести себя одинаково. При этом повышение качества детектирования таких файлов антивирусными решениями часто упирается в определение похожести таких файлов (сходства файлов в соответствии с одной из мер похожести). Стоит отметить, что подобные полиморфные вредоносные файлы могут быть не только файлами формата PE (Portable Executable), но и любыми другими файлами, формат которых позволяет внедрить в файл вредоносный код, который будет исполнен тем или иным образом, например файлы формата Portable Document Format, Microsoft Compound File Binary (OLE2 файлы) или одного из форматов Office Open XML (DOCX, PPTX и т.п.).

Так, в патенте US9396334 описан подход к обнаружению вредоносных файлов, которые могут быть тем или иным образом запущены на стековой виртуальной машине (а именно: файлы содержат код, который может быть интерпретирован). Технология позволяет эффективно обнаруживать, например, вредоносные файлы (группы похожих вредоносных файлов), содержащие Action Script сценарии (или скрипты, англ. script). Однако в приведенной публикации не описаны подходы эффективного обнаружения составных файлов (Compound File), например файлов формата DOC.

Хотя рассмотренные подходы направлены на решение определенных задач в области защиты вычислительных устройств, они не решают задачу обнаружения вредоносных составных файлов или решают недостаточно эффективно. Настоящее изобретение позволяет более эффективно решить задачу обнаружения вредоносных составных файлов.

Раскрытие изобретения

Настоящее изобретение предназначено для обнаружения вредоносных составных файлов.

Технический результат настоящего изобретения заключается в обеспечении защиты вычислительного устройства от вредоносных программ, который достигается за счет обнаружения составного вредоносного файла, при этом составной файл признается вредоносным, если вычисленный хеш составного файла совпадает с хешем вредоносного

составного файла. Технический результат достигается при помощи способа обнаружения вредоносных составных файлов, в котором: определяют при помощи средства вычисления, является ли первый файл составным (compound file); выделяют при помощи средства вычисления первый набор признаков из заголовка первого составного файла, если ранее этапе было определено, что первый файл является составным; выделяют при помощи средства вычисления второй набор признаков из по меньшей мере одной директории (directory entry) первого составного файла, если ранее этапе было определено, что файл является составным; вычисляют при помощи средства вычисления хеш (hash) первого составного файла с использованием по меньшей мере первого и второго набора признаков; признают при помощи средства сравнения составной файл вредоносным, если вычисленный хеш составного файла совпадает с хешем вредоносного составного файла; при этом хеш вредоносного составного файла хранится в базе данных хешей.

В частном случае реализации способа составным файлом является файл формата Microsoft Compound File Binary File Format.

В еще одном частном случае реализации способа составным файлом является файл формата OLE или OLE2.

В другом частном случае реализации способа первый набор признаков включает в себя поля структуры заголовка составного файла.

В еще одном частном случае реализации способа второй набор признаков включает в себя поля структуры директории составного файла.

Краткое описание чертежей

Дополнительные цели, признаки и преимущества настоящего изобретения будут очевидными из прочтения последующего описания осуществления изобретения со ссылкой на прилагаемые чертежи, на которых:

Фиг. 1 показывает пример компонентов системы, с помощью которой может быть реализован способ настоящего изобретения.

Фиг. 2 иллюстрирует элементы системы, участвующие формировании базы данных хешей.

Фиг. 3 иллюстрирует примерный вариант реализации способа изобретения.

Фиг. 4 показывает пример компьютерной системы общего назначения.

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 формуле.

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 не чем иным, как конкретными деталями, необходимыми для помощи специалисту в области техники в исчерпывающем понимании изобретения, и настоящее изобретение определяется в объеме приложенной формулы.

Введем ряд определений и понятий, которые будут использоваться при описании вариантов осуществления изобретения.

Вредоносное приложение - приложение, способное нанести вред компьютеру или

данным пользователя компьютера, например: сетевой червь, клавиатурный шпион, компьютерный вирус. В качестве нанесенного вреда может выступать неправомерный доступ к ресурсам компьютера, в том числе к данным, хранящимся на компьютере, с целью хищения, а также неправомерное использование ресурсов, в том числе для хранения данных, проведения вычислений и т.п.

