ПРЕДУПРЕЖДЕНИЕ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Бутусов И.В[1], Романов А.А.[2]

Цель статьи: *поддержка процессов предупреждения инцидентов информационной безопасности в условиях высокой неопределенности.*

Метод: *методы математической (теоретической) информатики и теории нечетких множеств.*

Полученный результат. *Инцидент информационной безопасности, включая инцидент компьютерный, рассматривается как факт нарушения или прекращения функционирования автоматизированной информационной системы и (или) нарушения хранимой и обрабатываемой в этой системе информации, в том числе вызванный компьютерной атакой. Информационные описания представлены в виде структурированных данных о признаках компьютерных атак. Структурированные данные являются конечными последовательностями строк символов формального языка. В качестве метрики измерения расстояния между строками символов из определенного алфавита предложено использовать редакционное расстояние Дамерау-Левенштейна. Обоснована возможность представления семантики информационных описаний признаков атак в виде нечетких множеств. Определены пороги (степени) разделения нечетких информационных описаний. Оценено влияние семантической определенности информационных описаний признаков (степеней размытости нечетких информационных описаний) на принятие решений об их идентичности (похожести). Показано, что семантическая составляющая информационных описаний признаков компьютерных атак предполагает наличие некоторой семантической метрики (для ее измерения и интерпретации) и которая, как правило, формально плохо определена, неоднозначно интерпретирована и характеризуется неопределенностью типа нечеткости, наличием семантической информации и невозможностью непосредственного применения вероятностной меры для определения степеней сходства входных и хранимых информационных описаний признаков. Предложен подход к идентификации нечетких информационных описаний признаков компьютерных атак и применения методов разделения элементов опорных множеств, на которых определены эти информационные описания. Показано, что результаты процедуры идентификации нечетких информационных описаний признаков компьютерных атак зависят от степеней разделения опорных множеств и от показателей семантической неопределенности этих описаний.*

Ключевые слова: *признаки компьютерной атаки, информационное описание, семантика, неопределенность, нечеткое множество, функция принадлежности, степень разделения, сведения, алфавит, строки символов, редакционное расстояние.*

DOI:10.21681/2311-3456-2020-05-45-51

Введение

Под инцидентом информационной безопасности (ИБ), включая инцидент компьютерный, будем понимать факт нарушения или прекращения функционирования автоматизированной информационной системы (АИС) и (или) нарушения хранимой и обрабатываемой в этой системе информации, в том числе вызванный компьютерной атакой (КА, проект ФЗ № 47571-7 «О безопасности критической информационной инфраструктуры Российской Федерации»[3] [1]). Согласно[4,5] и [1] КА – это целенаправленное несанкционированное воздействие на информацию, на защищаемый ресурс АИС или получение несанкционированного доступа к ним с применением программных или программно-аппаратных средств. Предупреждение инцидентов ИБ связано с оценкой рисков нарушения критически важных свойств (целостность, доступность и конфиденциальность) защищаемых ресурсов АИС с различными категориями важности в результате воздействия КА [2]. В этих целях соответствующая система мониторинга RF должна распознавать (идентифицировать) возник-

1 Бутусов Игорь Викторович, начальник научно-исследовательского управления АО «Концерн «Системпром», Москва, Россия. E mail: butusigor@yandex.ru

2 Романов Александр Анатольевич, доктор технических наук, главный специалист АО «Концерн «Системпром», Москва, Россия. E mail: ralexhome@yandex.ru

3 Проект ФЗ № 47571-7 О безопасности критической информационной инфраструктуры Российской Федерации.

4 ГОСТ Р 51275-2006 Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения (п.3.11).

5 Р 50.1.056-2005 Техническая защита информации. Основные термины и определения (п.3.2.8).

шую угрозу безопасности информации (БИ) в динамике функционирования АИС.

КА можно представить в виде совокупности информационных описаний признаков, которую необходимо идентифицировать в процессах технологического цикла, реализуемого АИС при попытках нарушителя вывода ее из строя или снижения эффективности применения.

Совокупность информационных описаний признаков КА формируется, декомпозируется и систематизируется на основе априорных знаний об опыте эксплуатации АИС и классификации угроз безопасности информации (БИ), а затем уточняется по апостериорной информации. Основными средствами и источниками информации для идентификации информационных описаний признаков КА и предупреждения возникновения инцидентов ИБ в АИС являются датчики средств противодействия. Средства противодействия КА совместно со средствами мониторинга опасности воздействия КА осуществляют сбор данных от датчиков для формирования совокупности информационных описаний признаков атак и оценки возможного нарушения устойчивости функционирования АИС и степени опасности [3, 4].

