Global

Global: A briefing on HTML email

Contributed by RobBrandt on Apr 17, 2006 - 10:11 PM

I have this discussion often with the various groups of people that I work with, and usually they don't believe me, and I come off as an old tech fogey, resistant to "modern technology". That's OK, because I know that one day they'll have a bad experience and realize that I'm right :). And I've given up trying to change how people use email anyway; that horse has left the barn, and it's too late to close the doors. So now I'm just going to refer people to this briefing and let it go at that; they'll be duly warned, and will at least be able to anticipate the problems they may encounter. Here's an Executive Summary if you're not interested in the details:

- You can't assume that viewing incoming html email won't infect your computer with a virus;
- You can't assume that a sender of an html email isn't tracking whether you are looking at the email - or even previewing it;
- You can't assume that your formatted message will look good to your recipients;
- You can't assume that your recipients will even be able to read your formatted message;
- You can't assume that the automatically created plain text version of your message will look good, or even work at all.

If you are interested in the details, read on by clicking the title of this article...
I have this discussion often with the various groups of people that I work with, and usually they don't believe me, and I come off as an old tech fogey, resistant to "modern technology". That's OK, because I know that one day they'll have a bad experience and realize that I'm right :). And I've given up trying to change how people use email anyway; that horse has left the barn, and it's too late to close the doors. So now I'm just going to refer people to this briefing and let it go at that; they'll be duly warned, and will at least be able to anticipate the problems they may encounter. Here's an Executive Summary if you're not interested in the details:

- You can't assume that viewing incoming html email won't infect your computer with a virus;
- You can't assume that a sender of an html email isn't tracking whether you are looking at the email - or even previewing it;
- You can't assume that your formatted message will look good to your recipients;
- You can't assume that your recipients will even be able to read your formatted message;
- You can't assume that the automatically created plain text version of your message will look good, or even work at all.

There are two issues with html email. The first is if you want to view email you receive in html, and the 2nd is if you want to send email formatted with html. You usually have a choice in these matters, but it depends on what email program you use. Most people use Microsoft Outlook, which uses Internet Explorer as it's internal html viewing engine, but there are many other options. Each program usually has a setting to control whether you view email in html or send email formatted with html or both.

Viewing email you receive in html

I admit, it's really nice to receive a neatly formatted email with colors, italics, different fonts, etc. There are HUGE problems with doing so though, and in my estimation it's just not worth it. Typically (but not always) each html email you receive has a plain text version with it, so it is your choice, not the sender's, whether you see the message in html or plain text. Here's why I don't think viewing the formatted html version is worth it:

It can carry viruses. When you see an email in plain text (not html), you can see exactly what's been sent to you. There's no where to hide anything, because you are viewing the exact content of the email. With the html version, you are only viewing what the creator of the email wants you to see. Typically, there is no bad intent, but a virus

writer can easily hide viruses or a script to retrieve a virus from a web server within the html. Most up to date email programs watch for these tricks, but none are perfect. New attack methods are being developed all the time. If one hole is found in the email program's security, other security protections may be disabled by the virus without your knowledge, and then your system may be wide open. No security is perfect. The odds of a particular email causing a problem are extremely small, but the virus writers are playing the odds here. Suppose the odds of a virus infecting your PC are only 1 in 1,000. Many people I know complain of getting a dozen virus loaded emails per day. That means they will get their 1,000th virus in 83 days. Which means that you will likely get a virus within 83 days of starting to receive html email. I don't like those odds at all. With plain text, this isn't even possible no matter what the odds.

You could be a victim of a scam (commonly called "phishing"). This is a variation on the virus theme above except it involves a bit of deception to get you to go along. Have you ever gotten an email from an online service you use, such as PayPal, that asks you for a response? These are mostly legitimate, but a phishing scam looks the same but is not legitimate. When you get an email asking you to go to http://www.paypal.com, it seems fair to assume that when you click on it, you will actually go there. But with html email, it ain't necessarily so. A legitimate company will send out an html email with a link written like this: http://www.paypal.com. For those of you who are not html savvy, the href= part is actually what determines where your browser goes, and the part between > and is what is displayed. There is no connection between the two; what you see and what you get is only the same if that's what the composer of the html email wants. Phishers don't. So they construct a link that looks like it's going one place, and it's really going another. For example I can make an "paypal" link that comes back to this web site, like this: http://www.paypal.com. A phisher will send a link that looks like it's going to paypal, but will instead go to a site they control, make the site look like paypal's, and present a login form where the unsuspecting "mark" will use their password to try and login; ***Viola!*** they now know the login id and password to get into paypal and empty your account. With plain text, there is no hiding where the browser is going to go, so when you get an email from a phisher you can see quite clearly that it's not going to go to paypal, and you don't click on it (if you have any brains).

