ОБЗОР СИСТЕМ ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА В ТЕЛЕКОММУНИКАЦИОННОМ ПРЕДПРИЯТИИ

A REVIEW OF FRAUD DETECTION SYSTEMS IN TELECOMMUNICATION COMPANY

T. Nguyen

Summary. In the present fraud detection and prevention are properly mechanism to protect against fraud. This article presents the problems of implementation of fraud detection systems in the telecommunications company, methods for the detection of fraud and the proposed architecture of the fraud detection system in real time using big data technology.

Keywords: telecommunication fraud, fraud detection, machine learning, real time processing, batch processing, big data technique, data mining.

Нгуен Туан Ань
Аспирант, Волгоградский государственный технический университет
anhtuank37@gmail.com

Аннотация. В настоящий день обнаружение и предотвращение мошенничества являются надлежащим механизмом защиты от мошенничества. В данной статье представлены проблемы реализации системы обнаружения мошенничества в телекоммуникационном предприятии, также методы обнаружения мошенничества, и предложена архитектура системы обнаружения мошенничество в реальном времени с использованием технологии больших данных.

Ключевые слова: телекоммуникационные мошенничества, обнаружения мошенничества, машинное обучение, обработка данных в реальном времени, пакетная обработка данных, методы интеллектуального анализа данных.

Введение

Мошенничество телекоммуникаций — это проблема, которая становится актуальной проблемой за последние десять лет [1]. Мошенничество в области мобильных телекоммуникаций — это сложная и динамическая задача для операторов связи. Так как эти мошенничества угрожают предполётные и пост-платные услуги. Кроме этого, мошенничество может быть совершенным на стационарные и мобильные телефонные линии [2]. Мошенничество стационарной телефонной линии совершенно противно телефонных компаний; Это как мошенник, который получает доступ к коммутатору и продаёт другим людям возможность, чтобы совершал звонки через коммутатор [3]. Мобильное мошенничество — несанкционированное использование, искажение или манипулирование для сотового телефона или услуги. Как правило, основная цель за совершение мошенничества в оба вида связи (фиксированной, мобильной линии) для получения услуги и звонков незаконными способами [2].

На основе обзора потери глобального мошенничества, объявленного коммуникацией ассоциацией по борьбе с мошенничеством (CFCA), в 2013 году потери от мошенничества был зафиксирован в 46,3 млрд. долларов, что на 15% больше, чем в 2011 году. В процентном отношении к глобальным телекоммуникационным доходам убытки от мошенничества составляют примерно 2,09% — увеличение 0.21% по сравнению с 2011 г. [4]. Это связано с большим количеством мошенничества телекоммуникаций, записанного из разных категорий.

В данной работе будут представлены проблемы мошенничества телекоммуникаций и также методы обнаружения мошенничества телекоммуникаций в практике.

Классификация типов мошенничества в телекоммуникации

В [5] мошенничества телекоммуникаций сгруппированы на четыре категории:

- Договорное мошенничество (Contractual fraud): мошенник использует услуги связи без намерения оплатить услугу, например, абонентский фрод (subscription fraud) и Premium Rate Fraud.
- Хакерское мошенничество (hacking fraud) — проникновение в компьютерную систему безопасности для удаления механизмов защиты или переконфигурирования системы с целью несанкционированного использования сети, например, мошенничество учрежденческой автоматической телефонной станции (УАТС) и сетевая атака.
- Техническое мошенничество (technical fraud): мошенники в этой категории капитализируют на слабости, существующие в технологии мобильной системы. Такое мошенничество требует высоких технических знаний. Примеры такого мошенничества — клонирование (Cloning) и внутрикорпоративное техническое мошенничество (Technical Internal fraud). В последнем случае

Рис. 1. Механизмы защиты от мошенничества

мошенник получает возможность пользования услугами связи по сниженной цене за счёт незаконного доступа к корпоративной системе. При осторожном использовании этот способ мошенничества наиболее сложен для обнаружения.

- Процедурное мошенничество (procedural fraud) — Мошенники в этой группе участвуют нападения на процедуры, которые используются для уменьшения риска мошенничества, и часто нападают на слабые места в бизнес-процедуры (например, биллинг), используемые для предоставления доступа к системе с целью уменьшения оплаты услуг связи. Примером такого мошенничества являются мошенничества в роуминге (Roaming fraud), дублирование ИД ваучера (Voucher ID duplication), и неисправные ваучеры (Faulty vouchers).

С другой стороны, работа [6] классифицирует мошенничества электросвязи по трём областям, которые являются:

Мотивом: основная причина за совершение мошенничества.

