СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

УДК 681.5.03 DOI: 10.17587/mau.22.227-236

И. А. Каляев, академик РАН, д-р техн. наук, проф., науч. руководитель направления ЮФУ, ikalyaev@sfedu.ru,
Южный федеральный университет, г. Таганрог,
Э. В. Мельник, д-р техн. наук, зав. лаб., evm17@mail.ru,
Федеральный исследовательский центр Южный научный центр Российской академии наук, г. Ростов-на-Дону

Доверенные системы управления

В современных условиях проблема обеспечения безопасности систем с критической миссией приобрела особую актуальность. Причина тому — возросшие возможности несанкционированного воздействия на такие системы через аппаратное и программное обеспечение, а также через коммуникационные сети. Это подтверждается целым рядом аварий, когда оборудование выводилось из строя за счет закладных элементов и вирусов. В настоящее время в Российской Федерации на зарубежных аппаратно-программных платформах построена подавляющая часть систем управления, используемых, в том числе, на стратегических предприятиях и объектах с критической миссией. При этом доля используемых в них зарубежных микроэлектронных компонентов превышает 85 %.

Статья посвящена развитию научных основ и методик оценки степени доверия к системам управления объектов с критической миссией. Показано, что степень доверия к системе управления — это более широкий показатель, чем просто показатели ее надежности и отказоустойчивости, который должен объединить разнородные свидетельства и утверждения, как объективные, основанные на физически и математически обоснованных методах оценки степени их истинности, так и субъективные, основанные на опыте экспертов. В работе предложен метод оценки степени доверия к системе управления объектов с критической миссией, основанный на схеме Шортлиффа (E. Shortliffe), используемой в теории нечеткой логики для оценки степени доверия к некоторой гипотезе на основе разнородных свидетельств и утверждений. Важным преимуществом схемы Шортлиффа является то, что набор свидетельств может расширяться и дополняться (например, на основе вновь полученного опыта), что позволяет уточнять значение коэффициента уверенности.

Предложены методы оценки степени истинности терминальных утверждений различных типов, в том числе таких, которые требуют сочетания как объективных, так и субъективных методов оценки степени их истинности. Использование предложенного метода оценки доверия при формировании национальных стандартов разработки и создания систем управления объектов с критической миссией позволит существенно повысить их функциональную защищенность.

Ключевые слова: *системы управления, доверие, нечеткая логика, объект с критической миссией*

Введение

Летом 2010 г. 1368 центрифуг обогащения урана, используемых в иранской ядерной программе, были выведены из строя без какого-либо внешнего физического воздействия. Это было осуществлено путем внедрения в их системы управления (СУ) компьютерного вируса Stuxnet. Вирус был ориентирован для работы с контроллерами Siemens, используемыми для управления центрифугами.

Суть конструкции вируса была такова, что он перехватывал управление зараженным контроллером и сам начинал отдавать команды, но так, чтобы у операторов сохранялась иллюзия контроля над ситуацией. Получив контроль над СУ центрифуг, вирус начал незаметно менять режимы их работы. Центрифуги резко разгонялись и так же резко тормозили. При этом операторы оставались в неведении о происходящем, поскольку показатели, выводимые на экран, вирус фальсифицировал. В результате в один момент центрифуги в Натанзе (Иран) начали массово выходить из строя, что нанесло очень сильный удар по иранской ядерной программе.

Дальнейшим развитием технологии Stuxnet стал вирус Flame, появившийся в 2012 г., активация которого происходит только в определенных географических зонах.

В настоящее время в Российской Федерации на зарубежных аппаратно-программных платформах построена подавляющая часть СУ, используемых, в том числе, на стратегических предприятиях и объектах с критической миссией. При этом доля используемых в них зарубежных микроэлектронных компонентов превышает 85 %. Все это наталкивает на мысль о том, что приведенные выше технологии вполне могут быть использованы для вывода из строя контроллеров зарубежных компаний, применя-

емых на российских стратегических предприятиях или объектах с критической миссией, что, в частности, подтверждается недавними сообщениями газеты Нью-Йорк Таймс о планируемых МО США кибератаках на энергетическую инфраструктуру России.

С учетом вышеизложенного возникает вопрос, насколько мы можем доверять той или иной СУ, используемой на объектах с критической миссией, и каким образом можно измерить степень доверия к ней?

В настоящее время в РФ существует ряд документов, регламентирующих понятие доверенных аппаратно-программных систем и методов обеспечения доверенности, такие как ГОСТ Р 54581-2011, ГОСТ Р 54582-2011 и ГОСТ Р 54583-2011 "Методы и средства обеспечения безопасности. Основы доверия к безопасности информационных технологий" [1—3], ГОСТ Р ИСО/МЭК 15408-3-2013 "Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий", ГОСТ Р ИСО/МЭК 18045-013 "Методы и средства обеспечения безопасности. Методология оценки безопасности информационных технологий", ГОСТ Р ИСО/МЭК 25010-2015 "Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов" [4—6].

Однако эти документы в основном определяют организационные механизмы обеспечения доверия к информационно-управляющим системам, связанные с контролем их разработки, производства, проведения испытаний и т. п., и в то же время не учитывают многие трудно формализуемые аспекты доверия, такие как субъективные мнения специалистов, накопленный ими опыт применения подобных систем и т. п. Поэтому несмотря на значительное число регламентирующих документов задача развития методологии оценки доверия остается актуальной (например, в ГОСТ Р ИСО/МЭК 15408-3-2013 говорится о целесообразности и возможности включения в данный стандарт альтернативных методов достижения доверия).

Указанные выше проблемы порождают необходимость развития научных основ доверенных СУ и методов оценки степени доверия к ним, которые могут быть использованы при формировании национальных стандартов разработки и создания СУ объектов с критической миссией. Именно этому вопросу и будет посвящена настоящая статья.

