FY2020's 200x Rise in Spam, and Tackling Password-protected ZIP Files

1.1 Introduction

A year has passed since the sudden rise of telework, from both home and satellite offices, as a means of dealing with COVID-19. While the world has been transformed over the past year, the importance of email for enterprises remains unchanged.

In this article, we look back on events in the messaging space in 2020.

1.2 FY2020 Spam and Virus Data

Figure 1 shows the number of spam emails received by honeypots run by IIJ over April 2020 – March 2021.

In the first half of fiscal 2020, we intermittently observed unprecedented levels of spam. It is difficult to see because the vertical axis accommodates the maximum value, but the average for April is set to 1. In the first wave in early May, we received 10 times [1] as many spam emails, the second wave in mid-May brought around 40 times [2], and around 60 times as many came in late May through early June. This continued intermittently, with around 200 times as many spam emails received at end-July. I think it's rare for the equipment designs of organizations in general to call for investment in equipment capable of withstanding 200 times normal levels, so calling this a DDoS attack would not really be an exaggeration.

Let's also look at the virus figures. Figure 2 shows the total number of viruses arriving at IIJ's honeypots during the same period.

The average for April is again set to 1 here, and 1,000 times that many viruses were received in June. But the other figures are hard to see because the range of counts observed over the year is too wide, so Figure 3 plots the same data logarithmically.

There is always something going on, but unlike with spam, incoming virus volumes are concentrated over short time intervals several times a year.

A look at a sample email received in June reveals the subject line "Look at this photo!", and the name of the attached ZIP file—IMG135123.jpg.js.zip—makes you think it could contain an image. Upon unzipping, we find it contains JavaScript that downloads malware. The file itself is not malicious per se, but if the user runs the JavaScript, it will download the malware. Getting a user to open a file by making them think it's an image is a classic old technique for those in the know, but we must take note of this and not regard it as an antiquated attack.

The second largest peak is in September. The virus observed on this occasion was Emotet. A characteristic of

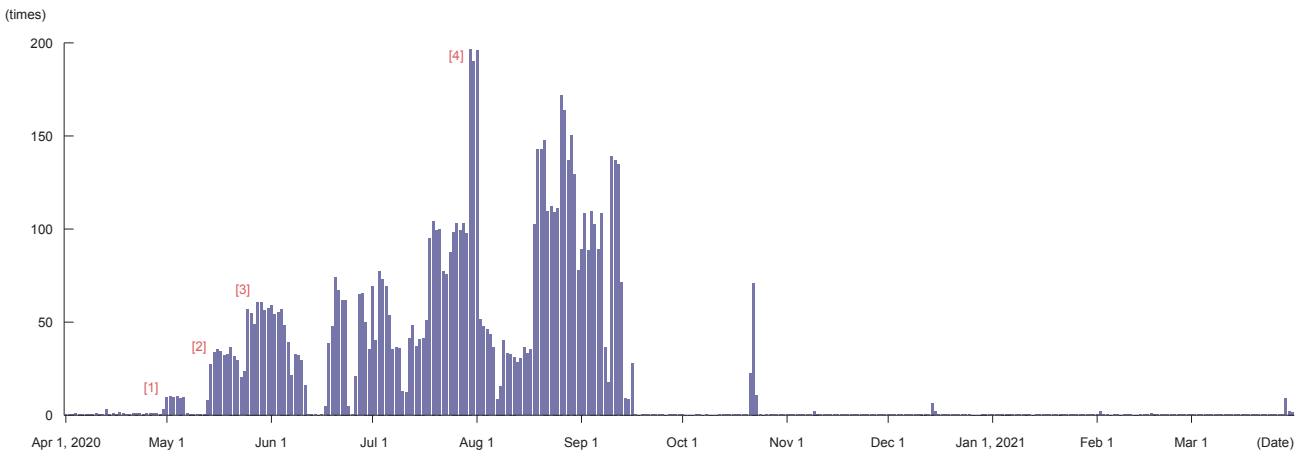

Figure 1: Spam Arriving at IIJ Honeypots (Apr 2020 – Mar 2021)

Emotet, which ran rampant in 2020, is that it encrypts itself in a password-protected ZIP file, but the observations in September revealed a previously unseen characteristic. The output in Figure 4 is part of a sample showing the detailed structure of the Emotet received (encrypted in ZIP format).

