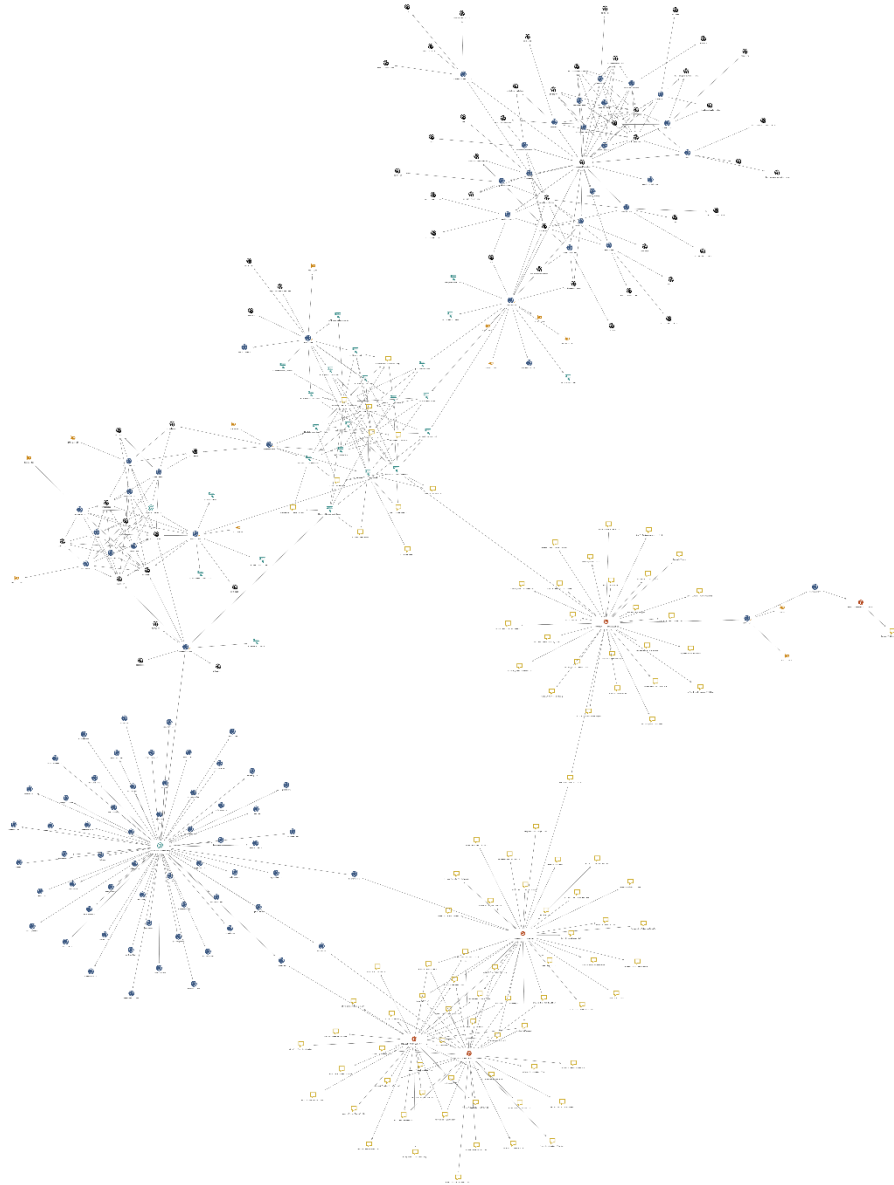


# Using Maltego to Enrich Tickets for Incident Response



Chris Jenks and Michael Mitama

Based on malware analysis done by Brad Duncan at Malware-Traffic-Analysis.net and J at Techhelplist.com

## Table of Contents

1. Why do this? .....	3
2. Maltego .....	3
Setting up Maltego .....	3
Starting a New Graph.....	4
Couple of notes on Maltego .....	5
Under the investigation tab:.....	5
Under the Collections tab: .....	5
Rate limiting .....	6
Changes over time .....	6
3. Domain Searches .....	6
Starting the investigation: .....	6
VTPub Domain Resolutions .....	7
VTPub Communication and Download Samples.....	8
4. IP Address Searches.....	13
VTPub IP Resolutions .....	13
5. Whois .....	18
6. Manual Linking.....	24
Linking Nodes .....	24
Combining Graphs .....	27
7. Example of a finished case .....	28
8. Exporting Data.....	30
CSV .....	30
PNG.....	31
PDF.....	33

## 1. Why do link analysis to enrich tickets?

Link analysis allows analysts to have more information about items they are investigating. Analysts can look at domains, find all the related files hashes and map out the connection for each item. This allows analysts to map all the domains to an owner, which allows for blocking future sites hosting malware. Link analysis can create a map for management, showing the who, and the how of attempted network attacks.

Using the data provide by different Open Source Intelligence sites, allows analysts to make decisions about the items being investigated. Some Open Source Intelligence will give the analyst a historical view of the domains to see when the last time it was updated, and if the site previously blocked is safe to unblock.

Exporting the linked data in a CSV file creates a list of possible Indicators of Compromise for the Hunt Team to go search. IR staff can take the list and create network blocks, either email, proxy, or other.

Link analysis track and map information on domain owners, hosting sites, and name servers. This data can be captured as artifacts for tracking in the tickets. The artifact collection can be used to show if attacks are one offs or targets of opportunity vs targeted campaigns.

### The Walkthroughs

The data used in the examples and walk through the data came from blog posts on Malware Traffic Analysis and Tech Help List:

<http://www.malware-traffic-analysis.net/2017/02/28/index3.html>

<https://techhelplist.com/spam-list/1107-2017-02-28-shipment-status-change-notification-for-parcel-malware>

## 2. Maltego

Maltego is a program for doing link analysis. Links can be created manually by the user, or by “Maltego transforms” which scrape publically available data. As the transform runs, the graph will automatically update, linking new related entity nodes.

Several third parties have transforms that can be accessed from Maltego. Some will require accounts on their sites and API key use their transforms.

Maltego also has some pre-built “machines” that will automate most of the searching functions based on what the user selects and fills in. These are outside the scope of this document. Make sure to un-click show at start up, or it will ask every Maltego starts.

### Setting up Maltego

For enhancing tickets, the following transforms sets are recommended. Some will require API Keys, which come with either free or paid accounts. The transforms that use the free keys, usually have a lower daily limit. Paid transforms usually have a larger limit. The ones suggested in this document are:

- Paterva CTAS
- VirusTotal Public API (Needs API Key)
- Shodan (Needs API Key)
- ThreatMiner
- ThreatCrowd
- Passive Total (Needs API Key)
- ThreatGrid (Needs an API Key)

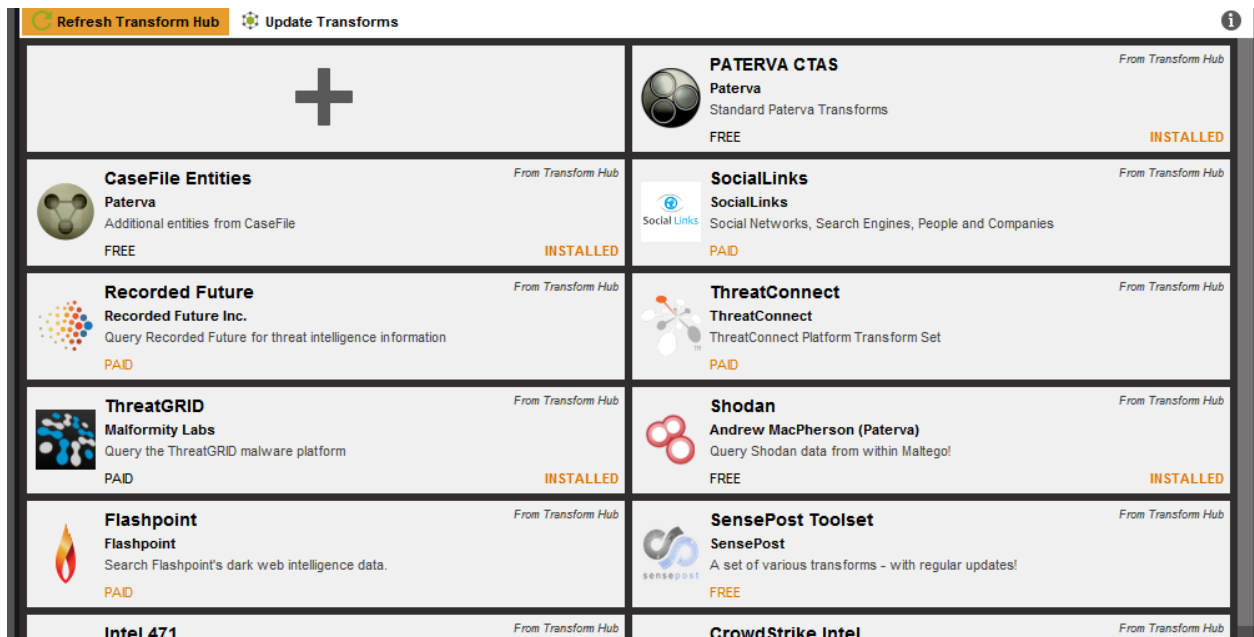


Figure 1 Example of Transforms

In the above screen shot, paid means that the API key is part of a Paid account. However some of them also depend on what service level is purchased. Some, like Threat Grid require an additional cost to use the API key on top of their normal sandbox tool.

### Starting a New Graph

Start a New graph and set your Number of Results from each Transform (search). CE is limited to 12, while Classic has 10,000, and XL 1,000,000. In classic 255 seems to be a good middle ground to work with. The new graph and number of results can be found in the upper left corner.

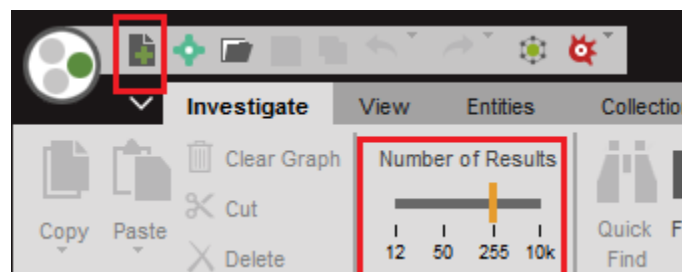
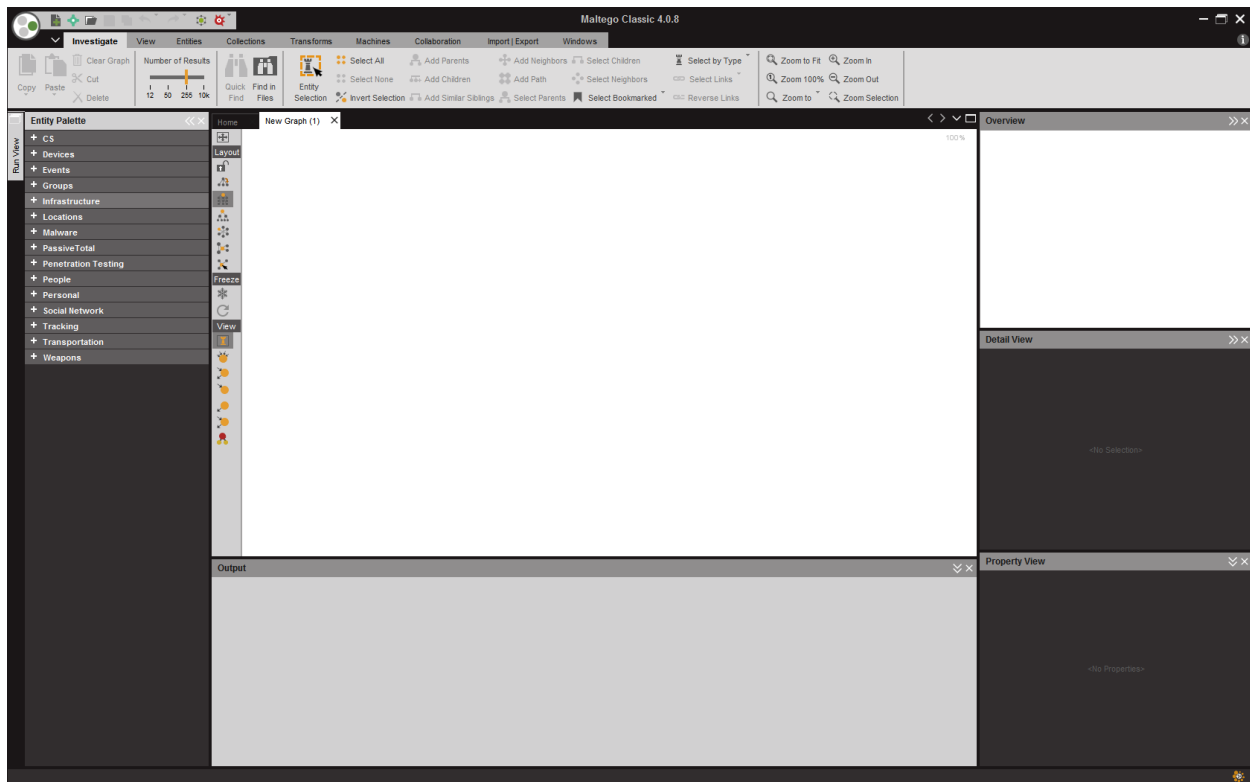


Figure 2 New Graph and Number of Results

Once started, the screen will look similar to this.



