УДК 004.056 Вестник УрФО № 1(35) / 2020, с. 12–23

Гайдамакин Н. А., Синадский Н. И., Сушков П. В.

DOI: 10.14529/secur200102

КОМПЛЕКСНЫЙ ИМИТАЦИОННО-СТАТИСТИЧЕСКИЙ МЕТОД СИНТЕЗА МАССИВОВ УСЛОВНО-РЕАЛЬНЫХ ДАННЫХ НА ОСНОВЕ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СЕРВИСОВ

В статье представлено решение задачи синтеза учебных заданий и массивов данных при организации компьютерного полигона для проведения практических занятий по расследованию инцидентов информационной безопасности. Предложены два основных этапа синтеза фоновой и ситуационной составляющих массивов условно-реальных данных на основе структурно-параметрической модели взаимодействия пользователей информационно-телекоммуникационных сервисов: формирование статических и динамических компонентов.

Статические компоненты синтезируются на основе метода формирования структуры социальных графов, использующего композицию моделей построения сложных сетей с различными структурными параметрами: для сервиса мобильной связи применяется модель Ваттса-Строгатца, для сервиса социальных сетей – модель Барабаши-Альберт. Для сохранения взаимосвязи между пользователями в различных сервисах предложен метод определения наибольшей общей части социальных

графов, основанный на взаимной дифференциации вершин и выделении частичного изоморфизма сравниваемых графов. При формировании атрибутов вершин применяется метод поиска социальных групп (семей), основанный на алгоритме Брона-Кербоша по поиску клики заданного размера в графе.

Для синтеза динамических компонентов массивов данных, описывающих совершение коммуникационных событий, используется математический аппарат цветных сетей Петри. Событие взаимодействия в информационно-телекоммуникационных сервисах представляется в виде метки сети Петри, которая содержит необходимый набор параметров, зависящий от типа сервиса. Для формирования начальной разметки сети Петри предложено использовать структурные, событийные, социальные и временные статистические характеристики реальных информационно-телекоммуникационных сервисов.

Ключевые слова: *массив условно-реальных данных, информационно-телекоммуникационные сервисы, модели сложных сетей, цветная сеть Петри, учебный компьютерный полигон.*

Gaidamakin N. A., Sinadsky N. I., Sushkov P. V.

COMPLEX SIMULATION-STATISTICAL METHOD FOR SYNTHESIZING CONDITIONALLY REAL DATA ARRAYS BASED ON A STRUCTURAL-PARAMETRIC MODEL OF INTERACTION BETWEEN USERS OF INFORMATION AND TELECOMMUNICATION SERVICES

The article presents a solution to the problem of synthesizing training tasks and data arrays when organizing a computer training platform for conducting practical exercises to investigate information security incidents. Two main stages of the synthesis of the background and situational components of the conditionally real data arrays based on the structural-parametric model of interaction between users of information and telecommunication services are proposed: the formation of static and dynamic components.

Static components are synthesized based on the method of forming the structure of social graphs using a composition of models for constructing complex networks with various structural parameters: the Watts-Strogatz model is used for a mobile communication service, and the Barabashi-Albert model is used for a social network service. To preserve the relationship between users in various services, a method is proposed for determining the largest common part of social graphs, based on the mutual differentiation of vertices and the allocation of a partial isomorphism of the compared graphs. When generating vertex attributes, the method of searching for social groups (families) is used, based on the Bron-Kerbosch algorithm for finding a click of a given size in a graph.

For the synthesis of dynamic components of data arrays that describe the performance of communication events, the mathematical apparatus of color Petri nets is used. An interaction event in information and telecommunication services is represented in the form of a Petri net label, which contains the necessary set of parameters, depending on the type of service. It is proposed to use the structural, event, social and temporal statistical characteristics of real information and telecommunication services to form the initial marking of the Petri net.

Keywords: *array of conditionally real data, information and telecommunication services, models of complex networks, color Petri net, training computer training platform.*

Вступление в силу Федерального закона «О безопасности критической информационной инфраструктуры Российской Федерации» свидетельствует об актуальности и значимости решения задачи по обнаружению, предупреждению и ликвидации последствий компьютерных атак и реагирования на компьютерные инциденты. Условием качественного решения данных задач является практико-ориентированная профессиональная подготовка соответствующих специалистов.

Для организации и проведения практических занятий по расследованию инцидентов информационной безопасности в сетях документальной электросвязи и сети Интернет как на потоках магистратуры по направлению «Информационная безопасность», так и на потоках специалитета «Информационная безопасность телекоммуникационных систем» и «Информационно-аналитические системы безопасности» необходимо создание учебного компьютерного полигона, оснащенного современными образцами информационно-аналитических систем безопасности (далее – ИАСБ), такими как IBM I21, МФИ СОФТ «Январь»2, Lampyre3, Gephi4 и др.

ИАСБ – это аппаратно-программные комплексы для проведения поисково-аналитической работы, имеющие возможность накапливать и анализировать данные о взаимодействии пользователей информационно-телекоммуникационных сервисов (далее – ИТ-сервисов). Однако подключение учебных ИАСБ к действующему оборудованию операторов связи, являющемуся источником информации о взаимодействии пользователей ИТ-сервисов, невозможно в соответствии со ст. 64 Федерального закона «О связи». Также отсутствует возможность применения настоящих массивов биллинговой информации в силу того, что такие массивы содержат персональные данные пользователей, а доступ к ним ограничен законодательно.

