TELECOM FRAUD 101:

Fraud Types, Fraud Methods, & Fraud Technology

Authored by:
Dr. Ian Howells, Dr. Volkmar Scharf-Katz, and Padraig Stapleton
# TABLE OF CONTENTS

## INTRODUCTION

## TYPES OF FRAUD

<table>
<thead>
<tr>
<th>Type of Fraud</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbitrage</td>
<td>8</td>
</tr>
<tr>
<td>Call and SMS Spamming</td>
<td>9</td>
</tr>
<tr>
<td>Domestic Revenue Share Fraud</td>
<td>10</td>
</tr>
<tr>
<td>International Revenue Share Fraud</td>
<td>11</td>
</tr>
<tr>
<td>Phishing</td>
<td>12</td>
</tr>
<tr>
<td>Premium Rate Service Fraud</td>
<td>13</td>
</tr>
<tr>
<td>Roaming Fraud</td>
<td>14</td>
</tr>
</tbody>
</table>

## FRAUD METHODS

<table>
<thead>
<tr>
<th>Method</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cramming/Slamming</td>
<td>16</td>
</tr>
<tr>
<td>PBX Hacking</td>
<td>17</td>
</tr>
<tr>
<td>SMS Phishing</td>
<td>18</td>
</tr>
<tr>
<td>Subscription Fraud</td>
<td>19</td>
</tr>
<tr>
<td>Wangiri Fraud</td>
<td>20</td>
</tr>
</tbody>
</table>

## HADOOP-BASED FRAUD TECHNOLOGY

<table>
<thead>
<tr>
<th>Technology</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache Accumulo</td>
<td>22</td>
</tr>
<tr>
<td>Apache Hadoop</td>
<td>23</td>
</tr>
<tr>
<td>Data Ingestion Analytics</td>
<td>24</td>
</tr>
<tr>
<td>Deep Packet Inspection</td>
<td>25</td>
</tr>
<tr>
<td>Presto</td>
<td>26</td>
</tr>
<tr>
<td>Schema</td>
<td>27</td>
</tr>
</tbody>
</table>

## ABOUT THE AUTHORS

## ABOUT ARGYLE DATA
INTRODUCTION

The growing technical sophistication of criminals is leading to an arms race to see who can scale more quickly to outmaneuver the other side. Cybercriminals are increasingly adopting hyper-scale techniques to help them perpetrate fraud faster and more efficiently than ever before. Criminals are highly adaptive and continually evolving, able to work around old technology and old approaches. Give them a 24-hour batch window and they will complete their fraud in under a day. Give them thresholds or rules and they will fly under the radar or overwhelm you with false positives.

How big of a problem is fraud? The Communications Fraud Control Association (CFCA) “2013 Global Fraud Loss Survey” analyzed the $2.2 trillion mobile communications industry. It estimated that the industry is losing $46.3 billion per year from fraud, increasing at a rate of 15% from 2011. This amounts to 2.09% of revenue. Here is a quick breakdown of costs by type of fraud:

- Roaming Fraud – $6.11B annually
- Premium Rate Service Fraud – $4.73B annually
- Abuse or Arbitrage Fraud – $2.2B annually
- International Revenue Share Fraud (IRSF) - $1.8B annually
- Domestic Revenue Share Fraud (DRSF) – $0.8B annually
- Call and SMS Spamming – $0.8B annually
- Subscription Fraud – $5.2B annually
- Wangiri Fraud – $2B annually
- SMS Phishing/Pharming – $1.7B annually
- (V)PBX Fraud – $4.42B annually

At Argyle Data, we work with fraud-prevention specialists on a daily basis to keep up to date on the latest techniques being employed by fraudsters. We know that not everyone can be an expert on all aspects of fraud, and we’ve come across many professionals in the telecom industry (especially those in less-technical roles) who have expressed interest in learning more about telecommunications fraud.