*Figure 3 Maltego Ready to start*

## Couple of notes on Maltego

### Under the investigation tab:

The number of results is set to 255. This means regardless of how many results there are, searches will only bring back 255 results. This can be changed, but will impact how much can be searched. Some APIs, like VirusTotal, have daily limits. So far 255 has seemed to be a good break point.

### Under the Collections tab:

Related nodes will be grouped, based on what is set under simplify graph. This number will change throughout an investigation. During the initial parts, when still running transforms, it is best to leave this set to 25. This will make it easier to run transforms on large groups of data, since the graph changes in real time. Selecting a collection and then the transform will run the search on all grouped items.

Once the analyst has finished running all the transforms, it is best to break up the collections to get a fuller picture. This can be done by changing the collection slider to never, which is the fastest way in graphs with lots of nodes. The other option is to use the push pin in the corner of the collection. This pins the individual nodes to the graph. To undo this, unpin items either one at a time or by selecting multiple nodes and unpinning.

### Rate limiting

To prevent abuse, some of the API transforms are rate limited. For example VirusTotal's documentation says that the API is limited to 4 requests a minute. Running a transform on a collection will take that in to account and do 4 every minute. For example running a transform on 100 items in a collection will take about 25 minutes to complete.

### Changes over time

Lastly changes do appear over time. Running a search on the same three times IP Address has created 2 completely different maps, and instance with no map. The first map had nodes associated with multiple malware samples. On the second map everything appeared to be associated with only one malware sample. The last map was created during an AWS / Amazon S3 outage, which may be part of the reason for the lack of results.

Because the data changes, take screenshots, export to CVS, and save the graph, to attach to the ticket.

## 3. Domain Searches

Normally when reviewing domains, there is only one from the ticket to look at. In this case all the associated domains will be looked at, and linked together if needed.

The two domains from the blog posts we will not look at are `api[.]ipify[.]org` and `checkup[.]dyndns[.]org`. All links shared have been defanged with `[]` around the dots. To use these domains, remove the brackets. These were done to for the sole purpose to check outbound traffic, and possibly to see if it was a known address range for malware research.

### Starting the investigation:

Select the domain icon from under Infrastructure and drag it in to the new graph.

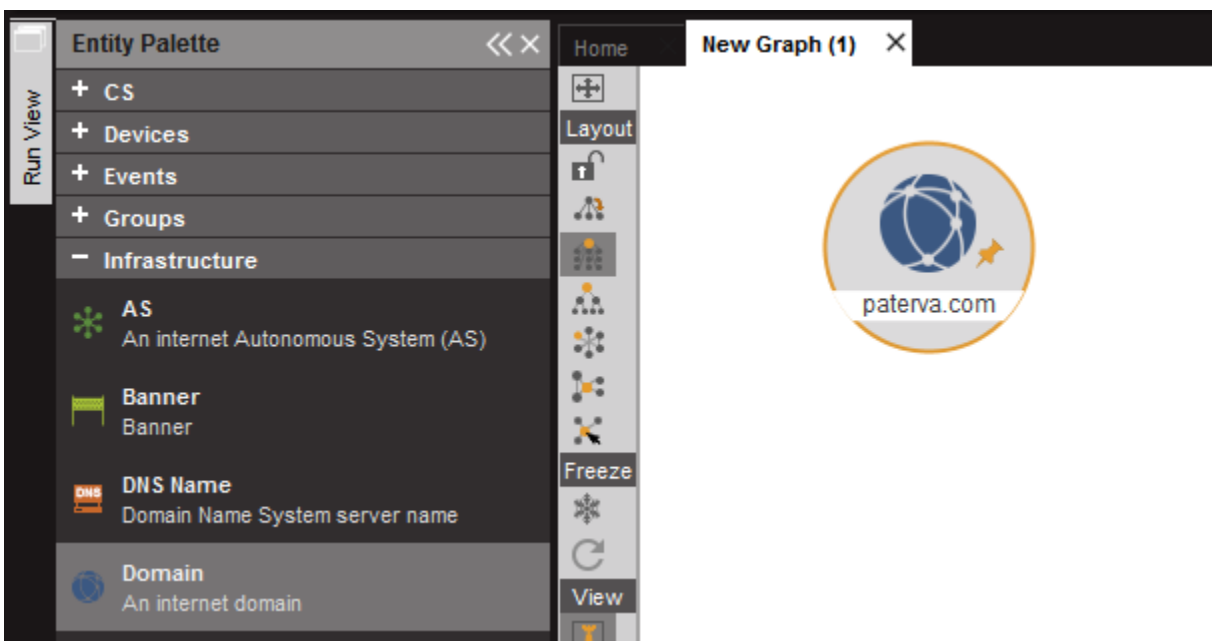


Figure 4 New Domain Lookup

Change the name from paterva.com to the node that is being looked up. In this case change the domain to **canmake[.]vn**. This step will be repeated later with the other domains on the site, and the domains linked based on comments in the blog posts.

### VTPub Domain Resolutions

The first one to start with is Domain Resolutions under the VirusTotal Public API transforms. Listed here are all the transforms that can be ran against VirusTotal for domains.

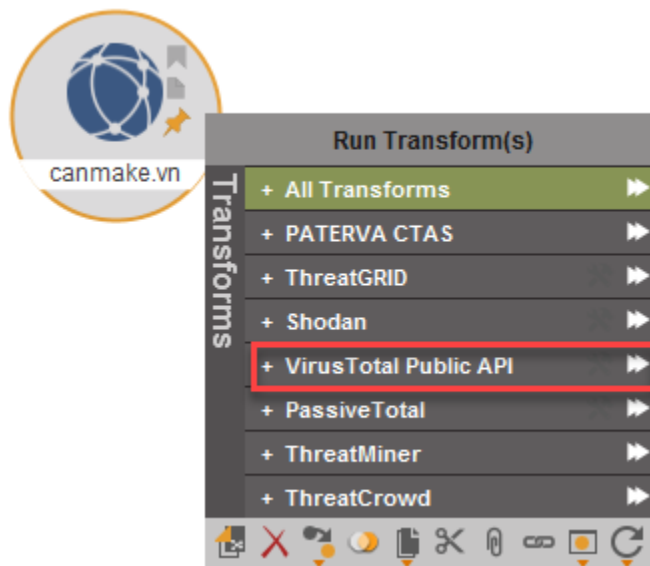


Figure 5 Loaded transforms for domains

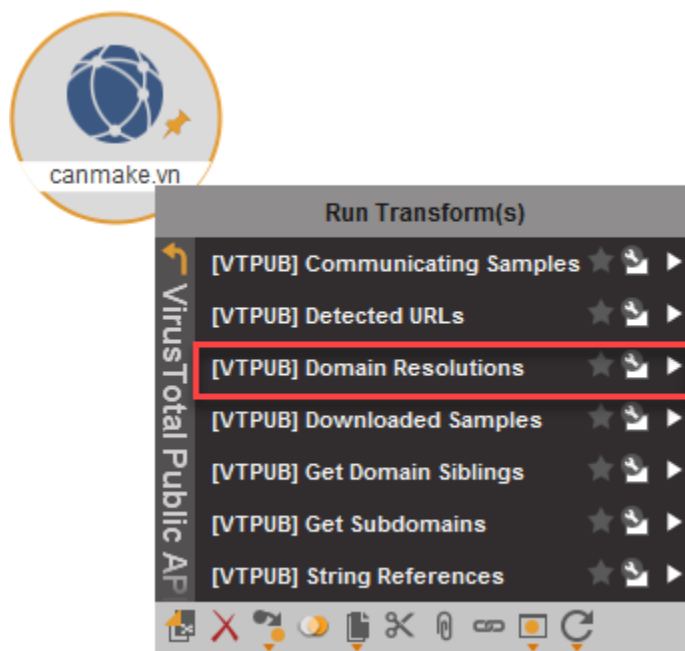
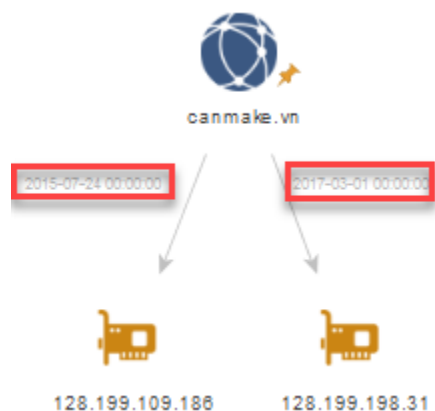


Figure 6 VT Transform List for Domains

This should return a list of the IP addresses known by VirusTotal for this domain. In the case of **canmake[.]vn**, VirusTotal knows two different IP Addresses for the domain. The links from the domain node to the IP address nodes contain dates of when those IP addresses were seen by VirusTotal.



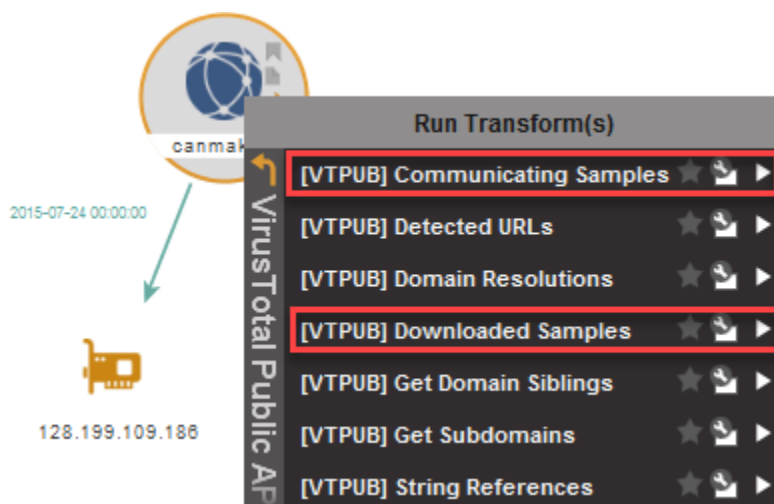
*Figure 7 Domain Resolution transform results*

This may be something useful later, if there are no communication or download samples associated with the domain, or sub domains. In that case, the investigation will come back to the IP addresses to search for the Samples.

### VTPub Communication and Download Samples

The next step is to look at the Communication and Download Samples. This may be more complicated than just running those two transforms on the domain. If there are no results, then the subdomain transform will need to be run as well. It may be worth running anyway depending on investigation needs.

