Агрегированные оценки вероятности успеха социоинженерной атаки: устойчивость структуры политик доступа

А. А. Азаров[1], А. В. Суворова[2]
Санкт-Петербургский институт информатики и автоматизации РАН
[1]artur-azarov@yandex.ru, [2]suvalv@gmail.com

***Аннотация*. Доклад посвящен разработке метода анализа устойчивости структуры политик доступа пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника; под анализом устойчивости в данном случае понимается расчет агрегированных оценок вероятности успеха социоинженерного атакующего воздействия злоумышленника, другими словами – вероятности получения злоумышленником доступа к тем или иным конфиденциальным данным, хранящимся в информационной системе. Подход основывается на данных о степени выраженности уязвимостей пользователей информационных систем, расчете вероятностных оценок, а также формировании более устойчивой к социоинженерным атакующим воздействиям структуры политик доступа пользователей, и строится на основании генетических алгоритмов. Такой подход позволяет определить наиболее устойчивую к социоинженерным атакующим воздействиям злоумышленника конфигурацию пользователей информационной системы. В докладе описаны общие принципы предлагаемого подхода и приведены результаты вычислительных экспериментов.**

Ключевые слова: социоинженерная атака; вероятностная модель; генетический алгоритм

I. Введение

Современные темпы развития цифровых коммуникаций, связывающие в единое целое все большее количество различных частей жизнедеятельности человека, вызывают серьезные опасения у специалистов, ответственных за обеспечение информационной безопасности [6].

Внедрение цифровых коммуникаций происходит не только на уровне повседневной жизни человека, но и в корпоративной среде. Корпоративная информация приобретает все большую ценность, соответствующим образом растут и бюджеты на защиту такой информации. Для успешного решения задач защиты информации разрабатываются все более совершенные методы защиты [5–7, 10, 11]. Вместе с тем, гораздо в меньшей степени изучаются вопросы, связанные с различными воздействиями на пользователей информационных систем, нацеленными на получение конфиденциальных корпоративных данных [9]. Подобного рода воздействия могут быть объединены термином социоинженерная атака злоумышленника, а отдельно взятое воздействие – социоинженерное атакующее воздействие злоумышленника.

В докладе рассматривается метод анализа устойчивости структуры политик доступа пользователей корпоративной информационной системы к социоинженерным атакующим воздействиям злоумышленника. Под анализом устойчивости в данном случае понимается расчет агрегированных оценок вероятности успеха социоинженерного атакующего воздействия злоумышленника, другими словами – вероятности получения злоумышленником доступа к тем или иным конфиденциальным данным, хранящимся в информационной системе. Подход основывается на данных о степени выраженности уязвимостей пользователей информационных систем, расчете вероятностных оценок, а также формировании более устойчивой к социоинженерным атакующим воздействиям структуры политик доступа пользователей, и строится на основании генетических алгоритмов. Такой подход позволяет определить наиболее устойчивую к социоинженерным атакующим воздействиям злоумышленника конфигурацию пользователей информационной системы.

II. Модели комплекса «Критичные документы – информационная система – персонал – злоумышленник»

В докладе рассматриваются модели, ранее предложенные в работах [1–4]. Общий комплекс состоит из моделей различных элементов информационной системы в целом, включающих, в том числе, модель пользователя, модель документов, модель системы, модель злоумышленника. Более подробно остановимся на модели пользователя, параметрами которой служат доступ пользователя к тем или иным критичным документам, уровень доступа к этим документам, а также связи между пользователями. На основании данных о связях между пользователями может быть построен граф социальных связей пользователей с нагруженными двунаправленными дугами и нагруженными вершинами. Весами дуг являются вероятности перехода от пользователя к пользователю,

Работа выполнена при финансовой поддержке РФФИ, проект №18-37-00340, и частичной поддержке по проекту по государственному заданию СПИИРАН № 0073-2018-0001

