В диссертационный совет 24.1.206.01 на базе Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Израилова Константина Евгеньевича на тему **«Генетический реверс-инжиниринг для нейтрализации уязвимостей программного обеспечения»**, представленную на соискание ученой степени доктора технических наук по специальностям 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» и 2.3.6 «Методы и системы защиты информации, информационная безопасность»

1. Актуальность темы диссертационного исследования

Функционирование современного общества напрямую связано с использованием программных систем (как совокупности разнородных и сложно-взаимодействующих программ), уязвимости в которых существенно повышают успешность проводимых кибератак. Безопасное использование таких систем основано не только на обнаружении уязвимостей, но и на последующем их устранении, что в диссертационной работе названо процессом их нейтрализации.

Уязвимости могут внедряться на всех этапах создания программы – начиная от проектирования ее архитектуры (или даже концепции) и заканчивая выполняемым (т.е. машинным или байт-) кодом. Также возможна злонамеренная модификация уже готового программного продукта. При этом обнаружение уязвимостей целесообразно производить в представлении программы именно на том этапе, где она была внедрена; например, поиск архитектурных ошибок по выполняемому коду будет явно неэффективен. Однако, существующие методы и средства, как правило, направлены на получение исходного кода из выполняемого – т.е. на его декомпиляцию. Дальнейшие же преобразования для получения алгоритмов или архитектуры – т.е. проведение полного цикла реверс-инжиниринга – как правило, не рассматриваются.

Получение представления программы, которое не гарантированно соответствует выполняемому коду, мало результативно для устранения уязвимостей, поскольку из полученного таким образом безопасного кода будет практически невозможно получить требуемую программную систему. Однако, большое количество существующих декомпиляторов предназначены для получения псевдоисходного кода, который не обязан

даже компилироваться, поскольку в основном предназначен для понимания общей логики работы программы.

Наличие достаточно большого количества процессорных архитектур, языков программирования, нотаций записи алгоритмов программ, их архитектур требует такое же количество средств преобразования представлений. Данная потребность, очевидно, не соответствует существующему уровню развития технологии реверс-инжиниринга (включающей, в том числе, и декомпиляцию).

Для разрешения вышеуказанных противоречий автором предлагается гипотеза касательно сведения задачи декомпиляции (а в более общем случае – реверс-инжиниринга) к оптимизационной (с целевой фитнес-функцией, как близостью программ в одном представлении), в результате решения которой конструкции программы в более раннем представлении «подбираются» к такой их совокупности, которая позволит получить программу в текущем представлении. В результате могут быть получены все предыдущие представления программы, которые однозначно соответствуют ее выполняемому коду, и при этом, в той форме, которая наиболее предпочтительна для обнаружения (экспертно или специализированными программными средствами) и устранения уязвимостей. В качестве основы подхода к решению задачи автор обоснованно предлагает применение генетических алгоритмов. Разрешаемая таким образом проблема определяется несоответствием уровня развития теории и технологий реверс-инжиниринга для нейтрализации уязвимостей программного обеспечения потребностям практики. Следовательно, применение реверс-инжиниринга для нейтрализации уязвимостей программного обеспечения на базе генетических алгоритмов является актуальной темой и представляет научный и практический интерес.

2. Достоверность и новизна результатов

Достоверность результатов, основные выводы диссертационной работы обеспечиваются корректностью постановки научной проблемы исследования, строго обоснованной совокупностью ограничений и допущений, представительным библиографическим материалом, опорой на современную научную базу, корректным применением апробированных классических и современных методов исследования и подтверждается непротиворечивостью полученных результатов и их практическим применением для реверс-инжиниринга программных систем, деэволюции представлений программ и поиска в них уязвимостей. Результаты прошли апробацию на представительных российских и зарубежных научных форумах; также, на них было получено 14 свидетельств о государственной регистрации программ для ЭВМ.

Научная новизна работы определяется новизной ее *результатов* и состоит в следующем:

1. Создана оригинальная *методология реверс-инжиниринга программной системы в интересах обнаружения уязвимостей*, которая отличается охватом и ориентацией всего спектра существующих (и потенциально необходимых) методов и

средств на синтез требуемых представлений программы с последующим их анализом (п. 1 паспорта специальности 2.3.6).

2. Синтезирована *модель жизненного цикла программы с разноуровневыми уязвимостями с позиции эволюции представлений*, которая отличается от аналогов высоким уровнем абстракции, наличием единого набора признаков представлений и учетом различных сценариев процесса; также, на базе элементов модели построена новая классификация уязвимостей согласно их структурному уровню (п. 3 паспорта специальности 2.3.6).

