УДК 004.056.53(045)

**Стасюк А.И., Корченко А.А.**

# БАЗОВАЯ МОДЕЛЬ ПАРАМЕТРОВ ДЛЯ ПОСТРОЕНИЯ СИСТЕМ ВЫЯВЛЕНИЯ АТАК

Предложена обобщенная модель параметров, которая ориентирована на построение систем выявления атак, основанных на идентификации аномального состояния в информационной системе. Модель основывается на трех множествах – возможных атак, возможных параметров и наборов лингвистических связок. Для построения соответствующих систем обнаружения формируются множества пар – "атака→параметры" и "атака→набор лингвистических связок", на основании которых строятся лингвистические переменные, эталоны параметров и логические правила, ориентированные на обнаружение вторжений в информационную систему.

Ключевые слова: атака, аномалия, аномальное состояние, нечеткая среда, системы выявления атак, защита информации.

Интенсивное развитие информационных технологий оказало положительное влияние на все сферы человеческой деятельности. Вместе с этим наблюдаются и побочные эффекты, один из которых связан с увеличением атак на ресурсы информационных систем (РИС). Особого внимания заслуживают DoS-атаки, которые являются одними из самых опасных и простых в организации, дешевыми по стоимости и очень сложными в защите. Так, например, в Украине они направлялись на сайты органов государственной власти (Президента, Кабинета Министров, СБУ, МВД) и при этом было установлено около 5 тысяч активных пользователей, осуществлявших это нападение [1]. Произошедшие события вскрыли неготовность систем безопасности к такого типа спланированным атакам, и даже на государственном уровне не оказалось (для реализации соответствующих выявляющих и блокирующих действий) достаточно эффективных средств защиты, к которым относятся и системы выявления атак (СВА). В связи с этим, создание методов и моделей, позволяющих разрабатывать эффективные СВА является актуальной научной задачей.

Системы, основанные на сигнатурном подходе, как правило, позволяют идентифицировать только известные формы вторжений, а обнаружения неизвестных в большей степени осуществляется с помощью СВА, основанных на идентификации аномального состояния. Они, как правило, направлены на функционирование в слабоформализованной нечетко определенной среде, для которой требуется определить набор параметров необходимый для выявления атак, породивших аномалии в информационной системе (ИС). В работах [2-4] показана эффективность применения нечетких множеств для решения различных задач защиты информации, а также, на основе логико-лингвистического подхода, рассмотрен пример формирования нечетких параметров для построения систем выявления такого вида атак как сканирования портов.

Использование математического аппарата нечетких множеств для формализации подхода к рациональному формированию необходимых (для решения подобных задач) параметров, позволит повысить эффективность разрабатываемых СВА. В этой связи, целью данной работы является создание обобщенной модели параметров, позволяющей синтезировать эффективно функционирующие системы обнаружения вторжений по аномальному состоянию параметров (например, сетевого трафика), характеризующих состояние среды окружения ИС.

Для создания базовой модели параметров введем два множества – множество возможных атак на РИС

$$\boldsymbol{AT}=\bigcup_{i=1}^{n} AT_i = \{AT_1, AT_2, \ldots, AT_n\} \quad (1)$$

и множество возможных параметров

$$\boldsymbol{P}=\bigcup_{i=1}^{m} P_i = \{P_1, P_2, \ldots, P_m\}, \quad (2)$$

характеризующих состояние среды окружения. По значениям параметров из выражения (2) можно выявить аномальное состояние в ИС, порождаемое определенным элементом из множества $\boldsymbol{AT}$ из формулы (1), где $n$ определяет количество возможных атак, а $m$ – общее количество возможных параметров. Например, при $n=3$ (1) можно определить как:

$$\boldsymbol{AT}=\bigcup_{i=1}^{3} AT_i = \{AT_1, AT_2, AT_3\}=\{SN, DS, SP\}, \quad (3)$$