We have prepared this guide as an introduction for anyone who is interested in learning more about fraud: the common types of fraud, the common techniques employed by fraudsters, and the current fraud technology being utilized to prevent fraud.
Chapter 1:

TYPES OF FRAUD
ARBITRAGE

Telecom arbitrage fraud is the exploitation of the differences in settlement rates between countries. Phone carriers often charge different interconnection rates according to the type of call or service provider involved. International calls cannot be processed and completed through one phone carrier, so the originating carrier routes traffic via an intermediary phone carrier for an additional fee, called settlement rates. Settlement rates are what phone companies pay to each other for completing their calls and, until recent regulation, they were much higher than the actual cost of completing the calls. Different companies have different settlement rates according to each country.
CALL AND SMS SPAMMING

Mobile phone spam is similar to email spam. Like with junk emails, mobile phone users receive unwanted texts and calls about special rewards or deals in the form of a simple message, a link to a number to call or text, or a link to a website.

Two laws – the Telephone Consumer Protection Act (TCPA) and the Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act – address email and mobile spam. The TCPA was passed in 1991 restricting telemarketing and the use of automated telephone equipment, such as autodialers, prerecorded calls, and SMS messages. The law also demands specific identification and contact information for autodialers and prerecorded calls.

As of March 2005, SMS spamming is illegal to send to users who haven’t specifically asked for them. However there is a loophole in the law: solicitors are only prohibited from sending unwanted messages to cell phones from Internet domains. They can still send these messages from a cell phone. So rather than receiving an SMS from an email address, you would receive it from a shortened number like “55533.”

Albeit less pervasive than email spam, call and SMS spamming is particularly more annoying for two reasons: first, there is no filter for calls and text messages like users have for junk email. The second is that call and SMS spamming victims may be charged a fee for every text message received. If users respond “stop” to unsubscribe from legitimate telemarketers, they can effectively end the unwanted messages, but if the spamming is coming from illegal scammers, victims are notifying them that the number is in fact valid and would be paying for two messages (and many more to come.) SMS spam charges range from $0.10 or $0.20 for each text. Unlimited texters beware, the message could also be linked to a premium rate service line, which is not included in unlimited texting plans.

Luckily, most cell phone carriers have ways you can report spam text messages. Victims can simply forward the message to their cell phone carrier. For most carriers (including AT&T, T-Mobile, Sprint, and Verizon), they can send the spam to the phone number “7726” (which spells out SPAM). Another preventative measure is to place the targeted phone number on the Federal Trade Commission’s Do Not Call list. If the spamming persists, victims file a complaint with the FTC at the same website.
DOMESTIC REVENUE SHARE FRAUD

Domestic revenue share fraud is one of top five emerging fraud types globally. It accounted for $800 million in losses in 2013, according to the Communications Control Fraud Association, and generated a quarter-billion dollars in revenue losses in North America alone.

Domestic revenue share fraud pertains to the abuse of carrier interconnect agreements and is very similar to international revenue share fraud and premium rate service fraud. In all three scenarios, there is an artificial inflation of traffic to a premium rate phone number. The scheme is fairly simple: A fraudster gets hold of a premium rate service number – a phone number where a portion of the charges goes to the operator and not only the phone carrier like with regular phone numbers – and inflates the traffic to the service to generate more revenue.

There are many different ways this is done and they range from very simple to calculated and organized. One of the simplest methods is by dialing a phone number just long enough to place a missed call on victims’ phones but not long enough for them to pick up, so as to lure them into calling back. This fraud method has become popularly known as “One Ring” or, in its more advanced variant, Wangiri fraud.

More sophisticated methods of artificial traffic inflation like PBX and voicemail hacking are very common today. And there are now Bluetooth-based attacks that can replace mobile phone numbers with premium rate phone numbers. This not only increases the fraud revenue, but also enables fraudsters to listen into phone conversations. VoIP hacking is also common, where hackers introduce their premium number into victims’ communications as a call-through service. With the evolution of technology, it’s only normal that fraud becomes more organized.
International revenue share fraud is a fairly new type of fraud that started out on the Internet and migrated to telecommunications. It generates $1.8 billion of losses globally and $510 million in North America every year. It is seen as one of the top emerging threats because new technology has opened up opportunities for scammers. The basis for international revenue share services is similar to premium rate services, the biggest difference being that one is international and the other is domestic. International revenue share numbers are used for services in countries where domestic premium rate numbers do not exist.

