А.А. Мурзакова, Е.А. Мурзакова, Т.И. Паюсова

Тюменский государственный университет, г. Тюмень

УДК 004.056

ПРОЕКТИРОВАНИЕ ОПТИМАЛЬНОЙ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Аннотация. В статье представлено решение многокритериальной оптимизационной задачи построения наиболее эффективной системы защиты информации с помощью генетического алгоритма в рамках программного комплекса, выполненного на языке C#.

Ключевые слова: система защиты информации, многокритериальная оптимизация, оптимальная система защиты, математическое программирование, генетический алгоритм.

Защита информации является одной из первостепенных задач в любой предметной области. Однако такая задача выступает достаточно сложной. Путем проб и ошибок компании приходят к тем или иным решениям в обеспечении информационной безопасности (ИБ), при этом теряя в процессе пути время и денежные средства. Актуальной является возможность применения для своей организации грамотно построенной системы защиты, которая была бы спроектирована и разработана согласно построенной математической модели, учитывающей все критерии и нормы систем ИБ и подтвержденной математическими вычислениями.

Эффективность системы защиты информации (СЗИ) определяется соотношением вложенных сил в её реализацию, например, с точки зрения финансовых показателей, временных и трудозатрат, и качеством полученного результата, который, например, можно определить по скорости реагирования на инциденты ИБ, по уровню снижения рисков ИБ, по умению адаптироваться к условиям изменяющейся внешней среды. Таким образом,

построение эффективной системы защиты является многокритериальной задачей, от успешного решения которой зависит эффективность работы всей информационной системы в целом.

Целью работы выступило построение математической модели оптимальной системы защиты информации и разработка программного приложения, решающего эту задачу.

Чтобы обеспечить требуемый уровень информационной защиты, каждая комплексная система защиты информации должна содержать следующие элементы: правовые, организационные, инженерно-технические, программно-математические.

Существуют различные программные и технические средства защиты информации (СрЗИ), отличающиеся такими параметрами, как стоимость, быстродействие, надежность, требуемые ресурсы, универсальность, сложность, степень защиты, время преодоления защиты, стойкость шифра, возможность сетевого администрирования, требования к квалификации пользователей и др. Поэтому, чтобы получить оптимальный комплекс СрЗИ, обеспечивающий перекрытие всех выявленных каналов несанкционированного доступа (НСД) и удовлетворяющий заданным нормам эффективности, требованиям той или иной организации, необходимо из множества имеющихся СрЗИ выбрать те, что помогут получить рациональную структуру системы защиты (осуществить сравнение различных вариантов по критериям и выбрать наилучший на основе множества показателей).

У построенной СЗИ необходимо оценить комплексную эффективность и при необходимости внести изменения [1].

Математическая постановка задачи

Рассматриваемая задача представляет собой задачу многокритериальной оптимизации, так как система защиты характеризуется целым рядом параметров, которые необходимо оптимизировать. Происходит

одновременная оптимизация двух или более конфликтующих целевых функций в заданной области определения.

Разберем математическую постановку задачи построения оптимальной системы защиты информации.

Требуется максимизировать уровень защищенности (эффективность применяемых средств защиты, предотвращенный ущерб от воздействия угроз), минимизировать стоимость системы защиты (затраты на защиту), минимизировать влияние системы защиты на производительность системы (минимизировать ресурсозатратность или максимизировать производительность системы). Выполнение этих требований позволит подобрать оптимальный вариант реализации СЗИ, рационально спроектировать комплексную систему защиты.

Разберем последовательно. Введем понятие коэффициента защищенности системы R. Коэффициент защищенности может задаваться следующей формулой:

$$R = 1 - \frac{\sum_1^n L_i * I_i * (1 - p_i)}{\sum_1^n L_i * I_i},$$

где L_i – затраты (потери) от взлома угрозой -го вида (стоимость защищаемой информации, i-го актива); I_i – интенсивность потока взломов i-го вида угроз; p_i – вероятность отражения угроз i-го вида системой защиты.