In this case, the original link in the phishing email associated with the blog entry went to a web page running on **canmake[.]vn**.



*Figure 8 the two samples to run for data*



In the case of **canmake[.]vn** no communicating samples were seen and there is only one download sample.

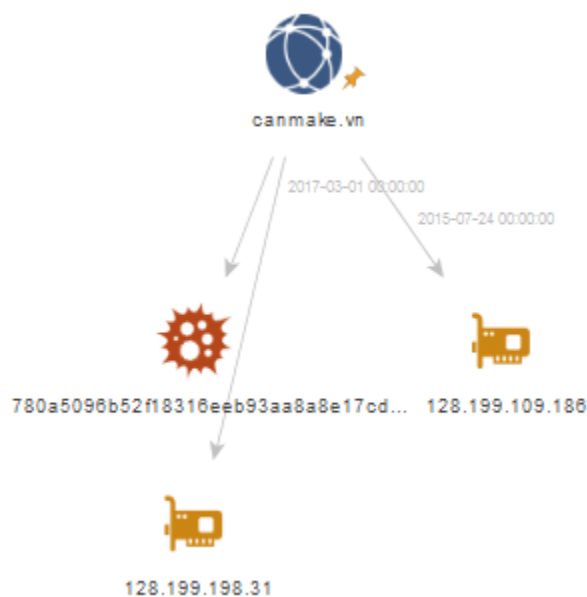


Figure 9 VT Sample Results

```
Running transform [VTPUB] Communicating Samples on 1 entities (from entity "canmake.vn")
Transform [VTPUB] Communicating Samples returned with 0 entities (from entity "canmake.vn")
Transform [VTPUB] Communicating Samples done (from entity "canmake.vn")
Running transform [VTPUB] Downloaded Samples on 1 entities (from entity "canmake.vn")
Transform [VTPUB] Downloaded Samples returned with 1 entities (from entity "canmake.vn")
Transform [VTPUB] Downloaded Samples done (from entity "canmake.vn")
```

Figure 10 VT Samples Transform output

Before moving on, run the subdomain transform since there was so few samples.

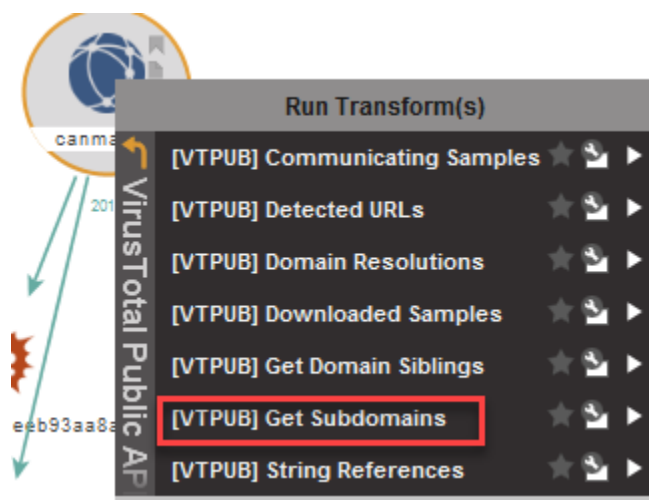


Figure 11 VT Get Subdomains

This transform returned one subdomain. That subdomain was **www[.]canmake[.]vn**. Re-running the two “samples” transform only returned one additional file hash.

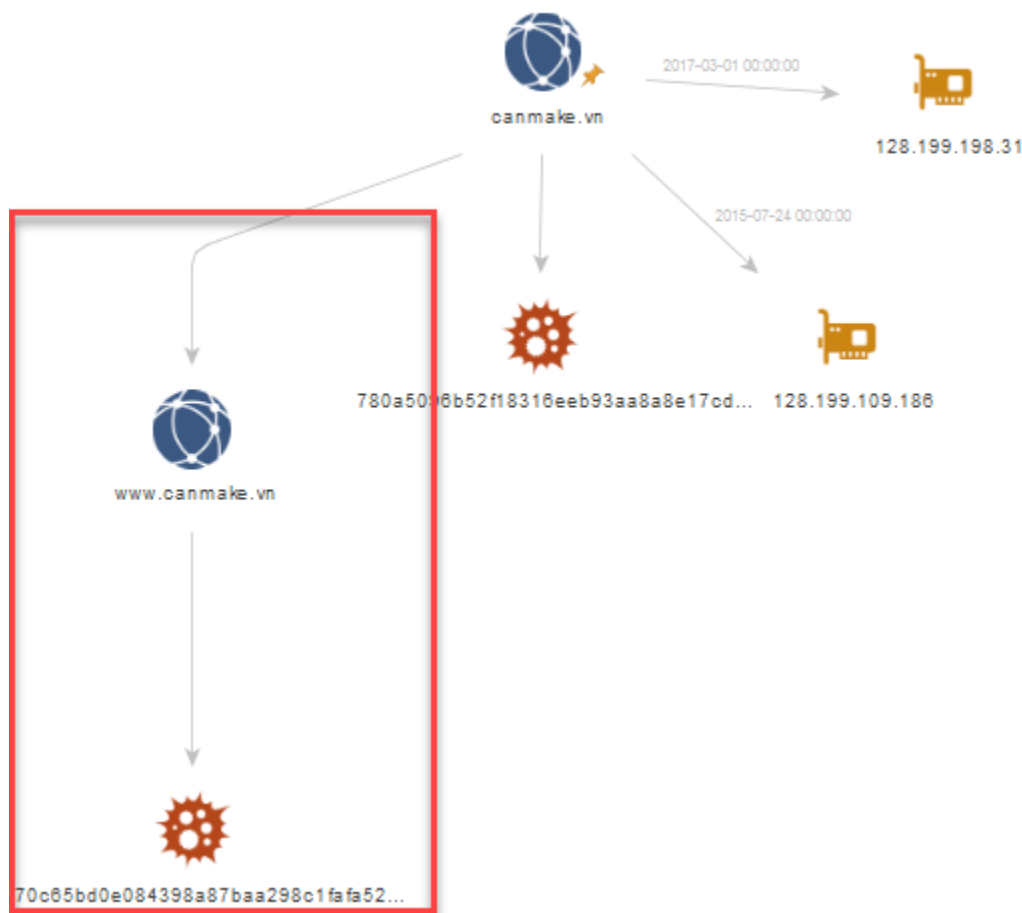


Figure 12 New subdomain with single hash

The next step is to select the two hashes nodes and run [VTPub] Check Hash Report. Since these nodes are on different branches, Maltego makes it hard to select both. To get around this, under the Investigate tab at the top, select “Select by Type” hash. Both nodes are selected and can be searched with a single right click on either node.

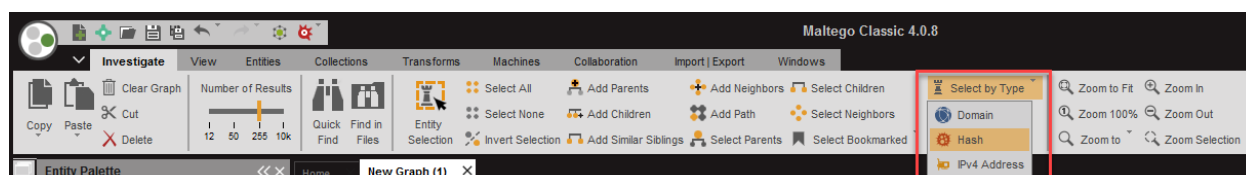


Figure 13 Select by type

Running this transform will copy the details of the hash, and make the information available on the graph. The transform in this case is based on the API being used. Public API uses VTPUB, while a private API would use the vtprivCheckHash.

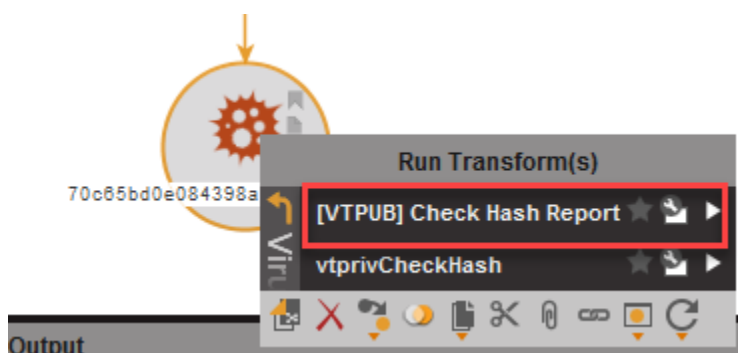


Figure 14 Check Hash Report

The hash reports contains different information on the files associated with the hashes. Such as what kind of file the hash is in the name and the date it was seen by VirusTotal.

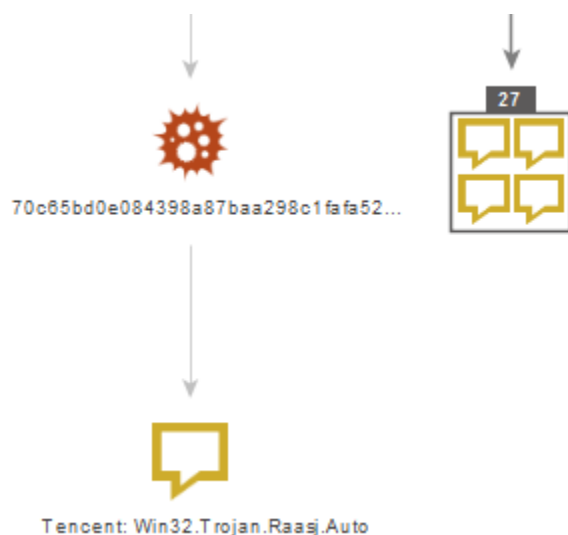


Figure 15 Hash information

Notice above, the box with 27, and the multiple phrase nodes in it. That is a collection. On the graph, zooming in to the area around that box in will expand the collection slightly, into a scrollable list as. See below.



Figure 16 Zoomed Collection

In the final product that would be expanded, so there are no collection.

At this point the analyst has to make a decision. Do they want to investigate the other domains in the same graph or make a new one and go there? The steps in either instance would be the same ones that lead to here, but doing it one graph might find additional overlap between files, phrases, or IP addresses.

Example of the finished graph, using just **canmake[.]vn**. In this graph, the collection was set to never, and the set layout mode to organic was selected. Normally an analyst would highlight different aspects that they wanted to call out, such as when malware type changes, when it's modified but still the same family, etc.

