Meaningful CDR Statistics For Identifying Scammers on VoIP Networks

Prior to 2019 the VoIP industry could be accurately described as the “Wild West”. Anyone with a Gmail address and a PayPal account could sign up with any number of intermediate carriers and be running high volumes of minutes with very little oversight and even less enforcement.

Well, it's not 2019 anymore and if the regulators determine that your network is participating in illicit traffic then 2022 could turn into a really bad year, really quickly. Your best defense is strong KYC and active monitoring of any traffic that is on your network already.

The purpose of this document is to describe common metrics used to monitor existing traffic and identify potentially bad actors on your network. Once a potentially bad actor has been identified this document offers recommendations on how to determine if the traffic is legitimate or not. The metrics described in this document come from CDR that any class 4 switching platform would have available.

In each section we will identify what we look for and what metrics indicate potentially undesirable traffic. The source code is also provided to assist in your own analysis.

If you have any questions about the statements made in this document please do not hesitate to contact us directly. Contact information can be found on the last page.

Version	Date	Author	Description
1.0	2022-05-18	Dean Hansen	Initial draft.
1.1	2022-05-21	Dean Hansen	Added the FTC DNC/Robocall API section.
1.2	2022-05-24	Dean Hansen	Added the Neighborhood Spoofing section.
1.3	2022-05-25	Dean Hansen	Added the Reflection Spoofing section.
1.4	2022-05-26	Dean Hansen	Added the Call Duration Distribution section.
1.5	2022-06-06	Dean Hansen	Added STIR/SHAKEN Attestation Distribution section.
1.6	2022-06-08	Dean Hansen	Added the Detecting Wangiri Premium Rate Number Scams section.

Get the latest online version here.

Overview

There is no individual metric in Call Detail Records ("CDR") that provides confident determination on whether a customer is sending scam traffic over your network or not. That's not to say CDR don't have meaningful data, they absolutely do, but all you can derive is which customers deserve to be investigated. The following sections will describe metrics we employ to analyze network traffic in order to determine which customers deserve your fraud team's attention.

Once we identify possible scam traffic via CDR then we engage external tools to determine whether or not it is scam traffic. This second step is very important. You must listen to what the originator is saying to callees to determine if something is a scam. Statistics alone are not evidence of wrong-doing.

Steps to investigate suspicious traffic are described in the *Is It a Scam* section.

Audiences

This document was written with 2 primary audiences in mind.

Wholesale / Aggregator - This is one of the most dangerous places to be since the call originator can be multiple hops away and scammers will be banking on blending their nastiness into legitimate traffic so they can fly below the radar. In this scenario we recommend paying particular attention to the media-level queries as it will help you unblend the traffic and view into the calling behaviors of your customer's customers.

Call Center Service Provider - This is a pretty good place to be when you're hunting scam traffic. The originator is your direct customer and you have the control necessary to terminate the traffic at its source if you find scams originating on your infrastructure. In this scenario we recommend focusing on account-level statistics, media IP-level and campaign-level statistics.

Before We Get Going

- This document should not be interpreted as legal advice.
- YouMail is referenced often in this document. We have no direct affiliation with YouMail. We use their transcript / audio file features (YouMail Watchlist) during investigations.

Definitions

Answer Seizure Ratio (“ASR”) - Total number of answered calls divided by the total number of attempts (seizures). Represented as a %. E.g. 100 attempts with 90 answered calls = 90% ASR.

Automatic Number Identification (“ANI”) is a telephony service that allows the receiver of a phone call to capture and display the phone number of the phone that originated the call. It is often referred to as the CallerID, CID, DID or Source Number.

Average Call Duration (“ACD”) - The average length of connected calls (in seconds) for the given entity being measured.

Average Length of Call (“ALOC”) - See Average Call Duration.

Call Detail Records (“CDR”) - A call detail record is a data record produced by telecom equipment that details the properties of each call. CDR are used for billing and reporting purposes.

Callee - One who receives a telephone call.

Dialed Number Identification Service (“DNIS”) - The number that was dialed. The called number.

Do Not Call Registry (“DNC”) - The DNC is a database implemented by the Federal Trade Commission, in compliance with the Telephone Consumer Protection Act of 1991, in 2003 that allows people to register their numbers requesting telemarketers do not call them.