Метод оценки степени доверия к системам управления

До недавнего времени качество работы СУ объекта с критической миссией (т. е. способность выполнить возложенную на нее задачу) оценивалась в основном с помощью объективных (количественных) показателей надежности, например таких, как вероятность безотказной работы, гамма-процентная наработка на отказ и т. п. [7—12]. Однако такие объективные показатели не учитывают целый ряд факторов, которые могут существенно влиять на работоспособность СУ при выполнении поставленной перед ней задачи, например, наличие тех или иных не декларируемых возможностей ("закладок") аппаратно-программного обеспечения СУ, способных снизить качество управления формируемого СУ или даже вывести ее из строя, возможностей деструктивных внешних воздействий на систему управления и т. п. Такие возможности очень трудно и даже практически невозможно оценить с помощью каких-либо объективных показателей, т. е. они могут быть оценены только субъективно на основе мнения экспертов. Введение таких субъективных показателей доверия крайне актуально именно сейчас, когда глобализация рынка, а также крайне нестабильная международная обстановка приводят к тому, что уровень доверия к той или иной системе управления должен зависеть не только от объективных показателей, но и, например, от того, на сколько мы доверяем фирмам-производителям электронной компонентной базы и программного обеспечения, используемых при создании данной конкретной СУ, какие у нас отношения со страной, где осуществлялась сборка данной СУ и т. п.

Проблема оценивания качества работы СУ с помощью объективных показателей надежности становится еще более проблематичной в случае, если СУ представляет собой человеко-машинную систему, в контуре управления которой задействован естественный (человеческий) интеллект. Действительно, мы же не можем утверждать, что "вероятность безотказной работы водителя троллейбуса Сидорова составляет 0,999". Мы можем только, например, сказать, что мы доверяем водителю Сидорову больше, чем водителю Петрову, поскольку Сидоров обладает бо́льшим опытом и не злоупотребляет спиртными напитками.

Еще сложнее дело обстоит в случае, если в контуре управления объекта используются

подсистемы, построенные на основе технологий искусственного интеллекта (ИИ). Как правило, в ИИ процесс принятия решения полностью скрыт от внешнего наблюдателя и поэтому труднопредсказуем. Поэтому оценить качество работы СУ, использующей технологии ИИ, с помощью объективных показателей надежности, таких как, например, вероятность безотказной работы, практически не представляется возможным.

Приведенные выше примеры говорят о том, что степень доверия к СУ — это более широкий показатель, чем просто показатели ее надежности, который должен объединять в себе как объективные, так и субъективные факторы качества работы СУ, а также накопленный ранее опыт. Введение данного показателя и стандартизация методик его оценки позволит существенно повысить функциональную защищенность систем управления, используемых на отечественных объектах с критической миссией.

Для представления и обработки разнородных и неопределенных данных и знаний предложены различные формальные модели, например такие, как:

— теория свидетельств Демпстера—Шефера [13];

— баейсовские сети доверия [14, 15];

— теория возможностей [16].

В то же время наиболее распространенным подходом к формализации и обработке разнородных и неопределенных данных и знаний является подход, основанный на нечеткой логике [17, 18]. Предложенная в 1965 г. Лотфи Заде, нечеткая логика является на сегодняшний день одним из самых эффективных методов обработки неполной и неточной информации. Нечеткая логика предлагает использование градаций или степеней принадлежности элемента множеству, находящихся в интервале от 0 до 1, позволяя тем самым отобразить степень уверенности эксперта. Это свойство нечеткой логики позволяет моделировать сомнение эксперта, а также является оптимальным средством оперирования лингвистическими понятиями, содержащими расплывчатость и неполноту информации в своей основе.

В нечеткой логике существует понятие коэффициента уверенности, с помощью которого измеряется степень доверия к некоторой гипотезе (утверждению) на основе имеющихся свидетельств (опыта) [19]. Впервые понятие коэффициента уверенности было введено Шортлиффом (E. Shortliffe) для оценки степени доверия к решению, выдаваемому некоторой экспертной системой [20, 21]. Он же предложил и схему (формулу) для определения значения коэффициента уверенности:

$$KU[H{:}E] = MD[H{:}E] - MND[H{:}E], \qquad (1)$$

где $KU[H{:}E]$ — коэффициент уверенности в гипотезе (утверждении) H с учетом свидетельств (опыта) E; $MD[H{:}E]$ — мера доверия к гипотезе H при заданных свидетельствах E; $MND[H{:}E]$ — мера недоверия к гипотезе H при заданных свидетельствах E.

При этом KU, MD и MND не являются вероятностными мерами. Значение KU изменяется в пределах от −1 до +1, причем −1 соответствует абсолютной лжи, +1 абсолютной истине, а 0 — означает полное незнание.

Значения MD и MND изменяются в пределах от 0 до 1. Использование коэффициента KU позволяет упорядочить выдвигаемые гипотезы по степени их обоснованности.

Понятие коэффициента уверенности может стать синонимом степени доверия, если под гипотезой понимать утверждение вида: "СУ обеспечит оптимальное и безотказное управление объектом O на интервале времени $[t_K, t_T]$", а под свидетельствами понимать различного рода объективные и субъективные данные и знания, подтверждающие или опровергающие данную гипотезу.

Важным преимуществом схемы Шортлифа является то, что набор свидетельств может расширяться и дополняться, (например на основе вновь полученного опыта), что позволяет уточнять значение коэффициента уверенности. В целях такого уточнения Шортлифф предложил формулу для взвешивания различных свидетельств, которая позволяет непосредственно сочетать новые свидетельства со старыми. Она применяется к мерам доверия и недоверия, связанным с каждым свидетельством, и имеет вид

$$MD[H{:}E_1, E_2] - MD[H{:}E_1] + \\ + MD[H{:}E_2] \cdot (1 - MD[H{:}E_1]); \qquad (2)$$

$$MND[H{:}E_1, E_2] - MND[H{:}E_1] + \\ + MND[H{:}E_2] \cdot (1 - MND[H{:}E_1]), \qquad (3)$$

где запятая между E_1 и E_2 говорит о том, что свидетельство E_2 следует за свидетельством E_1.

Смысл данных формул заключается в том, что эффект от свидетельства E_2 на гипотезу при заданном свидетельстве E_1 заключается в смещении значения MD в сторону полной определенности на расстояние, зависящее от свидетельства E_2.