Если же изначально заданы вместо интенсивности потоков угроз вероятности их появления, то формула примет следующий вид:

$$R = 1 - \frac{\sum_1^n L_i * C_i * (1 - p_i)}{\sum_1^n L_i * C_i},$$

где C_i – вероятность появления угроз -го вида в общем потоке попыток НСД к информации.

Зависимость интенсивности потока взломов от вероятности появления угроз выражается как $C_i = \frac{I_i}{I_{all}}$, где I_{all} – общая интенсивность потока НСД к информации.

Теперь рассмотрим затраты на построение системы защиты S (стоимость приобретения, внедрения, поддержки). Их можно рассчитать по следующей формуле:

$$S = \sum_{i \in F} \sum_{j \in A} S_{ij} * x_{ij} + \sum_{i \in F} S_i * Y_i,$$

$$\sum_{i \in F} x_{ij} = 1,\ \forall_j \in A\ , x_{ij} \in \{0; 1\}, y_i \in \{0; 1\},$$

где S_{ij} – затраты на защиту -го информационного актива i-м средством; S_i – затраты, общие для всех информационных активов, на защиту i-м средством; F – множество СрЗИ на предприятии; A – множество защищаемых информационных активов; x_{ij} – булева переменная: «1», если i-е СрЗИ используется для защиты j-го информационного актива, «0» - иначе. Допускается, что -е средство используется для защиты от i-ой угрозы; y_i – булева переменная: «1» - i-е СрЗИ используется в системе защиты, «0» - иначе. Причем -ое средство защиты в системе может быть использовано только один раз.

Производительность системы (количество прикладных задач, решаемых в единицу времени) можно рассчитать с применением моделей и методов теории массового обслуживания и теории расписаний. Лучше вместо «производительности» использовать понятие «снижение производительности» системы от установки системы защиты, так как часто изначальным условием задается граница снижения производительности.

Пусть E- производительность системы. Итоговая производительность системы может быть найдена из следующей формулы:

$$E = \frac{N_З * V_Д}{3600} * K_{ТОТ} * (1 - K_{ИЗБ}),$$

где $K_{ИЗБ}$ – коэффициент программной избыточности (часть вычислительной мощности системы, расходуемой на проверку входных данных для защиты от непредумышленных, случайных искажений

вычислительного процесса); $N_З$ – число запросов в час; $V_Д$ – объем запрашиваемых данных, бит; $K_{ТОТ}$ – коэффициент готовности системы.

Ограничение на ресурсозатратность зададим следующим образом. Пусть имеется -точек (сервера, рабочие станции и др.), на которых размещаются СрЗИ, причем каждая точка принадлежит одному из Z-классов; n-средств информационной защиты, причем каждое средство принадлежит какому-либо классу $ClassISF_i, i = \overline{1,k}$ и имеет стоимость w_j и ресурсозатратность q_j, а в отдельном классе находятся эквивалентные средства защиты, и одно из них должно присутствовать в системе. Ресурсозатратность Q комплекса средств защиты определяется по формуле:

$$Q = \sum_{i=1}^{k} \sum_{j \in ClassISF_i} \sum_{g \in ClassISF_i} q_{ijg} * x_{ijg},$$

где x_{ijg} – булева переменная («1» - используется средство защиты, «0» - не используется); q_{ijg} – ресурсозатратность отдельного СрЗИ.

Таким образом, целевую функцию и ограничения задачи проектирования системы защиты информации можно представить следующим образом:

$$R \to max$$
$$\begin{cases} S \le S_{доп} \\ E \ge E_{доп}, \\ Q \le Q_{доп} \end{cases}$$

где $S_{доп}$ определяется финансовой состоятельностью предприятия и рисками (ущербом) от реализации атак на инфраструктуру ЗИ (заданное ограничение на затраты по защите); $E_{доп}$ – заданное ограничение на снижение производительности; $Q_{доп}$ – заданное значение ресурсозатратности предприятия [2-5].