Доверенное приложение - приложение, которое не наносит вреда компьютеру или его пользователю. Доверенным приложением может считаться приложение, разработанное доверенным производителем ПО, загруженное из доверенного источника (например, сайт, занесенный в базу данных доверенных сайтов) или приложение, идентификатор которого (например, MD5 файла приложения) хранится в базе данных доверенных приложений. Идентификатор производителя, например, цифровой сертификат, может также храниться в базе данных доверенных приложений.

Недоверенное приложение - приложение, которое не является доверенным, но также не признано вредоносным, например, при помощи антивирусного приложения.

Вредоносный файл - файл, являющийся компонентом вредоносного приложения.

Недоверенный файл - файл, являющийся компонентом недоверенного приложения.

Доверенный файл - файл, являющийся компонентом доверенного приложения.

Сигнатурный анализ - технология поиска совпадений какого-либо участка кода программы с известным кодом - сигнатурой - из базы данных сигнатур вредоносных программ с целью обнаружения вредоносной программы. Стоит отметить, что данная технология может применяться как для обнаружения файлов вредоносной программы, так и для обнаружения вредоносного кода в оперативной памяти вычислительного устройства.

Эвристический анализ - технология, заключающаяся в эмуляции работы программы (выполнение кода программы, предназначенного для выполнения с использованием одних программно-аппаратных средств, с использованием других программно-аппаратных средств, отличных от первых), создании журнала вызовов API-функций и поиске совпадений данных из созданного журнала вызова API-функций с данными из базы данных эмуляций вредоносных программ с целью обнаружения вредоносной программы.

Антивирусная запись - информация, необходимая антивирусному приложению для обнаружения вредоносных файлов. Может быть представлена, например, в виде сигнатуры, эвристического правила (которое, например, может быть использовано эвристическим и проактивным анализом), контрольной суммы файла - хеша (в том числе и гибкого хеша - англ. locality sensitive hash - хеша файла, инвариантного к небольшим изменениям файла).

Составной файл (файл-контейнер) - файл формата Microsoft Compound File Binary File Format (MS-CFB), например файл формата OLE или OLE2. Примерами составных файлов могут являться файлы форматов: DOC, PPT, XLS (на ОС Windows файлы имеют расширения .doc, .ppt и .xls соответственно).

Под средствами системы обнаружения вредоносных составных файлов в настоящем изобретении понимаются реальные устройства, системы, компоненты, группы компонентов, реализованные с использованием аппаратных средств, таких как интегральные микросхемы (англ. application-specific integrated circuit, ASIC) или программируемые вентильные матрицы (англ. field-programmable gate array, FPGA) или, например, в виде комбинации программных и аппаратных средств, таких как микропроцессорная система и набор программных инструкций, а также на нейроморфных чипах (англ. neurosynaptic chips). Функциональность указанных средств

системы может быть реализована исключительно аппаратными средствами, а также в виде комбинации, где часть функциональности средств системы реализована программными средствами, а часть аппаратными. В некоторых вариантах реализации часть средств, или все средства, могут быть исполнены на процессоре компьютера общего назначения (например, который изображен на Фиг. 4). При этом компоненты системы могут быть реализованы в рамках как одного вычислительного устройства, так и разнесены между несколькими, связанными между собой вычислительными устройствами.

Фиг. 1 изображает примерный набор компонентов системы, необходимых для реализации способа изобретения. Система обнаружения вредоносных составных файлов в частном случае реализации включает в себя следующие компоненты: средство вычисления 110, средство сравнения 120 и базу данных хешей 130. В предпочтительном варианте реализации изобретения средство вычисления 110 расположено на вычислительном устройстве пользователя, а средство сравнения 120 и база данных хешей 130 - на удаленном сервере (не показано на Фиг. 1). В еще одном варианте реализации средство сравнения 120 и база данных хешей 130 расположены также на вычислительном устройстве пользователя.

