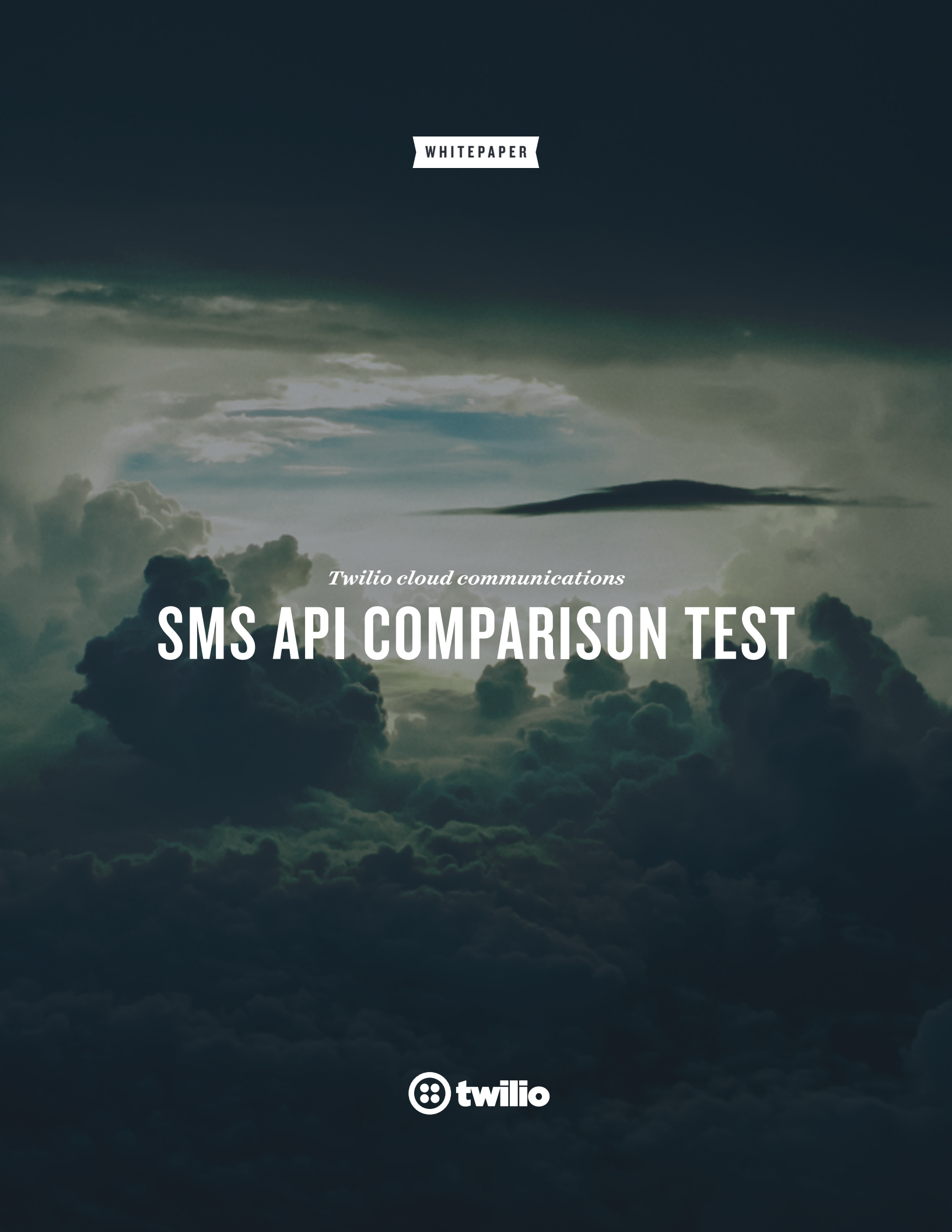

WHITEPAPER
Twilio cloud communications
SMS API COMPARISON TEST
twilio

SMS API COMPARISON TEST

Thanks to cloud communications, SMS, a decades-old technology, is now accessible to the global developer community. The impact can be experienced everywhere: in on-demand marketplaces like Uber and Airbnb and established enterprises like Nike and ING bank. Developers are using cloud-based SMS APIs to reframe how companies communicate with customers and employees.

