

Factores que caracterizan el tráfico gris en compañías de telefonía móvil¹

Gina María Valladares Hernández ²

Alejandro Josué Calderón Torres

UNITEC, Tegcigalpa, Honduras.

(Recibido: Diciembre, 2014 y Aceptado: Junio, 2015)

Resumen:

Actualmente en Honduras existe una deficiencia en la detección de líneas telefónicas utilizadas para tráfico gris en redes GSM, debido a la variabilidad de los factores principales considerados y la rigidez de los métodos tradicionales de detección. El presente estudio tiene como propósito conocer factores característicos de líneas telefónicas utilizadas para tráfico gris en Honduras, con la finalidad de utilizar el conocimiento de dicho perfil para la detección de casos de fraude mediante una red neuronal con aprendizaje supervisado. Los resultados del estudio se presentan de forma separada para el enfoque cualitativo y cuantitativo, de forma que se comprendan los factores más significativos encontrados en cada análisis. Con la caracterización tanto del perfil de línea telefónica de tráfico gris como de línea telefónica legal, se procedió a construir un modelo de red neuronal con un 99% de efectividad de detección, y para efectos prácticos de aplicabilidad se describen las bases para la posterior construcción de un modelo de detección de tráfico gris.

Palabras Claves: Perceptrón multicapa, Red GSM, Red neuronal, Tráfico gris.

Abstract:

Honduras currently suffers a deficiency in detection of mobile phone lines used for bypass fraud in GSM networks, due to the variable nature of the main factors considered, and the inflexibility of traditional detection methods. This study aims to find determinant factors of telephone lines used for bypass fraud, in order to use this knowledge to detect these cases through the use of a neural network with supervised learning. The results of this study are presented separately for the qualitative and quantitative approach, for a better understanding of the most significant factors discovered. With the characterization of user profiles for both the bypass telephone line and the legal telephone line, the study proceeds to build a neural network model with 99% detection efficiency, and for applicability matters, the basic considerations that need to be taken into account are described for the subsequent construction of a bypass detection model.

Keywords: Multilayer perceptron, GSM network, Neural network, Bypass.

1. Introducción

En la investigación presentada se expone la problemática del tráfico gris en el rubro de la telefonía móvil en Honduras, con el propósito de comprender a profundidad los escenarios de fraude a través de éste método, y así poder exponer los puntos de enfoque en aras de minimizar las incidencias del mismo.

¹ Artículo ganador del **Primer Lugar** en la categoría de *Estudiantes de Postgrado* del Concurso de Investigación UNITEC-CEUTEC 2014.

² Autor para correspondencia. Email: ginamvh@gmail.com

1.1 Antecedentes del problema

De acuerdo a los datos de (CONATEL, 2011, p.4) “En Honduras existen más de siete millones de líneas de telefonía móvil, estas se distribuyen entre las tres compañías proveedoras de dichos servicios: Claro, Hondutel y CELTEL” Las tres compañías de telecomunicaciones en Honduras cuentan con el servicio de llamadas de voz internacionales que permiten la comunicación rápida y efectiva entre habitantes de regiones geográficas distintas.

Debido a la demanda del servicio de llamadas internacionales, éste representa importantes utilidades para los proveedores, dado que el precio por minuto de una de estas llamadas es mayor al de una llamada local, y las tarifas varían de acuerdo al país u operador destino de la llamada.

Un factor que afecta de forma negativa la percepción de utilidades por parte de las empresas de telecomunicaciones debido a las llamadas internacionales, es el llamado tráfico gris, el cual disfraza las llamadas internacionales como locales, evadiendo así los costos asociados y por ende, haciendo fraude al proveedor. El servicio de telefonía móvil tiene cobertura en los dieciocho departamentos del país, tanto en zonas rurales como urbanas, permitiendo acceder al servicio de llamadas internacionales desde cualquiera de estas zonas.

Tráfico gris en Honduras

El fraude en la telefonía móvil afecta en gran medida a las compañías proveedoras de este servicio a nivel global, debido a que en éstos casos se afecta de forma monetaria y también se ve afectada la imagen de la empresa, como consecuencia de verse comprometida la integridad y seguridad de la red de servicios. En la Figura 1 se muestra los involucrados en el transporte de llamadas de tráfico gris hacia empresas de telefonía móvil en Honduras.

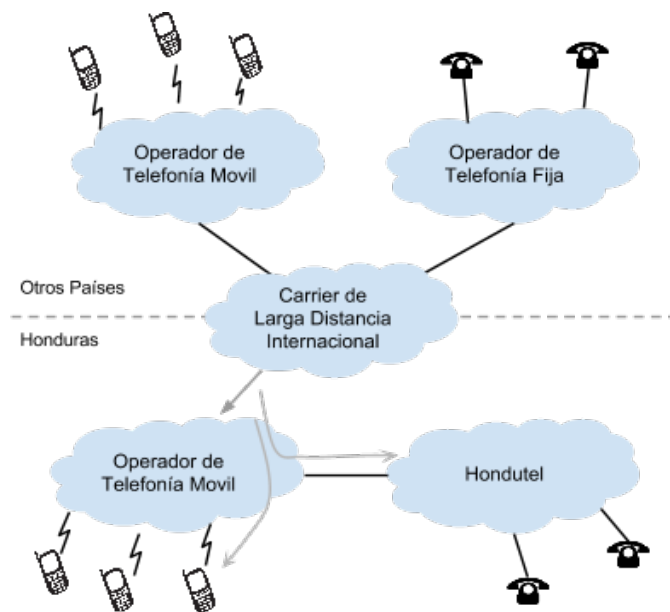


Figura 1. Flujo de tráfico gris mediante un operador de telefonía móvil

Para la estimación de la magnitud del tráfico gris, se debe analizar el tráfico internacional de entrada registrado en Honduras, ya que el tráfico gris sigue rutas no habituales para entrar al país, de modo que no es registrado por los operadores locales. Sin embargo, el mismo tráfico puede ser considerado como legal en otros países, de modo que en el origen del tráfico entrante a Honduras sí es registrado por los carriers y reportado ante los entes reguladores. De esta manera, una aproximación a la cantidad de tráfico gris corresponde a la diferencia entre el tráfico reportado por un país externo, versus el tráfico reportado por Honduras. En la Tabla 1 se muestra la información del Tribunal Superior de Cuentas al año 2005, tomando los Estados Unidos de América como ejemplo.

