МАСКИРОВАНИЕ ТОПОЛОГИЧЕСКИХ СВОЙСТВ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ В УСЛОВИЯХ СЕТЕВОЙ РАЗВЕДКИ. Часть 1

Горбачев А. А.[1]

DOI: 10.21681/2311-3456-2024-6-130-139

Цель исследования: *исследование моделей случайных графов и генетических алгоритмов для решения задачи синтеза ложной структуры для маскирования топологических свойств вычислительных сетей при генерации ложного сетевого трафика и применении ложных сетевых информационных объектов, с учетом степени сходства топологических свойств реальных вычислительных сетей с ложными, а также с учетом показателя защищенности вычислительных сетей.*

Используемые методы: *генетический алгоритм оптимизации, метод линейной свертки, модель Эрдеша-Реньи, Барбаши, Харари.*

Результат исследования: *синтез ложной структуры вычислительной сети на основе моделей случайных графов и эволюционных алгоритмов оптимизации позволяет повысить результативность защиты вычислительной сети за счет снижения возможностей злоумышленника по идентификации ее критических узлов посредством анализа сетевого трафика. В качестве показателя близости топологических характеристик вычислительных сетей выступает коэффициент Жаккара между множествами ребер истинной и ложной вычислительных сетей, а в качестве аппроксимации дистанции между истинными и ложными критическими узлами выступает среднее кратчайшее расстояние. Генетические алгоритмы позволяют решить задачу оптимальной параметризации моделей случайных графов с точки зрения выбранной функции приспособленности, а также при явной комбинаторной оптимизации ложной топологии. Экспоненциальный рост переборного пространства не позволяет решать задачу комбинаторной оптимизации матрицы смежности графа, характеризующего топологию вычислительной сети большого размера, что приводит к необходимости использования методов снижения размерности и параметрических моделей при маскировании топологических свойств составных вычислительных сетей.*

Научная новизна: *заключается решении задачи синтеза топологических свойств ложной вычислительной сети с использованием генетических алгоритмов и моделей случайных графов, параметризованных с учетом скалярной целевой функции приспособленности, включающей показатель близости ложной и истинной топологической структуры вычислительной сети, а также аппроксимацию расстояния между истинными и ложными критическими узлами вычислительной сети.*

Ключевые слова: *анализ сетевого трафика, проактивная защита, ложные сетевые информационные объекты, эволюционные алгоритмы оптимизации, критические узлы.*

Введение

Критически важным этапом реализации кибератак является сетевая (компьютерная) разведка (*network reconnaissance*). Определение структурно-функциональных характеристик узлов вычислительной сети (*IP*-адресов, *MAC*-адресов, состояний *TCP*, *UDP*-портов), версий программного обеспечения, служб и сервисов, версий используемых сетевых протоколов, настроек межсетевых экранов и средств антивирусной защиты, а также других сведений, позволяет выявить уязвимости или подобрать соответствующие алгоритмы для реализации сетевой (компьютерной) атаки [1, 2].

Большинство вычислительных сетей, функционирующих в интересах различных организаций имеют статичную структуру (топологию), которая может быть идентифицирована с использованием различных методов сетевой разведки. Наиболее распространенным методом построения топологии вычислительной сети является анализ сетевого трафика (сниффинг, прослушивание), проходящего через интерфейсы коммутационного оборудования, прокси-серверов и другие контролируемые злоумышленниками узлы вычислительной сети, а также активное сканирование вычислительной сети с использованием таких инструментов как *Nmap*, *Nessus*, *Wireshark*, *ZENMap*, *Sparta*, *OpenVAS* (соответствующие тактики и техники реализации описаны в матрице *MITRE ATT&CK*: *TA*0043, *T*1595, *T*1590[2], и в методике оценки угроз информационной безопасности ФСТЭК России: Т1.3, Т1.4). В связи с тем, что различные типы активного сканирования сетевых узлов вычислительной сети содержатся в сигнатурах

1 Горбачев Александр Александрович, кандидат технических наук, Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С. М. Штеменко, г. Краснодар, Россия. E-mail: infosec23.00@mail.ru

2 MITRE ATT&CK. Enterprice: Reconnaissance. https://attack.mitre.org/tactics/TA0043/ (дата обращения 01.07.24 г.)

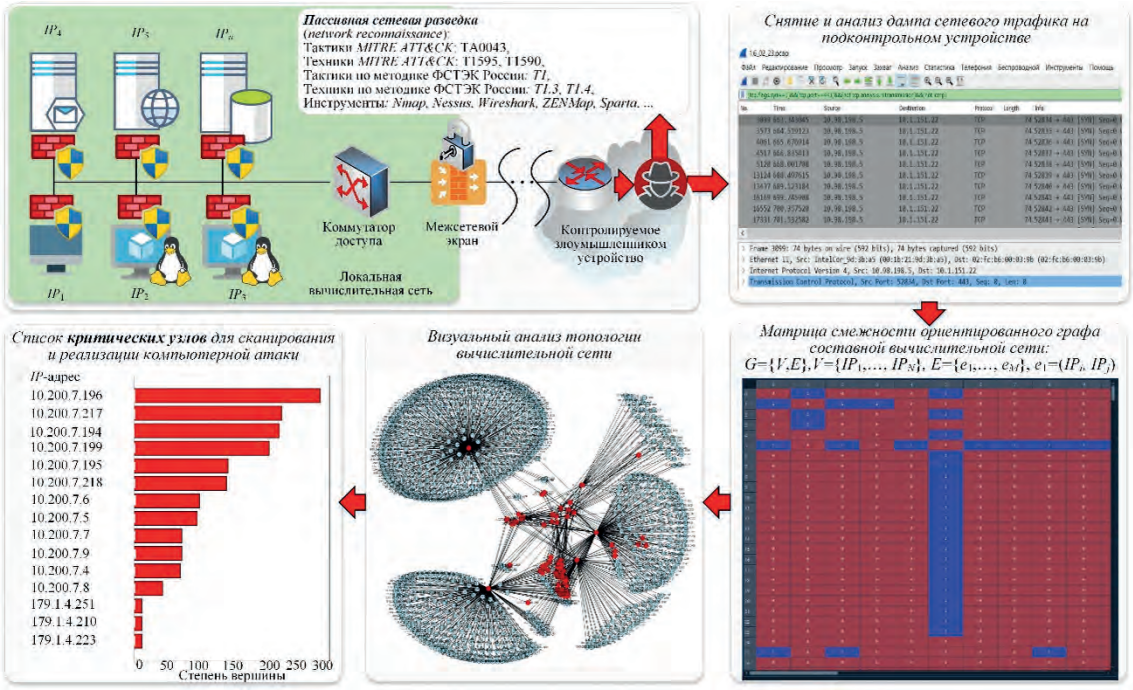

Рис. 1. Процесс пассивной сетевой разведки злоумышленника, направленной на вскрытие топологических свойств вычислительной сети

