

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

*В статье рассмотрен программный комплекс, реализующий алгоритм синтеза массивов биллинговой информации на основе разработанной пространственно-временной статистико-событийной модели, приводятся статистические характеристики синтезируемых массивов и результаты их обработки; комплекс предназначен для тестирования информационно-аналитических систем безопасности и обеспечения практических занятий по дисциплинам направления «Информационная безопасность».*

**Ключевые слова:** синтез массивов биллинговой информации, информационно-аналитические системы безопасности.

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

---

# SYNTHESIS BILLING INFORMATION ARRAYS BASED ON THE STATISTICAL EVENT MODEL OF INTERACTION OF CELLULAR NETWORKS SUBSCRIBERS

*The article introduces the software complex that implements the algorithm for synthesizing an array of billing information based on a spatio-temporal statistical-event model, statistical characteristics of synthesized arrays and the results of their processing. The complex is designed for testing information and analytical security systems and providing practical classes in the disciplines of the specialty Information Security.*

**Keywords:** Synthesis of arrays of billing information, information-analytical security systems.

Вступление в силу Федерального закона 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» свидетельствует об актуальности и значимости разработки программно-аппаратных систем, предназначенных для решения задач по обнаружению, предупреждению и ликвидации последствий компьютерных атак на критическую информационную инфраструктуру (КИИ) Российской Федерации. Задачи обнаружения компьютерных атак и расследования компьютерных инцидентов решаются, в том числе информационно-аналитическими системами безопасности (ИАСБ), одним из примеров которых является система DATAPK [1], представляющая собой комплекс оперативного мониторинга и контроля защищенности объектов КИИ.

В рамках обеспечения информационной безопасности на каналах сотовой связи применяются и активно разрабатываются поисково-аналитические системы, такие как IBM I2<sup>1</sup>, МФИ СОФТ «Январь»<sup>2</sup> и др., использующие сложные алгоритмы поиска взаимосвязей элементов современных телекоммуникационных систем. Эти программы применяются службами информационной безопасности операторов сетей сотовой связи с целью выявления, в том числе, фактов мошенничества при оплате услуг.

Для тестирования корректности реализации аналитических алгоритмов в системах различного типа требуются массивы биллинговой информации. Аналогичные массивы должны использоваться для обеспечения образовательного процесса в рамках изучения ИАСБ как на потоках магистратуры по направлению «Информационная безопасность», так и на потоках специалитета «Информационная безопасность телекоммуникационных систем» и «Информационно-аналитические системы безопасности».

При этом использование настоящих массивов биллинговой информации о взаимодействии абонентов сетей сотовой связи невозможно, в том числе в силу ограничений, накладываемых ФЗ «О связи». Кроме того, для анализа алгоритмов ИАСБ необходим эталонный трафик, статистические распределения, характер связей элементов и конкретные связи между заданными элементами которого должны быть точно известны.

<sup>1</sup> <http://www-03.ibm.com/software/products/ru/analysts-notebook>

<sup>2</sup> <http://www.mfisoft.ru/direction/sorm/sorm-3/>

Целью работы является разработка программного обеспечения для синтеза массивов биллинговой информации, состоящих из фонового биллинга, являющегося информационным наполнением, на которое накладывается массив ситуационных задач (тестов), с целью формирования у студентов практических навыков по поиску и анализу взаимосвязей элементов современных телекоммуникационных систем.

Создан программный комплекс, позволяющий осуществлять синтез массивов биллинговой информации и массивов данных, имитирующих взаимодействие пользователей в социальных сетях. В основе комплекса лежит разработанная авторами пространственно-временная статистико-событийная модель взаимодействия пользователей в сетях операторов связи [2–5].

Программная реализация предложенной модели позволяет синтезировать массив биллинговой информации по заранее заданным статистическим распределениям и создавать требуемые связи между элементами сети.

Вариативность статистических распределений, описывающих характеристики фонового массива биллинга, позволяет создавать тестовые массивы как общего вида, так и для конкретной ситуационной задачи, используемой для постановки поисково-аналитических задач при изучении ИАСБ.

### Структура массива биллинговой информации

Массив биллинговой информации представляет собой упорядоченную по временной метке совокупность записей. Каждая запись (строка) является структурой данных, в которую входят 12 полей, которые могут быть как заполненными, так и пустыми. Основными значимыми для генерации полями являются: AbonentIMSI (номер SIM-карты абонента), AbonentIMEI (тип и серийный номер устройства абонента), AbonentPhone (MSISDN – номер телефона), LAC и CellID (поля, указывающие на привязку к базовой станции), BillTime (время события), CallDuration (продолжительность соединения) и BillingType (тип события) (рис. 1).