While sending and receiving a short message using an API is relatively straight-forward with any cloud communications vendor, making it work at scale is far more challenging. You can find yourself mired in country regulations, unreliable delivery receipts, strange handset-specific behavior, carrier filtering, phone number selection, message queuing, and the list spirals on and on. Ultimately, it's the developer who pays the price when trying to accommodate for every corner case. Unlike SMS APIs that simply resell carrier services, Twilio uses software to improve the reliability and to make it easier for developers to operate at scale.

But talk is cheap. We wanted to actually quantify the impact of Twilio's unique software-based approach to SMS. So we carried out a series of tests to see how Twilio stacked up against more conventional SMS APIs. We ran tests in the following three areas:

CONTENT ACCURACY

When you send someone a message, you expect that same message to show up intact on the recipient's device without errors, missing characters, or breaking into multiple messages. With over 100,000 characters representing the world's languages and multiple variants of character encoding and concatenation by global carriers, this is not a trivial undertaking.

DELIVERABILITY

When a message doesn't reach your customer, it can damage marketing campaigns, cause unnecessary support calls, and erode your carefully crafted communication workflow. SMS platforms must have a high degree of platform availability and be resilient against outages across hundreds of mobile carriers worldwide.

OPERATING AT SCALE

It's critical that the SMS platform abstracts away as much development complexity as possible, eliminating the need for telecom and messaging engineering expertise. You won't run into these issues building a proof of concept, but the workload can be massive when you go to scale, secure, and operate that solution globally.

While there is no de facto benchmark for determining what's an "acceptable" level of performance in each category, we lean on our experience with over one million developers to identify some standards; standards that a typical business would expect from a SMS platform.

CONTENTS

1. CONTENT ACCURACY

WHAT IS CONTENT ACCURACY?

By content accuracy, we mean that the message letters, numbers, and characters (content) that you send is the same thing (accuracy) that the recipient receives. The measurement shows how well an SMS platform provider is able to encode the message so that it correctly displays on the subscriber handset.

As a short message makes its journey from your application to the SMS cloud-based platform to the destination network of the user's mobile device, there's quite a bit that can go wrong. And when something goes wrong in that chain, it usually leads to out-of-sequence and garbled messages. Different encoding models used by carriers, inconsistent support for SMS features, and even variation of the phones that are sending and receiving these messages can all affect content accuracy.

CHALLENGES WITH OTHER OFFERINGS?

Most SMS platform providers claim support for standards and encoding of common content. However, what we continue to find and what bears out in this report is that, providers make this claim only for the bounds of their service. They only take responsibility for what happens between receiving a messaging request and sending it to the next hop on the message path.

What people expect when using an SMS platform is something much different. They expect the platform to handle encoding differences for them and mitigate any challenges that may exist between carriers or even phones. The platform needs to contain content intelligence that dynamically adapts to conditions outside of its direct control.

WHAT TWILIO CAPABILITIES ARE WE VALIDATING?

Unlike carrier resellers, Twilio takes a software approach to the problem using SMS content intelligence built into the platform to accommodate for variation in encoding, concatenation, and feature support by carriers as well as handset capabilities. With Twilio, the intention is to allow you to just send everything in UTF-8 and Twilio works behind the scenes to properly encode content and sort out the idiosyncrasies of each carrier so messages arrive intact. And with unicode, every written language, plus emoji, is supported.

WHAT WE TESTED

To examine content accuracy, we looked at how different platforms performed in common and special use cases across networks in North America and Europe. The use cases include general and special character encoding and the handling of concatenated messages — i.e. messages that exceed the number of bytes that can be sent in a single SMS. We tested Twilio and two other well known alternatives to Twilio — we'll refer to as Vendor B and Vendor C.

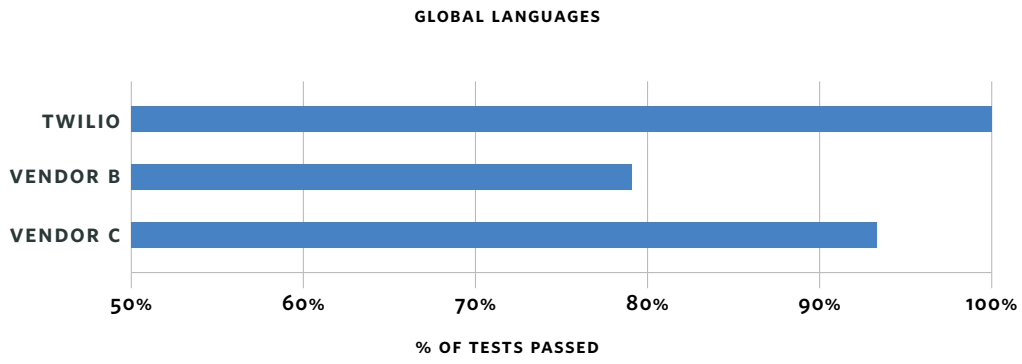
TEST 1: GLOBAL LANGUAGES

An SMS platform should support unicode and accurately represent global languages as they're sent between different geographic locations and across carriers.

