УДК 004.056

А.А. Петров[1]

[1]*к.т.н, доцент, ВНУ им. В. Даля, Луганск*

МЕТОД ОЦЕНКИ ВЕРОЯТНОСТИ ВОЗНИКНОВЕНИЯ АНОМАЛЬНЫХ СОБЫТИЙ В КОМПЬЮТЕРНОЙ СЕТИ, ОСНОВЫВАЮЩИЙСЯ НА СИСТЕМАХ НЕЧЕТКИХ МНОЖЕСТВ

В статье предложен метод оценки вероятности возникновения аномальных событий в компьютерной сети, основывающийся на системах нечетких множеств.
Ключевые слова: комьютерная сеть, нечеткое множество, защита информации, аномальное событие.

Введение

В условиях информационного противоборства для принятия решения о варианте реагирования на аномальные события в сети необходимы не только регулярный контроль, оперативный анализ, но и *обобщение* данных от различных обнаружителей о подозрительной активности, численное *оценивание* вероятности того, что эта активность является атакой, что позволяет повысить степень достоверности установления факта атаки, снизить риск принятия нерациональных решений.

Объединение нечеткого логического вывода и экспертных оценок является одним из перспективных подходов к организации систем динамического мониторинга обнаружения атак, способствующих повышению надежности защиты информации [1, 2].

Теория *нечетких множеств* может быть использована как средство сбора и обработки нечетной информации, представленной экспертом, особенно те аспекты теории, которые связаны с лингвистической неопределенностью, часто возникающей при работе с экспертами на естественном языке. Под лингвистической неопределенностью подразумеваются качественные оценки естественного языка для логического вывода, принятия решений.

Важными преимуществами моделей реальных систем, построенных на основе нечеткой математики, являются большая гибкость и адекватность реальному миру, а также, по сравнению с традиционными моделями, более быстрое получение окончательного результата через специфическое построение и простоту используемых нечетких операций [1].

Несмотря на известную аналогию с методами теории вероятности, существенное отличие метода теории нечетких множеств состоит в том, что неопределенность связана не со случайностью, а с имеющимися неточностями и размытостями.

Теории лингвистических переменных и приближенных рассуждений опираются на понятие нечеткого множества, систему операций над нечеткими множествами и методы построения *функций принадлежности* [1, 2].

Одним из основных понятий, используемых в лингвистических моделях, является понятие *лингвистической переменной* (ЛП), значениями которой являются не числа, а слова или предложения.

Лингвистическая переменная «число событий информационной безопасности (ИБ) на пути распространения атаки» может принимать значения: мало, средне, выше среднего, много, очень много и т.д. Эти термы − лингвистические значения переменной. Множество допустимых значений лингвистической переменной называется терм-множеством. ЛП задается набором из пяти компонентов [3]:

$$\text{ЛП} = \langle \text{A}, \text{T}, \text{X}, \text{Q}, \text{W} \rangle, \tag{1}$$

где A − имя лингвистической переменной («число событий ИБ»);

T − множество термов (значений) ЛП, которое есть наименования *нечеткой переменной* (НП);

X − область, на которой определены значения лингвистической переменной, определяется экспертом на основе анализа путей распространения атак;

Q − описывает операции по порождению производных значений лингвистической переменной на основе тех, что входят в терм-множество; с помощью правил из Q можно расширить терм-множество.

Каждому значению лингвистической переменной A соответствует нечеткое множество X_a, являющееся подмножеством X.

Компонент W образует набор семантических правил, с помощью которых происходит отображение значения ЛП в нечеткие множества X_a и выполняются обратные преобразования. Эти правила обеспечивают формализацию качественных утверждений экспертов при формировании проблемной области в памяти интеллектуального средства.

В основе понятия ЛП лежит термин нечеткая переменная, которая означает нечеткое множество (НМ), которому присвоено некоторое название. НП определяется кортежем

$$\text{НП} = \langle \text{A}, \text{X}, \widetilde{\text{A}} \rangle, \tag{2}$$

где A − идентификатор НП,

X = {x} − область определения,

$\widetilde{\text{A}}$ − нечеткое множество на X, которое задает ограничения на набор числовых значений НП $\widetilde{\text{A}}$.