BillTime	CallDuration	BillingType	LAC	CellID	PhoneB	AbonentIMEI	AbonentIMSI	AbonentPhone
06.06.2014 0:00	4	Normal LocUpd	3901	39762		357331049414210	250023622333809	79222244716
06.06.2014 0:00	3	GPRES	3901	39762		351180406097420	250280422334439	79052485247
06.06.2014 0:00	4	GPRES	3907	39781		357881046123760	250027122338724	7912249642
06.06.2014 0:00	1	SMS In	3907	3024	79122241006	12536006306700	250011422337848	79822248766
06.06.2014 0:00	4	SMS In	3907	3024	79122253586	357008045261270	25001152326032	7912236947
06.06.2014 0:00	4	GPRES	3901	39762		35929056323320	250284722325348	79052236267
06.06.2014 0:00	4	Normal LocUpd	3907	3024		357881045500410	250027022325792	79122336705
06.06.2014 0:00	2	Normal LocUpd	3907	39781		357331040962830	250026262326087	7922237908
06.06.2014 0:00	0	SMS Out	3901	39762	79922256258	357881043528250	250027122345377	79222256280

Рис. 1. Строки массива биллинговой информации

Под термином «абонент» понимается элемент сети (узел графа) сотовой связи, инициирующий или принимающий соединения. Соединение — это направленная ветвь от одного узла графа к другому. Инициатором соединения называется исходный для этой ветви узел, принимающим абонентом — конечный.

Массив биллинга формируется для коммутатора, принадлежащего одному из операторов сетей сотовой связи, и представляет собой совокупность записей, поступающих с различных базовых станций о совершаемых абонентскими устройствами действиях.

### Пространственно-временная статистико-событийная модель биллинговой информации

Разработанный программный комплекс позволяет на основе имеющихся статистических распределений генерировать массивы биллинговой информации.

С целью синтеза массивов биллинговой информации предложена модель жителей населенного пункта в терминах телекоммуникаций [2]. В качестве основы модели выбрана популяция абонентов операторов сетей сотовой связи, проживающих на вымышленной территории, перемещающихся по заданным правилам в течение генерируемого периода времени и осуществляющих взаимодействие путем совершения звонков и отправки СМС (SMS)-сообщений в соответствии с задаваемыми статистическими распределениями.

Пространственно-временная статистико-событийная модель  $M$  синтеза биллинговой информации, основанная на поведении абонентов с точки зрения сети операторов связи, является совокупностью модели перемещений абонентов в течение заданного промежутка времени  $MSH$ , модели соединений  $MS$  и модели  $MTG$  отправки сообщений о подключениях к базовым станциям для передачи координат и получения GPRS-трафика.

На данном этапе разработки предполагается, что один типовой абонент имеет один телефонный аппарат (IMEI) и одну идентифицирующую его SIM-карту (IMSI). Генерируются только значимые для последующего анализа события: служебные записи, не влияющие на анализ, при генерации игнорируются.

Модель соединений  $MS$  учитывает статистические распределения биллинговой информации и социальные характеристики абонентов операторов сотовой связи (круг обще-

няя: наличие родственных, личных связей). Социальные характеристики представлены шаблонами поведения для различных социальных групп, включая характеристики взаимодействия в сетях сотовой связи и характеристики перемещения людей в рамках населенного пункта.

В состав модели соединений  $MS$  входят множества абонентов оператора сотовой связи  $H: H = \{h_i\}_{i=1}^{n_h}$  и «семей»  $W: W = \{w_i | w_i \subseteq H\}_{i=1}^{n_w}$ ,  $n_h$  — количество абонентов,  $n_w$  — «семей».

Абонентские терминалы представлены в модели множествами номеров IMSI, IMEI и MSISDN с количеством элементов, равным  $n_h$ :  $N_{imsi} = \{n_{imsi_i}\}_{i=1}^{n_h}$ ,  $N_{imei} = \{n_{imei_i}\}_{i=1}^{n_h}$ ,  $N_{msisdn} = \{n_{msisdn_i}\}_{i=1}^{n_h}$ . Для формирования элементов множеств используется алгоритм генерации неповторяющихся чисел требуемого формата.

Совокупность абонентов и соответствующих им абонентских терминалов формирует абонентскую базу (базу принадлежности).

Модель перемещений  $MSH$  предназначена для генерации полей, описывающих географические координаты совершения события, имитирует перемещения абонентов в заданном временном интервале в рамках населенного пункта, который представляется квадратом, состоящим из массива клеток (рис. 2). Для каждой клетки задаются модифицируемые списками параметры LAC (Location Area Code — код локальной зоны) и CellID (уникальный номер, предназначенный для идентификации базовых станций).