For messages to be readable, the general expectation that we see across our customer base is that language characters and global symbols should be accurately represented at least 95% of the time. Anything less and the message may be unreadable.

TESTING RESULTS

In 120 unique different messaging tests to European subscribers, only Twilio displayed correctly each one of the 120 messages on the recipient's phone (100% success).Vendor B resulted in garbled messages in 35 of the 120 tests (79% success); vendor C failed 8 tests (93% success). Characters commonly missed include the euro (€) and greek characters; long messages had trouble with accuracy at the concatenation boundary.



All platforms performed well when it came to accurately receiving and processing the message characters sent to their platform — e.g. sending a message with a euro and other symbols resulted in message logs showing accurate receipt and encoding within the platform. However, the results show that when the message was verified at the handset, it was a much different story and didn't display as expected.

It's worth mentioning, that these results are particularly interesting because we see a significant improvement over legacy SMS offerings whose services are more tightly aligned to telecom infrastructure and not fully adept at handling the encoding differences.

TEST 2: CURRENCY AND CHARACTERS WITH ENCODING CONFLICTS

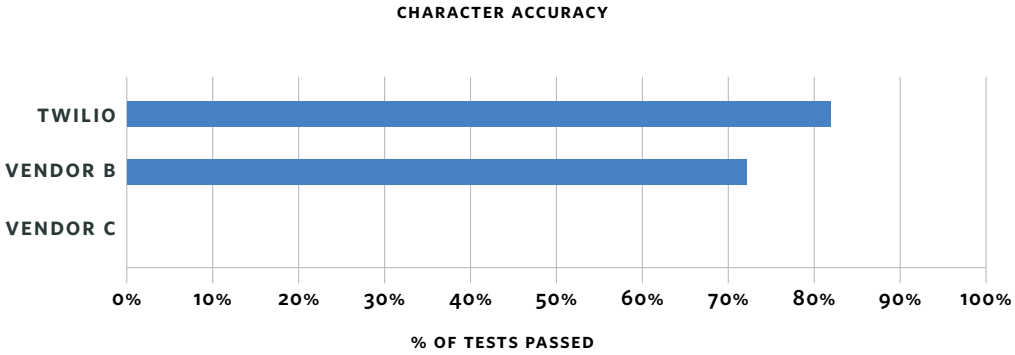
In SMS, there is a set of characters considered particularly difficult to handle. This includes things like currency symbols (e.g. ¥ £ € ﷼ ฿ $) and characters that are not shared between standard SMS character encodings (e.g. where the ASCII and GSM representation are different such as with é ò Ç å Φ ff Ψ Σ Θ ¿ ö ñ).

The expectation is that an SMS cloud communications platform would understand these gaps and make the appropriate translation in software before exchanging the message.

While it's quite common for these symbols to be misrepresented, and maybe not such a big deal in personal messaging conversations, it can cause big problems for businesses when say, a currency symbol is wrong. The general expectation that we see across our customer base is that language characters and global symbols should be accurately represented 75-80% of the time.

TESTING RESULT

In the 'difficult character' tests performed on each platform, Twilio passed 82% of tests, and Vendor B passed 72% of tests.



Vendor C was inconclusive. While it did OK handling well-known strings of difficult characters, it mishandled those exact same characters when they were used in different contexts. It's unclear whether the vendor actually encoded the characters correctly or just hard-coded these exact character strings, which are useless in real world communication.

TEST 3: LONG MESSAGES

Length restrictions are not always obvious to someone sending a text message. And the rules can change based on factors such as the language and characters used in the message. The SMS platform should hide this complexity from the user, seamlessly segmenting and reassembling long messages as needed. People expect both short and long messages to just work, without arriving chopped up and out of order. In fact, long messages are very common in most use cases.