и правилах *SIEM*-систем (*Security Information and Event Management*), средств обнаружения вторжений и межсетевых экранов, то наиболее безопасным и скрытным способом сетевой разведки для злоумышленников является пассивный анализ сетевого трафика [3].

После реконструкции топологии вычислительной сети злоумышленником может быть составлен перечень наиболее важных для дальнейшего исследования и информационно-технического воздействия узлов вычислительной сети с точки зрения их топологических свойств (рисунок 1).

Работа посвящена методам генерации ложного сетевого трафика с заданной ложной топологией защищаемой вычислительной сети с целью снижения эффективности действий злоумышленников, реализующих сетевую разведку посредством пассивного анализа сетевого трафика. В зарубежной литературе аналогичный метод введения злоумышленника в заблуждение называется *киберобманом* посредством *обфускации топологии* вычислительной сети, при этом синтез ложной топологии осуществляется посредством: случайного изменения параметров маршрутизации заголовков пакетов в соответствии с заданными критериями и ограничениями[3] [4], добавления логических связей между узлами вычислительной сети с учетом распределения степеней вершин [5], использования виртуальных маршрутизаторов[4] и сетевых ловушек [6, 7]. В отечественной литературе маскирование истинной топологии вычислительной сети рассматривалось посредством трансляции сетевых адресов, расширения адресного пространства, введения ложных объектов[5] [8, 9], обеспечения работоспособности сложных систем в условиях деструктивных информационных воздействий [10, 11].

Тем не менее, анализ качества различных моделей, алгоритмов и постановок задач синтеза ложной топологии вычислительной сети с учетом размерности в полной мере не был реализован. В приведенном материале синтез ложной топологии вычислительной сети осуществляется посредством решения задачи комбинаторной оптимизации матрицы смежности ориентированного графа с использованием генетического алгоритма для вычислительных сетей низкой размерности, а также с использованием классических моделей случайных графов, реализующих синтез ложной структуры вычислительной сети с заданными свойствами близости к топологии реальной сети и защищенности от деструктивных воздействий на критические узлы, качество ложной структуры также оценивается как степень пересечения множества ложных и истинных критических узлов вычислительной сети.

3 Roland Meier, Petar Tsankov, Vincent Lenders, Laurent Vanbever, Martin Vechev. NetHide: Secure and Practical Network Topology Obfuscation. Proceedings of the 27-th USENIX Security Symposium. 2018. pp. 693–709.

4 Stefan Achleitner, Thomas F. La Porta, Patrick McDaniel, Shridatt Sugrim, Srikanth V. Krishnamurthy, Ritu Chadha. Deceiving Network Reconnaissance Using SDN-Based Virtual Topologies // IEEE Transactions On Network And Service Management, Vol. 14, No. 4, 2017. pp. 1098–1112.

5 Шерстобитов Р. С., Шарифуллин Р. С., Максимов Р. В. Маскирование интегрированных сетей связи ведомственного назначения // Системы управления, связи и безопасности. 2018. № 4. С. 136–175.

Анализ объекта исследования

Вычислительная сеть (компьютерная сеть, сеть передачи данных) представляет собой совокупность узлов (компьютеров) и аппаратуры передачи данных, ветви которых являются линиями передачи данных. Под топологией вычислительной сети в контексте данной работы будет пониматься ориентированный и невзвешенный граф (орграф).

Ориентированный граф, характеризующий топологию вычислительной сети с N узлами может быть однозначно определен компонентами (выражение 1):

$$G = \{V,E\}, \quad (1)$$

где $V = \{v_1, \ldots, v_N\}$ – вершины графа, применительно к рассматриваемой области $V = \{IP_1, \ldots, IP_N\}$ вершинами графа являются *IP*-адреса сетевых узлов; $E = \{e_1, \ldots, e_N\}$ – ребра графа, причем $e_1 = (v_i, v_j)$ – направленное ребро от узла v_i до ребра v_j или для вычислительной сети $e_1 = (IP_i, IP_j)$.

Для математического моделирования ориентированных графов как правило используют матричное представление графа в виде *матрицы смежности* или *матрицы инциденций*. Матрица смежности представляет собой квадратную матрицу размерностью N с бинарными значениями $A \in \{0,1\}^{N\times N}$, причем для орграфов матрица смежности в общем случае не является симметричной.

Под топологическими свойствами в работе понимаются характеристики ориентированного графа, которые с точки зрения формы их представления можно разделить на *скалярные* и *векторные* (*распределения*). Скалярные характеристики представляют собой функционал от компонентов орграфа G.

С целью создания ***правдоподобной ложной топологии*** вычислительной сети могут быть использованы различные характеристики *близости* топологических свойств орграфов, характеризующих реальную и ложную вычислительные сети. С одной стороны, указанная близость может быть оценена как отклонение между функционалами (*норма Фробениуса*) или как *мера схожести* (*коэффициент Жаккара*, *d-мера*, *δ-мера*) от матриц смежности реальной сети A_{real} и ложной сети A_{synt}. С другой стороны, характеристиками близости могут выступать любые скалярные или векторные характеристики орграфа, к примеру, количество ребер $M = |E|$, вершин $N = |V|$, спектральный радиус, коэффициент кластеризации, распределение входящих и исходящих степеней вершин. Для оценки качества аппроксимации векторных характеристик графов могут быть использованы различные метрики дистанций распределений (дивергенция Кульбака-Лейблера, статистические критерии проверки гипотез об однородности распределений). Вопрос использования тех или иных показателей близости структур является дискуссионным, но с точки зрения вычислительной сложности и физической интерпретируемости в качестве показателей качества аппроксимации топологических свойств реальной вычислительной сети посредством ложной в работе рассматриваетс я коэффициент Жаккара $J(A_{real}, A_{synt})$ в соответствии с выражением (2). Для удобства решения задачи оптимизации за счет минимизации целевой функции, характеризующей степень близости ложной и реальной вычислительной сети, форма коэффициента Жаккара имеет вид:

$$K_{sym1} = J(A_{real}, A_{synt}) = 1 - \frac{|E_{real} \cap E_{synt}|}{|E_{real} \cup E_{synt}|}, \quad (2)$$