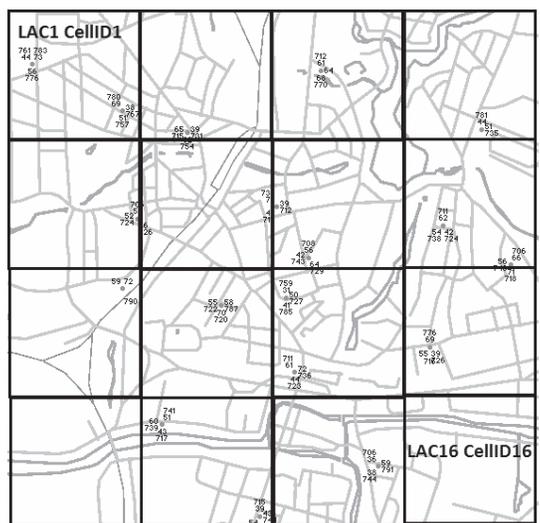


Рис. 2. Схема разделения моделируемой территории на области действия базовых станций

Перемещения абонентов могут моделироваться в двух совмещаемых режимах: на основе частотных распределений по популярности конкретных зон или на основе шаблонов перемещений. В первом случае перемещения абонентов имеют большую случайность, возможность быстрого редактирования характера перемещений абонентов путем изменения приоритетов конкретных зон (ячеек карты), во втором перемещения предсказуемы и направляемы (пользователь может менять маршруты абонентов и указывать конкретные точки маршрута по своему усмотрению), что требуется при создании ситуационной задачи.

Передвижение абонентов в рамках населенного пункта описывается шаблонами перемещений  $SH$ . Количество шаблонов произвольное, задается значением  $n_{sh}$ ,  $SH = \{sh\}_{i=1}^{n_{sh}}$ . Шаблон перемещений представляет собой упорядоченный по времени список LAC и CellID, по которому отслеживается передвижение абонента в течение заданного времени.

Населенный пункт характеризуется массивом ячеек его карты: двумерным вектором  $map$ ,  $n_{map}$  – его длина (количество ячеек карты), в каждом  $map \in (\overline{1, n_{map}})$  из элементов которого содержатся два параметра (LAC и CellID) базовой станции, обслуживающей участок территории.

При генерации массива абонентов  $H$  каждому абоненту  $h_i$  присваивается номер, соответствующий шаблону перемещений  $sh_i$  из массива шаблонов  $SH$ . Соответственно, каждому абоненту назначается начальная точка на карте и перечень точек его местонахождения в последующие моменты времени.

В синтезируемом массиве строк биллинга  $D$ :  $D = \{d\}_{i=1}^{n_d}$  ( $n_d$  – количество строк) каждая строка соответствует событию (действию) в сети. Строки массива  $D$  формируются из массивов, описывающих события различного типа –  $T, G, C$  и  $S$ :

$T$  – массив событий отправки сообщений о подключениях к базовым станциям для передачи координат:  $T = \{t\}_{i=1}^{n_t}$ ;

$G$  – массив событий получения GPRS-трафика:  $G = \{g\}_{i=1}^{n_g}$ ;

$C$  – массив звонков  $C = \{c\}_{i=1}^{n_c}$ , состоит из массивов исходящих  $C_o$  и входящих  $C_i$  звонков:  $C = C_i + C_o$ ;

$S$  – массив СМС-сообщений  $S = \{s\}_{i=1}^{n_s}$ , состоит из массивов исходящих  $S_o$  и входящих  $S_i$  СМС-сообщений:  $S = S_i + S_o$  (где  $n_t$  – количество

перемещений,  $n_g$  – подключений,  $n_c$ ,  $n_{co}$  и  $n_{ci}$  – звонков,  $n_s$ ,  $n_{si}$  и  $n_{so}$  – сообщений, причём  $n_c = n_{ci} + n_{co}$ ,  $n_s = n_{si} + n_{so}$ ).

Массив соединений  $Y$  является совокупностью массива звонков и массива СМС-сообщений:  $Y = C + S$ , где  $n_y$  – количество соединений ( $n_y = n_c + n_s$ ). Характеристики элементов массива соединений  $Y$ :

- идентификатор типа соединения абонентов  $A \in (\overline{1, n_a})$ ,  $n_a$  – количество различных типов соединений;

- идентификаторы абонентов, участвующих в соединении – абоненты  $h_1, h_2 \in H$  – источник и получатель соединения;

- идентификаторы абонентов  $h_1$  и  $h_2$  в сети сотовой связи (IMSI, MSISDN, IMEI), где  $n_p$  – предельное значение параметра:  $p_1, p_2 \in (\overline{1, n_p})$ .

При формировании массива соединений  $Y$  учитываются статистические характеристики биллинговой информации:

- $K_0 = \langle F_{time}, F_{dur}, F_n, F_a \rangle$  – 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;

- $K_1 = \langle F_l, F_t \rangle$  – связанные с адресацией соединения:  $F_l$  – функция распределения выбора получателей соединения, а  $F_t$  – функция распределения промежутков времени между началами инициализации двух последовательных соединений.

Таким образом, структура соединений определена вектором  $\langle Y_i \rangle_{i=1}^{n_y}$  и описывается выражением:  $Y_i = \langle A, h_1, h_2, p_1, p_2, K_0, K_1 \rangle$ .

