Математические вопросы кибернетики и вычислительной техники 24, 2005, 125-132.

О статистическом подходе к обнаружению нарушений в телефонных сетях

Карен А. Мкртчян

Институт проблем информатики и автоматизации НАН РА
e-mail: kamkrtchyan@ipia.sci.am, kamkrtchyan@yahoo.com

Аннотация

В статье предлагается метод моделирования образов объектов, позволяющий обнаруживать нарушения в телефонной сети, с применением статистического анализа.

1. Введение

Проблема обнаружения нарушений (fraud) в телефонных сетях с каждым годом становится все актуальней. В связи с большим разнообразием вредоносных действий в сфере телекоммуникаций, эта индустрия ежегодно несет большие финансовые потери. В Китае, согласно отчетам, в 2001 году, по причине нарушений в сфере телекоммуникаций, были нанесены убытки в размере 20 миллиардов юаней [1]. В западноевропейских странах, в зависимости от времени и специфики предоставляемых услуг, потери оцениваются в размере 3% - 6% от доходов [2]. Вместе с появлением новых технологий, оборудования и услуг нарушения в сфере телекоммуникаций становятся все изощренней [3], [4], [5], [6]. Примерами нарушений могут быть нарушения типа: By Pass, Call Back, Launching, Landing, 3rd Country Refile, Information Distortion, и т.д, более подробно о этом можно прочитать в [7]. Соответственно выявление вредоносных действий требует разработки методологии, учитывающей специфику каждого нарушения и, в частности, включающей проведение мониторинга [3], [4], [5], [6]. Но, зачастую, из-за различных организационных, технических, финансовых сложностей, часть исследуемых абонентов остается вне поля наблюдения, либо не все их характеристики удается зарегистрировать. В таких случаях, о тех или иных характеристиках интересующих нас абонентов, не зарегистрированных при мониторинге, можно судить опосредствованно, путем систематизации и анализа параметров и характеристик других аналогичных абонентов, попавших в поле зрения при мониторинге за какой-то промежуток времени.

Цель данной статьи - продемонстрировать возможность выявления нарушений в телефонной сети путем их моделирования с использованием статистического метода.

Предлагается, используя статистические методы, создать математическую модель категорий абонентов телефонной компании, с помощью которой можно будет вычислять образы их поведения. Изменяя наборы атрибутов и задавая им различные значения, можно будет получать типичные образы поведения абонентов разных категорий.

Сталкиваясь с абонентами, имеющими аналогичные атрибуты и образ поведения, можно будет, с некоторой вероятностью, судить об их принадлежности к тому или иному типу, и в дальнейшем, с помощью целенаправленного мониторинга, выявить истинных нарушителей.

2. Формулировка необходимых понятий

Ниже будет использована следующая терминология.

Телефонная компания (*оператор связи*) это организация, предоставляющая услуги телефонной связи. Термином *абонент* мы называем частное лицо, либо организацию, пользующихся услугами телефонной компании. Сокращением *PSTN* (Public Switched Telephone Network) обозначается коммутируемая телефонная сеть общего пользования (сеть, для доступа к которой используются обычные телефонные аппараты, мини-АТС, оборудование передачи данных). Под выражением *атрибуты абонента* будем понимать сведения об абоненте, зарегистрированные в информационной базе телефонной компании. Например, информационная база для каждого абонента может содержать следующие атрибуты:

- идентификатор абонента: частное лицо, организация "такси сервис", производственная компания, интернет провайдер, и т.д.,
- паспортные данные для частных лиц и аналогичная информация для организаций или компаний,
- список предоставленных номеров телефонов и их адресов,
- права и обязанности обусловленные договором между телефонной компанией и абонентом.

Часть сведений об абоненте предоставляется им самим при оформлении договора, а часть собирается соответствующими службами телефонной компании в дальнейшем по ходу функционирования. Например: своевременность выплат, случаи нарушений договора, какие это были нарушения, соучастники нарушений, какие действия были предприняты и т.д.

Термином *категория абонентов* обозначается группа абонентов, имеющая одинаковый список некоторых атрибутов. Примеры категорий абонентов: частные пользователи, организации такси-сервис, компании интернет-провайдеров, и т.д.

