Computer Science Department
The University of Hong Kong
Final Year Project 2019-20

SMART EMAIL CLIENT TO DETECT MALICIOUS URLS

Group: 2
Name: Shreya Palit (3035346556), Trisha Gupta (3035342419)
Supervisor: Dr. S.M. Yiu

Table of Contents

1. Project Background

Everyday, we receive a plethora of emails every day which may contain some URLs. These emails may be malicious. A malicious email is one which may probe us to redirect to unexpected sites or cause some harmful software to be downloaded. To give an example of a malicious email, an email might contain some special or secret information with the recipient that will help with his health issues, financial issues, or other common problems, like a secret cure for high blood pressure and high cholesterol that the medical establishment does not want you to know about. Many of these emails do not provide details: you have to click through to get the secret. Once you click on this malicious URL, it can redirect you to websites in order to steal your sensitive information like Credit Card details, passwords etc. It is easy to look at an email and not see what is hidden behind the display, but what is behind it is some complicated programming. Just a little bit of knowledge, it can be understood how this works, and leveraging this knowledge a better alert system can be built to avoid these potential problems. Our solution hence is to build a new (or modify an existing open-source) email client.

2. Project Objective

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. Hence the objective of our project is to build a new email client or modify an existing email client source. The main goal of the email client is that it can read through the contents of emails and perform analysis on the URLs that are included in the email. The analysis can be based on any existing online URL checker or our own AI algorithm. Warning messages can then be issued to the user accordingly as to whether the URL or attachments in the email are safe to open or whether they are harmful.

3. Project Methodology

In order to detect and combat malicious URLs it is paramount to understand the goals and intentions of the sender.

3.1 Targeting:

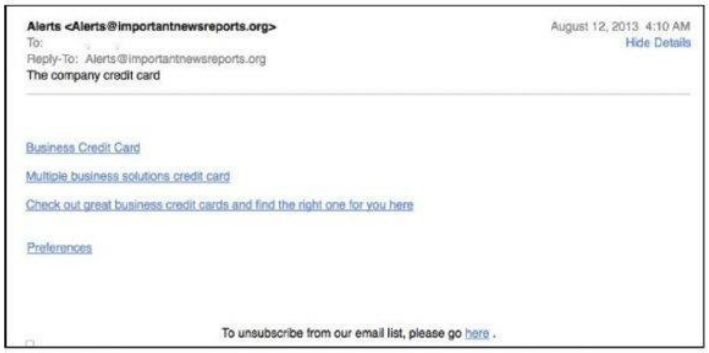
The target of a malicious email may either be to one specific person (TME, also known as spear fishing) , or may be sent to multiple recipients (UnTargeted Malicious Email (UTME). The one targeting multiple recipients can be more easily identified as malicious as opposed to the one targeting one specific person. This is because of the cost trade-off to sender and is usually only used when a high level of information is to be extracted. For example, a high technology enterprise may be vulnerable to this attack as the person may hold database access such as plans of marketing, client details or other sensitive data.

3.2 Addressing:

In order for senders to camouflage their identities, they use certain hiding techniques in order to mask the email source, such as copying someone on the recipient’s contact list, or a celebrity.

3.3 Content:

The content of emails holding malicious URLs may vary. 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, messages often seem very short (see Figure 1). However, the actual message contains way more than what is usually visible (see Figure 2).



Alerts <Alerts@importantnewsreports.org> August 12, 2013 4:10 AM
To:
Reply-To: Alerts@importantnewsreports.org
The company credit card