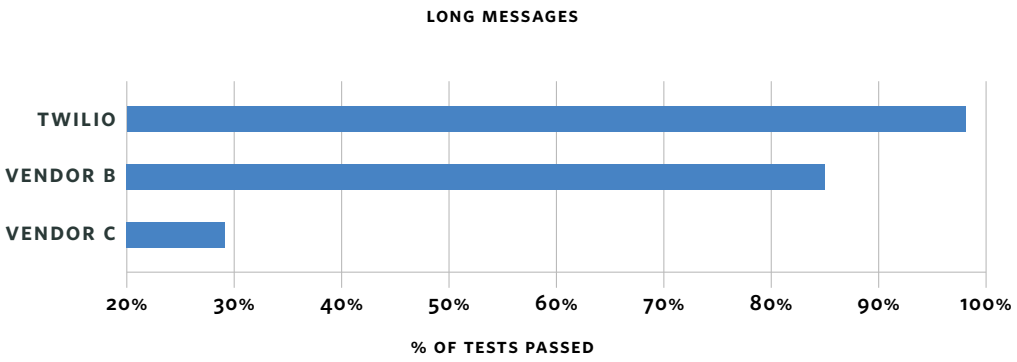
This test examines how well each platform handles long messages, the encoding of content as multiple linked messages, and following the industry standard SMS specification. The long message can then be reassembled as a single message in the right order on the recipient's phone.

We expected each platform to deliver each long message in one delivery instead of several and to avoid any encoding errors at the boundaries of each segment.

The tests included messages with over 700 characters in some cases and other that blended character types, mixing GSM and non-GSM ASCII characters. A non-GSM ASCII character can be something as simple as a backtick (`) and should be handled gracefully.

TESTING RESULT

Twilio correctly encoded and displayed messages 98% of the time. Vendor B had a fairly high error rate, passing only 85% of the tests. Most of the errors happened with message encoding, leaving random and garbled characters at the end of each individual message segment. When the message was reassembled at the end handset, it showed up as random gibberish in the middle of the long message. Vendor C passed 29% of tests, almost always inaccurately encoding the messages.



The big difference in how well each platform handled long messages was surprising considering that the specification for long messages was completed almost two decades ago in 1999.

TAKEAWAYS FROM CONTENT ACCURACY TESTING

Getting encoding right, all the way to the phone, is especially hard because the delivery path is largely outside the direct control of the SMS platform. A platform provider has to accommodate for all the features and idiosyncrasies of operator SMSCs (Short Message Service Centers) as well as the capabilities of the mobile phone itself.

Prior to SMS platforms, customers had to figure out the right way to encode messages on their own, spending developer cycles finding the right combination of SMS flags and byte encoding for each mobile network. Now SMS platforms promise to do that for you, but it's clear that a simple API repackaging of a SMS provider isn't enough. The difference is in the software and how each platform uses it to accommodate for all these variables is what makes or breaks that platform.

Here is a summary of how the platforms performed in each test

TEST	TWILIO	VENDOR B	VENDOR C
GLOBAL LANGUAGES	100%	79%	93%
DIFFICULT CHARACTERS	82%	72%	N/A
LONG MESSAGES	98%	85%	29%

When looking at content accuracy, there is a wide gap in performance between Twilio and others. This likely comes down to the emphasis that Twilio puts on software quality and accommodating for as many global conditions as possible.

2. DELIVERABILITY

WHAT IS SMS DELIVERABILITY?

Deliverability is the likelihood that a message will reach its destination intact. There are many factors at play in regards to deliverability — ranging from infrastructure resiliency to content-based message filtering to handset related issues. Carriers may return delivery receipts but these numbers are notoriously unreliable and only report on one aspect of deliverability.

Deliverability is a complex problem and even more difficult to build simple, repeatable comparison tests for SMS platforms because so many factors can change from minute to minute. The most reliable measurement we could devise for this report was to evaluate different aspects of availability.

GLOBAL AVAILABILITY

This refers to how often the SMS infrastructure is operating without problems and delivering timely messages to every destination where it claims coverage.

PLATFORM AVAILABILITY

Is the length of time the platform API is up and responding normally to requests for message delivery.

One other important factor to consider is phone number availability. Most messaging use cases depend on instant availability of new phone numbers in the regions where they operate.

