

FOR RELEASE NOVEMBER 29, 2017

# Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates

*Fully 57% of comments used temporary or duplicate email addresses, and seven popular comments accounted for 38% of all submissions*

**BY** Paul Hitlin, Kenneth Olmstead and Skye Toor

**FOR MEDIA OR OTHER INQUIRIES:**

Lee Rainie, Director, Internet and Technology Research  
Aaron Smith, Associate Director, Research  
Tom Caiazza, Communications Manager

202.419.4372

[www.pewresearch.org](http://www.pewresearch.org)

**RECOMMENDED CITATION**

Pew Research Center, November, 2017, "Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates"

## About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping America and the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at [www.pewresearch.org](http://www.pewresearch.org). Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

© Pew Research Center 2017

## Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates

*Fully 57% of comments used temporary or duplicate email addresses, and seven popular comments accounted for 38% of all submissions*

*Correction: This report initially noted that 450,000 comments were submitted to the FCC during its previous open comment period on net neutrality. That data point was based only on the initial comment period, spanning Feb. 9–July 18, 2014. The FCC subsequently reopened the comment period through Sept. 15, 2014, and the report now reflects the total number of comments received during the entirety of the 2014 public comment period. In addition, a reference to John Oliver in a sentence referring to the most popular pro-net-neutrality comment has been removed. Pew Research Center has issued a [statement](#) regarding concerns raised about this analysis.*

For the second time in less than four years, the U.S. Federal Communications Commission (FCC) is considering regulations regarding net neutrality – the principle that internet service providers must treat all data the same, regardless of the origin or purpose of that data. Opponents of net neutrality regulations argue that ISPs should have the right to prioritize traffic and charge for their services as they wish. Meanwhile, supporters of net neutrality suggest that so-called fast lanes are anti-competitive and would prevent start-ups and smaller companies from competing with more well-established companies that can afford to pay for prioritized web traffic.

From April 27 to Aug. 30, 2017, the FCC allowed members of the public to formally submit comments on the subject. In total, 21.7 million comments were submitted electronically and posted online for review. This figure dwarfs the number received during the [initial comment period](#) when the FCC last accepted comments on this topic in 2014, as well as the nearly four million [total submissions](#) received during the entirety of the comment process that year.<sup>1</sup> Net neutrality regulations underpin the digital lives of many Americans, yet it is challenging to survey the public on such an inherently complex and technical subject. For this reason, Pew Research Center set out to analyze the opinions of those who had taken the time to submit their thoughts to the FCC.

However, the Center’s analysis of these submissions finds that the comments present challenges to anyone hoping to understand the attitudes of the concerned public regarding net neutrality. It also

---

<sup>1</sup> The 2014 Pew Research Center study was released after the initial comment period was closed and the initial data had been made available to the public. The FCC subsequently extended the comment period. Additionally, the FCC allows for submissions by phone or letter, but those comments are not publicly accessible and are excluded from this report.

highlights the ways in which individuals and groups are using modern digital tools to engage in the long-standing practice of speaking out in order to influence government policy decisions. Among the most notable findings:

- **Many submissions seemed to include false or misleading personal information.** Some 57% of the comments utilized either duplicate email addresses or temporary email addresses created with the intention of being used for a short period of time and then discarded. In addition, many individual names appeared thousands of times in the submissions. As a result, it is often difficult to determine if any given comment came from a specific citizen or from an unknown person (or entity) submitting multiple comments using unverified names and email addresses.
- **There is clear evidence of organized campaigns to flood the comments with repeated messages.** Of the 21.7 million comments posted, 6% were unique. The other 94% were submitted multiple times – in some cases, hundreds of thousands of times. In fact, the seven most-submitted comments (six of which argued against net neutrality regulations) comprise 38% of all the submissions over the four-month comment period.
- **Often, thousands of comments were submitted at precisely the same moment.** On nine different occasions, more than 75,000 comments were submitted at the very same second – often including identical or highly similar comments. Three of these nine instances featured variations of a popular pro-net-neutrality message, while the others promoted several different anti-net-neutrality statements.