Если КА реализует известную угрозу БИ, то можно применить методы анализа (сравнения) информационных описаний характерных признаков атаки (хранимые информационные описания признаков КА – сигнатуры атак), используемых для их обнаружения.

Методы анализа сигнатур предназначены для идентификации известных атак и основаны на контроле программ и данных в системе и эталонной сверке последовательности символов и событий в сети с базой данных сигнатур атак.

Недостатком методов анализа сигнатур является невозможность обнаружения новых (модифицированных атак) без строгой формализации ключевых слов сетевого трафика и обновления базы данных сигнатур атак.

Если выявляется аномальное поведение системы отличное от типичного (информационные описания признаков КА отсутствуют в базе сигнатур), то на основании этого факта можно принять решение о возможном наличии атаки (обнаружена неизвестная атака). Обнаружение аномальных отклонений в сети осуществляется по признакам КА — как правило используют синтаксические сигнатуры, взятые непосредственно из тела атаки.

Так как информационные описания признаков реализуемой КА и хранимые информационные описания признаков КА представляют собой синтаксические конструкции, то для их сравнения имеет смысл использовать метрику, представляющую собой функцию расстояния между двумя словами (строками), позволяющую оценить степень их сходства в данном контексте. Строгое математическое определение метрики включает в себя необходимость соответствия условию неравенства треугольника (X – множество слов, p – метрика) [4]:

$$(\, p(x,y) \le p(x,z) + p(z,y)\,, \quad x,y,z \in X$$

Между тем, в большинстве случаев под метрикой подразумевается более общее понятие, не требующее выполнения такого условия, но которое можно также рассматривать как расстояние. К числу наиболее известных из таких метрик относятся расстояния Хемминга, Левенштейна и Дамерау-Левенштейна. Следует отметить, что расстояние Хемминга является метрикой только на множестве слов одинаковой длины, что сильно ограничивает область его применения [5].

1. Основные задачи, решаемые в процессе предупреждения инцидентов ИБ

- обоснование возможности представления семантики информационных описаний признаков КА в виде нечетких множеств (НМ).
- определение порогов (степеней) разделения нечетких информационных описаний.
- оценка влияния семантической определенности информационных описаний признаков (степеней размытости нечетких информационных описаний) на принятие решений об их идентичности (похожести).

2. Содержательная постановка задачи предупреждения инцидентов ИБ

Идентификации информационных описаний признаков КА: поиск по входному информационному описанию похожих на него хранимых информационных описаний в условиях возможных искажений, неполного состава и частичного нарушения порядка следования данных во входном и (или) хранимых описаниях (высокая неопределенность).

Под информационным описанием будем понимать структурированные данные о признаках КА. Данные представлены значениями (конечными последовательностями строк символов формального языка) соответствующих реквизитов (атрибутов). Реквизиты (атрибуты) могут быть сгруппированы по некоторому семантическому признаку, например, по классам КА.

3. Особенности задачи: 1) обработка и анализ строк (подстрок) символов (модель идентификации информационных описаний, в том числе и в условиях неполных и (или) искаженных сведений о них); 2) произвольное количество ошибок при анализе символьных строк (произвольное количество операций редактирования – вставка, замена, удаление, перестановка); 3) клавиатурная модель ошибок с определением средневзвешенного редакционного расстояния между строками; 4) наличие весов ошибок, зависящих от символа, над которым производится операция (особенности формального языка, определяемого выбранным алфавитом); вес ошибки зависит от местоположения символа в слове; учет веса ошибки с клавиатурно близким символом; 5) множество операций над подстроками (замена, вставка, удаление подстрок) для обработки ошибок транскрибирования; 6) наличие нечеткой меры (из-за учета весов ошибок); неопределенность типа нечеткости; нечеткость, размытость соотнесения входного информационного описания к выделяемому подмножеству хранимых; переход от полной идентичности (эквивалентности) входного описания хранимым

к полной не идентичности происходит постепенно (нечеткость = неточность + неопределенность).