The encryption method had changed from ZipCrypto, the ZIP file standard, to AES 256-bit.

AES offers stronger encryption than the ZIP standard and is supported by some archivers such as 7-Zip, but Windows

```
$ zipdetails 'OXEFVNG1 20209月16.zip'

0000 LOCAL HEADER #1       04034B50
0004 Extract Zip Spec      33 '5.1'
0005 Extract OS            00 'MS-DOS'
0006 General Purpose Flag  0803
     [Bit  0]              1 'Encryption'
     [Bit 11]              1 'Language Encoding'
0008 Compression Method    0063 'AES Encryption'
003D   Encryption Strength 03 '256-bit encryption key'
```

Figure 4: Part of Sample Showing Detailed Structure of Received Emotet (encrypted in ZIP format)

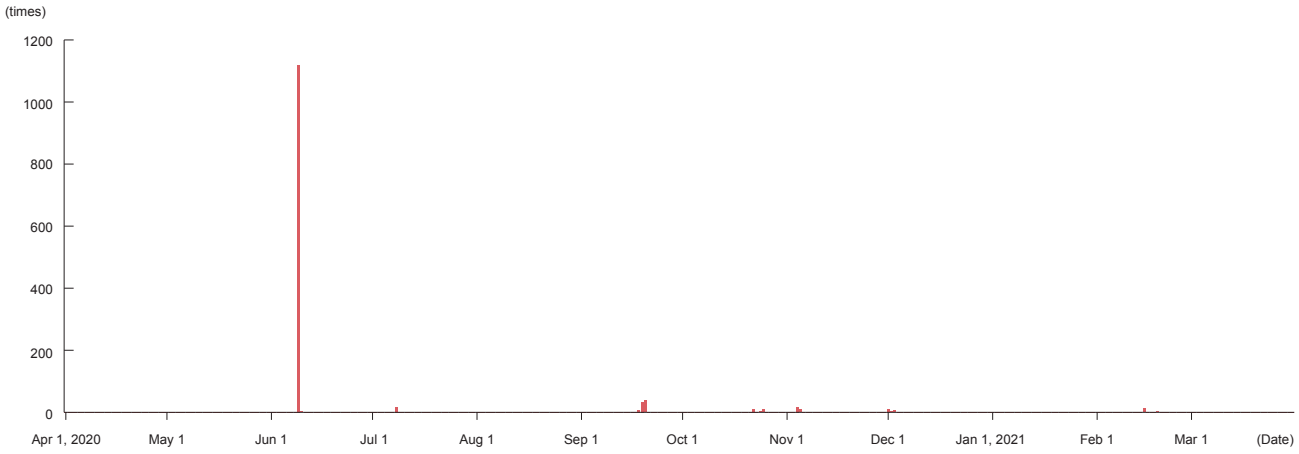

Figure 2: Viruses Arriving at IIJ Honeypots (Apr 2020 – Mar 2021)