Средство вычисления 110 предназначено для вычисления хеша (англ. hash, hash sum) составного файла 105. В общем случае средство вычисления 110 вычисляет хеш составного файла 105, входящий в множество похожих составных файлов, таким образом, чтобы для каждого составного файла из множества похожих составных файлов значение такого хеша совпадало. В частном случае реализации составные файлы являются похожими, если расстояние между составными файлами не превышает порогового значения. Расстояние между файлами - мера похожести, вычисленная одним из известных из уровня техники способов. При этом расстояние между составными файлами может быть Евклидовым расстоянием, Хеммингловым расстоянием или любой другой мерой расстояния, примененной, например, к байтовому представлению составного файла или к блокам байт заданного размера. В еще одном частном случае реализации составные файлы являются похожими, если коэффициент сходства между такими файлами превышает пороговое значение. При этом коэффициентом сходства может являться один из известных из уровня техники коэффициентов сходства, например коэффициент Жаккара. В еще одном частном случае реализации составные файлы являются похожими, если они признаны таковыми специалистом в области информационных технологий.

В частном случае реализации составные файлы являются похожими, если они созданы путем применения средств автоматизированной генерации (создания) файлов к одним и тем же исходным данным. Примером использования такой автоматизированной генерации может быть создание множества составных файлов, например документов формата DOC, на основании одного и того же текста (созданные файлы будут содержать один и тот же текст). Еще одним примером может являться создание множества вредоносных файлов приложения при помощи, например, среды разработки с использованием одного и того же исходного кода вредоносного приложения. Такие вредоносные файлы также называются полиморфными (англ. polymorph malware).

Особенностью использования вышеупомянутых средств автоматизированной генерации является то, что созданные из одних и тех же исходных данных файлы имеют разное байтовое представление, однако полезная нагрузка (англ. payload) остается неизменной. В частном случае реализации под полезной нагрузкой файла подразумевается последовательность команд, исполняющаяся при открытии файла. В

еще одном частном случае под полезной нагрузкой файла подразумевается информация, отображаемая пользователю при открытии файла приложением, предназначенным для открытия таких файлов. Соответственно, в частном случае реализации предназначение средства вычисления 110 - вычислять хеши составных файлов, полученных путем автоматизированной генерации, таким образом, чтобы хеши таких сгенерированных файлов совпадали.

Для вычисления хеша, удовлетворяющего вышеуказанному условию (один и тот же хеш для похожих, например сгенерированных, файлов), средство вычисления 110 извлекает из составного файла 105 признаки файла в два этапа.

На первом этапе средство вычисления 110 выделяет из составного файла 105 первый набор признаков. Первый набор признаков состоит из признаков, расположенных в заголовке составного файла 105 (Compound File Header). Признаками, расположенными в заголовке составного файла 105, в частном случае реализации являются поля структуры заголовка составного файла 105 в соответствии со спецификацией формата MS-CFB (можно ознакомиться на странице https://msdn.microsoft.com/en-us/library/dd941946.aspx). В частном случае реализации первый набор признаков включает в себя полный перечень полей структуры заголовка составного файла 105 (например, поля "Header Signature", "Minor Version", "Major Version" и другие). В еще одном частном случае реализации первый набор признаков включает в себя все поля структуры заголовка составного файла 105, начинающиеся с поля "Number of FAT Sectors", которое находится по смещению $0x2C$ от начала заголовка (поля заголовка, начинающиеся по смещению $0x2C$ и заканчивающиеся в конце заголовка составного файла 105).

На втором этапе средство вычисления 110 выделяет из составного файла 105 второй набор признаков. Второй набор признаков состоит из признаков, расположенных в по меньшей мере одной директории составного файла 105 (Compound File Directory Entry). В частном случае реализации под такими признаками понимают полный перечень полей структуры директории составного файла 105 (например, поля " Directory Entry Name", "Color Flag" и другие). В частном случае реализации изобретения второй набор признаков состоит из признаков, расположенных в первой директории из массива директорий составного файла 105 (Compound File Directory Array). В еще одном частном случае реализации изобретения второй набор признаков состоит из признаков, расположенных в первых четырех директориях (по порядку из массива директорий) составного файла 105 (Compound File Directory Array).