где $AT_1=SN$, $AT_2=DS$ и $AT_3=SP$ соответственно являются идентификаторами атак типа "Сканирование портов", "Отказ в обслуживании (DoS)" и "Спуфинг". Например, при $m=6$ выражение (2) принимает следующий вид:

$$\boldsymbol{P}=\bigcup_{i=1}^{6} P_i = \{P_1, P_2, P_3, P_4, P_5, P_6\}=\{КВК, ВВК, КОП, СОЗ, ЗМЗ, КПОА\}, \quad (4)$$

где $P_1=КВК$, $P_2=ВВК$, $P_3=КОП$, $P_4=СОЗ$, $P_5=ЗМЗ$, и $P_6=КПОА$ соответственно являются идентификаторами параметров типа "Количество виртуальных каналов", "Возраст виртуального канала", "Количество одновременных подключений к серверу", "Скорость обработки запросов от клиентов", "Задержка между запросами от одного пользователя" и "Количество пакетов с одинаковым адресом отправителя и получателя".

Каждому элементу (типу атаки) множества $\boldsymbol{AT}$ ставится в соответствие подмножество набора параметров $\boldsymbol{P_n}$ (необходимого для обнаружения аномалий), составленного из элементов множества $\boldsymbol{P}$. Таким образом формируется множество пар – "атака→параметры", т.е.

$$\boldsymbol{AT}\rightarrow\boldsymbol{P_n}=\bigcup_{i=1}^{n}\Big( AT_i \rightarrow \bigcup_{j=1}^{k_i} P_{ij} \Big)=\{(AT_1\rightarrow\{P_{11}, P_{12}, \ldots P_{1k_1}\}),$$
$$(AT_2\rightarrow\{P_{21}, P_{22}, \ldots P_{2k_2}\}), \ldots (AT_i\rightarrow\{P_{i1}, P_{i2}, \ldots P_{ik_j}\})\}. \quad (5)$$

Например, при $k_1=k_3=2$, и $k_2=3$, с учетом формулы (3) определим, что $P_{11}=P_1$, $P_{12}=P_2$, $P_{21}=P_3$, $P_{22}=P_4$, $P_{23}=P_5$, $P_{31}=P_3$, и $P_{32}=P_6$ и тогда выражение (5) с учетом формулы (4) будет иметь следующий вид:

$$\boldsymbol{AT}\rightarrow\boldsymbol{P_n}=\bigcup_{i=1}^{3}\Big( AT_i \rightarrow \bigcup_{j=1}^{k_i} P_{ij} \Big)=\{(AT_1\rightarrow\{P_1, P_2\}), (AT_2\rightarrow\{P_3, P_4, P_5\}), (AT_i\rightarrow\{P_3, P_6\})\}=$$
$$=\{(SN\rightarrow\{КВК, ВВК\}), (DS\rightarrow\{КОП, СОЗ, ЗМЗ\}), (SP\rightarrow\{КОП, КПОА\})\}. \quad (6)$$

Каждый $P_{ij}$ (с учетом [2, 3]) удобно отображать лингвистическими переменными (ЛП), каждая из которых представляется кортежем

$$<P_{ij}, T_{ij}, U_{ij}> (i=\overline{1,n}; j=\overline{1,k_i}), \quad (7)$$

где $P_{ij}$ идентификатор (имя) ЛП, $T_{ij}$ – базовое терм-множество (содержит термы $T_{ij}^k$ $(k=\overline{1,r})$), $U_{ij}$ – универсальное множество, являющееся областью определения для $T_{ij}$. Отметим, что $m$ и $P_{ij}$ определяются исходя из специфики реализации атаки и количества признаков, по которым можно определить аномальное состояние в среде окружения информационной системы. Например, с учетом выражений (6) и (7) набор кортежей, отображающих соответствующие значения ЛП для: $P_{11}$ и $P_{12}$ имеет вид $<P_{11}, T_{11}, U_{11}>$, $<P_{12}, T_{12}, U_{12}>$, т.е. $<КВК, T_{КВК}, U_{КВК}>$, $<ВВК, T_{ВВК}, U_{ВВК}>$; $P_{21}$, $P_{22}$ и $P_{23}$ – $<P_{21}, T_{21}, U_{21}>$, $<P_{22}, T_{22}, U_{22}>$, $<P_{23}, T_{23}, U_{23}>$, т.е. $<КОП, T_{КОП}, U_{КОП}>$, $<СОЗ, T_{СОЗ}, U_{СОЗ}>$, $<ЗМЗ, T_{ЗМЗ}, U_{ЗМЗ}>$; $P_{31}$ и $P_{32}$ – $<P_{31}, T_{31}, U_{31}>$, $<P_{32}, T_{32}, U_{32}>$, т.е. $<КОП, T_{КОП}, U_{КОП}>$, $<КПОА, T_{КПОА}, U_{КПОА}>$.