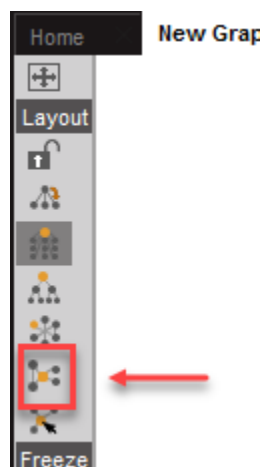
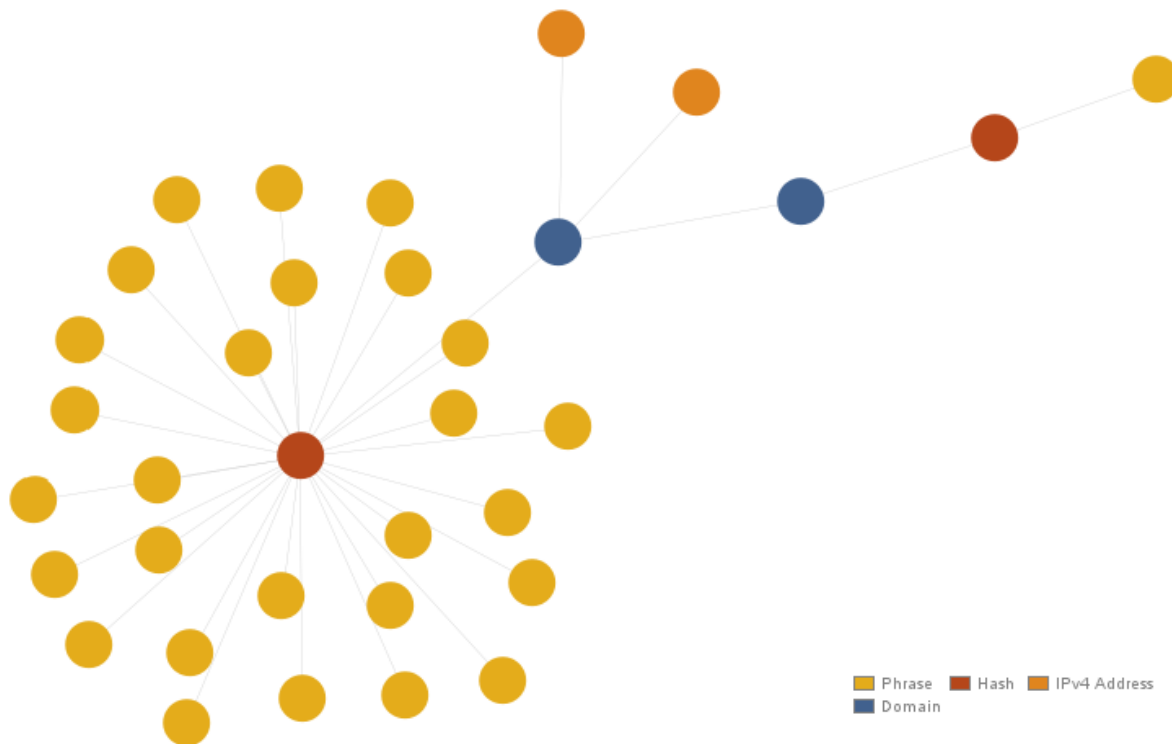


Figure 17 Layout selection for Organic

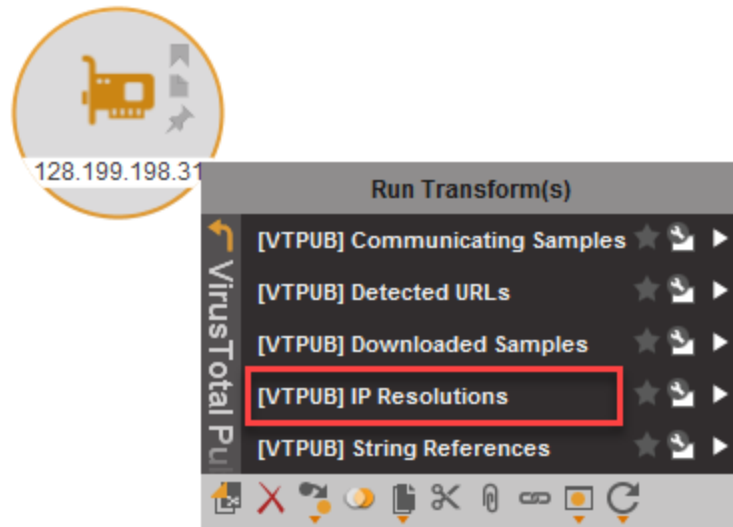


*Figure 18 Organic representation of malicious files associated with canmake[.]vn*

## 4. IP Address Searches

### VTPub IP Resolutions

In some cases an analyst starts with an IP address instead of a domain. For this example, the IP Address is the newest one seen from the **canmake[.]vn** IP address.



*Figure 19 VT IP Resolutions*

This resulted in the IP Address resolving to 13 different DNS nodes. Some were domains, and some were sub-domains. While they all overlap, notice in the picture below that all the links have dates associated with them. These are the dates that VirusTotal associated the IP Address and the domain.

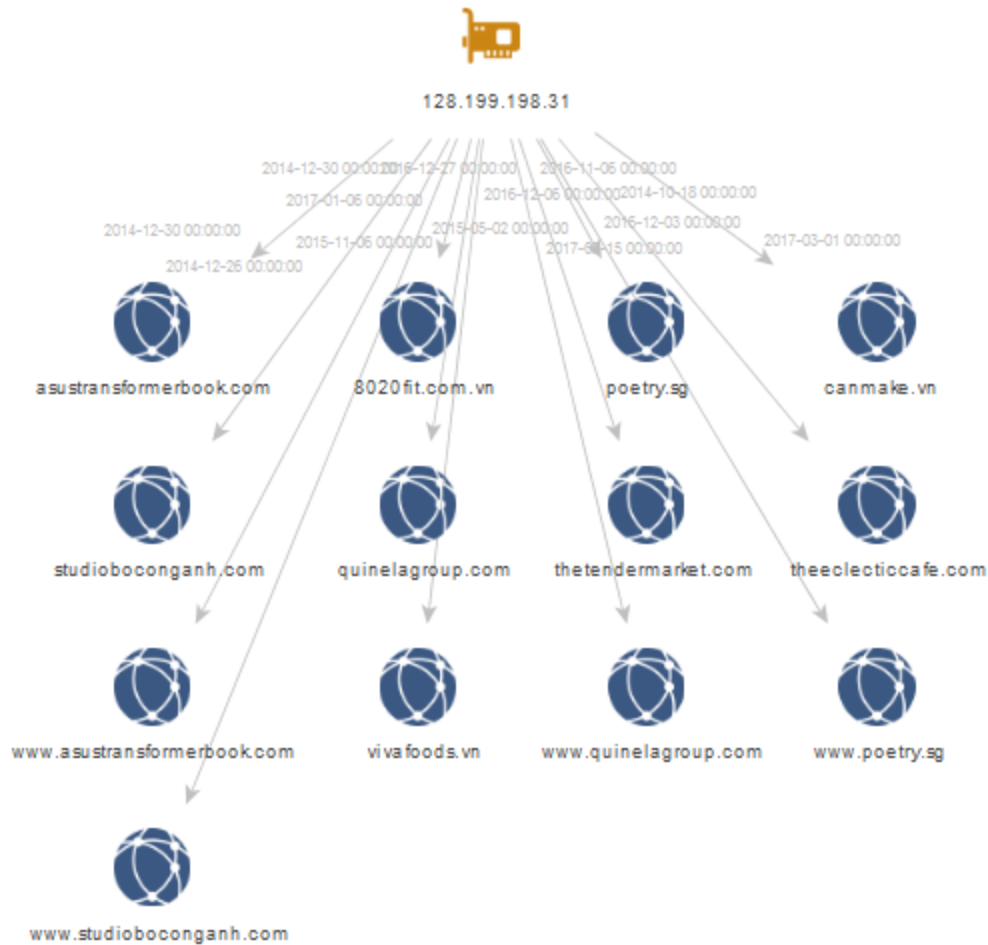


Figure 20 Results of IP Resolutions

In this case, interest is only in the most recent domains associated. To find those, use the Domain entity under Select by Type in the Investigate tab, and then Incoming. This will select all the links, as seen in the next two screenshots.

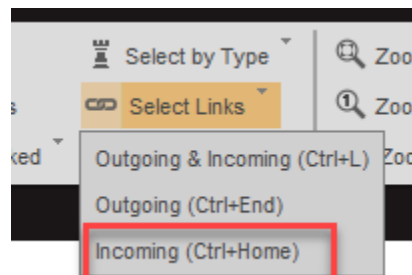


Figure 21 Select links

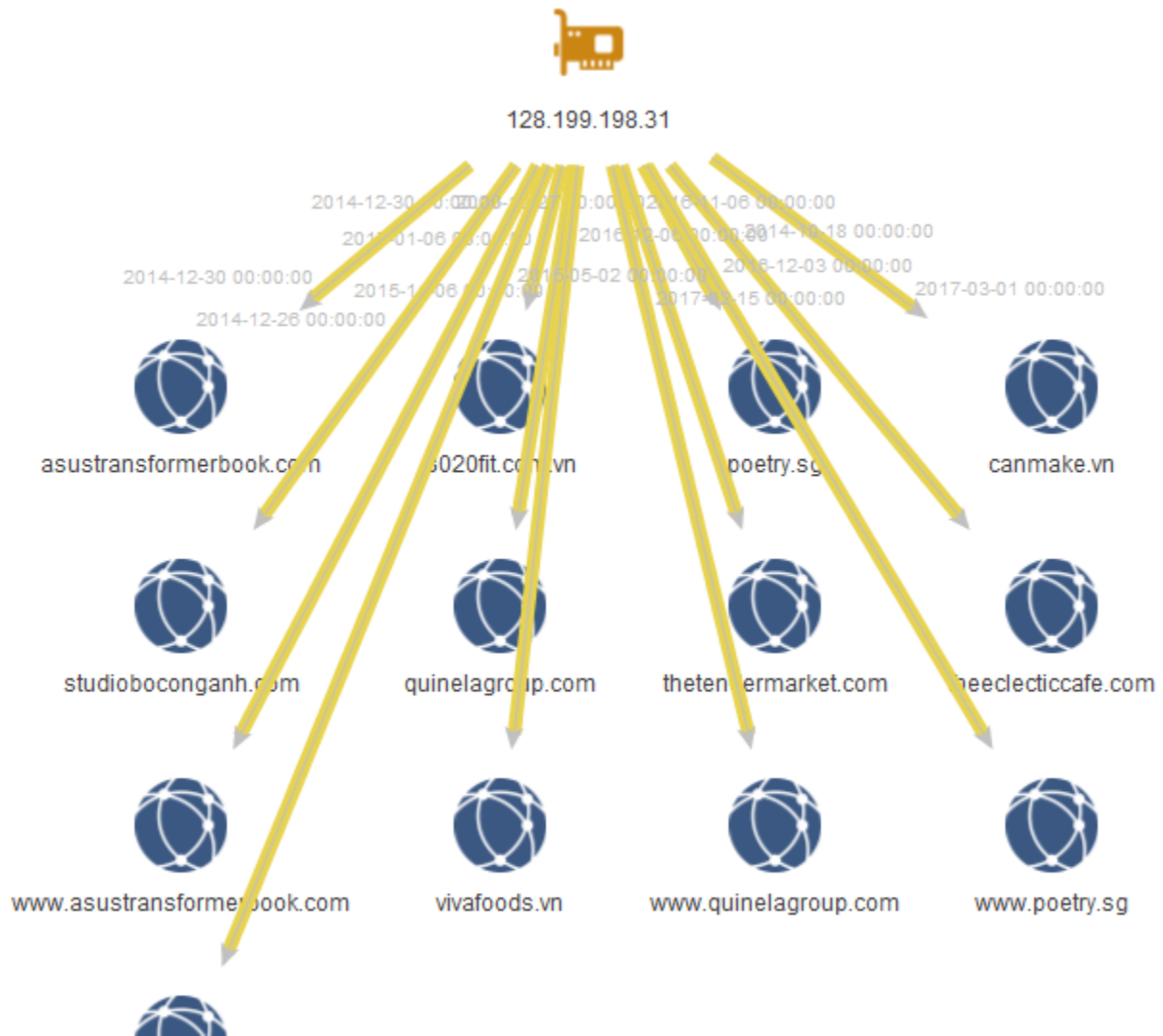
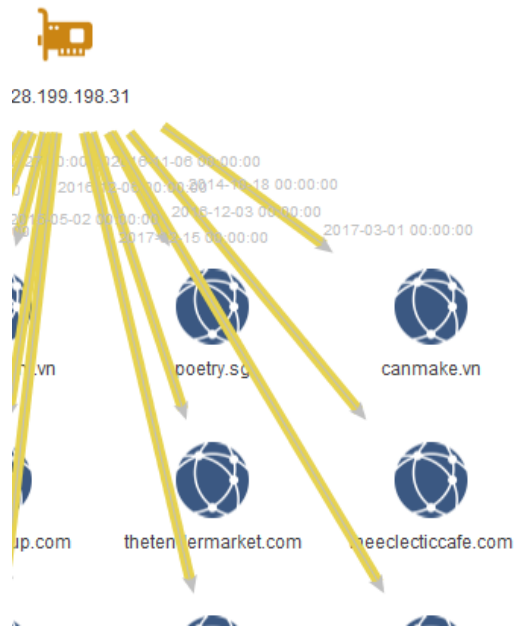


Figure 22 All Links Selected

Using the detail view will show the link data, date or transform ran, the source node, and the target node.