Нечетким множеством $\widetilde{\text{A}}$ на множестве $\text{X}(\widetilde{\text{A}} \subseteq \text{X})$ называется совокупность пар

$$\widetilde{\text{A}} = (\text{x}, \mu_{\widetilde{\text{A}}}(\text{x})) = \int_{\text{x}} \mu_{\widetilde{\text{A}}}(\text{x}) \mid \text{x}, \text{x} \in \text{X}, \mu_{\widetilde{\text{A}}}(\text{x}) \in [0,1], \tag{3}$$

где интервал обозначает операцию объединения одноточечных нечетких множеств.

Функция $\mu_{\widetilde{\text{A}}}(\text{x}) : \text{X} \to \text{R}$, отображающая универсальное множество X в пространство R, $\text{R} \in [0,1]$, называется функцией принадлежности нечеткого множества

$\widetilde{A}$. Значение $\mu_{\widetilde{A}}$ называется степенью принадлежности x нечеткому множеству $\widetilde{A}$. Функция принадлежности выражает субъективную возможность наличия у элемента x свойств, позволяющих отнести его к множеству $\widetilde{A}$. Носителем нечеткого множества $\widetilde{A}$ называется множество элементов $x \in X$ универсального множества, которые имеют ненулевую степень принадлежности

$$\mu_{\widetilde{A}}(x) > 0 : \sup A = \left\{x : x \in X, \mu_{\widetilde{A}}(x) > 0\right\} \tag{4}$$

Нечеткое множество $\widetilde{A}$ называют нормальным, если верхняя граница его функции принадлежности равна единице:

$$\sup \mu_{\widetilde{A}}(x) = 1, \qquad x \in X. \tag{5}$$

Рассмотрим нечеткое множество $\widetilde{A}$, отвечающее нечеткому понятию «среднее число сетевых событий ИБ для данного пути распространения атаки». Носителем данного нечеткого множества является конечное множество, элементы которого представляют собой значения числа индикаторов $\{1,2,...x_i,...x_I\}$, нечеткое множество может иметь вид

$$\widetilde{A} = \{0,1 \mid 1 : 0,7 \mid 2; ...1 \mid x_{i-1}; 0,3 \mid x_{i+1}; ...0,1 \mid x_I\}. \tag{6}$$

Для эксперта понятию «среднее число сетевых событий ИБ» соответствует число индикаторов x_{i-1} и x_i, в меньшей степени − число индикаторов 1, 2 и от x_{i+1} до x_I. Число индикаторов, большее x_I, понятием «среднее» охарактеризованы быть не могут, т.е. они не являются носителем данного нечеткого множества.

Введем лингвистические переменные:

«число сетевых событий ИБ на пути распространения атаки»;

«число событий ИБ на хосте»;

«число событий ИБ на периметре»;

«вероятность того, что подозрительная активность в сети является атакой».

Введем в рассмотрение нечеткие множества A, B, C, D с функциями принадлежности $\mu_{\widetilde{A}}, \mu_{\widetilde{B}}, \mu_{\widetilde{C}}, \mu_{\widetilde{D}}$.

$$
\begin{aligned}
A &= \{\mu_{\tilde{A}}(x) \mid x : \mu_{\tilde{A}}(x) \in [0,1], x \in X\},\\
B &= \{\mu_{\tilde{B}}(x) \mid x : \mu_{\tilde{B}}(x) \in [0,1], x \in X\},\\
C &= \{\mu_{\tilde{C}}(x) \mid x : \mu_{\tilde{C}}(x) \in [0,1], x \in X\},\\
D &= \{\mu_{\tilde{D}}(p) \mid p : \mu_{\tilde{D}}(p) \in [0,1], p \in [0,1]\}.
\end{aligned} \tag{7}
$$

Над нечеткими множествами выполняются операции, введенные для использования нечетких множеств в задачах принятия решений. Выбор конкретного вида операции зависит от смысла, вкладываемого в эти операции.

Основные операции над нечеткими множествами: объединение и пересечение множеств.

Объединением нечетких множеств A и B в X называется нечеткое множество $A \cup B$ с функцией принадлежности:

$$
\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}, x \in X. \tag{8}
$$

Пересечением нечетких множеств A и B в X называется нечеткое множество $A \cap B$ с функцией принадлежности:

$$
\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}, x \in X. \tag{9}
$$

Для носителей этих множеств выполняются следующие равенства:

$$
\begin{aligned}
\sup(A \cup B) &= (\sup A) \cup (\sup B),\\
\sup(A \cap B) &= (\sup A) \cap (\sup B).
\end{aligned} \tag{10}
$$

Логической связке «И» соответствует операция пересечения, логической связке «ИЛИ» − операция объединения множеств.

