УДК 004.93

Теленик С.Ф., д-р. техн. наук
Ролик А.И., канд. техн. наук
Малюков П.Н.
Пищаева Н.Н.

ВЫЯВЛЕНИЕ НЕЛЕГИТИМНОГО ИСПОЛЬЗОВАНИЯ РЕСУРСОВ ОПЕРАТОРОВ МОБИЛЬНОЙ СВЯЗИ

Национальный технический университет Украины «КПИ»

Предложен подход для быстрого и эффективного анализа временных рядов, отображающих конкретные процессы в работе провайдера мобильной связи. Подход основан на методе сегментации кривых и использовании скользящего окна для организации мониторинга ряда. Для анализа трафика предложена модификация алгоритма голосования оценок с применением метрики сходства

Введение

В настоящее время происходят глобальная информатизация общества, создание и повсеместное внедрение технических и технологических условий для коммуникации людей. Наряду с системами традиционной телефонии широкое распространение получили системы и средства *IP*-телефонии, мобильной связи и другие средства осуществления телекоммуникационного взаимодействия. Это позволяет людям осуществлять обмен голосовыми сообщениями, независимо от их месторасположения, например, мобильная связь делает возможным деловое общение сотрудников, находящихся не только на работе или дома, но и на отдыхе, и в дороге.

Эффективность систем телекоммуникационного взаимодействия непосредственно зависит от эффективности работы различных информационных технологий, призванных обеспечить поддержку бизнес-процессов операторов и провайдеров телекоммуникационных услуг. Для качественного предоставления услуг и поддержания своей конкурентоспособности операторы телекоммуникаций должны эффективно использовать имеющиеся у них всевозможные телекоммуникационные ресурсы, не допуская или сводя к минимуму нелегитимное использование каналов связи. Это особенно важно для операторов мобильной связи, использующих дорогостоящие каналы связи с ограниченной пропускной способностью. В этом случае мошенническое или недобросовестное использование каналов связи и соответствующего вспомогательного оборудования, кроме снижения качества предоставляемых абонентам услуг, может привести к большим финансовым потерям и серьезным долговым обязательствам при взаиморасчетах между операторами мобильной связи или мобильной и фиксированной связи [1].

В свою очередь снижение прибыли операторов мобильной связи влечет за собой повышение себестоимости и ограничение средств на модернизацию и развитие сетевой инфраструктуры. Поэтому оперативное определение и своевременное пресечение попыток мошеннического использования каналов связи является одной из наиболее значимых задач для операторов телекоммуникаций, а данная работа, посвященная разработке моделей и методов для выявления нелегитимного использования каналов связи, является актуальной.

Анализ проблемы

Для повышения эффективности деятельности операторов телекоммуникаций создаются и внедряются различные системы управления, осуществляющие автоматизацию управления функционированием предприятия оператора и управление технологическим оборудованием телекоммуникационных сетей. Стремление операторов минимизировать нелегитим-

ное использование каналов связи, своевременно выявлять и пресекать разнообразные способы мошенничества и хакерских действий вынуждает их искать или самостоятельно разрабатывать различные специализированные решения.

Учитывая многообразие телекоммуникационных технологий, широкий спектр методов мошенничества и способов нелегитимного использования ресурсов операторов нельзя создать универсальную систему, способную выявлять все возможные существующие методы мошенничества и методы, которые могут появиться в будущем, пресекать способы злоупотреблений, появляющиеся как реакция на рекламные акции, проводимые операторами и пр. Поэтому возникает необходимость разработки моделей, методов и алгоритмов анализа использования коммуникационных ресурсов абонентами мобильной связи для создания автоматизированных средств выявления случаев нелегитимного использования оборудования и каналов связи, применяемых в системах поддержки операций ИТ-персонала, занимающегося вопросами сохранения доходности бизнеса оператора мобильной связи.

Целью статьи является повышение эффективности использования коммуникационных ресурсов за счет оперативного выявления и пресечения попыток нелегитимного использования каналов мобильной связи.

Постановка задачи

Для контроля объемных и динамических процессов, которые имеют место в сетях операторов связи, необходимо использование быстродействующих алгоритмов и методов мониторинга и анализа трафика. Представление трафика в виде временного ряда позволяет использовать инструментарий прикладного анализа данных.

При этом целесообразно отдельно рассматривать и анализировать ряды с различными аргументами, параметрами или показателями. Например, при рассмотрении трафика операторов сотовой связи это могут быть ряды, отображающие загруженность каналов связи в фиксированные моменты времени, активность отдельных абонентов во времени, количество и продолжительность звонков, проходящих через конкретную базовую станцию, активность абонентов отдельного тарифного плана, трафик роуминга и пр.