4. Формальный язык [6]. Определен некоторый алфавит $\aleph$ и $\aleph^*$ – множество определенных над этим алфавитом строк $\xi_k \in \aleph^*, k = \overline{1, N^{str}}$. Строка $\xi_k = (a_1,...,a_n)$-- конечная последовательность символов $a_i \in \aleph$. . Алфавит $\aleph$ включает в себя символ «пробел» («_»), а $\aleph^*$ – пустую строку ε. .

Если $str_1 \in \aleph*$ – строка длины m, а $str_2 \in \aleph*$ – строка длины n, то их объединение $str_1 \cup str_2$, обозначаемое как $str_1 str_2$, представляет собой строку длины $m+n$, в которой первые m символов составляют строку, совпадающую со строкой str_1, а последние n символов составляют строку, совпадающую со строкой str_2. Если $str_3 = str_1 str_2$, где $str_3, str_1, str_2 \in \aleph^*$, то строка str_1 – префикс строки str_3, а строка – str_2 суффикс строки str_3. Построение префиксных деревьев существенно сокращает время поиска строк символов (слов).

Префиксное дерево имеет много полезных свойств для решения задачи идентификации информационных описаний.

Первое. При обходе дерева любой префикс любого имени проходится только один раз. Это не только экономит время просмотра набора уникальных имен. Можно построить такой алгоритм, который будет использовать состояние таблицы динамического программирования в предыдущей вершине для вычисления состояния в текущей вершине. Другими словами, состояние таблицы динамического программирования для вычисления оптимального редакционного расстояния всегда соответствует префиксу какого-либо слова из заданного алфавита.

Второе. Если префикс какого-либо имени уже содержит число ошибок редактирования больше, чем допустимое количество k, то все выравнивания всех имен, начинающихся с этого префикса (и находящихся в поддереве этого префикса), можно не проводить.

Третье. Можно ввести весовой порог на оптимально-взвешенное редакционное расстояние. Тогда, если префикс какого-либо имени весит больше, чем порог, то все выравнивания всех имен, начинающихся с этого префикса (и находящихся в поддереве этого префикса) можно не проводить.

Четвертое. Применение префиксного дерева делает тривиальным поиск сокращенных признаков, так как сокращение есть не что иное, как префикс и нужно просто достать все продолжения этого префикса из поддерева этого префикса.

Произвольное подмножество множества $\aleph^*$ называется формальным языком Λ.

Если строка *str* может быть получена из строки str' с помощью одного или нескольких правил вывода, то $str' \Rightarrow str$.

5. Операции по переводу одной строки в другую Расстояние Левенштейна (также редакционное расстояние или дистанция редактирования) [5] в теории информации и компьютерной лингвистике – это мера разницы двух последовательностей символов (строк) относительно минимального количества операций вставки, удаления и замены, необходимых для перевода одной строки в другую.

Две строки str_1 и str_2 называются похожими, если редакционное расстояние между ними не превышает заданного допустимого количества ошибок k.

Редакционное предписание – последовательность действий, необходимых для получения из первой строки второй кратчайшим образом. Обычно действия обозначаются так: D (англ. delete) – удалить, I (англ. insert) – вставить, R (англ. replace) – заменить, M (англ. match) – совпадение.

Каждая операция редактирования имеет свой вес: I – p, D – q, R – r, T – v. Тогда взвешенное редакционное расстояние σ между двумя строками str_1 и str_2 определяется как суммарный вес минимального числа редакционных операций, необходимых для преобразования str_1 в str_2 (совпадения не засчитываются).

Если $\sigma^{П}$ порог на взвешенное редакционное расстояние, то две строки str_1 и str_2 похожи, если редакционное расстояние между ними не превышает допустимого количества ошибок k, а взвешенное редакционное расстояние не превышает $\sigma^{П}$.

Степень похожести s_1 и s_2 определяется величиной σ. Чем меньше σ, тем str_1 более похожа на str_2.

Расстояние Дамерау-Левенштейна – это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и перестановки соседних символов, необходимых для перевода одной строки в другую. Является модификацией расстояния Левенштейна, отличается от него добавлением операции перестановки.

6. Семантика информационного описания [7]. Семантика всякого сведения о признаке предполагает наличие следующих четырех величин: опорного множества X описаний, семантического указателя x одного из описаний X, т.е. $x \in X$, подмножества δ признаков из X, т.е. $\delta \subset X$ и семантической достоверности p, которая характеризует достоверность (определенность) выполнения главного условия $x \in \delta$.