Задаётся область, на которой определены значения каждой лингвистической переменной для пути распространения атаки:

$$
X = \{1,2,\ldots x_i,\ldots x_I\}, \tag{11}
$$

где X − множество числа индикаторов событий ИБ.

В работе применяются числовые лингвистические переменные − нечеткие числа (НЧ). У первых трех ЛП областью определения является интервал, соответствующий оси целых чисел. Множество X области определения НЧ является счетным.

Областью определения ЛП «вероятность того, что подозрительная активность в сети является атакой» является интервал, соответствующий действительной оси от 0 до 1.

Определение степеней принадлежности элементов множества и построение на их основе функции принадлежности (ФП) − основной вопрос, решаемый экспертом. Построение ФП − формализация и интеграция нечетких исходных данных, сформирован-

ных экспертом в процессе оценивания параметров событий безопасности в реальных системах защиты информации (СЗИ). Для эффективного решения задачи необходимо правильно выбрать нужный метод формирования ФП:

метод опроса предполагает наличие группы экспертов;

метод лингвистических термов использует статистические данные, в качестве степени принадлежности элемента множеству принимается оценка частоты использования понятия, которое задается НМ; собирается статистика, характеризующая частоту использования экспертом ЛП для отображения выводов экспертизы;

метод назначения параметров − параметрический метод построения ФП удобен для практического использования, он позволяет формировать трапециевидные и треугольные ФП. Используется следующая экспертная информация о параметре: название параметра, диапазон [a, c] определения параметра, количество лингвистических термов т, с помощью которых оценивается параметр, название каждого лингвистического терма.

Для формирования ФП будем использовать треугольные и трапециевидные формы.

Трапециевидную форму параметрического НЧ $\widetilde{A}$ определяет четверка:

$$A = (a, b_1, b_2, c)_{LR}, \tag{53}$$

где a(c) — нижняя (верхняя) граница НЧ $\widetilde{A}$ на нулевом α-уровне, L и R − линейные функции. Носителем НЧ $\widetilde{A}$ будет интервал [a,c], ядром − [b1,b2].

Аналитический вид трапециевидной ФП

$$\mu_{\widetilde{A}}(x) = \begin{cases} 0, \text{если } x < a, \\ (x-a)/(b_1-a), \text{если } a \le x \le b_1, \\ 1, \text{если } b_1 \le x \le b_2, \\ (c-x)/(c-b_2), \text{если } b_2 \le x \le c, \\ 0, \text{если } x > c. \end{cases} \tag{12}$$

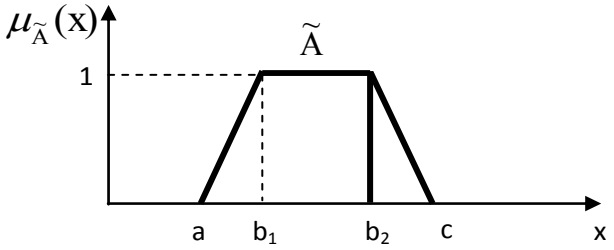

Рис. 1 − Нечеткое число $\widetilde{A}$ с трапециевидной ФП

Треугольную форму НЧ определяет тройка вида

$$\tilde{A} = (a, b, c)_{LR}, \tag{13}$$

где a(c) - нижняя (верхняя) граница НЧ $\tilde{A}$ на нулевом α-уровне, b-значение НЧ $\tilde{A}$ на единичном α-уровне, L и R − линейные функции.

Такому описанию отвечает ФП, имеющая аналитический вид:

$$\mu_{\tilde{A}} = \begin{cases} 0, \text{если } x < a, \\ (x - a)/(b - a), \text{если } a \le x < b, \\ (c - x)/(c - b), \text{если } b \le x < c, \\ 0, \text{если } x > c. \end{cases} \tag{14}$$

Носителем НЧ $\tilde{A}$ в этом случае является интервал [a,c], ядром − число b, интервал [a,c] называют пессимистической оценкой параметра A, число b – оптимистической оценкой параметра A.