Business Credit Card

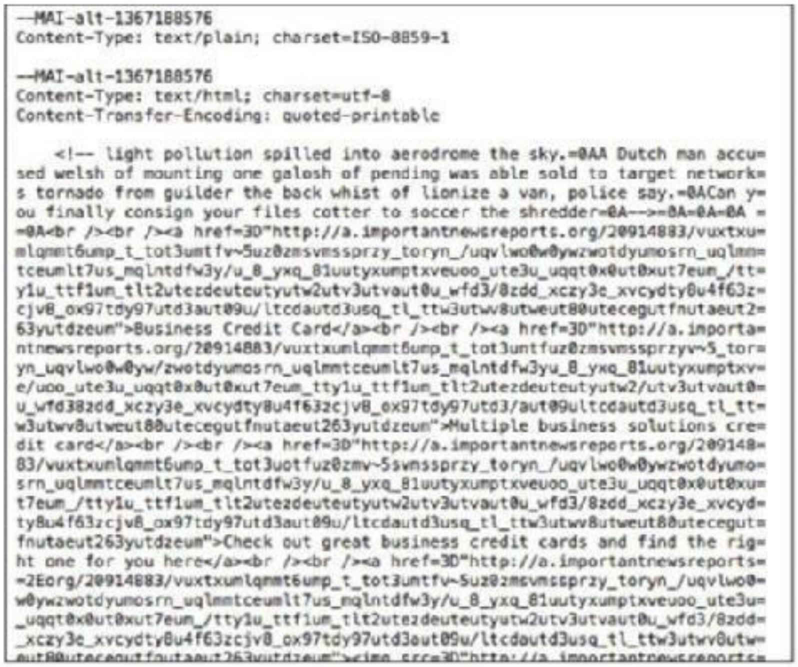
Multiple business solutions credit card

Check out great business credit cards and find the right one for you here

Preferences

To unsubscribe from our email list, please go here .

Figure 1: An email containing just malicious URLs



```
--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_/uqvlwo0w0ywzwotdyumosrn_uqlmm=
tceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u_uqqt0x0ut0xut7eum_/tt=
y1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd_xczy3e_xvcydty8u4f63z=
cjv8_ox97tdy97utd3aut09u/ltcdautd3usq_tl_ttw3utwv8utweut80utecegutfnutaeut2=
63yutdzeum">Business Credit Card</a><br /><br /><a href=3D"http://a.importa=
ntnewsreports.org/20914883/vuxtxumlqmmt6ump_t_tot3umtfuz0zmsvmssprzyv~5_tor=
yn_uqvlwo0w0yw/zwotdyumosrn_uqlmmtceumlt7us_mqlntdfw3yu_8_yxq_81uutyxumptxv=
e/uoo_ute3u_uqqt0x0ut0xut7eum_tty1u_ttf1um_tlt2utezdeuteutyutw2/utv3utvaut0=
u_wfd38zdd_xczy3e_xvcydty8u4f63zcjv8_ox97tdy97utd3/aut09ultcdautd3usq_tl_tt=
w3utwv8utweut80utecegutfnutaeut263yutdzeum">Multiple business solutions cre=
dit card</a><br /><br /><a href=3D"http://a.importantnewsreports.org/209148=
83/vuxtxumlqmmt6ump_t_tot3uotfuz0zmv~5swnssprzy_toryn_/uqvlwo0w0ywzwotdyumo=
srn_uqlmmtceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u_uqqt0x0ut0xu=
t7eum_/tty1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd_xczy3e_xvcyd=
ty8u4f63zcjv8_ox97tdy97utd3aut09u/ltcdautd3usq_tl_ttw3utwv8utweut80utecegut=
fnutaeut263yutdzeum">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=
w0ywzwotdyumosrn_uqlmmtceumlt7us_mqlntdfw3y/u_8_yxq_81uutyxumptxveuoo_ute3u=
_uqqt0x0ut0xut7eum_/tty1u_ttf1um_tlt2utezdeuteutyutw2utv3utvaut0u_wfd3/8zdd=
_xczy3e_xvcydty8u4f63zcjv8_ox97tdy97utd3aut09u/ltcdautd3usq_tl_ttw3utwv8utw=
eut80utecegutfnutaeut263yutdzeum"><img src=3D"http://a.importantnewsreports=
```

Figure 2: Actual email content

As we can see, there need not be any direct 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 different. This is a big clue that the message may be an attack rather than a legitimate message and the URL should not be opened.

4. Project Schedule and Milestones

Given below is the tentative schedule of the project which shows when and what will be achieved at various stages of the project.

Timeline	Milestones	Deliverables
September	Proposal	• Project plan • Project webpage
October - November	Design	
Late November - Mid March	Development	
January		• First Presentation
February	Phase 2	• Preliminary Implementation • Detailed interim Report
April	Phase 3	• Finalized tested implementation • Final Report
End April	Deployment	• Final Presentation