The Center conducted its analysis by downloading all the comments from the [FCC's publicly available API](#). All data and comments used in this report are stored on the FCC's site and are freely available to the public. Researchers then used various data analysis techniques to summarize the comments and to look for duplicates or invalid information. Most notably, the Center utilized a measure of textual similarity to determine the share of highly similar comments that were submitted multiple times.<sup>2</sup> Full details of the contents of this dataset and the techniques used in this analysis can be found in the [methodology](#) at the end of this report.

## Many submissions contained false or misleading personal information

Collecting large-scale data from the public is always challenging. It is difficult to ensure that a person online is indeed who they claim to be, and falsification of someone's personal information can be accomplished with relatively minimal effort. The Center's analysis finds evidence that many

---

<sup>2</sup> The analysis used is known as cosine similarity, which measures the distance between characters in different documents. Throughout this report (unless explicitly noted) comments with a cosine similarity of at least .95 on a scale from 0-1 are grouped together and considered the same.

people did not reveal their true identities when submitting comments to the FCC. Some of these instances may have been accidental, but in many cases patterns in the comments indicate those submitting the comments intentionally entered false or misleading personal information.

### **Many common names – as well as other words – appeared thousands of times as “authors” of comments**

The most common “name” included as an author was not, in fact, a name. In nearly 17,000 instances, the name of the commenter filing their views on the FCC site was written as “Net Neutrality” (this term also appeared as the author of more than 5,000 comments in lower-case form). “The Internet” also appeared as the name in almost 7,500 submissions. Of the top 15 names that appeared in the FCC submissions, eight included the common last names of “Smith” or “Johnson,” and four were not names at all.

### **These submissions often featured email addresses that were nonfunctional, frequently repeated, or disposable**

In theory, the process for submitting a comment to the FCC included a validation technique to ensure the email address submitted with each comment came from a legitimate account. The submission form clearly states that all information submitted, including names and addresses, would be publicly available via the [FCC site](#).

---

### **Thousands of submissions to FCC featured duplicate ‘names’**

*Top 15 most-common names provided on posts submitted to the FCC’s net neutrality comment website*

	<b>Name listed on the FCC’s comment site</b>	<b>Number of submissions</b>
1	Net Neutrality	16,983
2	The Internet	7,470
3	Pat M	5,910
4	net neutrality	5,153
5	John Smith	3,997
6	John Johnson	3,133
7	Smith	3,121
8	Michael Smith	2,708
9	Mary Johnson	2,635
10	James Smith	2,625
11	David Smith	2,576
12	Robert Smith	2,442
13	17-108	2,439
14	Michael Williams	2,237
15	MyNameJeff	2,167

Source: Pew Research Center analysis of online submissions to the FCC’s net neutrality public comment form from April 27 to Aug. 30, 2017.

“Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates”

**PEW RESEARCH CENTER**

---

However, the Center’s analysis shows that the FCC site does not appear to have utilized this email verification process on a consistent basis. According to this analysis of the data from the FCC, only 3% of the comments definitively went through this validation process. In the vast majority of cases, it is unclear whether any attempt was made to validate the email address provided.

As a result, in many cases commenters were able to use generic or bogus email addresses and still have their comments accepted by the FCC and posted online. For instance, the email address *example@example.com* appeared in 7,513 comments, making it the most common email address to appear. The email address *john\_oliver@yahoo.com* (television host John Oliver advocated on his [show](#) for net neutrality earlier this year) was also used in 1,002 comments. All told, the Center's analysis identified 1.4 million email addresses that appeared multiple times in the comments.

Additionally, in 9,190 cases the email address supplied did not contain the "@" character necessary to serve as a functioning email account. Moreover, 10% of the comments submitted did not include an email address at all.

Along with using duplicate or potentially fraudulent addresses, the Center's analysis finds more than 8 million submissions included email addresses from temporary email accounts designed to disappear within hours and leave no trace of email exchanges behind.<sup>3</sup> Taken together, some 57% of the comments submitted to the FCC either utilized a temporary email address or an email address that was also included with at least one other comment.

