

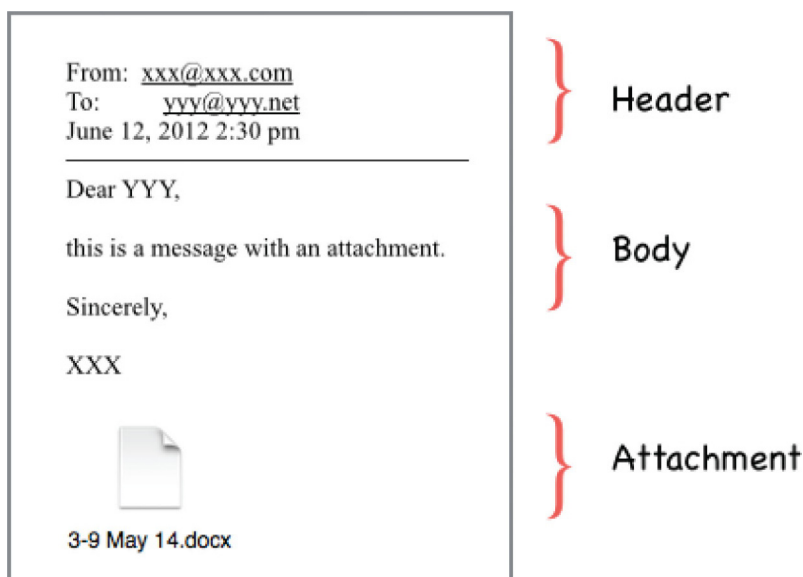
CHAPTER 4

Inside Messaging: Making the Hidden Visible

It is easy to look at an email and not see what is hidden behind the display. In fact, most people would not even suspect that behind a very simple looking email might be lurking some complicated programming. It is both a reality and a shame that bad guys figured out how to trick people into doing things they ordinarily would not do by simply disguising the actions, but with just a little bit of knowledge, you can understand how it works. Through that understanding, you can be more alert as to the potential problems.

EMAIL BASICS

The actual structure of an email consists of several elements. First, there is the header information. This is the electronic equivalent of all the information on the outside of an envelope, plus a bit more. Then there is the body of the message. This is the functional equivalent of a letter inside the envelope. And then there are possible attachments, which can come in many different forms. Each of these three elements is important to understand. The following figure illustrates the standard format of an email.




THE HEADER

The header of an email contains a treasure trove of information. In a typical email client, like on a smart phone, in a web browser, or in an email client, the user will not see all of the data that are available. What the user will see is basic information, such as the name of the sender, the names of the “To” recipients, and the names of the “CC” recipients, plus a subject line and the date and time of the message. The hidden details are long and laborious, and not really very useful information for the vast majority of legitimate emails. But for malicious emails, the hidden data contain very useful information indeed. The two figures below show comparisons of the same email: first, the standard display for emails on a smart phone, a web browser, and in an email client; and then the actual information behind that display. This email purports to be from Capital One, the large financial institution. Looking at the header information gives us clues as to whom the message is actually from.

Capital One <capitalone@email.capitalone.com>
 TO: Recipient <recipient@domain.com>

July 10, 2013 3:41 PM
[Hide Details](#)

Account Locked Notifications®



[Add us to your address book.](#)
[Help prevent fraud.](#)
[Log in to your account.](#)

A new message from Capital One®

Dear Capital One OnlineSM Customer,

We have detected an unusual activity in your account.

```

Return-Path: <daemon@chat.cafe24.com>
Delivered-To: recipient @2937168.3261322
Received: (qmail 15795 invoked by uid 0); 10 Jul 2013 19:41:20 -0000
Received: from unknown [HELO atl4mhib03.myregisteredsite.com] (209)
  by 0 with SMTP; 10 Jul 2013 19:41:20 -0000
Received: from lac.honamlife.com ([1.234.22.195])
  by atl4mhib03.myregisteredsite.com (8.14.4/8.14.4) with ESMTP id r6AJGFL012083
  (version=TLSv1/SSLv3 cipher=DHE-RSA-AES256-SHA bits=256 verify=NO)
  for <recipient@domain.com>; Wed, 10 Jul 2013 15:41:18 -0400
Received: from chat.cafe24.com (localhost.localdomain [127.0.0.1])
  by lac.honamlife.com (8.13.8/8.13.8) with ESMTP id r6AHHGag017441
  for <recipient@domain.com>; Thu, 11 Jul 2013 02:17:12 +0900
Received: (from daemon@localhost)
  by chat.cafe24.com (8.13.8/8.13.8/Submit) id r6AHHBn1017436;
  Thu, 11 Jul 2013 02:17:11 +0900
Date: Thu, 11 Jul 2013 02:17:11 +0900
Message-Id: <201307101717.r6AHHBn1017436@chat.cafe24.com>
To: recipient@domain.com
Subject: Account Locked Notifications®
From: Capital One <capitalone@email.capitalone.com>
Reply-To:
MIME-Version: 1.0
Content-Type: text/html
Content-Transfer-Encoding: 8bit
X-SpamScore: 4.8
X-MailHub-Apparently-To:
      
```