Обзор литературы позволяет сделать вывод об отсутствии готовых методов и алгоритмов генерации массивов данных, отражающих взаимодействие пользователей ИТ-сервисов.

Для решения данной проблемы при создании учебного компьютерного полигона по расследованию инцидентов информационной безопасности в учебно-научном центре «Информационная безопасность» ИРИТ-РтФ УрФУ им. первого Президента России Б.Н. Ельцина разработано программное обеспечение (далее – ПО). Данное ПО работает в соответствии с созданной структурно-параметрической моделью взаимодействия пользователей в ИТ-сервисах и позволяет синтезировать массивы условно-реальных данных. Файлы, содержащие сгенерированные массивы, в дальнейшем загружаются в базу данных ИАСБ для проведения практических занятий по решению поисково-аналитических задач. На текущем этапе разработки ПО способно генерировать массивы условно-реальных данных о взаимодействии пользователей социальных сетей и сетей мобильной связи.

[1] http://www.ibm.com/software/ products/ru/analysts-notebook

[2] http://www.mfisoft.ru/direction/sorm/sorm-3/

[3] http://www.lampyre.io

[4] http://www.gephi.org

Структура массивов условно-реальных данных

Данные, подлежащие анализу, накапливаются в ИАСБ в формате совокупности таблиц взаимодействия пользователей. В зависимости от вида ИАСБ и типа ИТ-сервиса таблицы, имеющие различные поля хранимой информации, описываются множеством записей вида:

$$D=\{d_1, d_2, \ldots, d_n\},$$

где d_n – элемент таблицы (строка, запись), описывающий одно коммуникационное событие. Для сервиса мобильной связи строка параметров, хранящихся в памяти ИАСБ, может иметь следующий вид [1]:

$$d_{Mobile} \in D_{Mobile} = \{AbonentIMSI, AbonentIMEI, AbonentPhone, LAC, CellID, Time, CallDuration, BillingType, PhoneB\},$$

где *AbonentIMSI* — идентификатор абонента, *AbonentIMEI* — серийный номер устройства абонента, *AbonentPhone* — номер телефона абонента, *BillTime* — время фиксации соединения, *CallDuration* — продолжительность соединения в секундах, *BillingType* — тип соединения, *LAC* — код локальной зоны, *CellId* — идентификатор соты, *PhoneB* — номер телефона принимающей соединение стороны.

Для сервиса социальных сетей – вид:

$$dS_{ocial} \in D_{Social} = \{AbonentPhone, AbonentUID, AbonentIP, AbonentLogin, Time, Type, UidB\},$$

где *AbonentPhone* — номер телефона пользователя, *AbonentUID* — идентификатор учетной записи пользователя, *AbonentIP* — IP-адрес пользователя, *AbonentLogin* — логин Интернет-соединения пользователя, *Time* — время фиксации события, *Type* — тип события, *UidB* — идентификатор учетной записи второго пользователя, участвующего в событии.

Для решения поисково-аналитических задач в структуре массива данных должны быть представлены фоновые и ситуационные компоненты [1]:

$$D = D_{task} \cup D_{feed},$$

где D_{task} – ситуационные задачи и D_{feed} – фоновые события.

Ситуационная задача предназначена для последующего выявления взаимодействия пользователей при решении комплексной учебной поисково-аналитической задачи, описывается преподавателем через интерфейс разработанного ПО. Фоновый массив данных имитирует социальную активность в целом и содержит события пользователей, подчиняющиеся некоторым статистическим законам реальных ИТ-сервисов.

Синтез фонового массива данных основан на комплексном имитационно-статистическом методе, включающем последовательное формирование статических и динамических компонентов с использованием различных математических подходов:

1. На первом этапе происходит формирование статических компонентов записей d_{Mobile} и d_{Social}, к которым относятся персональные идентификаторы пользователей (*AbonentIMSI, AbonentIMEI, AbonentPhone, AbonentUID, AbonentIP, AbonentLogin*) и структура их взаимосвязей внутри ИТ-сервиса (в том числе, информация о социальных группах). Для создания структурной основы, которая описывает существование социальных связей между пользователями, используется композиция существующих моделей построения сложных сетей на основе статистических распределений структурных параметров сервисов мобильной связи и социальных сетей.

2. На втором этапе на основе существующих социальных связей пользователей происходит генерация динамических компонентов коммуникационного события (тип, участники, место, время начала и продолжительность). Для формирования различных полей записей d_{Mobile} и d_{Social} требуются статистические распределения определенных параметров, которые можно разбить на категории:

– событийные (поля *BillingType, Type*);

– социальные (поля *PhoneB, UIDB*);

– пространственные (поля *LAC, CellID*);

– временные (поля *Time, CallDuration*).