Если x_0 – точка X и δ – некоторое непустое подмножество X, которое определяется свойством δ, то факт принадлежности $x_0 \in \delta$ или истинное высказывание «точка x_0 из X обладает свойством δ» записывается в виде в виде одноместного предиката $\delta(x_0), x_0 \in X$ ..

Если P– решетка достоверностей, X – опорное множество, $x_0 \in X$. и $\delta \subset X$, и если про точку x_0 известно с семантической достоверностью $p \in P$, что $x_0 \in \delta$, , то имеется сведение о точке $x_0 \in X$., которое записывается в форме триады $(p)\ \delta(x_0)$ которая интерпретируется как принадлежность $x_0 \in \delta$, с достоверностью p или высказывание «точка x_0 из X обладает свойством δ с семантической достоверностью p».

Если более точно, то согласно основным положениям математической информатики [8] под информационным описанием $\Delta(x)$ произвольного признака x понимается структурированная совокупность сведений вида $(p)\delta(x)$:

$$\Delta(x) = \{(p_i)\delta_i\}(x), i = \overline{1, N_\delta}. \qquad (1)$$

В (1) сведения $\{(p_i)\ \delta_i\}(x)$ интерпретируются как «признак x из X характеризуется свойством δ_i с семан-

тической достоверностью p_i», δ_i – подмножество признаков из X, характеризующихся одноименным свойством $\delta_i \subset X$, а p_i – семантическая достоверность того, что $x \in \delta_i$.

7. Нечеткие множества и семантика информационного описания. Рассмотрим информационное описание $\Delta(x_i)=\{(p)\delta_p; 0<p\leq 1\}(x_i)$, где δ_p – семейство подмножеств таких, что $x_i \in \delta_p, 0 < p \leq 1$, с семантической достоверностью p. Для δ_p можно подобрать информационные описания $\Delta(x_j)=\{(p')\delta_{p'}; 0 < p' \leq 1\}(x_j)$, $i \neq j$, такие, что $\delta_p = \{x_j \mid x_j \in \delta_{p'}, p' \geq p\}$. Следовательно, δ_p и $\delta_{p'}$ – множества уровней p и p' нечеткого свойства (множества) $\tilde{\delta} : \tilde{\delta} = \{x, p_{\tilde{\delta}}(x)\}$, $p_{\tilde{\delta}}(x) = \mu_{\tilde{\delta}}(x) : X \to [0,1]$ – функция принадлежности по Л. Заде. Следовательно, семантика информационного описания $\Delta(x_i)$ и $\Delta(x_j)$ в соответствии с математической информатикой (данные об признаках в смысле (1)) формально представлена нечетким свойством $\tilde{\delta}$.

Таким образом, семантика информационного описания признака может быть представлена НМ (формализация через множества уровней) и, следовательно, может обрабатываться с использованием известных операций над НМ.

Отметим, что семантическая составляющая информационных описаний признаков предполагает наличие некоторой семантической метрики (для ее измерения и интерпретации) и, как правило, формально плохо определена, неоднозначно интерпретирована. Таким образом, информационные описания признаков характеризуются неопределенностью типа нечеткости, наличием семантической информации и, как следствие, невозможностью непосредственного применения вероятностной меры для определения степеней сходства информационных описаний различных признаков.

Напомним, что если некоторый признак $x_0 \in X$ обладает свойством $\tilde{\delta}$, но лишь в частичной мере – $0 < \mu_{\tilde{\delta}}(x_0) < 1$, то внутренняя неопределенность, двусмысленность признака x_0 по отношению к свойству $\tilde{\delta}$ проявляется в том, что он, хотя и в разной степени, принадлежит сразу двум противоположным классам: классу признаков «обладающих свойством $\tilde{\delta}$» и классу признаков, «не обладающих свойством $\tilde{\delta}$». Эта двусмысленность признака x_0 по отношению к свойству $\tilde{\delta}$ максимальна, когда степени принадлежности признака x_0 к классам «$\tilde{\delta}$» и «не $\tilde{\delta}$» равны, т.е. $\mu_{\tilde{\delta}}(x_0) = 0{,}5$ и $\mu_{не\tilde{\delta}}(x_0) = 1 - \mu_{\tilde{\delta}}(x_0) = 0{,}5$. И наоборот, двусмысленность признака минимальна, когда признак принадлежит только к одному из этих классов, т.е. либо $\mu_{\tilde{\delta}}(x_0) = 1$, $\mu_{не\tilde{\delta}}(x_0) = 0$, либо $\mu_{\tilde{\delta}}(x_0) = 0$, $\mu_{не\tilde{\delta}}(x_0) = 1$.

