UDC 004.7:62-519:621.391 (045)

Райба С.В., Райба Т.Я, Райф П.Б. Моделювальні дослідження випадкового керування доступом в безпровідній сенсорній мережі

Анотація. *У статті представлено результати дослідження роботи мережі БСМ з випадковим керуванням доступом, використовуючи систему PASTA (середнє значення за час спостереження надходження пуассонівського потоку). Наведено результати моделювальних тестувань ефективності функціонування мережі для двох випадків: 1) коли всі вузли мережі передають протоколи з тим самим середнім часом між передачами, 2) якщо вузли мережі поділено на групи, які мають різні середні часи між передачами. Цей підхід має багато практичних переваг, які висвітлено у статті..*

Ключові слова: *безпровідна сенсорна мережа (БСМ), середнє значення за час спостереження надходження пуассонівського потоку (система PASTA), імовірність колізії, випадкове керування, моделювання.*

Райба С.В., Райба Т.Я, Райф П.Б. Моделирующие исследования случайного управления доступом в беспроводной сенсорной сети

Аннотация. *В статье представлено результаты исследования работы сети БСС со случайным управлением доступом, используя систему PASTA (среднее значение за время наблюдения поступления пуассоновского потока). Приведено результаты моделирующих тестирований эффективности функционирования сети для двух случаев: 1) если все узлы сети передают протоколы с тем же средним временем между передачами, 2) если узлы сети разделены на группы, которые имеют разные средние времена между передачами. Этот подход имеет много практических преимуществ, которые отражено в статье.*

Ключевые слова: *беспроводная сенсорная сеть (БСС), среднее значение за время наблюдения поступления пуассоновского потока (система PASTA), вероятность коллизии, случайное управление, моделирование.*

Отримано *15 січня 2013 року*, затверджено редколегією *19 лютого 2013 року*

МОДЕЛІ ЕТАЛОНІВ ЛІНГВІСТИЧНИХ ЗМІННИХ ДЛЯ СИСТЕМ ВИЯВЛЕННЯ ТА ІДЕНТИФІКАЦІЇ ПОРУШНИКА ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Владислава Волянська, Андрій Гізун, Віктор Гнатюк

Національний авіаційний університет, Україна

ВОЛЯНСЬКА Владислава Вікторівна

Рік та місце народження: 1977 рік, м. Плавськ, Росія.
Освіта: Національний авіаційний університет, 2004 рік.
Посада: IT-Manager for network infrastructure, Apogeum Sp. z o.o. Poland, здобувач кафедри безпеки інформаційних технологій.
Наукові інтереси: інформаційна безпека операційних систем, управління інцидентами інформаційної безпеки, комплексні системи захисту інформації.
Публікації: більше 20 наукових публікацій, серед яких наукові статті, матеріали і тези доповідей на конференціях.
E-mail: volyanska.vladyslava@gmail.com

ГІЗУН Андрій Іванович

Рік та місце народження: 1987 рік, м. Нетішин, Хмельницька область, Україна.
Освіта: Національний авіаційний університет, 2010 рік.
Посада: асистент кафедри безпеки інформаційних технологій з 2012 року.
Наукові інтереси: інформаційна безпека, управління інцидентами інформаційної безпеки, штучні імунні системи, управління безперервністю бізнесу та правове забезпечення захисту інформації.
Публікації: більше 20 наукових публікацій, серед яких наукові статті, матеріали і тези доповідей на конференціях, авторські свідоцтва.
E-mail: caesar07@meta.ua

ГНАТЮК Віктор Олександрович

Рік та місце народження: 1990 рік, м. Нетішин, Хмельницька область, Україна.
Освіта: Хмельницький національний університет, 2012 рік.
Посада: аспірант кафедри безпеки інформаційних технологій з 2012 року.
Наукові інтереси: інформаційна безпека, управління інцидентами інформаційної безпеки.
Публікації: більше 10 наукових публікацій, серед яких наукові статті, тези та матеріали доповідей на конференціях, авторські свідоцтва.
E-mail: viktorgnatyuk@meta.ua

Анотація. *Вивчення факту порушення інформаційної безпеки та особистості порушника має велике наукове та практичне значення. З цих позицій, формалізація параметрів, які можуть бути використані для ідентифікації порушників, є актуальною науковою задачею. Чітке визначення повної множини параметрів дозволить підвищити ефективність превентивних заходів та систем захисту. В роботі запропонована модель еталонів лінгвістичних змінних, орієнтована на побудову системи виявлення порушника (вторгнення). Для виявлення порушників використовується ряд хостових та мережевих параметрів, більшість з яких мають нечітку природу. На основі проведеного експерименту побудовані моделі еталонів цих параметрів з використанням нечітких чисел. Отримані результати можуть бути базисом для побудови системи виявлення вторгнень на основі технології honeypot.*

Ключові слова: *порушник інформаційної безпеки, система виявлення порушника, параметри, ідентифікація, лінгвістичні змінні, нечітка логіка, еталони параметрів.*

Вступ. Розвиток інформаційних технологій породжує нові види загроз інформаційній безпеці, серед яких чільне місце посідає вторгнення порушника в комп'ютерні системи та мережі. Для ефективної протидії цій загрозі розробляють системи виявлення порушника (СВП), які дають змогу виявити факт вторгнення в систему порушника та ідентифікувати його.

У процесі атаки порушник, діючи на систему, змінює певні її параметри, створює або припиняє властиві їй процеси тощо. Всі ці дії відображаються на стані системи. Оцінюючи ці параметри можна провести виявлення факту вторгнення в систему. Саме на такому принципі ґрунтується робота сучасних СВП. Так система NIDES виконує аудит таких процесів як вхід у систему, робота з файлами та процесами, адміністрування та фіксація помилок та збоїв. У роботі [1] описані параметри, за якими здійснюється ідентифікація порушника розроблюваною системою. До них відносяться: Ім'я користувача при вході, *UID;* Час входу в систему, *Tlog;* Частота запитів на вхід у систему, *Nlog;* Час затрачений на вхід в систему, *TSlog;* Інтенсивність дій, *I;* Процесорний час/завантаженість процесора, *CPU;* Об'єм завантаженої оперативної пам'яті, *Muse;* Кількість виконуваних файлів, *NEF;* Тип використовуваних файлів при атаці, *AtEF;* Кількість збоїв та помилок, *NEr;* Час виконання процесу/файла, *RTPr/F;* Невластиві процеси, *UPr;* Передача файлу в систему, *TrFin;* зміна файлів, *ModF,* копіювання/передача фалів з системи, *TrFout;* Натиснення клавіш клавіатури, *KS* – хостові параметри та ряд мережевих параметрів – характеристики *ARP-, IP-, ICMP-* та *TCP*-пакетів.