CHALLENGES WITH OTHER OFFERINGS

Every SMS platform claims high deliverability, but the devil is in the details. There may be persistent routing issues in a region, lack of carrier redundancy, or just plain API downtime. Because most SMS platforms are simply reselling carrier services, they are as reliable as the weakest mobile provider they connect to.

A proof of concept won't reveal the limitations of the platform until you hit scale and by then it's too late. This is when deliverability issues cause the most damage, and can result in a cascade of business-related problems, customer complaints, and revenue loss.

WHAT TWILIO CAPABILITIES ARE WE VALIDATING?

Many of the capabilities that help Twilio increase deliverability are under the covers — functioning when you need them most. Unlike carrier resellers, Twilio uses infrastructure redundancy, a mix of direct and indirect routes and software to orchestrate quality and ensure high deliverability.

While also an important metric, we did not examine deliverability latency — the time it takes between sending a message to a platform and delivery to the mobile phone — because of the technical challenges of creating repeatable and fair tests for comparison.

WHAT WE TESTED

To examine deliverability, we tested uptime and availability of the infrastructure and the underlying messaging service. We focused on tests that had clear metrics and that were repeatable, meaning that a developer could run the testing again and would likely get a similar result. Testing was done across a range of use cases in North America and Europe.

TEST 4: GLOBAL AVAILABILITY

Most SMS platforms can reach users globally, but very few offer predictable delivery in every country. The reality is, temperamental routes exist everywhere and sometimes even stable infrastructure goes down. The platform needs to dynamically recover when routing issues occur.

For example in the US, Twilio is the only vendor to use two SMS aggregators. When aggregator SAP (Sybase) had an outage in 2016, they were unable to route messages to the 4 US Mobile Carriers; Verizon, AT&T, T-Mobile, and Sprint. Twilio automatically routed messages through a second aggregator, Syniverse, and was able to maintain deliverability. Every other SMS API, except Twilio, was hard down with no messages sent or received for roughly 4 hours.

Twilio has built software to monitor the downstream aggregators and to dynamically change routing. This intelligence, coupled with more carrier interconnects, improves global delivery.

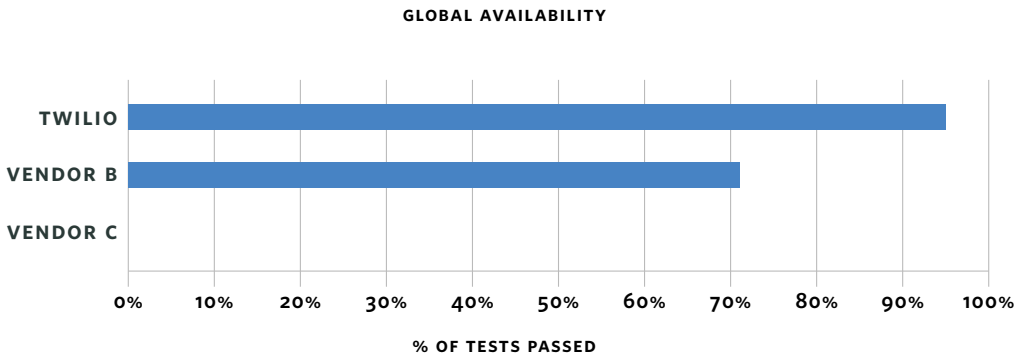
Other vendors don't follow this redundant routing approach because it requires significant software engineering resources to implement dynamic inspection and routing. Additionally, it costs more per

message. Splitting messaging traffic equates to lower volume through each aggregator and therefore a higher price per SMS.

In this test, we are measuring how often an SMS platform delivers to every country where it claims coverage. To remove any subjectivity, we measured availability by tracking each vendor's status reporting.

TESTING RESULT

In a recent six-month period, Twilio's global availability (i.e. the time the platform was not having issues delivering messages to any destination) was 95%. Over the same period, Vendor B had issues 29% of the time (71% global availability). The total time Vendor B's SMS network was having problems was greater than 50 days. This included 49 SMS outages to 31 countries with the average outage lasting over 26 hours. We were unable to obtain reliable data over the period examined for Vendor C. This was due to inconsistencies found in Vendor C's reporting vs. actual outages.



The results here highlight the effectiveness of Twilio's multiple-path, direct and indirect connectivity strategy for SMS deliverability.