Полученные на вышеописанных этапах (первом и втором) результаты средство вычисления 110 использует для вычисления хеша составного файла 105. В частном случае реализации средство вычисления 110 вычисляет хеш составного файла 105 с использованием первого и второго набора признаков следующим образом: байтовое представление (конкатенация байтовых значений каждого признака из набора признаков) первого набора признаков объединяется с байтовым представлением второго набора признаков при помощи конкатенации, в результате чего формируется байтовое представление признаков составного файла 105. В частном случае реализации упомянутое байтовое представление признаков составного файла 105 дополняется (посредством конкатенации) размером составного файла 105 (байтовое представление значения размера). Далее применяется хеш-функция к полученному байтовому представлению признаков составного файла 105. В частном случае реализации в качестве хеш-функции применяется MD5. В еще одном частном случае реализации применяется SHA-0, SHA-1 или одна из хеш-функций семейства AES. Полученное после применения

вышеупомянутой хеш-функции значение (пусть размер полученного значения - N бит) разбивается на две части равного размера (в битах) - одна часть, характеризующая первые N/2 бит полученного значения, и вторая, характеризующая вторые N/2 бит значения (например, для хеш-функции MD5 такими частями будут первые и вторые 8 байт значения, полученного в результате вычисления хеш-функции). Средство вычисления 110 применяет логическую операцию XOR к двум упомянутым частям. Результат вычисления операции XOR средством вычисления 110 является хешем составного файла 105.

Характерной особенность хеша составного файла 105, вычисленного в соответствии с вышеописанным алгоритмом (выделение признаков заголовка составного файла 105, а также признаков первых четырех директорий из массива директорий составного файла 105 с последующим применением хеш-функции) является то, что хеши полученных путем автоматизированной генерации составных файлов 105 совпадают. Эта особенность вытекает из того, что составные файлы, полученные путем автоматизированной генерации, имеют разное байтовое представление содержимого (а именно секторов составных файлов, содержащих полезную нагрузку файла -секторов FAT, Mini FAT и DIFAT), но структура таких файлов, которая описывается директориями составных файлов (данные, хранящиеся в по крайней мере первых четырех директориях из массива директорий составного файла), идентична. При этом вычисление хеша вышеупомянутым способом является более быстрой операцией, чем вычисление MD5 файла, т.к. даже на этапе использование MD5 (или аналогичных хеш-функций) хеш-функция применяется не ко всему телу файла, а лишь к данным небольшого размера.

Средство сравнения 120 предназначено для сравнения хешей составных файлов 105. В частном случае реализации изобретения, если сравниваемые хеши двух составных файлов 105 совпадают, то средство сравнения 120 признает эти составные файлы 105 похожими. В еще одном частном случае реализации, если хеш составного файла 105 (хеш первого составного файла 105), вычисленный при помощи средства вычисления 110, совпадает с хешем вредоносного составного файла 105, то первый составной файл 105 признается вредоносным. В еще одном частном случае реализации, если хеш составного файла 105 (хеш первого составного файла 105), вычисленный при помощи средства вычисления 110, совпадает с хешем доверенного составного файла 105, то первый составной файл 105 признается доверенным. В еще одном частном случае реализации, если хеш составного файла 105 (хеш первого составного файла 105), вычисленный при помощи средства вычисления 110, совпадает с хешем недоверенного составного файла 105, то первый составной файл 105 признается недоверенным. Хеши вредоносных, доверенных и недоверенных составных файлов хранятся в базе данных хешей 130.

На Фиг. 2 изображены средства, которые используются для формирования данных, хранящихся в базе данных хешей 130. База данных хешей 130 предназначена для хранения хешей составных файлов 105, а также информации о том, какой составной файл 105 использовался для вычисления хеша составного файла 105. Упомянутая информация о составном файле 105 - является ли он вредоносным, доверенным или недоверенным. Таким образом, если при вычислении хеша составного файла 105 использовался вредоносный составной файл из базы данных файлов 210, то полученный хеш будем называть хешем вредоносного составного файла 105. Если при вычислении хеша составного файла 105 использовался доверенный составной файл 105 из базы данных файлов 210, то полученный хеш будем называть хешем доверенного составного файла 105. Если при вычислении хеша составного файла 105 использовался