Далее для каждой ЛП формируются $r$ нечетких термов

$$T_{ij}=\bigcup_{k=1}^{r} T_{ij}^k = \{T_{ij}^1, T_{ij}^2, \ldots, T_{ij}^k\}, \quad (8)$$

которые могут отображаться на универсальное множество $U_{ij}$ с областью определения $[P_{ij}^{\min}, P_{ij}^{\max}]$, где $P_{ij}^{\min}$ и $P_{ij}^{\max}$ соответственно нижняя и верхняя границы значений $T_{ij}$. Например, если ЛП $P_{11}$ определяется пятью термами $(r=5)$, а $P_{12}$ тремя $(r=3)$, то с учетом выражения (8) базовое терм-множество для $P_{11}$ определяется как:

$$T_{11}=\bigcup_{k=1}^{5} T_{11}^{k}=\{T_{11}^{1},\ T_{11}^{2},\ \ldots,\ T_{11}^{5}\}=\{T_{КВК}^{1},\ T_{КВК}^{2},\ \ldots,\ T_{КВК}^{5}\}=$$
$$=\{\text{“ОЧЕНЬ МАЛОЕ” (ОМ), “МАЛОЕ” (М), “СРЕДНЕЕ” (С),}$$
$$\text{“БОЛЬШОЕ” (Б), “ОЧЕНЬ БОЛЬШОЕ” (ОБ)}\}, \quad (9)$$

и может быть отображено на универсальном множестве $U_{ij}$ с областью определения $[P_{11}^{\min}, P_{11}^{\max}]=[0, 256]$, а для $P_{12}$ – как:

$$T_{12}=\bigcup_{k=1}^{3} T_{12}^{k}=\{T_{12}^{1},\ \ldots,\ T_{12}^{3}\}=\{T_{ВВК}^{1},\ \ldots, T_{ВВК}^{3}\}=$$
$$=\{\text{“МОЛОДОЙ” (М), “СРЕДНИЙ” (СР), “СТАРЫЙ” (СТ)}\}, \quad (10)$$

которые могут быть отображены на универсальном множестве $U_{ij}$ с областью определения $[P_{12}^{\min}, P_{12}^{\max}]=[0, 250]$. Отметим, что множество термов $T_{ij}$ $(i=\overline{1,n},\ j=\overline{1,m})$ отображается $r$ нечеткими числами (НЧ)

$$T_{ij}\in\bigcup_{f=1}^{r} \underset{\sim}{T}_{ij}^{f}=\{\underset{\sim}{T}_{ij}^{1},\ \underset{\sim}{T}_{ij}^{2},\ \ldots,\ \underset{\sim}{T}_{ij}^{r}\}, \quad (11)$$