Таким образом, показатель семантической неопределенности нечеткого свойства $\tilde{\delta}$ можно представить в виде показателя размытости (меры энтропии) соответствующего НМ [8].

В частности, показатель семантической неопределенности нечетких свойств $\tilde{\delta}$ можно определить по аналогии с шенноновской энтропией теории информации в виде логарифмической энтропии:

$$d(\tilde{A}) = k\sum_{j=1}^{N} S(\mu_{\tilde{A}}(x_j)),$$

где S - функция Шеннона $S(y) = -y\ln y - (1-y)\ln(1-y)$ и k- положительная константа.

Свертка векторных нечетких отношений предпочтения имеет вид:

$$\mu_P(x,y) = \sum_{j=1}^{m} \lambda_j \mu_j(x,y),$$

где λ_j - коэффициент важности для нечеткого отношения предпочтения с функцией принадлежности μ_j.

Содержательная постановка задачи идентификации информационных описаний предполагает применение методов разделения элементов опорных множеств, на которых определены нечеткие свойства в зависимости степеней разделения и от показателей семантической неопределенности нечетких свойств.

Показатель семантической неопределенности нечетких свойств был рассмотрен ранее. Рассмотрим понятие степени разделения нечетких информационных описаний [9].

8. Степень разделения нечетких информационных описаний. Пусть на опорном множестве признаков X определены нечеткие свойства $\tilde{\delta}_i = \{(x, \mu_{\tilde{\delta}}(x))\}$.

Поскольку любые два нечетких свойства $\tilde{\delta}$ и $\tilde{\delta}_j$, $i \neq j$, ограничены степенями принадлежности $\sup \mu_{\tilde{\delta}_i}(x)$ и $\sup \mu_{\tilde{\delta}_j}(x)$ в точках α_i и α_j соответственно, то их пересечение $\tilde{\delta}_i \cap \tilde{\delta}_j$ принимает максимальное значение $\sup_x \mu_{\tilde{\delta}_i \cap \tilde{\delta}_j}(x)$ в γ. Наивысшая степень разделения нечетких свойств $\tilde{\delta}_i$ и $\tilde{\delta}_j$ достигается в точке $\gamma = 1 - \sup_x \mu_{\tilde{\delta}_i \cap \tilde{\delta}_j}(x)$.

Для разделения элементов, отвечающих свойствам $\tilde{\delta}_i$ или $\tilde{\delta}_j$ можно использовать понятие порога разделения. В рассматриваемом случае порог разделения Pr ограничен условием

$$\Pr < \max_x \min[\mu_{\tilde{\delta}_i}(x), \mu_{\tilde{\delta}_j}(x)] = \sup_x \mu_{\tilde{\delta}_i \cap \tilde{\delta}_j}(x).$$

Таким образом, для выбранного порога разделения Pr области M_i и M_j разделения сравниваемых элементов относительно предпочтений любых двух свойств (далее – области разделения) $\tilde{\delta}_i$ или $\tilde{\delta}_j$, $i \neq j$, определяются нечеткими подмножествами уровня Pr:

$$M_i = \{x \mid \mu_{\tilde{\delta}_i}(x) \geq \max_x \min[\mu_{\tilde{\delta}_i}(x), \mu_{\tilde{\delta}_j}(x)]\} \text{ и}$$
$$M_j = \{x \mid \mu_{\tilde{\delta}_j}(x) \geq \max_x \min[\mu_{\tilde{\delta}_i}(x), \mu_{\tilde{\delta}_j}(x)]\}$$
для всех $x \in M_i$.

Представленная выше модель обобщается на $N_{\tilde{\delta}}$ нечетких свойств $\tilde{\delta}_i$, $i=\overline{1,N_{\tilde{\delta}}}$.