сформированные на основании типа связей пользователей, а весом каждой вершины – полная вероятность успеха социоинженерного атакующего воздействия на пользователя информационной системы, построенная на основании степени выраженности уязвимостей пользователя [10]. На основании полной вероятности успеха социоинженерного атакующего воздействия злоумышленника предлагается построить вероятность доступа злоумышленника к критичным документам через пользователя, на которого совершено социоинженерное атакующее воздействие. С учетом того, что согласно политикам доступа, у пользователей может быть различный уровень доступа к файлам (чтение – создание/модификация – удаление – полный доступ), представляется целесообразным для каждого типа уровня доступа построить такую вероятность. Соответствующие вероятности могут быть обозначены как p_r, p_c, p_d, p_f. Для каждого отдельного элемента критичной информации (документа) может быть найдена полная вероятность доступа к нему хотя бы через одного пользователя информационной системы.

В качестве примера рассмотрим вероятность доступа к критичной информации с типом доступа «чтение». Вероятность может быть выражена как $p^r_{\text{doc-acc}} = 1-(1-p^r_{\text{doc1}})...(1-p^r_{\text{docN}})$ для каждого типа доступа, где $p^r_{\text{doc1}} ... p^r_{\text{docN}}$ – вероятности полного доступа к критичному документам $1..N$ у пользователей с типом доступа «чтение». Таким образом могут быть получены оценки вероятности защищенности критичной информации, хранимой в информационной системе, как для каждого документа отдельно, так и всей информации в совокупности, если будет рассмотрена общая вероятность доступа ко всем документам. Для повышения уровня защищенности могут быть применены меры по изменению политик доступа пользователей к тем или иным файлам. В данном случае информация о вероятности успеха социоинженерного атакующего воздействия на пользователя является маркером, по которому может быть произведена коррекция политик доступа пользователей к критичной информации. Для реализации алгоритмов коррекции предлагается использовать генетические алгоритмы.

III. Оценка качества работы генетического алгоритма

Для оценки качества работы генетического алгоритма предлагается использовать критерий оптимизации по параметру защищенности критичной информации, хранимой в информационной системе, по типу доступа к такой информации, путем изменения политик доступа к ним. Очевидным образом должен быть сформирован набор документов, необходимый пользователям для постоянной работы, в противном случае могут быть критическим образом нарушены бизнес-процессы организации. Поэтому может быть сформировано множество документов $D = \left\{ \left\{ Doc_i \right\}_{i=1}^{n}, \left\{ Us_i, \left\{ UsR_j \right\}_{j=1}^{4} \right\}_{i=1}^{m} \right\}$, где $\left\{ Doc_i \right\}_{i=1}^{n}$ – это набор из n критичных документов, к которым должен быть предоставлен постоянный доступ, Us_i – это набор из m пользователей, которые должны обладать доступом UsR_j к документы. При этом пользователи могут обладать сразу несколькими типами доступа. Следует отметить, что оптимизация доступа сразу до этого уровня вызовет существенное увеличение времени протекания бизнес-процессов в организации, т.к. пользователи будут вынуждены запрашивать дополнительный доступ каждый раз, когда им понадобится документ, к которым у них отсутствует заранее предопределенный мандатный доступ. Формирование вероятности защищенности критичной информации, хранимой в информационной системе, на основании множества D позволит сформировать оценку нижней границы допустимой оптимизации. Данная граница, которая может быть обозначена как $Q_{\min}$, является множеством границ, рассчитанным для каждого типа доступа $Q_{\min} = \left\{ Q^r_{\min}, Q^c_{\min}, Q^d_{\min}, Q^f_{\min} \right\}$. Вместе с тем, собственники информационной системы могут назначить допустимый уровень риска, то есть минимально допустимый уровень защищенности информационной системы, к достижению которого должен стремиться процесс оптимизации. Обозначим это значение как $Q_{\max}$, для этой границы также верно утверждение, что это множество границ, т.е. $Q_{\max} = \left\{ Q^r_{\max}, Q^c_{\max}, Q^d_{\max}, Q^f_{\max} \right\}$.