TEST 5: PLATFORM AVAILABILITY

A basic metric for any cloud-based platform is the percentage of time the API is reachable and responding normally to requests. Practically speaking, when a SMS platform is not reachable, you are not able to send a text message.

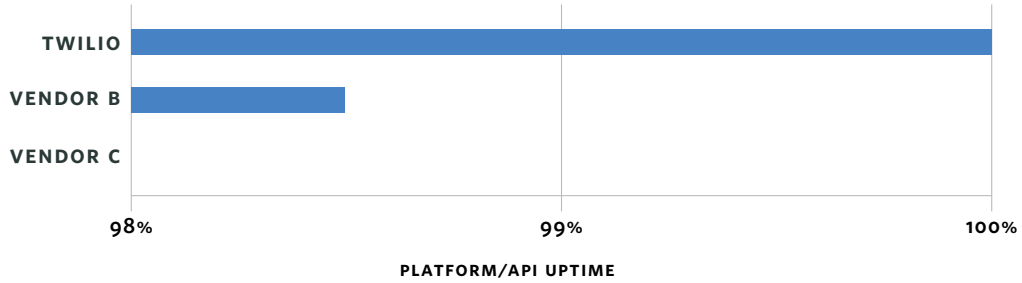
To measure availability we analyzed published outage reports over the same six-month period. The research was conducted during most of this time frame, letting us validate first-hand the correspondence between reporting and actual availability.

TESTING RESULT

Twilio's SMS API availability during the six month test period was greater than 99.999% equating to under two minutes of downtime. Five nines is a well-known operational standard and places Twilio on par with leading cloud platform providers across industries.

During the same time period, Vendor B's SMS API availability was approximately 98.5%. Vendor B experienced availability issues for a combined total of 48 hours and 29 minutes. To their credit, Vendor B was responsive in communicating their outages.

Due to observed inconsistencies between actual and reported downtime by Vendor C during testing, we decided that historical information was unreliable and removed them from this test. Specifically, during the period we were measuring downtime, there were two concerns. 1) They didn't report scheduled downtime as actual "downtime" and 2) they didn't report all of the unscheduled downtime we experienced.



TEST 6: PHONE NUMBER AVAILABILITY

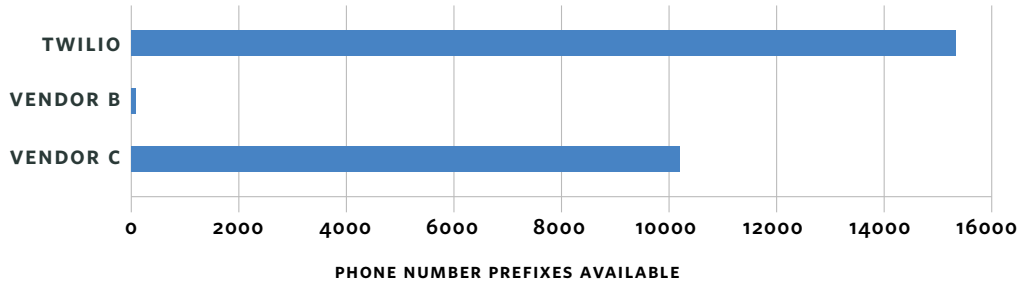
Another function of availability is how many phone numbers are actually available in a region. Some refer to this as phone number "depth". If you have a large group of customers in one area, but can't provision enough phone numbers using that local prefix, then a platform won't serve your needs.

We also found some vendors that claimed to have availability of mobile phone numbers, not only had just one area code available in that country, but only had 10-15 actual phone numbers available! Clearly, they were just providing the bare minimum in order to claim support in that country and were not able to support a real scalable business need in that region. We did not include this test in our report because it is variable day-by-day and difficult for someone to independently verify.

For this test, we looked at phone number availability in the US and in a sample of European and Asian countries

TESTING RESULT

For US phone number availability, the results varied dramatically. Twilio offers 15,335 phone number prefixes (e. g. in the US that refers to the first three digits after the area code). Vendor B offers only 39 and vendor C offers 10,211.



For global phone number availability, we tested a sample of seven countries and every major region in the US. In these tests Twilio and Vendor B each had at least one mobile phone number area code in four countries. Only Twilio offered multiple area code options in countries outside the US, specifically in France and Germany.