In the early days, fraudsters would commit international revenue share fraud by obtaining SIM cards (either from stolen phones or subscription fraud) and using them to call international revenue share numbers either in roaming or international areas. Call records would take as much as 36 hours to get processed through the phone company, which would then terminate that traffic. But in that window of opportunity, fraudsters would dial as many international revenue share numbers as they could, generating thousands of dollars in profits that the fraudsters would pocket.

Some countries are much more vulnerable to international revenue share fraud because of their high call interconnect or termination fees. International termination rates in North America and Europe frequently amount to less than one cent per minute. These fees are highest in the Caribbean and many small countries in the Pacific, which is where fraudsters target.

Over time, technology has made it easier for fraudsters to artificially inflate traffic to these international revenue share numbers, and the attacks have become much more organized. Call forwarding and conference calling, for example, can drastically increase the cost of each call by having multiple international revenue share numbers dialed into the same call. Other technological advancements such as SIM cloning have become a significant threat. The emergence of SMS spamming, PBX hacking, and Wangiri fraud has cemented international revenue share fraud as one of the biggest threats in telecommunication fraud.
PHISHING

Phishing is a very popular form of hacking. It is simply the attempt to acquire personal information such as usernames, passwords, credit card account information, and other sensitive information by posing as a legitimate company. This can be done via email in its most popular form, phone calls, or even text messages. Phishing attacks in 2012 accounted for an estimated $1.5 billion in losses.

Email phishing is when hackers send fake emails that are often almost identical to emails that you would receive from legitimate financial, e-commerce, or social websites. These emails often contain links that direct users to websites that either contain malware or to websites with login pages that look very similar to the login pages of legitimate companies.

In this increasingly digital world, users can use online services for paying bills, making purchases, applying for a loan, paying that loan back, paying taxes, paying for traffic violations, and so many other things. Because of this rise in online transactions, phishing is increasingly prevalent. What’s more, email phishing is easy to automate. Everything can be done online and on a massive scale, attacking thousands of users at once.

SMS phishing (sometimes called “smishing”) is particularly easy to manifest as there is no junk filter like with email and SMS messages are not as intricate as email spoofing. What’s more, users are charged for receiving these texts. Luckily though, it’s fairly easy to report fraudulent texts to your phone carrier or to the FTC (see “Call and SMS Spamming”).
PREMIUM RATE SERVICE FRAUD

Premium rate service fraud is the second largest contributor to the $46.3 billion problem of mobile fraud. It rakes in $4.73 billion globally and $1.35 in North America of losses for subscribers annually. This type of fraud directly attacks subscribers by getting them to make calls to a premium rate telephone number.

Premium rate numbers (such as 1-900 numbers) are used for upcharging telephone calls in exchange for certain services provided such as technical support and TV show voting. With a normal call, all of the call charge goes to the cell phone carrier. However, with a premium-rate phone call, a portion of that call’s charge goes to the service provider, so businesses can turn a profit through these calls. Premium rate service fraud happens when customers are unaware of the additional charges linked to a call they’ve made.

This is how adult chat lines and phone psychics generate profit; lower profit-driven services such as recorded weather forecast lines do this as well. Most of these services were very lucrative in the 1980’s and 1990’s, but since the rise of the Internet they have been fairly unprofitable. In Europe, the premium-rate service extends further, as it is fairly common for businesses in France, Germany, and the UK to charge users for their customer service lines, unlike in North America.

The most common occurrences of premium rate service fraud directly attack phone companies through the subscription fraud method. It is a fairly basic scheme that takes advantage of phone billing cycles. Fraudsters set up a premium-rate phone number through a carrier and subscribe for one or multiple phone lines through a different carrier using false information. They then run autodialers on the subscriber lines that call the premium rate numbers, running up extremely large bills. They don’t pay the subscription bills, but receive the profits from the premium-rate line. This goes on until the phone company begins to investigate a bill for non-payment, and then the fraudsters simply close out their services – leaving the bills unpaid at the expense of the phone company.