3. Создан *комплекс методов генетической деэволюции представлений программы*, который отличается от аналогичных методов, непосредственно преобразующих текущее представление в предыдущее, тем, что задает противоположную направленность, поскольку выполняет множественные преобразования программы из предыдущего представления в текущее; при этом методы комплекса отличаются инвариантностью к представлениям, поскольку не зависят от их синтаксиса (в т. ч. от процессорных архитектур) (пп. 1 и 4 паспорта специальности 2.3.5).

4. Реализован *программный инструментарий генетической декомпиляции машинного кода*, состоящий из 7 инструментов, обладающих собственной новизной: *метод генетической декомпиляции представлений программы* отличается от аналогичных методов декомпиляции интеллектуальной основой в виде генетических алгоритмов; *модель представления исходного кода* отличается от аналогичных возможностью записи программы в виде последовательности чисел, определяющих альтернативные ветки по графу синтаксических правил; алгоритмы для работы с моделью исходного кода и его генерации являются новыми, поскольку функционируют на базе предложенного модельного представления программы, как полного пути по графу синтаксических правил; алгоритмы прогнозирования размера исходного кода по размеру его ассемблерного кода и предсказания константных значений в исходном коде получены впервые; метрика сравнения двух ассемблерных кодов отличается от аналогичных чувствительностью к близости их текстового представления и независимостью от их форматов, а также наличием коэффициентов учета влияния позиций строк и символов (п. 3 паспорта специальности 2.3.5).

5. Спроектирована *архитектура системы генетического реверс-инжиниринга представлений программы с интеллектуальным поиском уязвимостей*, являющаяся результатом программного проектирования и реализации предложенного авторского комплекса методов генетической деэволюции представлений программ, а также необходимого для него инструментария. Введен модуль поиска уязвимостей, использующий новый интеллектуальный способ сигнатурного поиска на базе генетического представления программы; при этом, генетическая сигнатура уязвимости в отличие от существующих контекстно-зависимых аналогов, не зависит от исходного представления программы, а определяется выбранным графом синтаксических правил для текущего. *Программно реализован центральный компонент* предложенной архитектуры системы генетического реверс-инжиниринга (п. 17 паспорта специальности 2.3.6).

Таким образом, диссертационная работа безусловно обладает новизной и относится к двум специальностям – 2.3.5 и 2.3.6.

Теоретическая значимость результатов состоит в следующем: обобщены и систематизированы основные методы и средства, составляющие процесс реверс-инжиниринга программной системы, в виде графо-аналитической схемы; установлена двунаправленная (де)эволюционная взаимосвязь между различными представлениями программы в соответствии с этапами и сценариями ее разработки; обоснована возможность и целесообразность проведения реверс-инжиниринга программ с помощью генетических алгоритмов и получена аналитическая запись механизма, лежащего в основе соответствующего комплекса методов; предложена синтаксически-оптимальная модель исходного кода программы в виде пути по графу синтаксических правил для представления; установлена статистическая зависимость между размерами программ в двух соседних представлениях – исходного и машинного кодов; вычислена статистическая вероятность появления константных значений в исходном коде программ; получена формула определения близости двух ассемблерных кодов, представленных как списки символьных строк; осуществлено качественное развитие сигнатурных способов поиска уязвимостей (как и способов их хранения) в сторону генетического представления программ.

Практическая значимость результатов состоит в возможности: проектирования и непосредственного проведения реверс-инжиниринга современных программных систем, имеющих различную структуру, аппаратное расположение и особенности функционирования, что существенно повышает возможности нейтрализации в них уязвимостей; проведения деэволюцию соседних представлений программы без непосредственного участия эксперта; получения исходного кода программы из ее ассемблерного кода для дальнейшего поиска уязвимостей как предложенным, так и существующими способами (затем обнаруженные уязвимости могут быть устранены с получением соответствующей безопасной версии программы).

3. Степень обоснованности научных положений, выводов и рекомендаций

Содержание и структура диссертации Израилова К.Е. соответствует теме исследования, поставленной цели и решаемым задачам. Диссертация написана на русском языке, состоит из введения, семи глав, заключения, списка сокращений и условных обозначений, списка литературы из 370 источников, двадцати шести приложений (включая акты внедрения, реализации и использования, а также список работ автора по теме диссертации). Полный объем диссертации составляет 470 страниц (в том числе, 417 страниц основного текста), включая 94 рисунка и 26 таблиц. Структура диссертации логична, а ее текст, рисунки и таблицы оформлены в соответствии со стандартами ГОСТ.