Задача синтеза динамической составляющей фонового массива данных, предполагающая наличие большого количества активных объектов с отчетливо выраженным индивидуальным поведением, относится к категории синтеза сложных систем. Одним из подходов к синтезу сложных систем является агентное моделирование, позволяющее учесть структуру и взаимодействие пользователей ИТ-сервисов. В качестве общепринятого математического аппарата решения задач агентного имитационного моделирования применяются цветные сети Петри (далее – ЦСП).

Формирование статических компонентов массивов условно-реальных данных

ИТ-сервисы целесообразно рассматривать в виде социальных графов. В общем случае структура социального графа представляется в виде $G=(U, E)$, где U — множество вершин графа. Обозначим через G_M и G_S социальные графы ИТ-сервисов мобильной связи и социальных сетей соответственно. Ситуационная задача t описывается шаблоном взаимодействия пользователей $G_t = (U_t, E_t)$.

С целью создания основы для описания взаимодействия пользователей ИТ-сервисов предлагается использовать метод формирования статической структуры социальных графов G_M и G_S на основе композиции существующих моделей построения сложных сетей с учетом заданного шаблона взаимодействия пользователей G_t.

В рамках данного исследования наибольший интерес представляют три модели построения сложных сетей, позволяющие описывать взаимодействие между людьми:

– модель Эрдёша-Реньи (случайные графы) [2];

– модель Ваттса-Строгатца (сети тесного мира) [3];

– модель Барабаши-Альберт (сети предпочтительного присоединения) [4].

В таблице 1 представлены основные структурные свойства, которые являются определяющими при выборе модели для формирования структуры ИТ-сервиса.

сетей определены некоторые структурные свойства [4 ,5]:

– закон распределения степеней вершин близкий к Пуассоновскому;

– малая длина пути между вершинами;

– высокий коэффициент кластеризации;

– децентрализованная структура.

Построение структуры мобильных сетей начинается с некоторого фиксированного числа вершин, которые затем случайным образом связываются или меняют связи. Однако социальные сети носят открытый характер, что подразумевает рост благодаря непрерывному добавлению новых узлов. Начиная с небольшого ядра количество узлов увеличивается на протяжении всего времени жизни сети путем последующего добавления новых. Кроме того, большинство социальных сетей демонстрируют предпочтительное соединение, такое, что вероятность подключения к узлу зависит от его степени.

Указанные свойства определяют особенность социальных сетей по сравнению с сетями взаимодействия в реальном мире: закон распределения степеней вершин – показательный. Кроме того, показательному закону распределения соответствует сильноцентрализованная структура. Остальные свойства (малая длина пути между вершинами, высокий коэффициент кластеризации) остаются неизменными.

Таблица 1

Структурные свойства моделей

	Модель Эрдёша-Реньи	**Модель Ваттса-Строгатца**	**Модель Барабаши-Альберт**
Закон распределения степеней вершин	Пуассоновский	близкий к Пуассоновскому	показательный
Расстояние между вершинами	малое	малое	малое
Коэффициент кластеризации	низкий	высокий	средний
Структура	децентрализованная	децентрализованная	централизованная

Процесс формирования статических структур социальных графов целесообразно начинать с построения структуры сервиса мобильной связи, т.к. количество абонентов будет превышать количество пользователей социальных сетей.

Структура взаимосвязей абонентов ИТ-сервиса сотовой связи будет рассматриваться как сеть простого вербального общения людей в реальном мире без применения каких-либо технических средств, т.к. для данных

Анализ свойств моделей и реальных сетей позволяет сделать вывод, что для описания взаимодействия абонентов мобильной связи наилучшим образом подходит модель Ваттса-Строгатца. С другой стороны, активность между пользователями социальных сетей имеет отличительные особенности, наиболее полно отраженные в модели Барабаши-Альберт.

Процесс формирования структуры социального графа G_M начинается с генерации регу-

лярной решетки со степенью вершин *K*. Затем происходит выбор случайным образом соседних вершин в количестве $|U_t|$, с распределением значений вершин и ребер в соответствии с U_t и E_t. На последнем этапе выполняется перераспределение каждого ребра с вероятностью *p* на случайную вершину. Назначенные в соответствии с E_t ребра остаются неизменными.

Процесс формирования структуры социального графа G_S начинается с генерации случайного графа с количеством вершин m_0. Данный граф выступает в роли начального ядра будущего социального графа. Случайный граф строится в соответствии с моделью Эрдёша-Реньи. Исходными данными на этом этапе являются количество вершин m_0 и вероятность p_0, с которой между двумя произвольными вершинами образуется ребро.

После создания структурного ядра графа происходит последовательное добавление вершин в количестве m_{max}, определенном изначально. За один шаг создается одна вершина. Количество ребер, с которым данная вершина добавляется, зависит от коэффициента *C*, значение которого определено заранее и остается постоянным. Данный коэффициент призван снизить разрежённость формируемого графа, что как следствие позволит увеличить коэффициент кластеризации.

В процессе формирования атрибутивных параметров вершин социальных графов возникает необходимость нахождения «семей», т.е. групп пользователей, объединенных в полносвязный граф. Такая задача существует и в теории графов и носит название «Задача о клике». Кликой в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа. Иными словами, это полный подграф первоначального графа. Задача о клике существует в двух вариантах: в задаче распознавания требуется определить, существует ли в заданном графе *G* клика размера *k*, в то время как в вычислительном варианте требуется найти в заданном графе *G* клику максимального размера.