In other cases, scammers attack consumers by getting dial-up Internet users to download a dialer (a program that connects to the Internet). The malicious dialer software disconnects the telephone connection to the Internet and redials the premium-rate number. Once a connection is established, unsuspecting users have very few, if any, indication of it. They would not be notified until they receive their monthly phone bill, which would show astronomical charges due to the connection to a premium rate number.
Roaming is one of the highest revenue earners in the telecommunications industry, which means that it is also the most vulnerable to fraudulent attacks. Every year, the telecommunications industry loses $46.3 billion to fraud, and roaming fraud takes the biggest hit: about $6 billion globally and almost $2 billion in North America, according to the Communications Fraud Control Association’s most recent survey. These losses can contribute to rises in cell phone carrier rates, which in turn has repercussions on a company’s brand and customer satisfaction.

Put simply, roaming is the automatic connection to a visiting network when the user’s home network is unavailable. Subscribers cannot connect to just any network; this connection can only be accessed if the two networks have negotiated a roaming contract together. When roaming, the call data is collected by the visiting network, which can be a nearby operator just as easily as it could be one halfway across the world. The call detail record (CDR) for these roaming charges doesn’t arrive to the home network until days (sometimes weeks) later, leaving a large window of opportunity for fraudulent attacks.

Roaming fraud can happen when a subscriber that used the services of the visiting network refuses to pay for them either by claiming ignorance, insufficient knowledge of the additional costs, or by claiming that the service was never requested. It is fraud in its most basic form, and also its most common.

There are other more organized methods to committing roaming fraud, such as subscription fraud committed by subscribers, or PBX and Wangiri fraud committed by third parties, or even cramming and slamming committed by the phone companies themselves.

There are many different kinds of roaming fraud. These attacks can be a result of many factors but they are mostly due to cell phone providers not having the proper fraud management tools to prevent them. A key to preventing roaming fraud is to speed up the exchange of data between network partners, removing the window of opportunity for fraudsters to attack and leave without detection.
Chapter 2: FRAUD METHODS
Anonymous criminals are not the only purporters of fraud. Telephone businesses have been known to take advantage of consumers through billing. Cramming and slamming are two kinds of telephone scams that phone carriers use on customers to extort additional money from them.

Slamming refers to when phone carriers illegally change customers’ telephone service without their permission. Telephone service providers are obligated by law to obtain customers’ permission before switching them to a different provider. Cramming refers to when phone carriers illegally add charges to customers’ telephone bills for services they did not authorize. Similar to slamming, telephone service providers are obligated by law to obtain customers’ permission before placing charges on their telephone bill.

Cramming and slamming are so prevalent because of customers’ lack of knowledge on scams. Not only do customers not take the necessary precautions to ensure that their bills are accurate, but phone bills are confusing and difficult to understand.

Telephone service customers can avoid cramming by paying close attention to their monthly bill. Look for companies you don’t recognize and any additional services you did not order. Make sure that you made all the calls listed on your phone bill (this will also help to identify other forms of fraud like premium rate service fraud and subscription fraud). Read and (most importantly) understand the fine print on promotional offers before signing up for them.
PBX HACKING

Private Branch Exchange (PBX) is a private telephone network used within a company. Rather than requiring a line for each employee, which can be costly, a PBX system switches calls between users on local lines while allowing them to share several external phone lines for making calls outside of the PBX. Phone calls within the company are typically made by dialing a three or four digit extension. The term PBX was first introduced in the time of switchboards, where operators would manually connect calls but over time, the process has become standardized.

As useful as PBX systems are, they are particularly vulnerable to hacking. PBX fraud has caused $4.4 billion in losses globally and $1.36 billion in North America. Hacking into PBX systems can cost businesses significant amounts of money because often the owners of PBX systems aren't aware of the fraud until they receive their monthly telephone bill, in which case by then the business is obligated to pay the extra costs.