Модель соединений  $MS$  определяется выражением:  $MS = \langle H, W, \langle Y_i \rangle_{i=1}^{n_y} \rangle$ .

Модель отправки сообщений о подключениях к базовым станциям для передачи координат и получения GPRS-трафика  $MTG$  формирует массивы  $T$  и  $G$  аналогично модели соединений, за исключением необходимости выбора получателя соединения (ввиду направленности события) и совершения обратного соединения.

С учетом перемещений абонентов по заданным шаблонам пространственно-временная статистико-событийная модель  $M$  взаимодействия абонентов в сетях операторов сотовой связи описывается выражением:  $M = \langle MS + MSH + MTG \rangle$ .

На основе разработанной модели создан программный комплекс, решающий задачу синтеза массивов биллинговой информации, в котором для построения модели соедине-

ний  $MS$  на основе модели перемещений  $MSH$  использован алгоритм сетей Петри. Модель соединений  $MS$  (направленных событий) описывается в терминах сетей Петри, где в качестве множества узлов  $P = \{p_i\}_{i=1}^{n_p}$  ( $n_p$  – количество) принимаются участвующие в соединениях абоненты, а в качестве индикатора наличия соединения между абонентами – вес связывающего перехода.

Каждый узел связан переходом с каждым: если  $i \neq j$ , то узлы  $p_i$  и  $p_j$  связаны переходом. Множество всех разрешенных (на которых имеется фишка) переходов сети  $T = \{t_{ij}\}_{j=1}^{n_{ti}}$  обозначим ( $n_{ij}$  – их количество на  $p_i$ ,  $t_{ij}$  –  $j$ -й переход от узла  $p_i$ ). Переход называется разрешенным, если его вес больше 0. Каждый переход  $t_{ij}$  обладает весом  $m(t_{ij})$ , имеющим смысл количества совершенных соединений между абонентами, которым соответствуют начальный  $p_i$  и конечный  $p_k$  узлы перехода. Количество элементов множества весов переходов графа равно количеству переходов  $\sum_{i=1}^{n_p} n_{ti}$ .

Множество всех разрешенных переходов сети обозначим  $T_w$ ,  $n_{wti}$  – их количество на  $p_i$ ,  $t_{wij}$  –  $j$ -й переход от узла  $p_i$ , тогда  $T_w = \{t_{wti}\}_{j=1}^{n_{wti}}$ .

Моделирование ситуации на основе заданного шаблона происходит в три этапа. На первом этапе из шаблона выделяются все  $n_p$  узлов, участвующих в соединениях, т. е. образуется множество  $P = \{p_i\}_{i=1}^{n_p}$ . После этого из шаблона выделяются все соединения между узлами и записываются в виде переходов, формируется множество  $T_w = \{t_{wti}\}_{j=1}^{n_{wti}}$ .

Заметим, что в ходе построения модели ситуации множество узлов, описывающих ситуацию, может увеличиваться, но не может уменьшаться. В то же время количество переходов может как уменьшаться, так и увеличиваться.

На втором этапе сеть дополняется узлами и переходами, часть переходов запрещается или разрешается. Каждому узлу  $p_i$  случайно задается вес 0, 1 или 2. Условие  $y_k$  задания узлу конкретного значения веса зависит от полученного из шаблона количества разрешенных переходов с него  $t_{wij}$ . Условие  $y_{i1}$  задания узлу  $p_i$  веса 1 определяется по формуле

$$y_{i1} = \frac{n_{wti} - n_{wmin}}{n_{wmax} - n_{wmin}}$$

где,  $n_{wmin} = \min \left\{ \{t_{wti}\}_{j=1}^{n_{wti}} \right\}_{i=1}^{n_p}$  т. е. минимальное количество разрешенных переходов

из узла в шаблоне, а

$$n_{wmax} = \max \left\{ \{t_{wti}\}_{j=1}^{n_{wti}} \right\}_{i=1}^{n_p}$$

т. е. максимальное количество разрешенных переходов из узла в шаблоне. Условие  $y_{i0}$  задания узлу веса 0 и  $y_{i2}$  задания веса 2 определяются по формуле  $y_{i0} = y_{i2} = \frac{1 - y_{i1}}{2}$ .

Узел  $p_j$ , получивший вес 0, удаляется. Для всех узлов  $p_i \neq p_j$  не являющихся удаляемыми, выполняются следующие действия: в случае, если с узла  $p_i$  на удаленный узел  $p_j$  и с узла  $p_j$  на узел  $p_k \neq p_i$  переход был разрешен, то переход с узла  $p_i$  на узел  $p_k$  разрешается (вес увеличивается на 1). Если до этого действия переход узла  $p_i$  на узел  $p_k$  уже был разрешен, то к его весу прибавляется 1.