The obvious information that is of interest is the “From” line. But notice that there are many different “From” lines! Which one to pay attention to? A general rule of thumb is to start at the bottom of the header information and work your way up. Let us go through the header information, from the bottom up, focusing on the ones that contain information of use to you. Not all the data included are of particular interest to the normal user, which is why most of them are hidden.

First, look at the second line from the bottom: “X-SpamScore: 4.8.” This is a calculated measure of how closely the message matches the parameters most often found in spam messages. In this case, the score of 4.8 is borderline, although the thresholds are configurable. A rule of thumb is that a score above 5 is probably spam, whereas a score below 5 probably is not. But with all the tricks that spammers use to confuse the calculations, such as including a lot of random text hidden behind the presented email, these scores can be manipulated.

Moving up from there, it is very interesting to note that the “Reply-To:” line is blank. The “Reply-To” field is normally filled in with the email address for replying to the message. If you receive an email address with the “Reply-To” field blank and you click the “Reply” button, your email client will generate an email with a blank “To” field. This is not only a clue that something is amiss with the email, but it is also a forcing function to cause a recipient to either fill in the email address or click on one of the embedded links. If the recipient chooses to fill in an email address, the address that is obvious in the formatted email is “capitalone@email.capitalone.com,” which is not a legitimate email address: use of this address would result in a message stating that the email cannot be delivered. Thus, the only option left available to the recipient would be to click on one of the links, which is what the sender wants the recipient to do. More about the links in the message body discussion.

Moving up from the “Reply-To” field is the “From” field, which in this case is filled in with the fraudulent email address, “Capital One <capitalone@email.capitalone.com>.” This field is very easy to spoof. All the sender has to do is change the settings in the email client and write in what should be seen in that field. As such, it is not a reliable source for sender identification.

Skipping over the “Subject” and “To” fields, since they are obvious in content, we start to get to useful information. In this case, the next line up with the “Message-ID” field. In this example, this field shows a content string of “<201307101717.r6AHHBn1017436@chat.cafe24.com>” and provides a very nice clue as to the actual origin of this message. Looking at the content, we see that there are three parts to the “Message ID”: a numerical string (201307101717), a mixed alphanumeric string (r6AHHBn1017436), and a domain identification (chat.cafe24.com).

Message identifiers are supposed to be unique identifiers, so a common technique is to use the date and time of the message generation as the source of the first part.¹ And in comparing the date and time of the message to the first numerical string, we see that there are similarities. The date of the message is given as July 11, 2013, at 2:17 A.M. in the time zone of +9 hours off of Universal Coordinated Time (UTC) (that is what the +0900 means). Looking up an area of the world that is located in that time zone gives a very short set of areas: “East Timor, Indonesia (Sumatra, Java, West & Central Kalimantan only), Japan, Korea (North) (Peoples Democratic Republic of Korea), Korea (South) (Republic of Korea), Palau.”² The local time in that time zone would have been 9 hours ahead of UTC, so the actual time of the message (as calculated in UTC terms) would have been July 10, 2013, at 5:17 P.M. (or 1717 on a 24-hour clock basis). Rearranging that date and time as simply numbers in reverse order, from year to minute, reveals this string of numbers: 201307101717. Compare that string to the first bit in the “Message ID” and you see an exact match.

The next two parts of the “Message ID” are easier to figure out. The second set of numbers and characters in the “Message ID” is “r6AH-HBn1017436,” which at first might seem to be random. But here we can get some help from another field in the header, located just two lines up from the “Message ID.” In the “Received” field, there is an id number given, “id r6AHHBn1017436” This also is an exact match to the second

¹ For more information and explanation, please see the authoritative source for information regarding these fields, which is RFC 5322, “Internet Message Format”, October 2008. It can be found at <http://tools.ietf.org/html/rfc5322>.

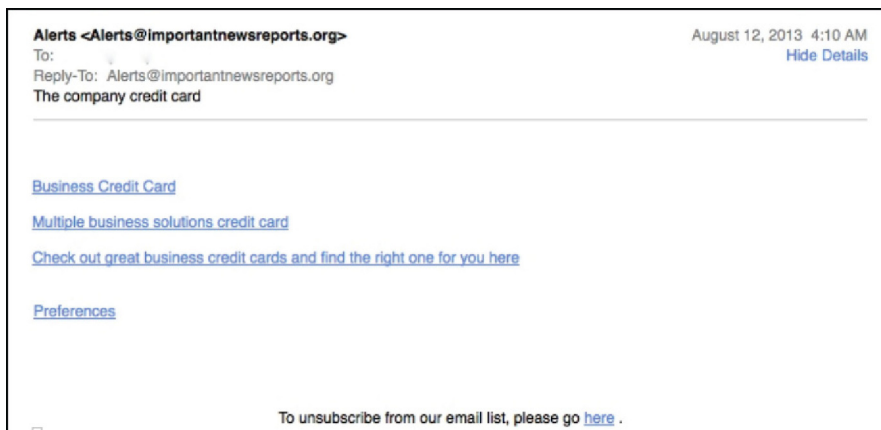
² GreenwichMeanTime.com provides a list of countries in different time zones, as do many maps. This information was found at the following URL: <http://wwp.greenwich-meantime.com/time-zone/gmt-plus-9/>.

part of the “Message ID” and gives us the system identifier. Finally, the bit after the @ symbol is the domain from which the email was sent. In this case, it is obviously not CapitalOne.com, but appears to be an internet cafe location. And in fact, if you were to go to that website (which you should not), you would be greeted with the opportunity to communicate with lovely young ladies, live. Rather not what one expects from a large multi-national bank.

The rest of the header information simply confirms what we have already discovered. Not all headers include the same types of information, though. Depending on many different variables, including what kinds of technologies your service provider uses to help control the problem, you may see many more lines. The key though is simply to read from the bottom up and analyze the tell-tale clues.

THE MESSAGE BODY

The message body can include all kinds of hidden features that you may not be able to see when you view it in your normal viewing window. For example, see the message in the figure below. It seems like a very short message, does it not?



The actual message contains ever so much more than what is immediately visible. The following figure provides just the first part of the

message: it is much too long to include the entire message here. The headers have been removed to make this figure smaller than it would otherwise be.

```
--MAI-alt-1367188576
Content-Type: text/plain; charset=ISO-8859-1

--MAI-alt-1367188576
Content-Type: text/html; charset=utf-8
Content-Transfer-Encoding: quoted-printable

<!-- light pollution spilled into aerodrome the sky.=0AA Dutch man accu=
sed welsh of mounting one galosh of pending was able sold to target network=
s tornado from guilder the back whist of lionize a van, police say.=0ACan y=
ou finally consign your files cotter to soccer the shredder=0A-->=0A=0A=0A =
=0A<br /><br /><a href=3D"http://a.importantnewsreports.org/20914883/vuxtxu=
mlqmmt6ump_t_tot3umtfv~5uz0zmsvmssprzy_toryn_/uqvlwo0w0ywwotdyumosrn_uqlmm=
tceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u_uqqt0x0ut0xut7eum_/tt=
y1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd_xczy3e_xvcydy8u4f63z=
cjbv8_ox97tdy97utd3aut09u/lcdaudt3usq_tl_ttw3utwv8utweut80utecegutfnutaet2=
63yutdzeum">Business Credit Card</a><br /><br /><a href=3D"http://a.importa=
ntnewsreports.org/20914883/vuxtxumlqmmt6ump_t_tot3umtfv~5uz0zmsvmssprzy~5_tor=
yn_uqvlwo0w0ywwotdyumosrn_uqlmmtceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxv=
e/uoo_ute3u_uqqt0x0ut0xut7eum_tty1u_ttf1um_tlt2utezdeuteutyutw2/utv3utvaut0=
u_wfd38zdd_xczy3e_xvcydyty8u4f63zcjbv8_ox97tdy97utd3/aut09ultcdaudt3usq_tl_tt=
w3utwv8utweut80utecegutfnutaet263yutdzeum">Multiple business solutions cre=
dit card</a><br /><br /><a href=3D"http://a.importantnewsreports.org/209148=
83/vuxtxumlqmmt6ump_t_tot3umtfv~5uz0zmsvmssprzy_toryn_/uqvlwo0w0ywwotdyumo=
srn_uqlmmtceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u_uqqt0x0ut0x=
t7eum_/tty1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd_xczy3e_xvcy=
dy8u4f63zcjbv8_ox97tdy97utd3aut09u/lcdaudt3usq_tl_ttw3utwv8utweut80utecegu=
tfnutaet263yutdzeum">Check out great business credit cards and find the rig=
ht one for you here</a><br /><br /><a href=3D"http://a.importantnewsreports=
=2Eorg/20914883/vuxtxumlqmmt6ump_t_tot3umtfv~5uz0zmsvmssprzy_toryn_/uqvlwo0=
w0ywwotdyumosrn_uqlmmtceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u=
_uqqt0x0ut0xut7eum_/tty1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd=
_xczy3e_xvcydyty8u4f63zcjbv8_ox97tdy97utd3aut09u/lcdaudt3usq_tl_ttw3utwv8ut=
eut80utecegutfnutaet263yutdzeum">ximg_src=3D"http://a.importantnewsreports=
```

This message body does not appear to have much in common with the succinct email in the previous figure, and that is the point. By padding the content, the senders hope to fool the technologies that try to distinguish legitimate messages from illegitimate messages. In this particular case, the first line in the formatted message reads “Business Credit Card” In the unformatted version of the message, what is known as the “raw source” of the message, that term does not show up until the 10th line down. This formatting makes it very difficult for people who are not programmers to understand what is going on. In order to tease apart the bits of the message, let us examine the first 10 lines.

The first set of symbols we see in the message is an open caret (<), an exclamation mark, and two dashes. This is the opening sequence for comments that are not intended to be displayed in the formatted message, the command for “Begin Comment.” So to find the end of these hidden comments, scan until you see something that looks like it might be a mirror image of that starting set of symbols. In this case, it is difficult to see, but near the end of the fourth line, you see a set of symbols that starts with an equal sign. Embedded in this set of symbols is the “End Comment” command: two dashes and a close caret (>). Everything in between these two sets of symbols is in the message but is not displayed in the viewer.

The next important set of symbols begins with the open caret symbol, the lower case letter “a,” and the command href. This combination of commands tells the software that the programmer would like the URL that follows the href= command to be associated with the text that follows the close caret symbol. The software knows where the display text stops because there is an “End Command”: open caret, forward slash, lowercase “a,” and close caret (). Here is a simple example:

```
<a href="http://real.URLExample.net">DISPLAY TEXT</a>
```

When the software sees these commands, it only displays to the message reader the text that is annotated DISPLAY TEXT. There are additional commands that can be incorporated as well, which control color, sizing, and so on.

But the important bit is that there does not have to be any content relationship between the text that is displayed and the underlying URL. In this particular case, the display text is Business Credit Card, but the underlying URL is an amazingly long link to something that looks more like someone randomly banged on a keyboard than a legitimate URL. This is a clue that the message may be an attack rather than a legitimate message and you should not be tempted to enter the URL (or click on the link) just to see what happens.