Federal Trade Commission (“FTC”) - The FTC is an independent agency of the United States government whose mission is the enforcement of civil U.S. antitrust laws and the promotion of consumer protection.

Internet Traceback Group (“ITG”) - A collaborative effort of VoIP, wireliness, wireline and cable companies that trace and identify the source of illegal robocalls.

Know Your Customer (“KYC”) - Information collected by the service provider (you), typically in the form of a questionnaire, in order to verify the identity of and assess risk of, prospective customers. A strong KYC is highly recommended.

Local Exchange Routing Guide (“LERG”) - Refers to the iconectiv document and datasets used by telecom carriers to identify NPX/NXX routing and homing information as well as element and equipment designations.

Neighborhood Spoofing - Neighborhood Spoofing is when you get a call from a number that has the same NPANXX/Area Code + Exchange/first 6 digits of the caller that matches your number.

North American Number Plan (“NANP”) - The telephone numbering plan that governs twenty countries. Primarily the United States, Canada and the Caribbean.

Numbering Plan Area (“NPA”) - Commonly known as the Area Code. It is a 3-digit number referring to the service region.

NXX - Often referred to as the Exchange. It is the 4th, 5th and 6th digits of the 10-digit NANP phone number.

Originator - This is the person or entity that initiated the call attempt. Depending on where you sit in the call path the originator could be your direct customer or they could be multiple hops away. For the purpose of identifying and stopping scam traffic, the closer you are to the originator (in hops) the more effective you can be.

Reflection Spoofing - Reflection Spoofing occurs when the calling number and the dialed number are the name. Scammers use this method to prompt people to answer out of curiosity and concern. We identify this behavior via our ANI/DNIS Match Ratio metric.

Ringless VoiceMail (“RVM”) - A method to drop a pre-recorded message on a voicemail box without ringing the telephone. Also known as a voicemail drop.

Structured Query Language (“SQL”) - A standards-based programming language used in databases and data management systems.

Telephone Consumer Protection Act (“TCPA”) - The TCPA was passed by the United States Congress in 1991 in response to consumer concerns about telemarketing. The act sets guidelines for telemarketing practices and prescribes penalties for violations.

Meaningful Metrics

The metrics described in this section should not be considered absolute proof that an account is sending scam traffic. These metrics are ones that are commonly observed in scam traffic and should be used as indicators of what accounts require your attention. If any of the metrics below exceed your thresholds then we recommend referring to the "*Is It a Scam?"* section of this document in order to determine whether the traffic is, or isn't, a scam.

Source code is provided for all of the metrics described in this paper.

ANI Cardinality Ratio ("ACR")

ACR is the measure of uniqueness in the ANI values in a dataset. For example, if a customer made 1,000 call attempts and there were 1,000 unique ANIs then the ACR would be 100%.

Oftentimes, scammers will use a random number generator as the ANI in order to 1) get around ANI profiling, 2) save money and 3) avoid receiving call backs and 4) prompt the callee to answer. Random number generators create a very high ACR. A high ACR value does not conclusively prove it is scam traffic but it is definitely something that should be investigated. If your scan reveals an ACR greater than 90% on any given account, campaign, signal IP or media IP then we recommend investigating the traffic source.

Invalid ANI Ratio ("IAR")

IAR is the measure of invalid ANI values within a dataset. For example, if a customer made 1,000 call attempts and 100 had invalid ANIs then their IAR would be 10%. Whether or not an ANI is valid is done by comparing the ANI to the LERG to validate the NPA and NXX. People using random numbers will often simply use a function to create a random number between 1,000,000,000 and 9,999,999,999 but will never actually check to make sure the NPA and NXX of the random number are valid.

Statistically, if someone is using a random number generator then ~13% of the ANIs generated will be invalid. If you have an account or media IP with high ACR and IAR then you can feel confident that the customer is up to no good. We recommend investigating immediately.

ANI/DNIS Match Ratio / Reflection Spoofing

This is defined as a call attempt wherein the ANI and the DNIS are the same value. Also known as Reflection Spoofing. This is akin to the random number generator in that the originator is avoiding using a valid ANI that receives callbacks. Anyone that is avoiding receiving callbacks is probably doing something they shouldn't be doing. We can usually trace it back to a scammer that is trying to bypass ANI profilers and is using the DNIS as the ANI instead of using a random function generator to create an ANI. In some ways you could say it is better than a random number generator since it won't generate an invalid ANI.