Если $\tilde{\delta}_i=\{(x,\mu_{\tilde{\delta}_i}(x))\}$, $i=\overline{1,N_{\tilde{\delta}}}$, – ограниченные выпуклые нечеткие свойства, то нечеткие подмножества $\tilde{\delta}_1\cap\tilde{\delta}_2,\tilde{\delta}_1\cap\tilde{\delta}_3,...,\tilde{\delta}_1\cap\tilde{\delta}_Q,\tilde{\delta}_2\cap\tilde{\delta}_3,...,\tilde{\delta}_{Q-1}\cap\tilde{\delta}_m$ будут также выпуклыми и ограниченными. Таким образом, можно определить области разделения $\underline{M}_1,\underline{M}_2,...,\underline{M}_{N_{\tilde{\delta}}}$ и $\overline{M}_1,\overline{M}_2,...,\overline{M}_{N_{\tilde{\delta}}}$ элементов с использованием нижнего и верхнего порогов разделения:

$$\underline{\Pr}<\min_{i,j}\max_{x}\min[\mu_{\tilde{\delta}_i}(x),\mu_{\tilde{\delta}_j}(x)] \text{ и}$$

$$\overline{\Pr}<\max_{i,j}\max_{x}\min[\mu_{\tilde{\delta}_i}(x),\mu_{\tilde{\delta}_j}(x)].$$

Области разделения $\underline{M}_i$ и $\overline{M}_i$ признаков будут опять нечеткими уровневыми множествами, определяемыми соотношениями

$$\underline{M}_i=\{x\,|\,\mu_{\tilde{\delta}_i}(x)\ge\min_{i,j}\max_{x}\min[\mu_{\tilde{\delta}_i}(x),\mu_{\tilde{\delta}_j}(x)]\} \text{ и}$$

$$\overline{M}_i=\{x\,|\,\mu_{\tilde{\delta}_i}(x)\ge\max_{i,j}\max_{x}\min[\mu_{\tilde{\delta}_i}(x),\mu_{\tilde{\delta}_j}(x)]\}$$

для всех $x\in\overline{M}_i,\underline{M}_i$.

Если $\mu_{\tilde{\delta}_i}(x)=\underline{\Pr}$, то $(\underline{\Pr})\delta_{i_{\underline{\Pr}}}(x)$ – сведение о том, что признак x обладает свойством $\tilde{\delta}_1$ с семантической достоверностью $\underline{\Pr}$. Здесь $\delta_{i_{\underline{\Pr}}}$ – множество уровня $\underline{\Pr}$ нечеткого свойства $\tilde{\delta}_1$. Сказанное справедливо и для $\underline{\Pr}$.

Пусть $\Delta(x_{so})=\{(p_i^{so})\delta_i^{so}\}(x_{so})$ – хранимые информационные описания, $i=\overline{1,Q},so=\overline{1,N_{so}}$, Q – число нечетких свойств в информационном описании, N_{so} – число хранимых информационных описаний, и $\Delta(x_0)=\{(p_i^0)(\delta_i^0)\}(x_0)$ – информационное описание входных признаков. Здесь δ_i^0 – значение i-го свойства во входном информационном описании x_0, δ_i^{so} – значение i-го свойства хранимого информационного описания x_{so}.

Формальное решение задачи идентификации КА предполагает

1) формирование семантических метрик для измерения и интерпретации нечетких свойств входного и хранимых информационного описания;

2) определение значения

$$(\mu_{\tilde{\delta}_i}(x_0,x_{so}))(\delta_i^0\sim\delta_i^{so})(x_{so}), \qquad (2)$$

где $\mu_{\tilde{\delta}_i}(x_0,x_{so})$ – степень сходства хранимого x_{so} и входного x_0 описаний относительно значений свойства δ_i хранимого δ_i^{so} и входного δ_i^0 информационного описания;

3) определение верхнего $\overline{\Pr}$ и нижнего $\underline{\Pr}$ порогов разделения хранимых информационных описаний с учетом степеней сходства $\mu_{\tilde{\delta}_i}(x_0,x_{so})$ с входным информационным описанием относительно нечетких свойств δ_i;

4) классифицирование хранимых информационных описаний по степеням сходства с входным информационным описанием:

хранимые информационные описания эквивалентные входному информационному описанию

$$M_э=\{x_i\,|\,\mu_P(x_0,x_i)\ge\Pr\}, \qquad (3)$$

$$M_п=\{x_i\,|\,\underline{\Pr}\le\mu_P(x_0,x_i)\le\overline{\Pr}\}, \qquad (4)$$

$$M_н=\{x_i\,|\,\mu_P(x_0,x_i)\le\underline{\Pr}\}. \qquad (5)$$

5) по значениям (2) определить предпочтения

$$\mu_P(x_0,x_i)=f((^0p_i^{so})(\delta_i^0\sim\delta_i^{so})(x_{so})). \qquad (6)$$

Определим эквивалентные, идентичные, похожие и непохожие хранимые ИО по отношению входного ИО.