недоверенный составной файл 105 из базы данных файлов 210, то полученный хеш будем называть хешем недоверенного составного файла 105. Для вычисления хешей составных файлов 105 средство вычисления 110 использует составные файлы 105 из базы данных файлов 210. База данных файлов 210 содержит составные файлы 105, а также информацию о каждом составном файле 105 - является ли он вредоносным, доверенным или недоверенным. В частном случае реализации изобретения данные (составные файлы и информация о них), которые хранятся в базе данных файлов 210, вносятся в базу данных файлов 210 человеком, например специалистом в области информационных технологий. В еще одном частном случае реализации изобретения данные сохраняются в базу данных файлов 210 антивирусным приложением (не показано на Фиг. 2), которое анализирует набор составных файлов, например составных файлов, хранящихся на удаленном сервере, с целью обнаружения вредоносных составных файлов и записывает в базу данных файлов 210 файлы и информацию о них, полученную в результате анализа - являются ли они, например, вредоносными, доверенными или недоверенными. Средство вычисления 110 получает из базы данных файлов 210 составной файл 105 и информацию о нем, вычисляет для этого составного файла 105 хеш составного файла 105 и сохраняет полученную информацию в базе данных хешей 130: хеш составного файла 105 и информацию о файле 105, использованном для вычисления хеша.

Стоит отметить, что в частном случае реализации этап наполнения базы данных хешей 130 вышеупомянутым способом предшествует шагам, раскрытым при описании Фиг. 1.

На Фиг. 3 изображен пример способа обнаружения вредоносного составного файла 105. На этапе 301 средство вычисления 110 выделяет из составного файла 105 первый набор признаков из составного файла 105. В частном случае реализации таким составным файлом 105 может быть любой из составных файлов 105, хранящихся, например, на устройстве хранения данных, расположенном на вычислительном устройстве пользователя. В еще одном частном случае реализации информация о таком составном файле 105 не хранится в базе данных файлов 210. В частном случае реализации будем считать, что относительно такого составного файла 105 не известно, является ли он похожим на файлы из базы данных файлов 210 или же является ли он вредоносным, доверенным или недоверенным. На этапе 302 средство вычисления 110 выделяет из составного файла 105 второй набор признаков составного файла 105. Далее, на этапе 303, средство вычисления 110 вычисляет хеш составного файла 105 на основании первого и второго набора признаков составного файла 105. На этапе 306 средство сравнения 120 сравнивает вычисленный хеш составного файла 105 с хешем второго составного файла. Составной файл 105, хеш которого был вычислен на этапе 303, будем называть первым составным файлом 105, соответственно его хеш - хеш первого составного файла 105. В частном случае реализации хеш второго составного файла 105 вычисляется путем выполнения этапов 301 -303 в отношении второго составного файла 105. В частном случае реализации вторым составным файлом 105 может быть любой из составных файлов 105, хранящихся, например, на устройстве хранения данных на вычислительном устройстве пользователя, и отличный от первого составного файла 105 (в частном случае реализации под отличием понимают по меньшей мере различие каталогов расположения файлов на устройстве хранения данных). В еще одном частном случае реализации второй составной файл 105 - файл, хеш и информация о котором (например, является ли составной файл вредоносным или доверенным) хранятся в базе данных хешей 130 (в частном случае реализации такой файл присутствует в базе данных файлов

210). Если хеши первого и второго составных файлов 105 совпадают, то средство сравнения 120 на этапе 307 признает первый и второй составные файлы 105 похожими. При этом, если хеш второго составного файла 105 хранится в базе данных хешей 130 и является хешем вредоносного составного файла 105, то первый составной файл 105 признается вредоносным. В еще одном частном случае реализации, если хеш второго составного файла 105 хранится в базе данных хешей 130 и является хешем доверенного составного файла 105, то первый составной файл 105 признается доверенным. В еще одном частном случае реализации, если хеш второго составного файла 105 хранится в базе данных хешей 130 и является хешем недоверенного составного файла 105, то первый составной файл 105 признается недоверенным. Если же хеши первого и второго составных файлов 105 не совпадают, то средство сравнения 120 на этапе 308 не признает первый и второй составные файлы 105 похожими.

В частном случае реализации после признания составного файла 105 вредоносным средство сравнения 120 удаляет данный составной файл 105 на устройстве хранения данных, расположенном на вычислительном устройстве пользователя. В еще одном частном случае реализации средство сравнения 130 не удаляет вредоносный составной файл 105, но помещает его в карантин.