Далее планируется применить традиционный генетический алгоритм, начальной популяцией которого является начальная конфигурация информационной системы, с имеющимся распределением ролей пользователей и их доступом к документам. Для каждого состояния системы формируется граничная оценка $Q = \left\{ Q^r, Q^c, Q^d, Q^f \right\}$, также происходит сравнение на выполнение ограничений множества D. Затем производится проверка выполнения критерия $Q_{\min} < Q < Q_{\max}$.

IV. Вычислительные эксперименты

Разберем пример оптимизации прав доступа пользователей на основе информационной системы, к которой имеют доступ 8 пользователей и в которой содержатся 12 документов. Для упрощения расчетов будем рассматривать только один тип доступа – «чтение». Ниже представлена матрица доступа пользователей информационной системы к документам. Строки матрицы – пользователи информационной системы (u1–u8), а столбцы – документы (табл. I).

Предположим, что для каждого пользователя уже рассчитана полная вероятность успеха социоинженерного атакующего воздействия злоумышленника (с учетом социальных связей среди пользователей информационной системы): $p(u1) = 0,05$, $p(u2) = 0,16$, $p(u3) = 0,12$,

$p(u4) = 0{,}14$, $p(u5) = 0{,}17$, $p(u6) = 0{,}13$, $p(u7) = 0{,}11$, $p(u8) = 0{,}01$.

ТАБЛИЦА I ИСХОДНОЕ РАСПРЕДЕЛЕНИЕ ПРАВ

Пользователи	Документы											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
u1	+	+	+	+	+		+	+	+	+	+	
u2	+	+	+			+	+	+	+	+	+	+
u3			+	+	+	+	+	+	+	+	+	+
u4	+	+	+	+	+	+					+	+
u5	+	+				+	+	+	+	+	+	+
u6	+			+	+	+	+	+	+	+	+	+
u7	+	+	+	+	+	+	+	+	+			+
u8	+	+	+	+	+		+		+	+	+	+

Соответствующие вероятности будут сопоставлены документам как p_r в соответствии с тем, к каким документам есть доступ у каждого пользователя, то есть итоговая матрица является результатом покомпонентного произведения вектора вероятностей успеха атакующего воздействия и матрицы доступа.

Для каждого документа может быть рассчитана полная вероятность доступа к такому файлу (в предположении, что доступ каждого пользователя независим), затем может быть рассчитана вероятность $p^r_{\text{doc-acc}}$. Данная величина будет критерием оптимизации Q, а матрица будет являться начальным элементом популяции. В текущей задаче данная вероятность равна 0,99.

Зададим множество *D*. В силу того, что в модельном примере рассматривается только тип доступа чтение, данное множество также может быть задано матрицей (табл. II).

ТАБЛИЦА II МИНИМАЛЬНЫЕ ПРАВА ДОСТУПА

Пользователи	Документы											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
u1	+											
u2										+		
u3											+	+
u4	+					+						
u5								+				
u6						+						
u7				+	+							
u8			+				+		+			

В этом случае для каждого документа может быть аналогично рассчитана полная вероятность доступа к каждому файлу, а затем может быть рассчитана вероятность $p^r_{\text{doc-acc}}$. Данная величина является граничным значением $Q_{\min}$. В текущей задаче данная вероятность равна 0,75.

Каждый новый элемент популяции, генерируемый генетическим алгоритмом, изменяет одну из связей «пользователя – документ», т.е. отменяет/добавляет связь или изменяет ее тип. В случае рассматриваемого примера, связь может быть отменена или добавлена, так как рассматривается только один тип связи. Для каждого нового элемента популяции производится расчет вероятности Q, проверяется выполнение условия *D*, а также вхождение вероятности Q в диапазон $Q_{\min} < Q < Q_{\max}$. В случае, если это условие выполняется, то работа генетического алгоритма прекращается, а найденный элемент популяции рассматривается как приемлемая оптимизация прав доступа пользователей информационной системы к критичной информации.