## Many submissions highlight organized efforts to influence the commentary period

The Center's analysis of these data suggests the net neutrality comment period was marked by several organized efforts aimed at conveying the public's feelings on this subject.

### **Some 6% of the comments posted were unique submissions. Six of the seven most-common submissions in the remaining 94% argued against net neutrality and can be traced back to websites of a handful of organizations**

This analysis finds that 6% of the 21.7 million comments were submitted a single time. The remaining 94% were each submitted multiple times, in some cases numbering in the thousands. In fact, five comments were submitted more than 800,000 times each. Taken together, these seven comments alone account for more than 8 million submissions, representing 38% of the total over the entirety of the comment period.

---

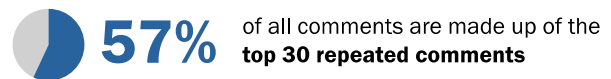
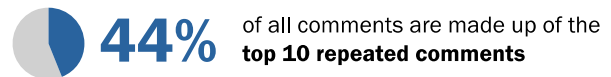
<sup>3</sup> The Center identified disposable email addresses by matching their domains to a list of known providers of temporary or disposable email accounts. The specific domains include @pornhub.com, as well as the 10 domains provided as an option by the site FakeMailGenerator.com (which include @gustr.com and @armyspy.com, among others). Other sites, such as 10minutemail.com, also offer disposable accounts. But these sites utilize random domain names that cannot be tracked and as a result there was no way to identify these accounts for the purposes of this analysis.

The single comment submitted more times than any other was a pro-net-neutrality statement that appeared 2.8 million times, accounting for 13% of all submissions. At the same time, seven of the top 10 comments argued *against* net neutrality and encouraged the FCC to roll back Title II regulations.<sup>4</sup> The seven most-popular anti-net-neutrality posts made up 27% of all the comments submitted, while the three most-popular comments in favor of net neutrality made up 17% of the total submitted.

Whether they argued for or against net neutrality, the text of many of the top comments can be traced back to a small number of organizations. For example, the single most-popular comment was a pro-net-neutrality statement that appeared as a submission form on the website [battleforthenet.com](http://battleforthenet.com). Similarly, the wording for three popular comments opposing net neutrality (representing the second-, sixth- and ninth-most submitted overall) appeared on the website for an organization known as the [Taxpayers Protection Alliance](#). Combined, the text from these three suggested comments appeared in almost 2.4 million submissions, making up 11% of the total.

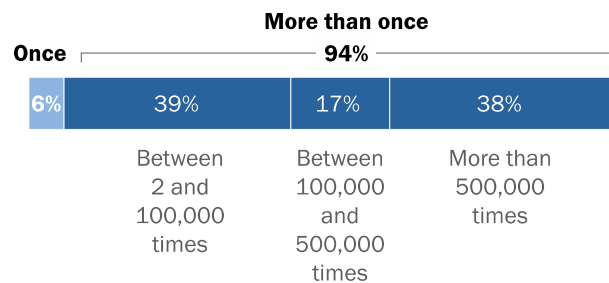
## Small number of FCC net neutrality comments made up a large portion of the total submitted

*% of FCC net neutrality comments in each group*



## Most appeared more than once; only 6% of comments were unique

*% of FCC net neutrality comments that appeared ...*



Source: Pew Research Center analysis of online submissions to the FCC's net neutrality public comment form from April 27 to Aug. 30, 2017.

"Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates"

PEW RESEARCH CENTER

<sup>4</sup> The shorthand "Title II" is often used to refer to how the FCC implemented net neutrality regulations in 2014. The FCC reclassified internet providers as common carriers under Title II of the Communications Act of 1934. This decision allowed the FCC to implement net neutrality regulations.