для которых необходимо сформировать функции принадлежности (ФП) одним из известных методов [4]. Например, термы $T_{11}$ $(при\ r=5)$ и $T_{12}$ (при $r=3$) с учетом формул (10) и (11) можно соответственно отобразить НЧ $\underset{\sim}{T}_{11}^{1}$, $\underset{\sim}{T}_{11}^{2}$, $\underset{\sim}{T}_{11}^{3}$, $\underset{\sim}{T}_{11}^{4}$, $\underset{\sim}{T}_{11}^{5}$ (т.е. $\underset{\sim}{ОМ}$, $\underset{\sim}{М}$, $\underset{\sim}{С}$, $\underset{\sim}{Б}$, $\underset{\sim}{ОБ}$) и $\underset{\sim}{T}_{12}^{1}$, $\underset{\sim}{T}_{12}^{2}$, $\underset{\sim}{T}_{12}^{3}$ (т.е. $\underset{\sim}{М}$, $\underset{\sim}{СР}$, $\underset{\sim}{СТ}$), для которых формируются ФП. Получить ФП можно, например, на основе метода лингвистических термов с использованием статистических данных (МЛТС) [2], при помощи которого для любого из заданных термов определяется $l$ номеров интервалов $N_{ij}^{1}$, $N_{ij}^{2}$, ..., $N_{ij}^{l}$ возможных значений с соответствующими граничными параметрами $P_{ij}^{\min}$ и $P_{ij}^{\max}$. Здесь в качестве исходных данных может использоваться статистическая, аналитическая, экспертная и другая информация, обычно применяемая для построения нечетких эталонов, т.е. эталонов параметров, с помощью которых осуществляется классификация текущего состояния величин в аномальной среде. Например, для $T_{11}$ при $l=5$ значениям номеров $N_{1j}^{1}=N_{11}^{1}$, $N_{1j}^{2}=N_{11}^{2}$, $N_{1j}^{3}=N_{11}^{3}$, $N_{1j}^{4}=N_{11}^{4}$, $N_{1j}^{5}=N_{11}^{5}$ будут соответствовать интервалы $[P_{11}^{\min}=P_{11}^{0},\ P_{11}^{1}],\ ]P_{11}^{1},\ P_{11}^{2}],\ ]P_{11}^{2},\ P_{11}^{3}],\ \ldots,\ ]P_{11}^{4},\ P_{11}^{5}=P_{11}^{\max}]$ т.е. *[0; 2], ]2; 8], ]8; 16], ]16; 64], ]64; 256]*, а для $T_{12}$ при $l=3$ номерам интервалов $N_{1j}^{1}=N_{12}^{1}$, $N_{1j}^{2}=N_{12}^{2}$, $N_{1j}^{3}=N_{12}^{3}$ соответствуют $[P_{12}^{\min}=P_{12}^{0},\ P_{12}^{1}],\ ]P_{12}^{1},\ P_{12}^{2}],\ ]P_{12}^{2},\ P_{12}^{3}=P_{12}^{\max}]$ т.е. *[0; 30], ]30; 100], ]100; 250]*.

На основе полученных значений ФП НЧ $\underset{\sim}{T}_{ij}^{f}$ $(f=\overline{1,r})$ для каждой $P_{ij}$ формируются эталоны параметров $\underset{\sim}{T}_{ij}^{ef}$ $(f=\overline{1,r};\ i=\overline{1,n};\ j=\overline{1,m})$, по принципу определенного класса НЧ на основе признаков нормальности, модальности, выпуклости, непрерывности и параметричности [2]. Например, для $P_{11}=КВК$ и $P_{12}=ВВК$ значения $\underset{\sim}{T}_{11}^{ef}=\underset{\sim}{T}_{КВК}^{ef}$, $(f=\overline{1,5})$ и