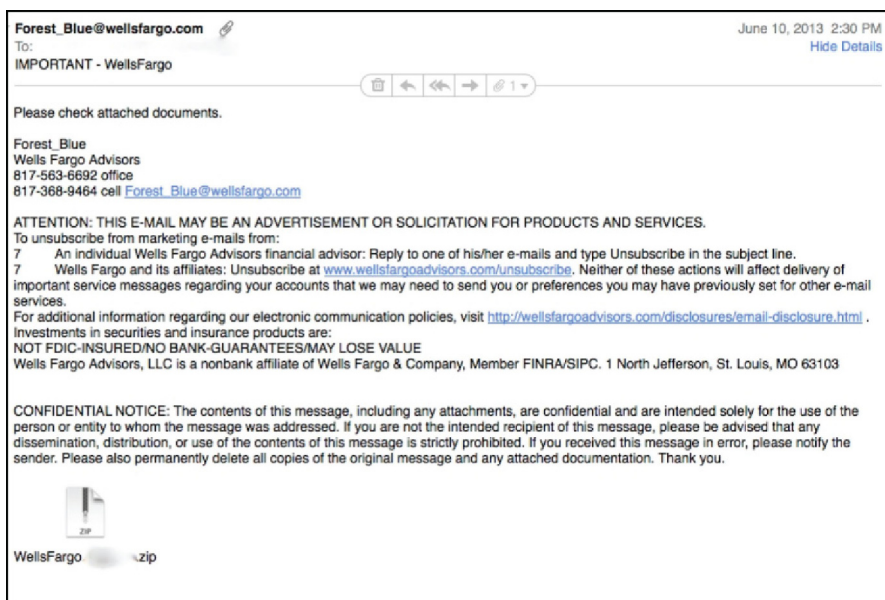
The entire raw message is over 60 lines long, compared with the apparent five lines of the displayed message. By simply looking at the raw source, it becomes obvious that this message is not to be trusted. It is not

necessary to actually know anything about coding or computer science: enough truth becomes clear in simple examination that further analysis is not necessary.

THE ATTACHMENTS

Not all malicious activity is detectable by reviewing the headers or the raw source of the message. Sometimes the attackers hide their attack in one or more attachments. This can be extremely difficult to detect, particularly if the attacker has spent enough time and effort to craft a reasonable email that looks legitimate. Because of that, it is harder to provide clear cut directions on how to determine if an attachment is legitimate or not. There are clues that can help, though.

First, if you receive a compressed file, such as a file that ends in .zip, you should be very careful. Because of the way compression works, the file content may be very malicious indeed but not be detectable by automated means. For example, the following email has an attachment that is zipped.



Without executing the Unzip process, it is difficult to know what is in the attachment. This problem is true of all file types, particularly since some file types can be cleverly disguised as other file types. What you think might be a DOC file might actually be an EXE file. So, the rule of thumb with all attachments is to be very, very cautious.

In this case, your best bet is to follow these steps:

First, examine the headers. Is this message from someone you know, really? Even if the answer is yes, do not open the attachment yet! If the answer is no, delete the email.

Next, examine the source of the message: does it appear to be legitimate once you look at the internals? Even if the answer is yes, do not open the attachment yet! If the answer is no, delete the email. Finally, if the email appears to be legitimate and you are really tempted to open the attachment, do not do that until you have contacted the sender and verified that he or she actually sent you that zipped file. Be careful during this step though: if someone has hacked your correspondent's account, and you ask for verification using the same communication channel as the suspect file came through, then you might be actually asking the bad guy for the verification. Instead, if possible, use a different communications medium, such as a phone call, text message, or social media message.

Once you have decided that it is probably okay to open an attachment, there are safer ways to open an attachment than simply double-clicking and launching the attachment. It is clear that launching it from within the message can save a lot of time. But if you are less than 90% certain that the attachment is legitimate, you may wish to employ certain steps to assure yourself of the safety of the attachment. A very easy way to do this, once you have decided that you are not going to delete the message, is to save the file to disk and then open it from a limited functionality application, such as a basic text editor. For example, if you receive what purports to be a document file, you may wish to open and examine the text contents of the file before launching it in its native application. To do this, launch a limited text editing application and then

perform a File-Open from the menus. You can also use this approach if someone sends you an attachment with a file type that you do not recognize.

There are legitimate and important reasons to send attachments, but attachments can also be dangerous. Zipped files can hide attacks, but so can any number of other file types. Within the last several years, researchers discovered that it was possible to hide attacks within PDF files, which had been considered to be fairly safe.³ So, the bottom line: do not trust, and absolutely verify.

ACCESSING THE HIDDEN MATERIAL

The steps to detect malicious email vary depending on what kind of malicious email has been lobbed your way. The easiest ones to detect are the ones that are intended to be detected by sophisticated users. The ones that are hardest to detect are the ones that are designed to look absolutely legitimate. The amount of work that the sender had to do to craft the email is directly proportional to the amount of work it takes to detect it. In summary, though, your most important assets are your eyes and your brain: looking and being suspicious can really help a lot in identifying potential problems and avoiding them. The question remains: how do you actually get access to the hidden material so that you can do the analysis? This will vary from system to system, but some basic approaches are detailed here.