Узел, получивший вес 1, не изменяется. Переходы на него или с него могут изменяться лишь в том случае, если они изменяются вследствие изменения соседних (связанных с описываемыми разрешенными переходами) узлов.

Узел  $p_j$ , получивший вес 2, сохраняется. Кроме того, во множество  $P$  добавляется дополнительный узел  $p_k$ , связанный запрещенными переходами со всеми узлами из множества  $P$ , и только с узлом  $p_j$  – двумя разрешенными, причем один переход направлен от узла  $p_j$  к узлу  $p_k$ , а второй – в обратном направлении. После добавления узла  $p_k$  для него разрешаются переходы на такие  $p_i$ , переходы на которые с узла  $p_j$  являются разрешенными. Вес добавленного перехода с узла  $p_k$  на узел  $p_i$  равен весу перехода с узла  $p_j$  на узел  $p_i$ .

На третьем этапе всем разрешенным переходам задаются временные отметки в соответствии с распределением соединений по времени. При этом количество временных отметок, присваиваемых переходу, равно его весу.

После этого выполняется запись соединений, информация о которых хранится в сети Петри: последовательно для всех узлов из множества  $P$  строками в биллинге описываются их вес (количество строк), направленность (исходный узел – иницирующий абонент, конечный узел – принимающий абонент) и время (временная отметка в биллинге соответствует временному значению времени перехода, если значений времени перехода несколько, то для генерации каждой следующей строки они берутся последовательно).

## Алгоритм формирования абонентской базы

Для генерации массива биллинговой информации необходимо иметь список абонентов, принимающих в нем участие. Для этого написана программа, генерирующая массив абонентов заданного размера.

На вход программе подаются списки фамилий, мужских и женских имен в порядке уменьшения статистической частоты появления в реальном мире, наборы 8-значных чисел, являющиеся 8 первыми цифрами IMEI сотовых телефонов и описывающими модель и происхождение телефона.

Описание каждого абонента хранится в отдельной строке. Каждый абонент имеет уникальный номер (от 0 до требуемого количества абонентов). Абоненты разделены по однопоколенным «семьям» – группам от 1 до 5 абонентов. Каждый абонент имеет «социальный статус» — поле, содержащее двузначное число, цифра из разряда десятков которого обозначает количество «человек» в «семье» абонента, а вторая цифра — «социальную роль» абонента (1 – отец, 2 – мать, 3, 4, 5 – дети).

Доля семей каждого типа (по количеству членов) в общем количестве генерируемых семей вычисляется, исходя из требуемого количества абонентов и распределения количества одиноких людей и реальных семей с различным количеством детей (0–3). Фамилия у всех членов семьи одна, отчество детей зависит от имени отца.

Каждому абоненту присваиваются номер оператора связи, которому он принадлежит, уникальные значения IMSI (номер SIM-карты), IMEI (номер аппарата) и MSISDN. Значения IMSI генерируется в зависимости от выбранного оператора связи, причем первому абоненту присваивается минимальное значение IMSI, а всем последующим — следующее большее со случайным дискретом от 1 до 100. Значения IMEI генерируются случайно, первыми двумя цифрами могут быть 35 или 01, следующие 6 цифр — номер модели, цифры с 9 по 14 — серийный номер аппарата, генерируется аналогично IMSI. В значениях MSISDN первая цифра всегда 7 (Россия), от оператора связи зависят 3 цифры, остальные 7 цифр генерируются аналогично IMSI.

Алгоритм позволяет создавать до 100 миллионов уникальных номеров IMSI, IMEI и MSISDN.

## Алгоритм и интерфейс программы для создания ситуационных задач

Для создания массива ситуационного биллинга оператору необходимо загрузить в приложение базу абонентов, в которой записаны их ФИО, IMSI, IMEI, MSISDN и т. д.

После загрузки базы в приложение появляется возможность сортировки базы и поиска абонента, который удовлетворяет критериям поиска: семейное положение и количество «человек» в «семье». Таким образом, оператор может выбрать несколько абонентов, которые в дальнейшем будут участниками ситуационного биллинга. Для абонентов в фоновом массиве биллинга задаются четыре типа событий (LocUpd, Call, SMS, GPRS). Все параметры соединений задаются в отдельной форме (рис. 3).

Рис. 3. Процесс добавления абонентов для формирования ситуационной задачи

В итоге в окне приложения (рис. 4) будет изображен граф, вершинами которого являются выбранные заранее абоненты, а ребрами — настроенные соединения. После чего каждое соединение с учетом всех его параметров записывается в отдельный файл для каждого из коммутаторов, который описывает коммуникации группы людей за определенный период времени.