TAKEAWAYS FROM DELIVERABILITY TESTING

A lot has to go right so that every single one of your messages reaches its destination. It's not just about access to carrier connections — it's also about how your message is managed across those connections.

Differences in deliverability and overall platform quality are often reflected in pricing, because the cost structure of high deliverability requires things like dual-peering to carriers and extensive investment in software to orchestrate failover. Software can not only react to routing issues but can use collective intelligence from delivering billions of messages across the world to act predictively, adjusting routes, and avoiding problems before they happen.

Previous to SMS platforms, customers had to establish their own peering relationships with global carriers and monitor and optimize routes on their own. Now, we expect the SMS platform to handle this for us and select the best route to deliver each and every message. Here is a summary of how each platform did.

TEST	TWILIO	VENDOR B	VENDOR C
GLOBAL SMS AVAILABILITY	95%	71%	N/A
PLATFORM AVAILABILITY	99.999%	98.5%	N/A
NUMBER PREFIX AVAILABILITY	15,335	39	10,211

Another benefit not mentioned is that redundant infrastructure and dual-peering can also improve delivery feedback information. For example, most APIs will indicate a message is "sent" when the intermediary aggregator receives the message. The Twilio API instead reports the delivery status from the mobile carrier.

3. OPERATING AT SCALE

WHAT DO WE MEAN BY OPERATING AT SCALE?

Using an SMS API needs to be accessible to every web and mobile developer. Regardless of whether you're building a contact center chat experience or something a bit more straightforward like appointment reminder notifications. Operating at scale means making it easy for the developer to go to production, with a wide range of tools, helper libraries, documentation, and support services. This way you can build at scale and avoid getting stuck in the process. It also looks at what common workflows are abstracted from the developer, so you don't have to code things like opt-in logic and phone number selection. Instead you can focus on building and iterating the right solution for the task.

CHALLENGES WITH OTHER OFFERINGS

Every SMS platform will make it easy to use their HTTP SMS APIs to send a message in a matter of minutes and even to build a basic proof of concept. However, the measure here is how well the SMS API can be used for building, scaling, and operating at production.

WHAT WE TESTED

The testing in this section uses more subjective evaluation, but we also attempted to find specific features and capabilities that contribute to a good developer experience.

TEST 7: API RESPONSE

The APIs from all three SMS platforms were tested against both basic and more complex use cases. The goal was to understand the ease of use for developers and the utility of the API.

In this test, we looked at the responses returned from all three platforms when we submitted SMS. All three platforms support JSON and XML response formats and follow proper REST organization of HTTP status codes. However, the content of the response (i.e. the choice of return parameters in the response body) was inconsistent and takes developers significantly more time to work with.

TEST RESULTS

In general, we found the API Response from the three platforms to be:

TWILIO	★★★★★
VENDOR B	★★★☆☆
VENDOR C	★☆☆☆☆

Twilio was designed to have obvious response elements and content (e.g. 'status' and 'message' elements for both success and error responses) and a consistent response body format that is more intuitive to understand.

Vendor B's response elements and content were OK, but they used different formats in their response body depending on what happened with the message. An issue with this approach is that the developer doesn't know ahead of time what the particular outcome will be and therefore has to code for the presence or absence of elements in the response body in order to take the correct action. This is unnecessarily cumbersome and error-prone.

Vendor C chose the least developer-friendly path using differing response body formats depending on request outcome and also using non-obvious or unintuitive elements and content — e.g. 'status: 7'— requiring the team to write additional code to handle their API and semantic idiosyncrasies.

TEST 8: OPERATIONAL USABILITY

A critical element of developing and operating successful SMS programs is having insight into message outcomes. Typically, this includes a combination of log access, alerts, and dashboards.

All three platforms provide access to message logs via a customer dashboard and through call back handlers as part of their APIs. Similarly, exporting results to csv was readily available and simple for each of the platforms. However, our research revealed differences between the usefulness of the tools to give customers real insight into their messaging activity.

In this case, we tested the ability of each platform to search the account message history using query parameters a person would naturally expect for messaging: message status, date range, sender, receiver, etc. While vendors will differ in how much history they make available to their customers, providing the ability to search that history with the above parameters will cover the great majority of use cases for customers wanting to understand what is happening with their messages.