Мы рассмотрим лишь два *типа абонентов*: *нормальные* и *нарушители*. Здесь мы ограничимся рассмотрением нарушений (fraud), связаных с действиями абонентов, в результате которых телефонная компания несет убытки (получает оплату, неадекватную предоставленным ею услугам).

Под термином *мониторинг* будем понимать процедуру наблюдения за функционированием телефонной сети, с целью выявления:

- нарушений,
- механизмов их реализации,
- нарушителей.

Организационно-технические средства телефонной компании для регистрации всех видов услуг и их продолжительности по каждому телефонному номеру абонента образуют *биллинговую систему*.

Характеристиками абонента, регистрируемыми в биллинговой системе, являются значения типов услуг абоненту. Например, в случае звонка абонента, в качестве *характеристики абонента* биллинговой системой, будет зарегистрирован сам факт звонка — *исходящий (outgoing call)* - номер телефона, с которого был совершен звонок, *входящий (incoming call)* - номер телефона, на который был совершен звонок, продолжительность связи, и т.д.

Образ поведения это набор характеристик абонента.

Математическая модель категорий абонентов это приближенное описание групп абонентов телефонной компании математическими средствами с целью предсказания их образа поведени [5].

3. Классификация моделей

Как известно, классифицировать модели можно по разным критериям [6]. По характеру решаемых проблем модели могут быть разделены на *функциональные* и *структурные*.

В первом случае все величины, характеризующие явление или объект, выражаются количественно. При этом одни из них рассматриваются как независимые переменные, а другие — как функции от этих величин. Фактически математическая модель представляет собой систему уравнений (дифференциальных, алгебраических и т. д.), устанавливающих количественные зависимости между рассматриваемыми величинами.

Во втором случае модель характеризует структуру сложного объекта, состоящего из отдельных частей, между которыми существуют определенные связи. Для построения таких моделей удобно использовать теорию графов.

По характеру исходных данных и результатов предсказания модели могут быть разделены на *детерминистические* и *вероятностно-статистические*.

Детерминистические модели дают определенные, однозначные предсказания, а вероятностно-статистические модели основаны на статистической информации, предсказания, полученные с их помощью имеют вероятностный характер.

В соответствии с приведенной классификацией, предлагаемая нами математическая модель, по характеру решаемых проблем является *функциональной и/или структурной*, а по характеру исходных данных — *вероятностно-статистической*.

4. Основные этапы математического моделирования

На этапе *построения модели* анализируются различные категории абонентов, выявляются их особенности и устанавливаются математические выражения связей между атрибутами и характеристиками абонентов выбраной категории. Затем *решается математическая задача, к которой приводит исследование модели*. На этом этапе разрабатываются алгоритмы и численные методы решения. *Интерпретация следствий полученных с помощью математической модели* — следующий этап. Для *проверки адекватности модели* выясняется, согласуются ли интерпретированные результаты с результатами, полученными традиционным путем, в пределах определенной точности. Далее производится *модификация модели* - корректировка, с тем, чтобы она была более адекватной действительности.

5. Постановка задачи

Задача состоит в разработке методов математического-статистического моделирования типов объектов с целью выявления нарушителей.

Более подробно предлагаемый подход продемонстрируем на примере выявления нарушения одного типа.

6. Развернутый анализ нарушения типа Landing.

Сравним положительные и отрицательные стороны процедур обычно используемых при мониторинге и путем моделирования. Как будет показано далее, нарушение типа Landing выбрано в связи с тем, что, для его мониторинга требуются затраты, существенно превышающие затраты, связанные с применением статистического метода.

Схематически телефонную сеть (PSTN) какой-либо компании (оператора связи) можно представить в виде группы автоматических телефонных станций *(АТС)*, соединенных друг с другом каналами связи и обеспечивающих организацию связи между абонентами (см. Рис. 1).

Одна или несколько АТС одного оператора связи предназначаются для связи с другими операторами и их абонентами, такими АТС на Рис. 1 являются АТС k, оператора 1, АТС r, оператора 2, АТС 2, оператора 3.