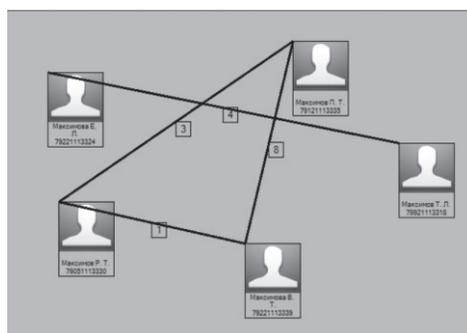


Рис. 4. Интерфейс модуля формирования ситуационной задачи

Считается, что одному оператору связи соответствует один коммутатор, при этом в стандартном населенном пункте присутствуют четыре основных оператора сотовой связи.

После генерации происходит встраивание массива ситуационного биллинга в фоновый.

### **Алгоритм синтеза фонового массива биллинговой информации**

При синтезе фонового массива биллинговой информации используются абонентская база, созданная ранее, файл с LAC и CellID базовых станций выбранной местности, статистические распределения видов событий по времени и продолжительности, распределение количества звонков по дням недели, заранее сгенерированный массив ситуационного биллинга.

После выбора пользователем требуемого количества абонентов, участвующих в биллинге, часть абонентской базы записывается в оперативную память. Также из файлов считываются необходимые распределения и информация о расположении базовых станций, которая используется для создания «карты».

Перемещения генерируются в соответствии с заданными шаблонами. В шаблоне указывается соответствие часа суток и ячейки «карты», в которой находится абонент.

Затем происходит генерация «событий» – записей в массиве биллинга.

События генерируются поочередно для каждого из абонентов. Количество событий каждого вида, совершенных абонентом, выбирается по загруженным ранее статистическим распределениям.

Сначала по заданному в абонентской базе шаблону с использованием карты генерируются все перемещения, затем все GPRS-подключения.

После этого генерируются исходящие звонки. Их количество, продолжительность и время совершения события выбираются по загруженному ранее статистическому распределению. Абонент, которому направлен исходящий вызов, определяется следующим образом: с вероятностью, зависящей от количества «человек» в «семье» абонента и заданной ему при считывании абонентской базы количественной характеристики круга общения, происходит выбор одного из двух направлений соединения: абоненту из «семьи» или случайному. В случае звонка «семье» абонент гарантированно звонит «родственни-

кам», в случае случайного собеседника – с вероятностью  $p = \frac{\text{количество человек в семье} - 1}{\text{общее количество абонентов} - 1}$

«семье» и с вероятностью обратного события – иному абоненту.

После создания записи исходящего звонка от абонента А к абоненту В на коммутаторе абонента А создается запись входящего звонка для абонента В на коммутаторе абонента В. Этим достигается отсутствие необходимости отдельной генерации входящих звонков.

Исходящие и входящие СМС генерируются аналогично звонкам, отличие лишь в типе события и используемых распределениях.

После генерации всех событий для всех абонентов из базы принадлежностей, содержащей информацию об абонентах, в соответствующие массивы добавляются события, сгенерированные при создании ситуационной задачи, и начинается этап сортировки.

Сортировка выполняется для каждого коммутатора отдельно. Все (исходящие и входящие) события объединяются в единый массив. С целью оптимизации сортировки по времени создается дополнительный массив структур, состоящий только из двух полей: номера события (уникального для каждой записи) в исходном массиве и времени совершения события. После этого дополнительный массив сортируется по увеличению времени совершения события методом выбора. Затем информация из исходного массива в порядке номеров из дополнительного, уже отсортированного по времени, массива выводится в файл.

После повторения вышеописанных действий для каждого коммутатора массив биллинговой информации считается сформированным.

### **Статистические характеристики синтезируемых массивов**

Для генерации фонового биллинга необходимы статистические распределения видов событий по времени и продолжительности, распределение количества звонков по дням недели и другие. Данные распределения были собраны на основе личного биллинга, обработаны и представлены в формате графиков.

Статистический анализ синтезируемого биллинга показывает, что его характеристики схожи с исходным массивом. В частности, распределение количества событий по дням недели демонстрирует всплеск активности в

пятничный вечер и падение на период выходных (рис. 5).

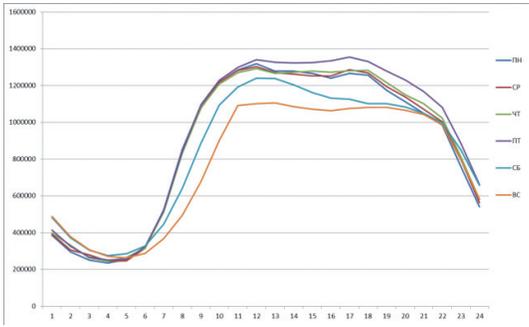


Рис. 5. Статистическое распределение объема биллинга по дням недели

Распределение по продолжительности звонков указывает на то, что большинство звонков длится менее 10 секунд (рис. 6).