Tabla 1. Estimación de la magnitud del tráfico gris desde EUA (millones de min)

Tipo de Tráfico	2002	2003	2004	2005
Internacional de Entrada	325.4	404.2	513.4	611.4
Internacional desde los EUA	296.1	367.8	467.2	556.4
Registrado por la FCC (regulador EUA)	370.7	423.6	580.4	711.4
Pérdida (tráfico gris, minutos)	74.6	55.8	113.2	155.0
Pérdida (tráfico gris, en porcentaje)	20.1%	13.2%	19.5%	21.8%

Tal como lo establece el Tribunal Superior de Cuentas de Honduras (2005), se concluye que “Un volumen de tráfico gris equivalente al 20% de tráfico enviado desde EUA, significó pérdidas del orden de USD 35 millones anuales” (p.19). Estas pérdidas afectan tanto al operador estatal como a los operadores de telefonía móvil.

1.2 Formulación del problema

Existe una deficiencia en la detección de líneas telefónicas utilizadas para tráfico gris, dado que los factores principales estudiados son de carácter variable y los métodos tradicionales de detección carecen de la flexibilidad requerida para tal efecto.

1.3 Objetivos de la investigación

Objetivo general

Identificar factores significativos para la detección de líneas telefónicas utilizadas para tráfico gris mediante el análisis de líneas identificadas como fraudulentas para la detección de nuevos casos de fraude.

Objetivos específicos

- Conocer las características que definen el comportamiento de una línea telefónica legal.
- Analizar las características que definen el comportamiento de una línea telefónica que realiza tráfico gris.
- Determinar un mecanismo automatizado que permita diferenciar los perfiles previamente establecidos en base a comportamientos reconocidos como ilegales.

1.4 Hipótesis de investigación

Con el propósito de enmarcar el problema que se desea abordar, y para dar más claridad a los objetivos de la investigación, se formula a continuación la hipótesis:

Hi: Mediante la selección de los factores apropiados, la efectividad de detección de tráfico gris será mayor a 90%.

Ho: Mediante la selección de los factores apropiados, la efectividad de detección de tráfico gris será menor o igual a 90%.

1.5 Justificación

La detección de tráfico gris puede representar la diferencia entre percibir o dejar de percibir una importante suma de dinero para el proveedor de servicios de telefonía móvil, dado que la tarifa de llamadas internacionales es mayor a una llamada local, y el bloqueo de tráfico fraudulento permite que las líneas telefónicas deban recurrir al método tradicional de llamada internacional para poder efectuar la misma, lo que resulta en una mayor percepción de ingresos debido a este servicio.

2. Marco Teórico

La diversidad de tipos de fraude se relaciona estrechamente con los métodos de detección a utilizar, dado que se requiere conocer factores determinantes y patrones de conducta, tanto de uso legal como ilegal; como punto de partida para inferir sobre la forma en la que se identificará y subsecuentemente bloqueará el uso fraudulento del servicio de llamadas. Los métodos actuales utilizados son poco flexibles, debido a que son basados en umbrales de tolerancia, los cuales se fijan de forma estática (W. Baluja & S. Llanes, 2005, p.50). Este enfoque es poco adaptativo y no considera márgenes de variación para adecuarse a escenarios con cambios condicionales, por lo que tienden a desfasarse respecto al comportamiento real del tráfico de llamadas.

2.1 Fraude en telefonía móvil

Con el surgimiento de nuevas tecnologías de comunicaciones, también se crean nuevas formas de fraude. Este es el caso del fraude llamado “SIM Box Bypass” el cual se basa en la tecnología de Voz Sobre IP (VoIP), donde un individuo o empresa utiliza cientos de líneas en un aparato especializado para evitar cargos internacionales en sus llamadas. Las investigaciones existentes sobre fraudes se enfocan generalmente en los fraudes por suscripciones, los cuales son los tipos dominantes de fraude a nivel global en la industria de las telecomunicaciones. Sin embargo, el tipo de fraude llamado SIM Box Bypass, se ha convertido en un amenaza que representa un reto para numerosas compañías de telefonía móvil (Elmi, Ibrahim, Sallehuddin, 2013, p.576). En Honduras y en otros países de Centro y Sur América, éste tipo de fraude es muy común.

El éxito del fraude por SIM box depende de la obtención de grandes cantidades de tarjetas SIM, por lo que los efectos varían dependiendo de las regulaciones correspondientes a cada país. En países donde las tarjetas SIM no registradas no son permitidas y las leyes de gobierno reconocen a los dispositivos SIM box como equipo ilegal, el efecto es mucho menor comparado con los países donde la obtención de tarjetas SIM por clientes es barata e incluso puede llegar a ser gratuita, y dónde las leyes del gobierno no prohíben usuarios

no registrados. Según (CONATEL, 2012) “En Honduras, la ley requiere que todo usuario de telefonía móvil se encuentre registrado” (p.13).

El fraude por SIM box toma lugar cuando individuos u organizaciones adquieren miles de tarjetas SIM, ofreciendo llamadas gratuitas o de bajo costo a líneas móviles de operadores extranjeros. Las tarjetas SIM son utilizadas para canalizar llamadas nacionales o internacionales fuera del operador de red móvil y entregarlas como llamadas locales (Elmi, Ibrahim, Sallehuddin, 2013, p.576). Esta práctica fraudulenta motiva a los suscriptores ofreciéndoles un servicio de bajo costo por el servicio de llamadas internacionales, las cuales, dependiendo del destino, pueden llegar a ser muy costosas.