Пусть абонент E сети оператора 3 должен позвонить абоненту A сети оператора 1. В обычном случае этот звонок должен пройти по трассе:

Абонент E→АТС t (опер.3)→АТС 2 (опер.3)→АТС k (опер.1)→АТС n (опер.1)→Абонент A.

В случае Landing-а трасса будет выглядеть следующим образом:

Абонент E→АТС t (опер.3)→АТС 1 (опер.3)→Абонент G→Router2→ Satellite dish2→ Satellite dish1→Router1→Абонент F→АТС m (опер.1)→АТС n (опер.1)→Абонент A.

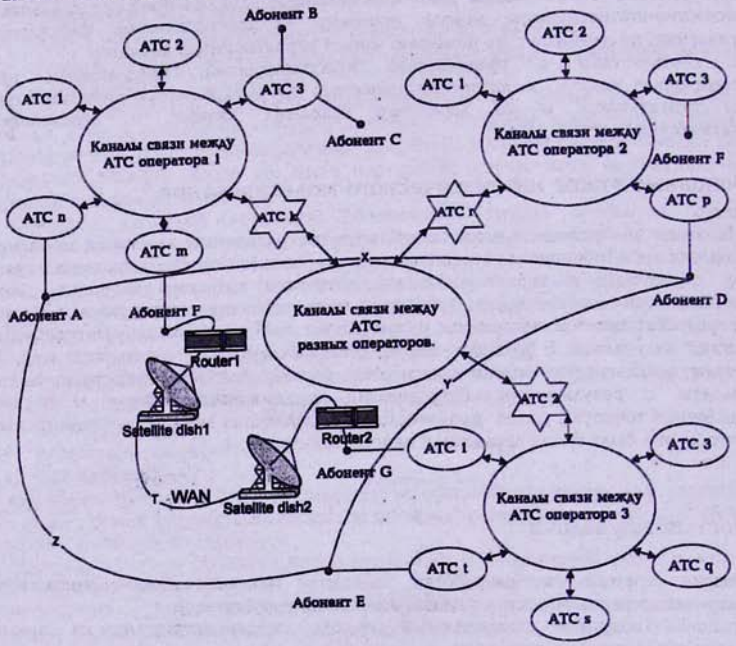

Рис. 1. Организация связи между абонентами различных операторов.

Каждая АТС имеет свою биллинговую систему. На современных АТС биллинговые системы (БС) компьютеризированы и представляют собой информационные АСУ. Биллинговые системы АТС, принадлежащих одной телефонной компании, объединяются в единую сеть, тем самым образуя единую БС телефонной компании (см. Рис. 2).

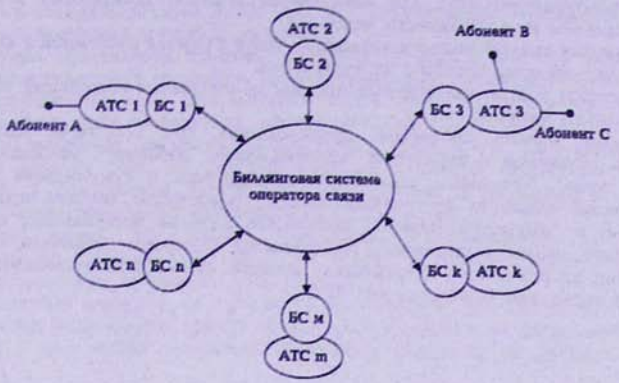

Рис. 2. Биллинговая система оператора связи.

В обычном случае звонок должен быть зарегистрирован в БД АТС k и БД АТС n оператора 1. В случае Landing-а звонок будет зарегистрирован на локальных АТС m и АТС n, и будет отсутствовать в БД АТС k.

Участок трассы router2 -> Satellite dish2 -> Satellite dish1 -> router1 -> представляет собой альтернативный, более дешевый, канал связи. Тариф звонка, проходящего через АТС 2 (опер.3) —> АТС k (опер.1) выше тарифа через Satellite dish2 -> Satellite dish1. Такого типа связь может быть организована абонентами F и G. Получив звонок от абонента - оператора G через Satellite dish1 и имея доступ к PSTN оператора 1, абонент F производит соединение с любым абонентом этой сети. Типичным примером такой связи может быть связь типа Voice over IP (VoIP). Фактически абоненты F и G в этом случае становятся операторами связи и зарабатывают на разнице в размере тарифов междугородной и VoIP связи, принося убытки операторам связи 1 и 3.

