Генетический алгоритм подбора оптимальной конфигурации системы защиты информации

С.О. Иванов
Факультет информатики и вычислительной
Чувашский государственный университет имени И.Н. Ульянова
Чебоксары, Россия
e-mail: v101-11@mail.ru

Д.В. Ильин
Факультет информатики и вычислительной
Чувашский государственный университет имени И.Н. Ульянова
Чебоксары, Россия
e-mail: destr@mail.ru

Л.А. Ильина
Факультет информатики и вычислительной
Чувашский государственный университет имени И.Н. Ульянова
Чебоксары, Россия
e-mail: larisai2009@gmail.com

Аннотация

Статья посвящена применению генетического алгоритма для подбора и оптимизации конфигурации системы защиты информации. В качестве критериев используются суммарная величина риска и стоимость системы защиты информации. Описана модель системы защиты информации, которая применяется для расчета рисков. Приведены базовые функции генетического алгоритма. Работа алгоритма демонстрируется на примере расчета конфигурации защиты для типовой фирмы.

1. Введение

В настоящее время существует множество средств защиты информации, правил и рекомендаций по формированию систем защиты. Многие из них противоречат друг другу или не могут быть совместимы.

Одной из задач при построении системы защиты информации является составление набора защитных мер. При этом в необходимо учитывать не только требования безопасности [4], но и затраты на внедрение и поддержание системы защиты информации.

Обычно при составлении набора защитных мер используется метод, основанный на экспертном подходе [2], который сводится к предпочтениям экспертов, но так как количество возможных вариантов конфигурации системы защиты превышает мыслительные возможности, то в итоге к типовым схемам.

В данной работе предлагается применение генетического алгоритма для подбора и оптимизации конфигурации системы защиты информации. Генетические алгоритмы являются частным случаем эволюционных методов приближенного (эвристического) решения задач оптимизации и структурного синтеза [1]. Генетические алгоритмы (ГА) основаны на поиске лучших решений с помощью наследования и усиления полезных свойств множества объектов определённого приложения в процессе имитации их эволюции [9].

2.1. Модель системы защиты информации

Для решения вышеперечисленных проблем, была разработана модель распространения последствий [7, 8]. Элементы исследуемой среды непосредственно отображаются в модели с помощью субъектов и их свойств, а логика ее развития с помощью функций реакций на импульсы воздействия. Более подробно устройство модели рассмотрено в работах [7]. Проблема недостаточности информации решается с помощью симуляции, что позволяет получить нужные данные в любой момент (или интервал) времени.

Рассмотрим модель системы защиты информации, основанную на риск-ориентированном подходе, изучением которого занимается дисциплина риск-менеджмент [3]. Свяжем основные элементы модели взаимодействия с понятиями и терминами риск-менеджмента.

Структура модели и основные связи элементов представлены на рисунке 1.

Труды Шестой всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28-31 мая, Уфа-Ставрополь, Россия, 2018