где $E_{real} = \{e^{real}_1,\ldots,e^{real}_k\}$ – множество ребер орграфа реальной вычислительной сети, $E_{synt} = \{e^{synt}_1,\ldots,e^{synt}_n\}$ – множество ребер орграфа ложной вычислительной сети.

В контексте оценки ***защищенности*** вычислительных сетей наиболее важными топологическими характеристиками ориентированных графов являются: *характеристики устойчивости структуры в целом к деструктивным воздействиям и характеристики важности отдельных узлов*.

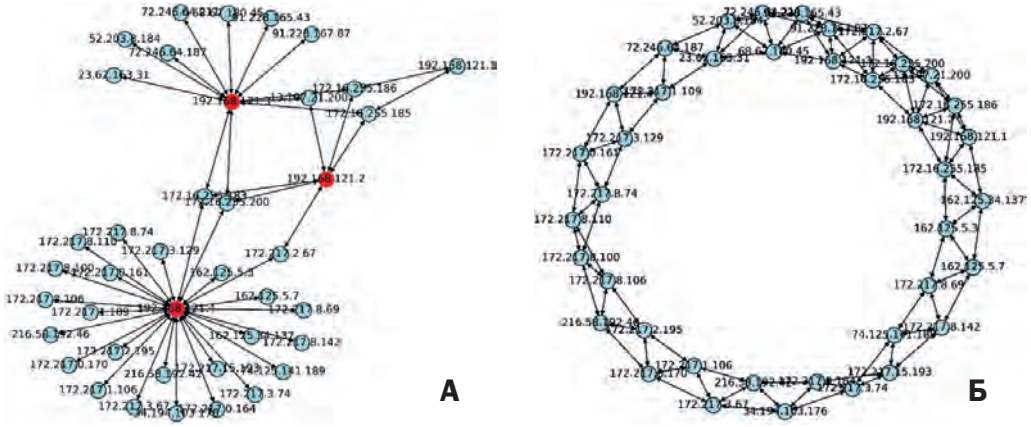

Рис. 2. Граф Харари (б) с количеством вершин N = 40 и количеством ребер M = 92, соответствующий реальной вычислительной сети (а). Средний коэффициент кластеризации исходного графа (а) – 0,0; графа Харари (б) – 0,5375.

Коэффициент кластеризации позволяет оценить общую устойчивость топологии вычислительной сети к деструктивным воздействиям. Например, при высоком коэффициенте кластеризации орграфа удаление случайного ребра или вершины приведет к незначительным негативным последствиям. Граф *Харари* (рисунок 2, б) характеризуется высоким коэффициентом кластеризации и соответственно высокой устойчивостью к деструктивным воздействиям, направленным на нарушение связности сети при фиксированном количестве вершин и ребер графа, поэтому данную модель топологии также целесообразно использовать при построении ложной топологии вычислительной сети, которая с точки зрения злоумышленника будет иметь меньше уязвимых узлов.

Несмотря на то, что адекватную модель злоумышленника в общем случае построить не представляется возможным, существует возможность выделить несколько предпочтений, которые могут быть сделаны на основе анализа нарушителями топологических свойств вычислительной сети. В связи с ограниченностью ресурсов планирование и реализация атаки злоумышленниками ведется на подмножество наиболее важных или в некотором смысле ***критических узлов*** в вычислительной сети. Атака на критические узлы позволяет нарушить работоспособность сетевых сервисов для целых кластеров (подмножеств) вычислительной сети. Критичность узла может быть определена по нескольким топологическим признакам:

- критические узлы с наибольшим значением *степени вершины* или значением степени, значительно превосходящим степени других узлов (вершин). Интуитивные соображения подсказывают, что важные узлы вычислительной сети реализуют информационный обмен с большим числом узлов в сети;
- критические узлы с наибольшим значением *коэффициента связности*, значительно превосходящим другие узлы;
- критические узлы как *артикуляционные узлы* графа. Вершина графа $v_k = \{IP_k\}$ называется артикуляционной тогда и только тогда, когда некоторое наименьшее вершинное покрытие графа содержит эту вершину[6]. Наименьшее вершинное покрытие графа — это минимальный набор вершин, таких, что каждое ребро графа инцидентно хотя бы одной вершине из этого набора. Это означает, что каждая вершина в наименьшем вершинном покрытии является важной для обеспечения связности между элементами графа. Удаление артикуляционных узлов приведет к нарушению связи у целых подсетей или подмножеств подсетей. Одним из распространенных и эффективных способов нахождения артикуляционных узлов является алгоритм *Тарьяна* с линейной временной сложностью $O(|V|+|E|)$[7].