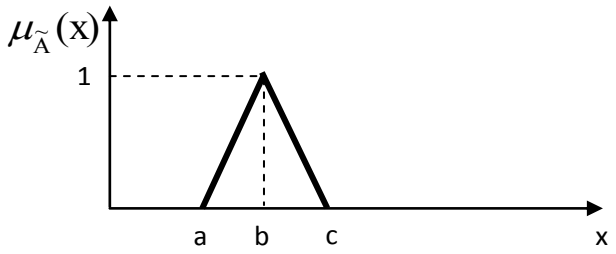

Рис. 2 – Нечеткое число $\tilde{A}$ с треугольной ФП

Как вытекает из условия метода, в процессе формирования ФП её параметры назначает эксперт.

На рисунках 3, 4 и 5 представлены функции принадлежности входных переменных.

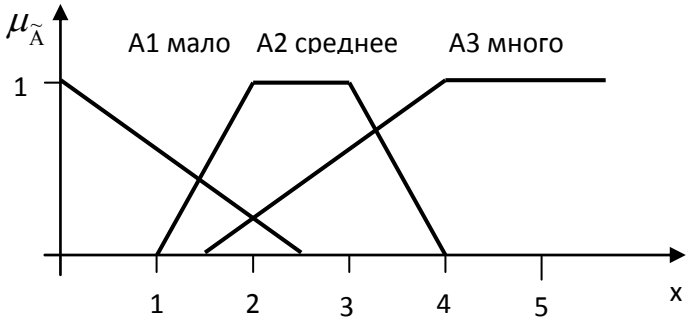

Рис. 3 − Функции принадлежности лингвистических переменных «число сетевых событий ИБ»

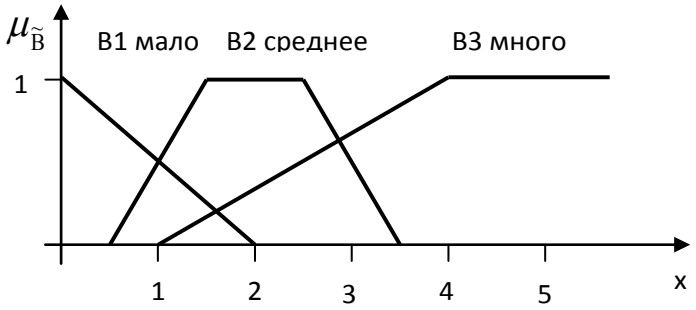

Рис. 4 − Функции принадлежности лингвистической переменной «число событий ИБ на хосте»

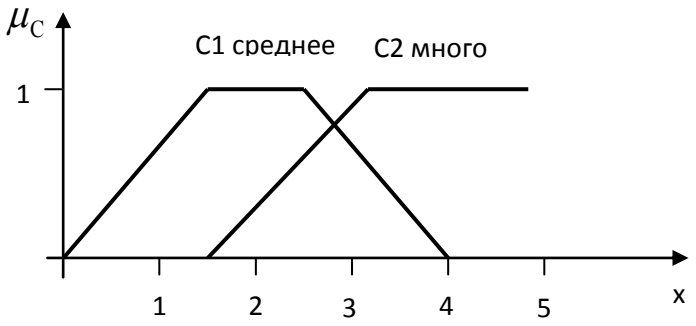

Рис. 5 − Функции принадлежности лингвистической переменной «число событий ИБ на периметре»

Экспертом задаются функции принадлежности лингвистической переменной «вероятность того, что подозрительная активность в сети является атакой». Область, на которой определена лингвистическая переменная, $P \in [0,1]$. Терм – множество лингвистической переменной: низкая, ниже средней, средняя, выше средней, высокая.

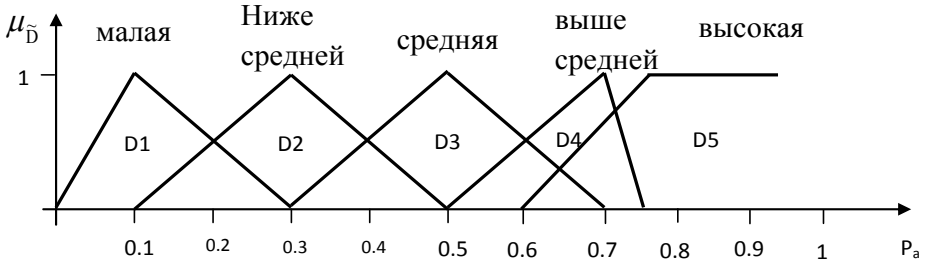

Рис. 6 – Функции принадлежности выходной переменной

На рисунке 7 представлена общая схема механизма нечеткого логического вывода.