PBX hacking happens when the networks don't have strong security systems. Hackers can penetrate weak PBX systems by direct inward system access (DISA) to make international, long distance, or premium rate phone calls. They can also listen to voicemails and phone conversations, change the call routing configurations and passwords, delete or add extensions, or even shut down the PBX entirely.

Many voicemail systems on a PBX can be accessed remotely and programmed to make external calls – a useful tool for business people who are on the road a lot – but they can also be abused by criminals. Hackers are able to create a “phantom” mailbox that will allow them to make external phone calls from anywhere. Since they've already accessed the voicemail system, they also have the capacity to listen to or delete voicemail messages.
SMS phishing (often called “smishing”) operates in a similar fashion as its more well-known cousin, email phishing. Instead of using an email to bait victims to send sensitive information such as bank, credit card, or Social Security numbers, and usernames or passwords, fraudsters use text messages. In 2011, customers in the U.S. received 4.5 billion spam text messages – more than double the previous year.

Imagine getting a text message from your bank urging you to immediately resolve a problem on your card by texting important information back to them, going to a spoofed website, or calling a premium phone number. Instead of a bank on the other end, however, it is a fraudster who can quickly convert that money into cash.

Smishing can be surprisingly effective because of the lack of context surrounding texts. Where phishing emails can (usually) be spotted easily by looking at the details in the email, text messages are just simple text that come from a “5000” number and might more easily be mistaken as legitimate.

Smishing attacks can come in many different forms, from offers for a gift card from a major retailer, or even deals on mortgages, to the aforementioned alerts about accounts or cards. A more sophisticated attack that happened last year in China affected more than 100,000 Android users by infecting and automatically installing SMS malware on their phones which then sent more than 20 million texts over a five-day period, costing each user an average of $5 each.
Subscription fraud is one of the most common methods of telephone fraud. It also generates the most losses, accounting for $5.2 billion globally and $1.5 billion in North America. Subscription fraud happens with a scammer takes a user’s information illegally – or uses fake information altogether – to open up a phone account. Monthly phone bills are sent either to the victim or to a fake contact, making it difficult to identify the scammer due to a large window of opportunity.

In the case of identity theft subscription fraud, it is often difficult for the victim to resolve the fraud as he or she may not discover it for a long time due to the nature of monthly phone bills. Additionally, phone carriers are weary of customers claiming subscription fraud and victims may find it difficult to prove that they did not actually make the calls they were billed for.

The reason why subscription fraud is so pervasive is that it lends itself to many different types of mobile fraud. For example, scammers gain access to a phone line via subscription fraud to rack up charges on roaming networks, which is called roaming fraud.

In a more organized type of subscription fraud, premium rate service fraud is particularly prone to subscription fraud. It is one of the most common methods that scammers use to commit premium rate service fraud. Scammers subscribe to one or multiple phone lines through a certain carrier and generate hundreds, sometimes thousands, of calls a day to a premium rate service line that they had already set up. Scammers pocket the premium rate revenues from the artificial traffic inflation and then cancel the phone lines before a victim or phone carrier catches wind of the fraud.
WANGIRI FRAUD

In Japanese, “wan” means “one” and “giri” means “hang up.” This form of fraud, also known as “one ring and cut,” targets millions of mobile phone users by making random calls from premium-rate phone lines, letting the call ring once, and then hanging up. By leaving a “missed call” message on a user’s phone, the scammers hope that the users will call back. When they do, they find themselves listening to advertisements like subscriptions to premium chat lines or Internet services. The scammer pockets the revenue from the call, since they are hardly charged for receiving a phone call. The charge goes onto the user’s phone bill – which often times isn’t seen until weeks or months later, if ever. This extremely successful fraud method that originated in Japan has caused $2 billion of losses globally and $570 million in North America in the last year.

Wangiri fraud is a fairly new form of mobile fraud, especially in its more organized forms. When paired with international revenue share fraud, Wangiri fraud can cause serious damage. Mobile phone users see a missed call from a domestic-looking number (usually one with a three-digit area code), but the number is in fact connected to an international premium-rate service line. Many consumers are not aware that many three-digit area codes connect callers to international lines (often in the Caribbean), which is why this generates exponentially more revenue for the scammers. Additionally, Wangiri fraud used to be operated manually, but has been automated. Specialized fraud firms were common in the early 2000s, making thousand of calls a day with the use of autodialers.

