Система моделювання процесу зараження мережі та її протидії

Автор: студент 2-го курсу групи КА-83
Гаркавенко Денис
Науковий керівник: Ігнатенко О.П.

Актуальність роботи

Актуальність задачі полягає у недосконалості існуючих систем виявлення вторгнень та відсутності досліджень на тему моделювання атак на систему. Метою дисертації була розробка системи моделювання атак на мережу та системи протидії цим атакам.

Постановка задачі магістерської роботи

Проаналізувати існуючі методи забезпечення мережевої безпеки.

Оцінити існуючі підходи моделювання сценаріїв зараження мережі.

Реалізувати системи для моделювання дій та стратегій порушника і захищаючого.

Розробити систему моделювання процесів зараження мережі та її протидій.

Метою запропонованої системи є моделювання реальних загроз безпеці комп’ютерної мережі та створення системи протидії цим загрозам.

Предметом дослідження є побудова теоріко-ігрової моделі протидії системи виявлення вторгнень процесу атаки на мережу.

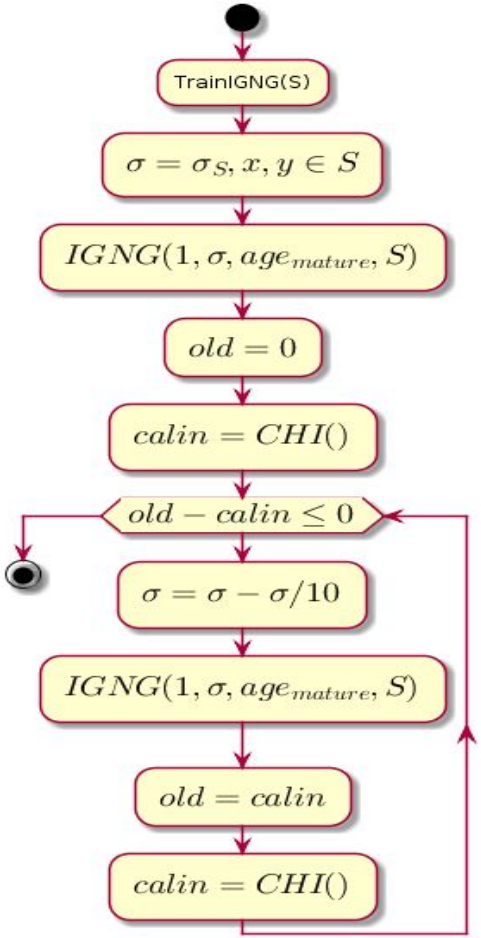
Об’єктом дослідження є створення системи виявлення вторгнень на основі даних нормальної та аномальної роботи мережі.

Алгоритм роботи детектора.

CHI - індекс Калінського-Харабаза, що показує якість кластерізації.

$$CHI = \frac{B/(c-1)}{W/(n-c)}$$

n - кількість семплів даних.
c - кількість кластерів.
B - матриця внутрішній дисперсії.
W - матриця зовнішньої дисперсії.



Алгоритм роботи реалізованої СВВ: IGNG

Умови останови алгоритма:

- Не знайдено жодних нейронів.
- Знайдено один нейрон-переможець.
- Знайдено два нейрона-переможця.

Цикл роботи алгоритму починається з пустого графу. Параметр ідентифікується як середньоквадратичне відхилення по навчальній вибі|

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Основний цикл зменшує значення , що є порогом близькості, та розраховує різницю між попереднім рівнем якості кластерізації та рівнем, який отримали у результаті кластерізації процедурою IGNG.

Алгоритм роботи реалізованої СВВ: IGNG

На першому кроці виконується пошук нейрона, що найкраще наближає семпл даних:

$$c_1 = min\ (dist(\xi, w_c)$$

Якщо не було знайдено жодного задовільняючого умові нейрона – створюють новий неактивний нейрон з координатами семплу у просторі даних. Якщо такий нейрон було знайдено, то проводиться пошук другого нейрону аналогічним чином. Якщо його немає – він створюється.

Алгоритм роботи реалізованої СВВ: IGNG

Якщо знайдено два нейрона, що задовольняють умовам, їх координати корегуються наступним чином:

$$\epsilon(t)h_{c,c_i} = \begin{cases} \epsilon_b, \text{якщо } c = c_i \\ \epsilon_n, \text{якщо є зв'язок } c \text{ -- } c_i \\ 0, \text{у інших випадках} \end{cases}$$

$$\Delta w_c = \epsilon(t)h_{c,c_i}\|\xi - w_c\|$$

$$w_c = w_c + \Delta w_c$$

На наступному кроці створюється чи оновлюється дуга між нейронами-переможцями, а вік усіх інших дуг зростає.