Detail View

Link	Source	Target
2014-10-18 00:00:00	128.199.198.31	theeclecticcafe.com
2014-12-26 00:00:00	128.199.198.31	studioboconganh.com
2014-12-30 00:00:00	128.199.198.31	www.asustransformerbo
2014-12-30 00:00:00	128.199.198.31	asustransformerbook.com
2015-05-02 00:00:00	128.199.198.31	www.studioboconganh.c
2015-11-06 00:00:00	128.199.198.31	quinelagroup.com
2016-11-06 00:00:00	128.199.198.31	thetendermarket.com
2016-12-03 00:00:00	128.199.198.31	www.poetry.sg
2016-12-06 00:00:00	128.199.198.31	vivafoods.vn

Figure 23 Selected links and the Detail View

Unfortunately there isn't a way to select an item in the detail view and have the associated target node deleted from the graph. Using the dates in the Link Field, find the corresponding target node to the graph and delete it.

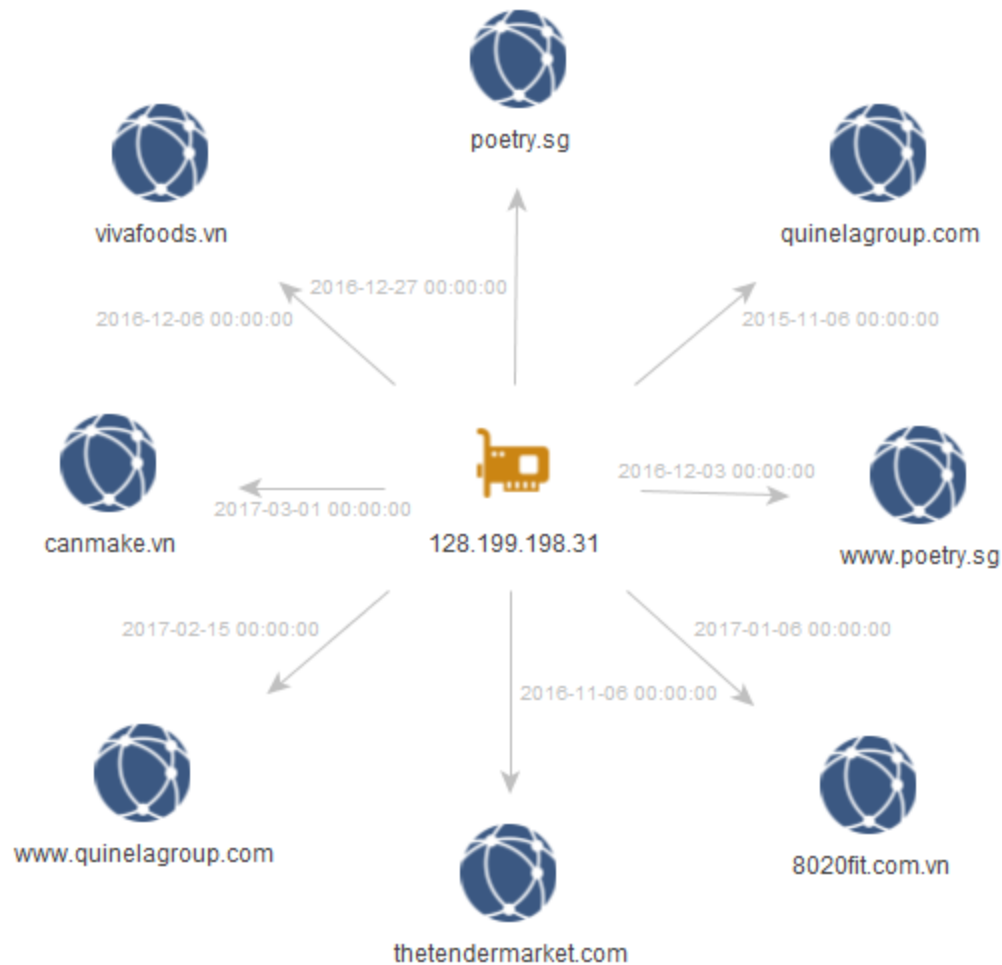


Figure 24 After removing older data

As seen above, there is one domain older than 6 months in the list, the quinelagroup[.]com node, because there is a matching subdomain node from 2017. From this point, it becomes a case of Section 3 Domain Searches.

## 5. WHOIS

Several of the transforms allow for WHOIS data lookup. In some cases these can be made into nodes on the graph, or text data stored in the domain node's properties. Like above that is a decision for the analyst to make. Is the data needed?

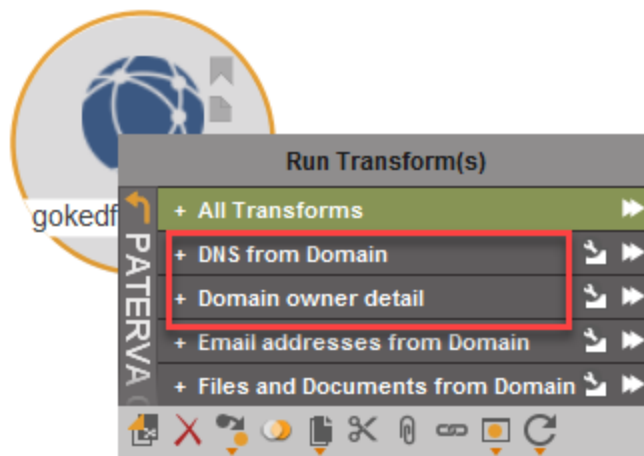
From the blog post, some of the questions to answer are: How long has the domain been registered? Who owns it? What else do they own? Are there any samples from them?

In this case, canmake didn't return data in VirusTotal. That data will be done by hand in the manual link section. For these examples one of the domains that canmake called out to will be used. After creating a node, start at PATERVA's transform list.



*Figure 25 Start with PATERVA CTAS transforms*

The two used here are highlighted below. Domain owner and DNS From domain. Start with Domain Owner to get the email address and phone number. These will be used to search other WHOIS records for the same information to find domains with the same owner.



*Figure 26 Paterva transforms*

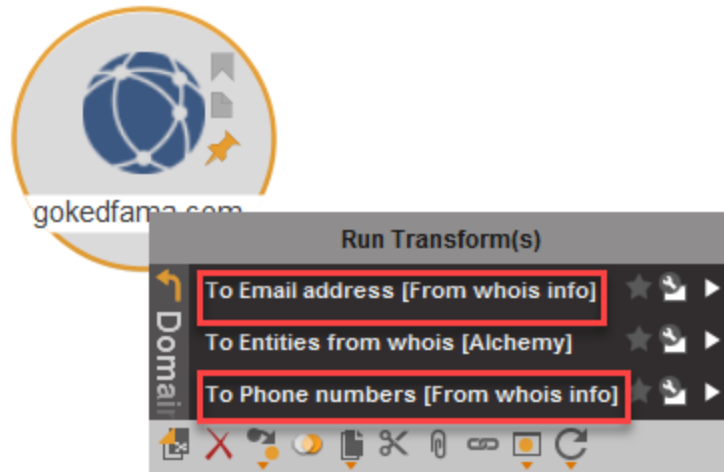


Figure 27 WHOIS email and phone

In this case the phone one didn't work out. Even though the phone number was in the WHOIS record, and is captured in the node details (double click the node), it wasn't picked up by the transform and added to the graph.

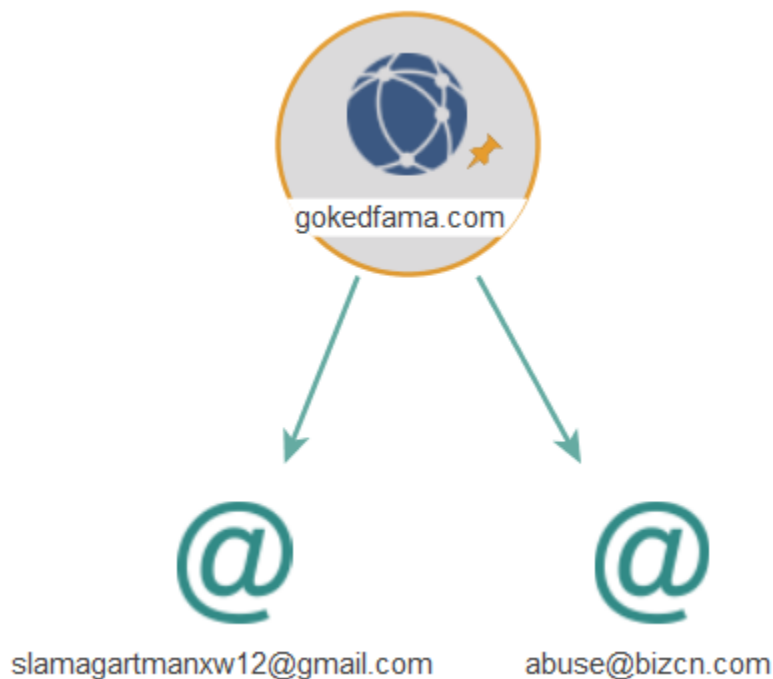


Figure 28 Paterva Transform only saw emails

Double clicking the node opens the detail pane for that node. The Property Tab contains the WHOIS Info. Copying this whole field and pasting in to a text editor showed the whole info, including the phone numbers that were not picked up.

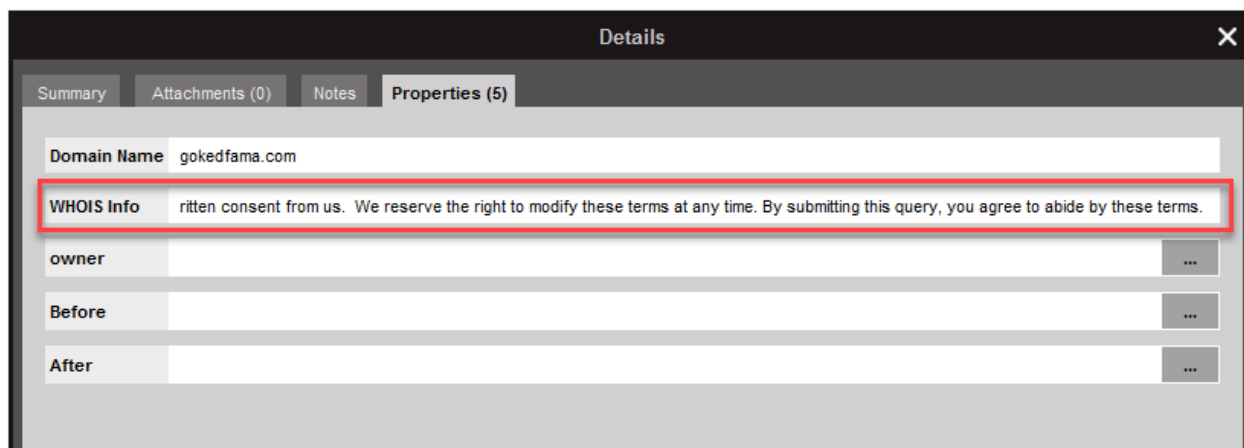


Figure 29 Details > Properties > Copy WHOIS Info

Because only emails were captured with the way the Paterva transform is written, a second transform is can be ran against the domain node. This one is the WHOIS by PassiveTotal.

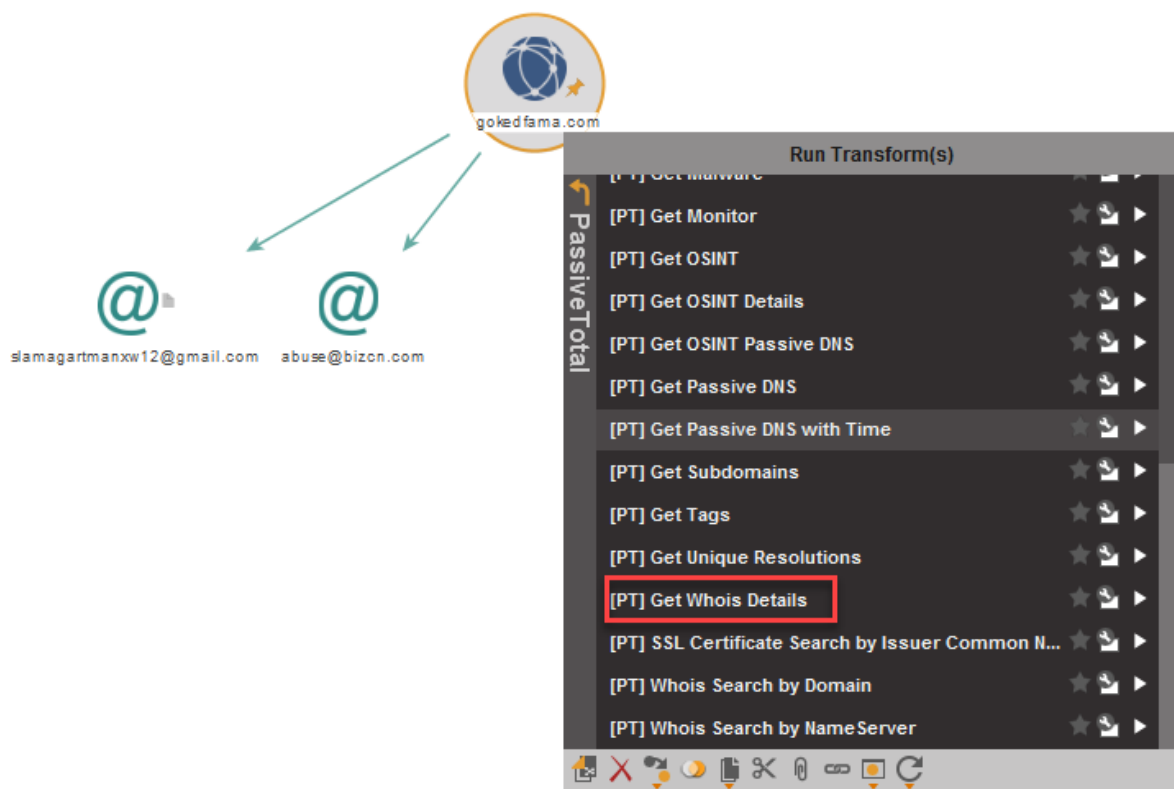


Figure 30 PassiveTotal WHOIS

This provided a second copy of the email address, and other information about the domain. Not all of which is needed for this investigation.

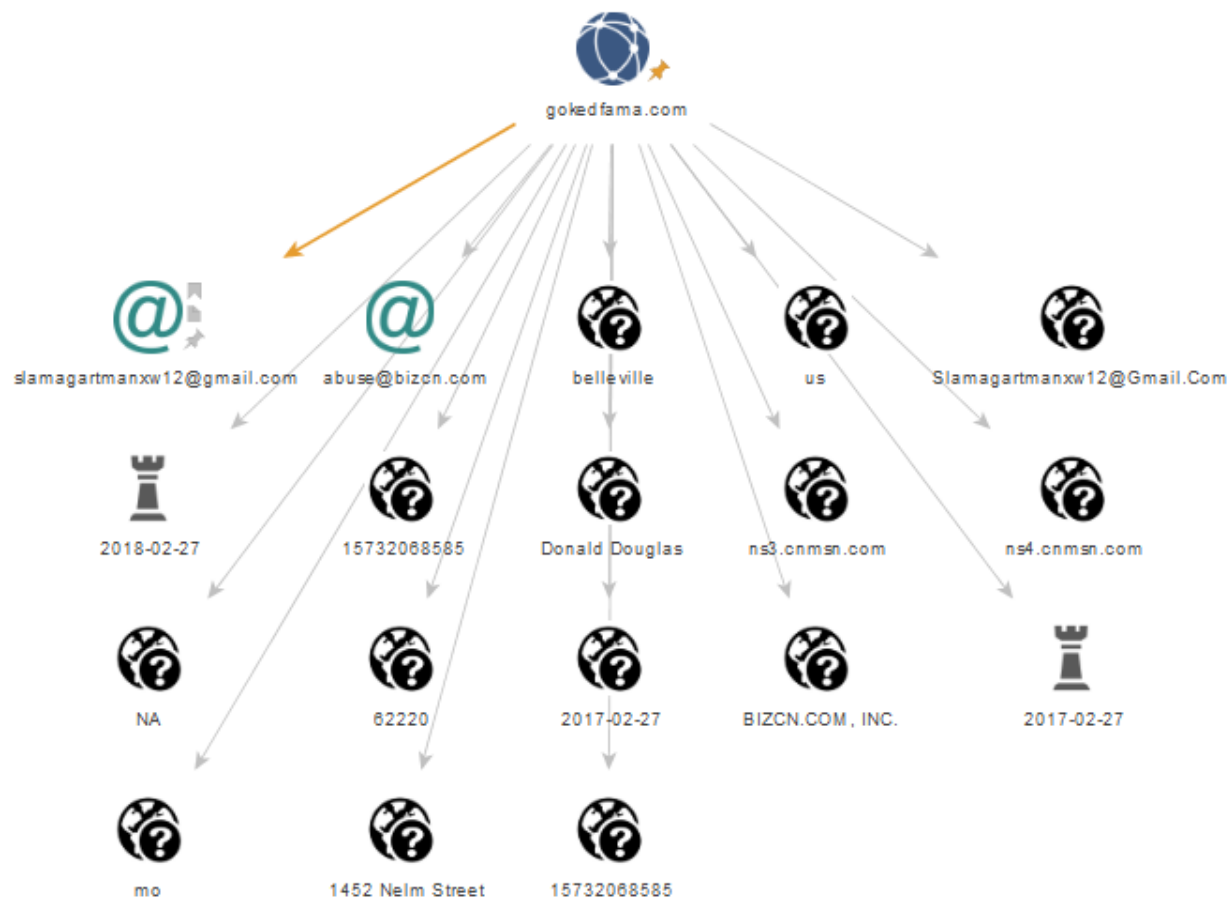


Figure 31 PassiveTotal Whois Results

The nodes not needed for the investigation, were removed, which cleaned the graph up. Then the PassiveTotal WHOIS by email was ran to get a list of all domains associated with that email address.

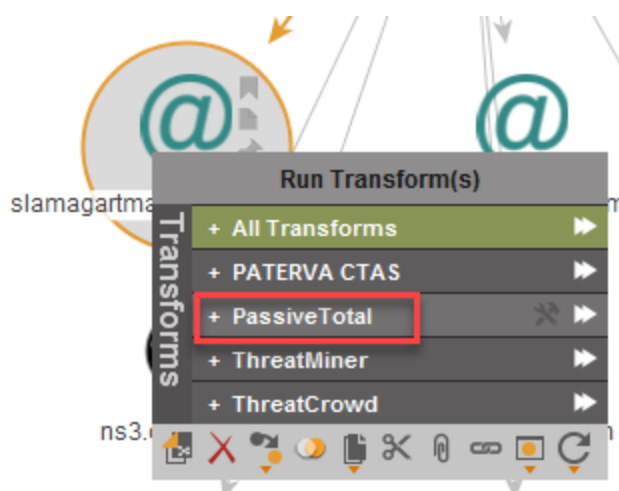


Figure 32 Back to Passive Total

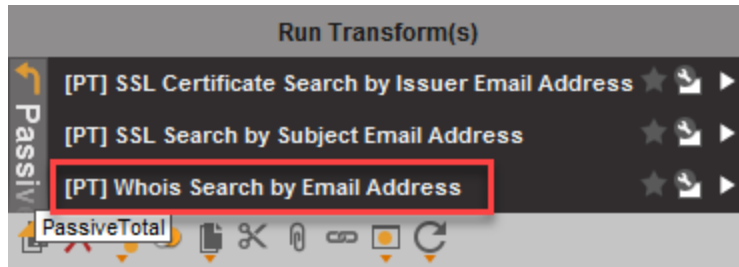


Figure 33 Run WHOIS by email

This lead to 8 additional domains, which needed to be investigated.

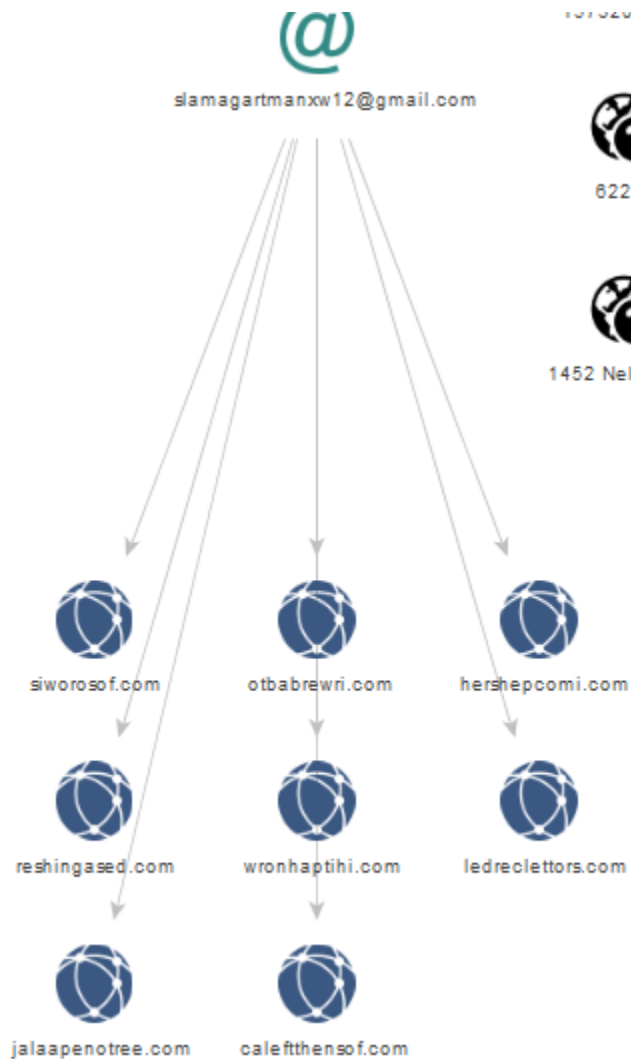


Figure 34 8 more domains to the same actor

These 8 domain nodes had the same WHOIS and sampling steps ran on them which created a graph like the one below. The data in these graphs shows who the owners are, when the domains were registered, the phone number, and the email. Some of the domains don't have IP Addresses yet, but could be blocked (in a black list environment) before being used.

None of the below domains contained any samples at the time of the searches.