Оскільки процес виявлення та ідентифікації порушника відбувається в умовах невизначеності, а ряд наведених параметрів СВП носять нечіткий характер, то функціонування такої системи має грунтуватись на нечіткій логіці. Для ідентифікації порушника можна використовувати логіко-лінгвістичний підхід і базову модель параметрів, частково описану в [1], які й будуть основою побудови розроблюваної СВП. Наприклад, для виявлення процесу сканування портів в роботах [2, 3] використовуються лінгвістичні зміні (ЛЗ) "Кількість віртуальних каналів" та "Вік віртуальних каналів", а в роботі [4] ЛЗ "Кількість одночасних підключень", "Швидкість обробки запитів", "Затримка між запитами" та "Кількість пакетів з однаковим адресом відправника та одержувача" - для виявлення DDOS-атак та спуфінгу.

Процес виявлення та ідентифікації порушника вимагає визначення необхідних параметрів і їх властивостей. В зв'язку з цим основною метою даної роботи є побудова моделей еталонів параметрів, необхідних для функціонування СВП в нечітко визначеному, слабоформалізованому середовищі.

Основна частина. Розглянемо метод лінгвістичних термів з використанням статистичних даних (МЛТС) [5,6], де в якості міри належності елемента множині приймається оцінка частоти використання поняття, яке задається нечіткою множиною для характеристики елемента. Для цього на універсальній шкалі [0;1] розміщуються значення лінгвістичної змінної (ЛЗ) $X=\{x_1, x_2, \ldots, x_n\}$. Метод грунтується на тій умові, що в кожний інтервал шкали потрапляє однакова кількість експериментів, але на практиці цього, як правило, не дотримуються. У реальних умовах складається емпірична таблиця, в якій експерименти можуть бути нерівномірно розподілені за інтервалами. Деякі з них можуть бути взагалі не задіяні, тоді дані обробляють за допомогою матриці підказок [6].

Нехай необхідно оцінити в значеннях ЛЗ відхилення параметра $\Delta B \in [0, B]$ (B - максимально можливе відхилення), яке характеризує поточні виміри. Далі для n = 5 визначимо значення ЛЗ {x1, x2,

x3, x4, x5}. Інтервал [0, B] і ΔB/B (оцінюване відношення) розділені на k відрізків (наприклад, 5), по якими збирається статистика, що характеризує частоту використання експертом значення ЛЗ для відображення своїх висновків. Далі дані заносять у таблицю і обробляються так, щоб зменшити похибки, внесені в процесі експерименту: з таблиці видаляються окремі елементи, по ліву сторону і по праву сторону від яких у рядку стоять нулі. Матриця підказок являє собою рядок, елементи якої обчислюють за формулою

$$k_j = \sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{5} b_{ij},\ j = \overline{1,5}. \quad (1)$$

Далі в отриманому рядку матриці вибирається максимальний елемент $k_{max} = \max k_j$, і потім всі елементи таблиці перетворюються за виразом

$$c_{ij} = b_{ij} k_{max}/k_j,\ i = \overline{1,5};\ j = \overline{1,5}, \quad (2)$$

а для стовпців, де $k_j = 0$ застосовується лінійна апроксимація $c_{ij} = (c_{ij-1} + c_{ij+1})/2,\ i = \overline{1,5};\ j = \overline{1,5}$. Далі обчислюється значення ФП за формулою

$$м_{ij} = c_{ij}/c_{imax},$$

де $c_{imax} = \max_j c_{ij},\ i = \overline{1,5};\ j = \overline{1,5}$. (3)

В описаному методі використовуються дані статистичних досліджень. Їх обробка досить трудомістка, оскільки для побудови функції приналежності (ФП) одного терму потрібно проводити статистичні дослідження всіх термів ЛЗ [5].

Побудуємо модель еталонів лінгвістичних змінних для нечітких параметрів ідентифікації порушника з множини параметрів, визначених в роботі [1].

Час входу в систему, Tlog. Даний параметр заснований на тому, що активність ІС і користувачів цих систем залежить від часу доби. Зазвичай більша активність користувачів щодо входу в систему виявляється в денний час, менша - в нічний, але можлива інша статистика, яка визначається режимом роботи організації, до яких належать ІС. Природа цього параметра нечітка, адже неможливо однозначно зробити висновок про нелегальну активність порушника. Так в організаціях з часом роботи з 08.00 до 16.00 імовірність того, що користувач, який авторизувався, - порушник найнижча в 08.00 і з часом зростає, досягаючи максимуму в години після 16.00. Однак, слід відмітити, що в концепції honeypot-технологій цей параметр дещо втрачає свою вагу, так як будь-яка активність на них вважається зловмисною.

Оцінимо ЛЗ "Рівень легітимності за часом". Визначимо значення лінгвістичної змінної {x_1, x_2, x_3}, що відповідають {легітимний, підозрілий, нелегітимний}. Тобто

$$T_{Tlog} = \bigcup_{i=1}^{3} T^{i}_{Tlog} = \{\ \text{легітимний, підозрілий, нелегітимний}\ \}.$$

Використаємо статистику за B=24 години. Доцільно загальний інтервал розбити на 4 інтервали [00:00;06:00], [06:00;12:00], [12:00;18:00], [18:00;24:00].