2.2 Técnicas de inteligencia artificial

La naturaleza dinámica y cambiante del fraude en las telecomunicaciones ha llevado a las empresas a abordar el problema utilizando diversas técnicas avanzadas de análisis y reconocimiento de información. Estas técnicas se basan en el estudio de los registros generados por la actividad de los suscriptores de telefonía y en la detección de patrones anormales en su comportamiento.

Debido a que los modelos tradicionales basados en reglas estáticas son muy rígidos, se han creado métodos de detección más flexibles basados en técnicas de inteligencia artificial. Las técnicas más utilizadas para este fin se basan en conceptos como las redes neuronales artificiales, la minería de datos y la lógica difusa. El presente estudio está orientado a la utilización de un modelo basado en redes neuronales artificiales, de forma más específica, una red perceptrón multicapa.

Redes neuronales artificiales

Una de las técnicas de inteligencia artificial más utilizadas en la detección de patrones son las redes neuronales artificiales. La definición de red neuronal artificial propuesta por Graupe (2007) es la siguiente:

Las redes neuronales artificiales son, como su nombre lo indica, redes computacionales que intentan simular, de forma básica, las redes de células nerviosas (neuronas) del sistema biológico central (humano o animal) (...) permiten el uso de operaciones de cálculo muy simples (adiciones, multiplicación y elementos lógicos fundamentales) para resolver problemas matemáticos complejos, problemas no lineales o problemas estocásticos (p.1).

Las redes neuronales artificiales sirven como apoyo en la resolución de problemas complejos, sin embargo, no todo tipo de problema puede ser adaptado a una solución con redes neuronales. Heaton (2008) afirma:

Las redes neuronales a menudo no son adecuadas para los problemas en los que se debe saber exactamente cómo se obtuvo la solución. Una red neuronal puede ser muy útil para resolver el problema para la que fue entrenada, pero la red neuronal no puede explicar su razonamiento. La red neuronal sabe algo porque fue entrenada para saberlo. La red neuronal no puede explicar cómo se siguió una serie de pasos para obtener la respuesta (p.43).

Sin embargo, a pesar de que las redes neuronales no son adaptables a cualquier tipo de problema, el reconocimiento de patrones es una de las áreas en las que más se utiliza este enfoque:

El reconocimiento de patrones es quizá el uso más común de las redes neuronales. Para este tipo de problema, se presenta a la red neuronal un patrón. Este puede ser una imagen, un sonido o cualquier otro tipo de dato. La red neuronal entonces trata de determinar si el dato de entrada coincide con el patrón que ha sido entrenada para reconocer (Heaton, 2008, p.44).

Para la resolución de problemas de reconocimiento de patrones utilizando redes neuronales, se debe entrenar la red previamente para reconocer el patrón de interés. Es decir, la red neuronal primero debe aprender a reconocer el patrón. Según (Yeung, Cloete, Shi y Ng, 2010) “En el contexto de redes neuronales, el aprendizaje está definido como un proceso por el cual los parámetros libres de una red neuronal se adaptan a través de un proceso continuo de la estimulación por el medio ambiente” (p.5).

3. Metodología

En el presente capítulo, se detalla el esquema de diseño del estudio, describiendo los enfoques y métodos que conforman la metodología de investigación utilizada, con el objetivo de contar con una guía a seguir para realizar el trabajo investigativo y analítico.

3.1 Enfoque y Métodos

El estudio se enmarca dentro de un enfoque mixto, caracterizado por su composición variada de criterios de investigación, procesos sistemáticos y empíricos en donde resulta necesario analizar de forma estructurada los datos recolectados; de manera que se pueda buscar comportamientos y relaciones entre elementos, para lograr demostrar la hipótesis propuesta, mediante la integración y discusión de los datos cualitativos y cuantitativos recolectados.

El enfoque utilizado en la investigación es de carácter mixto pero con dominancia cuantitativa, ya que ésta es complementada por el enfoque cualitativo, el cual aporta valiosos puntos de vista para el estudio. El enfoque cualitativo marca la pauta sobre la aproximación a tomar para el estudio cuantitativo, ya que es en el estudio cualitativo donde se analiza la teoría que fundamenta la investigación, y luego se selecciona a un grupo de expertos en el área de aseguramiento de ingresos, específicamente en el estudio de tráfico gris, para ser entrevistados en relación a factores determinantes, caracterización de perfiles, métodos existentes, y otros aspectos. Las calificaciones de los expertos entrevistados incluyen un grado de especialización en sistemas computacionales, múltiples cursos de sistemas de seguridad de la información y técnicas de prevención de fraude informático, experiencia laboral comprobada por un mínimo de 5 años en el rubro de las telecomunicaciones, con sólidos conocimientos de la arquitectura de red móvil y el flujo de control de las llamadas de voz.

En el enfoque cuantitativo se recolectan datos para probar la hipótesis propuesta, en base a mediciones numéricas y análisis con métodos de aprendizaje supervisado, con los cuales se establecen patrones de comportamiento y se predicen futuros casos positivos o negativos de utilización de tráfico gris en líneas móviles. El enfoque cuantitativo en este estudio es de carácter observacional, ya que las variables independientes no se manipulan en ningún aspecto de forma deliberada, al contrario, las acciones se limitan a observar los datos en su comportamiento natural. El estudio es transversal, la recolección de información toma lugar en una única ocasión.

3.2 Diseño de la investigación

La clasificación del estudio de acuerdo a la intervención de los investigadores puede considerarse como cuasi experimental, debido a que los factores de estudio no serán elegidos al azar, y son de carácter externo a este estudio. Los métodos observacionales se llevan a cabo en situaciones no controladas, por las que no es factible realizar el presente estudio con ese tipo de aproximación por parte de los investigadores.

El proceso de investigación se desarrolla en las siguientes fases:

Fase 1: Identificación de características utilizadas por expertos para la caracterización de líneas telefónicas móviles utilizadas para tráfico gris.

Fase 2: Identificación de técnicas comúnmente utilizadas para la detección de líneas telefónicas móviles que cursan tráfico gris.