If your CDR has instances of matching ANI and DNIS then we would definitely recommend investigating the customer in question. If you are a wholesaler then the signal IP and media IP queries will help you isolate it.

403 Ratio

403 / Forbidden is normally used to indicate the source of the attempt is not authenticated. For example, if you tried to send an attempt to a carrier but your IP was not configured on their end you would receive a 403 back. In this context you get 100% 403s (not authenticated) or 0% 403s (authenticated). In the past year some carriers have started using 403s as an ANI filter instead so now it is becoming common to see a small percentage of 403s in your return codes.

As a general rule, any time 403s at the account level go above 3% then we will use the ANI-level query to determine which ANIs are getting the 403s and then investigate those.

404/Disconnected Number Ratio

This is defined as the percentage of 404s in a customer's call attempts. For example, if the customer made 100 attempts and 5 of them resulted in a SIP code of 404 then their 404 Ratio is 5%. We encounter 2 common reasons for high 404s.

1. Call centers that manage their data poorly.
2. Scammers that are brute-force dialing.

Now, a call center that manages their data poorly is not necessarily doing anything illegal. It just means there is an opportunity to improve. We've encountered no shortage of call centers that were perfectly legitimate businesses. They just needed some help managing their data more effectively. This can be an opportunity to help your customer with their business and improve performance for both you and them.

On the other side of the 404 coin is the scammer that is brute-forcing numbers. Typically these scammers are using random DNIS generators and/or pulling number lists from anywhere they can get them. This behavior results in an unusually high 404% being returned from your underlying carriers.

We typically investigate accounts with a 404 ratio above 2%. Even if it turns out the customer is not running scam traffic it's still good to get 404s off your network. It's never a waste of time to chase 404s.

486 Ratio

SIP Code 486 was originally implemented as a “Busy Here” indicator. Over the last couple of years it has also been implemented by some carriers as an ANI blocker and can aptly be described as a “I’m not necessarily busy right now but I don’t want to talk to you.” response from the callee.

In organic traffic, the 486 ratio is normally under 2%. Scammers, by their nature, tend to do unpleasant things at high volume and it is reflected in their higher 486 ratio. If we see an account with a 486 ratio greater than 3% then we check 486% by ANI and put the offending ANIs into a YouMail Watchlist to see what the originator is saying to people.

603/607/608 Ratio

607 / Unwanted is a fairly new SIP code introduced to communicate that the callee does not want a call from this caller and that any future attempts from the same caller are likely to also result in 607 / Unwanted. SIP code 608 is similar to 607 but indicates the code was returned by an intermediate carrier on behalf of the callee. Some switching platforms are not currently compatible with SIP codes 607/608 so they are using SIP code 603 in lieu. For the purpose of this analysis we can treat 603, 607 and 608 as the same response.

During our analysis, any account showing cumulative 603/607/608 greater than 3% will get their individual ANIs analyzed and any ANI breaking that threshold will be loaded into YouMail for investigation.

For 603/607/608, you will also want to check the STIR/SHAKEN attestation level on the calls. No attestation or “C” attestation will generate 603/607/608.

YouMail Scam Ratio

YouMail offers a dataset via API that identifies ANIs that have received scam reports. This dataset is updated throughout the day based on call events examined by YouMail's sensor network. If your switching platform is capable of ingesting the dataset then it can be used as an additional means of blocking bad calls in real-time. Ideally, any call blocked by this filter would also record a CDR that identifies the block reason so it can be reported on as well.

If the switching platform can't ingest external data for filtering then you can perform a simple join of your CDR to the YouMail dataset to see if any accounts have any scam reports per YouMail's dataset. If any ANIs are returned in the report then you can use YouMail's Watchlist feature to see the transcripts and listen to the recordings associated with the account.

Although this paper focuses very much on the scam calls inundating the US telephony network; YouMail can score in 3 dimensions. Spam, Fraud and Unlawful Telemarketing. These scoring dimensions will allow you to get a birds eye view of your customer's traffic makeup.

Repeated DNIS Distribution

Repeated dialing is a classic brute force move employed by scammers. Hit them hard and hit them often. This metric tends to be one of the more damning ones since there are very few legitimate reasons for an originator to call the same number many times over the course of a day.

When heavy repeated dialing is detected on an account then we recommend loading the account's entire ANI list into YouMail and investigating any Call Event that it detects.