Дані для ЛЗ Tlog Таблиця 1

Значення ЛЗ	Інтервал			
	№1	№2	№3	№4
Високий	0	8	6	1
Середній	2	1	2	3
Низький	6	1	1	4

За допомогою виразу (1) визначимо $k_j = \|8\ 10\ 9\ 8\|$, де k_{max} = 10, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 0 & 8 & 6{,}66 & 1{,}25 \\ 2{,}5 & 1 & 2{,}22 & 3{,}75 \\ 7{,}5 & 1 & 1{,}11 & 5 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\| = \begin{Vmatrix} 0 & 1 & 0{,}83 & 0{,}16 \\ 0{,}66 & 0{,}26 & 0{,}59 & 1 \\ 1 & 0{,}13 & 0{,}15 & 0{,}66 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{3} \mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{3} ДВ_i/B = \{0{,}25;\ 0{,}5;\ 0{,}75;\ 1\}$ і отримуємо наступні нечіткі числа:

$Л = \{0/0{,}25;\ 1/0{,}5;\ 0{,}83/0{,}75;\ 0{,}16/1\}$,

$П = \{0{,}66/0{,}25;\ 0{,}26/0{,}5;\ 0{,}59/0{,}75;\ 1/1\}$,

$Н = \{1/0{,}25;\ 0{,}13/0{,}5;\ 0{,}15/0{,}75;\ 0{,}66/1\}$.

Графік ФП термів ЛЗ Час входу в систему показаний на рис. 1.

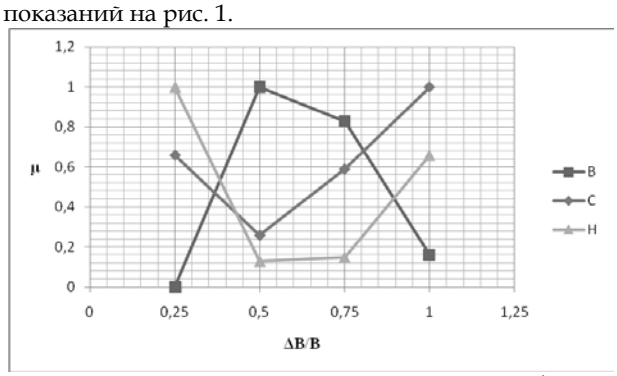

Рис. 1. Лінгвістичні еталони нечітких чисел для Tlog

Частота запитів на вхід у систему, Nlog. Зрозуміло, що найвища частота запитів на вхід буде відмічатися при атаках системи ботами (зокрема ботами-зломщиками, так як спамери не потребують входу в систему). Порушник-людина теж відзначається підвищеною частотою запитів внаслідок намагання обійти захист і теоретичного припущення що він не володіє легітимним логіном та паролем, тому буде змушений робити як мінімум декілька спроб. Причому чим більше число спроб, тим більша ймовірність що в ІС дійсно намагається ввійти порушник. Зрозуміло, що цей параметр теж є нечітким.

Оцінимо ЛЗ "Частота запитів на вхід у систему". Визначимо значення лінгвістичної змінної {x_1, x_2, x_3, x_4, x_5 }, що відповідають {низька, нижче середньої, середня, вище середньої, висока}. Тобто

$T_{Nlog} = \bigcup_{i=1}^{5} T^{i}_{Nlog} = \{$ низька, нижче середньої, середня, вище середньої, висока $\}$.

Частота запитів на вхід у систему звичайного користувача зазвичай мінімальна (найчастіше легітимний користувач вводить логін і пароль один раз), а сучасні програми для підбору паролів здатні перебрати 5310986 паролів/с [7]. Проте для визначення термів даної ЛЗ буде достатньо обмежитись значенням B=100 запитів/с, адже людина не здатна пройти процедуру аутентифікації більше 10-15 раз за хвилину. Доцільно загальний інтервал розбити на 5 інтервали [0;1], [1;10], [10;40], [40;80], [80;100].

Дані для ЛЗ Nlog Таблиця 2

Значення ЛЗ	Інтервал				
	№1	№2	№3	№4	№5
Низька	8	0	0	0	0
Нижче середньої	5	2	0	0	0
Середня	1	6	4	0	0
Вище середньої	0	2	8	1	0
Висока	0	0	1	6	6

За допомогою виразу (1) визначимо $k_j = \|14 \quad 10 \quad 13 \quad 7 \quad 6\|$, де $k_{max} = 14$, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 8 & 0 & 0 & 0 & 0 \\ 5 & 2{,}8 & 0 & 0 & 0 \\ 1 & 8{,}4 & 4{,}31 & 0 & 0 \\ 0 & 2{,}8 & 8{,}62 & 2 & 0 \\ 0 & 0 & 1{,}08 & 12 & 16 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0{,}56 & 0 & 0 & 0 \\ 0{,}12 & 1 & 0{,}51 & 0 & 0 \\ 0 & 0{,}32 & 1 & 0{,}23 & 0 \\ 0 & 0 & 0{,}07 & 0{,}75 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{5} \mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{5} ДВ_i/В = \{0{,}01;\ 0{,}1;\ 0{,}4;\ 0{,}8;\ 1\}$ і отримуємо наступні нечіткі числа:

$H = \{1/0{,}01;\ 0/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\}$,

$НС = \{1/0{,}01;\ 0{,}56/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\}$,

$С = \{0{,}12/0{,}01;\ 1/0{,}1;\ 0{,}51/0{,}4;\ 0/0{,}8;\ 0/1\}$,

$ВС = \{0/0{,}01;\ 0{,}32/0{,}1;\ 1/0{,}4;\ 0{,}23/0{,}8;\ 0/1\}$,

$В = \{0/0{,}01;\ 0/0{,}1;\ 0{,}07/0{,}4;\ 0{,}75/0{,}8;\ 1/1\}$.

Графік ФП термів ЛЗ Частота запитів на вхід у систему показаний на рис. 2.