## The seven most-prevalent comments to the FCC were submitted more than 500,000 times each, and six opposed net neutrality regulations

*Seven most-common submissions to the FCC's net neutrality comment form*

Beginning of common text	Number of times submitted
1 The FCC's Open Internet Rules (net neutrality rules) are extremely important to me. I urge you to protect them. I don't want ISPs to have the power to block websites, slow them down, give some sites an advantage over others, or split the Internet into "fast lanes" for companies that pay and "slow lanes" for the rest. Now is not the time to let giant ISPs censor what we see and do online. Censorship by ISPs is a serious problem. Comcast has throttled Netflix, AT&T blocked FaceTime, Time Warner Cable throttled the popular game League of Legends and Verizon admitted it will introduce fast lanes for sites that pay – and slow lanes for everyone else – if the FCC lifts the rules. This hurts consumers and businesses large and small ...	2,803,359
2 In 2015 Chairman Tom Wheeler's Federal Communications Commission (FCC) imposed restrictive Title II utility-style regulations under the guise of an open internet. Not only have these regulations inhibited innovation in the internet ecosystem, they hurt taxpayers and consumers by expanding the regulatory reach of the FCC and limiting investment in internet infrastructure. We cannot allow this revolutionary tool to be bogged down with excessive government interference. It is past time for the FCC, an agency that is funded by American taxpayers, to free the internet of burdensome regulations ...	1,306,644
3 Before leaving office the Obama Administration rammed through a massive scheme that gave the federal government broad regulatory control over the internet. That misguided policy decision is threatening innovation and hurting broadband investment in one of the largest and most important sectors of the U.S. economy. I support the Federal Communications Commission's decision to roll back Title II and allow for free market principles to guide our digital economy.	1,270,140
4 The current FCC regulatory scheme known as "Title II" represents an unprecedented increase in government control over the internet. Such over-regulation is hurting our economy and suffocating innovation. I support Chairman Pai's plan to return to a commonsense regulatory framework that allows for the internet to grow without useless government interference. The internet has flourished for decades without the heavy hand of government over-regulation. Its time we return to what works.	898,633
5 The unprecedented regulatory power the Obama Administration imposed on the internet is smothering innovation, damaging the American economy and obstructing job creation. I urge the Federal Communications Commission to end the bureaucratic regulatory overreach of the internet known as Title II and restore the bipartisan light-touch regulatory consensus that enabled the internet to flourish for more than 20 years ...	818,832
6 As a concerned taxpayer and consumer I am writing to urge the FCC to set the internet free and remove the inappropriate, unnecessary and overly vast regulations currently holding back the full potential of the internet. Due to the grip of the utility-style regulations imposed under the previous Commission, taxpayers have been put at risk, the threat of new fees on consumer bills still looms large, investment in internet infrastructure has not realized its full potential, innovations have gone undeveloped and unrealized, and twenty years of the appropriate level of oversight of the internet has been reversed ...	657,018
7 The Obama-era FCC regulations known as "Title II" enable the federal government to exert an extraordinary and unnecessary amount of regulatory control over the internet. This bureaucratic overreach impedes innovation, stifles investment and continues to create economic uncertainty for one of the largest sectors of the U.S. economy ...	525,007

Source: Pew Research Center analysis of online submissions to the FCC's net neutrality public comment form from April 27 to Aug. 30, 2017. "Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates"

PEW RESEARCH CENTER

**In many instances, thousands of comments were submitted simultaneously – down to the second.**

[Other research](#) has suggested that some share of the FCC comments may have been submitted in bulk using automated processes, such as organized bot campaigns. The Center's analysis finds support for this argument, based on the fact that many comments were submitted at precisely the



same instant. The FCC assigned a precise timestamp to each comment as it was submitted, and an analysis of those timestamps shows that on numerous occasions, thousands of posts were submitted at exactly the same time – a sign that these submissions were likely automated.

On more than 100 different occasions, 25,000 or more comments were submitted to the FCC at the same precise second. And on nine different occasions, 75,000 messages or more were posted simultaneously. The three most numerous of these nine moments featured variations of the most popular pro-net-neutrality message. The remaining six included several different anti-net-neutrality statements.