Fase 3: Indagación sobre el comportamiento promedio de un usuario regular (no fraudulento) en una empresa de telefonía móvil.

Fase 4: Recopilación e investigación de casos históricos reales de tráfico gris y casos de tráfico legal en una empresa de telefonía móvil de Honduras.

Fase 5: Construcción de un modelo para la detección de tráfico gris, basado en redes neuronales (perceptrón multicapa).

Fase 6: Entrenamiento de la red neuronal con los casos históricos de fraude y legales previamente obtenidos.

Fase 7: Ejecución de las pruebas de detección de casos de tráfico gris, utilizando la red neuronal entrenada en la fase previa.

Fase 8: Análisis de los resultados y conclusiones del estudio.

3.3 Población y muestra

Como parte de la población elegida para el presente estudio, se consideran los registros de casos donde se ha identificado positivamente la utilización de las líneas telefónicas para tráfico gris, es decir, se recopilaron datos históricos reales sobre casos positivos de fraude de este tipo. Para efectos de muestreo y de comprobación de las herramientas utilizadas, se utiliza una población que reproduzca de la mejor manera los factores significativos, y por ello debe elegirse con un muestreo intencional-discrecional, caracterizado por el esfuerzo deliberado de obtener muestras representativas, incluyendo casos típicos de tráfico gris.

3.4 Técnicas e instrumentos aplicados

Los instrumentos utilizados en la investigación para la recopilación de los datos necesarios para el análisis son la entrevista y las bases de datos. Por medio de una entrevista a personal experto del área en una empresa de telefonía móvil de Honduras se logró conocer de forma más cercana los criterios aplicados para el análisis de casos de tráfico gris.

El estudio tiene como fuentes de información las bases de datos en las que se recopilan registros de llamadas mediante un procesamiento de archivos de registros de llamadas. Con la centralización de dicha información, se procede a tomar una muestra de la base total para hacer un análisis de comportamiento sobre ella, y así poder proponer una base de comportamiento legal. No obstante, para efectos de un análisis objetivo y

fundamentado, se requiere contar con los aportes de los especialistas en el área, por lo que el estudio considera como una técnica importante la entrevista a personas expertas en el área de aseguramiento de ingresos, como sustento al enfoque de análisis utilizado sobre la técnica de recolección de datos en la base de datos.

Para enriquecer la investigación, se utilizaron diversas fuentes de información, dado que existen numerosas investigaciones, libros de texto, revistas científicas y expertos en la materia que aportaron significativamente a comprender mejor el ambiente de estudio, en este caso, los servicios de telefonía móvil, tráfico gris y redes neuronales.

4. Resultados y Análisis

4.1 Análisis cualitativo

Los resultados del análisis cualitativo se explicarán a detalle en esta sección. Se ha recopilado información de expertos en el área de aseguramiento de ingresos en telefonía móvil, específicamente con experiencia en detección de tráfico gris, participando en actividades de recolección, procesamiento y análisis de datos y uso de herramientas especializadas, lo cual da una mayor amplitud al estudio en las áreas de conocimiento relacionadas a procesos de detección de tráfico gris.

La herramienta utilizada para la recopilación de datos cualitativos es la encuesta, la cual en este caso específico consta de 10 preguntas abiertas, para dar una mayor amplitud a la experiencia que puede compartir el experto al momento de ser entrevistado. Se analizarán las respuestas a cada una de las preguntas de la entrevista, para concluir destacando los hallazgos principales de la investigación.

El fundamento teórico indica que existen múltiples técnicas conocidas utilizadas por los individuos que realizan tráfico gris, y los expertos entrevistados señalan que las más utilizadas en Honduras varían entre dispositivos especializados tales como los SIM Box, simulación de comportamiento humano, automatización de actividades originadoras de tráfico, adquisición masiva y rotación de tarjetas SIM por parte de un mismo suscriptor, e incluso suplantación de líneas de usuarios genuinos. La simulación del comportamiento humano está vinculada a la automatización de las actividades que originan tráfico, enmarcadas en el contexto de usuario de telefonía móvil. Las actividades pueden ser recargas electrónicas de saldo, consulta de saldo vía SMS o USSD, llamadas de voz, navegación en red de datos móvil, etc. La Figura 2 ilustra las técnicas que resaltan los expertos entrevistados.



Figura 2. Técnicas relacionadas al uso de tráfico gris en telefonía móvil

En aras de modelar un perfil de usuario de una línea telefónica utilizada para tráfico gris, se consulta a los expertos sobre su experiencia en la caracterización de dicho perfil. Por medio de la entrevista, es notable que existen múltiples factores a considerar, que varían dependiendo de la amplitud del estudio a realizar y el enfoque hacia el tipo de red; en este caso, una red de telefonía móvil. Las características de usuario que resaltan en la entrevista son ilustradas en la Figura 3, y son descritas a continuación:

- Alta cantidad de llamadas: específicamente, puede ser que se den muchas llamadas en un corto periodo de tiempo, que difiere del comportamiento regular de un usuario no fraudulento.
- Mayor duración de llamadas: Este puede significar un factor irregular, cuando ocurren duraciones sustanciosas de llamadas salientes.
- Relación desproporcionada de llamadas entrantes versus llamadas salientes: la línea telefónica utilizada para tráfico gris presenta un patrón de consumo anómalo entre su tráfico de llamadas entrantes y salientes. La variación entre cada tipo de llamada suele ser mucho mayor cuando la línea se utiliza para éste tipo de fraude, especialmente notable en el aumento significativo de llamadas salientes.
- Desvío significativo de una actividad específica: Un usuario regular varía sus actividades entre los diferentes productos que ofrece el operador de telefonía móvil, por ejemplo, llamadas de voz, mensajes de texto (SMS), navegación en red de datos Móvil o Internet, llamadas a servicios USSD, llamadas a códigos cortos de servicios como consulta de saldo, pagos vía operador móvil, etc. Sin embargo, una línea telefónica utilizada para tráfico gris tiene un comportamiento que genera una cantidad irregular de transacciones de un solo tipo en un corto periodo de tiempo, dada la repetición de un patrón específico de comportamiento por tiempo prolongado.
- Ubicación típicamente invariable: la línea telefónica utilizada para tráfico gris tiende a ubicarse en la misma zona geográfica la mayor parte de tiempo, como consecuencia de utilizarse en SIM Box. El estudio de este comportamiento se realiza con la obtención de la ubicación de usuario que provee los registros de la llamada, y es notable la poca movilidad que tienen las líneas fraudulentas.