Методы решения поставленной задачи

Раз задача относится к многокритериальной, то ее можно решать методами многокритериальной оптимизации. Методами решения выступают интерактивные методы (по информации систем поддержки принятия решений эксперты делают свой выбор и оценку), эволюционные методы (применение генетических алгоритмов) и метод исследования пространства параметров (построение допустимого и Парето-оптимального множеств решений) [6].

При решении задачи построения оптимальной СЗИ необходимо учитывать большое число показателей СрЗИ при оценке и выборе их рационального варианта; многие показатели имеют преимущественно качественный характер, что вызывает сложность перевода их в количественные значения при решении; некоторые требования имеют противоречивый характер; имеется трудность получения исходных данных на ранних этапах проектирования, да и сама постановка задачи может быть нечеткой. В связи с этим практически невозможно применение традиционных математических методов для решения задачи [8], например, таких как методы теории принятия решений [5], метод последовательных уступок [3, 4], оптимизационно-имитационный подход [7].

Во многих работах указано [3, 5, 7, 8], что вероятность проявления угрозы, возможность предотвращения угрозы с помощью определенного средства защиты, величина ущерба, важность выполнения некоторого требования для устранения угрозы и другие данные, необходимые при оценке и выборе, задаются на основе статистики, математического моделирования, экспериментальных исследований и данных от экспертов. Необходимы подходы, которые могут работать с экспертной информацией и неполной информацией. Лингвистический подход на базе теории нечетких множеств и генетические алгоритмы как раз работают с таким видом информации.

Поэтому для решения поставленной задачи было решено использовать генетический алгоритм, который помогает найти приемлемое значение (приблизительное оптимальное решение) за относительно короткое время по сравнению с другими методами.

Генетические алгоритмы

Метод эволюционных вычислений опирается на модель естественного отбора. Все понятия заимствуются из биологии, как и суть самих эволюционных вычислений, которая состоит в том, что особи, которые более приспособлены к внешним условиям, имеют преимущество в выживании и размножении перед остальными и дают большее число потомства, которое так же явится более сильным и приспособленным, что приведет к увеличению процента приспособленных особей в виде и по истечении времени станет выше общая приспособленность вида. Таким образом, в методе эволюционных вычислений производят выборку параметров системы, приносящих наибольший вклад в решение выдвинутой задачи.

Эволюционные вычисления, применительно к обеспечению ИБ, используют в качестве оптимизирующих компонентов, предназначенных для того, чтобы ускорить процесс обнаружения угроз и повысить его эффективность. Опираясь на принятые меры защиты, такая система путем проб и ошибок осуществляет поиск сбалансированного комплекса защитных средств и мер, обладающего свойством приспосабливаемости к изменению ситуации, то есть к возникновению новых угроз, атак и других непредвиденных обстоятельств.

Основными этапами генетического алгоритма являются: задание исходной популяции хромосом; оценка приспособленности хромосом в популяции; проверка условия остановки алгоритма; выбор хромосом-родителей для создания потомков; применение генетических операторов – мутации и скрещивания; формирование новой популяции; выбор «наилучшей» хромосомы.

Теперь проведем аналогию элементов математической модели задачи оптимизации с терминами, используемыми в генетических алгоритмах.

Наименьшей неделимой единицей биологического вида является особь a_k^t, где k – номер особи, t – некоторый момент времени эволюционного процесса. Тогда в качестве особи можно взять $\bar{x} \in D$ (произвольное допустимое решение, где $\bar{x} = (x_1, x_2, \dots, x_n)$).

Пусть $S(\overline{x}) = (s_1, s_2, \dots, s_n)$ – хромосома из -членов. При заполнении хромосомы значениями появляется понятие «генотип» $S(\overline{x}) \in S$. Генофонд – конечное множество всех возможных генотипов, S (пространство поиска).

Качественные признаки, характеристики $(g_1, g_2, \dots, g_m)$, в модели оптимизационной задачи можно заменить понятием фенотипа $g(a_k^t)$.

Критерий оптимальности Q можно рассмотреть как оценку фенотипа, приспособленность $m(a_k^t)$ особи a_k^t. Чем выше приспособленность, тем лучше особь адаптирована к внешней среде. Ареал – это область поиска D. Поиск осуществляется среди популяции $P^t = (a_1^t, a_2^t, \dots, a_n^t)$, где n – численность популяции [9-11].