In the most prolific example, 475,482 comments were submitted on July 19 at precisely 2:57:15 p.m. EDT. Almost all of those comments were pro-net-neutrality and offered variations of text that appeared on the site [battleforthenet.com](http://battleforthenet.com). In some cases, the only difference was the name of the submitter: the same text was “signed” 286 times by “Andrew,” 265 times by “Michael” and 235 times by “Ryan,” among other names.

A deeper analysis of these simultaneous comments highlights several variations in how they were submitted. In some cases, the comments were highly similar but with minor variations. The 86,237 comments submitted at precisely 7:18:04 p.m. on May 24 offer an example of this approach. No two were exactly the same, but all featured consistent patterns. Most began with variations of a similar theme, such as: “Dear [FCC Chairman] Mr. Pai, I am a voter worried about regulations on the Internet,” “Dear Chairman Pai, I am a voter worried about Title 2 and net neutrality,” or “Dear Commissioners: I’m concerned about Internet regulation and net neutrality.”

The body of these comments also featured similar phrases. One post charged, “Obama’s policy to take over the web is a betrayal of net neutrality. It reversed a free-market policy that functioned supremely well for decades with both parties’ backing.” While another stated, “The previous

---

### More than 75,000 net neutrality comments were submitted simultaneously on nine different occasions

*Number of FCC net neutrality comments*

	Date	Time (EDT)	Number of comments
1	July 19	2:57:15 p.m.	475,482
2	July 17	4:53:08 p.m.	129,281
3	July 17	4:39:21 p.m.	90,673
4	May 28	8:23:51 p.m.	90,458
5	May 24	7:18:04 p.m.	86,237
6	May 18	8:34:00 p.m.	86,237
7	May 24	7:09:20 p.m.	81,837
8	May 18	8:53:00 p.m.	80,479
9	May 23	6:12:57 p.m.	76,512

Source: Pew Research Center analysis of online submissions to the FCC’s net neutrality public comment form from April 27 to Aug. 30, 2017.

“Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates”

PEW RESEARCH CENTER

---

administration's policy to control the Internet is a betrayal of the open Internet. It disrupted a free-market system that functioned fabulously smoothly for decades with bipartisan approval.”

In other cases, the content of these simultaneous submissions was entirely identical. On May 28 at exactly 8:23:51 p.m. EDT, the FCC received 90,458 comments with this exact message: “Title II is a Depression-era regulatory framework designed for a telephone monopoly that no longer exists. It was wrong to apply it to the Internet and the FCC should repeal it and go back to the free-market approach that worked so well.” Indeed, this example was not an isolated incident. The Center identified at least five separate occasions when the exact same text was submitted more than 24,000 times at precisely the same moment.

---

### **On at least five occasions, more than 24,000 exactly identical net neutrality comments were submitted to the FCC in a single second**

*Number of FCC net neutrality comments*

	<b>Date</b>	<b>Time (EDT)</b>	<b>Number of submissions</b>	<b>Beginning of comment text</b>
1	July 17	4:53:08 p.m.	120,425	The FCC's Open Internet Rules (net neutrality rules) are extremely important to me. I urge you to protect them ...
2	May 28	8:23:51 p.m.	90,458	Title II is a Depression-era regulatory framework designed for a telephone monopoly that no longer exists. It was wrong to apply it to the Internet and the FCC should repeal it and go back to the free-market approach that worked so well.
3	July 17	4:39:21 p.m.	84,731	The FCC's Open Internet Rules (net neutrality rules) are extremely important to me. I urge you to protect them ...
4	May 26	3:25:35 p.m.	74,810	Obama's Net Neutrality order was the corrupt result of a corrupt process controlled by Silicon Valley special interests. It gives some of the biggest companies in the world a free ride at the expense of consumers and should be immediately repealed!
5	May 18	8:53:00 p.m.	24,163	In 2015, wealthy leftist billionaires and powerful Silicon Valley monopolies took the internet out of the hands of the people and placed it firmly under the thumb of the federal government, monopolies like Google and global billionaires like George Soros.