Средством: характер или форма мошенничества, используются для удовлетворения мотива.

Методами: средства и инструменты, которые используются для совершения мошенничества.

Существует множество видов мошенничества, которые угрожают телекоммуникационным секторам, которые считаются наиболее популярное место мошенничество. Подсчитано, что существует более 200 вариантов мошенничества телекоммуникаций, существующие в телекоммуникационной отрасли [7]. Существует много методов для решения этих проблем мошенничества [8].

Процессы обнаружения мошенничества в телекоммуникации

Мошенничество резко возрастает с прогрессированием современных технологий и глобальной коммуникации. В результате этого борьба с мошенничеством стала важным вопросом, который предлагается в [9] [10]. Как указано на рис. 1 механизмы обнаружения и профилактики используются по большей части для борьбы с мошенничеством. Следующие подразделы объяснят механизмы защиты от мошенничества.

Система предотвращения мошенничества является первым уровнем защиты, чтобы обезопасить технологические системы от мошенничества. Цель этой фазы состоит в том, что остановить мошенничество от происходящих в первую очередь. Механизмы в этой фазе ограничат, подавляют, уничтожают, разрушают, контролируют, удаляют, или предотвратят возникновение кибер-атак, в компьютерных системах, сетей и данных. Пример функционирования такого механизма включает в себя использование алгоритма шифрования, который

применяется для скремблирования данных. Другой механизм — межсетевой экран, где она образует блокады между внутренней частной сетью и внешними сетями. Она не только обеспечивает безопасность системы от несанкционированного доступа, но и позволяет организовать реализации политики сетевой безопасности трафика, проходящего между его сетью и интернетом [11] [10]. Однако этот уровень не всегда эффективный и сильный [12]. В ряде случаев уровень профилактики может быть нарушены мошенниками.

Следует, что необходимо создать системы обнаружения мошенничества (СОМ).

Система обнаружения мошенничества-это следующий уровень защиты. Обнаружение мошенничества пытается обнаружить и выявить мошеннические действия, когда они входят в системы и сообщать о них системному администратору [13]. В последние годы ручные методы аудита мошенничества, такие как выборка (sampling) обнаружения были использованы для обнаружения мошенничества, например, в [13]. Эти сложные и трудоемкие методы совершают операции с различными областями знаний, как экономикой, финансам, правом и деловыми практиками. Поэтому, для повышения эффективности обнаружения, компьютеризированной и автоматизированной СОМ был изобретён.

Однако, возможности СОМ были ограничены, поскольку обнаружение принципиально зависит от предопределённых правил, которые заявили эксперты [14]. Более сложные СОМ, интегрирующие широкий спектр методов интеллектуального анализа данных, требуются и разрабатываются для эффективного выявления случаев мошенничества [15] [16] [17]. Интеллектуальный анализ данных (ИАД) включает статистические, математические, искусственного интеллекта и методы машинного обучения для извлечения и выявления полезной информации и последующего знаний из больших баз данных (систем поддержки принятия решений и интеллектуальных систем). Эти системы имеют несколько основных преимуществ: (1) Схема мошенничества получаются автоматически из данных; (2) уточнение « вероятность мошенничества» для каждого конкретного случая, следовательно, что усилия в расследовании подозрительных случаев могут быть приоритетными; и (3) выявление новых видов мошенничества, которые не были определены [14].

Методы интеллектуального анализа данных состоит из шести основных категорий: классификации, кластеризации, регрессии, обнаружение выбросов, визуализация и прогнозирование [18]. Каждый из этих методов поддерживается конкретными методами. Например, искусственная нейронная сеть (ИНС) и метод опорных векторов (англ. SVM, support vector machine) используются для метода классификации. метод К-средних используется для кластеризации данных.

Кроме того, ИАД включил в себя многие приёмы из других областей, таких как статистика, машинное обучение, распознавание образов, базы данных и системы хранилища данных, информационный поиск, визуализация, алгоритмы, высокопроизводительные вычисления, и многие домены приложений [19]. В последнее время, обнаружение мошенничества интегрирует подход обнаружения на основе аномалии и неправильного использования методами ИАД [20].

В практике популярно используются методы обнаружения мошенничества на основе аномалия. Подход обнаружения аномалий и выбросов используется СОМ, и он опирается на поведенческие методы профилирования, где он моделирует каждую модель поведения индивидуума, контролирующегося на какие-либо отклонения от нормы [21]. СОМ на основе аномалии принимается многочисленными авторами в разных областях мошенничества [22]. СОМ на основе аномалия имеют потенциал для обнаружения нового мошенничества. Поэтому он в основном используется в литературы СОМ [23]. Этот метод может быть дополнительно классифицирован на три вида: контролируемый, без учителя, частичного обучения [24].