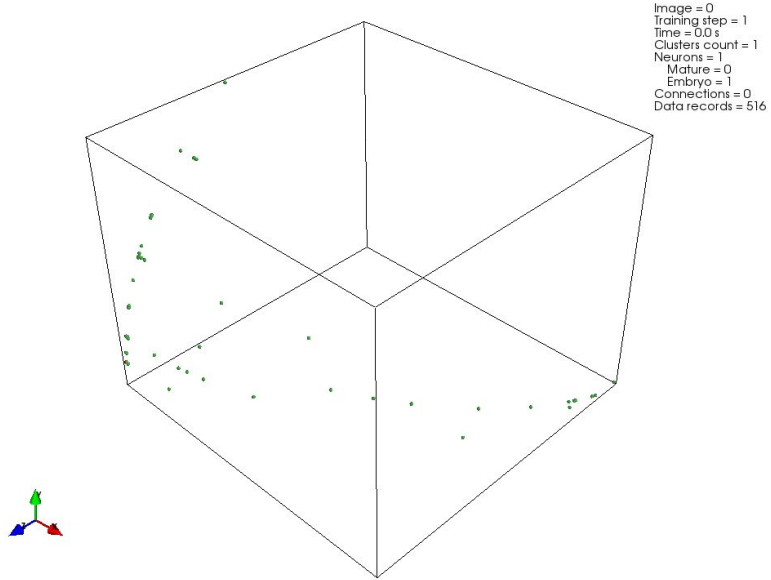
Видаляють усі дуги, якщо їх вік перевищує максимально допустимий вік .

Видаляють усі активні нейрони, з котрих не виходять ніякі дуги, а вік усіх нейронів-сусідів нейрона-переможця зростає.

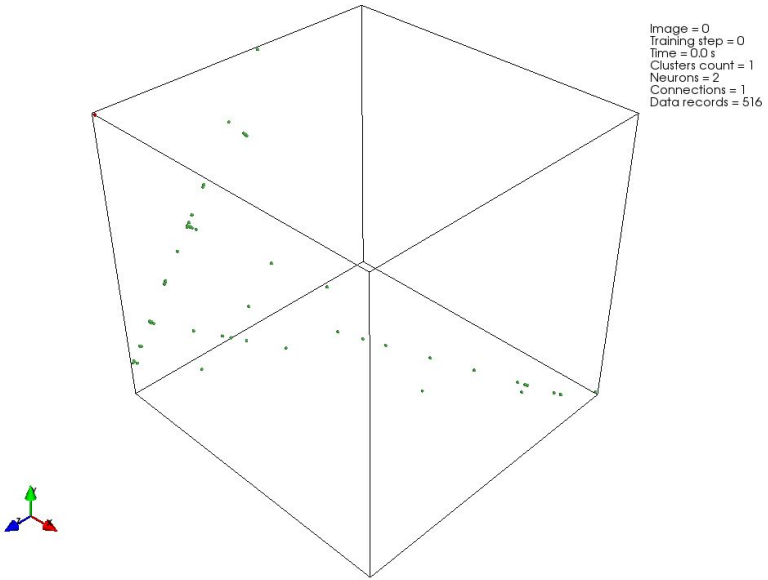
Якщо вік неактивного нейрона перевищує максимально допустимий вік , він становиться активним.

Процес навчання СВВ

Incremental Growing Neuron Gas for the network anomalies detection



Growing Neural Gas for the network anomalies detection



Результати роботи СВВ

```
----------------------------------------------------------------------
Applying detector to the normal activity using the testing set without adaptive learning...
----------------------------------------------------------------------
Reading normal activity from the file "NSL_KDD/KDDTest-21.txt" [generated host data was included]...
Records count: 2152
Abnormal records = 1, Normal records = 0, Detection time = 0.0 s, Time per record = 0 s
Abnormal records = 165, Normal records = 836, Detection time = 1.13 s, Time per record = 0.00113072776794 s
Abnormal records = 343, Normal records = 1658, Detection time = 2.25 s, Time per record = 0.00112573599815 s
Anomalies weren't detected [abnormal records = 371, normal records = 1781, detection time = 2.42 s, time per record = 0.00112456171929 s]
```