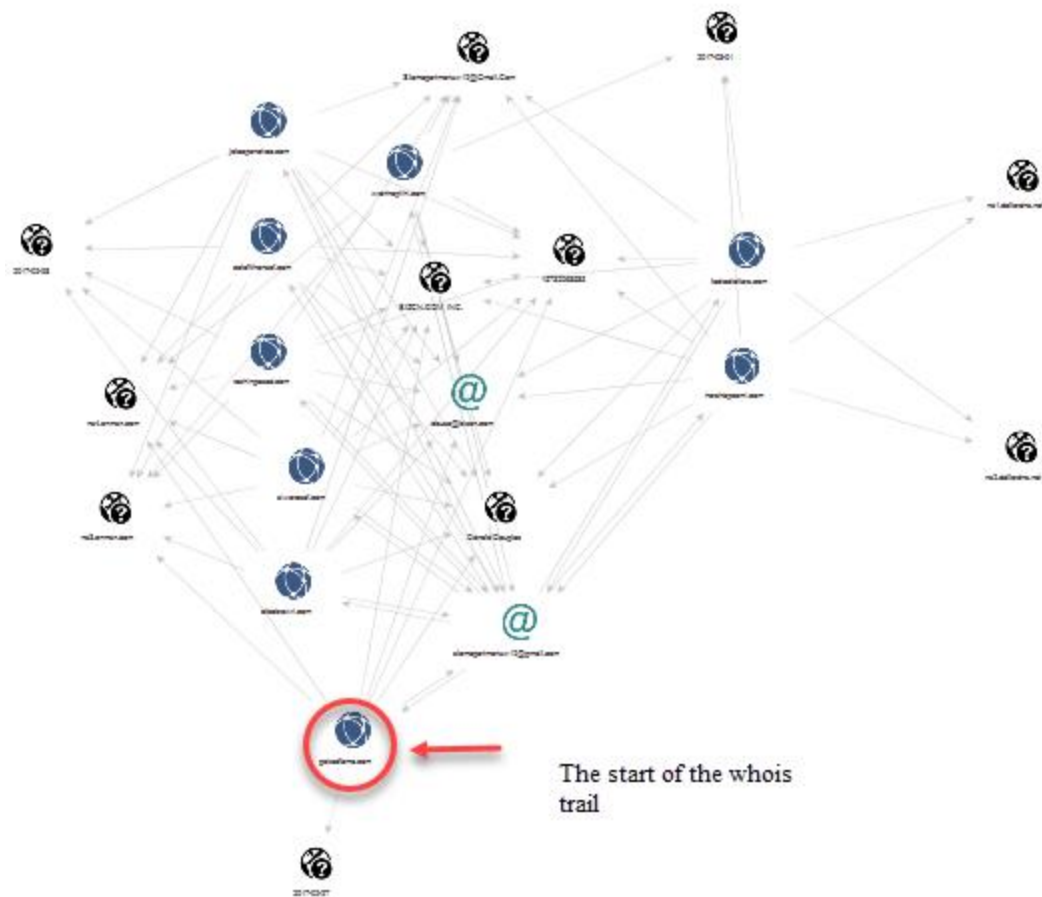


Figure 35 Full WHOIS from one domain.

The next step would be to integrate this graph with the exiting graph, and do some manual linking.

## 6. Manual Linking

Manual linking is used when the transforms doesn't link for you. There are two types of manual links that will be discussed here.

## Linking Nodes

The first one is linking nodes, for when a connection needs to be made and the transforms don't know to make them. This could be creating a link between a malware sample and the domain it calls out to, or connecting a URL from a C2 like location, to the malware dropper location it redirects to

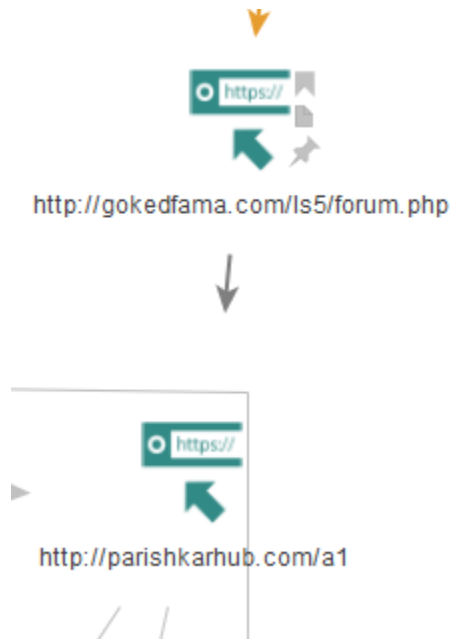




*Figure 36 Malware redirector and Dropper needing a connection*

In this case the original malware document does a check in to the forum, and the response tells the computer where to go download the malware. VT and Maltego did not know enough to actually link the sites. So a manual link is needed.

To manually link nodes, click one of the nodes and drag the mouse to the node that needs to be connected. Multiple nodes can be selected, creating multiple arrows at one time.



*Figure 37 clicking the forum and drag to A1*

Once the arrow is over the node, or one of the selected nodes in a multi-node connection, let go. This will create the link and open the link property box.

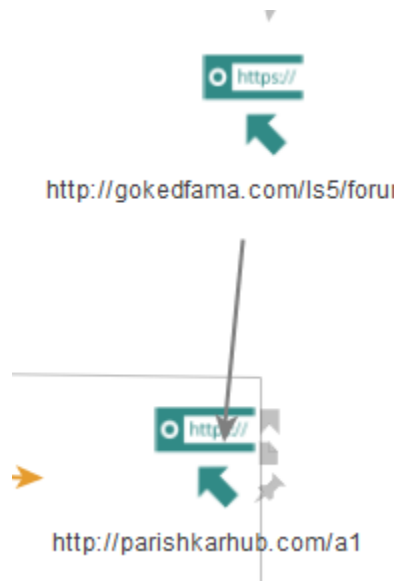


Figure 38 Link creation

The property box will ask what label to use. In this case the label box will be changed to “additional dropper”. Line thickness, style (solid, dashed, other), and color can also be accessed from this page.

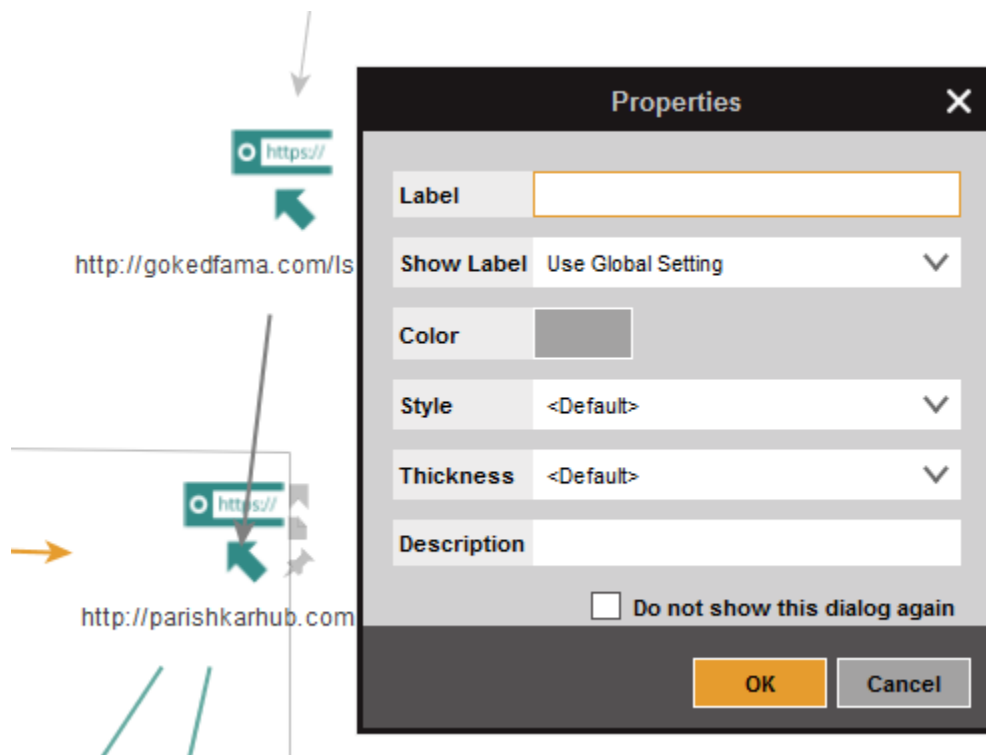
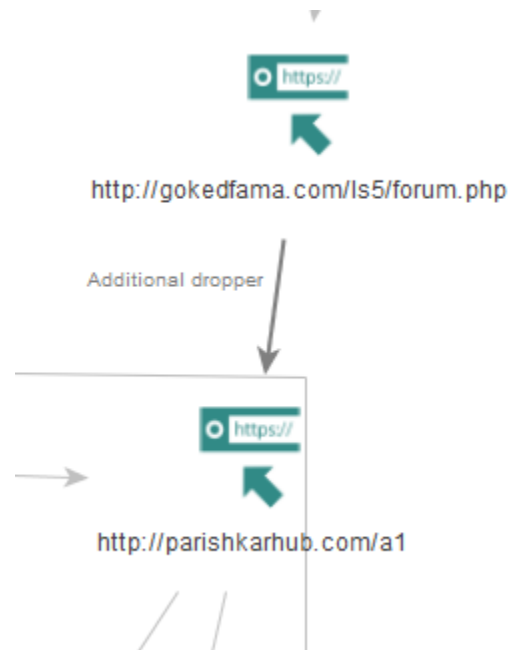


Figure 39 Link Properties

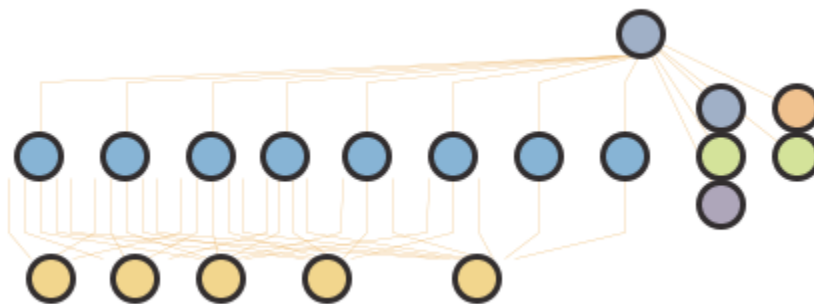
To make changes to a manual link, hold the control key and drag a box over the link with the mouse. This will select all the links that the box is drawn over. Double Clicking allows changes to the link. Delete will delete the links.



*Figure 40 completed manual link*

## Combining Graphs

The second form of linking is combining graphs. In this case, all the WHOIS, resolution, and sample transforms were ran on one node on a graph by itself. The main reason was to make it easier to delete the un-needed nodes from the WHOIS. After the graph was ready to be added, all nodes were highlighted with ctrl-a.



*Figure 41 Select all Nodes with Ctrl-a*

Copy the graph with ctrl-c. Select the tab for the graph to copy too, click where you want the copied graph to go, and press ctrl-v. If the graphs share entities, it will ask which one to keep.