If you have any customers consistently calling the same number more than 2-3 times in a day then it wouldn't hurt to see why they would be doing that.

Neighborhood Spoofing Ratio

Neighborhood Spoofing occurs when the NPANXX / Area Code + Exchange / first 6 digits of the caller's number matches the dialed number. This is a common tactic employed by scammers to make callees think that the caller is "just around the corner". It preys on a person's inherent local trust and it is used to increase the likelihood of a callee answering. On its own, a local call is not necessarily illegal but if a traffic source has an unusually high Neighborhood Spoofing Ratio then it warrants investigation.

STIR/SHAKEN Attestation Distribution

STIR/SHAKEN's primary function is to eliminate the spoofing behavior that has plagued consumers for years. STIR/SHAKEN works by having the originating carrier apply an attestation level to each outbound call. Attestation can be interpreted as a level of trustworthiness assigned to the call by the signing carrier. By June 30th, 2022; all originating carriers will be required to implement STIR/SHAKEN on their networks.

Attestation provides the following improvements:

- Eliminate spoofing by requiring the originating carrier to assign a level of trustworthiness to the call. The efficacy of this relies on the legitimacy of the signing provider. Hopefully, illegitimate signers will lose the certificates with extreme prejudice.

- The terminating carrier (the receiving party) can use attestation level in the blocking algorithms.

- Law Enforcement will be able to quickly see who is signing bad calls instead of using the existing traceback method which is more time-consuming. This increases accountability considerably.

Attestation levels include:

A - Full Attestation - The signing service provider has determined that the caller is authorized to use the calling number. This is the highest level of attestation.

B - Partial Attestation - The signing service provider has authenticated the call origin (customer) but cannot confirm they are authorized to use the calling number.

C - Gateway Attestation - The signing service provider can authenticate the network the call was received from but cannot authenticate the call originator nor the calling number. This is the lowest level of attestation. The only thing the signing party is really attesting to is the fact that the call traversed their network.

Unsigned - This is not really an attestation level. It signifies no one was willing to sign the call. Many carriers will outright block these.

As an example, if “Dynamic Telecom[1]” signed a call with attestation level A then it conveys more trust than if they have signed it with a C. In a perfect world, all calls would have A-level attestation.

[1] This is a fake carrier.

Our Attestation Distribution report will breakdown the levels of attestation present on your network. As a network provider, you would naturally want the highest level of attestation on the calls crossing your network. We would recommend starting with our Attestation Distribution report at the account level. If you determine any given account does not have a favorable attestation distribution then you can use the signal IP and Media IP reports to isolate its source.

When hunting scam traffic, B and C-level calls are where you are most likely to find it. [2]

Caveat:

- Some unscrupulous vendors will sign calls with a higher level of attestation than they should. Regulators, hopefully, will terminate their certificates so they can no longer sign calls.

Call Length Distribution

The most common misconception we encounter when talking to network operators is that “short traffic is bad (scam) and long traffic is good (not scam).”

This is a widespread assumption and the scammers know it. Smart scammers know they only have to extend the calls by a few billing intervals in order to fly under the radar. The SIP40 project provides code that breaks traffic sources into duration buckets (1-6 seconds, 7-12 seconds, etc) by distribution percentage but it is not intended to be a primary indicator of scam.

Call Length is a very broad metric and if used by itself will lead to many false positives. We recommend only using Call Length Distribution in concert with our other metrics.

[2] https://transnexus.com/blog/2022/shaken-statistics-april/

Detecting Ringless VoiceMail ("RVM")

Until recently, if you asked 10 people their opinion on whether or not RVM triggered TCPA, you would get 11 varying opinions.

On February 2nd, 2022[3][4] Federal Communications Chairwoman Jessica Rosenworcel made it clear that the FCC intends to clarify that ringless voicemails are subject to the TCPA. If the action is adopted then consumer's consent will be required. This does not inherently make RVM illegal but it does mean rules apply, like any other call.

The *Source Code* section of this document provides a query that will identify any of your accounts that are running RVM over your network.