Формула уточнения имеет три важных свойства:

1) она симметрична в том смысле, что порядок E_1 и E_2 несущественен;

2) по мере накопления подкрепляющих свидетельств значение *MD* (или *MND*) смещается к определенности;

3) наличие слагаемого *MND* позволяет не "потерять" на фоне многих положительных свидетельств существенные отрицательные свидетельства (о важности этого говорится, например, в ГОСТ 15467-79).

Схема Шортлиффа допускает также возможность того, что свидетельства, как и данные, могут быть ненадежными. Это позволяет описывать более широкий класс ситуаций. Каждое свидетельство снабжается так называемым коэффициентом ослабления, принимающим значения от 0 до 1 и показывающим надежность (доверие) правила. Кроме того, вводится так называемый порог уверенности (*PU*) — число от 0 до 1. Если *KU* некоторой гипотезы (утверждения) меньше, чем *PU*, то такой гипотезой можно пренебречь.

Схема Шортлиффа может лечь в основу методики определения степени доверия к СУ объектов с критической миссией. Для этого необходимо определить правила формирования и обработки свидетельств, подтверждающих либо опровергающих основную гипотезу о том, что "СУ обеспечит оптимальное и безотказное управление объектом *O* на интервале $[t_O, t_K]$".

Под свидетельством будем понимать набор некоторых утверждений Y_i, объединенных логическими условиями вида

$$P_1 \wedge P_2 = \min(P_1, P_2); \quad (4)$$

$$P_1 \vee P_2 = \max(P_1, P_2); \quad (5)$$

$$\bar{P}_1 = (1 - P_1), \quad (6)$$

где P_1 — степень истинности утверждения Y_1; P_2 — степень истинности утверждения Y_2.

Если степень истинности утверждения Y_i известна, то такое утверждение будем называть *терминальным*. Иными словами, *терминальное утверждение* — это утверждение, степень истинности которого подтверждена либо объективными данными, либо субъективными (экспертными) оценками и заключениями.

Терминальные утверждения могут быть двух типов: объективные и субъективные. *Объективные* терминальные утверждения — это утверждения, степень истинности которых определена (измерена) на основе объективных (физически и математически обоснованных) фактов (параметров). Примером объективного терминального утверждения может служить утверждение вида: "вероятность безотказной работы аппаратной платформы СУ на интервале времени $[t_H, t_K]$ равна 0,95".

Очевидно, что степень истинности *P* последнего утверждения может быть рассчитана с помощью классических методов оценки надежности на основе объективных фактов о применяемой элементной базе.

Субъективные терминальные утверждения — это утверждения, степень истинности которых может быть установлена только на основе субъективных оценок экспертов, например: "аппаратная платформа СУ, изготовленная в стране X, не содержит недекларируемые возможности (закладки), которые могут повлиять на ее работоспособность".

Очевидно, что степень истинности последнего утверждения может быть получена только экспертным путем, на основе знаний и опыта экспертов.

Если степень истинности P_i некоторого утверждения Y_i неизвестна, то данное утверждение должно рассматриваться как гипотеза и, соответственно, обосновываться свидетельствами более низкого уровня. Так, утверждение типа: "аппаратная платформа СУ отработает без сбоев и отказов на интервале времени $[t_H, t_K]$" не является терминальным, поскольку степень его истинности зависит от некоторого множества свидетельств более низкого уровня, например:

Свидетельство

Если

- *утверждение*: "вероятность безотказной работы аппаратной платформы СУ на интервале $[t_H, t_K]$ равна 0,95" (степень истинности) и
- *утверждение*: "аппаратная платформа СУ, изготовленная в стране X, не содержит не декларируемые возможности (закладки), которые могут повлиять на ее работоспособность (степень истинности P_2),

то *гипотеза (утверждение)*

- "аппаратная платформа СУ отрабатывает без сбоев и отказов на интервале $[t_H, t_K]$" верна (степень истинности $P_1 \wedge P_2$), т. е. в данном случае утверждение о том, что "аппаратная платформа СУ отработает без сбоев и отказов на интервале $[t_H, t_K]$" становится гипотезой, которая должна подтверждаться или опровергаться утверждениями более низкого уровня.

Таким образом, для того чтобы определить значение коэффициента уверенности (степень доверия) к какой-либо сложной гипотезе (например, гипотезе "СУ обеспечит оптимальное и безотказное управление и безотказное управление объектом O на интервале $[t_H, t_K]$ с") необходимо построить дерево свидетельств (рис. 1), вершиной которого является основная (главная) гипотеза, т. е. гипотеза, коэффициент уверенности в которой надо определить.

Основная (главная) гипотеза должна подкрепляться набором свидетельств первого уровня, каждое из которых содержит набор логически связанных утверждений. Если все утверждения свидетельства являются терминальными (т. е. для них известны степени истинности), то на их основе рассчитывается степень истинности данного свидетельства. Если же свидетельство содержит нетерминальные утверждения, то последние рассматриваются в качестве гипотез, которые должны подкрепляться свидетельствами более низкого (второго) уровня.

Процесс формирования дерева свидетельств продолжается до тех пор, пока все свидетельства i-го уровня не будут содержать только терминальные утверждения.

Далее с помощью схемы Шортлиффа рассчитываются степени истинности всех свидетельств, начиная со свидетельств самого нижнего уровня, причем степень истинности свидетельства i-го уровня принимается в качестве степени истинности соответствующего ему утверждения, используемого в свидетельстве $(i-1)$-го уровня, и так далее, вплоть до главной гипотезы, расположенной в вершине дерева.

В результате выполнения такой процедуры будет получено значение коэффициента уверенности KU главной гипотезы, которое можно принять в качестве степени доверия к данной СУ.