Получаемые таким образом временные ряды являются классическими объектами ПАД и для их анализа удобно использовать аддитивную математическую модель, содержащую в качестве компонент пять составляющих – тренд, сезонные и циклические колебания, шум и интервенции [4]. При анализе этих временных рядов предполагается выделение и формализация компонент – систематических и случайных составляющих. Классические методы анализа рядов, например, регрессионный и спектральный, используются для выделения и оценки, соответственно, апериодических и периодических составляющих ряда. Методы авторегрессии и сегментации кривых используются для анализа поведения ряда в локальной временной области. Эти методы направлены на поиск закономерностей в поведении ряда, решение задач идентификации и пр., но не позволяют определить причину изменения поведения ряда и мало пригодны для выявления интервенций – разовых или стохастических внедрений в ряд в результате злоумышленных действий.

Нелегитимное использование каналов связи сопровождается появлением аномальных участков во временном ряду, представляющем один из видов трафика оператора. Всплески или провалы, резкие переходы к более низкому или более высокому установившемуся значению могут свидетельствовать о случаях злоупотреблений или нетипичного поведения абонентов мобильной связи. Поэтому, кроме выявления участков трафика с аномальным поведением и определением источников, являющихся причиной появления таких участков, может потребоваться

анализ множества дополнительных признаков, проведение различных тестов, профилактических мероприятий и пр. для дифференциации нелегитимного использования ресурсов операторов от аномального поведения абонентов. Например, мошенническое оборудование, работающее в качестве *GSM*-шлюза, не будет реагировать на входящие звонки по номеру, который сопровождает исходящий трафик.

Таким образом, для автоматизации процесса выявления случаев нелегитимного использования ресурсов *GSM*-операторов необходимо разработать математические модели, эффективные для анализа временных рядов, содержащих интервенции и другие признаки, характерные для случаев злоупотреблений, автоматизировать анализ ряда дополнительных признаков и, при необходимости, задействовать оператора для принятия окончательного решения о необходимости проведения дополнительных тестов или блокирования недобросовестных абонентов.

Для анализа временных рядов в данной работе использован метод сегментации кривых [5], поскольку он наиболее полно отвечает существу решаемой задачи – выявлению и идентификации аномальных участков трафика.

Классическая реализация этого метода предполагает разбиение ряда на участки с помощью скользящего окна и вычисление на каждом участке функции сложности в виде скалярного произведения текущего окна на предыдущее и последующее. Участок ряда в текущем окне считается аномальным, если значения функции сложности в предыдущем и последующем окнах значительно отличаются друг от друга. Недостатком метода является необходимость многократного вычисления суммы произведений, что влечет за собой повышенные требования к вычислительным ресурсам и становится практически непригодным для обработки больших объемов информации, характерных для баз данных операторов мобильной связи.

Кроме того, в чистом виде метод не пригоден для анализа трафика в реальном масштабе времени, поскольку последующий участок ряда еще не существует. Поэтому в работе для мониторинга трафика используется механизм скользящего окна, а для анализа поведения ряда в окне рассчитываются функции сложности в виде меры близости сравниваемых участков ряда.

Предлагаемый подход к выявлению нелигитимного использования каналов связи

Трафик провайдера связи, представленный в виде кривой, можно разбить на две группы.

В первой группе объект под действием возмущений переходит из одного стабильного состояния в другое, что может свидетельствовать о нелегитимном использовании каналов связи. Задачей сегментации в этом случае является выделение точек, соответствующих этим моментам. Алгоритм сегментации [5] должен разбить кривую на ряд участков, вплотную примыкающих друг к другу и характеризующихся единством формы кривой в пределах каждого участка.

Во второй группе объект преимущественно находится в неизменном состоянии, лишь время от времени выходя из него под действием возмущений. Такие аномалии могут указывать на интервенции или атипичное поведение абонентов сети. Для идентификации злоупотреблений необходимо проанализировать дополнительные признаки. В этом случае информативные участки кривой рассматриваются как следы таких сравнительно коротких событий на фоне общего однородного протекания исследуемого процесса. На таких кривых алгоритм сегментации должен выделять лишь некоторые фрагменты, рассматриваемые как информативные, пропуская остальную часть кривой.

Кривые обеих групп являются записями дискретных упорядоченных после-

довательностей событий, а анализ кривых есть не что иное, как описание этих последовательностей. В качестве объекта рассматривался сегмент временного ряда, в качестве свойств – конкретные моменты времени, а для сравнения сходства или различия объектов использовали формальную метрику сходства — меру близости [6].

Мера близости двух объектов (X) по каждому из N свойств задается в следующем виде

$$B(X(1), X(2)) = \sum_{j=1}^{N} b(j, X(1, j), X(2, j)),$$

где близости объектов $b(j, X(1, j), X(2, j))$ по j-му свойству определяются следующим образом

$$b(j, X(l, j), X(2, J)) = \begin{cases} 1, & \text{если } |X(l, j) - X(2, j)| \le \varepsilon(j), \\ 0 & \text{в противном случае.} \end{cases}$$

Близость объектов определяется числом совпадений свойств сравниваемых объектов с точностью до априорно оговоренного (заданного) допуска $\varepsilon(j)$ по каждому признаку, в частности, при $\varepsilon(j) = 0$ предполагается эквивалентность соответствующих свойств сравниваемых объектов.