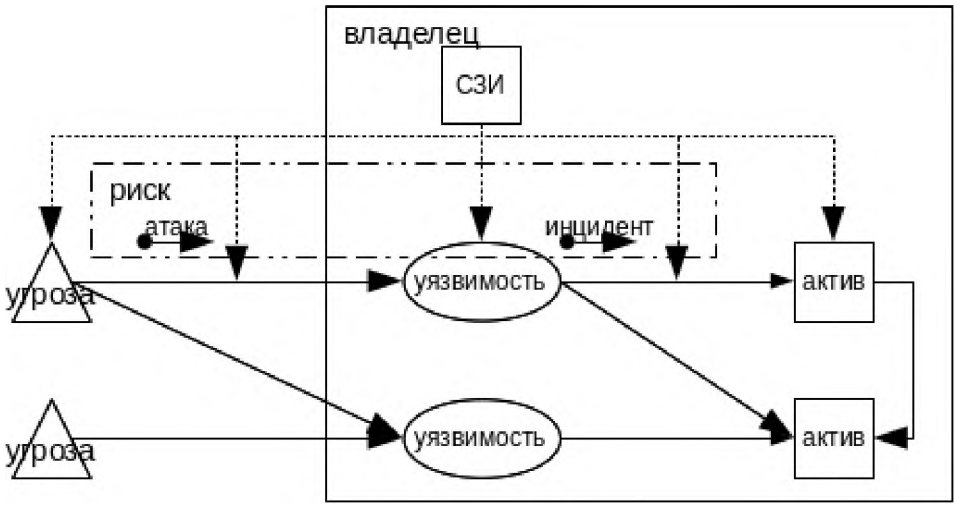

Рис. 1. Модель системы защиты информации

Эффекты от применения защитных мер, согласно модели, относятся к одной из пяти категорий:

- маскирующие – приводят к изменению вероятности атаки;
- избегающие – уменьшают актуальность угрозы;
- ограждающие – увеличивают защищенность от атак;
- укрепляющие – снижают ущерб от инцидента;
- компенсирующие – ликвидируют часть потерь;
- восстанавливающие – позволяют изменять потери.

Применение защитных мер могут иметь так же и отрицательные эффекты, из-за их влияния друг на друга и особенностей реализации.

Для упрощения компьютерного моделирования расчетов, представим все элементы в табличной форме: в строках располагаются угрозы, в столбцах – активы, а на пересечении - уязвимости (Табл.1).

Таблица 1. Табличное представление системы защиты информации

Угрозы/ Ценность:	**Ценность (A_1)**	…	**Ценность (A_M)**
Угроза (T_1)	уязвимость ($V_{1,1}$)/ ущерб ($D_{1,1}$)	…	уязвимость ($V_{1,M}$)/ ущерб ($D_{1,M}$)
…	…	…	…
Угроза (T_N)	уязвимость ($V_{N,1}$)/ ущерб ($D_{N,1}$)	…	уязвимость ($V_{N,M}$)/ ущерб ($D_{N,M}$)

К построенной таблице применяются защитные меры, которые изменяют параметры уязвимостей. Ущерб определяется по потерям от каждого инцидента с учетом длительности и стоимости восстановления, максимального времени простоя актива и его влияния на бизнес-функции.

Составление требуемых профилей атак и расчет величины риска осуществляет владелец риска – лицо или организация имеющая ответственность или полномочия по менеджменту риска. Цикл имитационного моделирования для оценки защищенности состоит из следующих шагов:

для каждой угрозы:

- определить вероятность возникновения атаки на текущем шаге (с учетом интенсивности угрозы),
- проверить актуальность атаки,
- если атака не актуальна, то перейти к следующей угрозе,
- для каждого актива текущей угрозы:
- определить вероятность возникновения инцидента (с учетом защищенности),
- при возникновении инцидента – добавить ущерб от него к потерям актива.

2.2. Оценка риска

В настоящее время риски используются для системного подхода к задачам анализа, разработки, оценки и управления системой безопасности организации.

Существует много подходов к определению риска:

- вероятностно-стоимостная оценка потерь,
- дисперсия случайной величины финансового баланса,
- сочетание вероятности и последствий наступления неблагоприятных событий,

- неопределённое событие или условие, которое в случае возникновения имеет позитивное или негативное воздействие на репутацию компании, приводит к приобретениям или потерям в денежном выражении,
- вероятностные последствия (отрицательные и положительные).

В разработанной модели системы защиты информации, для определения риска, используется тот факт, что величина потерь зависит от цепочки элементов: угроз, уязвимостей, средств защиты [8]. Таким образом, риск – путь от угрозы до актива (рис. 1, штрих-пунктир). В разработанной модели получается, что количество рисков – число комбинаций угроз и активов:

$$R = N \cdot M, \quad (1)$$

где N – количество угроз, M – количество активов.

