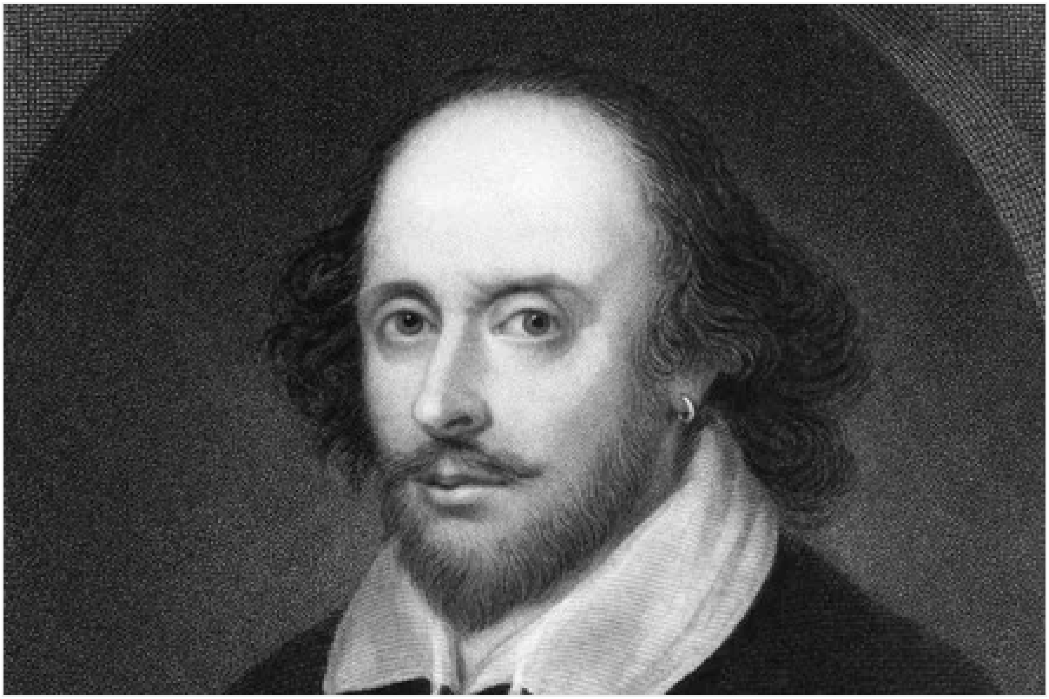
Software

# Tiny Twitter thumbnail tweaked to transport different file types

Troll preserve's images can be used to distribute code, PDFs and other stuff

By Thomas Claburn in San Francisco 31 Oct 2018 at 20:28

10 SHARE ▼



A picture turns out to be worth much more than a thousand words, at least on Twitter. For security researcher David Buchanan, it amounts at least 884,000, roughly the number words in the complete works of William Shakespeare.

Buchanan found that Twitter image uploads can be polyglot files, meaning they can be valid simultaneously in multiple formats, such as a `.jpg`, a `.rar` archive and a `.zip` archive. Using some Python code he wrote, he created a thumbnail image of William Shakespeare overlaid with the words, "Unzip Me" and posted it to Twitter.

The `.jpg` image is also a valid `.zip` file, so if you download it, you can unzip it and extract the contents, a multipart `.rar` archive of the text of Shakespeare's plays.

From the macOS command line, assuming you have unrar installed (`brew install unrar`) to handle the `.rar` files, this series of commands should work:

```
$ curl 'https://pbs.twimg.com/media/DqteCf6WsAAhqwV.jpg' >
bard.zip
$ unzip bard.zip
$ unrar e bard.part001.rar
```

Twitter performs some processing on uploaded images, which has the potential to mess with the data. But Buchanan found that his multi-format file survived this process. It may be that image itself (excluding the rather bulky metadata) is light enough not to trigger any compression or post-upload processing.

"The `.jpeg` format is made up of multiple segments," Buchanan explained, via Twitter DM. "One type of segment is reserved to define an 'ICC profile,' which is typically used for color calibration etc. Although Twitter strips most metadata (e.g. EXIF data), they do not strip ICC profiles. As it turns out, an ICC profile can be up to 16MB in size, and contain totally arbitrary data, with the slight limitation that it has to be split up into 64KB chunks (due to the nature of the JPEG/ICC formats)."