Механизм или алгоритм вывода является важной частью базовой архитектуры систем нечеткого вывода. Механизм вывода представляет собой конкретизацию методов прямого и обратного вывода заключений в системах нечетких продукций, в которых условия и заключения записаны в форме нечетких лингвистических переменных. Разработка и применение систем нечеткого вывода включает в себя ряд этапов.

Информацией, которая поступает на вход системы нечеткого вывода, являются измеренные некоторым образом входные переменные, – число признаков аномальных

событий. Эти переменные соответствуют реальным процессам в сети. Информация, которая формируется на выходе системы нечеткого вывода, соответствует выходной переменной, которая является коэффициентом уверенности в том, что аномальные события в сети являются атакой.

Система нечеткого вывода предназначена для преобразования значений входных переменных − информации о количестве индикаторов − в выходную переменную на основе использования нечетких правил продукций. Для этого система нечеткого вывода должна содержать базу правил нечетких продукций и реализовывать нечеткий вывод заключений на основе посылок или условий, представленных в форме нечетких логических высказываний.

Сформируем базу правил системы нечеткого вывода (таблицы 1 – 3).

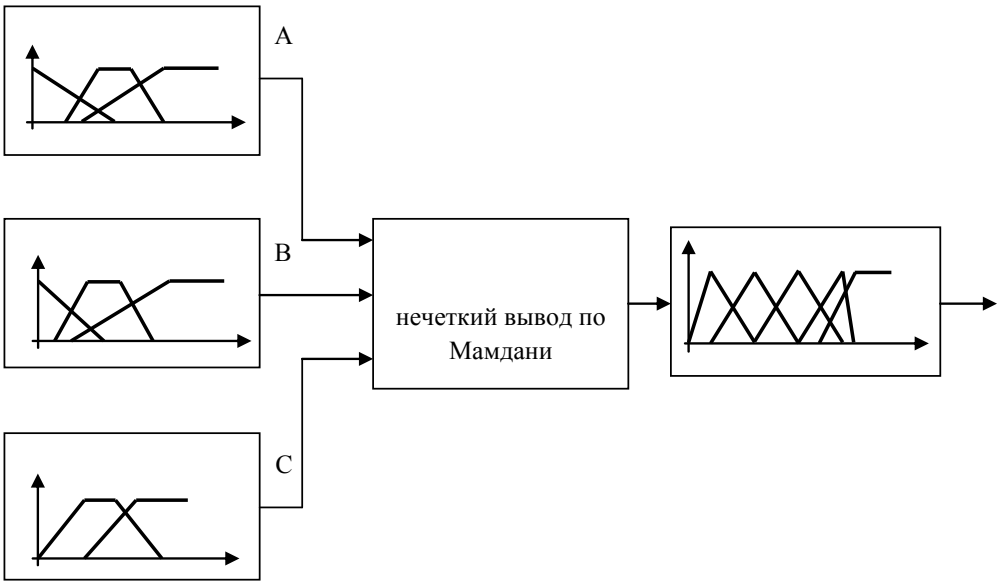

Рис. 7 – Общая схема механизма нечеткого логического вывода

Таблица 1

База нечетких экспертных правил для численной оценки вероятности внутренней угрозы (НЧ A и B)

№	Атрибуты и их значения	Результаты
П1	Если A есть A1 и B есть B1,	тогда D есть D1;
П2	Если A есть A2 и B есть B1 или A есть A1 и B есть B2,	тогда D есть D2;
П3	Если A есть A3 и B есть B1 или A есть A2 и B есть B2 или A если A1 и B если B3,	тогда D есть D3;
П4	Если A есть A3 и B есть B2 или A есть A2 и B есть B3,	тогда D есть D4;
П5	Если A есть A3 и B есть B3,	тогда D есть D5.