[3] https://www.fcc.gov/document/rosenworcel-proposes-ringless-voicemail-robocall-protections

[4] https://tcpaworld.com/2022/02/02/rvm-rip-new-fcc-chairwoman-sets-her-sights-on-ringless-voicemail/

Detecting Wangiri Premium Number Scam

“Wangiri” is a Japanese word for “one ring and cut”. It’s a scam where they trick people into calling back premium rate numbers by placing high volumes of calls and ringing just once so the callee sees a “missed call’ notice. When the victim calls back, they are billed a high per-minute rate for the call and the scammer shares the revenue with the phone company.

Spotting this scam in your CDR is quite easy. The calls will only have 1 ring from an international number and then get immediately canceled by the originator resulting in high 487 counts and almost no connected calls. Since none of the calls actually connect, scammers have to blend this traffic into other traffic so that the account-level 487% doesn’t tip off the service provider (you) that is unknowingly carrying the bad traffic. We unblend the traffic at the account level and then help you isolate which IP(s) it is coming from so you can stop it at the source.

Popular originating countries include, but are not limited to:

- New Caledonia
- Ecuador
- Myanmar
- Africa
- United Arab Emirates
- Lesotho
- Haiti
- Cook Islands
- Syria
- Latvia
- Switzerland

Federal Trade Commission DNC/Robocall API

In April 2020, the Federal Trade Commission (“FTC”) published an API endpoint that allows you to collect daily data about Do Not Call (“DNC”) and robocall complaints that are reported to them. On any given day this can be 12,000 to 15,000 complaints which makes it a meaningful dataset.

That said, don’t take this dataset as gospel. The FTC states plainly that these complaints have not been vetted. We recommend using the dataset as a litmus test. Meaning that you use the ANIs in the FTC’s dataset and compare them to the ANIs in your customer’s CDR as well as any DIDs you may have sold to customers. If you get a match then it becomes another indicator that the customer is worth investigating.

Keep in mind that spoofing is still a thing. Don’t declare a customer guilty over the odd match. Consistent matches; however, warrant your attention.

This data feed is also available via CSV download. We have built this feature into our open-source project.

Metric Levels

Account

Reviewing these metrics at the account level is meaningful if the majority of the customer's traffic comes from a single source. If the customer is a traffic aggregator then bad traffic from a small percentage of their customers can blend into the remaining clean customers. If your customers are aggregators (think of a wholesaler or software company that sells call center software) then we recommend digging into IP-level statistics, particularly the originating media IP[5].

T-SQL version available on GitHub.

Originating Signal IP

If you are a wholesaler and your direct customer is a switchless reseller (more common than you'd think) then your customer would be routing their customers directly to your network and each IP they signal from would effectively be one of their individual customers. In this scenario, account-level statistics could end up blending a small amount of scam traffic into the overall accounts statistics and it wouldn't be detectable. For this situation we would recommend using signal IP-level queries so you can easily identify the varying calling behavior between your customer's customers to isolate any of them that are behaving poorly.

T-SQL version available on GitHub.

Originating Media IP

If you are a wholesaler and your customer runs their own switching infrastructure then they may have many originating customers but you will only see a small number of signal IPs in the CDR. This can make it hard to isolate the individual behaviors of your customer's customers since it will all be blended together. In this scenario we recommend the use of the media IP-level queries so you can view the customer's customer independently for analysis and investigation.

T-SQL version available on GitHub.

[5] This may not be available if the media is being anchored somewhere between the originator and your network. If this is the case then you may have to identify bad ANIs and have your customer chase down the originator.

Campaign

Campaign-level statistics are suitable for the call center service provider that wants to analyze behavior of the different campaigns on a single customer.

T-SQL version available on GitHub.

ANI

ANI-based statistics offer the greatest level of detail when it comes to metrics although we don’t recommend starting by looking at ANI-level. Depending on the size of your network the volume of data it returns can be overwhelming. In most scenarios you would start with account, campaign, signal IP or media IP level metrics and then drill into ANI-level if you’ve identified something of concern.

T-SQL version available on GitHub.

Suggested Thresholds

We recommend the following thresholds. Violation of any of these parameters triggers an investigation using the methods described in the *Is It a Scam?* Section of this document.