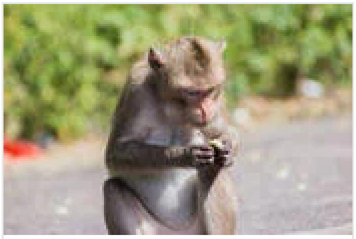
Adding files into this space is possible because the `.zip` format is surprisingly flexible, Buchanan said. "The 'Central Directory' is at the end of the file, and it 'points' to individual compressed files elsewhere in the overall `.zip` file," he said.

Thus a `.zip` file can still be valid with junk data at the beginning, middle, and even a bit on the end. "The `.rar` files are to circumvent the aforementioned 64KB chunk sizes, in a way that can easily be recombined," he said.

## Possibly a bigger issue

Buchanan said his technique works on image hosting service Imgur, but he hasn't tested it elsewhere. Some web services, like Shopify, strip ICC color profiles from images for the sake of color consistency and storage economy. The International Color Consortium (ICC) points out that while ICC profiles contain no executable code, there may be potential security issues arising from badly formed profiles.

*The Register* asked Twitter whether polyglot files of this sort pose a problem or violate the site's terms of service. We've not heard back.



**Give 1,000 monkeys typewriters, they'll write Shakespeare. Give them robot arms, and wait – they actually did that?**

READ MORE

"I originally reported this to Twitter's bug bounty program, via HackerOne, but they didn't seem very interested," said Buchanan. "I don't think this will become widespread enough to hurt Twitter's hosting costs, for example, although I can see it becoming a bit of a moderation nightmare."

He said people have suggested this technique could be used to distribute malware, but he's skeptical that it could become a practical attack vector. There is however precedent for using polyglot `.jpg` images to deliver malware (e.g. Stegosploit).

There are potential privacy issues with ICC profiles, Buchanan said, noting that they can be used to fingerprint devices.

Buchanan has made his source code available, again using a Twitter-hosted image as his distribution mechanism. To access it, you'd download the image, which depicts its underlying structure in the rendered words "source.pdf.zip.jpg." So if you download the image and unzip it (from the command line), you end up with the source code in `.pdf` and `.py` files. ®

Tips and corrections

10 Comments

**MORE** Security Software Twitter

## Whitepapers

### Digital Transformation Drives New Fast Data and Big Data Storage Platform Requirements

This white paper discusses the storage infrastructure requirements for fast data and big data platforms as well as the storage requirements around them that businesses will need to meet going forward.

### The Automated Enterprise eBook

Organizations are trying to optimize resources, speed development, and adapt faster to market changes.

### How CIOs Can Work Effectively With CFOs to Optimize Cost

In this Gartner research, find out how the relationship between the CIO and CFO is strengthened through strategic and ongoing spend optimization.

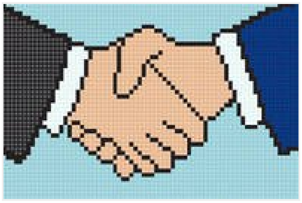
### The politics and practicalities of IT procurement

Having more options available can be quite liberating, but for some, the additional choices and decisions can become almost paralysing.

## More from The Register

### Facebook's security boss is offski. Not to worry, it has 'embedded security' in all divisions

Alex Stamos's replacement not yet announced



### Palo Alto Networks buys security startup Redlock for $173m

Threat detection outfit gets new owners

### Audit finds Department of Homeland Security's security is insecure

The agency that keeps America safe runs un-patched Flash, and worse besides

### Cisco drops a cool $2.3 billion on SaaSy outfit Duo Security

Switchzilla slurps trusted access into cloud to make it rain

### Mozilla-endorsed security plug-in accused of tracking users

Web Security says there's nothing nefarious to its URL collection

### Intel finds a cure for its software security pain: Window Snyder

Microsoft, Mozilla veteran will also handle external researcher work

### About us

Who we are

Under the hood

Contact us

Advertise with us

### More content

Week's headlines

Top 20 stories

Alerts

Whitepapers

### Situation Publishing

The Next Platform

Continuous Lifecycle London

M-cubed

Webinars

SITUATION PUBLISHING

**The Register** - Independent news and views for the tech community. Part of Situation Publishing

### Sign up to our Newsletters

Join our daily or weekly newsletters, subscribe to a specific section or set News alerts

Subscribe

Cookies Privacy Ts&Cs