Выбор инструментов разработки и функционала приложения

Планируется решение многокритериальной оптимизационной задачи построения эффективной СЗИ в рамках программного комплекса, выполненного в среде разработки Visual Studio 2015 на языке C# с применением системы управления базами данных MS SQL Server 2012.

Предполагаемый функционал приложения: задание параметров, целевой функции, ограничений, направления оптимизации (минимизация/максимизация); ввод параметров генетического алгоритма (численность начальной популяции, число «брачных» пар, число «мутантов», вероятность кроссовера, вероятность мутации, процент мутирующих битов внутри особи, число поколений, в течение которых осуществляется генетический поиск, максимальный возраст особи); решение

оптимизационной задачи с помощью генетического алгоритма; вывод полученного результата.

Приложение выступит оценочным и оптимизационным механизмом, обеспечивающим построение обоснованной по экономическим соображениям СЗИ.

Выводы

Таким образом, представлена математическая модель решения задачи проектирования оптимальной системы защиты информации, рассмотрены методы для решения задач многокритериальной оптимизации, в том числе применение генетических алгоритмов. Намечены этапы дальнейшей работы (реализация программного решения задачи).

Система защиты информации, спроектированная с применением методов оптимизации и генетического алгоритма, позволит упростить и улучшить деятельность организации в области принятия решений, позволит сэкономить время, снизить лишние расходы и разумно распределить финансы, обеспечивая оптимальный уровень защиты информации в компании.

СПИСОК ЛИТЕРАТУРЫ

1. Клюев С.Г. Математическая постановка задачи построения оптимальной системы защиты информации // Журнал научных публикаций аспирантов и докторантов: сетевой журн. 2008. URL: http://jurnal.org/articles/2008/inf4.html (дата обращения: 04.04.2018);

2. Расчет производительности информационной системы [Электронный ресурс]// Студопедия: [сайт]. [2014]. URL: https://studopedia.su/14_109076_raschet-proizvoditelnosti-informatsionnoy-sistemi.html (дата обращения: 07.04.2018);

3. Гатчин Ю.А., Жаринов И.О., Коробейников А.Г. Математические модели оценки инфраструктуры системы защиты информации на предприятии // Методы и системы защиты информации. 2012. №2. С.92-95.

4. Колесников К.В., Шадхин В.Е. Системный анализ критериев и параметров проектирования системы защиты // Радиоэлектронные и компьютерные системы. 2006. №6. С.87-90.

5. Кацупеев А.А., Щербакова Е.А., Воробьёв С.П. Постановка и формализация задачи формирования информационной защиты распределенных систем // Инженерный вестник Дона. 2015. №1. С.1-17.

6. Лотов А.В., Поспелова И.И. Конспект лекций по теории и методам многокритериальной оптимизации: учебное пособие. Москва, 2014. 127 с.

7. Овчинников А.И., Журавлев А.М., Медведев Н.В., Быков А.Ю. Математическая модель оптимального выбора средств защиты от угроз безопасности вычислительной сети предприятия // Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2007. №3. С.115-121.

8. Кудин Д., Корольков В. Количественные оценки качества функционирования системы защиты информации // Обеспечение компьютерной безопасности в государственных, банковских и других информационных системах. 2003. №6. С.25-28.

9. Батищев Д.И., Неймарк Е.А., Старостин Н.В. Применение генетических алгоритмов к решению задач дискретной оптимизации: учебно-методическое пособие. Нижний Новгород: ННГУ, 2007. 85 с.

10. Еремеев А.В. Генетические алгоритмы и оптимизация: учебное пособие. Омск: Изд-во Ом. гос. ун-та, 2008. 48 с.

11. Семенкин Е.С., Жукова М.Н., Жуков В.Г., Панфилов И.А., Тынченко В.В. Эволюционные методы моделирования и оптимизации сложных систем: учеб. пособие. Красноярск: Сибирский федеральный университет, 2007. 310 с.