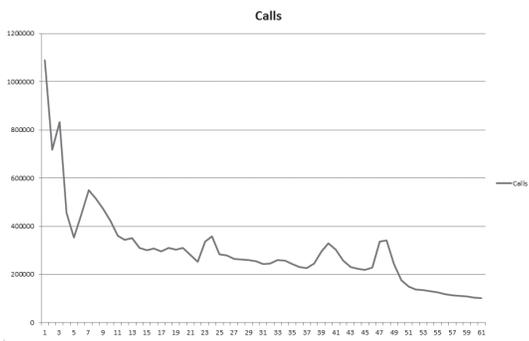


Рис. 6. График распределения продолжительности звонков

График распределения количества событий по времени суток в биллинге, сгенерированном для города с полутора миллионным населением (рис. 7) показывает, что события обновления местоположения абонента

(LocUpd) значительно превышают по количеству остальные типы событий (звонок, СМС-сообщение и подключение к GPRS).

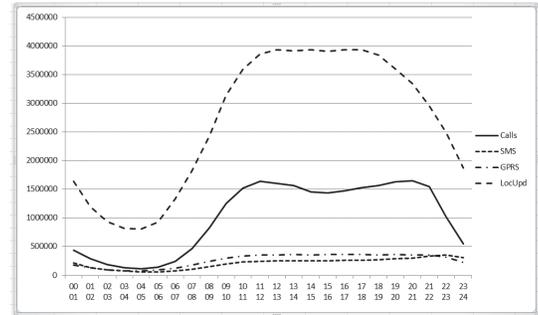


Рис. 7. График количественных отношений событий в биллинге в течение дня

Аналитические исследования [6] подтверждают корректность формируемых распределений. Таким образом, можно говорить о статистической схожести генерируемого биллинга и возможности его использования в практических приложениях.

### Результаты обработки сформированного массива поисково-аналитическими комплексами

Для анализа сформированного массива возможно использовать программу IBM I2 или её аналог – открытую библиотеку CodePlex GraphSharp (рис. 8).

Программа адекватно воспринимает сформированный массив биллинга и предлагает несколько вариантов визуализации. При увеличении количества абонентов, участвующих в моделируемой сети, сложность взаимосвязей увеличивается.

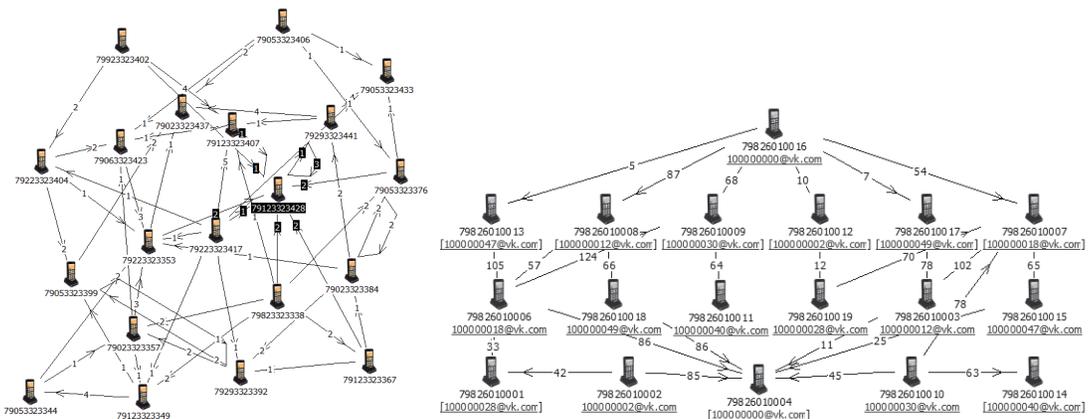


Рис. 8. Пример работы программы I2 при анализе сформированного биллинга на 20 абонентов

## Структура и характеристики программного комплекса

Программный комплекс разработан на базе ОС Windows и состоит из программы генерации базы принадлежностей абонентов, программы создания ситуационных задач, программы статистического анализа массивов биллинговой информации и программы синтеза массивов биллинговой информации.

В процессе разработки программы синтеза массивов применен ряд оптимизационных решений по представлению данных в оперативной памяти и ускорению процессов сортировки записей по временным меткам. Для синтеза массивов биллинговой информации, описывающей взаимодействие 50 тысяч абонентов в течение 1 суток в вымышленном населенном пункте, где работают четыре оператора сотовой связи, при использовании одного стандартного персонального компьютера, оснащенного 8 ГБ оперативной памяти и про-

цессором i7, требуется 1,5 часа. При этом в файлах биллинга формируется 30 миллионов строк, занимающих объем 2,7 Гб. Переносить информацию удобно в сжатом виде, тогда все файлы занимают не более 500 Мб. Формируемые объемы информации, а также время генерации являются приемлемыми для применения в образовательном процессе.

## Выводы

Разработанный программный комплекс позволяет синтезировать массивы фонового биллинга, по статистическим характеристикам идентичные массивам биллинга реальных сетей операторов связи, и интегрировать в них ситуационные задачи, что дает возможность использовать данные массивы для тестирования поисково-аналитических систем и обеспечения практических занятий по изучению информационно-аналитических систем безопасности.