С помощью генетического алгоритма ищется такая конфигурация прав доступа, что она соответствует критерию остановки (соблюдены минимальные права доступа и вероятность Q не превышает заданный порог $Q_{\max}$) и при этом получена из исходной конфигурации минимальным числом изменений. Все расчеты выполнены в среде R [8], пакет genalg [12].

Допустим, что заданный собственниками максимальный уровень риска, допустимый в информационной системе, равен 0,8. Тогда величина $Q_{\max} = 0{,}8$. Результаты работы алгоритма представлены на рис. 1, а итоговая конфигурация прав доступа в табл. III. Общая вероятность $p^r_{\text{doc-acc}}$ для указанной конфигурации равна 0,797.

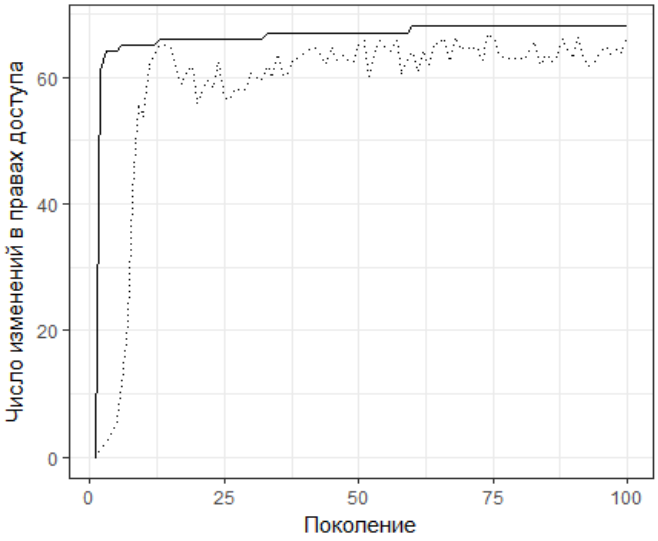

Рис. 1. Сходимость генетического алгоритма для $Q_{\max} = 0{,}8$.

ТАБЛИЦА III КОНФИГУРАЦИЯ ПРАВ ДОСТУПА ПРИ $Q_{\max} = 0{,}8$

Пользователи	Документы											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
u1	+					+						
u2										+		
u3											+	+
u4	+					+		+				
u5								+				
u6						+						
u7				+	+							
u8			+			+	+	+	+			

Рассмотрим еще один пример. Допустим, что заданный собственниками максимальный уровень риска,

допустимый в информационной системе, равен 0,88. Тогда величина $Q_{\max} = 0{,}88$. Результаты работы алгоритма представлены на рис. 2, а итоговая конфигурация прав доступа в табл. IV. Общая вероятность $p^{r}_{\text{doc-acc}}$ для указанной конфигурации равна 0,877.

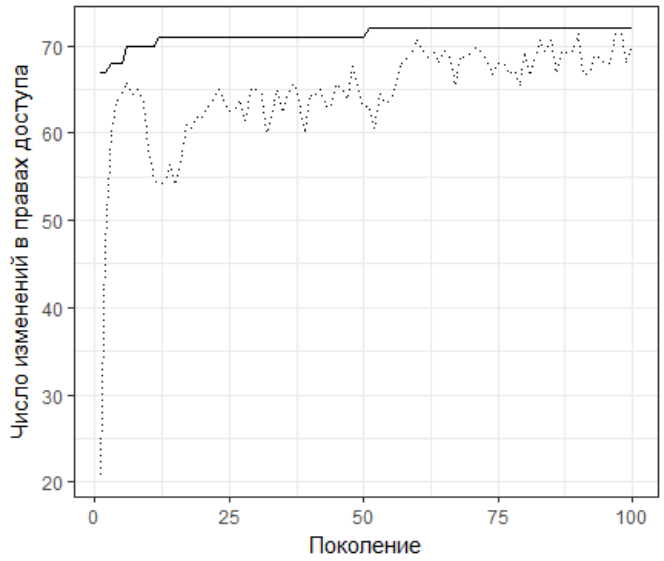

Рис. 2. Сходимость генетического алгоритма для $Q_{\max} = 0{,}88$

ТАБЛИЦА IV КОНФИГУРАЦИЯ ПРАВ ДОСТУПА ПРИ $Q_{\max} = 0{,}88$

Пользо-ватели	**Документы**											
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***
u1	+					+						+
u2				+						+		
u3	+										+	+
u4	+					+		+	+			
u5								+				
u6						+						
u7				+	+							
u8			+			+	+	+	+			