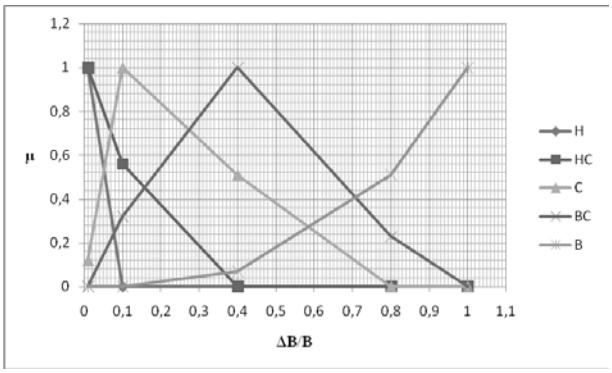

Рис. 2. Лінгвістичні еталони нечітких чисел для Nlog

Час затрачений на вхід в систему, TSlog. Параметр, який тісно пов'язаний з попереднім. Час, затрачений порушником, у більшості випадків більший за час, затрачений легітимним користувачем. Але він є нечітким, оскільки не дає змоги провести однозначну ідентифікацію.

Оцінимо ЛЗ "Час затрачений на вхід в систему". Визначимо значення лінгвістичної змінної $\{x_1, x_2, x_3, x_4, x_5\}$, що відповідають {дуже малий, малий, середній, великий, дуже великий}. Тобто $T_{Slog} = \bigcup_{i=1}^{5} T^{i}_{Slog} = \{$ дуже малий, малий, середній, великий, дуже великий $\}$.

Легітимний користувач у ІС, захищеній паролем, витрачає на ідентифікацію від декількох секунд до декількох хвилин. Натомість затрати часу нелегітимних користувачів на злам парольної системи порівняно великі. Так сучасні системи підбору паролів для зламу 8-символьного паролю, що складається з комбінації букв, цифр та спеціальних знаків, витрачають до 61 доби [7]. Тому візьмемо значенням B=60 діб = 5184000 с,. Доцільно загальний інтервал розбити на 5 інтервали [0 с; 30 с], [30 с; 5 хв], [5 хв; 1 год], [1 год; 1 доба], [1 доба; 60 діб].

Дані для ЛЗ TSlog Таблиця 3

Значення ЛЗ	Інтервал				
	№1	№2	№3	№4	№5
Дуже малий	9	3	0	0	0
Малий	5	10	1	0	0
Середній	1	7	5	0	0
Великий	0	1	2	9	2
Дуже великий	0	0	1	6	9

За допомогою виразу (1) визначимо $k_j = \|14 \quad 21 \quad 9 \quad 15 \quad 11\|$, де $k_{max} = 21$, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 13{,}5 & 3 & 0 & 0 & 0 \\ 7{,}5 & 10 & 2{,}33 & 0 & 0 \\ 1{,}5 & 7 & 11{,}67 & 0 & 0 \\ 0 & 1 & 4{,}67 & 12{,}6 & 3{,}82 \\ 0 & 0 & 2{,}33 & 8{,}4 & 17{,}18 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\left\|\mu_{ij}\right\| = \begin{Vmatrix} 1 & 0,22 & 0 & 0 & 0 \\ 0,75 & 1 & 0,23 & 0 & 0 \\ 0,13 & 0,6 & 1 & 0 & 0 \\ 0 & 0,08 & 0,37 & 1 & 0,3 \\ 0 & 0 & 0,14 & 0,49 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{5}\mu_{ij}$ відповідно знаходимо оціночні відношення

$\bigcup_{i=1}^{5}\text{ДВ}_i/\text{В} = \{5{,}79\cdot10^{-6};\ 5{,}79\cdot10^{-5};\ 6{,}94\cdot10^{-4};\ 0{,}02;\ 1\}$ і отримуємо наступні нечіткі числа:

$\text{ДМ} = \{1/5{,}79\cdot10^{-6};\ 0{,}22/5{,}79\cdot10^{-5};\ 0/6{,}94\cdot10^{-4};\ 0/0{,}02;\ 0/1\},$

$\text{М} = \{0{,}75/5{,}79\cdot10^{-6};\ 1/5{,}79\cdot10^{-5};\ 0{,}23/6{,}94\cdot10^{-4};\ 0/0{,}02;\ 0/1\},$

$\text{С} = \{0{,}13/5{,}79\cdot10^{-6};\ 0{,}6/5{,}79\cdot10^{-5};\ 1/6{,}94\cdot10^{-4};\ 0/0{,}02;\ 0/1\},$

$\text{В} = \{0/5{,}79\cdot10^{-6};\ 0{,}08/5{,}79\cdot10^{-5};\ 0{,}37/6{,}94\cdot10^{-4};\ 1/0{,}02;\ 0{,}3/1\},$

$\text{ДВ} = \{0/5{,}79\cdot10^{-6};\ 0/5{,}79\cdot10^{-5};\ 0{,}14/6{,}94\cdot10^{-4};\ 0{,}49/0{,}02;\ 1/1\}.$

Графік ФП термів ЛЗ показаний на рис. 3.

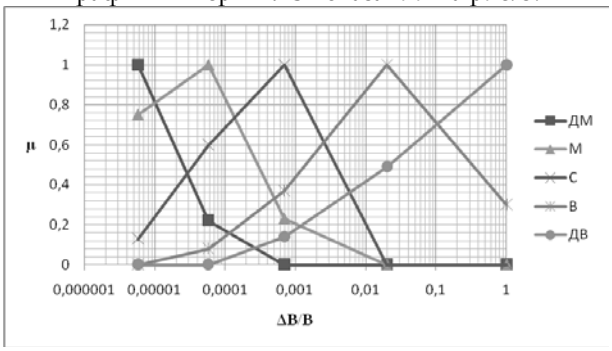

Рис. 3. Лінгвістичні еталони нечітких чисел для TSlog

Інтенсивність дій, I. Тут розуміється кількість будь-яких дії користувача, що включають в себе вхід/вихід з системи, передача, зміна, копіювання файлів, запуск/припинення процесів тощо, в одиницю часу. Інтенсивність може не відрізнятися в порушника-людини і в легітимного користувача, однак у ботів вона значно вища, тому найбільш суттєва для ідентифікації та розмежування категорій людина-бот. Хоча значне перевищення норми вказує на діяльність неавторизованих автоматичних систем-порушників (ботів), проте I – нечіткий параметр, оскільки нормальну величину показника інтенсивності визначити дуже важко.

Оцінимо ЛЗ " Інтенсивність дій ". Визначимо значення лінгвістичної змінної $\{x_1, x_2, x_3\}$, що відповідають {низька, середня, висока}.