I. Контролируемый (Supervised)

Методы контролируемого обучения требуют набор данных, которые были помечены как «мошенничество» и « не-мошенничество» и включают в себя обучение классификатор. Это самый распространённый подход к обучению. Основным преимуществом контролируемого обучения является то, что все выходы классов, манипулированы алгоритмом этого подхода, являются значимыми для человека, и он может быть легко использован для классификации образов и регрессионного анализа данных. Однако, контролируемое обучение имеет ряд ограничений. Первое обусловлено трудностями сбора наблюдения или меток. Когда есть огромный объем входных данных, это слишком дорого, чтобы маркировать все из них. Во-вторых, иногда это чрезвычайно трудно найти отличительные метки, потому что существует неопределённости и неясности в наблюдение или метки. Эти ограничения могут препятствовать реализации подходов к контролируемому обучению в некоторых случаях. Таким образом, обучение без учителя и частичное обучение используются для преодоления этих недостатков [25]. Контролируемое обучение включает в себя многие алгоритмы:

- Алгоритмы классификации, например, искусственные нейронные сети (artificial neural

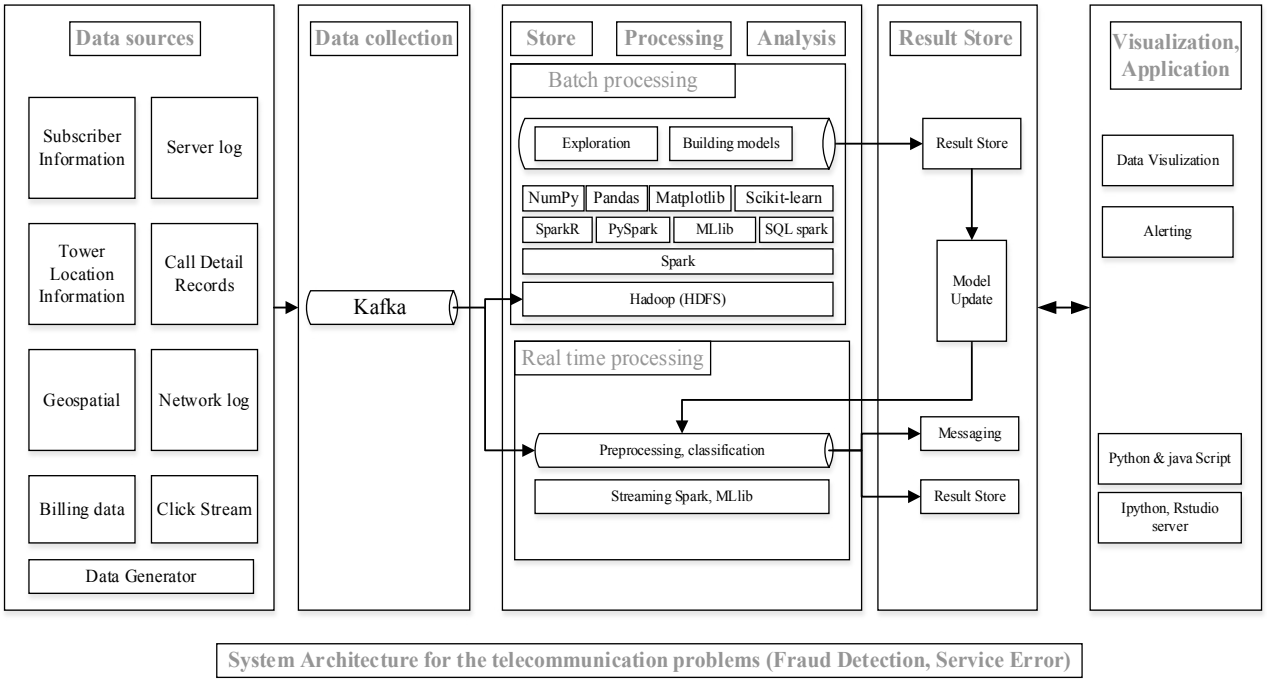

Рисунок 2. Архитектура системы обнаружения мошенничества в телекоммуникации

network, ANN), Метод k ближайших соседей (k-nearest neighbors algorithm, k-NN), Дерево принятия решений, логистическая регрессия (logistic regression, LR), Наивный байесовский классификатор (Naïve–Bayes, NB) и метод опорных векторов (support vector machine, SVM).