A Japanese telecommunications company, NTT, suffered serious revenue and reputation damage in 2002 when a Wangiri fraud attack paralyzed its telephone networks in Osaka for over four hours. The organized firm used high-tech systems to make up to 3,100 calls a minute, which almost tripled NTT’s normal volumes. NTT has taken measures to prevent this kind of fraud – like limiting large amounts of calls that cut off before they're answered – but fraud is constantly evolving. There are presently cases of Wangiri fraud where the call simply rings twice, knowing the first ring will be ignored.
Chapter 3:

HADOOP-BASED FRAUD TECHNOLOGY
The Accumulo key-value store is an Apache project based on Google’s BigTable that is built on top of Apache Hadoop, Apache ZooKeeper, and Apache Thrift.

The open source system provides the ability to store data in massive tables for fast, random access. Imagine an Excel spreadsheet that is “tall and wide”, with billions of rows and millions of columns filled with data. Apache Accumulo – “accumulo” is Italian for “backlog” – helps sort and distribute the data through horizontal scaling across hundreds of machines, offering extremely high performance – particularly for ingest and lookup. Apache Accumulo maps rows and column keys, in a similar way to BigTable, but also timestamps values, making the database a three-dimensional mapping system. Accumulo also has cell-level access control delivering fine-grained security. Rules are encoded directly into each individual data cell to strictly limit who can see the data. This enhanced multi-level security is important for organizations with complex policies. Apache Accumulo is commonly used in industries that require a combination of extreme performance and security – government, financial services, healthcare, and mobile communications.
Apache Hadoop is an open source software that allows for the distributed processing of large data sets across clusters of computers using simple programming models. A registered trademark of the Apache Software Foundation, it is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance. Rather than rely on high-end hardware, the flexibility of the clusters comes from the software's ability to detect and handle failures at the application layer.

Hadoop is an original application because it makes big data simpler to manage by changing the economics and the dynamics of large-scale computing. There are four defining characteristics that make Hadoop a successful organizational tool:

- **Scalable**: The program is designed to easily move up and down server scales from single servers to thousands of machines without needing to change data formats or the applications built on top of Hadoop;
- **Flexible**: Because of the application's lack of database schema, it can digest any type of data, no matter the structure, the source, or the multitude of sources;
- **Cost-effective**: Hadoop's ability to run on commodity servers considerably decreases the cost per terabyte of storage;
- **Resilient**: Falling in line with the program's flexibility, Hadoop is also fault tolerant and will simply continue to process when a node is lost.

For example, imagine you had a file that was larger than your computer could hold. You could not store that file, or do anything with it, right? Hadoop lets you store files bigger than what can be stored on one particular node or server. So you can store very, very large files in massive quantities.

The Apache Hadoop framework is composed of the following modules:

- **Hadoop Common** – contains libraries and utilities needed by other Hadoop modules.
- **Hadoop Distributed File System (HDFS)** – a distributed file system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster.
- **Hadoop YARN** – a resource management platform responsible for managing computer resources in clusters and using them for scheduling of users’ applications.
- **Hadoop MapReduce** – a programming model for large-scale data processing.
DATA INGESTION ANALYTICS

As the world moves further into the digital age, organizations must build big data strategies to sort through all of their data. These strategies have the capacity to capture, process, analyze, share, store, transfer, and search the massive quantities of data going through their systems. Data ingestion analytics are crucial in ensuring efficiency, organization, and security.

Data ingestion is the systematic practice of capturing and processing data into a database. This is particularly important when detecting fraud in the telecommunications or financial services industries. The “old world” architecture for fraud and security analytics consisted of batch ingestion of transactions and logging the files into a database for later use. In telecommunications, call data records are analyzed in a 24-hour window and billing systems run on a monthly cycle, leaving a wide window of opportunity for real-time fraud attacks such as Wangiri. Similarly in financial services, it can take up to a full business day for a majority of organizations to respond to and resolve attacks. Putting data into storage for future analysis is no longer practical due to the scale of the data being ingested.