---

## Литература

1. Программно-аппаратный комплекс DATAPK // DATAPK – комплекс оперативного мониторинга и контроля защищенности АСУ ТП. – URL: <http://www.uscc.ru/catalog/id/74> (дата обращения: 05.03.2018).
2. Семенищев И. А., Синадский А. Н., Синадский Н. И. Алгоритм формирования массива биллинговой информации на основе статистической модели поведения абонентов сотовой связи // Сборник материалов XV Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых «Безопасность информационного пространства». – Курган : Курганский ГУ, 2016. – С. 199–203.
3. Семенищев И. А., Синадский А. Н., Синадский Н. И. Статистические характеристики массива биллинговой информации при моделировании поведения абонентов сетей сотовой связи // Сборник материалов 12-ой международной молодежной научно-технической конференции «РТ–2016». – Севастополь : Севастопольский гос. ун-т, 2016. – С. 207.
4. Синадский А. Н., Синадский Н. И. Формальная математическая модель синтеза массива биллинговой информации // Сборник материалов XVI Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых «Безопасность информационного пространства». – Екатеринбург : УрФУ, 2017 (в печати).
5. Семенищев И. А., Синадский Н. И. Статистические характеристики синтезируемых массивов биллинговой информации // Сборник материалов XVI Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых «Безопасность информационного пространства». – Екатеринбург : УрФУ, 2017 (в печати).
6. Rao Zonghao, Yang Dongyuana, Duan Zhengyua. Resident Mobility Analysis Based on Mobile-Phone Billing Data // Procedia – Social and Behavioral Sciences, 96 (2013), С. 2032–2041.

### References

1. Programmno-apparatnyy kompleks DATAPK // DATAPK – kompleks operativnogo monitoringa i kontrolya zashchishchennosti ASU TP. URL: <http://www.uscc.ru/catalog/id/74> (data obrashcheniya: 05.03.2018).
2. Semenishchev I.A., Sinadskiy A.N., Sinadskiy N.I. Algoritm formirovaniya massiva billingovoy informatsii na osnove statisticheskoy modeli povedeniya abonentov sotovoy svyazi // Sbornik materialov XV Vserossiyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh «Bezopasnost' informatsionnogo prostranstva». – Kurgan: Kurganskiy GU, 2016. – S. 199-203.
3. Semenishchev I.A., Sinadskiy A.N., Sinadskiy N.I. Statisticheskiye kharakteristiki massiva billingovoy informatsii pri modelirovanii povedeniya abonentov setey sotovoy svyazi // Sbornik materialov 12-oy mezhdunarodnoy molodezhnoy nauchno-tekhnicheskoy konferentsii «RT–2016». – Sevastopol, Sevastopol'skiy gosudarstvennyy universitet, 2016. – S. 207.

4. Sinadskiy A.N., Sinadskiy N.I. Formal'naya matematicheskaya model' sinteza massiva billingovoy informatsii // Sbornik materialov XVI Vserossiyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh «Bezopasnost' informatsionnogo prostranstva». – Yekaterinburg: UrFU, 2017 (v pechati).

5. Semenishchev I.A., Sinadskiy N.I. Statisticheskiye kharakteristiki sinteziruyemykh massivov billingovoy informatsii // Sbornik materialov XVI Vserossiyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh «Bezopasnost' informatsionnogo prostranstva». – Yekaterinburg: UrFU, 2017 (v pechati).

6. Rao Zonghao, Yang Dongyuana, Duan Zhengyua. Resident Mobility Analysis Based on Mobile-Phone Billing Data // Procedia – Social and Behavioral Sciences, 96 (2013), S. 2032-2041.

---

**СЕМЕНИЩЕВ Игорь Алексеевич**, студент ИРИТ-РтФ УрФУ им. первого Президента России Б. Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19. E-mail: semenishchevigor@mail.ru.

**Синадский Алексей Николаевич**, студент ИРИТ-РтФ УрФУ им. первого Президента России Б. Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19. E-mail: alexsin@e1.ru.

**Синадский Николай Игоревич**, кандидат технических наук, доцент, доцент учебно-научного центра «Информационная безопасность» ИРИТ-РтФ УрФУ им. первого Президента России Б. Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19. E-mail: nickis@e1.ru.

**Сушков Павел Владимирович**, аспирант ИЕНМ УрФУ им. первого Президента России Б.Н. Ельцина; 620002, г. Екатеринбург, ул. Мира, 19. E-mail: сухарпукоб@gmail.com.

**Semenishchev Igor**, student of Institute of Radio electronics and Information Technologie,, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: semenishchevigor@mail.ru.

**Sinadsky Alexey**, student of 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:alexsin@e1.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: nickis@e1.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: сухарпукоб@gmail.com.