- Алгоритмы регрессии (Regression algorithms), например, Линейная регрессия (Linear regression), простая линейная регрессия и логистическая регрессия.

II. Без учителя (Unsupervised)

Методы обучения без учителя обнаруживают мошеннические в наборе немаркированных тестовых данных на основании предположения о том, что большинство случаев в наборе данных не является мошенничеством. В отличие от контролируемого метод, «без учителя» значит, что нет метки класса для построения модели. Основным преимуществом использования подхода без учителя является то, что оно не зависит на точной идентификации для данных метки, который часто не хватает или не существуют [26]. Существует два простых классических алгоритмов обучения без учителя:

- Алгоритмы кластеризации, такие как методы K-средних.
- Алгоритмы сокращения размерности, такие как метод главных компонент (principal component analysis, PCA)

III. Частичного обучения (Semi-supervised)

Частичное обучение лежит между обучением и без обучения, поскольку оно включает в себя небольшое количество маркированных образцов и большое количество немаркированных образцов. Основная цель подхода частичного обучения состоит в том, чтобы наблюдали подход, чтобы обучить классификатор с обеих маркированных и немаркированных данных [27]. Частичное обучение имеет большее преимущество по сравнению с контролируемым обучением, поскольку оно достигает более высокую производительность за счёт использования как маркированных, так и немаркированных данных, но с меньшим количеством маркированных случаев. Кроме того, частичного обучения также предоставляет вычислительные модели для обучения категории человека, где большинство входных самоочевидно немаркированный [28].

Представленная архитектура системы обнаружения мошенничества в телекоммуникации

На рисунке 2 представлена архитектура системы обнаружения ТКМ. Системы состоит из 5 подсистем: S1 Подсистема источников данных; S2 Подсистема сбора данных; S3 Подсистема обработки и анализа данных; S4 Подсистема хранения результатов обработки данных; S5 Подсистема визуализации результатов. Исходные данные представляются собой подробные записи о вы-

зовах клиентов. В практике такие данные сохраняются во внутренней базе данных телекоммуникационного предприятия, то есть они не находятся в общем доступе. Поэтому, для испытания была разработана подсистема генерации ПЗВ-логов клиентов на основе технологии Kafka. Kafka кластер состоит из множества брокеров для сбора данных, потребителей (Customers) и производителей (Producers). Потребители формируют подписку на определенные брокеры для сбора данных. В подсистеме S1 создано несколько производителей для генерации данных и в подсистеме S2 создано несколько потребителей данных. Эта модель позволяет эмулировать работу по сбору данных в режиме реального времени.

В подсистеме S3 процессы обработки данных выполняются в двух режимах: (1) пакетном режиме и (2) потоковом режиме. В пакетном режиме осуществляется построение модели обнаружения ТКМ на основе технологического стека Hadoop с использованием методов машинного обучения. Данные загружаются из Hadoop HDFS для обработки. Полученная модель сохраняется в подсистеме S4. В потоковом режиме осуществляется классификация ПМ на основе полученной в пакетном режиме модели. Также происходит обновление параметров модели на основе новых входных данных. В подсистеме S5 выполняется визуализация результатов в формате отчета «Прогнозирование ПМ».

Выводы

Таким образом, в данной статье представлены проблемы при реализации системы обнаружения мошенничества в телекоммуникации, классификации системы обнаружения мошенничества и также методы обнаружения мошенничества на основе методов анализа интеллектуального анализа данных. В результате работы предложена архитектура системы обнаружения мошенничества в телекоммуникации.