Using html email actually causes you to get more spam. This trick spammers use takes advantage of how the internet works. There are multiple "tools" for using the internet; email, web browsing, and file transfer are the most common. The last one, file transfer (known as "ftp") allows one computer to transfer a file from one computer to another. FTP requires a login and password, but it has a special "anonymous" mode where it is standard practice to use your email address as a password. Most web browsers do this automatically for you, often supplying your actual email address as the password. Some bright spammers realized that they could take advantage of this to build email lists of actual people with good email addresses who are interested in particular products. What they do is send you an html email with an image in it - maybe just a 1 pixel dot - and link it to their web site using ftp. So they compose the email offering a low rate on a mortgage in the title - it doesn't even have to have any content in it that makes any sense - and send it to you. If you view it, your browser goes out to his server to retrieve the image by ftp using your email address as the password, the password is logged, and when a few days have passed the log is retrieved and sold to a real mortgage spammer who knows that he's getting a list of good email addresses of people who are interested in mortgages (you looked at the email, right? So you must be interested). So simply by viewing the email in html mode, you've added yourself to a few spammer's lists. This might even happen in "email preview mode", meaning that you don't even have to open the email to trigger the ftp. With plain text, you'd have to actually click on the ftp links to get logged, and if you do that, well, "you're on your own, buddy".

So, in summary, viewing email in html increases the likelihood that you will get a virus, be a victim of a phishing scam, and increase the amount of spam you receive. Because this is so, savvy computer users do not view email in html. But if you insist, see the end of this message for ways to mitigate or limit the damage.

Sending email in html

First, html viewing isn't consistant. Formatted email causes problems. Period. The reason the internet works as well as it does is because it's based on standards that industry-wide committees and organizations have approved and published. These standards are free to use, and anybody who wants to write software that uses them knows

that the results will behave properly with other software written to make use of the standards. However, there is no standard available on how formatted email should be displayed in an email. So all the software makers do what's right in their own eyes, and email produced in one software program may not look right at all in another software program. When you are working in a homogeneous environment - perhaps at work - where everyone uses the same program, there is much less chance that there is going to be a problem, but once you go out of that environment the chances of a problem increase substantially. AOL is the worst offender in this regard, using style methods that are completely different than anyone else's; Outlook is better than it used to be, but still doesn't do some things right (as you'll see below); other lesser known applications have their own quirks. This will be the case until standards are approved, published and adopted.

Second, some people don't read formatted messages at all. Remember the conclusion of the section above; savvy email users don't view email in html. Think hard about how you want to target people who aren't savvy, and if you want to appear unsavvy to people who are. The html email problem becomes less of an issue as each year advances and people upgrade their mail software, but it will never go completely away because some people can't or won't read formatted email. Engineers particularly, who have been using email for decades, are wedded to their favorite text mode (think DOS) email software, and it simply has no ability to read formatting. Others - like myself - know that if a plain text version of the message isn't included, it's highly unlikely to be something I'm interested in and ignore it. It's most likely spam. Or a virus or scam. Again, see above. That doesn't mean I don't see *your* emails composed in html. Most email software (but not all - THERE ARE NO STANDARDS) that creates formatted email (like Outlook, AOL, Eudora and Mozilla Thunderbird) automatically creates a plain text version of the message at the same time for you in the background. If you are composing a formatted message, you never see this plain text message and may not even know it's there. What happens when you send the message (usually, but not always because THERE ARE NO STANDARDS) is that the plain text version gets sent as the "real" email, with the formatted version included as an attachment. Most software (but not all - THERE ARE NO STANDARDS) that receives an email like that will see the attachment and display that instead of the plain text version. As far as the sender is concerned, they sent a formatted message, and as far as the receiver knows, they received a formatted message. They never see the plain text version. Unless the receiver can't or won't look at formatted messages, in which case the plain text version is displayed, if there is one. If there isn't one, they never see the message. I run a lot of email mailing lists, and we have them set up so that this plain text version is what's sent (even if the message is html formatted), so all the messages will go through no matter how they are formatted, as long as the sender's software actually sends a plain text version. The formatted version (and most other attachments) are stripped off.

Third, you can't rely on the automatic plain text version. You don't get a chance to proof read it, sometimes it looks terrible, and sometimes it just doesn't work right. I can cite two recent examples of this. In the first example, someone sent out an announcement containing a link to a web site. She originally sent the message heavily formatted, and included an attachment of another message. Both of the attachments got stripped off, and the main content was in one of those attachments. Since that was an attachment, Outlook didn't have a chance to create a plain text version. So the first message went through without any content. An empty email - not very professional looking. On the second attempt, she copy/pasted the content of the other message into the formatted Outlook mail. Outlook therefore created a plain text version of the message, and that was what went through. It was readable, but didn't look great. There was lots of empty space and odd line breaks. The big problem though, was how Outlook transformed links from the formatted version to the plain text version. The formatted version of the mail contained a link to the web site page, which looked something like this: http://www.example.com. Outlook transformed the plain text link to this: http://www.example.com<a href="http://www.example.com> . I'm not sure why it does it twice since only one is necessary (the 2nd one is in brackets). But the real problem is that there are no spaces between the two. Outlook doesn't do a very good job of interpreting where a link URL starts & stops, so it thinks the whole thing - both versions - as a single web address. That's right, Outlook creates a link that Outlook can't read right. Did I *say* that THERE ARE NO STANDARDS? :). And so, since the message was created as formatted, she had no choice but to send a badly formatted plain text message, because it was automatically created and she couldn't see that plain text version to proof read it. Many recipients of the email complained that the page wasn't there (because that's what their browser reported) but in reality the real problem was that Outlook was confused about what the page address actually was; if the email had actually been composed in plain text, it would have worked for everybody. In the second example, the email sender had composed a really nice looking html template to use with all of her outgoing email. The key problem was the use of background images. She composed the template and linked directly to images on her own computer. Well, you and I can't see images on her

computer, so all of her emails that went out had missing images. But she didn't know there was a problem until I alerted her to the problem. Why? because her use of html was hiding from her the problems with her template. It worked on her computer because she could access all her files; but of course, the point of the email is for someone else to see it. The use of html hid her errors from her.

Fourth, use of html triggers spam filters. If you've viewed all your email in html mode until now, you know this makes sense. Spammers love html; big letters, lots of colors, etc. Spam filters have a difficult job to do; they have to be able to see an email they've never seen before, and figure out whether it's spam or not. There are many approaches to doing this, but most are either statistically based or "rule" based. With the statistical method, the system tracks what you and others identify as spam and not spam, and look for statistical similarities between a given email and what it's tracked as spam and not spam in the past. Since spam commonly uses lots of html and good email uses little or no html, a legitimate email formatted as html is more likely to be misidentified as spam the more html is included in it. "Rule based" filters work the same way; a given email accumulates points for every "rule" that is met. When a given number of points is accumulated, it is considered spam. The more email formatting is applied, the more points accumulated. I'll provide an example from my employer to illustrate the "rules" method:

Dear Botany 2005 Attendees,

1. If you have yet to book your accommodation...
2. Please try the new PERSONAL ITINERARY PLANNER
3. Oral Presentation AV Information
4. Poster Presentations Information
5. Student Projectionist openings

Message is on the web at: http://www.botany.org/emailmessages/Note-05-7-21.html

1. If you have not yet made your hotel reservations, the Hilton has extended the room guarantee date to July 23th. After that the price may go up...Reserve your room now! Go directly to Botany 2005 Hilton registration link at: http://www.hilton.com/en/hi/groups/private_groups/auscv_bso/index.jhtml

2. Please check out the new "**personal itinerary planner**" on the Botany 2005 meeting site: (http://www.2005.botanyconference.org/engine/search/index.php)

Have you ever had trouble trying to figure out and keep track of where you want to be, whom you want to hear, and what events you want to attend at a conference? For BOTANY 2005, weve helped you solve that problem! New this year, we have created a way for you to build your own customized itinerary for the conference. You can select which session you would like to attend, create your own schedule and print it off for the week.

a.) Go to the Conference Summary Page noted above.

b.) Login at the lower left hand corner of the menu using you conference login.

c.) Click on "**Custom Schedule"** under the Program/Schedule heading on the menu and follow the directions. From there youll be able to select the events/talks/meetings you would like to attend and and be able to print your schedule.