TESTING RESULTS

For operational usability, we rated the three platforms as:

TWILIO	★★★★★
VENDOR B	★★☆☆☆
VENDOR C	★★★★☆

Twilio and Vendor C provided the expected detail and search functionality, allowing search by date or date range, status, sender, recipient, and other relevant fields. That said, they differed in the user experience they provided.Twilio consistently requires less actions to obtain the same information.

Vendor B's approach required the particular phone number or message identifier in order to display message information. This can cause some confusion because the message identifier is a vendor-specific artifact and has no business-related meaning. Having a platform identifier is convenient for the vendor to troubleshoot issues, but isn't intuitive to the user. Only Vendor B made this a primary means while excluding other expected search parameters.

Another point of frustration was that Vendor B restricts searches for a recipient's phone number to one phone number per day. When multiple phone numbers were not receiving the correct message there was no simple way to see the set of messages that were having issues.

TEST 9: DOCS AND EXTERNAL RESOURCES

We know that documentation matters. It guides your work on any software project, prevents you from spiraling out of focus, and can make or break the final product. Important documentation should include detailed API reference documentation, use case tutorials, quick start guides, and other content to help with development.

Not only is it important to have a depth of resources maintained by the vendor, but it's also nice to have support from the greater developer community which is why we also looked at major developer watering holes like Stack Overflow and GitHub.

TESTING RESULTS

TWILIO ★★★★★

VENDOR B ★★☆☆☆

VENDOR C ★★☆☆☆

	/DOCS (API DOCS, QUICK STARTS, TUTORIALS, ETC.)	DISCUSSION RESULTS ON STACK OVERFLOW	REPRO RESULTS ON GITHUB
TWILIO	2,380	8,228	5,125
VENDOR B	135	315	364
VENDOR C	146	313	335

Even if we look at just the FAQ information at support.twilio.com (which is not included in the tally above), it would provide 3x the level of information that Vendor B & C offer across their entire docs site.

TAKEAWAYS FROM OPERATING AT SCALE TESTING

In comparison with Twilio, vendors B and C lack the features or software tools that overcome many of the inherent difficulties in creating large, global SMS programs. Issues such as consistent identity, optimized routing, sender scaling, and failover are all problems that large, global SMS programs contend with. Twilio addresses these in software, yielding two important results: first, this translates to more time software developers are spending building the business-relevant features into the program instead of working on solutions to SMS-specific issues; next, because these features are a part of the platform and are in wide use by customers around the world, it provides a high-degree of trust that these solutions have been well-thought through and that the services themselves are dependable.

APPENDIX

EVALUATION PROCESS

The rise of messaging as a core service for businesses has led to a number of recent entrants into the SMS API space. To ensure our research accurately reflected the best alternatives in the market, we selected companies for our evaluation based on:

LONGEVITY

Each company evaluated has been in business for over three years. Choosing companies with an established business history lets us evaluate platforms and user experiences that are mature, thus removing the feature and performance instability inherent in early-stage companies that could sway the overall results.

GLOBAL REACH

We chose companies that have global reach in their platform because of the increasing needs of businesses to work wtih clients, employees and partners in every corner of the world. Additionally, choosing global platforms — and not regional specialists — allowed us to evaluate sophistication evenly between providers.

PROGRAMMATIC (API) DRIVEN MESSAGING PLATFORM

Contemporary mobile messaging happens at a scale and speed that necessitates programmatic access in order for customers to achieve a reasonable ROI on their messaging programs. We chose API-based companies because this is the most pragmatic means a business has when implementing messaging beyond a trivial scope.

To maintain focus on the results, and out of respect for the competing providers, the names of evaluated platforms have been obscured.

METHODOLOGY

Each test was designed to control for variables known to affect potential outcomes and to minimize confounding factors. For example, when testing performance in a particular region, we programmatically ran that test across all platforms simultaneously to the same set of destination networks and phones. This minimized the chance that differences between platforms were due to time-based factors (severe network congestion, for example) or differences in the test content or API calls. We further tested multiple mobile networks per region and multiple phones per mobile network to control for differences peculiar to specific factors in the message path.

In an effort to accurately measure the real-world experience of customers, SMS testing was conducted to physical phones instead of virtual numbers, sims, or emulators.

For certain deliverability and availability analyses, data was obtained directly from the platforms using their internal customer tools and scraping publicly available performance metrics. All platforms were evaluated for the same six-month period of time to control for outside factors.