В этой формуле не учитывается сочетание нескольких угроз и внутреннее влияние между защитными мерами. Более подходящим способом определения риска, является метод, основанный на профилях атак – последовательности атак, состоящих из сочетания различных угроз.

Количество рисков в данном случае равно числу сочетаний угроз в последовательности:

$$R=(2^N)^T, \quad (2)$$

где T – количество атак.

Таким образом, величиной риска для заданного профиля атак в разработанной модели является суммарный ущерб от успешных атак.

При отсутствии цепных и побочных эффектов, расчет суммарного риска можно свести к математическому ожиданию ущерба для каждого актива от заданных атак

$$r = \sum V_{i,j} \cdot D_{i,j}, i = \overline{1,N}, j = \overline{1,M}, \quad (3)$$

где $V_{i,j}$ – вероятность возникновения инцидента от угрозы i с активом j, $D_{i,j}$ – величина ущерба от инцидента.

2.3. Описание генетического алгоритма выбора защитных мер

Построенная имитационная модель может использоваться для решения многих задач. Рассмотрим возможность ее использования для подбора оптимальной конфигурации (набора) защитных мер для системы защиты информации.

Формализуем задачу в терминах генетического алгоритма.

Хромосомой будет набор защитных мер, закодированный в виде двоичного числа [1]. Каждый двоичный разряд равный 1 означает включение в конфигурацию защитной меры с соответствующим номером. Диапазон изменения кода:

$$G = (d_0 d_1 \ldots d_L)_2 = (0 \ldots 2L)_{10}, \quad (4)$$

L – количество имеющихся защитных мер,

d_i – включение или не включение защитной меры в набор.

Популяция (P) состоит из экземпляров с различными хромосомами. Ее размер ограничен K – максимальным количеством экземпляров в популяции. Каждый экземпляр определяется совокупностью:

$$X = \{G, C, R\}, \quad (5)$$

где G – генетический код, C – стоимость защитных мер, R – суммарный риск с учетом принятых защитных мер.

Для расчета риска используется разработанная модель системы защиты. Для каждого носителя по его генетическому коду подбирается набор защитных мер, с использованием которого изменяются параметры модели, и вычисляется суммарный риск.

Функция полезности (fU) – оценка эффективности системы защиты для заданного профиля атаки, которую необходимо максимизировать.

Эффективность может оцениваться различными способами, представленными в [9, 10]. Наиболее простой – снижение риска.

$$fU(X) = R_0 - X.R, \quad (6)$$

где X – экземпляр набор защитных мер

R_0 – величина риска без использования защитных мер,

$X.R$ – величина риска с учетом набора защитных мер $X.G$.

С другой стороны достигнутый результат требует дополнительных расходов на защитные меры. Для учета влияния затрат на защиту используется следующее отношение:

$$fU(X) = (R_0 - X.R)/X.C, \quad (7)$$

где X – экземпляр набор защитных мер, R0 – величина риска без использования защитных мер,

$X.R$ – величина риска с учетом набора защитных мер $X.G$, $X.C$ – стоимость защитных мер.

Формула (7) показывает, как снижается (или увеличивается) риск на каждую вложенную единицу стоимости.

Функция кроссовера – порождение новых носителей. Для каждой пары создадим новый экземпляр, наследующий черты родителей:

func K(P):=

foreach X from P and foreach Y from P where X != Y do

R.add ({G: xor(Xi.G,Xj.G), C:, R:})

return R.add(P)

Функция мутации – варьирование генетического кода. Случайным образом инвертируем два двоичных разряда в хромосоме:

func M(P):=

foreach X from P do X.G = xor (X.G, 1 << rand(0,L))

где rand() – функция возвращающая случайное число.