Во введении обосновывается актуальность темы исследования, выявляется основное противоречие предметной области (декомпозируемое на 3 частных), выдвигается гипотеза, вводится используемая далее терминология, формулируется научная проблема и получаемое при ее разрешении научное достижение, ставится цель

и необходимые для ее достижения задачи, определяются объект и предмет исследования, новизна, теоретическая значимость и практическая ценность результатов, излагаются выносимые на защиту научные положения.

В первой главе рассматривается программная система с позиции путей повышения эффективности процесса нейтрализации уязвимостей в ее программах. Раскрывается основное противоречие предметной области и его декомпозиция на частные, с каждым из которых связан свой фактор результативности процесса. Вводится эффективностная модель нейтрализации уязвимостей в схематичном (иерархическом) и аналитическом виде. Дается формальная постановка цели и основной задачи исследования, выдвигается и обосновывается гипотеза исследования касательно сведения процесса реверс-инжиниринга программы к решению оптимизационной задачи.

Во второй главе рассматривается процесс реверс-инжиниринга программной системы с позиции обобщения и систематизации существующих методов и средств. Предлагается методология проведения реверс-инжиниринга программной системы (в интересах поиска уязвимостей) в схематичном и аналитическом виде; часть отсутствующих элементов методологии доводится до работоспособных программных прототипов.

В третьей главе рассматривается процесс создания программы с позиции выделения ее представлений, их форм, а также внедряемых уязвимостей и способов их поиска. Синтезируется модель жизненного цикла представлений программы, на базе которой автором вводятся собственное определение уязвимости и их классификация; проводится эксперимент, подтверждающий предложенные научные выкладки. Дается аналитическая запись всех элементов модели.

В четвертой главе рассматривается процесс деэволюции представлений программы с позиции его автоматизации (за счет применения генетических алгоритмов). Производится сравнение существующих подходов к декомпиляции машинного кода, исходя из чего исследуется гипотетический подход на основе полного перебора конструкций исходного кода; в результате развития которого предлагается подход к деэволюции представлений программы путем решения оптимизационной задачи. Осуществляется сравнение эвристических алгоритмов с позиции решения данной задачи, которое позволило сделать выбор в сторону генетических. Как следствие, создаются схемы реверс-инжиниринга представлений программы и ее центральной части, основанные на применении генетических алгоритмов. В интересах применения данного подхода предлагается необходимый иерархический комплекс методов генетической деэволюции, состоящий из метода проведения реверс-инжиниринга программы, который использует метод последовательной деэволюции соседних представлений программы, основанный на применении генетических алгоритмов (и их операций – вычисления фитнес-функции, скрещивания и мутации). Проведенный эксперимент позволил обосновать концепцию генетической деэволюции представлений.

В пятой главе рассматривается метод генетической деэволюции представлений программы с позиции его реализуемости. Для этого предлагается программный инструментарий генетической декомпиляции машинного кода и дается схема его

совместного функционирования. Приводится детальное описание каждого из семи разработанных инструментов, а также результаты экспериментов с их программными прототипами.

В шестой главе рассматривается процесс генетического реверс-инжиниринга с позиции проектирования единой системы его проведения в виде программной архитектуры. Предлагается общая и детализированная схема такой архитектуры, которая также имеет аналитическую запись. Описывается разработанный программный прототип центрального компонента архитектуры, осуществляющий генетическую декомпиляцию. Базовые эксперименты с прототипом позволили получить ряд характеристик процесса. Используя генетическое представление исходного кода, предлагается соответствующий сигнатурный способ поиска уязвимостей. Приводится методика применения системы генетического реверс-инжиниринга.

В седьмой главе рассматривается процесс генетического реверс-инжиниринга программы с позиции оценки его работоспособности и эффективности (включая сравнение с аналогами). Приводятся ход и результаты целой серии экспериментов, состоящих из следующих: поиск уязвимостей с помощью созданного сигнатурно-генетического способа и исследование соответствующей динамики значений фитнес-функции; проверка работоспособности прототипа генетической декомпиляции на базовом примере и получение значений его основных показателей; применение того же прототипа для крупного проекта. Производится обобщенная оценка эффективности нейтрализации уязвимостей с применением различных подходов к реверс-инжинирингу – генетической деэволюции и существующих аналогов; оценка показала преимущество подхода, предложенного автором в диссертации. Используя эффективностную модель нейтрализации уязвимостей, введенную в 1-й главе, производятся необходимые расчеты, которые также показали превосходство генетической деэволюции над другими подходами. Указаны рекомендации по развитию генетического реверс-инжиниринга программ.

В заключении формулируются итоги диссертационного исследования и основные выводы, указывается соотнесение результатов исследования с паспортами специальностей, даются рекомендации и перспективы дальнейшей разработки темы.