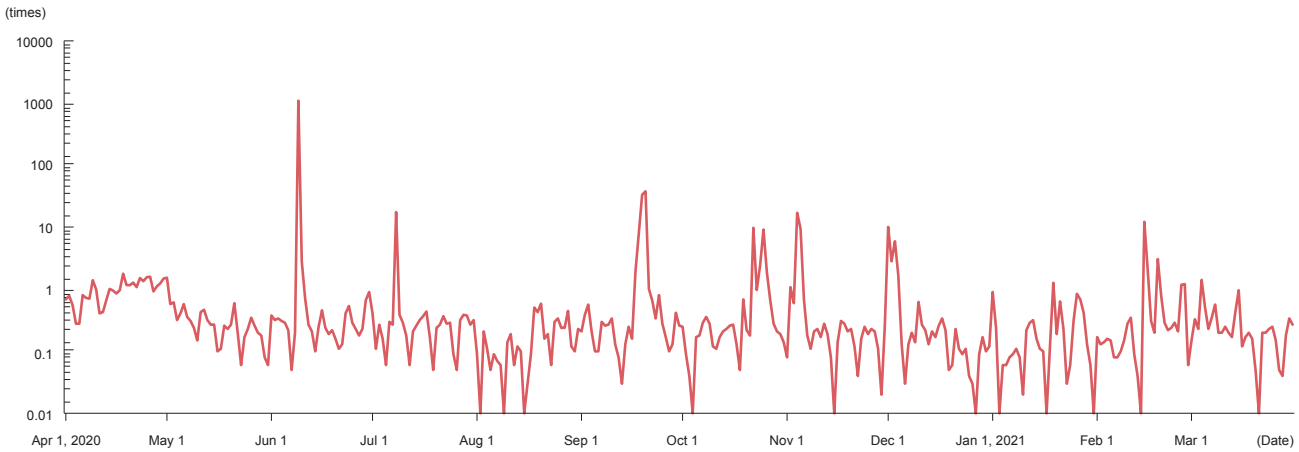

Figure 3: Viruses Arriving at IIJ Honeypots, Plotted Logarithmically (Apr 2020 – Mar 2021)

Compressed Folders do not support it. There would not seem to be much point in strengthening the encryption if your objective is to spread a virus, because you want as many targets to open it as possible. So why was AES being used?

This is just a guess, but it may be an attempt to avoid sandboxed detection. Some security products and services pull text strings likely to be passwords out of the email body and use them to try to open protected files in a sandbox. But the Windows standard does not support AES-encrypted ZIP files. Even if the correct password can be extracted from the email, the archive cannot be opened, so malicious behavior goes undetected and the file circumvents the sandbox protection.

So evidently, attacks that may appear the same on the surface are gradually changing tactics.

1.3 Calls to Abandon Encrypted ZIP Files

In November 2020, Takuya Hirai, Japan's Minister for Digital Transformation, spoke about the idea of "abandoning the use of encrypted ZIP files (a practice commonly referred to as PPAP in Japan)"[*1].

The phrase "encrypted ZIP files" here is referring to the practice of encrypting files in a password-protected ZIP archive and sending them to someone via email, and then sending the password in a separate email[*2]. The practice is somewhat unique to Japan and virtually nonexistent overseas.

Reasons cited for using this method include to prevent files being missent and for route encryption, and it is often stipulated mainly in organizations' IT security policies because it can be performed by individual users and the cost is low (the immediate cost at least, since no special applications are needed).

1.3.1 Problems with Encrypted ZIP Files

Why should encrypted ZIP files be abandoned? The answer is that the many risks outweigh the potential benefits of this method.

The most serious issue is that it prevents received emails from being scanned for viruses. As explained in Section 1.2, emails are still a useful attack vector for would-be attackers since anyone can send and receive them[*3]. So generally, many organizations implement virus protection on their gateways, Web servers, and the like to prevent infection by viruses attached to emails.

But you can easily circumvent such protection by encrypting attachments. This renders the antivirus product incapable of

*1 Cabinet Office, "Summary of press conference by Minister of State Takuya Hirai, November 17, 2020" (https://www.cao.go.jp/minister/2009_t_hirai/kaiken/20201117kaiken.html, in Japanese).

*2 At the time of publication, some emails sent by IIJ staff use this method. While work and customer circumstances mean we cannot completely abolish the practice immediately, we do intend to progressively discontinue encrypted ZIP files as preparations fall into place.

*3 Information security site for the Japanese public (Ministry of Internal Affairs and Communications), "Virus infection routes" (https://www.soumu.go.jp/main_sosiki/joho_tsusin/security/basic/risk/02-1.html, in Japanese).

analyzing the file contents[*4]. And this is perhaps an obvious outcome since the whole point of encrypting files is so that others can't read the contents.

Since around 2012, we have observed targeted attacks[*5] directed at specific organizations, individuals, and teams apparently designed to exploit this weakness. These attacks involved exchanging harmless emails on a number of occasions to gain the recipient's trust, culminating in a final email saying, "I've put all my questions into the attached file". The attached file would contain malware encrypted in ZIP format so as to deliver it to the recipient without a virus scan.

The Emotet version seen in 2020 would encrypt itself in a ZIP archive as a Word or Excel file with macros, which it attached to emails. To spread itself, the virus would read the address book and emails on infected devices and send out emails purporting to be replies coming from the infected device.

And Emotet is very sophisticated in that it comes in so many variants, spreading itself with email subject lines like "invoice" and "order" (in Japanese to target Japanese users) and timing its emails for the end/start of the month. Encrypted ZIP files are exchanged so commonly in Japan that unsuspecting users are likely to open them unless they are paying careful attention. This happening over and over again is probably what precipitated the virus's explosive growth.