Рассмотрим подробнее процесс обнаружения нарушений типа Landing средствами обычного мониторинга:

- выделяются абонентские телефонные номера, аналогичные номеру абонента А, на которые будут производиться тестовые звонки. Такие номера называются *Target* (мишень) номерами,
- на Target номерах устанавливаются номероопределители,
- организуются серии тестовых звонков с телефонов аналогичных телефону абонента Е на Target-номера. Такой процесс называется *Bombarding*,
- на номероопределителях Target-номеров регистрируются телефонные номера, с которых были получены звонки.
- если эти номера являются номерами локальных абонентов, аналогичных абоненту F, то их владельцы совершают нарушение типа Landing,

- далее проверяются абоненты, которым принадлежат зарегистрированые на номероопределителях телефонные номера, выясняется имеют ли они право, в рамках своего договора с оператором связи, на организацию связи такого типа...

После анализа всей накопленной информации, руководством телефонной компании, принимается решение о дальнейших действиях:

- о возбуждении иска для возмещения убытков понесенных компанией, в результате их деятельности, через суд,
- создании аналогичных конкурентных услуг в рамках собственной компании,
- другие средства борьбы с нарушителями.

Рассмотрим теперь аналогичный процесс обнаружения нарушений типа Landing статистическим методом.

Проведем анализ процессов, происходящих при Landing-е, например для категории абонентов с атрибутом "идентификатор абонента", имеющим значение "частное лицо". Установим зависимости между ним и суммарными значениями характеристик абонента "исходящий звонок" (количество звонков произведенных абонентом) и "входящий звонок" (количество звонков поступивших к абоненту). Графическое, соответствие между телефонными номерами абонентов, суммарными значениями их исходящих и входящих звонков за какой-то промежуток времени, например месяц, изображено на Рис. 3.

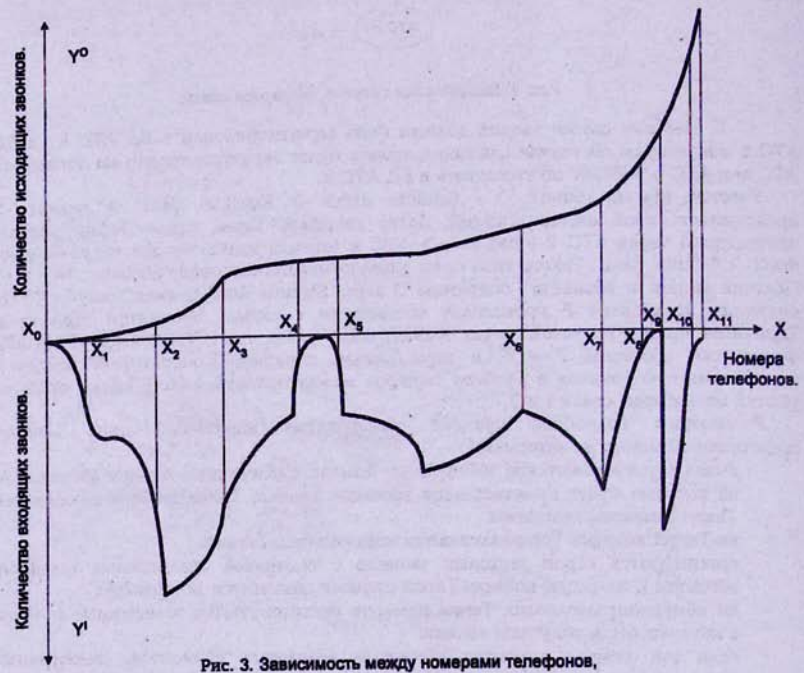

Рис. 3. Зависимость между номерами телефонов, их входящими и исходящими звонками.

По оси X, располагаются номера телефонов - X_m, абонентов. Вверх по оси Y^o располагаются соответствующие этим номерам суммарные значения характеристик "исходящий звонок" Y^o_{Xm}, вниз по оси Y^i - суммарные значения характеристик "входящий звонок" Y^i_{Xm}. Значения X_m располагаются в порядке возрастания Y^o_{Xm}.