Metric	Threshold	Description
ANI Cardinality Ratio	>= 90%	Greater than 90% ACR on any account, campaign, signal IP or medial IP indicates the use of a random number generator which should be investigated.
Invalid ANI Ratio	> 0%	Any invalid ANIs at the account, campaign, signal IP or media IP level indicates a problem. It could be a simple configuration issue by the customer but it is also a tactic used by scammers.
ANI/DNIS Match Ratio / Reflection Spoofing	> 0%	Any time a call attempt comes in with the ANI matching the DNIS is ungood.
403 Ratio	>= 3%	403s are a new form of ANI blocking.
404 Ratio	>= 2%	404s indicates unmanaged data or brute-force calling. Always worth it to investigate and remove 404s.
486 Ratio	>= 3%	486s are a new form of ANI blocking.
603/607/608 Ratio	>= 3%	603/607/608 codes indicate that the callee does not want to speak to the caller. High 603/607/608 ratio should be investigated using the ANI queries.
YouMail Scam Ratio	> 0%	We would recommend investigating any reports of scam identified via YouMail.
Repeated DNIS	More than 2-3 times per day.	Very few legitimate reasons to pound the same number all day long. Repeated dialing should be investigated.
FTC Complaints	>= 0.5%	The FTC is the first to state their complaints are not vetted but any

		meaningful amount of complaints should still be investigated for legitimacy.
Neighborhood Spoofing Ratio	>= 5%	Some calls are legitimately local calls but a high ratio of neighborhood calling warrants looking into.
Wangiri	>= 1%	Some short-rings can happen in organic calling. Not every short-ring is a scam. We recommend investigating the signal IP and media IPs of any account above 1%.

Is It a Scam?

When we identify a potential scammer through CDR analysis then it comes time to review the transcripts and audio files of the calls in order to determine if the traffic is a scam or not.

The YouMail Watchlist feature provides the ability to upload[6] a list of ANIs and then read and/or listen to the originator's audio. This is where the rubber meets the road. CDR metrics are great for spotting smoke but you need to know what an originator is saying to a callee to make a confident determination on whether or not the traffic is a scam.

The common scams (IRS, SSA, Amazon, CCIRR, Utility Disconnect, Direct TV Discount) are well documented by the Internet Traceback Group ("ITG"). They are pretty easy to recognize when you read the transcripts and you can feel confident about turning down the originator's service when you encounter them. If it is less obvious whether or not it is a scam then you can expand your investigation to include the following:

1. Call the ANI. A legitimate business calling for legitimate purposes should answer your call and be able to respond to basic inquiries from you.

2. Is the call recording robotic / pre-recorded? Often scammers will brute-force pre-recorded messages since it is much cheaper than using live humans. They are casting a wide, yet indiscriminate, net in hopes of getting someone on the phone.

3. Does the call recording seem to be of poor quality audio? We could never understand why so many scammers have such poor quality audio given how cheap proper hardware/bandwidth is but poor audio quality is certainly an indicator of something amiss. Legitimate companies rarely have audio quality that sounds like they are calling from the moon.

4. Is the originator using broken English? We're not suggesting all callers with poor pronunciation or poor grammar are automatically scammers but it is more prevalent amongst scammers. It is worth taking into consideration.

5. Is the caller being threatening and/or aggressive? Many scammers rely on threats or aggression in order to get a desired response from a callee. This is unlikely to be a legitimate business. Example: "This is Agent Smith calling from the Attorney General's office. There has been a warrant issued for your arrest…".

6. Do key personnel show up searches? Take the contact names off the customer's 499, Robocall Mitigation filings and your own KYC form and Google them with the keywords "fcc" and "attorney general" appended. If the customer has been charged by law enforcement then it will come up.

[6] A REST API is also available.

7. Is their domain new? Use https://lookup.icann.org/en/lookup to check the date it was registered. If it was registered yesterday, that's not a good sign.

8. Check the KYC paperwork. There have been numerous instances of copycats showing up lately. Scammers will sign up using 499, FCC and contact information for a legitimate company. E.g Legitimatevoip.com might be a legitimate VoIP provider but if "legitimatesvoip.com" signs up, that's trouble. The scammer will redirect their similarly-spelled scam domain to the legitimate company's domain so everything looks on the up and up on the surface. When you're investigating existing traffic that is scammy but the KYC looks legit, check to make sure someone didn't slip a scam domain past the goalie.

If your CDR analysis indicates a lot of negative feedback but your investigation determines that the traffic is not a scam then it wouldn't hurt to review the *Stop Pissing People Off Protocol* with your customer. It is not uncommon for legitimate callers to employ calling behaviors that garner negative feedback and increase the likelihood of being flagged as "scam" simply because they've done something to upset the people they are calling.