В приложениях приводятся вспомогательные материалы исследования (алгоритмы, отладочный вывод работы прототипов, формальные записи синтаксисов и др.), полученные акты (внедрения, реализации и использования) и список работ автора по теме диссертации (из 114 источников).

Диссертация является завершенным научно-исследовательским трудом. Задачи, поставленные автором, решены полностью, цель исследования достигнута. Полученные в работе научные положения, выводы, а также, рекомендации по развитию подхода генетического реверс-инжиниринга являются обоснованными.

Основные положения диссертации опубликованы в 114 научном труде, в том числе: 48 – в рецензируемых научных изданиях, входящих в перечень ВАК РФ (из них, 14 входят в ядро РИНЦ, 14 из Белого списка); 21 – в изданиях, входящих в международную систему цитирования (10 в журналах, 11 в сборниках трудов

конференций). Также, получено 14 свидетельств о государственной регистрации программ для ЭВМ. Результаты прошли широкую апробацию на представительных российских и зарубежных научных форумах.

4. Соответствие специальности

Выполненное соискателем научное исследование соответствует двум паспортам специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» (научные положения 3 и 4) и 2.3.6 «Методы и системы защиты информации, информационная безопасность» (научные положения 1, 2 и 5).

5. Замечания по диссертационной работе

Имеются следующие замечания к работе:

1) В условиях, когда большие языковые модели являются современным трендом (или даже «бумом») в области информационных технологий, при выборе подхода, основанного исключительно на генетических алгоритмах, желательно было бы сравнить их эффективность с альтернативными подходами.

2) Очевидно, что усложнение синтаксиса языка программирования негативно скажется на скорости работы генетического декомпилятора. Однако в работе отсутствует описание механизмов, препятствующих или компенсирующих падение производительности при таком усложнении.

3) Результаты эксперимента по применению прототипа генетического декомпилятора для крупного проекта (Параграф 7.2.2) хотя и не вызывают сомнения, однако наличие скриншотов и логов эксперимента усилило бы доказательную базу (например, аналогично приведенным в Параграфе 6.3).

4) Целесообразность введения новой терминологии (генетический реверс-инжиниринг, деэволюция и декомпиляция) должна быть раскрыта более подробно с точки зрения ее методологической полезности.

5) Модель жизненного цикла описывает идеализированный процесс создания программы, но не учитывает современные итеративные и гибкие методологии разработки программного обеспечения.

6) Возможность практического внедрения прототипа генетической деэволюции в процессы анализа защищенности программного обеспечения в государственном или корпоративном секторе требует дополнительного обсуждения.

7) Утверждение касательно инвариантности подхода генетической деэволюции к синтаксисам/нотациям не подкреплено строгим доказательством для произвольной пары представлений.

8) Объем основного текста диссертации (470 стр.) кажется избыточным, а часть содержания можно было бы без потери смысла отнести к приложениям (например, ход и результаты эксперимента в Параграфе 2.4.4, аналитическую запись модели в Параграфе 3.4.1, графические схемы в Параграфе 3.6.3, таблицы и гистограммы в Параграфе 5.8.4, листинги ассемблерного и исходного кода в Параграфе 6.3 и др.).

6. Заключение

Диссертация Израилова К.Е. «Генетический реверс-инжиниринг для нейтрализации уязвимостей программного обеспечения» представляет собой законченную научно-квалификационную работу, посвященную разрешению актуальной проблемы, имеющей важное значение в области безопасности программного обеспечения. Диссертация обладает научной новизной, имеет теоретическую значимость и практическую ценность. Полученные результаты в полной мере отражены в существенном наборе авторских публикаций. Автореферат полностью отражает содержание диссертации.

Диссертация соответствует критериям, предъявляемым ВАК Минобрнауки России к диссертациям на соискание ученой степени доктора технических наук (пп. 9–14 «Положения о присуждении ученых степеней», утвержденным постановлением Правительства Российской Федерации № 842 от 24 сентября 2013 г. (с изменениями и дополнениями), а ее автор, Израилов Константин Евгеньевич, заслуживает присвоения ученой степени доктора технических наук по специальностям 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» и 2.3.6 «Методы и системы защиты информации, информационная безопасность».

Официальный оппонент
Доктор технических наук, профессор, Заслуженный деятель науки РФ, заведующий кафедрой «Информационная безопасность» Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики»,
Шелухин Олег Иванович

111024, г. Москва, ул. Авиамоторная, 8а, МТУСИ
Тел. служ.: +7 (495) 957-79-17, email: mtuci@mtuci.ru

«09» 02 2026 г. _____ О.И. Шелухин

Подлинность подписи О.И. Шелухина удостоверяю

Заместитель проректора по научной работе _____ Н.В. Кленов