Figura 3. Características de líneas telefónicas fraudulentas

Como complemento para el modelamiento del perfil de línea telefónica utilizada para tráfico gris, también se consulta a los expertos sobre la caracterización de un perfil de una línea telefónica legal. Las características de usuario legal que resaltan en la entrevista son ilustradas en la Figura 4, y se describen a continuación:

- Comportamiento congruente con el promedio de uso de otros usuarios de la red: el comportamiento general de los usuarios legales se mantiene cerca de los márgenes promedio para todos los servicios: llamadas de voz, mensajes de texto, etc., por lo que los casos considerados legales son aquellos que se aproximan aceptablemente al comportamiento global de los usuarios de la red.
- Movilidad constante: El usuario legal no se mantiene fijo en una única ubicación geográfica, por lo contrario, tiene hábitos de vida que lo obligan a desplazarse por diferentes áreas, lo que se traduce al registro de múltiples zonas de cobertura en la generación de sus registros de llamadas.
- Variedad de transacciones: El usuario regular varía sus actividades entre los diferentes productos que ofrece el operador de telefonía móvil, por ejemplo, llamadas de voz, mensajes de texto (SMS), navegación en red de datos Móvil o Internet, llamadas a servicios USSD, llamadas a códigos cortos de servicios como consulta de saldo, pagos vía operador móvil, etc.
- Relación proporcionada de llamadas entrantes versus salientes: El usuario legal tiene una relación aceptable de llamadas entrantes versus llamadas salientes, por ejemplo, un 20% de las llamadas totales pueden ser llamadas entrantes.
- Llamadas de voz y mensajes de texto recurrentes con los mismos destinatarios: Aun cuando el número de llamadas salientes puede ser muy alta, el usuario legal no llama a una cantidad muy grande de números telefónicos distintos, por lo que se ven llamadas y mensajes de texto intercambiados regularmente con un número reducido de destinatarios.



Figura 4. Características de líneas telefónicas legales

Para efectos de la presente investigación, es importante conocer las características descritas anteriormente: características de líneas telefónicas fraudulentas y de líneas telefónicas legales. Tomando como base el aporte de los expertos mediante la entrevista, puede caracterizarse el perfil de usuario que se busca, considerando además que el usuario fraudulento siempre se encuentra en búsqueda de maneras de evadir la detección de su comportamiento ilegal. Los expertos indican que los perfiles de usuario que realizan tráfico gris tienden a ser muy dinámicos, cambian su comportamiento constantemente para evitar la detección. Las líneas telefónicas de usuarios fraudulentos buscan la manera de ser detectados de forma tardía, llegando incluso a simular comportamiento humano para evitar ser detectados.

Las principales dificultades en el trabajo de detección de fraude, de acuerdo al criterio de los expertos, son de diversa naturaleza. Por un lado, la disponibilidad de tiempo y esfuerzo requerido para la investigación resulta bastante demandante, debido a la limitación de

recursos en las oficinas de aseguramiento de ingresos. La investigación de tendencias actuales de detección y noticias sobre hallazgos en dicha área requieren dedicación y tiempo, para mantener al día los conocimientos y permanecer en constante aprendizaje de nuevos comportamientos fraudulentos. Los individuos que realizan el fraude tienden a encontrar nuevas formas ingeniosas para despistar a los algoritmos de detección, por ejemplo, reciclando las tarjetas SIM periódicamente, o cambiándose de ubicación en cortos periodos de tiempo.

El conocimiento de las técnicas de detección utilizadas en otros países resulta muy importante para la constante mejora de las técnicas empleadas, y los entrevistados indican que la comunicación entre operadoras hermanas es constante, y que los operadores de telefonía móvil monitorean frecuentemente las noticias del foro global de fraude de la GSMA, para mantenerse informados. Entre las técnicas de detección utilizadas en países diferentes a Honduras, existen los proveedores de pruebas de llamadas internacionales, y el análisis de Call Detail Records, el cual consiste en el procesamiento de los archivos de registro de datos para cada llamada efectuada en la red. En los archivos CDRs se encuentra información detallada sobre aspectos de duración, emisor y receptor de la llamada, y múltiples campos adicionales. Los entrevistados indican además que dichas técnicas pueden ser muy efectivas, pero no para la detección inmediata de casos de fraude, dado que requieren procesamiento de la información recolectada en cada caso.

La adaptación de técnicas de detección de tráfico gris utilizadas en otros países puede o no ser adaptable al caso específico de la telefonía móvil en Honduras. Por un lado, un par de expertos entrevistados indican que las técnicas empleadas en Honduras han sido mucho más complejas que las técnicas compartidas por otros países, y que otros métodos pueden resultar menos efectivos que el utilizado actualmente, mediante el análisis de CDRs. Sin embargo, otro experto entrevistado indica que sí podría ser conveniente la utilización de técnicas alternas, dada la complejidad y tamaño mayor de las operaciones de países con mayor cobertura geográfica, siempre adecuando la herramienta a los casos específicos de Honduras.

Respecto a la disposición de las empresas de telefonía móvil a invertir en nuevas soluciones para la detección de tráfico gris, los expertos entrevistados coinciden en que estos sí están dispuestos a invertir, debido a que tienen consciencia del riesgo al que se exponen, debido a las pérdidas en las que incurren por casos de fraude. Por tanto, en la medida en que se impacten sus ganancias, invertirán para mitigar dicha pérdida. Los expertos indican que el proceso de aseguramiento de ingresos es primordial en el área de las telecomunicaciones para blindar y evitar fugas innecesarias de ingresos, siempre considerando que es fundamental tener una relación de costo – beneficio. Mientras el fraude se mantenga en un margen aceptable, las empresas se limitan a continuar invirtiendo en nuevas tecnologías de detección de tráfico gris y casos de fraude.