$\underset{\sim}{T}^{ef}_{12} = \underset{\sim}{T}^{ef}_{BBK}$, $(f=\overline{1,3})$ могут быть определены нормальными унимодальными выпуклыми дискретными непараметрическими НЧ с произвольным числом носителей [2], т.е. $\underset{\sim}{T}^{e1}_{11} = \underset{\sim}{T}^{e1}_{KBK} = \underset{\sim}{OM}^{e} = \{0/0{,}008;\ 1/0{,}008;\ 0{,}33/0{,}031;\ 0/0{,}063\}$, $\underset{\sim}{T}^{e2}_{11} = \underset{\sim}{T}^{e2}_{KBK} = \underset{\sim}{M}^{e} = \{0/0{,}008;\ 0{,}5/0{,}008;\ 1/0{,}031;\ 0{,}5/0{,}063;\ 0/0{,}25\}$, $\underset{\sim}{T}^{e3}_{11} = \underset{\sim}{T}^{e3}_{KBK} = \underset{\sim}{C}^{e} = \{0/0{,}008;\ 0{,}33/0{,}031;\ 1/0{,}063;\ 0{,}67/0{,}25;\ 0/1\}$, $\underset{\sim}{T}^{e4}_{11} = \underset{\sim}{T}^{e4}_{KBK} = \underset{\sim}{Б}^{e} = \{0/0{,}063;\ 1/0{,}25;\ 0{,}75/1;\ 0/1\}$, $\underset{\sim}{T}^{e5}_{11} = \underset{\sim}{T}^{e5}_{KBK} = \underset{\sim}{OБ}^{e} = \{0/0{,}063;\ 0{,}2/0{,}25;\ 1/1;\ 0/1\}$ и соответственно $\underset{\sim}{T}^{e1}_{12} = \underset{\sim}{T}^{e1}_{BBK} = \underset{\sim}{M}^{e} = \{1/0;\ 1/0{,}12;\ 0{,}5/0{,}4;\ 0{,}25/1\}$, $\underset{\sim}{T}^{e2}_{12} = \underset{\sim}{T}^{e2}_{BBK} = \underset{\sim}{CP}^{e} = \{0{,}2/0;\ 0{,}2/0{,}12;\ 1/0{,}4;\ 0{,}4/1\}$, $\underset{\sim}{T}^{e3}_{12} = \underset{\sim}{T}^{e3}_{BBK} = \underset{\sim}{CT}^{e} = \{0/0{,}12;\ 0{,}17/0{,}4;\ 1/1\}$.

Принятие решения о том, что состояние среды характерно для процесса реализации атаки удобно осуществлять на основе наборов лингвистических связок, множество которых обозначим через:

$$\boldsymbol{LC} = \bigcup_{i=1}^{n}\left(\bigcup_{j=1}^{c_n} LC_{ij}\right) = \{(\boldsymbol{LC_1}), (\boldsymbol{LC_2}), \ldots, (\boldsymbol{LC_n})\} = \{(LC_{11}, LC_{12}, \ldots, LC_{1c_1}), (LC_{21}, LC_{22}, \ldots, LC_{2c_2}), \ldots (LC_{n1}, LC_{n2}, \ldots, LC_{nc_n})\}, \quad (12)$$

где $c_n$ – количество лингвистических связок в наборе, необходимых для составления правил направленных на выявление $n$-й атаки. Элементы $\boldsymbol{AT}$ в совокупности $\boldsymbol{LC}$ могут формировать множества пар – “атака→набор лингвистических связок”:

$$\boldsymbol{AT \rightarrow LC} = \left(\bigcup_{i=1}^{n} AT_i \rightarrow \bigcup_{j=1}^{c_i} LC_{ij}\right) = \{(AT_1 \rightarrow \boldsymbol{LC_1}), (AT_2 \rightarrow \boldsymbol{LC_2}), \ldots, (AT_n \rightarrow \boldsymbol{LC_n})\} = \{(AT_1 \rightarrow \{LC_{11}, LC_{12}, \ldots, LC_{1c_1}\}), (AT_2 \rightarrow \{LC_{21}, LC_{22}, \ldots, LC_{2c_2}\}), \ldots, (AT_i \rightarrow \{LC_{i1}, LC_{i2}, \ldots, LC_{ic_j}\})\}. \quad (13)$$

Например, при $c_1=c_2=c_3=5$ выражение (13) с учетом [5] примет следующий вид:

$$\boldsymbol{AT \rightarrow LC} = (AT_1 \rightarrow \{LC_{11}, LC_{12}, \ldots, LC_{15}\}), \ldots, (AT_5 \rightarrow \{LC_{51}, LC_{52}, \ldots, LC_{55}\}) =$$

$$\{(SN \rightarrow \{\text{“КВК } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{OM}^{e}\text{ ”, “КВК } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{M}^{e}\text{ ”, “КВК } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{C}^{e}\text{ ”, “КВК } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{Б}^{e}\text{ ”, “КВК } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{OБ}^{e}\text{ ”}\}),$$