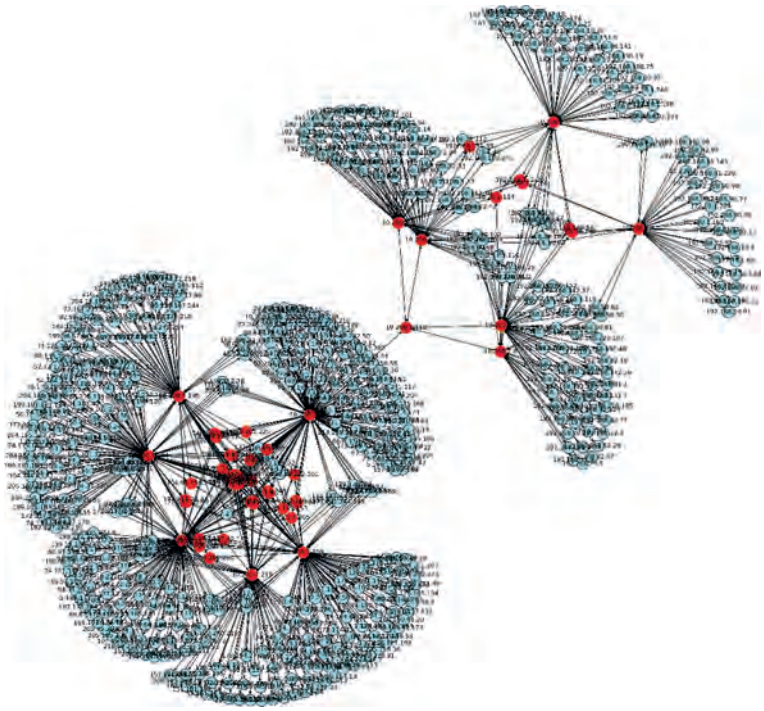

Рис. 3. Реконструкция топологии составной вычислительной сети из общедоступного дампа трафика[8]: N = 643 шт.; E = 1710 шт.; красным цветом обозначены критические узлы со степенями выше, чем 99-й процентиль степеней графа

Рассмотрим визуализацию топологии вычислительной сети, полученную посредством анализа дампа трафика из общедоступного репозитория (рисунок 3). Анализ сетевого трафика за длительный промежуток времени позволяет восстановить топологию составной вычислительной сети с учетом идентификации критических узлов (на рисунке 3 критические узлы определены, исходя из степеней вершин выше, чем 99-й процентиль всех степеней графа, то есть критическими являются узлы, степени которых больше, чем у 99 % остальных узлов сети). Как видно из рисунка, простейшее отсечение узлов с высокими степенями вершин позволяет идентифицировать наиболее важные с точки зрения связности и обеспечения доступности информационных ресурсов узлы вычислительной сети.

6 Харари Ф. Теория графов / Ф. Харари. М: 1973, 300 с.
7 Farima G. A linear time algorithm to compute the impact of all the articulation points // arXiv:1504.00341v3 [cs.DS]. – 2015 (дата обращения 01.07.24 г.)
8 Kaggle: IP Network Traffic Flows Labeled with 75 Apps. https://www.kaggle.com/datasets/jsrojas/ip-network-traffic-flows-labeled-with-87-apps (дата обращения 01.07.24 г.)

Решение задачи синтеза топологии ложной вычислительной сети

Для снижения возможностей злоумышленника по реализации информационно-технических воздействий на критические узлы исходя из топологических свойств вычислительной сети целесообразной является гипотеза о том, что синтезированная структура должна обладать критическими узлами, отличными от реальных критических узлов (по относительному расположению), более того необходимо расположить их на максимальной дистанции от реальных критических узлов, то есть максимизировать среднее кратчайшее расстояние от ложных до реальных критических узлов.

То есть вербальная постановка задачи на синтез ориентированного графа ложной вычислительной сети заключается в подборе такой модели и ее параметров, которая будет генерировать орграфы с максимальной степенью близости к исходному графу и при этом максимальным средним кратчайшим расстоянием от ложных до реальных узлов.

Решение указанной задачи синтеза ложной структуры вычислительной сети в общем виде можно определить в форме задачи многокритериальной оптимизации (выражение 3):

$$\begin{cases} K_{sim}(S, \theta, A_{real}) \rightarrow \underset{K_{sim},K_{def},S,\theta,A_{real} \in Q_1}{extr} \\ K_{def}(S, \theta, A_{real}) \rightarrow \underset{K_{sim},K_{def},S,\theta,A_{real} \in Q_1}{extr} \end{cases} \quad (3)$$

где S – форма модельного оператора, определяющая класс моделей, аппроксимирующих соответствующие свойства орграфа вычислительной сети; $\theta = (\theta_1, \ldots, \theta_m)$ – в общем случае вектор параметров соответствующего модельного оператора; Q_1 – допустимое множество значений целевых функций и аргументов; $K_{sim}(S, \theta, A_{real})$ – функция качества аппроксимации близости топологии реальной вычислительной сети; $K_{def}(S, \theta, A_{real})$ – функция, характеризующая количественную оценку защищенности реальной вычислительной сети при генерации заданной ложной топологии вычислительной сети.

Решение подобных задач сводится к нахождению Парето оптимального множества решений либо к выбору одного из методов *скаляризации*, к примеру, метода главного критерия (целевой функции), идеальной точки, линейной свертки целевой функции или использования других отношений предпочтения в критериальном пространстве. Причем, если $\theta = A$, то есть если в качестве параметров рассматривать элементы матрицы смежности, то синтез структуры осуществляется как решение задачи *комбинаторной оптимизации* матрицы смежности с количеством параметров N^2, а алгоритм нахождения экстремума целевой функции имеет нижнюю оценку вычислительной сложности $O(N^2)$. Размерность пространства решений задачи полного перебора значений элементов матрицы смежности составляет 2^{N^2}. Решение подобной задачи целесообразно лишь для графов малой размерности (при $N < 50$).

Рассмотрим синтез структуры вычислительной сети как задачу комбинаторной оптимизации скалярной целевой функции $f_1(\alpha_1, \alpha_2, S, \theta, A_{real})$, представляющей собой линейную свертку (взвешенную сумму) функций $K_{sim}(S, \theta, A_{real})$ и $K_{def}(S, \theta, A_{real})$. Постановка задачи с непосредственным нахождением элементов матрицы смежности удовлетворяет задаче (выражение 4):

$$f_1(\alpha_1, \alpha_2, S, \theta, A_{real}) = \alpha_1 \cdot K_{sim}(S, \theta, A_{real}) + \\ + \alpha_2 \cdot K_{def}(S, \theta, A_{real}) \rightarrow \underset{K_{sim},K_{def},S,\theta,A_{real},\alpha_1,\alpha_2 \in Q_1}{extr}, \quad (4)$$

где α_1, α_2 – коэффициенты значимости функций $K_{sim}(S, \theta, A_{real})$ и $K_{def}(S, \theta, A_{real})$ соответственно.