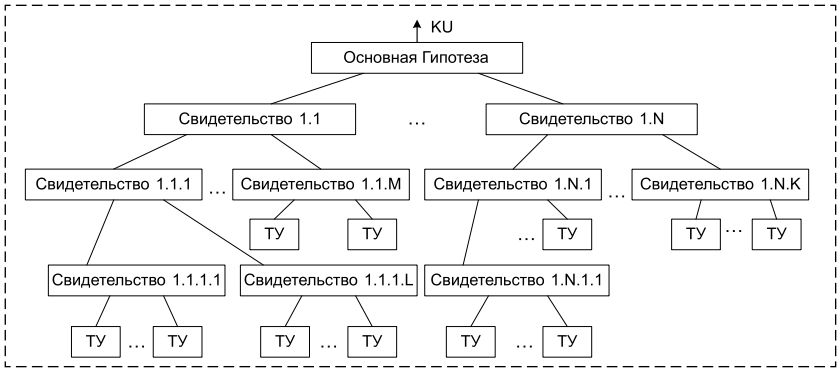

Рис. 1. Дерево свидетельств (ТУ — терминальные утверждения)
Fig. 1. Evidence tree

Рассмотрим пример формирования дерева свидетельств и определения на его основе значения коэффициента уверенности в главной гипотезе о гарантированной работоспособности СУ на интервале времени $[t_H, t_K]$.

Допустим, что степень доверия к СУ зависит от трех составляющих:

- доверия к используемой аппаратной платформе;
- доверия к используемому программному обеспечению;
- доверия к используемому алгоритмическому обеспечению.

Поэтому свидетельства, подтверждающие основную гипотезу, могут звучать следующим образом.

Свидетельство 1

Если

- *утверждение 1.1*: аппаратная платформа СУ отработает без сбоев и отказов на интервале $[t_H, t_K]$ (степень истинности P_{11})

и

- *утверждение 1.2*: ПО отработает без сбоев и отказов на интервале $[t_H, t_K]$ (степень истинности P_{12})

и

- *утверждение 1.3*: алгоритмическое обеспечение обеспечивает оптимальное управление объектом O на интервале $[t_H, t_K]$ (степень истинности P_{13}),

то *основная гипотеза* верна со степенью истинности $P_1 = P_{11} \wedge P_{12} \wedge P_{13}$.

Свидетельство 2

Если

- *утверждение 2.1*: фирме-производителю СУ можно доверять (степень истинности P_{21})

или

- *утверждение 2.2*: стране, в которой произведена данная СУ, можно доверять (степень истинности P_{22}),

то *основная гипотеза* верна (степень истинности $P_2 = P_{21} \vee P_{22}$).

Например, **Свидетельство 1** (Гипотеза 1) может быть подкреплено следующими свидетельствами.

Поскольку ни одно из утверждений, входящих в свидетельства 1 и 2, не является терминальным, то они рассматриваются как гипотезы и поэтому должны быть подкреплены свидетельствами следующего уровня.

Свидетельство 1.1

Если

- *утверждение 1.1.1*: "Вероятность безотказной работы аппаратной платформы СУ равна 0,95" (степень истинности P_{111})

и

- *утверждение 1.1.2*: аппаратная платформа СУ не содержит недекларируемых возможностей (закладок), способных повлиять на ее работоспособность (степень истинности P_{112}),

то *Гипотеза 1* верна (степень истинности $P_{11} = P_{111} \wedge P_{112}$).

При этом утверждение 1.1 является терминальным, поскольку значение P_{111} может быть посчитано с помощью объективных методик. В то же время утверждение 1.2 не является терминальным, и, соответственно, оно принимается в качестве Гипотезы 1.2, которая должна быть подкреплена свидетельствами следующего уровня, которые могут иметь следующий вид:

Свидетельство 1.2.1

- *утверждение 1.2.1.1*: элементная база аппаратной платформы СУ не содержит не декларируемых возможностей (закладок) (степень истинности P_{1211}),

и

- *утверждение 1.2.1.2*: в составе аппаратной платформы отсутствуют элементы, способные нарушить ее работоспособность (степень истинности P_{1212}),

и

- *утверждение 1.2.1.3*: печатная плата аппаратной платформы не содержит технологически непредусмотренных слоев (степень истинности P_{1213});

то *Гипотеза 1.2* верна (степень истинности $P_{121} = P_{1211} \wedge P_{1212} \wedge P_{1213}$).

Утверждения 1.2.1.1, 1.2.1.2 и 1.2.1.3 могут быть как терминальными, если заключения P_{1211}, P_{1212} и P_{1213} могут быть получены экспертным путем на основе данных об изготовителях электронной компонентной базы, печатных плат и т. п., либо не терминальными, если их подтверждение требует дополнительных, более детальных свидетельств.

Процесс формирования дерева свидетельств продолжается аналогичным образом сверху вниз от основной гипотезы вплоть до того, когда все входящие в него свидетельства будут подкреплены терминальными утверждениями.

При этом следует отметить, что сформированное таким образом дерево свидетельств может по мере надобности дополняться новыми свидетельствами, подтверждающими или опровергающими основную гипотезу, которые могут быть получены, например, в процессе эксплуатации СУ.

Терминальные утверждения

Как показано выше, в основе предложенной методики оценки степени доверия к СУ лежат терминальные утверждения двух типов — объективные и субъективные.

Объективные утверждения — это утверждения, степень истинности которых может быть определена с помощью объективных (физически и математически обоснованных) методов, например таких, как классические методы теории вероятности или теории надежности [9—12]. В отличие от объективных терминальных утверждений степень истинности субъективных терминальных утверждений не может быть оценена с помощью объективных (физически и математически обоснованных) методов, а может быть получена только на основе обобщения опыта экспертов. Процесс оценки степени истинности субъективного утверждения усложняется тем, что требует высокого уровня знаний и опыта эксперта, следовательно, требуется привлечения различных экспертов для получения объективной оценки. Для определения интегрального значения степени истинности субъективного терминального утверждения могут быть использованы методы обобщения мнения различных экспертов, разработанные в рамках теории экспертных систем [19].

В то же время следует отметить, что существуют терминальные утверждения, которые одновременно требуют применения как объективных, так и субъективных методов оценки степени их истинности.