Количество входящих звонков на телефонные номера, задействованные в Landinge-е, должно быть незначительно, поскольку они должны быть всегда свободны для выполнения запросов, поступающих от компаньонов из других *PSTN* на соединение с абонентами своей *PSTN*. Вместе с тем количество исходящих звонков с этих номеров телефонов должно быть велико.

Попробуем, анализируя график, составить представление об особенностях использования телефонов абонентов. Номера телефонов, расположенные между X_0 и X_1, либо отключены, либо используются очень мало, поскольку количество их входящих и исходящих звонков либо отсутствует, либо оно незначительно. Абоненты с номерами телефонов, расположенными между X_1-X_2, могут быть заподозрены в нарушении другого типа, не рассматриваемого в данной статье. Не вызывают подозрений, в совершении нарушений типа Landing, и абоненты с номерами телефонов расположенными в интервалах X_2-X_4, X_5-X_8, X_9-X_{10}.

Из графике видно, что номера телефонов, задействованных в Landing-е, расположены на участках X_4-X_5, X_8-X_9 и X_{10}-X_{11}. В первую очередь привлекают внимание абоненты из интервалов X_8-X_9 и X_{10}-X_{11}.

Особенностью номеров из группы X_4-X_5 является то, что либо частота их использования мала, либо входящие звонки распределяются по нескольким номерам телефонов с тем, чтобы показатели отдельного номера не выделялись из общей картины.

Анализ показал, что:

- количество исходящих звонков с телефонов интересующих нас абонентов, значительно превышает средние показатели,
- количество входящих звонков этих абонентов незначительно,

Таким образом алгоритм можно разбить на следующие этапы:

- выбор номеров телефонов обрабатываемой категории абонентов,
- вычисление типичных значений характеристик "исходящие звонки" и "входящие звонки" на основании статистической обработки,
- выделение абонентов, значения соответствующих характеристик которых значительно отличаются от типичных,
- дальнейшая тщательная проверка действий выделенных абонентов.

Подробное описание процедур реализации метода будет представлено в дальнейшем.

Литература

[1] L. Cao and other. Hybrid Strategy of Analysis and Control of Telecommunication Frauds. attend.it.uts.edu.au/icita05/ CDROM-ICITA04/papers/62-1.pdf 2004.

[2] Paul de Jager. "Introduction to using intelligent techniques for telecommunications fraud detection". http://www.eurescom.de/~pub/seminars/past/2001/SecurityFraud/12-Jager/

[3] James L. Johnson and other. Local Prosecutors' Experiences Fighting Telecommunications Fraud. www.ndaa-apri.org/pdf/sounds_too_good.pdf American Prosecutors Research Institute, 2004

[4] E. Lundin. Aspects of employing fraud and intrusion detection systems. Technical Report no. 2L, Department of Computer Engineering, Chalmers University of Technology, Geoteborg, Sweden, 2002.

[5] P. Gosset and M. Hyland. Classification, Detection and Prosecution of Fraud on Mobile Networks. 1998.
http://www.esat.kuleuven.ac.be/cosic/aspect/papers/mobsummit.doc
[6] E. Lundin. Combining fraud and intrusion detection - meeting new requirements. http://www.ce.chalmers.se/~emilie/ 2004
[7] Russell G. Smith, Stealing Telecommunications Services, Australian institute of criminology, http://www.aic.gov.au/publications/tandi/ti54.pdf 2005
[8] М. Скворцова. Математическое моделирование.
http://archive.1september.ru/mat/2003/14/no14_1.htm 2003.
[9] Е. Арутюнян и другие. Вероятность и прикладная статистика. (на армянском языке), Издательство "Гитутюн" НАН РА, Ереван 2000.

Հեռախոսային ցանցերում խախտումների բացահայտման վիճակագրական մի մոտեցման մասին

Կ. Սկրտչյան

Ամփոփում

Հոդվածում առաջարկվում է հեռախոսային ցանցերի օգտվողների մաթեմատիկական պատկերների վիճակագրական վերլուծության միջոցով մոդելավորման մի եղանակ՝ խախտումների բացահայտման համար: