Тема: Вредоносные программы: классификация, методы обнаружения

Одной из важнейших задач компьютерной безопасности является борьба с вредоносным программным обеспечением (ВПО) и в частности подзадача его обнаружения. Все методики обнаружения можно разделить на 2 типа: методики обнаружения известного ВПО и методики обнаружения неизвестного ВПО [].

В свете текущих тенденций развития вредоносных программ, все более актуальными становится задача создания эффективного средства обнаружения неизвестного ВПО. Автором ведется работа по созданию подобного средства. Целью разработки является повышение эффективности обнаружения за счет совершенствования существующих методик. Был проведен анализ имеющихся методик c целью выявления их характеристик. В результате этого исследования была сформирована классификация методик, которая будет представлена в данной работе (см. рис. 1).

Классификация методик обнаружения неизвестного ВПО Методику обнаружения неизвестного ВПО можно описать с помощью следующих параметров: данные, получаемые об исследуемом ПО, способы получения этих данных, математические методы, применяемые для анализа данных, и выявляемые признаки вредоносности []. Комбинация этих параметров определяет основные характеристики методики: уровень ошибок первого и второго рода, ресурсоемкость, вычислительную сложность, алгоритмическую сложность (трудоемкость реализации) и др. Поэтому построенная классификация рассматривает методики в контексте каждого из этих параметров. Классификация по характеру получаемых данных. По характеру получаемых данных методики принято разделять на структурный анализ и поведенческий анализ []. Структурный анализ [] учитывает тот факт, что некоторые виды ВПО (например, вирусы) имеют отличительные особенности в структуре: расположение точки входа, специфичные последовательности команд, а также многие признаки, обнаруживаемые при так называемом эвристическом анализе. Данный вид анализа выявляет в основном косвенные признаки вредоносности, которые напрямую не указывают на вредоносность ПО, но крайне редко наблюдаются в полезном ПО. Структурный анализ в большинстве случаев имеет высокую скорость работы (по причине небольшой вычислительной сложности). Главным минусом данного подхода является то, что не все типы ВПО имеют структурные отличия от полезного ПО. Таким образом, не все виды ВПО можно обнаруживать данным методом. Также к этой категории можно отнести методики, которые анализируют бинарное подобие исследуемого ПО и известных вредоносных программ []. Но на практике данные методики не дают приемлемых результатов. Поведенческий анализ [] исследует действия, выполняемые ПО, и их последствия. Такие методики определяют вредоносность программ по тем же признакам, что и человек – по их поведению. В идеальном исполнении система, реализующая данный подход, способна защитить от любого ВПО, но на практике создать такую систему невозможно. С одной стороны слежение за всеми действиями ПО – алгоритмически сложная, ресурсоемкая и в некоторых случаях невыполнимая задача. С другой стороны, невозможно полностью формализовать понятие «вредоносное поведение». На практике такие методики следят за ограниченным набором действий, выполняемых ПО, и пытаются выявить в них ограниченный набор признаков вредоносности.