Например, как показано выше, одним из базовых терминальных утверждений, используемых при оценке степени доверия к СУ, является утверждение следующего вида:

Утверждение 1. Алгоритм управления, используемый в СУ, гарантирует, что относительная погрешность качества управление объектом O на интервале времени $[t_H, t_K]$ по сравнению с оптимальным будет не больше величины C.

Очевидно, что, с одной стороны, это объективное утверждение, степень истинности которого зависит от объективных характеристик выбранного алгоритма управления и объекта O. С другой стороны, количественно оценить

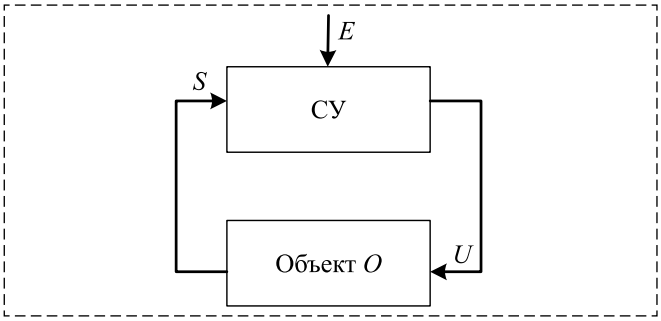

Рис. 2. Обобщенная схема управления
Fig. 2. General control scheme

значение относительной погрешности качества управления для всех возможных случаев, как правило, не представляется возможным, и поэтому необходимо использовать какие-то приблизительные, в том числе экспертные, оценки. Здесь можно предложить следующий подход.

Будем считать, что состояния объекта O описываются набором (кортежем) параметров $\boldsymbol{S}$, а сам объект функционирует в некоторой внешней среде E, состояние которой описывается набором (кортежем) параметров $\boldsymbol{E}$. С помощью СУ формируются управления $\boldsymbol{U}$ (рис. 2), которые переводят объект O в новое состояние, причем зависимость изменения состояния объекта O от управления $\boldsymbol{U}$ описывается некоторой системой дифференциальных управлений вида

$$\frac{d\boldsymbol{S}}{dt} = F(\boldsymbol{S}, \boldsymbol{E}, \boldsymbol{U}), \tag{7}$$

Под оптимальным управлением объектом O понимается такое управления $\boldsymbol{U}$, которое обеспечивает перевод объекта O из начального состояния $\boldsymbol{S}^0$ в целевое состояние $\boldsymbol{S}^K$ и при этом минимизирует некоторый функционал качества

$$H = \int_{t_T}^{t_K} R(\boldsymbol{S}, \boldsymbol{E}, \boldsymbol{U})dt \tag{8}$$

при граничных условиях $\boldsymbol{S}(t_0) = \boldsymbol{S}^0$ и $\boldsymbol{S}(t_K) = \boldsymbol{S}^K$, где t_0 — начальный момент времени; t_K — момент времени достижения объектом O целевого состояния $\boldsymbol{S}^K$.

Тогда **Утверждение 1** можно переформулировать следующим образом: "Алгоритм управления, используемый в СУ, гарантирует выполнение условия

$$\frac{H_\text{р} - H_\text{о}}{H_\text{р}} \leqslant C, \tag{9}$$

где $H_\text{о}$ — оптимальное (минимальное) значение функционала H при переходе объекта O из начального состояния $\boldsymbol{S}^0$ в целевое состояние $\boldsymbol{S}^K$; $H_\text{р}$ — реальное значение функционала H, получаемое при переходе объекта O из состояния $\boldsymbol{S}^0$ в состояние $\boldsymbol{S}^K$; C — допустимая относительная погрешность — отклонение — реального значения функционала $H_\text{р}$ от оптимального $H_\text{р}$ при переходе объекта O из состояния $\boldsymbol{S}^0$ в состояние $\boldsymbol{S}^K$.

Очевидно, что степень истинности последнего утверждения будет зависеть от отклонения значения $H_\text{р}$ критерия качества управления, вырабатываемого СУ, от оптимального $H_\text{о}$.

Если, например, алгоритм управления, используемый в СУ, построен на базе методов оптимального управления [22—24], то можно считать, что

$$H_\text{р} = H_\text{о}, \tag{10}$$

т. е. условие (9) будет гарантированно выполняться, и, следовательно, степень истинности P **Утверждения 1** будет равной 1. Однако, как известно, алгоритмы оптимального управления могут быть использованы далеко не всегда вследствие их огромной вычислительной сложности.

Поэтому, как правило, инженеры вынуждены применять иные подходы, основанные, в частности, на методах динамического программирования, эвристических методах и т. п. Все эти подходы позволяют значительно сократить пространство перебора при формировании управления объектом O, но, в то же время, не гарантируют его оптимальности, т. е. не гарантируют того, что $H_\text{р} = H_\text{о}$. При этом возникает вопрос: каким образом можно оценить степень истинности **Утверждения 1** при использовании такого рода алгоритмов управления?

Для этого оценим вероятность P того, что СУ обеспечит перевод объекта O из начального состояния $\boldsymbol{S}^0$ в состояние $\boldsymbol{S}^K$ при выполнении условии (9).

Перепишем условие (9) в следующем виде:

$$1 - \frac{H_\text{о}}{H_\text{р}} \leqslant C \tag{11}$$

или

$$\frac{H_\text{о}}{H_\text{р}} \geqslant 1 - C. \tag{12}$$

Учитывая соотношение (8), последнее выражение можно представить в следующем виде:

$$\frac{\int_{t_K}^{t_K} R(\boldsymbol{S}_\text{о}, \boldsymbol{E}_\text{о}, \boldsymbol{U}_\text{о})dt}{\int_{t_K}^{t_K} R(\boldsymbol{S}_\text{р}, \boldsymbol{E}_\text{р}, \boldsymbol{U}_\text{р})dt} \geqslant 1 - C, \tag{13}$$

где $\boldsymbol{S}_{\text{о}}$, $\boldsymbol{E}_{\text{о}}$, $\boldsymbol{U}_{\text{о}}$ — функции состояния объекта, среды и управления, соответствующие оптимальному переходу объекта O из состояния $\boldsymbol{S}_H$ в состояние $\boldsymbol{S}_K$; $\boldsymbol{S}_{\text{р}}$, $\boldsymbol{E}_{\text{р}}$, $\boldsymbol{U}_{\text{р}}$ — функции состояния объекта, среды и управления, соответствующие реальному переходу объекта O из состояния $\boldsymbol{S}_H$ в состояние $\boldsymbol{S}_K$.