В частном случае реализации система дополнительно включает средство контроля приложений (application control - известно из уровня техники), которое при доступе к составным файлам 105 на вычислительных устройствах пользователей передает составной файл средству вычисления 110, которое в свою очередь вычисляет хеш составного файла 105. Затем этот хеш сравнивается средством сравнения 120 с хешами составных файлов из базы данных хешей 130. Результаты сравнения - присутствует ли соответствующий хеш в базе данных 130, а также информация о том, является ли найденный (в случае, если таковой есть) хеш хешем вредоносного, доверенного или недоверенного составного файла 105 - передается средством сравнения 120 средству контроля приложений. При этом средство контроля приложений предоставляет пользователю вычислительного устройства доступ к упомянутому составному файлу 105 только в том случае, если средство сравнения 120 нашло в базе данных хешей 130 вычисленный хеш составного файла 105, и найденный хеш является хешем доверенного составного файла 105.

В еще одном частном случае реализации поучаемые при помощи средства вычисления 110 хеши составных файлов 105 могут использоваться в рамках технологии, описанной в патенте RU2415471 - для определения необходимости проведения антивирусной проверки составного файла. В этом случае получаемые при помощи средства вычисления 110 хеши составного файла 105 используются для определения изменений в составных файлах 105. Соответственно, при очередной антивирусной проверке, которая производится, например, антивирусным приложением (например, проверка устройства хранения данных с целью найти вредоносные файлы), средство вычисления 110 вычисляет хеш каждого просканированного (проверенного антивирусным приложением) составного файла 105. Если при следующей антивирусной проверке антивирусное приложение начнет проверку составного файла 105 и его хеш будет совпадать с ранее вычисленным хешем (вычисленным при предыдущей антивирусной проверке), то проверка в отношении этого составного файла 105 производиться не будет, т.е. необходимость в антивирусной проверке (проверка файла антивирусным приложением) отсутствует.

В частном случае реализации средство вычисления 110 может в рамках любого из вариантов функционирования получить (на вход, в качестве входных данных) любой

файл (не обязательно составной файл 105). В таком случае, прежде чем выделять наборы признаков составного файла 105, средство вычисления 110 определяет формат файла, полученного средством вычисления 110. Дальнейшее выделение наборов признаков осуществляется средством вычисления 110 только в том случае, если средство вычисления 110 определит, что полученный средством вычисления 110 файл является составным. В частном случае реализации изобретения определение того, является ли полученный файл составным или нет, производится на основании данных из файла: если файл начинается с последовательности байт "d0 cf 11 e0 a1 b1 1a e1" или " 0e 11 fc 0d d0 cf 11 0e", то файл считается составным. В другом частном случае реализации для определения того, является ли файл составным, может быть использован любой из известных из уровня техники способов определения составного файла.

Фиг. 4 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, сканнер 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. 4. 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.

(57) Формула изобретения

1. Способ обнаружения вредоносных составных файлов, в котором

a) определяют при помощи средства вычисления, является ли файл составным;

b) выделяют при помощи средства вычисления первый набор признаков из заголовка составного файла, если ранее на этапе было определено, что файл является составным;

c) выделяют при помощи средства вычисления второй набор признаков из по меньшей мере одной директории составного файла, если ранее на этапе было определено, что файл является составным;

d) вычисляют при помощи средства вычисления хеш составного файла с использованием по меньшей мере первого и второго набора признаков;

e) признают при помощи средства сравнения составной файл вредоносным, если вычисленный хеш составного файла совпадает с хешем вредоносного составного файла; при этом хеш вредоносного составного файла хранится в базе данных хешей.

2. Способ по п. 1, в котором составным файлом является файл формата Microsoft Compound File Binary File Format.

3. Способ по п. 2, в котором составным файлом является файл формата OLE или OLE2.