ЛИТЕРАТУРА

1. Hiyam A. E. T. Detecting Fraud in Cellular Telephone Networks (2010).
2. Claudio M. H. Subscription fraud prevention in telecommunications
3. using fuzzy rules and neural networks/ Claudio M. Held, Claudio A. Perez, and Pablo A. Este. 2001
4. Action Fraud. 2015. UK's national fraud and internet crime reporting
5. centre. 1 (2015), 1689–1699. DOI: http://dx.doi.org/10.1017/CBO9781107415324.004
6. CFCA. Communications fraud control association (cfca) announces results of worldwide telecom fraud survey// Commun. Fraud Control Assoc. (2013), 0–1
7. Gosset P. Classification, detection and prosecution of fraud in mobile networks/ P. Gosset and M. Hyland // Proceedings of ACTS Mobile Summit, Sorrento, Italy, June 1999.
8. Luis C. Fraud Management Systems in Telecommunications: a practical approach/ Luis C. Filipe Martins, António R., and Pedro C. (2005).
9. Fugee T. Applying manufacturing batch techniques to fraud detection with incomplete customer information/ Fugee T., Zhihong Z., and Wei J.// IIE Trans. 39, 6 (March 2007), 671–680. DOI: http://dx.doi.org/10.1080/07408170600897510
10. Mahuya G. Telecoms fraud/ Mahuya G.// Comput. Fraud Secur. 2010 — № 7- c. 14–17. DOI: http://dx.doi.org/10.1016/S1361–3723(10)70082–8
11. Yufeng K. Survey of fraud detection techniques / Yufeng K., Chang L., и Sirirat S. // Netw. Sens. Control 2004 IEEE Int. Conf. 2, 3 (2004), с. 749–754.
12. Asherry M. Security, Prevention and Detection of Cyber Crimes/ Asherry M. // Tumaini University Iringa University College. Cyber Crime. Prepared by Asherry Magalla (LL. M–ICT LAW-10919) Supervised by Dr. Puluru. (2013).
13. Rolf O. Internet Security: Firewalls and Beyond/ Rolf O.// Commun. ACM 40, 5 (May 1997), 92–102. DOI: http://dx.doi.org/10.1145/253769.253802
14. Orlando B. Applying User Signatures on Fraud Detection in Telecommunications Networks / Orlando B. и Carlos V. 2011 — с.286–299.
15. Mohammad B. Nature-Inspired Techniques in the Context of Fraud Detection. Mohammad B., Luigi B., Mohammed B., и Tim F.// IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev. 42, 6 (November 2012), 1273–1290. DOI: http://dx.doi.org/10.1109/TSMCC.2012.2215851
16. Guo T. A.O. Neural data mining for credit card fraud detection/ Guo T. A.O. and Gui-yang L.// July (2008), с. 12–15.
17. Sharon T. Claims auditing in automobile insurance: fraud detection and deterrence objectives/ Sharon T. and Pau F.// 2002- № 3 — с. 289–308.
18. John A. Data mining application for cyber credit-card fraud detection system/ John A. // In Lecture Notes in Engineering and Computer Science. 2013- с.1537–1542.
19. Hian C. K. Data Mining Applications in Healthcare/ Hian C. K. и Gerald T.// 2005–19, № 2 — с.64–72.
20. Saravanan P. Data Mining Approach For Subscription-Fraud Detection in Telecommunication Sector/ Saravanan P., Subramaniyaswamy V., Sivaramakrishnan N., Prakash M. A. And, и Arunkumar T.// 2014–7, № 11 с. 515–522.
21. Noor N. M.M. A Review on a Classification Framework for Supporting Decision Making in Crime Prevention/ Noor N. M.M., Hamid S. H.a, Mohemad R., Jalil M., и Hitam M. S.// J. Artif. Intell. (2015). DOI: http://dx.doi.org/10.3923/jai.2015.17.34
22. Jiawei H. Data Mining Concepts and Techniques/ Jiawei H., Micheline K., и Jian P.// In Jiawei HanMicheline Kamber & Jian Pei, eds. Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, 585–631. DOI: http://dx.doi.org/http://dx.doi.org/10.1016/B978–0–12–381479–1.00013–7
23. Sasirekha M. A Defense Mechanism for Credit Card Fraud Detection/ Sasirekha M.// Int. J. Cryptogr. Inf. Secur. 2–2012 — № 3 — с. 89–100.
24. DOI: http://dx.doi.org/10.5121/ijcis.2012.2308

25. Jyothsna V. A Review of Anomaly based IntrusionDetection Systems / Jyothsna V. и Prasad V. V. R.// 28–2011- № 7 — с. 26–35.
26. Brause R. Neural data mining for credit card fraud detection / Brause R., Langsdorf T., и Hepp M.// Tools with Artif. Intell. 1999. Proceedings. 11th IEEE Int. Conf. (1999), 103–106. DOI: http://dx.doi.org/10.1109/TAI.1999.809773
27. Bo S. Enhancing Security Using Mobility-Based Anomaly Detection in Cellular Mobile Networks/ Bo S., Fei Y., Kui W., Yang X.// 55, 2006- № 4 — с. 1385–1396.
28. Xu-ying L. Exploratory Undersampling for/ Xu-ying L., Jianxin W., and Zhi-hua Z.// 2012 — с.1–14.
29. Richard J. Unsupervised Profiling Methods for Fraud Detection/ Richard J. Bolton и David J.// Credit Scoring Credit Control. 2001- с.235–255.