Тобто $T_I = \bigcup_{i=1}^{3} T_I^i = \{\text{низька, середня, висока}\}$.

Інтенсивність дій звичайної людини дуже низька, зазвичай це від 3 до 10 дій на хвилину. Інтенсивність ж ботів в десятки разів більша. Для дослідження візьмемо верхню межу у 100 дій/хв, хоча у роботів цей показник може бути і більший. Доцільно загальний інтервал розбити на 4 інтервали [0;5], [5;10], [10;50], [50;100].

Дані для ЛЗ I Таблиця 4

Значення ЛЗ	Інтервал			
	№1	№2	№3	№4
Низька	7	5	1	0
Середня	0	7	4	0
Висока	0	1	5	7

За допомогою виразу (1) визначимо $k_j = \|7\ 13\ 10\ 7\|$, де k_{max} = 13, і відповідно до (2) обчислимо:

$$\left\|c_{ij}\right\| = \begin{Vmatrix} 13 & 5 & 1,3 & 0 \\ 0 & 7 & 5,2 & 0 \\ 0 & 1 & 6,5 & 13 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\left\|\mu_{ij}\right\| = \begin{Vmatrix} 1 & 0,38 & 0,1 & 0 \\ 0 & 1 & 0,74 & 0 \\ 0 & 0,08 & 0,5 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{3}\mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{3}\text{ДВ}_i/\text{В} = \{0{,}05;\ 0{,}1;\ 0{,}5;\ 1\}$ і отримуємо наступні нечіткі числа:

$\text{Н} = \{1/0{,}05;\ 0{,}38/0{,}1;\ 0{,}1/0{,}5;\ 0/1\},$

$\text{С} = \{0/0{,}05;\ 1/0{,}1;\ 0{,}74/0{,}5;\ 0/1\},$

$\text{В} = \{0/0{,}05;\ 0{,}08/0{,}1;\ 0{,}5/0{,}5;\ 1/1\}.$

Графік ФП термів ЛЗ Інтенсивність дій показаний на рис. 4.

Рис. 4. Лінгвістичні еталони нечітких чисел для I

Процесорний час/завантаженість процесора, CPU. Оскільки кількість активних процесів на honeypot-системах мусить бути мінімальною, то будь-яке збільшення навантаження є ознакою діяльності порушника в системі. В реальних ІС імовірність того, що активність спричинена саме порушником дещо нижча, а, зрозуміло, нормальна величина процесорного часу вища. Проте цей параметр все одно можна ефективно використовувати для ідентифікації факту порушення в системах виявлення вторгнень і СВП. Оскільки однозначну відповідь про порушника за даним параметром дати неможливо, впершу чергу через можливу діяльність вірусів, то процесорний CPU - нечіткий параметр.

Оцінимо ЛЗ "Процесорний час/завантаженість процесора". Визначимо значення

лінгвістичної змінної {x_1, x_2, x_3}, що відповідають {низька, середня, висока}. Тобто

$$T_{CPU} = \bigcup_{i=1}^{3} T^{i}_{CPU} = \{\text{низька, середня, висока}\}.$$

В нормальних умовах середньої експлуатації потужностей ПК і при відсутності впливу порушників чи шкідливого ПЗ середній показник завантаженості процесора становить 20-30%. Звичайно норма може дещо варіюватися залежно від ОС, встановленого ПЗ і виробничих завдань організації. Максимально можливий відсоток завантаженості CPU B=100%. Доцільно загальний інтервал розбити на 4 інтервали [0;20], [20;50], [50;75], [50;100].

Дані для ЛЗ CPU Таблиця 5

Значення ЛЗ	Інтервал			
	№1	№2	№3	№4
Низька	9	6	0	0
Середня	3	8	1	0
Висока	0	1	5	8

За допомогою виразу (1) визначимо $k_j = \|12\ 15\ 6\ 8\|$, де k_{max} = 15, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 11,25 & 6 & 0 & 0 \\ 3,75 & 8 & 2,5 & 0 \\ 0 & 1 & 12,5 & 15 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,53 & 0 & 0 \\ 0,47 & 1 & 0,31 & 0 \\ 0 & 0,07 & 0,83 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{3} \mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{3} ДВ_i/В = \{0,2;\ 0,5;\ 0,75;\ 1\}$, отримуємо наступні нечіткі числа:

$$H = \{1/0,2;\ 0,53/0,5;\ 0/0,75;\ 0/1\},$$

$$C = \{0,47/0,2;\ 1/0,5;\ 0,31/0,75;\ 0/1\},$$

$$B = \{0/0,2;\ 0,07/0,5;\ 0,83/0,75;\ 1/1\}.$$

Графік ФП термів ЛЗ Процесорний час/завантаженість процесора показаний на рис. 5.

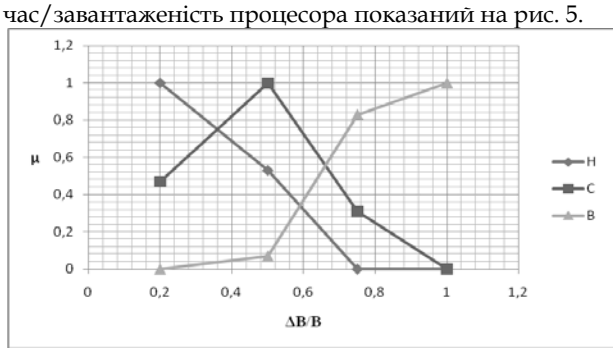

Рис. 5. Лінгвістичні еталони нечітких чисел для CPU

Об'єм завантаженої оперативної пам'яті, Muse. Аналогічний за суттю до попереднього і також є нечітким. ФП для даного параметра є практично ідентичною з ЛЗ "Процесорний час/завантаженість процесора".

Кількість виконуваних файлів/процесів, NEF. Також входить в групу нечітких параметрів. Факт дій зловмисника по цьому параметру визначається відхиленням від норми. Цей параметр включає в себе лише процеси та файли користувача, а системні - не враховуються. Так в кожній організації відповідно до політики безпеки та посадових обов'язків кожен легітимний користувач може використовувати певні файли в певний момент, причому одночасне використання одразу багатьох файлів чи процесів практично виключається. Це дає змогу виявити як зовнішнього, так і внутрішнього порушника, але з певною ймовірністю.

Оцінимо ЛЗ "Кількість виконуваних файлів". Визначимо значення лінгвістичної змінної {x_1, x_2, x_3}, що відповідають {дуже мала, мала, нормальна, велика, дуже велика}. Тобто

$$T_{NEF} = \bigcup_{i=1}^{5} T^{i}_{NEF} = \{\text{ дуже мала, мала,нормальна, велика, дуже велика }\}.$$

Нормальний показник кількості виконуваних файлів, як вже зазначалося, дуже залежний від галузі, в якій працює певна ІС, та від набору правил політики безпеки і посадових інструкцій організації. В основному це значення коливається від 7 до 10 процесів. Враховуючи також і технологічні обмеження ІС і типові політики безпеки для дослідження візьмемо максимальне значення в 20 процесів користувачів. Доцільно загальний інтервал розбити на 4 інтервали [0;4], [4;8], [8;12], [12;16], [16;20].

Дані для ЛЗ NEF Таблиця 6

Значення ЛЗ	Інтервал				
	№1	№2	№3	№4	№5
Дуже мала	8	5	1	0	0
Мала	4	7	2	0	0
Нормальна	1	4	9	3	1
Велика	0	0	4	8	2
Дуже велика	0	0	1	5	8

За допомогою виразу (1) визначимо $k_j = \|13\ \ 16\ \ 17\ \ 16\ \ 11\|$, де k_{max} = 17, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 10,46 & 5,31 & 1 & 0 & 0 \\ 5,23 & 7,44 & 2 & 0 & 0 \\ 1,31 & 4,25 & 9 & 3,19 & 1,55 \\ 0 & 0 & 4 & 8,5 & 3,09 \\ 0 & 0 & 1 & 5,31 & 12,36 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,51 & 0,1 & 0 & 0 \\ 0,7 & 1 & 0,27 & 0 & 0 \\ 0,15 & 0,47 & 1 & 0,35 & 0,17 \\ 0 & 0 & 0,47 & 1 & 0,36 \\ 0 & 0 & 0,08 & 0,43 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{5}\mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{5}\text{ДВ}_i/\text{В}=\{0,2;\ 0,4;\ 0,6;\ 0,8;\ 1\}$ і отримуємо наступні нечіткі числа:

$$\text{ДМ}=\{1/0,2;\ 0,51/0,4;\ 0,1/0,6;\ 0/0,8;\ 0/1\},$$

$$\text{М}=\{0,7/0,2;\ 1/0,4;\ 0,27/0,6;\ 0/0,8;\ 0/1\},$$

$$\text{Н}=\{0,15/0,2;\ 0,47/0,4;\ 1/0,6;\ 0,35/0,8;\ 0,17/1\},$$

$$\text{В}=\{0/0,2;\ 0/0,4;\ 0,47/0,6;\ 1/0,8;\ 0,36/1\},$$

$$\text{ДВ}=\{0/0,2;\ 0,0/0,4;\ 0,08/0,6;\ 0,43/0,8;\ 1/1\}.$$

Графік ФП термів ЛЗ Кількість виконуваних файлів/процесів показаний на рис. 6.

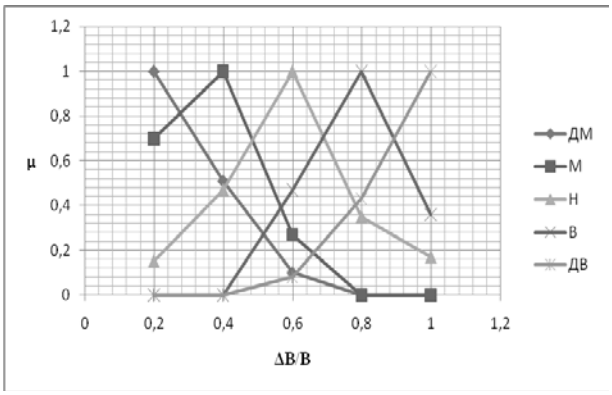

Рис. 6. Лінгвістичні еталони нечітких чисел для NEF

Кількість збоїв та помилок, NEr. Даний параметр є нечітким, оскільки збої та помилки можуть відбуватися під час роботи як авторизованого користувача так і порушника. Проте при частому повторені збоїв чи помилок можна зробити висновок з певною долею імовірності що система атакована. В цю групу входить широкий спектр подій від помилок при авторизації до збоїв при виконанні певних процесів або файлів. При активній роботі порушника, незалежно від його класу та категорії, частота появи несправностей буде дещо вищою. Слід також відмітити, що цілком можливо при ідентифікації порушника-робота ця частота буде ще вищою.

Оцінимо ЛЗ "Кількість збоїв та помилок". Визначимо значення лінгвістичної змінної $\{x_1, x_2, x_3\}$, що відповідають {низька, середня, висока}.

Тобто $\text{T}_{\text{NEr}}=\bigcup_{i=1}^{3}\text{T}^{\text{i}}{}_{\text{NEr}}=\{\text{низька, середня, висока}\}$.

Функціональність ІС вважається нормальною, якщо в її роботі відсутні помилки та збої взагалі. Проте поява невеликої кількості збоїв все ж можлива, в основному через недостатню кваліфікацію користувача, його халатність або застосування неякісного апаратного / програмного забезпечення. Офіційної статистики з даного питання немає, тому складно визначити нормальну величину цього параметру. В рамках дослідження встановимо максимальну кількість помилок та збоїв за добу В=10. Доцільно загальний інтервал розбити на 4 інтервали [0;1], [1;4], [4;8], [8;10].

Дані для ЛЗ NEr Таблиця 7

Значення ЛЗ	Інтервал			
	№1	№2	№3	№4
Низька	5	1	0	0
Середня	0	4	3	0
Висока	0	0	1	6

За допомогою виразу (1) визначимо $k_j=\|5\ 5\ 4\ 6\|$, де k_{max} = 6, і відповідно до (2) обчислимо:

$$\|c_{ij}\|=\begin{Vmatrix}6 & 1,2 & 0 & 0\\ 0 & 4,8 & 4,5 & 0\\ 0 & 0 & 1,5 & 6\end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\|=\begin{Vmatrix}1 & 0,2 & 0 & 0\\ 0 & 1 & 0,94 & 0\\ 0 & 0 & 0,25 & 1\end{Vmatrix}.$$

Для $\bigcup_{i=1}^{3}\mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{3}\text{ДВ}_i/\text{В}=\{0,1;\ 0,4;\ 0,8;\ 1\}$, отримуємо наступні нечіткі числа:

$$\text{Н}=\{1/0,1;\ 0,2/0,4;\ 0/0,8;\ 0/1\},$$

$$\text{С}=\{0/0,1;\ 1/0,4;\ 0,94/0,8;\ 0/1\},$$

$$\text{В}=\{0/0,1;\ 0/0,4;\ 0,25/0,8;\ 1/1\}.$$

Графік ФП термів ЛЗ Кількість збоїв та помилок показаний на рис. 7.

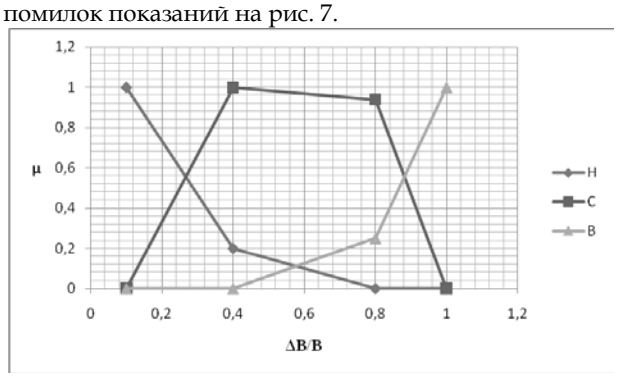

Рис. 7. Лінгвістичні еталони нечітких чисел для NEr

Час виконання процесу/файлу, RTPr/F. Досліджуючи статистику роботи ІС різних підприємств та організацій легко помітити, що залежно від специфіки роботи час, затрачений на виконання певної операції є приблизно однаковий для однотипних ІС та їх задач. В honeypot-системах в основному виконуються системні процеси, тобто ті, що підтримують роботу самого honeypot'а, або процеси адміністратора, що запускаються в певний час на деякий період. Таким чином при ідентифікації таких процесів можна зробити висновок про атаку системи порушником. Оскільки, такий стан речей може бути спричинений халатністю працівника, то висновок є неоднозначний і відповідно параметр носить нечіткий характер.

Оцінимо ЛЗ "Час виконання процесу/файл". Визначимо значення лінгвістичної змінної $\{x_1, x_2, x_3,$

x_4, x_5 }, що відповідають {дуже малий, малий, середній, великий, дуже великий}. Тобто $T_{RTPr/F} = \bigcup_{i=1}^{5} T^i_{RTPr/F} = \{$ дуже малий, малий, середній, великий, дуже великий $\}$.

Легітимний користувач у ІС в процесі виконання своїх посадових обов'язків працює з певним файлом чи процесом в процесі певного часу. Так середньостатистичний працівник працює з одним файлом чи процесом період часу від 30 хвилин до 3 годин. Якщо цей показник значно менший чи більший, то це може свідчити про підозрілу активність. Доцільно встановити максимальне значення даної змінної B=24 години загальний інтервал розбити на 5 інтервали [0 с; 1 хв], [1 хв; 30 хв], [30 хв; 3 год], [3 год; 6 год], [6 год; 24 год].

Дані для ЛЗ RTPr/F Таблиця 8

Значення ЛЗ	Інтервал				
	№1	№2	№3	№4	№5
Дуже малий	9	4	1	0	0
Малий	5	7	2	0	0
Середній	0	3	8	3	0
Великий	0	0	3	9	6
Дуже великий	0	0	1	4	9

За допомогою виразу (1) визначимо $k_j = \|14 \quad 14 \quad 15 \quad 16 \quad 15\|$, де k_{max} = 16, і відповідно до (2) обчислимо:

$$\|c_{ij}\| = \begin{Vmatrix} 10,29 & 4,57 & 1,07 & 0 & 0 \\ 5,71 & 8 & 2,13 & 0 & 0 \\ 0 & 3,43 & 8,53 & 3 & 0 \\ 0 & 0 & 3,2 & 9 & 6,4 \\ 0 & 0 & 1,07 & 4 & 9,6 \end{Vmatrix}.$$

Обчислимо ФП за формулою (3):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,44 & 0,1 & 0 & 0 \\ 0,71 & 1 & 0,27 & 0 & 0 \\ 0 & 0,4 & 1 & 0,35 & 0 \\ 0 & 0 & 0,36 & 1 & 0,71 \\ 0 & 0 & 0,11 & 0,42 & 1 \end{Vmatrix}.$$

Для $\bigcup_{i=1}^{5} \mu_{ij}$ відповідно знаходимо оціночні відношення $\bigcup_{i=1}^{5} ДВ_i/В = \{6,94 \cdot 10^{-4};\ 0,02;\ 0,125;\ 0,25;\ 1\}$ і отримуємо наступні нечіткі числа:

$ДМ = \{\ 1/6,94 \cdot 10^{-4};\ 0,44/0,02;\ 0,1/0,125;\ 0/0,25;\ 0/1\ \}$,

$М = \{\ 0,71/6,94 \cdot 10^{-4};\ 1/0,02;\ 0,27/0,125;\ 0/0,25;\ 0/1\ \}$,

$С = \{\ 0/6,94 \cdot 10^{-4};\ 0,4/0,02;\ 1/0,125;\ 0,35/0,25;\ 0/1\ \}$,

$В = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,36/0,125;\ 1/0,25;\ 0,71/1\ \}$,

$ДВ = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,11/0,125;\ 0,42/0,25;\ 1/1\ \}$.

Графік ФП термів ЛЗ Час виконання процесу/файлу показаний на рис. 8.

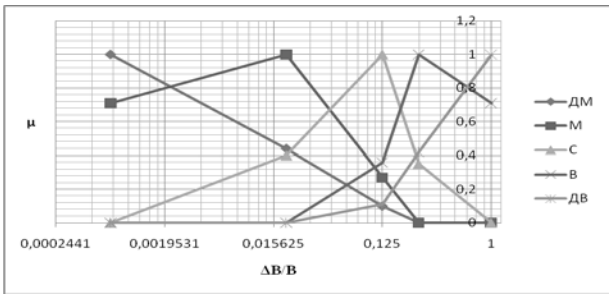

Рис. 8. Лінгвістичні еталони нечітких чисел для RTPr/F

Висновки

Таким чином, у роботі з використанням МЛТС були введені ЛЗ та побудовані моделі еталонів параметрів *Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F*. Також для кожної ЛЗ були розраховані ФП та побудовані графіки їх термів. Сформовані еталони необхідні для формування логічних правил, що дозволяють забезпечити функціонування СВП. Отримані результати у подальшому будуть використані для побудови IDS системи на базі технології honeypot.

Література

[1] Гізун А.І. Основні параметри для ідентифікації порушника інформаційної безпеки / А.І. Гізун, В.В. Волянська, В.О. Риндюк, С.О. Гнатюк // Захист інформації. — 2013. — №1 (58). — С.66-75.

[2] Корченко А.О. Система виявлення аномалій на основі нечітких моделей / А.О. Корченко, Є.В. Паціра, В.В. Волянська // Сучасні тренажерно-навчальні комплекси та системи : Зб. наук. праць. – Л.: Інституту проблем моделювання в енергетиці НАН України ім. Г.Є. Пухова, 2007. – Т.2. – С. 56 – 60.

[3] Програма захисту інформаційних ресурсів від атакуючих дій в комп'ютерних мережах : Комп'ютерна програма / Васюхін М.І., Гулевець В.Д., Корченко А.О., та інші — К. : Інститут кібернетики ім. В.М. Глушкова НАНУ, 2011. — Свідоцтво про реєстрацію авторського права на твір №37127 від 25.02.2011.

[4] Луцкий М.Г. Модели эталонов лингвистических переменных для систем выявления атак / М.Г. Луцький, А.В. Гавриленко, А.А. Корченко, А.А. Охрименко // Захист інформації. – 2012. –№2 (55). — С. 5-13

[5] Корченко А.Г. Построение систем защиты информации на нечетких множествах : Теория и практические решения / А.Г.Корченко. — К. : МК-Пресс, 2006. — 320 с.

[6] Сваровкий С.Т. Аппроксимация функций принадлежности значений лингвистической переменной / С.Т. Сваровский // Мат. вопр. анализа данных. — Новосибирск: ВЦ СО АН СССР — 1980. — С.127–131.

[7] Голуб В. Парольная защита [Електронний ресурс]: стаття / В. Голуб // Relga. – 01.12. 2009. – №17 (197). — Режим доступу: http://www.relga.ru/Environ/WebObjects/tgu-www.woa/wa/Main?textid=2516&level1=main&level2=articles

УДК 004.056.53:004.492.3 (045)

Гизун А.И., Волянская В.В., Гнатюк В.А. Модели эталонов лингвистических переменных для систем обнаружения и идентификации нарушителя информационной безопасности
Аннотация. *Изучение факта нарушения информационной безопасности и личности нарушителя имеет большое научное и практическое значение. С этих позиций, формализация параметров, которые могут быть использованы для идентификации нарушителей, является актуальной научной задачей. Четкое определение полного множества параметров позволит повысить эффективность превентивных мер и систем защиты. В работе предложена модель эталонов лингвистических переменных, ориентированная на построение системы обнаружения нарушителя (вторжение). Для обнаружения нарушителя используется ряд хостовых и сетевых параметров, большинство из которых имеют нечеткую природу. На основе проведенного эксперимента построены модели эталонов этих параметров с использованием нечетких чисел. Полученные результаты могут быть базисом для построения системы обнаружения вторжений на основе технологии honeypot.*
Ключевые слова: *нарушитель информационной безопасности, система обнаружения нарушителя, параметры, идентификация, лингвистические переменные, нечеткая логика, эталоны параметров.*

Gizun A.I., Volyanska V.V., Gnatyuk V.O. Etalon models of linguistic variables for information security intruders' detection and identification
Abstract. *Study of information security breach and intruder identity has a great scientific and practical importance. From this viewpoint parameters formalization for intrude identification is an actual research problem. Clearly definition of complete set parameters can give a possibility to increase preventive measures and security systems efficiency. In the paper etalon model of linguistic variables was proposed and oriented on IDS system construction. For intruder detection the set of host and network parameters are used. Most of them have a fuzzy nature. The etalon models were build using fuzzy numbers on basis of experiment. Given results can be the basis for IDS system based on honeypot-technology development.*
Key words: *information security intruder, intruder detection system, parameters, identification, linguistic variables, fuzzy logic, etalon parameters.*

Отримано 23 *січня 2012 року*, затверджено редколегією 27 *лютого 2013 року*

DATA PROTECTION FROM NETWORK ATTACKS

Gulnur Zhangissina, Erjan Kuldeev, Ajgul Shayhanova

Kazakh National Technical University named after K.I.Satpayev, Kazakhstan

ZHANGISSINA Gulnur D., Doctor of Science (DSc), Professor

Date and place of birth: 1958, Almaty, Kazakhstan.
Education: Kazakh National University named after Al-Farabi, 1980.
Research interests: information security, computer security, parallel computing, information systems and distance education.
Current position & Functions: Head of the Computer Science Department.
Publications: more than 200 publications, including 4 monographs, over 20 books & 180 papers.
E-mail: gul_zhd@mail.ru

KULDEEV Erjan I., PhD, Professor

Date and place of birth: 1973, Aralsk, Kyzyl-Orda region.
Education: Geological Department of the Kazakh Polytechnic Institute named after Lenin, in "Geophysical methods of exploration" for qualified mining engineer-geophysicist.
Research interests: information security, computer security, GIS.
Current position & Functions: Vice Rector for science and innovation.
E-mail: kuldeev@ntu.kz