Таблица 2

База нечетких экспертных правил для численной оценки вероятности внешней угрозы (НЧ A,B,C)

№	Атрибуты и их значения	Результаты
П1	Если A есть A1 и B есть B1 и C есть C1,	тогда D есть D1;

П2	Если А есть А2 и В есть В1 и С есть С1 или А есть А1 и В есть В2 и С есть С1 или А есть А1 и В есть 81 и С есть С2,	тогда D есть D2;
П3	Если А есть А2 и В есть В1 и С есть С2 или А есть А1 и В есть В2 и С есть С2,	тогда D есть D3;
П4	Если А есть А3 и В есть В1 и С есть С1 или А есть А3 и В есть В2 и С есть С1 или А есть А2 и В есть 82 и С есть С1 или А есть А2 и В есть В3 и С есть С1, или А есть А1 и В есть В3 и С есть С1, или А есть А3 и В есть В1 и С есть С2, или А есть А2 и В есть В2 и С есть С2,	тогда D есть D4;
П5	Если А есть А3 и В есть В2 и С есть С2 или А есть А3 и В есть В3 и С есть С2 или А есть А2 и В есть 83 и С есть С2 или А есть А3 и В есть В3 и С есть С1 или А есть А1 и В есть В3 и С есть С2,	тогда D есть D5;

Таблица 3

База нечетких экспертных правил для численной оценки вероятности внешней угрозы (НЧ А и С)

№	Атрибуты и их значения	Результаты
П1	Если А есть А1 и С есть С1,	тогда D есть D2;
П2	Если А есть А2 и С есть С1 или А есть А1 и С есть С2,	тогда D есть D3;
П3	Если А есть А3 и С есть С1 или А есть А2 и С есть С2,	тогда D есть D4;
П4	Если А есть А3 и С есть С2,	тогда D есть D5;

Следующий этап нечеткого вывода – фаззификация, под которой понимается процедура нахождения значений функций принадлежности нечетких множеств (термов) на основе обычных (не нечетких) исходных данных. Целью этапа фаззификации является установление соответствия между численными значениями отдельной входной переменной системы нечеткого вывода и значением функции принадлежности соответствующего ей терма входной лингвистической переменной. После завершения этого этапа для всех входных переменных должны быть определены конкретные значения функций принадлежности по каждому из лингвистических термов, которые используются в подусловиях базы правил.

На этапе агрегирования определяется степень истинности условий по каждому из правил системы нечеткого вывода. Если условие правила состоит, из нескольких подусловий, то определяется степень истинности сложного высказывания на основе расчетных формул нечеткой конъюнкции или нечеткой дизъюнкции. Этап агрегирования считается законченным, когда найдены все значения истинности для каждого из правил, входящих в рассматриваемую базу правил системы нечеткого вывода. Те правила, степень истинности условий которых отлична от нуля, считаются активными и используются для дальнейших расчетов, при этом для сокращения времени вывода учитываются только активные правила нечетких продукций.

Аккумуляция значений нечетких правил продукций осуществляется для объединения нечетких множеств, соответствующих термам подзаключений, относящихся к одним и тем же выходным лингвистическим переменным.

Дефаззификация в системах нечеткого вывода представляет собой процедуру или процесс нахождения обычного (не нечеткого) значения выходных лингвистических переменных. Дефаззификацию называют приведением к четкости. Действительно, применяемые в системах управления модули способны воспринимать команды в форме количественных значений соответствующих переменных. Традиционно используется метод центра тяжести. При дефаззификации методом центра тяжести обычное (не нечеткое) значение выходной переменной равно абсциссе центра тяжести площади, ограниченной графиком кривой функции принадлежности выходной переменной.

Примеры дефаззификации методом центра тяжести функции принадлежности выходной лингвистической переменной «вероятность того, что подозрительная активность в сети является атакой» изображены на рисунках 8 – 9.

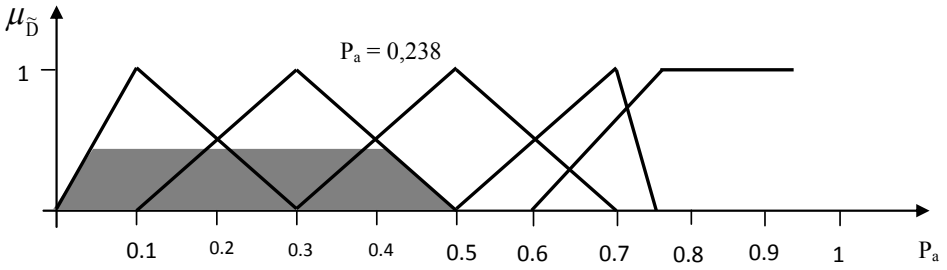

Рис. 8 - Пример дефаззификации функции принадлежности выходной лингвистической переменной в случае численной оценки вероятности удаленной атаки при следующих значениях входных НЧ: A=1 , B=1, C=1