Para el presente estudio es muy importante conocer la confianza que los expertos tienen en métodos estadísticos avanzados, y para conocer su opinión, se les entrevistó respecto a la aplicabilidad de métodos de reconocimiento de patrones, los cuales son fundamentales para los métodos de aprendizaje asistido. Los expertos entrevistados coinciden sin excepción en que se puede mejorar la efectividad de detección de métodos tradicionales, aplicando técnicas basadas en reconocimiento de patrones, dado que estas técnicas han probado ser altamente efectivas, y recalcan especialmente que cuando existe una detección temprana es bastante favorable, ya que la detección oportuna es sumamente relevante al medir efectividad de una herramienta de detección. Los expertos indican que,

con la tecnología actual, estas técnicas de reconocimiento de patrones pueden aplicar procedimientos computarizados complejos, mejorando así la efectividad de detección.

Concluyendo; en base a las respuestas brindadas en la entrevista a expertos, se logran resaltar factores fundamentales a considerar para la caracterización de un perfil de tráfico gris, ya sea fraudulento o legal según lo indican los expertos en la materia. Debido a la delimitación del alcance de la presente investigación, se estudiarán 5 variables importantes de acuerdo al criterio de expertos, resumidas en la Figura 5.



Figura 5. Opinión de expertos entrevistados versus variables de estudio

4.2 Análisis cuantitativo

Los criterios que fueron tomados en cuenta para llevar a cabo el análisis cuantitativo se basan en los resultados obtenidos a partir del análisis cualitativo, el cual contempla la opinión de expertos en el área de telefonía móvil con experiencia comprobada en el tema de estudio. La información obtenida de los expertos a través de entrevistas contribuyó de manera significativa a la selección de factores determinantes en la detección de fraude y a la caracterización de los perfiles de uso.

La principal fuente de datos para el análisis cuantitativo fue extraída a partir de la base de datos existente del operador de telefonía móvil, la cual almacena datos extraídos de los archivos CDR. Cuando un suscriptor realiza una llamada por medio del operador de telefonía móvil, se genera un registro con información completa de la llamada, incluyendo detalles como el número del suscriptor, el número de destino, la fecha y hora de la llamada, la duración de la llamada, el identificador de las celdas de telefonía celular y otra gran cantidad de detalles. Estos datos registran el comportamiento de un suscriptor a lo largo del tiempo; por tanto, representan una fuente muy importante para este tipo de estudio.

El conjunto de datos utilizado para el presente análisis está formado por los datos de 1,473,510 llamadas realizadas por 6,031 suscriptores de telefonía móvil a lo largo de 2 meses calendario. De este total de suscriptores, 2,210 son casos reconocidos de fraude y 3,821 son suscriptores legales.

Los casos fueron almacenados en una base de datos MySQL y posteriormente fueron procesados para calcular las variables necesarias para el estudio. Estas variables son el promedio de llamadas que un suscriptor realiza por día, el promedio de distintos números destino a los que un suscriptor llama por día, el promedio de segundos que un suscriptor llama por día, el promedio de duración de las llamadas de un suscriptor y la razón de la cantidad de números distintos a los que llama un suscriptor sobre el total de las llamadas que realiza.

Construcción del modelo

El análisis de datos está basado en la utilización de un modelo de red neuronal artificial para la clasificación automática de casos de fraude y casos legales. De forma más específica, se utilizó el modelo de red perceptrón multicapa, el cual consta de una capa

de neuronas de entrada, una o varias capas de neuronas ocultas y una capa de neuronas de salida.

Para construir el perceptrón multicapa se utilizó el módulo de redes neuronales de la herramienta IBM SPSS Statistics versión 22. Los 6,031 casos de estudio fueron divididos aleatoriamente en tres particiones de la siguiente forma:

- Partición de entrenamiento: Es el conjunto de casos utilizados para entrenar la red neuronal. Para el presente estudio se utilizó el 50% de los casos para el entrenamiento de la red.
- Partición de prueba: Es un conjunto independiente de casos que se reservan para realizar un seguimiento de los errores durante el entrenamiento, con la finalidad de evitar un exceso de entrenamiento (obtener una red muy estricta). Para el presente estudio se utilizó el 20% de los casos como muestra de prueba.
- Partición de reserva: Es un conjunto independiente de casos que se utilizan para evaluar la red neuronal final. Esta evaluación ofrece resultados más confiables ya que estos casos no son utilizados durante el entrenamiento de la red. Para el presente estudio se utilizó el 30% restante de los casos para la evaluación final.

Utilizando como referencia el estudio realizado por Elmi et al. (2013), se utilizó una arquitectura de red con dos capas ocultas y la función sigmoide como función de activación.

Para efectos de simplificación, en la elaboración del conjunto de datos se consideraron los siguientes nombres cortos para las variables de estudio:

Tabla 2. Nombres de variables utilizadas en el conjunto de datos

Nombre de Variable	Nombre Corto
Segundos Promedio de Duración de la Llamada	duracion_llamada
Destinatarios por Llamadas	numeros_llamadas
Total de Segundos	segundos_dia
Total de Llamadas	llamadas_dia
Total de Números Distintos	numeros_dia
Detección de Tráfico Gris	caso_bypass

Resultados

En esta sección se presentan los resultados obtenidos durante la fase de análisis, analizando el modelo con todas las variables independientes en conjunto y la capacidad del modelo final de clasificar correctamente los casos de fraude y los casos legales.

Modelo final

En esta sección se presenta el modelo de perceptrón multicapa construido utilizando todas las variables de estudio en conjunto. Este modelo permite una detección más realista ya que al utilizar varias variables se pueden detectar patrones de comportamiento más complejos. La arquitectura de red utilizada para el presente análisis se muestra en la Figura 6.