Stop Pissing People Off Protocol

Just because a call is unwanted doesn't make it illegal but in today's world of crowd-sourced feedback, poor behavior by callers will result in scam reports even if they are legitimate calls. We've encountered more than enough originators over the years whose traffic is perfectly legal but their calling behavior is more of a brute force approach. This results in negative feedback by callees and negative statistics accrued by the carriers. When this happens, no one, including you/the customer/the carriers, will be happy with the performance.

This is an opportunity to help your customer be better at what they do and improve the calling experience for all parties involved.

The first thing we recommend is loading the customer's ANIs into YouMail's Watchlist and letting their sensor engine cook on them for a day. YouMail provides transcripts and audio files for the calls made to their subscribers which will help you determine if the originator's calls are scams or not. I.e. if the audio contains "IRS", "Amazon", etc then it's a scam and getting the customer off your network and quickly as you can is best for everyone. If the transcripts and audio indicate the calls are legal then it is likely a poor calling behavior by the customer that is causing the negative feedback.

Below is a list of common behaviors we see.

1. Repeated Dialing - We see this one quite a bit. Perfectly legitimate callers sometimes feel the need to dial a number many times over the course of a day in order to reach someone. This might be fine in an emergency situation but most times it is very unnecessary and the only thing it accomplishes is negative feedback from your downstream carriers and the callee. If your customer is pounding the same numbers repeatedly then urge them to pump the brakes for everyone's sake.

2. Poor English/Bad Accent - Many companies will out-source their customer communications overseas. Based purely on the economics involved it's hard to disagree with that logic. Where the problem can occur is if the agent(s) have broken English or a bad accent then the callee will assume it's a scam and report it as such. Since a large percentage of scam traffic originates overseas it's pretty tough to blame the callee for assuming it's a scam. YouMail's service provides the audio files for the calls so you can easily determine if the agents are the source of the negative feedback. If they are, we recommend encouraging your customer to demand better quality agents.

3. Aggressive Audio - Hard-selling is a common tactic employed that is not recommended in this world of crowd-sourced feedback. Many scams employ hard-sell tactics (think IRS / Law Enforcement scams threatening fines/jail time). Callees are conditioned to assume it is a scam and report it. This type of negative feedback is easily avoided.

4. Calling the DNC - The DNC has been around for almost 2 decades but it is still widely ignored by some call centers. Calling people that have expressly stated “don’t call me without consent” is a sure fire way to receive complaints about your number(s) and get them blocked by carriers and local phone apps. DNC scrubbing is available by:
    a. Registering for your own copy with the FTC.
    b. Engage data service providers like DNC.com.

Source Code

The source code provided will return the metrics described in this document. We don't recommend running these queries on your production networks during peak hours. They can be quite resource-intensive. Running these queries on non-production hardware is recommended. If non-production hardware is not available then do it out of peak hours.

The source code is available on our GitHub.

Notes

- These queries assume you are querying the ingress record for each call, not the individual egress attempt records. If your platform records ingress and egress CDR in the same table then add a filter ("where" clause) to your queries so it only looks at the final record indicator.
- ANI and DNIS formats should be standardized in this table. I.e. all 1NNNNNNNNNN or all NNNNNNNNNN. Whatever format you choose, they all should be the same.
- The Structured Query Language ("SQL") examples use T-SQL syntax but can easily be converted to MySQL, Postgres, etc. Our GitHub will also publish new features in various SQL languages as they become available.

CDR Fields

Table: **cdr_in**

T-SQL version on GitHub. This table must be installed on your database.

Field	Datatype	Description
call_id	Text or Integer (depending on your platform)	A unique ID referencing the call record. Required.
attempt_date_time	Date/Time	Used to filter queries by period or breakdown result by period. Required.
account_id	Text or Integer (depending on your platform)	Required if performing account-level metrics.
campaign_id	Text or Integer (depending on your platform)	Required if performing campaign-level metrics.
signal_ip_orig	Text	Required if performing Signal IP-level metrics.
media_ip_orig	Text	Required if performing Media IP-level metrics.
ani	Text	Required if performing ANI-level metrics. Highly recommended.
dnis	Text	Required if scanning for RVM or Repeated Dialing.
sip_code	Integer	Required.
duration	Integer	Length of a connected call in seconds. Required for the Call Length Distribution metric.
attest_level	char(1)	Optional. Valid values are A, B, C, null, blank.
ring_time	Integer	The ring time, in milliseconds, of the call. Required if you want to scan for the Wangiri One Ring scam detection.