Таким образом, к плюсам поведенческого анализа можно отнести теоретическую возможность обнаружения любого типа ВПО, а также и возможность обнаружения ВПО в момент совершения вредоносного действия, а к минусам – практическую невозможность полного контроля системы, ресурсоемкость (чем больше контроль, тем сильнее замедляется работа всей системы). Среди часто обсуждаемых проблематик поведенческого анализа можно указать вопрос полноты информации получаемой о ПО, а так же задачу анализа потока информации с различными требованиями. Вопрос полноты полученной информации в основном возникает при попытке выяснения всех возможных действий, которые может выполнять программа. Данный вопрос особо остро проявляется при обнаружении так называемых «временных бомб». Этот вид ВПО выполняет вредоносные действия только при определенных условиях, например в конкретную дату. Таким образом, при невыполнении данного условия поведенческий анализатор не сможет обнаружить такое ВПО, и оно будет беспрепятственно себя распространять. Некоторые виды поведенческого анализа накладывают свои требования к методам анализа. Так для исследования потока исполнения (например, анализ системных вызовов на компьютере конечного пользователя) необходимы методы, которые будут его анализировать по мере поступления данных за гарантированный промежуток времени. В противном случае антивирус будет замедлять работу системы тем самым мешать пользователю. Классификация по способу получения данных. Существующие методики обнаружения неизвестного ВПО по способу получения данных об исследуемом ПО можно разделить на две категории: методики исполняющие и не исполняющие код программы. Методики, не исполняющие программный код, в основном применяются в структурном анализе, поэтому их основные достоинства и недостатки совпадают. Главный недостаток состоит в невозможности обнаружения ВПО, особенности которого проявляются только при исполнении кода. Главными преимуществами данного подхода являются высокая скорость, низкая ресурсоемкость и безопасность использования. Помимо того, что данные способы позволяют относительно быстро обнаруживать некоторые виды ВПО, они позволяют ускорить работу других способов. Так методика определения измененных файлов по контрольным суммам, также относящаяся к данному виду, позволяет ускорить анализ ПО с помощью других методов, например, не анализировать файл повторно, если он не был изменен. Методики, исполняющие программный код, применяются в основном в поведенческом анализе. К таким методикам относятся сбор данных при исполнении ПО на реальной системе (например, компьютер пользователя или «honeypot»), сбор данных при исполнении ПО на эмуляторах, а также смешанные способы. Каждый из указанных подвидов имеет свои достоинства, недостатки и области применения. При исследовании работы ПО в реальной системе обычно контролируют ряд действий во время выполнения ПО (системные вызовы, обращение к файлам) и изменения в системе после выполнения ПО (изменения в файловой системе).