So this method, intended by email senders to be a security measure, results in the recipient's email system unconditionally treating files as harmless, with the result that it actually exposes the recipient to risk.

The US CISA (Cybersecurity & Infrastructure Security Agency) has issued an alert on Emotet[*6]. The text states that blocking encrypted ZIP files that cannot be scanned by antivirus software is a risk mitigation measure.

1.3.2 Alternatives to Encrypted ZIP Files?

How should we send and receive files if we abolish the encrypted ZIP method?

There are moves within the government to use shared storage systems[*7]. Instead of sending an attachment, you send the path or a one-time URL to the file in shared storage.

This method has a number of advantages. Files can be scanned for viruses in the shared storage rather than in email[*8], missent files can be deleted as soon as the error is realized, and files that are too large to be sent via email can also be exchanged.

*4 Encrypted files are sometimes identified as viruses even when the contents cannot be analyzed. But because this is based on limited information, such as the file hash and filename, it is more difficult to identify viruses this way than with unencrypted files.

*5 Attacks that target a specific team or individual within an organization to allow an attacker to steal confidential information or gain a path to intrusion. While in technical terms such attacks do not differ from conventional methods such as sending people viruses or directing them to phishing sites, the emails do not end up in other organizations' or antivirus vendors' honeypots, so the attacks are harder to detect.

*6 CISA, "Alert (AA20-280A) Emotet Malware" (https://us-cert.cisa.gov/ncas/alerts/aa20-280a).

*7 When the Cabinet Office announced the policy, it did not disclose what sort of system would be used. But, unfortunately, a subsequent incident involving unauthorized access to shared storage revealed that a file transfer appliance installed within the Cabinet Office was being used. Cabinet Office, "Unauthorized access to shared file storage used by Cabinet Office staff" (https://www.cao.go.jp/others/csi/security/20210422notice.html, in Japanese).

*8 Antivirus technology these days generally updates patterns in real time to increase detection rates. This is also a big advantage because even if you can't detect a virus when the email is received, you may be able to detect it later when the recipient downloads or views the file.

These could be seen as weaknesses from an internal control perspective, however, so we can't leap on this option too enthusiastically. For example, many organizations commonly archive sent and received emails for tracing purposes. This provides important evidence in the event that an email exchange becomes a point of contention in a future litigation case or the like. But if attached files take the form of a URL, it may be difficult to trace exactly what files were sent and received.

Another risk is that it may create loopholes in measures to prevent information leaks and thus facilitate insider crime. Important documents could be uploaded to shared storage and taken outside of the organization via a one-time URL, with the file then deleted from shared storage. Administrators cannot realistically read through every email, so email auditing systems in some organizations archive emails for checking based on whether they have any attachments. But an email that only has a URL in the body text is no different from any other ordinary email, so it becomes very difficult to discover misconduct.

In the end, any system will have its advantages and disadvantages, so the key is how much of what sort of risks you are willing to accept.

Personally, I have been known to attach files to emails when I needn't have, so I can simply stop doing that, by which I mean reconsidering whether I really need to send certain information in an attached file, and if it really is necessary, sending it as an attachment and combining that with a temporary storage system and/or monitoring system depending on the risk. I think that is a reasonable middle ground. Either way, the call to abandon encrypted ZIP files will no doubt be an important turning point for the way we look at email systems and policies ahead.

1.4 Cautionary Notes on Online Conferencing Systems

As mentioned at the top, 2020 saw the rapid rise of telework. Online conferencing systems, only in use by a subset of users a year earlier, suddenly became widespread and reached a level of usability accommodating anyone and everyone. These online conferencing systems all share a particular mechanism in that users join a conference by clicking a one-time URL issued by the host. This system is very convenient because online conference participants simply need to click on the host's one-time URL received via email or the like.

But what if someone impersonates the online conference host and provides a phishing site URL instead? Users are accustomed to clicking on one-time URLs, so it's not hard to imagine them casually clicking on such a URL in an email. A fake login screen that looks identical to the genuine article could be used to steal user credentials including ID and password, allowing an attacker to retrieve internal company information and possibly gain access to other services if the user uses the same password for multiple services.