El perceptrón multicapa presenta cinco entradas (una para cada variable independiente) las cuales están conectadas a las cuatro neuronas de la primera capa oculta. Esta capa está conectada a una segunda capa oculta que contiene tres neuronas, las cuales a su vez están conectadas a las dos salidas de clasificación. Debido a que el aprendizaje es supervisado, el modelo también presenta tres unidades de sesgo, las cuales realizan los ajustes necesarios durante el proceso de aprendizaje.

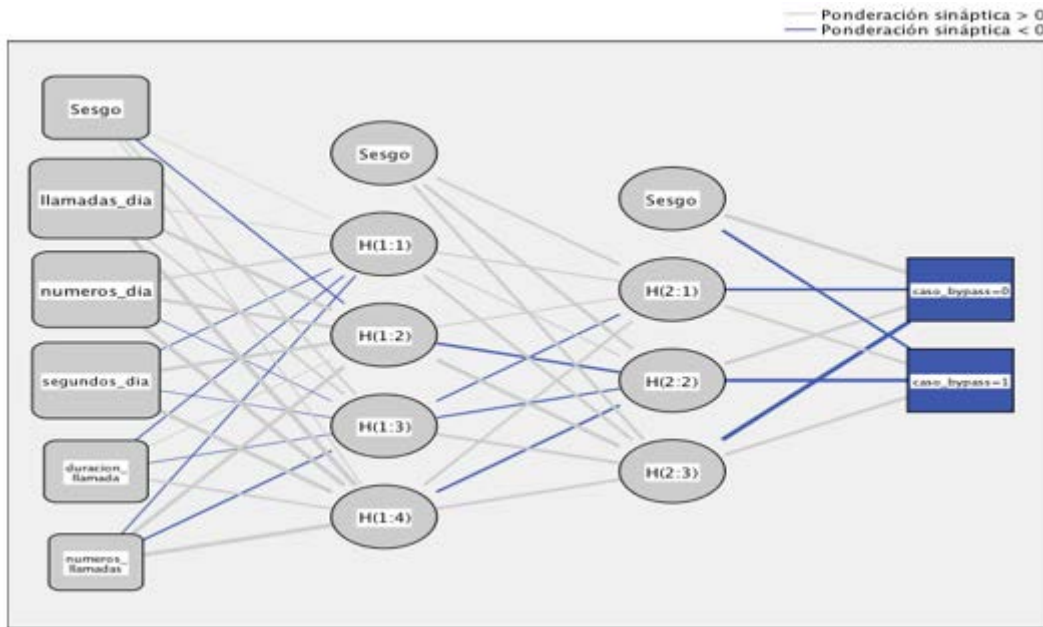


Figura 6. Arquitectura de red del modelo final

Tabla 3. Clasificación modelo final

Ejemplo	Observado	Pronosticado			
		Legal	Bypass	Datos Reales	Porcentaje correcto
Entrenamiento	Legal	1890	0	1890	100.0%
	Bypass	1	1107	1108	99.9%
	% Global	63.1%	36.9%	100%	100.0%
Pruebas	Legal	742	1	743	99.9%
	Bypass	1	455	456	99.8%
	% Global	62.0%	38.0%	100%	99.8%
Reserva	Legal	1187	1	1188	99.9%
	Bypass	1	645	646	99.8%
	% Global	64.8%	35.2%	100%	99.9%

La Tabla 3 muestra los resultados obtenidos en la etapa de clasificación, después de haber entrenado la red. Se puede observar que al utilizar la muestra de reserva el modelo final clasificó incorrectamente sólo 1 falso positivo y 1 falso negativo, lo cual representa un 99.9% de efectividad de detección. Este resultado comprueba la validez de la hipótesis de investigación, ya que utilizando los factores seleccionados se logró crear un modelo de clasificación con una efectividad de detección mayor al 90%.

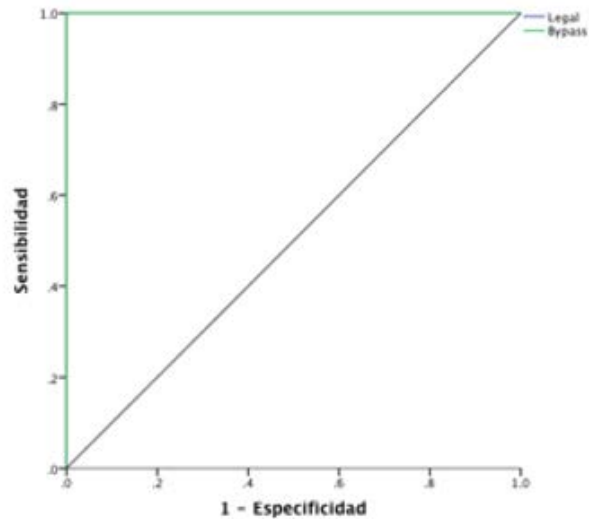


Figura 7. Curva COR del modelo final

El valor del área bajo la curva COR para el modelo de clasificación final es de aproximadamente 1.0, lo que categoriza al modelo como “excelente”, ilustrado en la Figura 7.

Para finalizar con el análisis, se muestra en la Figura 8 el gráfico de importancia normalizada de las variables independientes. El valor porcentual de cada variable es determinado en base a la importancia que toma cada una al momento de comparación con las muestras de entrenamiento y comprobación combinadas. La importancia normalizada indica el nivel de participación que tiene cada una de las variables independientes en la forma en que la red clasifica los casos. Se puede observar que las tres variables independientes que ejercen mayor influencia en el modelo final son el total de llamadas, el total de segundos y el total de números distintos. Para futuros estudios, se puede considerar el reemplazo de las variables menos significativas por otras variables que se consideren de interés.