Note: Data here only include matches with a cosine similarity value of 1 on a scale of 0-1.

Source: Pew Research Center analysis of online submissions to the FCC's net neutrality public comment form from April 27 to Aug. 30, 2017. "Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates "

**PEW RESEARCH CENTER**

---

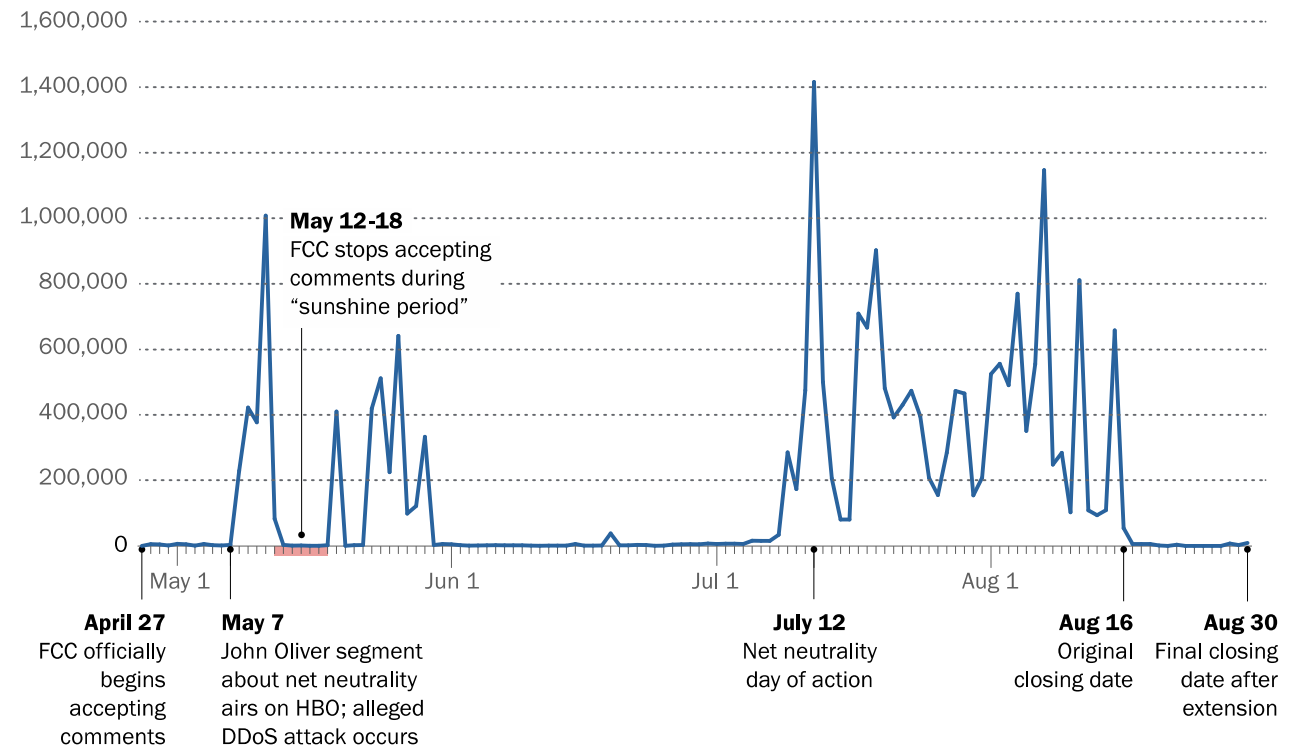


## The comment period was marked by bursts of intense activity and long stretches with few submissions

During the four-month period in which the FCC accepted comments on net neutrality, an average of 172,246 posts were submitted per day. But the comment period featured several long stretches with few submissions, punctuated by bursts of intense activity.