В работе в качестве характеристики близости структур $K_{sim}(S, \theta, A_{real})$ выступает коэффициент Жаккара $J(A_{real}, A_{synt})$ между генерируемой матрицей смежности A и матрицей смежности реального графа A_{real}. Характеристикой защищенности $K_{def}(S,\theta,A_{real})$ структуры является функция $D(A_{real}, A)$ как среднее кратчайшее расстояние между критическими узлами генерируемой матрицы смежности A и критическими узлами исходной матрицы A_{real} (выражение 5, 6):

$$f_1(\alpha_1, \alpha_2, S, \theta, A_{real}) = \alpha_1 \cdot J(A_{real}, A) + \\ + \alpha_2 \cdot (D(A_{real}, A) + \varepsilon)^{-1} \rightarrow \underset{A_{real},A,\alpha_1,\alpha_2 \in Q_3}{min} \quad (5)$$

$$Q_3 = \begin{cases} \alpha_1 \in [0,1], \alpha_2 \in [0,1], \\ J(A_{real}, A) \in [0,1], \\ D(A_{real}, A) \geq 0, \\ A \in \{0,1\}^{N \times N}, A_{real} \in \{0,1\}^{N \times N}, \\ N \leq 30, \varepsilon = 1{,}0. \end{cases} \quad (6)$$

где ε – коэффициент, предотвращающий деление на 0; $D(A_{real}, A)$ – функция, вычисляющая среднее кратчайшее расстояние между подмножеством критических узлов графа, восстановленного из матрицы смежности A_{real} реальной вычислительной сети и подмножеством критических узлов графа, восстановленного из матрицы смежности A ложной вычислительной сети.

Алгоритм вычисления функции $D(A_{real}, A)$ включает в себя:

- вычисление списка критических узлов реальной сети по матрице смежности A_{real}, для чего в зависимости от критерия важности узлов осуществляют сортировку списка критических узлов по степени важности (степени вершины или степени связности вершины);
- вычисление списка критических узлов ложной сети по матрице смежности A;

- вычисление кратчайшего расстояния между каждым критическим узлом реальной и ложной вычислительной сети с использованием алгоритма Дейкстры;
- вычисление среднего значения кратчайшего расстояния между критическими узлами реальной и ложной вычислительной сети.

Алгоритмическая реализация функции $D(A_{real}, A)$ может быть дополнена штрафными слагаемыми за пересечение списков критических узлов ложной и реальной вычислительной сети.

Для решения задач комбинаторной оптимизации с большим количеством переменных часто используют генетические алгоритмы оптимизации, демонстрирующие приближенное решение задач об N ферзях, о рюкзаке, коммивояжера и маршрутизации транспорта с допустимой точностью и за приемлемое время [12].

Использование простого генетического алгоритма для решения задач численной оптимизации включает в себя этапы:

- ***создание начальной популяции:*** начальная популяция представляет собой количество разглаженных векторов, содержащих значения 0 или 1, длиной N^2. В работе размер начальной популяции составил от 100 до 1000 индивидуумов (векторов) в зависимости от размерности матрицы смежности A_{real};
- ***вычисление приспособленности*** каждого индивидуума: в качестве функции приспособленности выступает скалярная целевая функция $f_1(\alpha_1, \alpha_2, A_{real}, A)$, весовые коэффициенты функций $\alpha_1 = \alpha_2 = 1$.
- ***отбор:*** включает в себя процедуру турнирного отбора индивидуумов-родителей для индивидуумов следующей итерации алгоритма. Размер турнира является гиперпараметром и составляет величину от 3 до 10;
- ***скрещивание:*** в работе реализовано одноточечным скрещиванием, при этом вероятность скрещивания каждой особи в поколении составляет величину 0,5;
- ***мутация:*** способ мутации – инвертирование бинарного значения элемента вектора (индивида), вероятность мутации составляет величину 0,2;
- ***проверка критерия остановки алгоритма:*** достижение заданного количества итераций расчета (от 100 до 1000 в зависимости от размерности N).
- ***выбор*** индивидуумов с максимальной приспособленностью (минимальным значением целевой функции f_1).

Для реализации вычислительного эксперимента по синтезу ложной структуры вычислительной сети с использованием генетического алгоритма и решения задачи комбинаторной оптимизации был использован дамп трафика из открытого источника[9], из которого была восстановлена топология вычислительной сети с заданным количеством узлов N.

Результаты вычислений, представленные на рисунке 4 показывают, что при использовании генетического алгоритма полученные структуры характеризуются более равномерной связностью (степенями вершин), а критические узлы, вычисленные по признаку степени вершины (значение степени выше 99-го процентиля степеней в графе) либо отсутствуют (рисунок 4, а), либо переместились на другие вершины (рисунок 4, б). Стоит отметить высокую вычислительную сложность процесса поиска оптимальной в указанном смысле структуры, которая в свою очередь ограничивает применимость генетических алгоритмов и решения задачи комбинаторной оптимизации элементов матрицы смежности A_{real} лишь для локальных вычислительных сетей с небольшим количеством узлов $N < 50$.

Оценка качества маскирования структуры вычислительной сети

Как предполагалось ранее, злоумышленник стремится к минимизации вероятности компрометации факта сетевой разведки узлов вычислительной сети, что приводит к тому, что в качестве мишеней для дальнейшего исследования (сканирования) и реализации компьютерных атак выбирается ограниченное подмножество узлов вычислительной сети, исходя из множества узлов, вскрытого посредством пассивного анализа сетевого трафика.

В качестве критерия отбора узлов по степени их важности рассмотрим степень вершины. В указанных условиях в качестве количественного показателя защищенности вычислительной сети может выступать *степень пересечения* множества критических узлов, отобранного злоумышленником на основе анализа ложной вычислительной сети с множеством критических узлов реальной вычислительной сети (выражение 7):

$$\Omega(S_{real}, S_{id}, N, \Delta) = \frac{|S_{real}(N, \Delta) \cap S_{id}(N, \Delta)|}{|S_{real}(N, \Delta)|} \quad (7)$$

где $S_{real} = \{IP^{real}{}_1,\ldots,IP^{real}{}_k\}$ – множество критических узлов реальной вычислительной сети, $S_{id} = \{IP^{id}{}_1,\ldots, IP^{id}{}_n\}$ – множество критических узлов, отобранных злоумышленником в качестве критических на основе анализа ложной топологии вычислительной сети, Δ – доля вершин графа, определяющая длину ранжированного по степени списка критических узлов графа, то есть, $\Delta = 0{,}1$, означает, что злоумышленником будет

9 Kaggle: Labeled Network Traffic flows – 141 Applications. https://www.kaggle.com/datasets/jsrojas/labeled-network-traffic-flows-114-applications/data (дата обращения 01.07.24 г.).