Если в СУ используется принцип терминального управления, когда текущее управление $\boldsymbol{U}^i$ формируется через дискретные промежутки времени Δ_t на основе данных о текущем состоянии $\boldsymbol{S}^i$ объекта O и состоянии $\boldsymbol{S}^i$ среды E, то выражение (13) можно переписать в следующем виде:

$$\frac{\sum\limits_{i=1}^{M} R(\boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{U}^i_{\text{о}})\Delta t}{\sum\limits_{i=1}^{M} R(\boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{U}^i_{\text{р}})\Delta t} \geqslant 1 - C, \qquad (14)$$

где M — число шагов формирования терминального управления $\boldsymbol{U}^T$ при переходе объекта O из состояния $\boldsymbol{S}_0$ в состояние $\boldsymbol{S}_K$ (иначе говоря, число шагов дискретизации времени $(t_0 - t_k)$; $\boldsymbol{S}^i$, $\boldsymbol{E}^i$ — текущие значения состояния объекта O и среды E на i-м шаге дискретизации; $\boldsymbol{U}^i_{\text{о}}$ — оптимальное управление объектом O на i-м шаге дискретизации; $\boldsymbol{U}^i_{\text{р}}$ — реальное значение управления $\boldsymbol{U}$, формируемое СУ на i-м шаге дискретизации.

Не уменьшая строгости неравенства (14), его можно заменить следующим выражением:

$$\frac{1}{B} \geqslant 1 - C, \qquad (15)$$

где $B = \max \dfrac{R(\boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{U}^i_{\text{р}})}{R(\boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{U}^i_{\text{о}})}$ — максимальное значение относительной "погрешности" качества управления, формируемого СУ, по сравнению с оптимальным.

Тогда значение вероятности P того, что используемый в СУ алгоритм управления обеспечит перевод объекта управления из состояния $\boldsymbol{S}_T$ в состояние $\boldsymbol{S}_K$ с относительным качеством не хуже, чем C, можно оценить с помощью следующего выражения:

$$P = \begin{cases} 1, \text{ если } \dfrac{1}{B} \geqslant 1 - C, \\ \dfrac{1}{(1-C)B}, \text{ если } \dfrac{1}{B} < 1 - C. \end{cases} \qquad (16)$$

Соответственно, последнее выражение может быть использовано для оценки степени истинности **Утверждения 1**. Правда, при этом остается открытым вопрос: каким образом можно оценить значение B для конкретного алгоритма управления объектом O, используемого в СУ? По-видимому, такая оценка как раз и должна быть получена экспертным путем, т. е. является субъективным параметром.

Сложнее дело обстоит в том случае, когда формализованного алгоритма управления объектом O просто не существует. В этом случае решение задачи управления объектом O может строиться на основе методов искусственного интеллекта, например таких, как машинное обучение и глубокие нейронные сети. При этом нейронная сеть (НС) должна быть предварительно обучена таким образом, чтобы формировать терминальное управление $\boldsymbol{U}^i$ объектом O в зависимости от значений параметров $\boldsymbol{S}^i$ и $\boldsymbol{E}^i$ текущего состояния объекта и окружающей его среды, а также значения целевого состояния $\boldsymbol{S}_K$ объекта O. Очевидно, что при таком подходе значение максимальной относительной погрешности B управления, формируемого с помощью НС, зависит от методики обучения НС.

Под пространством обучения НС будем понимать пространство $\boldsymbol{W}$, координатами w_i (i = = 1, 2, ..., N) которого являются параметры обучающей выборки. В нашем случае координатами пространства $\boldsymbol{W}$ будут являться параметры $\boldsymbol{S}^i$, $\boldsymbol{E}^i$, $\boldsymbol{S}^K$.

Процедура обучения НС, используемой для управления объектом O, заключается в предъявлении ей в качестве обучающих выборок некоторого подмножества $\boldsymbol{W}_y = \left\langle \boldsymbol{S}^i_y, \boldsymbol{E}^i_y, \boldsymbol{S}^K_y \right\rangle$ точек пространства $\boldsymbol{W}(\boldsymbol{W}_y \subseteq \boldsymbol{W})$ и соответствующих им значений оптимального управления $\boldsymbol{U}^i_{\text{о}}$.

После того, как НС таким образом обучена, она может быть использована для формирования терминального управления объектом O. Для этого на входы НС необходимо подать текущие значения параметров $\boldsymbol{S}^i$ и $\boldsymbol{E}$ состояния объекта управления и среды, а также целевого состояния $\boldsymbol{S}^K$. Тогда на выходах НС будет формироваться вектор $\boldsymbol{U}^i_{\text{р}}$ текущего управления.

При этом, поскольку в общем случае точка w_i с координатами $\left\langle \boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{S}^K \right\rangle$ может не попадать в подмножества $\boldsymbol{W}_y$ выборок, на которых проводилось обучение НС, то в качестве текущего управления $\boldsymbol{U}^i_{\text{р}}$ НС выдает управление $\boldsymbol{U}^i_{\text{о}}$, приписанное ближайшей к точке w_i точке $w^y_i = \left\langle \boldsymbol{S}^i_y, \boldsymbol{E}^i_y, \boldsymbol{S}^K_y \right\rangle$ подмножества $\boldsymbol{W}_y$.