Реализация данного подхода на компьютерах конечных пользователей следит за действиями программ и с одной стороны позволяет обнаружить ВПО в момент исполнения вредоносного действия, а с другой стороны замедляет работу системы. Достоинствами данного подхода по сравнению с созданием полноценных эмуляторов являются возможность исследования ПО в «естественных условиях», высокая скорость сбора данных и относительная простота реализации. Недостатки обусловлены следующими обстоятельствами. Существует поведение, данные о котором невозможно собирать напрямую – возможно только косвенное их получение. В первую очередь, это относится к работе на уровне ядра системы. Некоторые виды ВПО используют stealth-технологии, препятствующие контролю за действиями программы, или не выполняют вредоносных действий при обнаружении контроля за собой. Для данного подхода к исследованию ПО не найдено приемлемое решение задачи исследования всего функционала ПО. Если ПО исполняется на компьютере пользователя, существует вероятность, что к моменту обнаружения оно уже нанесет ущерб пользователю. Исследование работы ПО с помощью эмуляторов основано на эмуляции поведения системы: центрального процессора, операционной системы и пр. Фактически исследуемая программа выполняется не на реальной системе, а на специальном интерпретаторе. К достоинствам такого подхода относятся безопасность его использования, теоретическая возможность полного контроля над действиями программы, а также возможность исследования всего функционала ПО. Недостатками являются крайне высокая алгоритмическая сложность данного подхода и низкая скорость работы. Смешанные способы в основном предполагают выполнение программы на реальном процессоре, но в изолированной среде. Таким образом, они сочетают достоинства и недостатки обоих рассмотренных выше подходов. К смешанным способам относят. Использование виртуальных машин. Этот довольно ресурсоемкий подход, который полностью изолирует исследуемую программу от реальной системы, но в ряде задач он оказывается быстрее эмулятора. Использование «песочниц» («sandbox»). Данный подход является менее ресурсоемким и более прост в реализации, чем предыдущий, но он не гарантирует полную изоляцию ПО от реальной системы. Классификация методов анализа. Данная классификация основана на способах накопления знаний, которые используют методы анализа, и выделяет две группы методов: методы, основанные на экспертных знаниях, и различные методы машинного обучения; Методы, основанные на экспертных знаниях, используются для формализации понятия «вредоносности» и знаний экспертов в области исследования вредоносных программ. Знания могут быть связаны, например, с тем, какие действия являются вредоносными (поведенческий анализ), или какие особенности структуры могут говорить о вредоносности (структурный анализ). В дальнейшем эти формализованные знания применяются для обнаружения вредоносности в анализируемых данных. Представителями данной группы методом являются. Метод продукционных правил []. Это один из самых простых в реализации, но довольно эффективный метод. В его основу положена модель представления знаний в виде конструкций «ЕСЛИ-ТО». С помощью таких правил можно указать одиночные признаки вредоносности. Поиск поведенческих сигнатур []. Данный метод разработан для поведенческого анализа. За основу взят метод продукционных правил, который был адаптирован для обнаружения вредоносных последовательностей действий (т.е. определения перехода системы в «зараженное» состояние). Метод, основанный на нейро-нечетких сетях []. В основе данного метода также лежат правила, задаваемые экспертом. Используемый в нем нечеткий логический вывод позволяет определять комплексные признаки, а элементы искусственных нейронных сетей позволяют подстраивать правила на основе известных ВПО. Все методы данной группы довольно качественно определяют признаки, заданные экспертами, и для них характерен высокий процент обнаружения вредоносных программ, обладающих данными признаками. Тем не менее, более сложные признаки и новые техники, используемые ВПО, эти методы не определяют. Методы машинного обучения используются для «извлечения» знаний (признаков вредоносности) на основе анализа известных ВПО. Развитию данных методов способствует наличие большого количества известных ВПО и тот факт, что основная масса нового ВПО использует сходные технологии, а иногда являются модификацией известного ВПО. Основная задача методов машинного обучения состоит в определении зависимости между исследуемыми данными и выявляемыми признаками вредоносности. Исследования показали, что эффективность этих методов зависит от характера обнаруживаемых признаков, подбора входных данных и качества обучения, следовательно, в общем случае точность этих методов сравнить затруднительно. Однако можно выбрать наиболее подходящий метод для обнаружения конкретного признака при наличии конкретного набора данных и обучающей выборки. К методам машинного обучения относятся: методы, основанные на теореме Байеса[, ]; метод опорных векторов[]; деревья решений[, ]; искусственные нейронные сети[]; генетические алгоритмы[] и др. Классификация по выявляемым признакам вредоносности. В соответствии с данной классификацией выделяют два типа методик: обнаружение аномалий и обнаружение злоупотреблений. Методы обнаружения злоупотреблений основаны на описании вредоносных действий и попытке обнаружения этих действий в исследуемом ВПО. Данный подход подобен сигнатурному поиску, используемому для обнаружения известного ВПО. Сигнатурный поиск используется для нахождения совпадений по коду программы, а методы обнаружения злоупотреблений – для поиска совпадений, например, в поведении. Многие методики данной группы относительно легко реализуются и дают приемлемые уровни ошибок первого рода. Вместе с тем, данные методы неспособны различать новые техники работы ВПО, т.е. новые признаки вредоносности. В настоящее время обнаружение злоупотреблений является наиболее распространенным подходом к обнаружению неизвестного ВПО. Методы обнаружения аномалий основаны на описании нормальных (эталонных) особенностей программы (например, поведения) и попытке обнаружения отклонений от этого эталона. Описание эталона затрудняется наличием очень сложных программ, а также описанной выше проблемой полноты исследования программы. Данный подход алгоритмически более сложный и зачастую более ресурсоемкий, чем обнаружение злоупотреблений, однако он позволяет обнаруживать не только известные, но также и неизвестные новые признаки и техники. Для пояснения данной классификации рассмотрим также классификацию неизвестного ВПО относительно конкретной методики, в соответствии с которой можно выделить ВПО подобное тому, на котором обучалась методика и не являющееся таковым. Иными словами, такая классификация представляет собой деление по степени подобия набора признаков вредоносности, которыми обладает ВПО, и набора признаков, обнаруживаемых методикой. Существующие методики балансируют между: эффективным обнаружением подобного ВПО (эффективность определяется вероятностью ошибки 1-го рода) и низким обнаружением остальных его видов (методы обнаружения злоупотреблений); менее эффективным обнаружением подобного ВПО, но более эффективным обнаружением остальных его видов (методы обнаружения аномалий).