The key to packet ingestion, a framework more and more companies are moving towards in order to fight fraud, is closing the 24-hour window of opportunity. This is where real-time data ingestion analytics comes in. Real-time data ingestion technologies can tap into a network, perform deep packet inspection, and transform packets into database rows within seconds. This makes queries much faster to perform and therefore fraudulent attacks easier to resolve. Additionally, these technologies have the capability to adapt to different fraud patterns as they evolve through machine learning.
Deep Packet Inspection (DPI), also known as Complete Packet Inspection and Information eXtraction (IX), is a form of computer network packet filtering. Digital packets carry information from one location to its destination. Everything that Internet users do involves packets: sending email, visiting websites, and so on. If the World Wide Web was a human body, network packets would be cells.

This kind of filtering examines only the data from a packet as it passes an inspection point, searching for viruses, spam, intrusions, or protocol non-compliance. Other searches include collection of statistical information and defined criteria that decides whether the packet may pass through to continue onto its destination or if it needs to be routed to a different destination. DPI is a fancy term, but it simply means looking at the content of data traffic, not just the headers or the volume.

DPI enables advanced network management, user service, and securities functions as well as Internet data mining, eavesdropping, and Internet censorship. On the other hand, DPI skeptics fear that the technology would have implications on net neutrality because it would give Internet providers the means to engineer congestion according to the content types of packets being transferred.

However, DPIs are very useful in preventing attacks from viruses and worms at wire speeds and can be particularly effective against buffer overflow attacks, denial-of-service attacks (DoS), sophisticated intrusions, and a small percentage of worms that fit within a single packet.

In the context of fraud, the idea is to leverage the same DPI data that operators collect for service assurance, engineering, and network planning, to detect suspicious or abnormal behavior. Advanced fraud detection technology would be able to pick up on this activity, flag it, and eventually build alerts and cases around past experience.
Presto was initially built by Facebook for Facebook. It is an open source distributed SQL (Structured Query Language) query engine designed for ad-hoc analysis at interactive speed, from gigabytes to petabytes. As Facebook grew to petabyte scale, the company needed a query engine faster than Hive, which relies on relatively slow MapReduce processing.

Facebook’s data warehouse currently holds over 300 petabytes of data. From traditional batch processing to machine learning and real-time analytics, data processing and analytics are at the heart of building and improving Facebook. This is why Facebook built its own interactive query engine that operates at petabyte scale, from the ground up.

What separates Presto from many other fast SQL-on-Hadoop solutions is its large-scale testing. In addition, raising the bar on SQL scalability, Presto’s query response times can be as short as sub-seconds, rather than hours.

Facebook uses Presto for interactive queries that combine data from multiple sources and provide response times ranging from sub-seconds to minutes. Over 1,000 Facebook employees use Presto to run over 30,000 queries processing one petabyte every day. But Facebook wants other data-driven organizations to use and refine the open source query engine. The Hadoop community embraced Presto when Facebook open sourced the technology, and big Internet companies like Airbnb and Dropbox use the technology for their own analytics.
The term schema applies to many things. In its most basic definition, a schema is a framework. It can be used in psychology or mathematics, but in the context of technology, a database schema is the structure of a database system in its formal language, which is supported by the database management system (DBMS). A schema permits the organization and control of all data that a database processes into its system.

Traditionally, companies have used a Relational Database Management System (RDBMS) to manage their data. A RDBMS stores data into a predefined set of tables. Each table contains data categories in the columns and the data itself in the rows, similar to a spreadsheet. In a RDBMS, all of the data must be stored in the same predefined format. This means that each unique instance of data must be written to fit into the table’s set of categories. RDBMS has been used since the beginning of the Internet age and is generally considered outdated technology due to its rigid formatting rules. NoSQL is replacing RDBMS as it was built for big data, allowing quick and agile processing of information at a massive scale.