$$(SP \rightarrow \{\text{“КПОА } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{OM}^{e}\text{ ”, “КПОА } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{M}^{e}\text{ ”, “КПОА } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{C}^{e}\text{ ”, “КПОА } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{Б}^{e}\text{ ”, “КПОА } \underset{\sim}{M} \text{ соизмеримо с } \underset{\sim}{OБ}^{e}\text{ ”}\}),$$

$$(DS \rightarrow \{\text{“СОЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{OM}^{e}\text{ ”, “СОЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{M}^{e}\text{ ”, “СОЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{C}^{e}\text{ ”, “СОЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{Б}^{e}\text{ ”, “СОЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{OБ}^{e}\text{ ”}\}),$$

$$(DS \rightarrow \{\text{“ЗМЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{OM}^{e}\text{ ”, “ЗМЗ } \underset{\sim}{H} \text{ соизмеримо с } \underset{\sim}{M}^{e}\text{ ”, “ЗМЗ } \underset{\sim}{H} \quad (14)$$

*соизмеримо с $\underset{\sim}{С}^{e}$ ", "ЗМЗ $\underset{\sim}{Н}$ соизмеримо с $\underset{\sim}{Б}^{e}$ ", "ЗМЗ $\underset{\sim}{Н}$ соизмеримо с $\underset{\sim}{ОБ}^{e}$ "})}.*

На основе множества ***LC*** строятся логические правила типа – "Если $LC_{ij}$ то …" и например, для выражения (14) они будут иметь следующий вид: 1. Если $LC_{11}$, то возможность сканирования Н; 2. Если $LC_{12}$, то возможность сканирования БНВ; 3. Если $LC_{13}$, то возможность сканирования БВН; 4. Если $LC_{14}$, то возможность сканирования В; 5. Если $LC_{15}$, то возможность сканирования В, где Н – низкая, БНВ – больше низкая чем высокая, БВН – больше высокая чем низкая, В – высокая, а понятие "соизмеримо", используемое в лингвистических связках, может отображать минимальное расстояние Хемминга [4] между значениями используемых величин.

Таким образом, на основе предложений модели параметров, базирующейся на множествах возможных атак и возможных параметров, а также на множествах пар "атака→параметры" и "атака→набор лингвистических связок" можно строить модели систем выявления атак, позволяющих повысить эффективность соответствующих средств, основанных на идентификации аномального состояния параметров в информационной системе.

## ЛИТЕРАТУРА


1. СБУ нашла хакеров, мстивших за EX.UA. Ими оказались дети [Электронный ресурс] / Редакция «Зеркало недели. Украина» // ZN.UA : [Зеркало недели. Украина]. — Электрон. дан. — 2012. — 29 февраля. — Режим доступа: World Wide Web. — URL: http://news.zn.ua/SOCIETY/sbu_nashla_hakerov,_mstivshih_za_exua_imi_okazalis_deti-98040.html. — Загл. с титул. экрана.
2. Корченко О. Г. Построение систем защиты информации на нечетких множествах [Текст] : Теория и практические решения / О. Г. Корченко. — К. : МК-Пресс, 2006. — 320 с.
3. Корченко А. О. Система виявлення аномалій на основі нечітких моделей [Текст] / В. В. Волянська, А. О. Корченко, Є. В. Паціра // Зб. наук. пр. Інституту проблем моделювання в енергетиці НАН України ім. Г. Є Пухова. — Львів : ПП «Системи, технології, інформаційні послуги», 2007. — [Спец. випуск]. — Т.2. — С. 56–60.
4. Корченко О. Г. Системи захисту інформації [Текст] : Монографія / О. Г. Корченко. — К. : НАУ, 2004. — 264 с.
5. Охріменко А. О. Модель виявлення спуфінг-атак на ресурси інформаційних систем [Текст] / А. О. Охріменко, А. О. Корченко // Інформаційні технології та захист інформації : ІІІ міжнародна науково-практична конференція : [Збірник тез]. — Х. : ХУПС ім. І. Кожедуба, 2012. — С. 210.


Надійшла: 11.06.2012

Рецензент: д.т.н., проф. Жуков І.А.