Алгоритм Брона — Кербоша — метод ветвей и границ для поиска всех клик (а также максимальных по включению независимых множеств вершин) неориентированного графа [6]. Алгоритм использует тот факт, что всякая клика в графе является его максимальным по включению полным подграфом. Начиная с одиночной вершины (образующей полный подграф), алгоритм на каждом шаге пытается увеличить уже построенный полный подграф, добавляя в него вершины из множества кандидатов. Высокая скорость обеспечивается отсечением при переборе вариантов, не приводящих к построению клики, для чего используется дополнительное множество, в которое помещаются вершины, бывшие уже использованными для увеличения полного подграфа.

Алгоритм оперирует тремя множествами вершин графа:

1. Множество **compsub** — множество, содержащее на каждом шаге рекурсии полный подграф для данного шага. Строится рекурсивно.

2. Множество **candidates** — множество вершин, которые могут увеличить compsub.

3. Множество **not** — множество вершин, которые уже использовались для расширения compsub на предыдущих шагах алгоритма.

Алгоритм является рекурсивной процедурой, применяемой к этим трем множествам. Сам алгоритм можно описать следующей последовательностью действий:

ПРОЦЕДУРА *extend* (*candidates, not*):

ПОКА *candidates* не пусто **И** *not* не содержит вершины, соединенной со всеми вершинами из *candidates*,

ВЫПОЛНЯТЬ:

1 Выбираем вершину *v* из *candidates* и добавляем её в *compsub*

2 Формируем *new_candidates* и *new_not*, удаляя из *candidates* и *not* вершины, не соединенные с *v*

3 **ЕСЛИ** *new_candidates* и *new_not* пусты

4 **ТО** *compsub* – клика

5 **ИНАЧЕ** рекурсивно вызываем *extend* (*new_candidates, new_not*)

6 Удаляем *v* из *compsub* и *candidates*, и помещаем в *not*

Найденные таким образом клики заданного размера могут быть использованы для описания социальных групп («семей») без внесения дополнительных изменений в структуру графов, что позволит сохранить их структурные особенности.

Использование случайного закона при определении соответствия между вершинами различных социальных графов лишает реалистичности синтезируемые массивы условно-реальных данных. Для максимального сохранения взаимосвязей между объектами в различных сервисах предложено использовать метод анализа структур социальных гра-

фов на основе дифференциации вершин и определения частичного изоморфизма [7].

Задача определения сходства структур рассматривается как выделение в сравниваемых графах G_M и G_S наибольшей общей части — графа $G_{max} = (U_{max}, E_{max})$. Для решения данной задачи в работах [8,9] предлагается перебрать все возможные подстановки вершин исследуемых графов, и в каждой из них определить число совместившихся ребер, образующих общую часть. Подстановка, образующая граф G_{max} с наибольшим числом рёбер $|E_{max}|$, именуется подстановкой сходства, а сформированная на основе нее общая часть является наибольшей.

С целью уменьшения вычислительной сложности разрабатываемого алгоритма предлагается исключить перебор всех возможных подстановок вершин при формировании наибольшей общей части путем введения начальной подстановки, которая определяет всю дальнейшую подстановку сходства за счет применения метода дифференциации вершин.

Под начальной подстановкой будем понимать пару вершин, для которых принимается условие взаимно однозначного соответствия (биективного отображения), т.е. для пары вершин $u_1(G_M) \in U_M$ и условно соответствующей ей $u_1(G_S) \in U_S$ верно $u_1(G_M) \in u_1(G_S)$.

В качестве начальной подстановки $(u_x(G_M), u_y(G_S))$ предлагается использовать вершины с максимальными степенными параметрами в обоих графах G_M и G_S:

$$u_1(G_M) = u_x^k(G_M) \in U_M$$
$$u_1(G_S) = u_y^n(G_S) \in U_S,$$

где k и n обозначают максимальные степени вершин $u_x(G_M)$ и $u_y(G_S)$ графов G_M и G_S соответственно.

После определения начальной подстановки выполняется процедура взаимозависимой дифференциации вершин в обоих графах G_M и G_S. В результате данного этапа происходит присвоение одинаковых кодов различия двум вершинам из различных графов, которые максимально соответствуют друг другу структурно. Таким образом, часть взаимосвязей пользователей из социального графа U_M будет доступна и в графе U_S. После определения соответствия вершин U_S вершинам U_M при необходимости происходит добавление ребер для сохранения ситуационной задачи Gt.

Формирование динамических компонентов массивов условно-реальных данных

ЦСП [10] представляет собой кортеж $CPN = <P, T, TM, I, O, M>$, где P — конечное множество позиций, T — конечное множество переходов, TM — множество временных моментов для срабатывания переходов, I — конечное множество входящих в переходы дуг, O — конечное множество выходящих из переходов дуг, M — множество меток в начальный момент времени.