Account Queries

Account-level queries include the following metrics per account:

- Attempts
- ASR
- ANI Cardinality Ratio
- ANI/DNIS Match Ratio
- 403 Ratio
- 404 Ratio
- 486 Ratio
- 603/607/608 Ratio
- YouMail Scam Ratio
- Repeated Dial Distribution
- RVM Detection
- Wangiri Premium Number Scam Detection
- FTC DNC API Match Ratio
- Neighborhood Spoofing Ratio
- Call Length Distribution
- Invalid ANI Ratio
- Attestation Distribution

T-SQL version available on GitHub.

Signal IP Queries

Signal IP-level queries include the following metrics per IP:

- Attempts
- ASR
- ANI Cardinality Ratio
- ANI/DNIS Match Ratio
- 403 Ratio
- 404 Ratio
- 486 Ratio
- 603/607/608 Ratio
- YouMail Scam Ratio
- Repeated Dial Distribution
- RVM Detection
- Wangiri Premium Number Scam Detection
- FTC DNC API Match Ratio
- Neighborhood Spoofing Ratio
- Call Length Distribution
- Invalid ANI Ratio
- Attestation Distribution

T-SQL version available on GitHub.

Media IP Queries

Media IP-level queries include the following metrics per IP:

- Attempts
- ASR
- ANI Cardinality Ratio
- ANI/DNIS Match Ratio
- 403 Ratio
- 404 Ratio
- 486 Ratio
- 603/607/608 Ratio
- YouMail Scam Ratio
- Repeated Dial Distribution
- RVM Detection
- Wangiri Premium Number Scam Detection
- FTC DNC API Match Ratio
- Neighborhood Spoofing Ratio
- Call Length Distribution
- Invalid ANI Ratio
- Attestation Distribution

T-SQL version available on GitHub.

Campaign Queries

Campaign-level queries include the following metrics per campaign:

- Attempts
- ASR
- ANI Cardinality Ratio
- ANI/DNIS Match Ratio
- 403 Ratio
- 404 Ratio
- 486 Ratio
- 603/607/608 Ratio
- YouMail Scam Ratio
- Repeated Dial Distribution
- RVM Detection
- FTC DNC API Match Ratio
- Neighborhood Spoofing Ratio
- Call Length Distribution
- Invalid ANI Ratio

T-SQL version available on GitHub.

ANI Queries

ANI-level queries include the following metrics per ANI:

- Attempts
- ASR
- ANI/DNIS Match Ratio
- 403 Ratio
- 404 Ratio
- 486 Ratio
- 603/607/608 Ratio
- YouMail Scam Ratio
- Repeated Dial Distribution
- RVM Detection
- FTC DNC API Match Ratio
- Neighborhood Spoofing Ratio
- Call Length Distribution
- Invalid ANI List

T-SQL version available on GitHub.

Summary

Don’t under-estimate scammers. Most of them might be brute-force boneheads but some are extremely well organized and very clever. Scammers are also humans, not bots. They don’t all follow the same patterns or use the same methods. New scam methods are introduced all the time; however there is valuable information in your CDR to help isolate suspicious traffic. If it looks funny, investigate it. Failing to investigate suspicious traffic may put your company on the wrong side of an Attorney General.

We hope the details provided here assist you in your own analysis and investigations. If you run into a scammer using behaviors not identified in this paper we would love to hear about it!

Biography

I started doing database work in 1999 with a utility billing software company. The work primarily involved query optimization and database performance debugging. In 2007 I was introduced to Telephony and the seemingly unlimited amount of data (CDR) that comes with it. I dove in head first and opened my own company that specialized in telephony billing. The intellectual property for the billing platform was licensed externally in 2015.

While deciding what my next project would be I was given the opportunity to dig into vendor CDR looking for ways to improve routing and identify carrier performance issues. In 2019, the shift focused from vendor performance to customer behavior analytics. At present I spend most of my time looking for bad dudes doing bad things.

Contact Information

If you have any questions about the statements made in this paper please do not hesitate to reach out.

Dean Hansen
+1 613 297 0891 (mobile)
dean@sip40.com
Join us on Discord
LinkedIn