Исходя их этих соображений можно сделать вывод, что при использовании НС в контуре управления объектом O значение максималь-

ной относительной "погрешности" B качества терминального управления можно оценить как

$$B = \max \frac{R(\boldsymbol{S}^i, \boldsymbol{E}^i, \boldsymbol{U}^i_{\text{o}})}{R(\boldsymbol{S}^i_y, \boldsymbol{E}^i_y, \boldsymbol{U}^i_{\text{o}})}, \qquad (17)$$

где $\langle \boldsymbol{S}^i, \boldsymbol{E}^i \rangle$ — координаты некоторой произвольной точки w_i пространства $\boldsymbol{W}$; $\langle \boldsymbol{S}^i_y, \boldsymbol{E}^i_y \rangle$ — координаты ближайшей к ней точки w^i_y подпространства $\boldsymbol{W}_y \subseteq \boldsymbol{W}$; $\boldsymbol{U}^i_{\text{o}}$ — оптимальное управление, приписанное точке $w^y_i \in \boldsymbol{W}_y$.

Заключение

В работе предложен метод оценки степени доверия к СУ объектом с критической миссией, позволяющий объединить разнородные свидетельства и утверждения, как объективные, основанные на физически и математически обоснованных методах оценки степени их истинности, так и субъективные, основанные на опыте экспертов. В качестве математического аппарата, позволяющего оценить степень доверия к некоторой СУ на основе разнородных свидетельств и утверждений, предложено использовать схему Шортлиффа (E. Shortliffe), применяемую в теории нечеткой логики.

В основе предложенного метода оценки степени доверия к СУ лежат терминальные утверждения двух типов — объективные и субъективные. Объективные терминальные утверждения — это утверждения, степень истинности которых может быть определена с помощью объективных (физически и математически обоснованных) методов. В отличие от объективных терминальных утверждений степень истинности субъективных терминальных утверждений может быть оценена только на основе обобщения опыта экспертов. Предложены методы оценки степени истинности терминальных утверждений различных типов, в том числе таких, которые требуют сочетания как объективных, так и субъективных методов оценки степени их истинности.

Использование предложенного метода оценки доверия при формировании национальных стандартов разработки и создания СУ объектов с критической миссией позволит существенно повысить их функциональную защищенность.

Список литературы

1. **ГОСТ Р 54581-2011.** Информационная технология. Методы и средства обеспечения безопасности. Основы доверия к безопасности ИТ. Часть 1. Обзор и основы. М.: ФГУП "СТАНДАРТИНФОРМ", 2012. 23 с.

2. **ГОСТ Р 54582-2011.** Информационная технология. Методы и средства обеспечения безопасности. Основы доверия к безопасности информационных технологий. Часть 2. Методы доверия. М.: ФГУП "СТАНДАРТИНФОРМ", 2013. 47 с.

3. **ГОСТ Р 54583-2011.** Информационная технология. Методы и средства обеспечения безопасности. Основы доверия к безопасности информационных технологий. Часть 3. Анализ методов доверия. М.: ФГУП "СТАНДАРТИНФОРМ", 2013. 50 с.

4. **ГОСТ Р ИСО/МЭК 15408-3-2013.** Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 3. Компоненты доверия к безопасности. М: ФГУП "СТАНДАРТИНФОРМ", 2014. 267 с.

5. **ГОСТ Р ИСО/МЭК 18045-2013.** Информационная технология. Методы и средства обеспечения безопасности. Методология оценки безопасности информационных технологий. М.: ФГУП "СТАНДАРТИНФОРМ", 2014. 244 с.

6. **ГОСТ Р ИСО/МЭК 25010-2015.** Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов. М.: ФГУП "СТАНДАРТИНФОРМ", 2015. 30 с.

7. **ГОСТ 27.002-2015.** Надежность в технике. Основные понятия и определения. М.: ФГУП "СТАНДАРТИНФОРМ", 2016. 24 с.

8. **Труханов В. М.** Надежность в технике. М.: Машиностроение, 1999. 597 с.

9. **Викторова В. С., Степанянц А. С.** Модели и методы расчета надежности технических систем. М.: ЛЕНАНД, 2014. 256 с.

10. **Дорохов А. Н., Керножицкий В. А., Миронов А. Н., Шестопалова О. Л.** Обеспечение надежности сложных технических систем. М.: Лань, 2011. 352 с.

11. **Майерс Г.** Надежность программного обеспечения. М.: Мир, 2008. 360 с.

12. **Таейр Е., Липов М., Нельсое Э.** Надежность программного обеспечения. М.: ИЛ, 2008. 323с.

13. **Shafer G.** A Mathematical Theory of Evidence. Princeton University Press, 1976.

14. **Finn V. Jensen.** Bayesian Networks and Decision Graphs. Springer, New York, 2001. P. 268.

15. **Kevin B. Korb.** Bayesian Artificial Intelligence. CRC, London, 2004. P. 391.

16. **Пытьев Ю. П.** Возможность. Элементы теории и применения. М.: Эдиториал УРСС, 2000. 192с.

17. **Заде Л.** Понятие лингвистичекой переменной и его применение к принятию приближенных решений. М.: Мир, 1976, 166 с.

18. **Новак В., Перфильева И., Мочкрож И.** Математические принципы нечеткой логики. М.: Физматлит, 2006, 352 с.

19. **Джексон П.** Введение в экспертные системы. М.: Издательский дом "Вильямс", 2001, 624 с.

20. **Buchanan B. G., Shortliffe E. H.** Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, 1984.

21. **Моросанова Н.А, Соловьев С. Ю.** Формальные свойства схемы Шортлиффа // Управление большими системами. 2012. Т. 36. С. 5—38.

22. **Гамкрелидзе Р. В.** Основы оптимального управления. Тбилиси: Изд-во ТбГУ, 1977. 264 с.

23. **Иванов В. А., Медведев В. С.** Математические основы теории оптимального и логического управления: учеб. пособ. М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. 599 с.

24. **Алексеев В. М., Тихомиров В. М., Фомин С. В.** Оптимальное управление. М.: Физматлит, 2005. 408 с.