Представленные статические и динамические компоненты массивов данных соответствуют объектам сетей Петри: динамические — изображаются метками (фишками, маркерами) внутри позиций и статические — им соответствуют вершины и дуги сети Петри.

Для создания вершин $(p_k \in P, t_k \in T)$ и дуг $(i_k \in I, o_k \in O)$ сети Петри используются принципы моделирования сложных сетей. Выбор моделей основывается на соответствии структурных параметров создаваемых графов статистическим распределениям, полученным в результате анализа реальных ИТ-сервисов.

Для отражения динамических свойств в сеть Петри введено понятие разметки сети, которая реализуется с помощью меток $m_k \in M$, размещаемых в позициях. Метка представляет собой объект, содержащий несколько параметров, каждый из которых может принимать дискретный набор значений. В соответствии с этим метки различаются по типам параметров (переменных). Чтобы отличать метки различных типов, их можно окрашивать в различные цвета. В задаче моделирования взаимодействия между пользователями ИТ-сервисов метки представляют собой коммуникационные события. Одним из ключевых параметров метки выступает тип коммуникационного события. Для ИТ-сервиса мобильной связи целесообразно использовать следующие типы событий [1]:

- *B* — отправка сообщений о подключениях к базовым станциям для передачи координат;
- *G* — получение *GPRS*-трафика;
- *C* — совершение звонка (разделяется на исходящие и входящие звонки);
- *S* — отправка СМС-сообщения (разделяется на исходящие и входящие СМС-сообщения).

При определении типов коммуникационных событий для ИТ-сервиса социальных сетей учитывается возможность пользователя вести публичную (тип – *public*) и приватную (тип – *private*) переписку. В зависимости от со-

держимого сообщения выделяются следующие типы событий:

- *text* — текстовое сообщение;
- *picture* — передача изображения;
- *video* — передача видео;
- *link* — ссылка на источник.

Для различных типов коммуникационных событий могут быть использованы необходимые дополнительные параметры, которые также будут содержаться в метке (например, продолжительность телефонного звонка). Таким образом, тип коммуникационного события является составным. Все параметры метки определяют ее составной цвет. Представление же самой метки $m_k \in M$, окрашенной цветом c*olour={private, text}*, имеет следующий вид: $m_k^{colour:\{private,\ text\}}$.

Сеть Петри представляет собой асинхронную систему, в которой метки перемещаются по позициям через переходы. Переход может сработать (т.е. переместить метку из входной позиции в выходную для данного перехода), если во всех входных позициях для данного перехода присутствует хотя бы одна метка и выполнено логическое выражение, ограничивающее переход (*спусковая функция*).

Дуги могут иметь пометки в виде выражений (переменных, констант или функций), определенных для множества цветов, и использоваться либо для «вычленения» компонентов сложного цвета меток при определении условия срабатывания перехода, либо для изменения цвета метки следующей позиции после срабатывания перехода. Данное свойство позволяет упростить процесс разделения коммуникационных событий на «семейные» (т.е. между членами социальной группы) и «все остальные» путем добавления нового параметра метки в зависимости от социального статуса получателя сообщения (тип – *family/others*) и создания различных условий срабатывания перехода между «членами семьи» и другими пользователями. Обозначение дуги $i_k \in I$, обеспечивающей срабатывание перехода меткой $m_k^{colour:\{family\}}$ с параметром *colour:{family}*, имеет следующий вид: $i_k^{colour:\{family\}}$.

Чаще всего таблицы взаимодействия пользователей ИТ-сервисов наполняются записями в порядке их совершения, т.е. упорядочены по времени начала события. Время представляется в формате UTC, минимальным шагом изменения состояния является 1 секунда. Для анализа систем реального времени введен временной механизм, реализованный с помощью глобальных часов и так называемых *штампов*, которые несут метки. Временной штамп метки назначается при ее инициализации в начальной разметке и наращивается выражениями на переходах или дугах. В результате метка становится доступной для перехода, если ее штамп оказался меньше значения счетчика глобальных часов. Обозначение временного штампа метки $m_k \in M$, активирующей переход во временной момент *time:{z}*, $z \in TM$, принимает следующий вид: $m_k^{time:\{z\}}$. Часы наращивают свое значение, если на данный момент времени ни один переход сети не разрешен.

Конкретизация метки, находящейся в данной позиции, определяется инициализирующим выражением начальной разметки. Для выполнения начальной разметки требуется определить некоторые значения:

- общее количество меток $|M|=K$;
- набор параметров, определяющих цвет каждой метки $m_k^{colour:\{c1,\ c2,\ \ldots,\ cP\}}$, где *{c1, c2, …, cP}* — типовые и дополнительные параметры коммуникационного события *k, k=1, …, K*;
- временные штампы меток $m_k^{time:\{z\}}$, где $z \in TM$.

Для формирования начальной разметки ЦСП, описывающей взаимодействие пользователей ИТ-сервиса мобильных сетей, целесообразно использовать статистические характеристики биллинговой информации [1]:

- $K_0 = \langle F_{time}, F_{dur}, F_n, F_a \rangle$ — не связанные с адресацией соединения, описываемые функциями распределения: F_{time} — времени суток, F_{dur} — длительности события, F_n — количества соединений, F_a — вероятности генерации типа события;
- $F_1 \langle F_l, F_t \rangle$ — связанные с адресацией соединения: F_l — функция распределения выбора получателей соединения, а F_t — функция распределения промежутков времени между началами инициализации двух последовательных соединений.