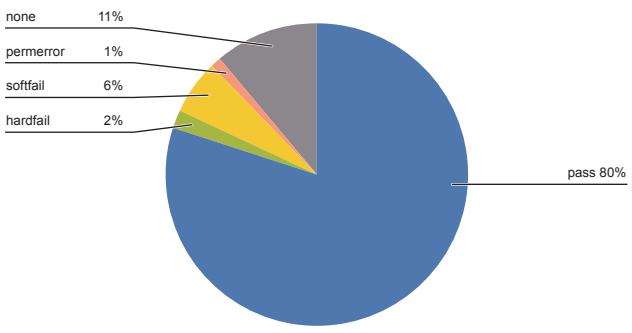

Figure 5: Breakdown of SPF Authentication Results

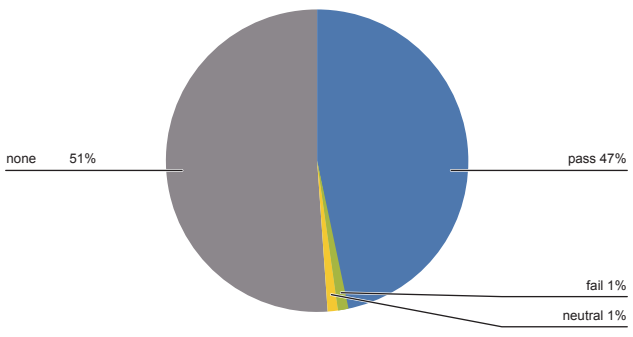

Figure 6: Breakdown of DKIM Authentication Results

Fortunately, email has a powerful framework called DMARC, which can authenticate the sender domain. Companies' and other organizations' domain administrators should ensure proper sender authentication is in place to combat spoofing.

Alongside system-based countermeasures, regularly reminding and educating users that some attacks start with emailed URLs can also further enhance the effectiveness of security measures.

1.5 The Rise of Sender Authentication

Figures 5 to 7 show breakdowns of sender authentication results aggregated across email services provided by IIJ over April 2020 – March 2021.

Sender authentication technology is undoubtedly effective against emails spoofing your domain and phishing emails impersonating well-known brands. The data here show sender authentication results as a proportion of all emails received, and it should be noted that large volumes of email, such as those sent out by mass email distributors, tend to swamp the rest of the data. Another point to note is that some spammers deal with sender authentication by buying domains for the purpose of sending out spam, so a "pass" result does not necessarily mean that an email is not spam.

Comparing these data with the results we reported in IIR Vol. 47 (https://www.iij.ad.jp/en/dev/iir/047.html), the proportion of "pass" results (successful authentication) is up a few percentage points for SPF, DKIM, and DMARC alike, and the proportion of "none" (no authentication information) is down for SPF and DMARC. DMARC has taken on the role of authenticating the domain found in the From header, so DKIM is really only used for authenticating email sent using an organization's own domain from a third-party SaaS, such as Salesforce and mass email distributors.

1.6 Conclusion

RFC822, specifying the current SMTP email protocol, was published in 1982. SMTP stands for Simple Mail Transfer Protocol, and it's hard not to be surprised that this same simple email protocol continues to run some 40 years later.

All sorts of communication tools such as online conferencing systems and text-based chat are popping up. But the reality is that email cannot replace all of them, so this situation is likely to persist for some time.

Looking ahead, IIJ will continue working to make the world safe for email.

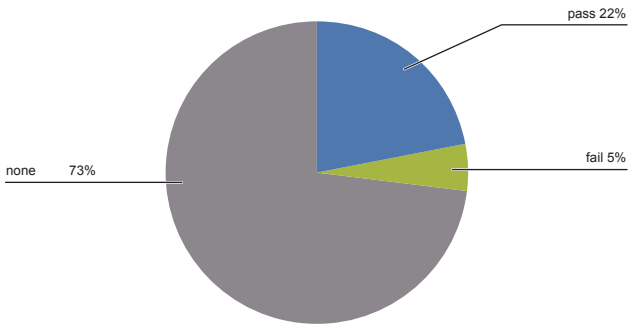

Figure 7: Breakdown of DMARC Authentication Results

Isamu Koga
Manager, Operation & Engineering Section, Application Service Department, Network Division, IIJ
Mr. Koga joined IIJ in 2007. He is engaged in the operation of email services and investigates email-related trends in the wild. To keep customers' email boxes safe, he communicates information about the latest attack methods, trends in spam, and countermeasures.