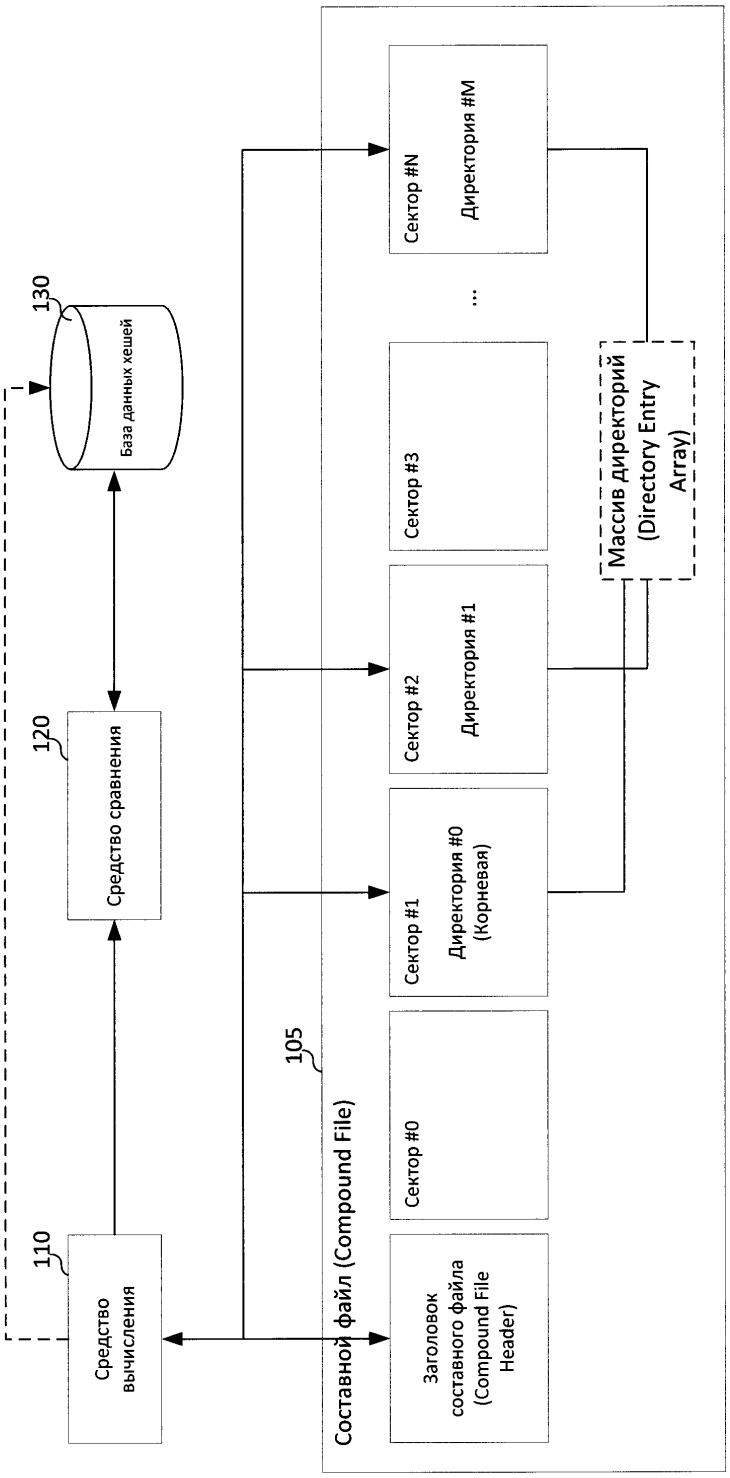
4. Способ по п. 1, в котором первый набор признаков включает в себя поля структуры

заголовка составного файла.

5. Способ по п. 1, в котором второй набор признаков включает в себя поля структуры директории составного файла.

1

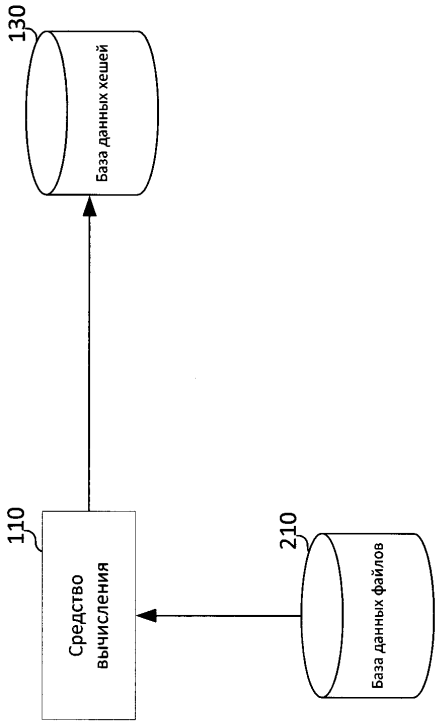
Способ обнаружения вредоносных составных файлов