In RDBMS, data storage is based on the schema-on-write approach. This means that the database schema is defined, then the data is interpreted before it can be loaded into the system. The data has to be written into the formal language of the schema. Once the data is written, it is read and analyzed into its designated table. The problem with this approach is that data can’t be added into the table until after the table’s schema has been declared, which would have the implication that if the data changes, the table would have to be dropped and the data reloaded.

This has been the protocol for decades. However, NoSQL, Hadoop, and other big data technologies use a different approach: schema-on-read. This novel approach immediately loads the data in its original form, then each user can apply his or her own lense to interpret the data. This allows for additional agility when dealing with complex, evolving data structures.

As systems are processing and storing more and more data, the schema-on-write approach becomes ineffective because not every user needs to read the same data the same way. In big database systems, data is increasingly shared among people with different roles. Therefore the data needs to be analyzed and queried in different ways. With schema-on-read, analysts have the ability to read the data according to what they are looking for, rather than going through the process of defining a schema and then writing the data, which may not be written in a way that
is useful for everyone else. Even if the schema is defined in a way that at the time is practical for all queries, a new query will undoubtedly emerge. With schema-on-read, that problem is eradicated, enhancing efficacy within a large database.

Long story short, the schema-on-read breaks past old barriers so that companies can easily load data as-is and get value from it immediately. This approach is built to work seamlessly with new big data technology such as Hadoop, to ensure quick and easy data organization.
ABOUT THE AUTHORS

Dr. Ian Howells, Chief Marketing Officer
ian.howells@argyledata.com

Ian Howells is a passionate technologist and serial entrepreneur with over 25 years experience building successful technology companies through IPO and acquisition. Prior to joining Argyle Data, Ian held senior marketing roles at Documentum from its early days through to IPO, SeeBeyond through its IPO and acquisition by Sun, Alfresco from a startup to the largest private open source company in the world, StorSimple, acquired by Microsoft, and Couchbase through a period of 400% growth. Ian has a Ph.D. in distributed databases and has published a number of papers and contributed to books on related topics. Ian has a passion for applying big data analytics techniques to marketing and open source. He previously penned the blog “Open Source Hearts and Minds” for Computerworld.

Dr. Volkmar Scharf-Katz, VP of Mobile Strategy and Solutions
volkmar.scharf-katz@argyledata.com

Dr. Volkmar Scharf-Katz has over 15 years of experience in the mobile industry specializing in analytics, big data platforms, and communication technologies. Previously, he was Chief IP Networks Partner, R&D and Head of Service Platform at Vodafone where he was responsible for building advanced computing environments to demonstrate real-time analytics, showcase artificial intelligence and develop mobile applications. Dr. Scharf-Katz has a Ph.D. in computer science and information technology from Stanford University. He has vast and successful international experience in creating new and disruptive solutions for communication networks and mobile banking, and he also holds several patents.

Padraig Stapleton, VP of Engineering
padraig.stapleton@argyledata.com

Padraig Stapleton brings years of industry-leading management and technical expertise across a number of areas including mobile telecommunications and big data. Most recently he was VP of Engineering and Operations for the Big Data group in AT&T responsible for development of their big data platform. Previously to that he was involved in a number of successful startups as VP of Engineering building development teams and delivering innovative products to the market place. Padraig has held senior leadership roles in various companies including Telephia, which was acquired by Nielsen, and InterWave Communications.
ABOUT ARGYLE DATA

Argyle Data is the leader in real-time fraud and security analytics at network speed and Hadoop scale, offering solutions for the largest data-driven companies in mobile communications and financial services. Argyle Data offers a real-time fraud and security analytics application built from the ground up on Hadoop using the latest big data, machine learning, and anomaly detection technology proven at Facebook and Google. It is able to detect fraud not detected by existing systems, discover fraud in minutes vs. days, discover both new and old fraud attack techniques, and dramatically reduce false positives.

ARGYLE DATA, INC.
2755 Campus Drive, Suite 165
San Mateo, CA 94403
USA
Tel: 800.695.6021
Email: info@argyledata.com
Web: www.argyledata.com

To learn more, please view our on-demand webinar:

Rewriting the Book on Mobile and Fixed-Line Fraud