В результате инициализации начальной разметки сети происходит распределение меток с заданными параметрами по определенным позициям в соответствии со статистическими распределениями различных характеристик рассматриваемых ИТ-сервисов. На основе выбранного временного интервала для описания взаимодействия пользователей ИТ-сервисов запускается механизм глобальных часов. После чего происходит последовательное перемещение меток между

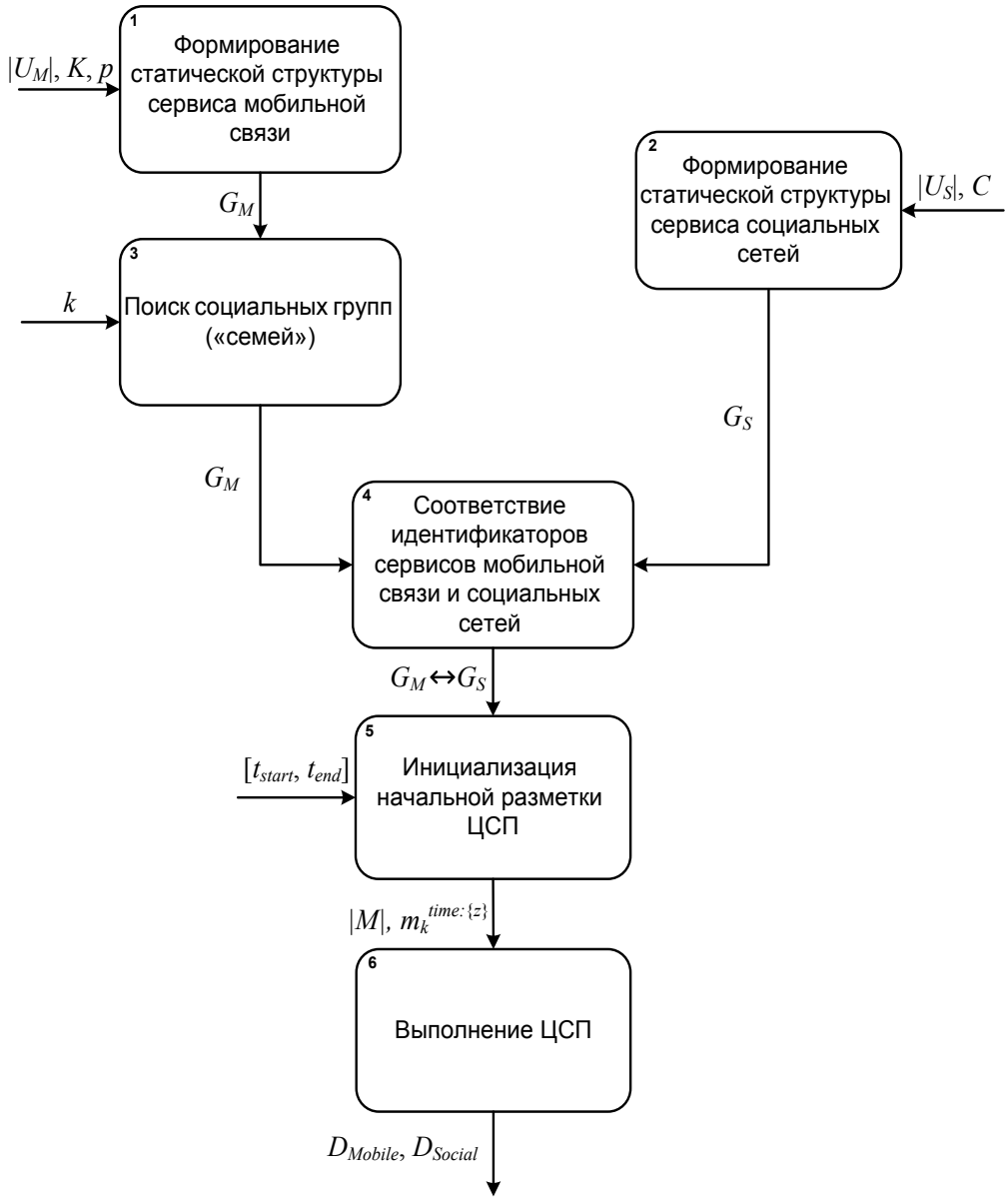

Рис. 1. Структурная схема комплексного метода синтеза массивов данных

позициями, что позволяет создавать строки d_{Mobile} и d_{Social} в массивах D_{Mobile} и D_{Social}.

Таким образом, комплексный метод синтеза массивов условно-реальных данных представляет собой последовательность математических подходов, имитирующих различные параметры процесса взаимодействия пользователей на основе статистических распределений требуемых показателей в реальных ИТ-сервисах. Структурная схема метода синтеза представлена на рис. 1.