Если $\overline{\Pr}=1$ и $M_{\sim}=\cap\overline{M}_i$ непустое множество, то $x\in M_{\sim}$ эквивалентные информационные описания.

Выводы

1. Показано, что семантическая составляющая информационных описаний признаков компьютерных атак предполагает наличие некоторой семантической метрики (для ее измерения и интерпретации) и, как правило, формально плохо определена, неоднозначно интерпретирована и характеризуются неопределенностью типа нечеткости, наличием семантической информации и невозможностью непосредственного применения вероятностной меры для определения степеней сходства информационных описаний признаков.

2. Предложен подход к идентификации нечетких информационных описаний признаков компьютерных атак и применения методов разделения элементов опорных множеств, на которых определены эти информационные описания.

3. Показано, что результаты процедуры идентификации нечетких информационных описаний признаков компьютерных атак зависят от степеней разделения опорных множеств и от показателей семантической неопределенности этих описаний.

DOI: 10.21681/2311-3456-2020-05-45-51

Литература:

1. Атагимова Э.И., Макаренко Г.И., Федичев В.А. Информационная безопасность. Терминологический словарь в определениях действующего законодательства/Федеральное бюджетное учреждение «Научный центр правовой информации при Министерстве юстиции Российской Федерации». Москва. 2017. (Издание 3-е). 448 с.
2. Butusov I.V., Romanov A.A. Methodology of security assessment automated systems as object critical information infrastructure // Вопросы кибербезопасности №1(24). 2018. С. 2-10 DOI: 10.21681/2311-3456-2018-1-2-10
3. Климов С.М., Сычев М.П., Астрахов А.В. Противодействие компьютерным атакам. Методические основы: Электронное учебное издание. М.: МГТУ имени Н.Э. Баумана, 2013. 108 с . // URL: http://wwwcdl.bmstu.ru/iu10/comp-atak-metod.htm
4. Марков А.С. Техническая защита информации. Курс лекций / учебное пособие. М. АИСНТ. 2020. 234 с., ISBN 978-5-9500695-3-1

   Новейшее учебное пособие, программа которого согласована со ФСТЭК России. Материал издания соответствует программе курса повышения квалификации «002.Техническая защита информации. Способы и средства защиты информации от несанкционированного доступа»
5. Нечёткий поиск в тексте и словаре. 2011. https://habr.com/ru/post/114997/

   В работе представлены результаты сравнительного тестирования качества и производительности семи алгоритмов нечеткого поиска (также известного как поиск по сходству или fuzzy string search), которые являются основой систем проверки орфографии и полноценных поисковых систем вроде Google или Yandex. На основе этого анализа для сравнения сходства двух символьных строк выбрана метрика Расстояние Дамерау-Левенштейна.
6. Рейуорд-Смит В. Дж. Теория формальных языков. Вводный курс: пер. с англ. М.: Радио и связь, 1988. 128 с.

   В области определения над некоторым алфавитом строк и построения префиксных деревьев пока нового ничего не придумано.
7. Чечкин А.В. Математическая информатика. М.: Наука, 1991. 416с.

   Это единственная известная нам отечественная монография (зарубежные аналоги нам не известны) в которой с позиций теоретической информатики формализуются такие понятия, как «сведения», «данные» и «информация» о точке, с использованием которых формируются информационные описания.
8. Аверкин А.Н., Батыршин И.З. и др. Нечеткие множества в моделях управления и искусственного интеллекта. М.: Наука. Гл. ред. физ-мат. лит., 1986. 312 с.

   Наиболее известная монография, в которой впервые в отечественной литературе изложена общая точка зрения на состояние работ в области нечетких множеств и их применений для решения задач управления и искусственного интеллекта
9. Нечеткие множества и теория возможностей. Последние достижения. Сборник научных статей / Под ред. Р.Р. Ягера.–М.: Радио и связь, 1986.– 408 с.

   В сборнике решена научная задача разделения элементов опорных множеств, на которых определены предпочтения в виде нечетких множеств (известная как задача разделения на торговые зоны). Имеется ряд интерпретаций этой задачи. В нашем случае предложен подход к идентификации нечетких информационных описаний признаков компьютерных атак и применению методов разделения элементов опорных множеств, на которых определены эти информационные описания.