3. AV Information is now available for Presenters of oral papers at http://www.2005.botanyconference.org/ScientificMeeting/AV.php. Be sure to double check your presentation before your scheduled time, in the Presentation Ready room. If you have any questions, don't hesitate to contact Johanne at johanne@botany.org

4. Information for Poster Presentations has also been updated at http://www.2005.botanyconference.org/ScientificMeeting/PosterInfo.php

5. There are still openings available for Student Projectionists....if you are able to volunteer to work, we will

reimburse your early student registration fee. Fill out the Application (http://www.2005.botanyconference.org/StudentSupport/StudProjApp.pdf) and fax it to Johannes at 614-895-7866.

Botany 2005 is promising to be an exciting week of plant science, education, sharing, and just plain fun! Austin is waiting for us!

See y'all there!

Thank you for your time!

Sincerely,

Now, the above email is certainly not spam, nor does it have a substantial amount of html formatting attached to it. In fact, there is very little in it that could not be done in plain text. I use a very popular spam filter on my server called SpamAssassin, and it uses a rule based system to evaluate email messages. Let's see how it evaluated this one:

X-Spam-Status: No, score=1.6 required=5.0 tests=DATE_IN_FUTURE_06_12, HTML_30_40,HTML_FONT_BIG,HTML_MESSAGE,MIME_BOUND_NEXTPART autolearn=no version=3.0.4

Now, that's not too bad! I have typical settings entered for SpamAssassin, meaning that it doesn't consider an email "spam" until it gets 5 points; the above email got 1.6 points for it's "score". It got points for the email having a date not within an expected time range because his PC's clock was slow (lots of spam has very old or future dates), for having html applied to 30-40% of the text, use of a large html font, being an html message in the first place, and having an particular kind of attachment. This email message is not considered spam. But it did get points, and he may have done some other things that push it closer to being spam, such as misspelling too many words, mentioning his "mortgage" or a "prescription", or other things. Bottom line, he's 1.6 points closer to being spam than if it were sent in plain text. Plus, some servers have their spam settings set differently, such as "required=3.0", which means that only 3 points are required to be considered spam. As an email sender, you can't control that. But you can control whether you get points for using html by not using it.

So, in summary, you shouldn't send email formatted with html because you have no idea if the email will look like you want it to because the recipient might be using a different email program than you are, because your recipient may not even be able to see the html version, because the automatically generated plain text version that goes along with it usually looks terrible and often doesn't even work, and because use of html might trigger a spam filter preventing your email from even being received and read!

"But I really really like html email" or "I'm required to for work"

Ok, we live in the real world and sometimes compromises are necessary. Despite the fact that it's a really bad idea to view email in html and there are huge disadvantage in sending html email, some of you are going to keep doing it for various reasons. What can you do to mitigate the damages?

Don't use Microsoft products. That's am oversimplification with exceptions. But since Microsoft tends to create products that work very well with it's own products, and not so well with others, if you anticipate sending and receiving email with others who might not be using Internet Explorer or Outlook you will most likely have the fewest problems with other products who know they have to work with all products. Also, since Internet Explorer is so closely "welded" into the Windows operating system, a security flaw in the browser is also a security flaw in the operating system, making such a flaw more serious.

Crank up the security. On Windows, go into your Internet Options control panel and set all the options for either Disallow or Prompt. There are a few that are inconsequential, but it's not my intent to go into it here. Most important are anything to do with scripting or iframes.

Turn off image viewing in your email and browser programs. This can be a huge sacrifice, especially in the browser, but in some circumstances it is the only way to prevent the ftp email address collection measures described above. Particularly in Internet Explorer/Outlook. In those programs, you can turn off image viewing, but you can't change the email address used for ftp file retrieval. In many other programs (such as Mozilla Firefox and Thunderbird), not only can you turn off image viewing, you can tell it to use a bogus email address to log in to an ftp server when you *do* want to view the images. Not possible in IE/Outlook.

If you are sending out a bulk email, use a program that will allow you to seperately compose the html and plain text versions. By doing this, you can control the look of the plain text version as well as the html version. There are many specialty bulk mailing programs available with this feature. If you are doing a lot of bulk emailing and want to be sure that people who want html email like what they get, and that people who want plain text actually get it and can read it, this is the only way.

Use a Mac or Linux based computer. Yeah, that's a lot to ask for some people, but few virus writers target Macs or Linux, and the software for them is usually done with more security in mind. This won't help with the sending part of it, but it will protect you better when receiving.