Функция селекции – отбор наилучших носителей. Для уменьшения популяции оставим только K носителей дающие наибольший результат fU:

func S(P):=

return P.sort().slice(1, K),

Перед применением селекции необходимо вычислить X.C и X.R для популяции. Начальная популяции должна состоять минимум из двух экземпляров. Каждая эпоха состоит из последовательного применения основных функций:

func E():=

P = K(P),

M(P),

P = S(P)

2.4. Демонстрация применения алгоритма

Рассмотрим типичную фирму (Рис 2.)

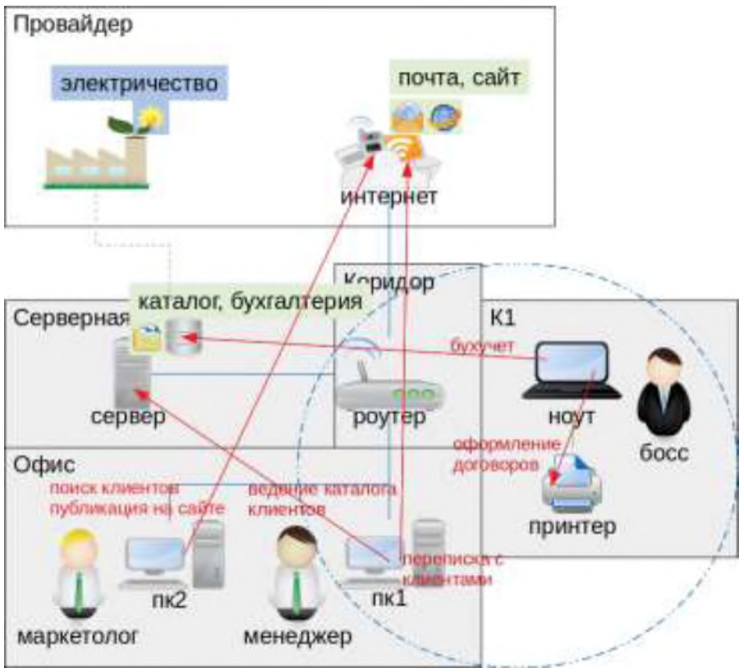

Рис. 2. Структура фирмы

На основании бизнес-функций (красные линии) составляется список ценностей. Используя стандарт [5, 6], идентифицируются типичные угрозы (Табл.2).

Список защитных мер приведен в таблице 3. В столбце с эффектами описаны действия над уязвимостью(V) и ущербом (D) в соответствующей ячейке таблицы с инцидентами (Таблица 2).

Используя [2] сформируются профили атак (таблица 4) .

Таблица 2. Инциденты

	Сервер	**Каталог**	**Бухгалтерия**	**ПК1**	**ПК2**	**Интернет**	**Почта**	**Электро-энергия**	**Договора**
Пожар в серверной	0,3/ 26000	0,1/ 99000	0,1/ 165000						
Воровство ПК				0,5/ 20000	0,5/ 15000				
Блекаут								1/ 500	
Разрыв						1/ 5000			

	Сервер	Каталог	Бухгалтерия	ПК1	ПК2	Интернет	Почта	Электро-энергия	Договора
Утечка содержимого договоров									0,01/ 500
Удаленный шпионаж			0,01/ 5000						
Взлом почты							0,1/ 1500		
Взлом роутера						0,1/ 5000			
Отказ сервера	0,2/ 5200								
Порча бухгалтерии			0,9/ 165000						

Таблица 3. Список защитных мер

№	Меры	Основная категория	эффект	Стоимость
1	Антивирус	ограждающие	V[5,9]*=0.1, V[7,7]-=0.05, V[10,3]*=0.5	1500
2	Резервная копия	восстанавливающие	D[1,3]*=0.1, D[7,7]*=0.1, D[10,3]*=0.1	4000
3	Резервный канал	укрепляющие	D[4,6]*=0.1	300*12
4	Облачное хранилище	избегающие	D[4,6]+=6000+7500, V[10,3]+=-0.8	700*12
5	Противопожарные меры	ограждающие	V[1,1]*=0.05	60000
6	Сетевой экран	ограждающие	V[5,9]-=0.05, V[7,7]-=0.05, V[8,6]*=0.6	7000
7	Правила ИБ	избегающие	V[5,9]*=0.3, V[6,3]*=0.3, V[10,3]*=0.3	7000

Таблица 4. Профили атак

№	Атаки	№	Атаки
1	0	5	4,5
2	1	6	6, 7
3	2	7	8
4	3	8	9

Величина риска в исходной конфигурации: $R_0 = 54425$.