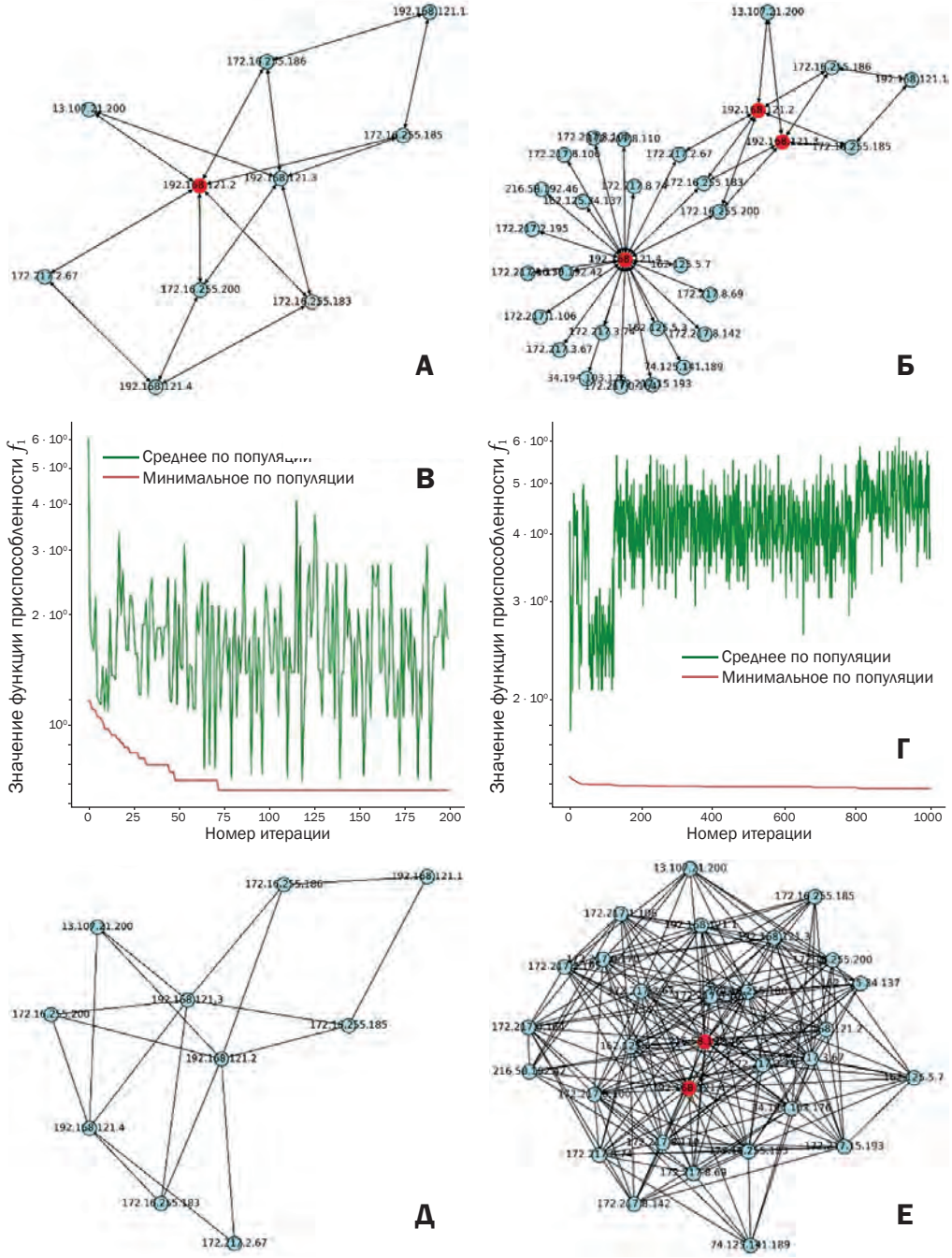

Рис. 4. Синтез топологии ложной вычислительной сети посредством решения задачи комбинаторной оптимизации генетическим алгоритмом в зависимости от количества вершин N: исходный граф для N = 10 шт. (а), для N = 30 шт. (б); процесс поиска минимума функции приспособленности для N = 10 шт. (в), для N = 30 шт. (г); результат синтеза ложной структуры для N = 10 шт. (д), время расчета – 6,481 с, для N = 30 шт. (е), время расчета – 839,432 с

составлен список критических узлов, ранжированный по степени и длина которого равна 10% от общего количества узлов подсети.

Для сравнительной характеристики эффективности максирования топологических свойств вычислительной сети были использованы классические модели случайных графов Эрдеша-Реньи, Барбаши и Харари. Параметры соответствующих моделей были оптимизированы с точки зрения показателя близости J структур, то есть, для модели Эрдеша-Реньи параметрическая идентификация имеет вид (выражение 8):

$$J_{ER}\left(p_{ER}, N_{ER}, A, A_{real}\right) \rightarrow \min_{N_{ER} = |N_{real}|,\ p_{ER} \in [0,1]}, \qquad (8)$$

где, p_{ER} – вероятность наличия ребра между вершинами графа Эрдеша-Реньи, N_{ER} – количество вершин графа Эрдеша-Реньи.

Для модели Барбаши параметрическая оптимизация имеет вид (выражение 9):

$$J_B(M_B, N_B, A, A_{real}) \rightarrow \min_{N_B = |N_{real}|,\ M_B \in Int,\ M_B \in [1, |N_{real}|-1]}, \quad (9)$$

где M_B – среднее количество ребер вершины графа Барбаши, N_B – количество вершин графа Барбаши.