Trusted Control Systems

I. A. Kalyaev, ikalyaev@sfedu.ru,
Southern Federal University, Taganrog, 347900, Russian Federation,
E. V. Melnik, evm17@mail.ru,
Federal Research Centre the Southern Scientific Centre of the Russian Academy of Sciences,
Rostov-on-Don, 344006, Russian Federation

Corresponding author: **Melnik Eduard V.**, Ph.D., Head of the laboratory of Information Technology and Control Processes, Federal Research Centre the Southern Scientic Centre of the Russian Academy of Sciences, Rostov-on-Don, 344006, Russian Federation, e-mail: *evm17@mail.ru*

Accepted on January 25, 2021

Abstract

Nowadays, the problem of ensuring security of systems with a critical mission has become particularly relevant. An increased opportunity for unauthorized exposure on such systems via hardware, software and communication networks is the main reason to discuss this problem. It is confirmed by a plenty of accidents when equipment is out of order by means of malicious embedded elements and viruses. Currently, in the Russian Federation the majority of control systems are based on foreign hardware and software platforms, including strategic enterprises and objects with a critical mission. Herewith, the proportion of foreign microelectronic components in such systems is more than 85 %. The article is devoted to the development of scientific basis and techniques of the assurance assessment to control systems of objects with a critical mission. It was shown, that assurance assessment to a control system is a broader index than its reliability and fault tolerance. Such index must integrate various evidences and approvals, which can be objective, based on physical and mathematical assurance assessment methods, as well as they can be subjective, based on the experts experience. A method of assurance assessment to a control system of objects with a critical mission, based on Shortliffe's scheme, was proposed in this paper. The Shortliffe's scheme is used in the theories of fuzzy logic for assurance assessment to a hypothesis on the basis of various evidences and statements. An important advantage of a Shortliffe's scheme is the set of evidences, which can be broadened and augmented (for instance, on the basis of obtained experience). It allows us to clarify a certainty factor. The assessment methods of truth degree of terminal statements of various types, including those, which require the combination of objective and subjective methods of their truth degree assessment, are proposed. The proposed assurance assessment method for national development and creation standards of control systems of objects with a critical mission allows to significantly increase their functional security.

Keywords*: control systems, assurance, fuzzy logic, objects with a critical mission*

For citation:

Kalyaev I. A., Melnik E. V. Trusted Control Systems, *Mekhatronika, Avtomatizatsiya, Upravlenie,* 2021, vol. 22, no. 5, pp. 227—236.

DOI: 10.17587/mau.22.227-236

References

1. **GOST R 54581-2011.** Information technology. Security techniques. A framework for IT security assurance. Part 1. Overview and framework, Moscow, FGUP "STANDARTINFORM", 2012, 23 p. (in Russian).

2. **GOST R 54582-2011.** Information technology. Security techniques. A framework for IT security assurance. Part 2. Assurance methods, Moscow, FGUP "STANDARTINFORM", 2013, 47 p. (in Russian).

3. **GOST R 54583-2011.** Information technology. Security techniques. A framework for IT security assurance. Part 3. Analysis of assurance methods, Moscow, FGUP "STANDARTINFORM", 2012, 50 p. (in Russian).

4. **GOST R ISO/MEK 15408-3-2013.** Information technology. Security techniques. Evaluation criteria for IT security. Part 3. Security assurance requirements, Moscow, FGUP "STANDARTINFORM", 2014, 267 p. (in Russian).

5. **GOST R ISO/MEK 18045-2013.** Information technology — Security techniques — Methodology for IT security evaluation, Moscow, FGUP "STANDARTINFORM", 2014, 244 p. (in Russian).

6. **GOST R ISO/MEK 25010-2015.** Information technology. Systems and software engineering. Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models, Moscow, FGUP "STANDARTINFORM", 2015, 30 p. (in Russian).

7. **GOST R 27.002-2015.** Dependability in technics. Terms and definitions, Moscow, FGUP "STANDARTINFORM", 2016, 24 p. (in Russian).

8. **Truhanov V. M.** Reliability in technique, Moscow, Mashinostroenie, 1999, 597 p. (in Russian).

9. **Viktorova V. S., Stepanyanc A. S.** Models and methods for calculating the reliability of technical systems, Moscow, LENAND, 2014, 256 p. (in Russian).

10. **Dorohov A. N., Kernozhickij V. A., Mironov A. N., Shestopalova O. L.** Ensuring the reliability of complex technical systems, Moscow, Lan', 2011, 352 p. (in Russian).

11. **Majers G.** Software reliability, Moscow, Mir, 2008, 360 p. (in Russian).

12. **Taejr E., Lipov M., Nel'soe E.** Software reliability, Moscow, IL, 2008, 323p. (in Russian).

13. **Shafer G. A.** Mathematical Theory of Evidence, Princeton University Press, 1976.

14. **Finn V. Jensen.** Bayesian Networks and Decision Graphs, Springer, New York, 2001, 268 p.

15. **Kevin B. Korb.** Bayesian Artificial Intelligence. CRC, London, 2004. P. 391.

16. **Pyt'ev Yu. P.** Opportunity. Elements of theory and application, Moscow, Editorial URSS, 2000, 192 p. (in Russian).

17. **Zade L.** The concept of a linguistic variable and its application to approximate decision-making, Moscow, Mir, 1976,166 p. (in Russian).

18. **Novak V., Perfil'eva I., Mochkrozh I.** Mathematical principles of fuzzy logic, Moscow, Fizmatlit, 2006, 352 p. (in Russian).

19. **Dzhekson P.** Introduction to expert systems, Moscow, Publishing house of "Vil'yams", 2001, 624 p. (in Russian).

20. **Buchanan B. G., Shortliffe E. H.** Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley, Reading, 1984.

21. **Morosanova N. A, Solov'ev S. Yu.** Formal properties of the Shortliffe scheme, *Upravlenie Bol'shimi Sistemami*, 2012, vol. 36, pp. 5—38 (in Russian).

22. **Gamkrelidze R. V.** Fundamentals of optimal control, Tbilisi, Publishing house of TbGU, 1977, 264 p. (in Russian).

23. **Ivanov V. A, Medvedev V. S.** Mathematical foundations of the theory of optimal and logical control, Moscow, Publishing house of MGTU im. N. E. Baumana, 2011, 599 p. (in Russian).

24. **Alekseev V. M., Tihomirov V. M., Fomin S. V.** Optimal control, Moscow, Fizmatlit, 2005, 408 p. (in Russian).