### FCC net neutrality comment period marked by bursts of intense activity, long periods of inactivity

Number of comments per day, April 27-Aug. 30, 2017



Source: Pew Research Center analysis of online submissions to the FCC's net neutrality public comment form. "Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates"

PEW RESEARCH CENTER

The comment period officially opened on April 27, and only 453 comments were submitted on that day. On Sunday, May 7, two major events occurred that coincided with a significant increase in submissions. That evening, comedian John Oliver broadcasted a nearly [20-minute segment](#) on his

HBO show Last Week Tonight defending net neutrality and encouraging his viewers to submit comments supporting his position. The last time the FCC considered net neutrality in 2014, a [Pew Research Center analysis showed](#) that John Oliver’s program also led to a spike in the number of comments submitted.

Also on May 7, the FCC issued a [news release](#) stating that a distributed denial of service attack (DDoS) occurred against the electronic filing system. Some critics have questioned whether an actual DDoS attack occurred, noting that the FCC did not [provide documentation](#) regarding the attack following a Freedom of Information Act [request](#) by the website Gizmodo. And two Democratic members of the House Energy and Commerce Committee have since [requested an investigation](#) into the matter.

More than 2.1 million comments were submitted in the five days following those two events (May 8-12). Those comments made up 10% of all the comments submitted during the entire submission period.

In response to this surge of submissions, the FCC released a [public notice](#) on May 11 that announced a “sunshine period” for the week spanning May 12-18 in which the FCC would temporarily stop taking public comments due to the large number of submissions. According to the FCC’s statement:

“This means that during this brief period of time, members of the public cannot make presentations to FCC employees who are working on the matter, and are likely to be involved in making a decision on it, if the underlying content of the communication concerns the outcome of the proceeding ... The Commission adopted these rules to provide FCC decision-makers with a period of repose during which they can reflect on the upcoming items.”

Although the FCC claimed it would not accept comments during this period, the Center’s analysis finds that more than 93,000 posts submitted on those days were included among the final database made available for public review.

The rate of comments slowed significantly over the next few weeks. From May 30 to July 8, the number of comments declined to an average of only 5,832 posts per day. In mid-July, activity increased dramatically and remained relatively high until the original date the comment period ended.

The single day with the most submissions occurred on July 12. Online activists dubbed the day “Net Neutrality Day of Action” or “Day of Action to Save Net Neutrality” and numerous sites altered their websites to include statements favoring net neutrality. On that day alone, 1.4 million comments were submitted electronically to the FCC.

## Acknowledgments

This report was made possible by The Pew Charitable Trusts. It is a collaborative effort based on the input and analysis of the following individuals:

### Primary researchers

Paul Hitlin, *Senior Researcher*

Kenneth Olmstead, *Research Associate*

Skye Toor, *Data Science Assistant*

### Research team

Lee Rainie, *Director, Internet and Technology Research*

Aaron Smith, *Associate Director, Research*

Claudia Deane, *Vice President, Research*

Solomon Messing, *Director, Data Labs*

Brian Broderick, *DevOps Engineer*

Adam Hughes, *Research Associate*

Andrew Perrin, *Research Analyst*

### Editorial and graphic design

Margaret Porteus, *Information Graphics Designer*

Shannon Greenwood, *Copy editor*

Peter Bell, *Design Director*

### Communications and web publishing

Tom Caiazza, *Communications Manager*

Shannon Greenwood, *Associate Digital Producer*

## Methodology

This study included all online comments submitted to the Federal Communications Commission (FCC) regarding the proposal called [Restoring Internet Freedom](#) (Docket FCC-17-108). All of the data downloaded and analyzed were originally submitted using the [FCC's Electronic Comment Filing System](#) (ECFS). All data and comments used in this report are stored on the FCC's site and are freely available to the public. When submitting a comment, the [FCC notified users](#) that all their information submitted, including names and addresses, would be publicly available via the web.

The FCC opened the docket for public comment on April 27, 2017. The comment period was initially scheduled to end on Aug. 16, 2017, but was extended to Aug. 30, 2017. Only comments submitted during that official period (April 27-Aug. 30) were included in this study.