PREVENTION OF INFORMATION SECURITY INCIDENTS IN AUTOMATED INFORMATION SYSTEM

Butusov I.V[6].,Romanov A.A[7].

The purpose of the article is *to support the processes of preventing information security incidents in conditions of high uncertainty.*

Method: *methods of mathematical (theoretical) computer science and fuzzy set theory.*

Result: *an information security Incident, including a computer incident, is considered as a violation or termination of the functioning of an automated information system and (or) a violation of information stored and processed in this system, including those caused by a computer attack. Information descriptions are presented in the form of structured data about signs of computer attacks. Structured data is the final sequence of strings of symbols in a formal language. The Damerau-Levenstein editorial rule is proposed as a metric for measuring the distance between strings of characters from a particular alphabet. The possibility of presenting the semantics of information descriptions of attack features in the form of fuzzy sets is proved. Thresholds (degrees) of separation of fuzzy information descriptions*

6 Igor V. Butusov, Head of Research Department JSC “concern SYSTEMPROM”, Moscow, Russia. E-mail: butusigor@yandex.ru

7 Aleksandr A. Romanov, Dr.Sc. (In Tech.), Chief specialist JSC “concern SYSTEMPROM”, Moscow, Russia. E-mail: ralexhome@yandex.ru

are defined. The influence of semantic certainty of information descriptions of features (degrees of blurring of fuzzy information descriptions) on the decision-making about their identity (similarity) is evaluated. It is shown that the semantic component of information descriptions of signs of computer attacks presupposes the presence of some semantic metric (for its measurement and interpretation), which, as a rule, is formally poorly defined, ambiguously interpreted and characterized by uncertainty of the type of fuzziness, the presence of semantic information and the inability to directly apply a probabilistic measure to determine the degree of similarity of input and stored information descriptions of signs. An approach is proposed to identify fuzzy information descriptions of computer attacks and to apply methods for separating elements of reference sets on which these information descriptions are defined. It is shown that the results of the procedure for identifying fuzzy information descriptions of computer attacks depend on the degree of separation of the reference sets and on the indicators of semantic uncertainty of these descriptions.

Keywords: *signs of a computer attack, information description, semantics, uncertainty, fuzzy set, membership function, degree of separation, information, alphabet, character strings, editorial distance.*

References:

1. Atagimova E`.I., Makarenko G.I., Fedichev V.A. Informatcionnaia bezopasnost`. Terminologicheskii` slovar` v opredeleniiakh dei`stvuiushchego zakonodatel`stva/Federal`noe biudzhetnoe uchrezhdenie «Nauchny`i` centr pravovoi` informatcii pri Ministerstve iustitcii Rossii`skoi` Federatcii». Moskva. 2017. (Izdanie 3-e). 448 s.
2. Butusov I.V., Romanov A.A. Methodology of security assessment automated systems as object critical information infrastructure // Voprosy` kiberbezopasnosti №1(24). 2018. S. 2-10 DOI: 10.21681/2311-3456-2018-1-2-10
3. Climov S.M., Sy`chev M.P., Astrahov A.V. Protivodei`stvie komp`iuterny`m atakam. Metodicheskie osnovy`: E`lektronnoe uchebnoe izdanie. M.: MGTU imeni N.E`. Baumana, 2013. 108 s . // URL: http://wwwcdl.bmstu.ru/iu10/comp-atak-metod.htm
4. Markov A.S. Tekhnicheskaia zashchita informatcii. Kurs lektcii` / uchebnoe posobie. M. AISNT. 2020. 234 s., ISBN 978-5-9500695-3-1
5. Nechyotkii` poisk v tekste i slovare. 2011. https://habr.com/ru/post/114997/
6. Rei`uord-Smith V. Dzh. Teoriia formal`ny`kh iazy`kov. Vvodny`i` kurs: per. s angl. M.: Radio i sviaz`, 1988. 128 s.
7. Chechkin A.V. Matematicheskaia informatika. M.: Nauka, 1991. 416s.
8. Averkin A.N., Baty`rshin I.Z. i dr. Nechetkie mnozhestva v modeliakh upravleniia i iskusstvennogo intellekta. M.: Nauka. Gl. red. fiz-mat. lit., 1986. 312 s.
9. Nechetkie mnozhestva i teoriia vozmozhnostei`. Poslednie dostizheniia. Sbornik nauchny`kh statei` / Pod red. R.R. Iagera.–M.: Radio i sviaz`, 1986.– 408 s.