V. ЗАКЛЮЧЕНИЕ

В статье рассмотрен метод анализа устойчивости структуры политик доступа пользователей корпоративной информационной системы при социоинженерных атакующих воздействиях злоумышленника. Подход основывается на данных о степени выраженности уязвимостей пользователей информационных систем, расчете вероятностных оценок, а также формировании более устойчивой к социоинженерным атакующим воздействиям конфигурации политик доступа пользователей на основании генетических алгоритмов. Отметим, что предложенный подход позволяет построить различные конфигурации политик доступа к конфиденциальной информации, хранимой в информационной системе, соблюдая ряд ограничений и требований. Также в статье приведены примеры расчета устойчивости графов.

СПИСОК ЛИТЕРАТУРЫ

[1] Абрамов М.В., Азаров А.А., Тулупьева Т.В., Тулупьев А.Л. Модель профиля компетенций злоумышленника в задаче анализа защищённости персонала информационных систем от социоинженерных атак // Информационно-управляющие системы. 2016. №4. С. 77–84.

[2] Азаров А.А., Тулупьев А.Л., Соловцов Н.Б., Тулупьева Т.В. SQL-представление реляционно-вероятностных моделей социо-инженерных атак в задачах расчета агрегированных оценок защищенности персонала информационной системы с учетом весов связей между пользователями // Труды СПИИРАН. 2013. Вып. 24. С. 41–53.

[3] Азаров А.А., Тулупьева Т.В., Суворова А.В., Тулупьев А.Л., Абрамов М.В., Юсупов Р.М. Социоинженерные атаки. Проблемы анализа. СПб.: Наука, 2016. 352 с.

[4] Azarov A.A., Abramov M.V., Tulupyeva T.V., Tulupyev A.L. Users' of Information System Protection Analysis from Malefactor's Social Engeneering Attacks Taking into Account Malefactor's Competence Profile // Biologically Inspired Cognitive Architectures (BICA) for Young Scientists. 2016. P. 25–30.

[5] Gupta B.B., Tewari A., Jain A.K., Agrawal D.P. Fighting against phishing attacks: state of the art and future challenges // Neural Computing and Applications. 2017. Vol. 28, No. 12. P. 3629–3654.

[6] Huda A.S.N., Živanović R. Accelerated distribution systems reliability evaluation by multilevel Monte Carlo simulation: implementation of two discretisation schemes // IET Generation, Transmission & Distribution. 2017. Vol. 11, No. 13. P. 3397–3405.

[7] Liu J., Lyu Q., Wang Q., Yu X. A digital memories based user authentication scheme with privacy preservation // PloS ONE. 2017. Vol. 12, No. 11. P. 0186925.

[8] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. URL: https://www.R-project.org/

[9] Schaik P., Jeske D., Onibokun J., Coventry L., Jansen J., Kusev P. Risk perceptions of cyber-security and precautionary behavior // Computers in Human Behavior. 2017. Vol. 62, Issue 11. P. 5678–5693.

[10] Struharik R., Vukobratović B. A system for hardware aided decision tree ensemble evolution // Journal of Parallel and Distrib-uted Computing. 2018. Vol. 112. P. 67–83.

[11] Terlizzi M.A., Meirelles F.S., Viegas Cortez da Cunha M.A. Behavior of Brazilian Banks Employees on Facebook and the Cybersecurity Governance // Journal of Applied Security Research. 2017. Vol. 12, No. 2. P. 224–252.

[12] Willighagen E., Ballings M. genalg: R Based Genetic Algorithm. R package version 0.2.0. 2015. URL: https://CRAN.R-project.org/package=genalg