Нахождение экстремумов функций J_{ER} и J_B является тривиальной задачи в связи с поиском оптимальных значений одного параметра в каждом случае, так как параметры $N_{ER} = N_B = N_{real}$ имеют фиксированное значение. Для модели Харари оптимальными условиями моделирования реальной вычислительной сети является равенство количества ребер $N_H = |N_{real}|$ и количества вершин $E_H = |E_{real}|$ графа Харари и реального графа соответственно.

Далее, используя указанные модели с оптимальными параметрами с точки зрения критерия близости, генерируется по 50 ложных вычислительных сетей с фиксированным количеством узлов вычислительной сети N и фиксированной долей Δ вершин вычислительной сети, относимой злоумышленником к критическим узлам (рисунок 5).

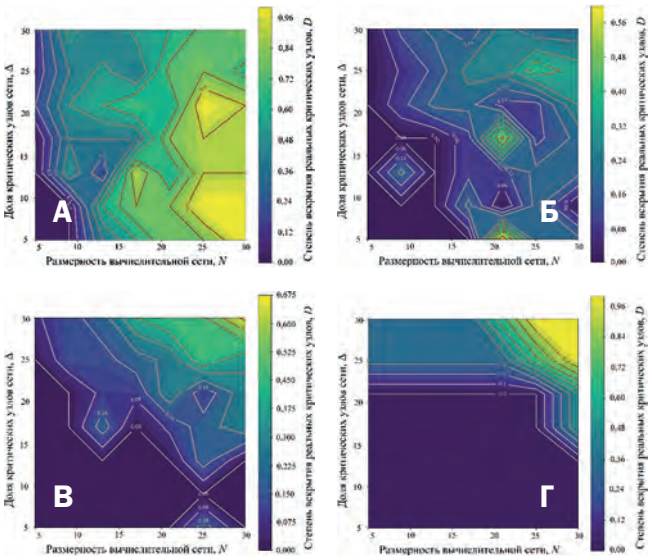

Рис. 5. Степень вскрытия злоумышленником реальных критических узлов вычислительной сети в зависимости от размерности сети N и доли Δ критических узлов в вычислительной сети: а) для генетического алгоритма; б) для модели Эрдеша-Реньи; в) для модели Барбаши; г) для модели Харари

Как видно из расчетов, при использовании генетического алгоритма (рисунок 5 а) степень вскрытия реальных критических узлов вычислительной сети D быстро растет с увеличением размерности сети N и доли Δ критических узлов в вычислительной сети, что, в свою очередь, может быть связано с тем, что поиск оптимальной структуры осуществляется исходя из двух показателей качества с одной стороны и приближенностью полученного решения с другой стороны. Использование моделей Эрдеша-Реньи и Барбаши (рисунок 5 б, в) показывают схожие результаты и лучшие по отношению к генетическому алгоритму с вышеуказанными диапазонами гиперпараметров. Также синтез ложной структуры по модели Харари позволяет синтезировать защищенную топологию и с точки зрения вскрытия критических узлов, и с точки зрения среднего коэффициента кластеризации. Также на результаты численного эксперимента оказывает влияние конкретный вид реальной вычислительной сети, по отношению к которой синтезируется ложная структура.

Полученные результаты позволяют сделать выводы:

- изолированное использование численных алгоритмов оптимизации, в частности, генетического алгоритма, при решении задачи комбинаторной оптимизации матрицы смежности нецелесообразно в связи с высокой вычислительной сложностью процесса поиска оптимальной структуры, обусловленной экспоненциальным ростом пространства возможных комбинаций;
- синтез структур с использованием классических моделей случайных графов, в частности моделей Эрдеша-Реньи, Барбаши, Харари, характеризуются относительно низкой вычислительной сложностью и при этом обеспечивают синтез структур вычислительных сетей, относительно более защищенных в рассмотренной метрике пересечения списка ложных и истинных критических узлов вычислительной сети;
- для решения задачи синтеза ложной топологии с большим количеством вершин ($N > 50$), что свойственно для структур, формируемых при анализе дампов трафика в информационно-телекоммуникационной сети общего пользования, целесообразно использовать либо относительно простые модели синтеза графов, рассмотренные в работе, либо методы снижения размерности поставленной задачи;
- дальнейшее направление исследований будет направлено на исследование эффективности совместного использования алгоритмов снижения размерности, машинного обучения, генетических алгоритмов и классических моделей случайных графов для синтеза ложных топологий вычислительных сетей большой размерности с заданными свойствами близости и защищенности.

Литература

1. Зегжда Д. П., Александрова Е. Б., Калинин М. О., Марков А. С. и др. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам / Под ред. профессора РАН, доктора техн. наук Д. П. Зегжды. – М.: Горячая линия – Телеком, 2019. – 560 с.
2. Stefan Marksteiner, Bernhard Jandl-Scherf and Harald Lernbeiß. Automatically Determining a Network Reconnaissance Scope Using Passive Scanning Techniques. Fourth International Congress on Information and Communication Technology. London. 2020. vol. 2. p. 117–127.
3. Дорофеев А. В., Марков А. С. Мониторинг событий информационной безопасности: технологии и методы контроля эффективности // Вестник военного инновационного технополиса «ЭРА». 2022. Т.3. № 4. С. 392–400.
4. Tao Hou, Tao Wang, Zhou Lu, Yao Liu. Combating Advarsarial Network Topology Inference by Proactive Topology Obfuscation. IEEE INFOCOM 2020. 2020. pp. 1–14.
5. Jinwoo Kim, Eduard Marin, Mauro Conti, Seungwon Shin. EqualNet: A Secure and Practical Defense for Long-term Network Topology Obfuscation. Network and Distributed Systems Security (NDSS) Symposium. 2022. pp. 1–18.
6. Rawski M. Network Topology Mutation as Moving Target Defense for Corporate Networks // INTL Journal Of Electronics And Telecommunications. 2019. Vol. 65, No. 4, pp. 571–577.
7. Hou T. et al. Proto: Proactive topology obfuscation against adversarial network topology inference // IEEE INFOCOM 2020-IEEE Conference on Computer Communications. – IEEE, 2020. Pp. 1598–1607.
8. Кучуров В. В., Максимов Р. В., Шерстобитов Р. С. Модель и методика маскирования адресации корреспондентов в киберпространстве // Вопросы кибербезопасности. 2020. № 6(40). С. 2–13. DOI:10.21681/2311-3456-2020-6-2-13
9. Теленьга А. П. Маскирование метаструктур информационных систем в киберпространстве // Вопросы кибербезопасности. 2024. № 5(57). С. 50–59. DOI:10.21681/2311-3456-2024-5-50-59
10. Зегжда Д. П. Интеллектуальные методы саморегуляции распределенных сетевых структур в условиях кибератак // XIV Всероссийская мультиконференция по проблемам управления МКПУ-2021. 2021. С. 16–19.
11. Лаврова Д. С., Зегжда Д. П., Зайцева Е. А. Моделирование сетевой инфраструктуры сложных объектов для решения задачи противодействия кибератакам // Вопросы кибербезопасности. 2019. №2(30). С. 13–20. DOI:10.21681/2311-3456-2019-2-13-20
12. Вирсански Э. Генетические алгоритмы на Python / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 286 с.: ил. ISBN 978-5-97060-857-9.