Результати роботи СВВ

Тип	l_time	te_l_time	te_t_time	g_l_perc	g_t_perc	f_l_perc	f_t_perc
GNG без host	1869	0.39	4.61	69.5	78.4	0	36.9
GNG з host	1839	0.42	4.72	60.2	72.5	0	47.7
IGNG без host	551	1.02	11.97	22.2	26.8	0	17.3
IGNG з host	567	1.14	13.34	28.9	27.0	0	20.8

l_time - Час навчання в секундах.

te_l_time - Час перевірки в секундах на наборі, з якого була сформована навчальна вибірка.

te_t_time - Час перевірки в секундах на повному наборі даних тестової вибірки.

g_l_perc - Відсоток знайдених аномалій для повного набору, з якого була сформована навчальна вибірка.

g_t_perc - Відсоток знайдених аномалій для повного набору даних тестової вибірки.

f_l_perc - Відсоток помилкових спрацьовувань для повного набору, з якого була сформована навчальна вибірка.

f_t_perc - Відсоток помилкових

Алгоритм роботи системи моделювання

До виявлення підозрілої активності система виявлення вторгнень спостерігає базовий набір критичних параметрів.

$$S(t) = \sum_{\omega \in \Omega} \sum_{j \in J_\omega} s(\omega, j, t) :$$

Функція корисності СВВ та атакуючого має наступний вигляд:

$$U^1_{mon} = \begin{cases} \alpha - R(t_m), & 0 \le t_a < t_m, NA, \\ -\alpha - R(t_m), & 0 \le t_m < t_a. \end{cases}$$

f - середній ваговий коефіцієнт. Визначає ціну одного спостережуваного параметра.

k - кількість спостережуваних пар «об'єкт - параметр».

S (t) - ціна додаткових ресурсів.

α - очікуваний виграш СВВ.

Результати роботи системи моделювання

	a_1	a_2	a_3	a_4
b_1	0; 5000	0; 4950	0; 4500	0; 2000
b_2	2000; -5000	-2000; 4950	-2000; 4500	-2000; 2000
b_3	1800; -5000	1800; -5050	-2200; 4500	-2200; 2000
b_4	1500; -5000	1500; -5050	1500; -5500	-2500; 2000
b_5	2000; -5000	2060; -5060	2550; -5550	-2000; 1985
b_6	2000; -5000	3050; -6050	-2000; 4450	-2000; 1985

Дії СВВ і атакуючого вибираються згідно рівноваги Неша. Одночасно з метою порівняння для кожного сценарію був підрахований відповідний обсяг ресурсів для традиційної реалізації СВВ.

Результати роботи системи моделювання

Стратегія атакуючого

b_1	b_2	b_3	b_4	b_5	b_6	очікувана корисність
0.6995816	0.02153313	0.0	0.0	0.2490239	0.02986128	0.0
0.6995493	0.0	0.0	0.0	0.2489484	0.05150225	0.0
0.6995799	0.02040104	0.0	0.0	0.2800189	0.0	0.0
0.6995493	0.0	0.0	0.0	0.3004506	0.0	0.0

Стратегія СВВ

a_1	a_2	a_3	a_4	очікувана корисність
0.5	0.0	0.0	0.5	1995.81672214
0.5	0.0	0.0	0.5	1995.49323986
0.5	0.0	0.0	0.5	1995.79971523
0.5	0.0	0.0	0.5	1995.49323986

Результати роботи системи моделювання

Кожен сценарій був реалізований 100 разів, а результати відображають середні значення ціни та відсотка виявлення. Дії СВВ та нападника обиралися відповідно до рівноваги Неша. У той же час для порівняння для кожного сценарію була розрахована відповідна кількість ресурсів для традиційного впровадження СВВ, що має 95-відсоткову точність.

Кількість ресурсів, необхідних для традиційної реалізації, перевищує запропонований підхід. З іншого боку, точність виявлення дещо поступається нормальній реалізації, але, тим не менше, дозволяє використовувати таку ігрово-теоретичну оптимізацію в системах з обмеженими ресурсами.

Перспективні напрямки розвитку продукту

Подальша оптимізація алгоритмів детектора СВВ.

Перехід з GNG та IGNG на Fast-GNG.

Навчання нейронного газу на даних, отриманих від взаємодії з honeypot.

Введення рефлексії для автоматичного вибору часу моніторингу системою різних нод мережі.