Таким образом, для вычисления количественной оценки сходства сравниваемых сегментов ряда достаточно подсчитать количество совпадений в описаниях сегментов в пределах априорно оговоренных допусков.

Применение меры близости в качестве функции сложности при сравнении сегментов анализируемого ряда позволяет:

– уменьшить требуемые вычислительные ресурсы,

– увеличить быстродействие и выполнять мониторинг в реальном масштабе времени,

– обеспечить исключительную гибкость настройки на характер выявляемой аномалии.

Для реализации метода сегментации кривых был выбран в рамках Г-модели алгоритм вычисления оценок. Алгоритм распознавания неизвестного объекта в алгоритме вычисления оценок включает следующие шаги:

– вычисление индивидуальных оценок близости распознаваемого объекта X_i к каждому из M, $i \in M$, объектов

$$\Gamma(X, X_i) = \sum_{j \in N} |X(j) - X(i, j)| \le \varepsilon(j);$$

– голосование индивидуальных оценок, т. е. вычисление по индивидуальным оценкам совокупных оценок близости распознаваемого объекта к каждому из существующих K классов, нормированные по количеству M_r объектов в r-м классе

$$\Gamma(X, R) = \frac{1}{M_r} \sum_{i \in R} \Gamma(X, X_i);$$

– в голосовании совокупных оценок, т. е. в вычислении на основании совокупных оценок окончательной оценки принадлежности распознаваемого объекта, соответствующего участку графика загруженности отдельного канала связи, к одному из априорно заданных классов поведения кривых.

Если аналогично оценке индивидуальных оценок определить частичные совокупные оценки близости объекта X к r-му классу обучающей выборки по остальным опорным множествам, совокупная оценка близости объекта к r-му классу по системе опорных множеств $S = \{S_q\}$, $q = 1, ..., K$ определяется как

$$\Gamma r(X) = \sum_{r \in K} \Gamma(X, R).$$

Решающее правило, реализующее голосование совокупных оценок близостей, отнесет неизвестный объект X к тому из классов, для которого получена максимальная совокупная оценка близости.

– модификация АВО, состоящая в трактовке участков трафика: участок, выделенный текущим положением скользящего окна, принимается за неизвестный объект, а участки, выделенные предыдущим и последующим положением сколь-

зящего окна (а в общем случае и любых других) рассматриваются в качестве объектов-прецедентов выборки.

Выводы

В работе предложен новый подход к решению задачи мониторинга и анализа трафика провайдера мобильной связи, основанный на методе сегментации кривых, а новизна его заключается в следующем:

– для мониторинга рядов использован метод сегментации кривых при помощи скользящего окна;

– для количественной оценки сравниваемых участков трафика, выделяемого скользящим окном, применена мера близости в качестве функции сложности, что позволило снизить расходы вычислительных ресурсов;

– для анализа трафика предложена модификация алгоритма голосования оценок, состоящая в трактовке участков трафика: участок, выделенный текущим положением скользящего окна, принимается за неизвестный объект, а участки, выделенные предыдущими положениями скользящего окна (а в общем случае и любых других), рассматриваются в качестве объектов-прецедентов.

Применение предложенных методов в подсистемах управления сетевой безопасности систем управления информационно-телекоммуникационными системами [7] и, в частности, управления сетью и инфраструктурой операторов мобильной связи позволяет существенно сократить потери операторов за счет своевременного пресечения попыток несанкционированного использования его сетевых ресурсов.

При этом может потребоваться дополнительно осуществлять анализ различных второстепенных признаков, таких как время пребывания абонента в качестве клиента оператора мобильной связи, среднее количество звонков абонента за заданный интервал времени, средняя длительность звонков, направление исходящего звонка, успешность его завершения, перемещение абонента между базовыми станциями за рассматриваемый интервал времени, нахождение абонентов в потенциально опасном с точки зрения возможных злоупотреблений тарифном плане и пр.

Список литературы

1. *Попков Д.* Tranzit-fraud, или Мошенничество по крупному ИнформКурьер-Связь, 2005. – № 2. – С. 55–56.

2. *Dong W., Quan-yu W.* Shou-yi Z. Feng-xia L. Da-zhen W. A feature extraction method for fraud detection in mobile communication networks//Intelligent Control and Automation. – 2004, June, vol.1. – P. 1853–1856.

3. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных. – М.: «Мир», 1989. – 526 с.

4. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. – Вып. 1, – 406 с.; вып. 2. – 197 с.

5. *Брaверманн Э. М., Мучник И. Б.* Структурные методы обработки эмпирических данных. – М.: Наука, 1983. – 464 с.

6. *Компанец Л.Ф., Краснопрошина А.А., Малюков Н.Н.* Математическое обеспечение научных исследований в автоматике и управлении. – К.: Вища шк., 1992. – 287 с.

7. *Теленик С.Ф., Ролік О.І., Букасов М.М., Соколовський Р.Л.* Система управління інформаційно-телекомунікаційною системою корпоративної АСУ // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка. – К.: «ВЕК+», 2006. – № 45. – С. 112–126.