MASKING OF TOPOLOGICAL PROPERTIES OF COMPUTER NETWORKS IN NETWORK RECONNAISSANCE CONDITIONS. Part 1

Gorbachev A. A.[10]

The purpose of the study: *to study models of random graphs and genetic algorithms for solving the problem of synthesizing a false structure to mask the topological properties of computer networks when generating false network traffic and using false network information objects, taking into account the degree of similarity of the topological properties of real computer networks with false ones, as well as taking into account the security index of computer networks.*

Methods used: *genetic optimization algorithm, linear convolution method, Erdos-Renyi, Barbashi, Harari model.*

The result of the study: *the synthesis of a false structure of a computer network based on random graph models and evolutionary optimization algorithms makes it possible to increase the effectiveness of protecting a computer network by reducing the ability of an attacker to identify its critical nodes through network traffic analysis. The Jacquard coefficient between the sets of edges of true and false computer networks acts as an indicator of the proximity of the topological characteristics of computer networks, and the average shortest distance acts as an approximation of the distance between true and false critical nodes. Genetic algorithms make it possible to solve the problem of optimal parameterization of random graph models from the point of view of the selected fitness function, as well as with explicit combinatorial optimization of a false topology. The exponential growth of the bulkhead space does not allow solving the problem of combinatorial optimization of the adjacency matrix of a graph characterizing the topology of a large computer network, which leads to the need to use dimensionality reduction methods and parametric models when masking the topological properties of composite computer networks.*

Scientific novelty: *it consists in solving the problem of synthesizing the topological properties of a false computer network using genetic algorithms and random graph models parameterized taking into account the scalar fitness objective function, which includes an indicator of the proximity of the false and true topological structure of the computer network, as well as approximating the distance between true and false critical nodes of the computer network.*

Keywords: *network traffic analysis, proactive protection, honeypots, evolutionary optimization algorithms, critical nodes.*

10 Alexander A. Gorbachev , Ph.D. Krasnodar Higher Military Orders of Zhukov and the October Revolution Red Banner School named after General of the Army S. M. Shtemenko, Krasnodar, Russia. E-mail: infosec23.00@mail.ru

References

1. *Zegzhda D. P., Aleksandrova E. B., Kalinin M. O., Markov A. S. i dr. Kiberbezopasnost' cifrovoj industrii. Teoriya i praktika funkcional'noj ustojchivosti k kiberatakam / Pod red. professora RAN, doktora texn. nauk D. P. Zegzhdy'. – M.: Goryachaya liniya – Telekom, 2019. – 560 p.*
2. *Stefan Marksteiner, Bernhard Jandl-Scherf and Harald Lernbeiß. Automatically Determining a Network Reconnaissance Scope Using Passive Scanning Techniques. Fourth International Congress on Information and Communication Technology. London. 2020. vol. 2. p. 117–127.*
3. *Dorofeev A. V., Markov A. S. Monitoring sobytij informacionnoj bezopasnosti: tekhnologii i metody kontrolya effektivnosti // Vestnik voennogo innovacionnogo tekhnopolisa «ERA». 2022. T.3. № 4. pp. 392–400.*
4. *Tao Hou, Tao Wang, Zhou Lu, Yao Liu. Combating Advarsarial Network Topology Inference by Proactive Topology Obfuscation. IEEE INFOCOM 2020. 2020. pp. 1–14.*
5. *Jinwoo Kim, Eduard Marin, Mauro Conti, Seungwon Shin. EqualNet: A Secure and Practical Defense for Long-term Network Topology Obfuscation. Network and Distributed Systems Security (NDSS) Symposium. 2022. pp. 1–18.*
6. *Rawski M. Network Topology Mutation as Moving Target Defense for Corporate Networks // INTL Journal Of Electronics And Telecommunications. 2019. Vol. 65, No. 4, pp. 571–577.*
7. *Hou T. et al. Proto: Proactive topology obfuscation against adversarial network topology inference // IEEE INFOCOM 2020-IEEE Conference on Computer Communications. – IEEE, 2020. pp. 1598–1607.*
8. *Kuchurov V. V., Maksimov R. V., Sherstobitov R. S. Model' i metodika maskirovaniya adresacii korrespondentov v kiberprostranstve // Voprosy kiberbezopasnosti. 2020. № 6(40). pp. 2–13.*
9. *Telen'ga A. P. Maskirovanie metastruktur informacionnyh sistem v kiberprostranstve // Voprosy kiberbezopasnosti. 2024. № 5(57). pp. 50–59.*
10. *Zegzhda D. P. Intellektual'ny'e metody' samoregulyacii raspredelenny'x setevy'x struktur v usloviyax kiberatak // XIV Vserossijskaya mul'tikonferenciya po problemam upravleniya MKPU-2021. 2021. pp. 16–19.*
11. *Lavrova D. S., Zegzhda D. P., Zajceva E. A. Modelirovanie setevoj infrastruktury' slozhny'x ob'ektov dlya resheniya zadachi protivodejstviya kiberatakam // Voprosy` kiberbezopasnosti. 2019. №2(30). pp. 13–20.*
12. *Virsanski E. Geneticheskie algoritmy na Python / per. s angl. A. A. Slinkina. – M.: DMK Press, 2020. – 286 p.: ISBN 978-5-97060-857-9.*