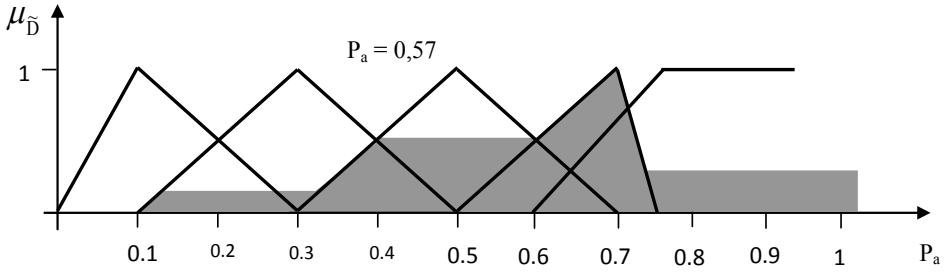

Рис. 9 – Пример дефаззификации функции принадлежности выходной лингвистической переменной в случае численной оценки вероятности внутренней атаки при следующих значениях входных НЧ: A= 2, B=3

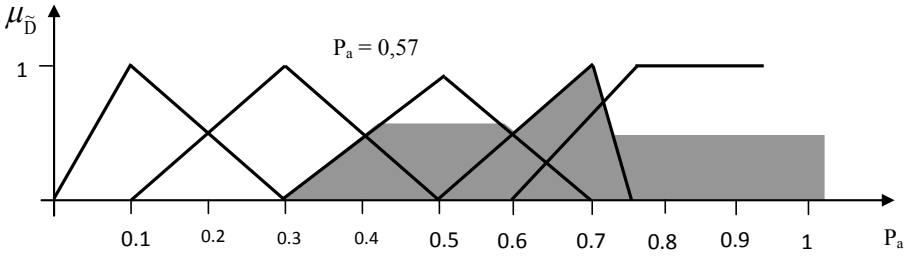

Рис. 10 – Пример дефаззификации функции принадлежности выходной лингвистической переменной в случае численной оценки вероятности удаленной атаки при следующих значениях входных НЧ: A= 3, C=3

Выводы

Аналогично приведенным в примерах расчетам, аппарат нечеткой логики можно использовать для численной оценки вероятности того, что аномальная активность в сети является атакой.

Преимущество предложенного подхода заключается, во-первых, в использовании опыта квалифицированных экспертов, обобщенного в виде системы продукционных правил, во-вторых, в возможности количественной оценки вероятности конкретной атаки.

При этом для каждого выявленного пути распространения атаки вероятность атаки, или «коэффициент уверенности», с которым аномальную активность можно отнести к атаке, рассчитывается на заданном экспертом интервале анализа. Временная ось раз-

бивается на интервалы анализа Δta, которые определяются для каждого заданного пути и скорости обнаружения аномальной активности подсистемами обнаружения.

Литература

1. Корченко, А. Г. Построение систем защиты информации на нечетких множествах. - Киев: МК-пресс, 2006. - 316 с.
2. Леоненков, А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. - СПб.: БХВ-Петербург, 2005. - 736 с.
3. Гаскаров, Д. В. Интеллектуальные информационные системы: учебник для вузов. - М.: Высш. шк., 200 - 431 с.

Надійшла до редколегії 15.02.2013 р.

Рецензент: д.т.н., проф. Хорошко В.О.

Петров А.О.

МЕТОД ОЦІНКИ ЙМОВІРНОСТІ ВИНИКНЕННЯ АНОМАЛЬНИХ ПОДІЙ В КОМП'ЮТЕРНІЙ МЕРЕЖІ, ЩО ГРУНТУЄТЬСЯ НА СИСТЕМАХ НЕЧІТКИХ МНОЖИН

У статті запропоновано метод оцінки ймовірності виникнення аномальних подій в комп'ютерній мережі, що грунтується на системах нечітких множин.

Ключові слова: комп'ютерна мережа, нечітка множина, захист інформації, аномальна подія.

Petrov A.A.

THE METHOD OF ESTIMATING THE PROBABILITY OF OCCURRENCE OF ABNORMAL EVENTS IN A COMPUTER NETWORK, BASED ON A SYSTEM OF FUZZY SETS

The paper proposed a method for estimating the probability of occurrence of abnormal events in a computer network, based on a system of fuzzy sets.

Keywords: computer network, fuzzy set, data protection, abnormal event.