Проведем вычислительный эксперимент для определения лучших конфигураций по снижению риска (Табл. 5) и по эффективности затрат (Табл. 6).

Для проверки проведем вычисления для всех вариантов генетического кода (Рис. 3, 4)

Таблица 5. Лучшие экземпляры по снижению потерь после 7 эпох

G	R	C	fU
119 (1110111)	32681.9	85100	21743.1
87 (1010111)	32896.9	78100	21528.1
118 (1110110)	32906	83600	21519
86 (1010110)	33121	76600	21304
55	33219.5	76100	21205.5
103	33422.9	25100	21002.1
23	33452	69100	20973
71	33637.9	18100	20787.1
102	33647	23600	20778
70	33862	16600	20563

G	R	C	fU
39	33960.5	16100	20464.5
54	33966.5	74600	20458.5
7	34193	9100	20232
22	34199	67600	20226
38	34707.5	14600	19717.5
6	34940	7600	19485
117	36239.15	81100	18185.85

Таблица 6. Лучшие экземпляры по эффективности затрат на защитные меры после 5 эпох

G	R	C	fU
1	46995.5	1500	4.953
…			
1	46995.5	1500	4.953

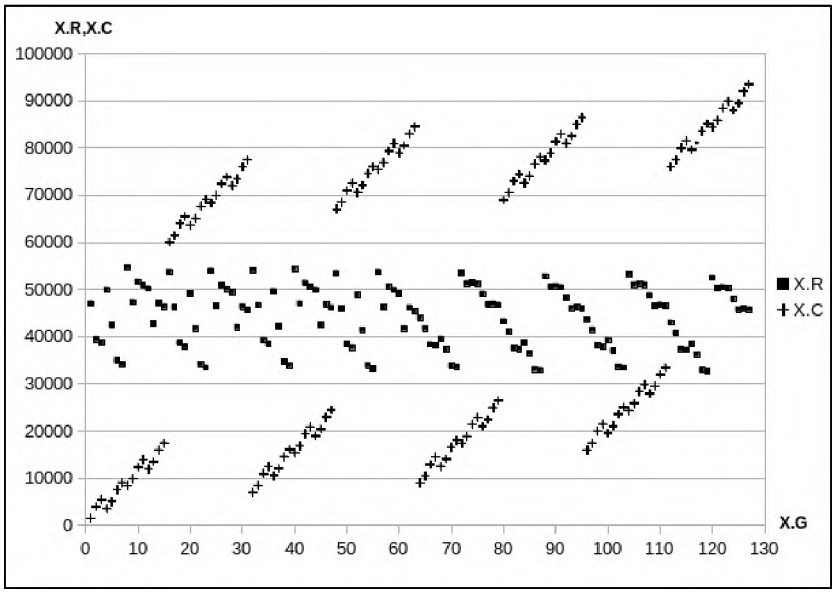

Рис 3. Величины риска и стоимости мер для каждого генетического кода

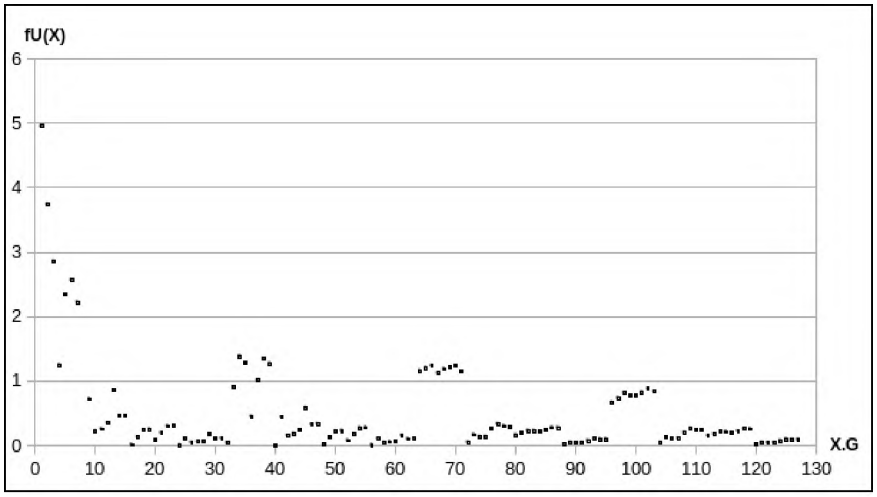

Рис 4. Эффективность защитных мер по затратам

Из полученных результатов можно сделать выводы:

- наиболее эффективная с точки зрения затрат система защиты состоит из противопожарных мер.
- для снижения рисков система защиты должна включать: Антивирус, Резервный канал, Противопожарные меры, Сетевой экран, а так же желательно Резервную копию и Правила ИБ.

Так же можно отметить, что в данном алгоритме осуществляется больший объем вычислений риска

$(2^L)^2$ по сравнению с перебором всех вариантов 2^L. Для 7 эпох было осуществлено 25030 вычислений величины риска, для 5 эпох – 8518, а при 128 для перебора всех вариантов. Данный недостаток можно сократить с помощью кеширования уже вычисленных значений и селекции только уникальных экземпляров.

4. Заключение