1. Для формирования статической структуры социального графа сервиса мобильной связи G_M в соответствии с моделью Ваттца-Строгатца определяются три основных параметра:

– численность пользователей имитируемого сервиса $|U_M|$,

– среднее количество социальных связей пользователей, описываемое через степень вершин K создаваемой регулярной решетки,

– вероятность перераспределения каждого ребра на случайную вершину p.

Модификация значений K и p позволяет менять такие структурные свойства создаваемого графа как минимальная длина пути между вершинами и коэффициент кластеризации. Для имитации сети вербального обще-

ния людей в соответствии со статистически определенными структурными параметрами в работах [4, 5] оптимальными значениями выбраны K=6 и p=0,1.

2. При создании статической структуры сервиса социальных сетей G_S в соответствии с моделью Барабаши-Альберт необходимо определить следующие параметры:

– численность пользователей имитируемого сервиса $|U_S|$,

– количество ребер C, с которыми новая вершина добавляется в структуру графа.

Параметр C определяет важное свойство создаваемого графа – степень кластеризации вершин. Для моделирования структуры социальных сетей в соответствии со статистическими параметрами в [5] достаточным значением является C=3.

3. Поскольку вовлеченность людей в сервис мобильной связи значительно выше, чем в социальные сети, именно в графе G_M производится поиск клик заданного размера в соответствии с алгоритмом Брона-Кербоша. Единственным параметром, задаваемым для алгоритма, является размер клики k. Для соответствия распределению количества одиноких людей и реальных семей с различным составом детей (0-3), последовательно производится поиск клик размерностью k=3, k=4, k=5 в требуемом соотношении.

4. Для имитации одновременного общения пользователей посредством различных ИТ-сервисов применяются методы анализа социальных графов G_M и G_S, направленные на установление соответствия персональных идентификаторов сервисов мобильной связи и социальных сетей *(AbonentIMSI, AbonentIMEI, AbonentPhone)* ↔ (AbonentUID, AbonentIP, AbonentLogin)

5. Для инициализации начальной разметки ЦСП необходимо определить временной интервал $[t_{start}, t_{end}]$, в течение которого происходит моделирование взаимодействия между пользователями. На основании данных параметров произойдет автоматическое формирование общего количества меток $|M|$ и временных штампов меток $m_k^{time:\{z\}}$, а также распределение их по позициям $p_k \in P$ в соответствии со статистическими распределениями активности пользователей для указанного временного промежутка.

6. При выполнении циклов ЦСП имитируется информационный обмен между заданным количеством пользователей $|U_M|$ и $|U_S|$ в определенный временной период $[t_{start}, t_{end}]$. Результатом выполнения всех циклов ЦСП являются синтезированные массивы D_{Mobile} и D_{Social} требуемого формата.

Анализ синтезированных массивов условно-реальных данных с применением ИАСБ на базе ПО Lampyre

Синтезированные массивы DMobile и DSocial имеют необходимый формат для интеграции в учебные ИАСБ компьютерного полигона. Обычно это файлы форматов *.xml и *.csv.

В составе учебного полигона включены различные аппаратно-программные комплексы для проведения поисково-аналитической работы, одним из которых является ПО Lampyre. Данная система позволяет обрабатывать большие массивы данных, строить графики взаимосвязей и накладывать их на карту и временной масштаб.

На компьютерном полигоне для анализа сетевого взаимодействия используются разрабатываемые студентами в процессе обучения аналитические методики, которые представляют собой программный код на языке Python, интегрируемый в ПО Lampyre. Данные методики и различные варианты визуализации графа связей, используемые в ПО Lampyre, применяются к синтезированному массиву данных. Они позволяют решить сквозную аналитическую задачу по поиску в массиве данных источника атакующего воздействия, ставшего причиной инцидента информационной безопасности, на основе анализа сетевого взаимодействия. На рис. 2 представлен интерфейс ПО Lampyre, визуализирующий граф связей условного пользователя при решении учебной аналитической задачи.

Заключение

В статье рассмотрен комплексный имитационно-статистический метод синтеза массивов условно-реальных данных позволяющий формировать статические и динамические компонентов. Предложенный композиционный метод, объединяющий несколько моделей построения сложных сетей для формирования статических структур моделируемых ИТ-сервисов, используются в качестве основы для создания коммуникационных событий. В результате применения алгоритмов ЦСП в процессе синтеза массивов данных имеется возможность создания записей о взаимодействии пользователей, наполнение которых зависит от структурных, событий-

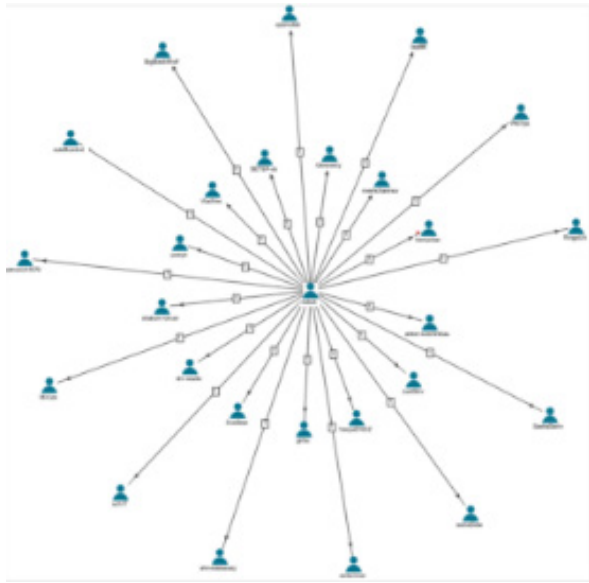

Рис. 2. Интерфейс ПО Lampyre

ных, временных и социальных параметров моделируемых ИТ-сервисов. Сгенерированный ситуационный массив условно-реальных данных применяется при отработке практических заданий по разработке поисково-аналитических методик на учебном компьютерном полигоне по расследованию инцидентов информационной безопасности.