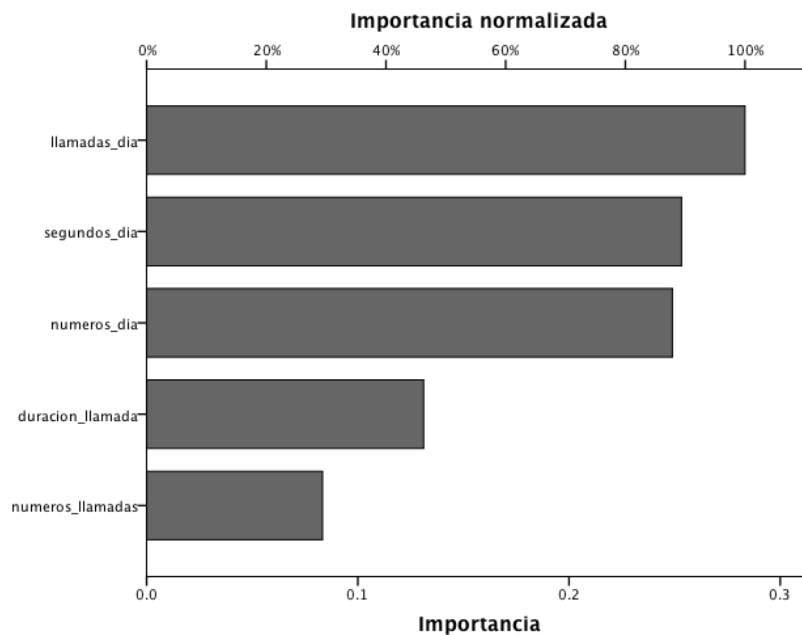


Figura 8. Importancia de variables independientes

5. Conclusiones

- En base a la entrevista realizada a expertos, se concluye que las características que definen una línea telefónica legal son el cumplimiento del perfil promedio de la red, la movilidad constante del usuario, la variedad de transacciones, una relación proporcionada de llamadas entrantes versus salientes, y la ocurrencia recurrente de llamadas hacia un grupo específico de destinatarios.
- De acuerdo a la experiencia compartida por los expertos mediante la entrevista, se concluye que las características que definen una línea telefónica utilizada para tráfico gris son una alta cantidad de llamadas, una alta duración de llamadas, la relación desproporcionada de llamadas entrantes versus salientes, la ejecución de una actividad específica de forma recurrente, y la ubicación invariable del usuario.
- Basados en los resultados de la investigación y del análisis cualitativo y cuantitativo del estudio, se concluye que los mecanismos de reconocimiento de patrones resultan efectivos como apoyo a la detección de líneas telefónicas utilizadas para tráfico gris. En el caso específico de la presente investigación, el perceptrón multicapa demostró una alta efectividad en la clasificación de líneas fraudulentas y líneas legales en base al estudio de su comportamiento.
- Mediante el análisis de casos previamente identificados como fraudulentos y la opinión de expertos, se logró identificar algunos de los factores determinantes de una línea telefónica utilizada para tráfico gris, y se construyó un modelo que puede ser utilizado posteriormente para la detección de nuevos casos de fraude.

5.1 Recomendaciones

- Para estudios futuros que pretendan modelar un perfil de usuario, se recomienda considerar factores adicionales de comportamiento, ya que en los archivos CDRs se registra una amplia variedad de campos que pueden ser utilizados para mayor precisión. La inclusión de más características puede llevar a una mayor efectividad, siempre analizando la relación de los factores a considerarse y su relevancia en el estudio.
- Para efectos prácticos, se recomienda aplicar al modelo conjuntos de datos más extensos, tanto en la cantidad de registros como en el rango de tiempo de la muestra. De esta forma, el modelo será capaz de aprender nuevos comportamientos de los suscriptores estudiados para mejorar la detección de cada perfil.
- Al utilizar este modelo en un ambiente de producción, se recomienda la revisión periódica del mismo para mantener los perfiles actualizados con respecto a nuevos comportamientos, a manera de mejorar continuamente el modelo para efectos de mantener la efectividad deseada.
- Aun cuando la automatización de la detección aporta mucho valor, es recomendable siempre referirse a la opinión de expertos que aporten mayor visibilidad a cada escenario, que propongan nuevas ideas y soluciones, para mejorar los métodos utilizados.
- Para Honduras, es altamente recomendable realizar una investigación de este mismo tipo aplicada al caso de telefonía fija, específicamente en la empresa

Hondureña de Telecomunicaciones, HONDUTEL, en el cual las pérdidas por tráfico gris pueden ser millonarias debido a la masividad del tráfico cursado.

Bibliografía

- CONATEL (2011). Informe de Actividades y Logros Realizadas por parte de CONATEL durante el año 2011.
- CONATEL (2012). Diario La Gaceta, Resolución NR002/12
- Elmi, A. H., Ibrahim, S., & Sallehuddin, R. (2013). Detecting SIM Box Fraud Using Neural Network. Lecture Notes in Electrical Engineering, p. 575.
- Galvan I. (2004). Redes de Neuronas Artificiales: Un Enfoque Práctico. Madrid: Pearson.
- Graupe D. (2007). Principles of Artificial Neural Networks. (2 ed.). Chicago, USA: World Scientific Publishing.
- GSMA (2013). The Mobile Economy
- Heaton J. (2008). Introduction to Neural Networks with Java. (2 ed.). St. Louis, USA: Heaton Research.
- Tribunal Superior de Cuentas (2005). Investigación Técnica Practicada al Control del Tráfico Gris en la Empresa Hondureña de Telecomunicaciones.
- W. Baluja & A. Llanes (2005). Estado Actual y Tendencias del Enfrentamiento del Fraude en las Redes de Telecomunicaciones.
- Yeung D.S., Cloete I., Shi D. & Ng W.W.Y. (2010). Sensitivity Analysis for Neural Networks. Berlin: Springer-Verlag.
- Yegnanarayana, B. (2006). Artificial Neural Networks.

“LA REVISTA INNOVARE NO SE HACE RESPONSABLE EN NINGÚN CASO DE LOS CONTENIDOS, DATOS, CONCLUSIONES U OPINIONES VERTIDAS EN LOS ARTÍCULOS PUBLICADOS, SIENDO ESTA RESPONSABILIDAD EXCLUSIVA DEL (DE LOS) AUTOR (AUTORES)”