В настоящее время существует множество средств защиты информации и правил обеспечения ИБ. Количество вариантов построения системы защиты возрастает экспоненциально (2). Применение ГА для поиска оптимальной конфигурации по заданным характеристикам обеспечивает приемлемый по скорости способ их нахождения. К недостаткам описанного алгоритма можно отнести – необходимость предварительного анализа и построения модели системы защиты информации, а так же сведение многокритериальной задачи к оптимизации одной характеристики.

Список используемых источников

1. Агибалов О.И., Золотарев А.А. Десятичный подход к кодированию значений в генетическом алгоритме // сборник трудов конференции «Современные информационные технологии: тенденции и перспективы развития». – 2014. – С. 31-33.
2. ГОСТ Р ИСО/МЭК ТО 13335-4-2007 Информационная технология. Методы и средства обеспечения безопасности. Часть 4. Выбор защитных мер
3. ГОСТ Р 51897-2011. Менеджмент риска. Термины и определения. - М.:Стандартинформ. - 2012.
4. ГОСТ Р ИСО/МЭК 27001-2006 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования
5. ГОСТ Р ИСО/МЭК 27002-2012 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Свод норм и правил менеджмента информационной безопасности
6. ГОСТ Р ИСО/МЭК 27005-2010 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности
7. Иванов С.О. Модель процесса взаимодействия // Вестник Российского университета кооперации. 2014. № 1(15). С. 132–137.
8. Иванов С.О., Ильин Д.В., Ильина Л.А. Методика анализа риска с использованием модели последствий // Вестник Чувашского университета. – 2015. – С.149-153
9. Тененев В.А. Решение задачи многокритериальной оптимизации генетическими алгоритмами // Интеллектуальные системы в производстве. – Ижевск: Ижевский государственный технический университет им. М.Т. Калашникова, 2006. – С.103-109
10. Граничин О.Н., Кияев В.А. Соотношение эффективности и рентабельности систем информационной безопасности // Безопасность информационных систем. – URL: https://www.intuit.ru/studies/courses/13845/1242/lecture/27501?page=4