Литература

1. Семенищев И.А., Синадский А.Н., Синадский Н.И., Сушков П.В. Синтез массивов биллинговой информации на основе статистико-событийной модели взаимодействия абонентов сетей сотовой связи. // Вестник УрФО. Безопасность в информационной сфере. — 2018. — № 1 (27). — С. 47–56.

2. Erdos, P., and A. Renyi, On Random Graphs. Publicationes Mathematicae (Debrecen), volume 6, 1959, pp. 290–297.

3. Watts, D. J., Small Worlds: The Dynamics of Networks between Order and Randomness (Princeton University, Princeton, NJ), 1999.

4. R. Albert, A-L. Barabasi. Statistical mechanics of complex networks. Reviews of modern physics, volume 74, January 2002.

5. Проект Edyo.ru. URL: http://edyo.ru (дата обращения 25.03.2020).

6. C. Bron and J. Kerbosch, Algorithm 457: Finding All Cliques of an Undirected Graph, Proceedings of the ACM, 1973, p. 575–577.

7. Синадский Н.И., Сушков П.В. Модификация методов анализа социальных графов на основе применения атрибутивных компонентов учетных записей для идентификации сообществ пользователей социальных сетей. — Вестник УрФО. Безопасность в информационной сфере. — 2017. — № 2 (24). — С. 32–40.

8. Погребной Ан.В., Погребной В.К. Метод дифференциации вершин графа и решение проблемы изоморфизма // Известия Томского политехнического университета. — 2015. — Т. 326. — № 6. — С. 34–45.

9. Погребной А.В. Метод определения сходства структур графов на основе выделения частичного изоморфизма в задачах геоинформатики // Известия Томского политехнического университета, 2015. — Т. 326. — № 11. — С. 56–66.

10. Питерсон Дж. Теория сетей Петри и моделирование систем: пер. с англ. — М.: Мир, 1984. — 264 с.

References

1. Semenishchev I.A., Sinadskiy A.N., Sinadskiy N.I., Sushkov P.V. Sintez massivov billingovoy informatsii na osnove statistiko-sobytiynoy modeli vzaimodeystviya abonentov setey sotovoy svyazi. — Vestnik UrFO. Bezopasnost' v informatsionnoy sfere. — 2018. № 1 (27). — S. 47-56.

5. Project Edyo.ru. URL: http://edyo.ru (data obrashcheniya 25.03.2020).

7. Sinadskiy N.I., Sushkov P.V. Modifikatsiya metodov analiza sotsial'nykh grafov na osnove primeneniya atributivnykh komponentov uchetnykh zapisey dlya identifikatsii soobshchestv pol'zovateley sotsial'nykh setey. — Vestnik UrFO. Bezopasnost' v informatsionnoy sfere. — 2017. № 2 (24). — S. 32-40.

8. Pogrebnoy An.V., Pogrebnoy V.K. Metod differentsiatsii vershin grafa i reshenie problemy izomorfizma // Izvestiya Tomskogo politekhniche-skogo universiteta. — 2015. — T. 326. — № 6. — S. 34–45.

9. Pogrebnoy A.V. Metod opredeleniya skhodstva struktur grafov na osnove vydeleniya chastichnogo izomorfizma v zadachakh geoinformatiki // Izvestiya Tomskogo politekhnicheskogo universiteta, 2015. — T. 326. — № 11. — S. 56-66.

10. Piterson Dzh. Teoriya setey Petri i modelirovanie sistem: Per. s angl. — M.: Mir, 1984. — 264 s.

ГАЙДАМАКИН Николай Александрович, доктор технических наук, профессор, профессор учебно-научного центра «Информационная безопасность», Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002; г. Екатеринбург, ул. Мира, 19. E-mail: n.a.gaidamakin@urfu.ru

СИНАДСКИЙ Николай Игоревич, кандидат технических наук, доцент, доцент учебно-научного центра «Информационная безопасность», Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: n.i.sinadsky@urfu.ru

СУШКОВ Павел Владимирович, аспирант ИЕНМ, Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: pavelsu@e1.ru

GAIDAMAKIN Nikolay, doctor of engineering, Professor, Educational and Scientific Center "Information Security", Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: n.a.gaidamakin@urfu.ru

SINADSKY Nikolay, candidate of technical sciences, associate Professor, Educational and Scientific Center "Information Security", Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: n.i.sinadsky@urfu.ru

SUSHKOV Pavel, post-graduate student of Institute of Natural Sciences and Mathematics, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: pavelsu@e1.ru