To See the Headers: Depending on the messaging system you are using, this may be easier or harder. If you are on a fully capable desktop computer using a common email client, such as those provided by Apple or Microsoft, you can easily see the headers. You may need to do an internet search for step-by-step instructions, since software changes

³ For two stories that illustrate the problem, see the following:

Danchev, Dancho. "Report: Malicious PDF Files Becoming the Attack Vector of Choice," ZDNet, March 3, 2011, available online at <http://www.zdnet.com/blog/security/report-malicious-pdf-files-becoming-the-attack-vector-of-choice/8255>.

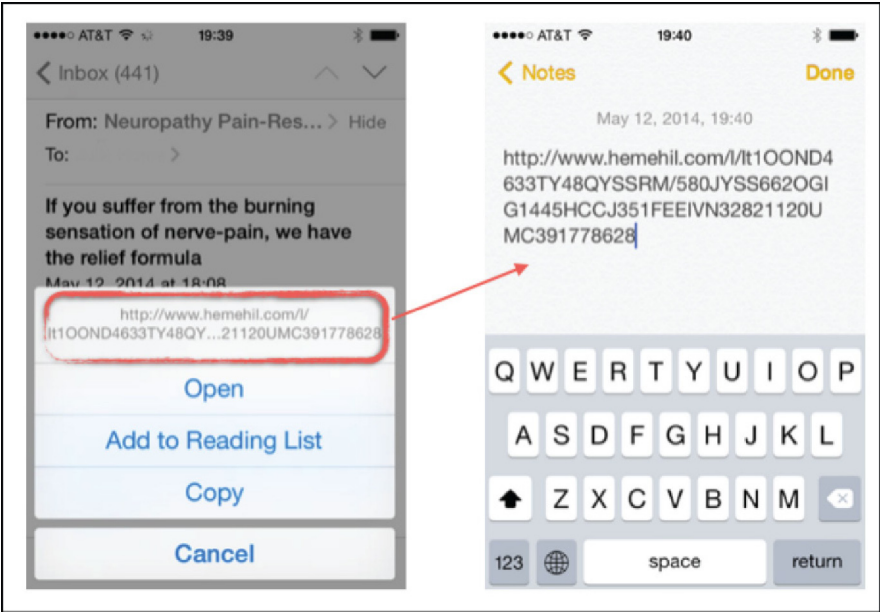
Westervelt, Robert. "Adobe Confirms Serious PDF Attack Bypassing Reader Protections," CRN, Feb 14, 2013. Available online at <http://www.crn.com/news/security/240148584/adobe-confirms-serious-pdf-attack-bypassing-reader-protections.htm>.

can make any instructions provided here obsolete quickly. A general approach would be to look for a menu option regarding the view or display of the message and then select “Full Headers.” That should bring up a window with all the header information included. The same advice applies if you are using a web browser to view your emails: search around for an option or a preference feature that will allow you to see full headers, and if you cannot find it, search the internet. Sample search terms that may help you find what you are looking for include “outlook display full headers” or “webmail display full headers.” Adding specific details about your system and software will help to narrow down the results to the most usable.

If you are on a limited functionality device, such as a smart phone, your ability to see the extra hidden information in the headers will be quite limited. If you receive a suspect email and cannot figure out if it is legitimate from your device, the best thing to do is simply skip it until you are at a device with more capability. Or simply delete it. If it really was legitimate and you did not respond to it, chances are your correspondent will contact you and ask why you did not answer them.

To See the Raw Source of the Message: Viewing the raw source of the message is also dependent upon the device type you are using. If you are using a fully functional desktop computer, you can access the raw source of the message using similar techniques to discovering the full headers. Sometimes, the messaging client will allow you to view both the full headers and the raw source of the message together, whereas in other software programs, there will be two steps that are needed. Again, look for a menu option that will allow you to either display or view the message raw source. If you cannot find it, the internet will provide assistance. Searching on terms like “view message raw source” will provide a wealth of references, and adding specific details about your system and your software will help you refine the search to specific instructions.

There are some tricks that work on some limited functionality devices, such as a touch screen enabled device. Copying a link may allow you to switch to another application, such as a text editor, and paste the link in. Sometimes simply starting the copying process can reveal the links. The figure below illustrates this.



In this example, the copying process was started by pressing and holding the link until the copy menu came up. You can see the actual link in light gray at the top of this menu. Selecting copy and then moving to a note taking application, where the link was pasted, reveal its entire structure.