The FCC also allows for submissions by phone or letter, but those comments are not publicly accessible and were excluded from this report.

Pew Research downloaded all online comments from the [FCC's public Application Program Interface \(API\)](#) using a Python script. The FCC assigned a unique ID number to each comment, although researchers discovered that some ID numbers appeared multiple times in the dataset. The Center removed all comments with duplicate ID numbers. In total, the Center collected and analyzed 21,706,195 comments.

The collection and analysis of the comments were conducted prior to the release of downloadable versions of the comments as [zip files](#) by the FCC in early November 2017.

In addition to filling out the FCC's form, commenters had the option to attach files such as a text document. In order to maintain consistency and to limit the size of the dataset, the Center removed all attachments. The Center also removed addresses.

The following data was collected and analyzed for each comment:

- Full text of the comment
- Confirmation number
- Author name
- Email address
- Email confirmation
- Date submitted
- Date received



## Text matching

Many comments included text that was nearly identical to other posts but differed in small ways. For example, many comments included the exact text as it appeared on a form or sample letter but signed with a different name. Some comments were identical except for an additional space or apostrophe, while others utilized the same sentences but were ordered in slightly different ways. For this study, Pew Research Center decided that comments of this nature were similar enough to be considered “matching” or “non-unique.”

To systematically identify these matching or non-unique posts, the Center used a measure known as “[cosine similarity](#)” to compare the text of all comments in the dataset. This technique takes two comments and compares the characters used. Comments were considered to be non-unique if the cosine similarity was .95 or above on a 0-1 scale. The .95 threshold is a conservative benchmark and ensured that only those comments that were nearly identical in content were counted as matching.

In addition to using cosine similarity to identify matching posts, the Center also performed a manual grouping process on the 100 most-submitted comments. In this process, researchers manually grouped together comments where the only differences involved line spacing, line breaks, word capitalization or the name used as the signoff in the text.

The following three examples offer a practical demonstration of the matching process employed for this analysis:

### Example 1:

Comments A and B below have a cosine similarity of 0.98 and accordingly are considered matching. In this case, the only differences between the two are the capitalization of the word “internet” and the order of the sentences.

#### Comment A:

*The unprecedented regulatory power the Obama Administration imposed on the internet is smothering innovation, damaging the American economy and obstructing job creation. I urge the Federal Communications Commission to end the bureaucratic regulatory overreach of the internet known as Title II and restore the bipartisan light-touch regulatory consensus that enabled the Internet to flourish for more than 20 years. The plan currently under consideration at the FCC to repeal Obama's Title II power grab is a*

*positive step forward and will help to promote a truly free and open internet for everyone.*

**Comment B:**

*The plan currently under consideration at the FCC to repeal Obama's Title II power grab is a positive step forward and will help to promote a truly free and open internet for everyone. The unprecedented regulatory power the Obama Administration imposed on the internet is smothering innovation, damaging the American economy and obstructing job creation. I urge the Federal Communications Commission to end the bureaucratic regulatory overreach of the internet known as Title II and restore the bipartisan light-touch regulatory consensus that enabled the internet to flourish for more than 20 years.*

**Example 2:**

Comments C and D have a cosine similarity of 0.95, which just meets the Center's threshold to be consolidated. The only difference is the additional sentence in the first comment.

**Comment C:**

*Don't kill net neutrality. We deserve a free and open Internet with strong Title II rules. This will ensure that the flow of data is determined by the interests of Internet users*

**Comment D:**

*We deserve a free and open Internet with strong Title II rules. This will ensure that the flow of data is determined by the interests of Internet users*

**Example 3:**

Comments E and F begin the same way, but the additional sentence in comment E causes them to fall short of the 0.95 threshold used in this report. Accordingly, these two comments would *not* be considered matching.

**Comment E:**

*The FCC's Open Internet Rules (net neutrality rules) are extremely important to me. I urge you to protect them.*

**Comment F:**

*The FCC's Open Internet Rules (net neutrality rules) are extremely important to me.*