Фиг. 1

2

Способ обнаружения вредоносных составных файлов



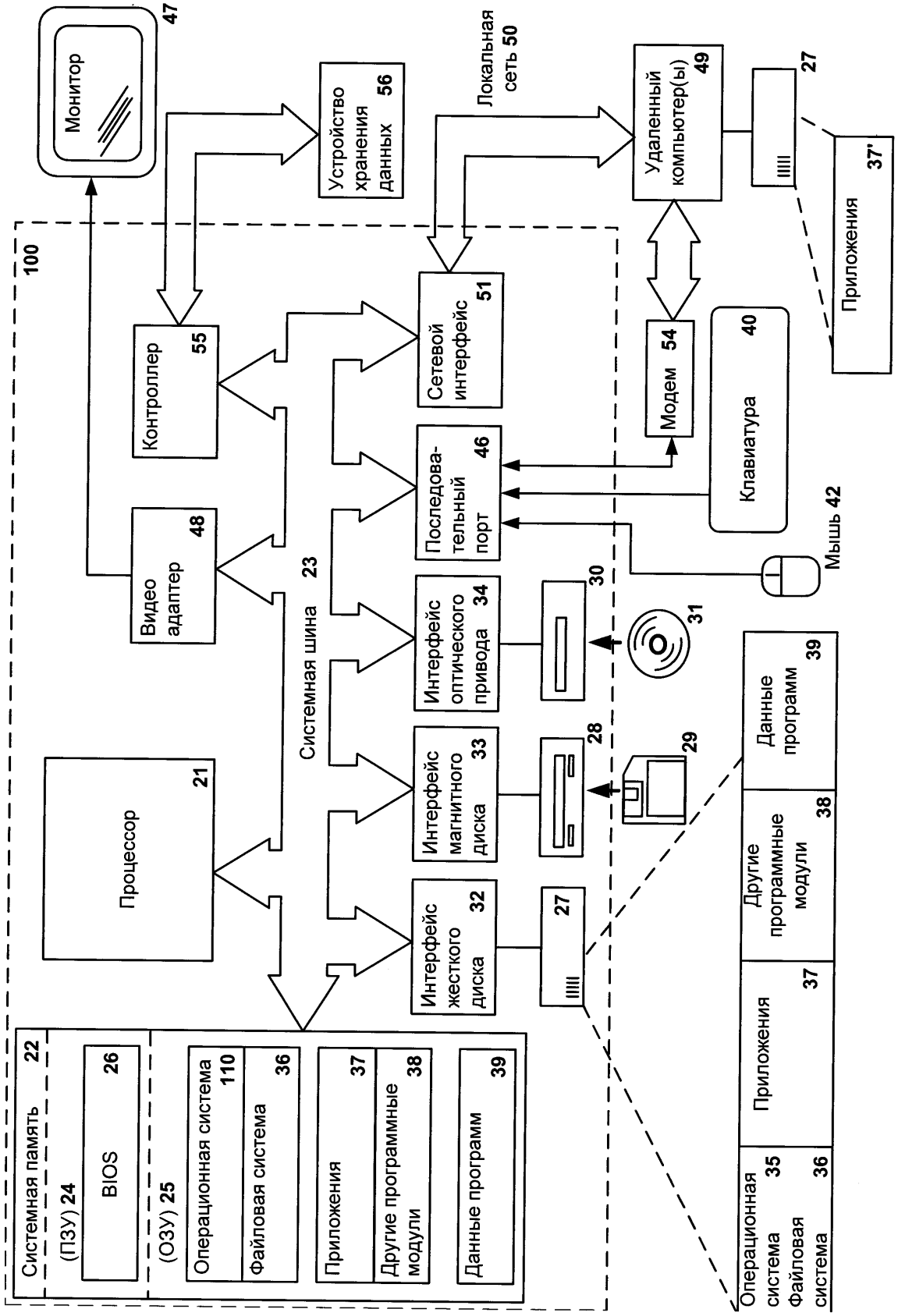
Фиг. 2

3

Способ обнаружения вредоносных составных файлов

Фиг. 3

4



Способ обнаружения вредоносных составных файлов Фиг. 4