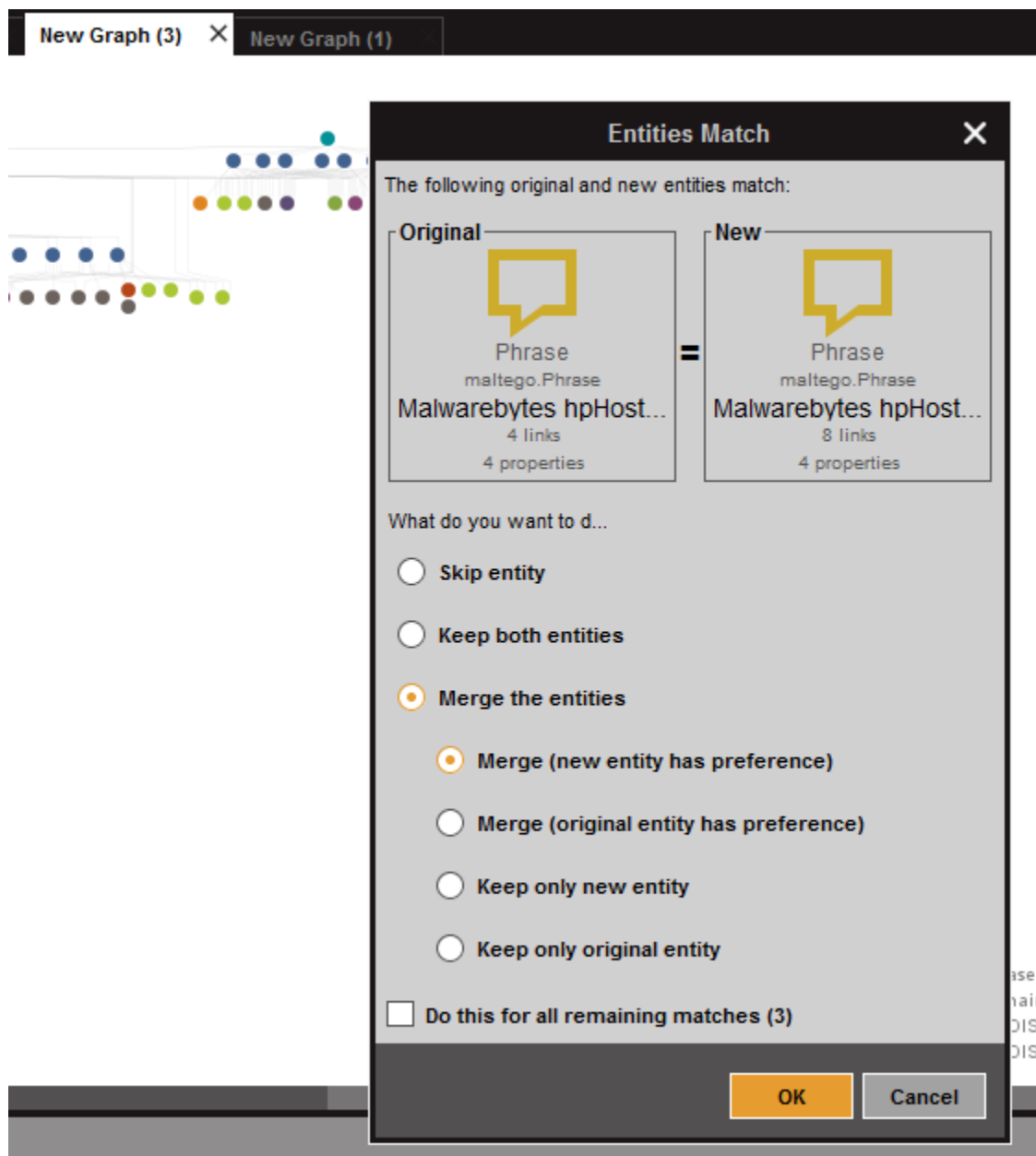
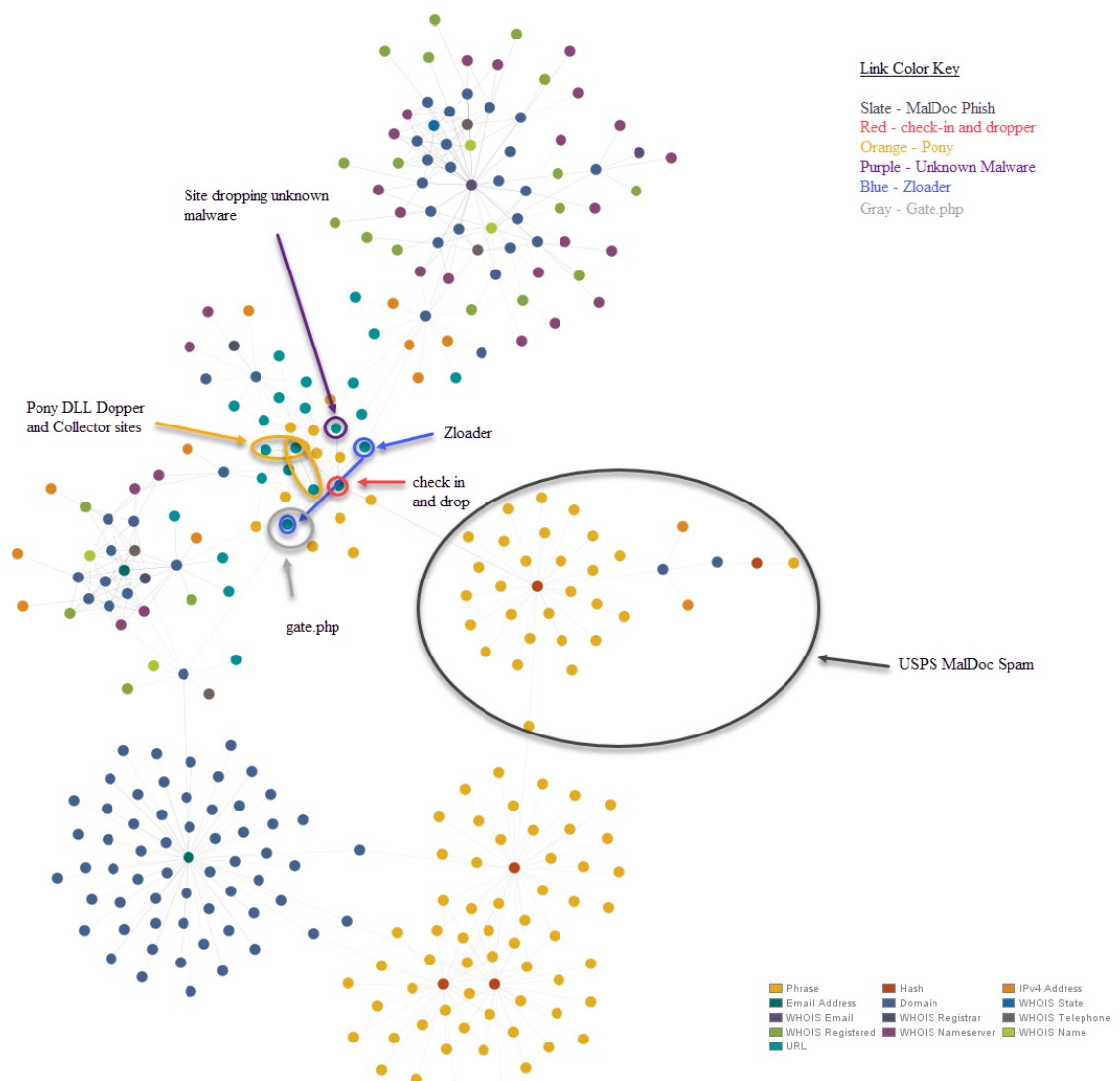


Figure 42 Select which to keep in the merge.

## 7. Example of a finished case

The screenshots for this section was done by using a screen capture tool and editing the files by hand. It is possible to export a zoom-able png file of the graph, but the colored lines won't be in the picture.

Below is the analyzed combined graph of all the domains from the Malware Traffic Analysis blog post and the THL blog posts. The note the color key in the top corner.



The following is a zoomed in section of the secondary droppers and their call outs. The colors from the zoomed out version is re-used in the zoomed in section. The manual links were colored to make them easier to see compared to the standard link color.

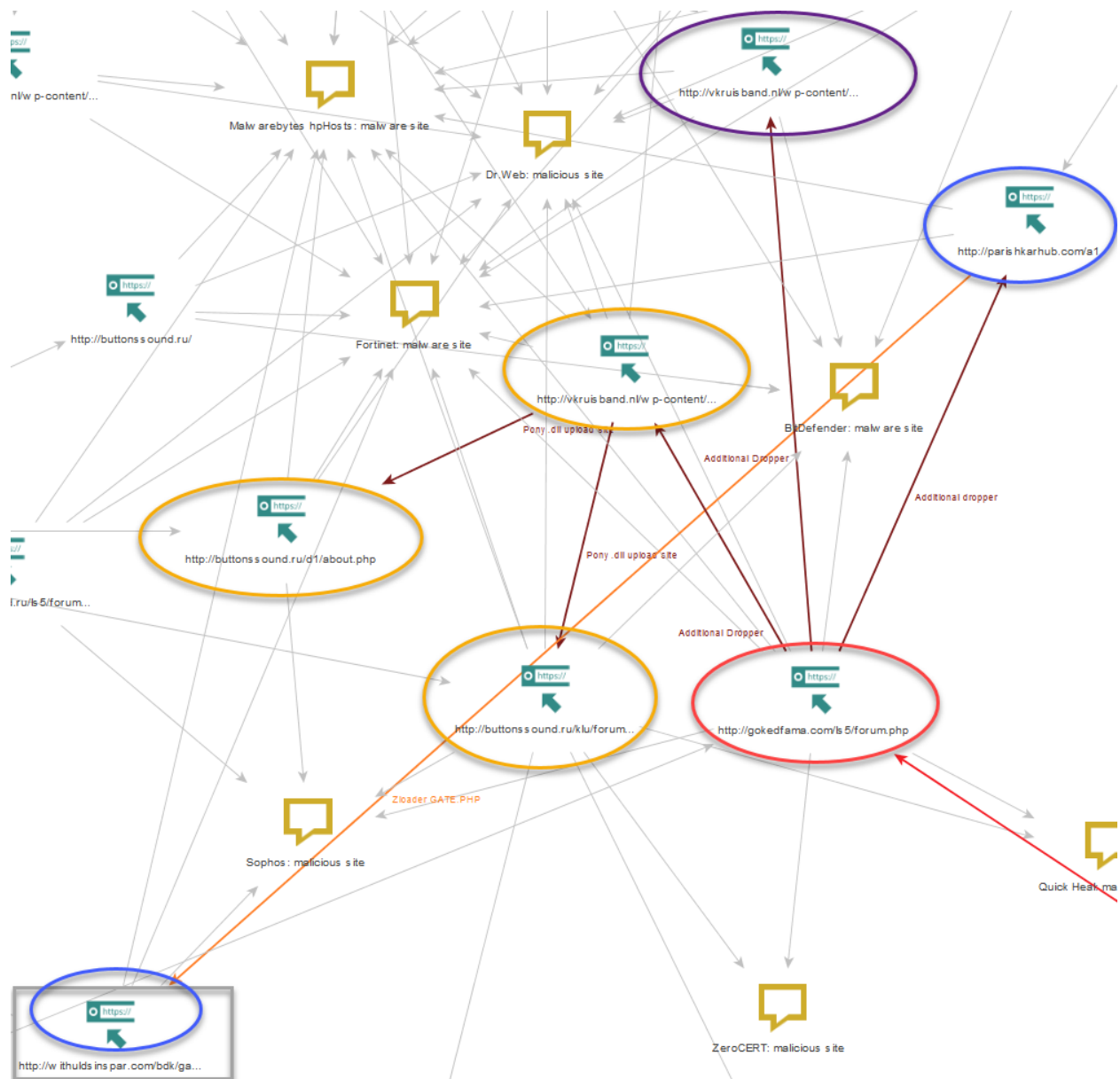


Figure 44 Zoomed in section of graph

## 8. Exporting Data

### CSV

The collected data can be exported to as a CVS file to give to the Incident Response and Threat Hunting teams to block, or hunt for in the network. The CSV file is a two column file that associates different links. For example, Hash, and what the “Phrases” domains, websites, etc that the hash is associated with. The email address, and what it is associated with the email address. Or the website to ip addresses, emails, name servers, etc.

	A	B
118	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	MicroWorld-eScan: Trojan.VBS.Downloader.ABZ
119	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	NANO-Antivirus: Trojan.Ole2.Vbs-heuristic.druvzi
120	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	Sophos: Troj/DocDrop-SV
121	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	Symantec: W97M.Downloader
122	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	TrendMicro-HouseCall: W2KM_DLOADR.YYSYJ
123	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	TrendMicro: W2KM_DLOADR.YYSYJ
124	780a5096b52f18316eeb93aa8a8e17cd7f8372f3dae08094bff39acae457fdc9	http://gokedfama.com/ls5/forum.php
125	amaravathidistrict.com	6/18/2015
126	amaravathidistrict.com	Andhra Pradesh
127	amaravathidistrict.com	Mukunda Rao
128	amaravathidistrict.com	mukundaraod@gmail.com
129	amaravathidistrict.com	ns1.topnameservers.com
130	amaravathidistrict.com	ns2.topnameservers.com

Figure 45 CSV Export example

To generate the CSV file, go to the Import | Export tab above the graph. Select Export Graph to Table, and follow the steps in the wizard.

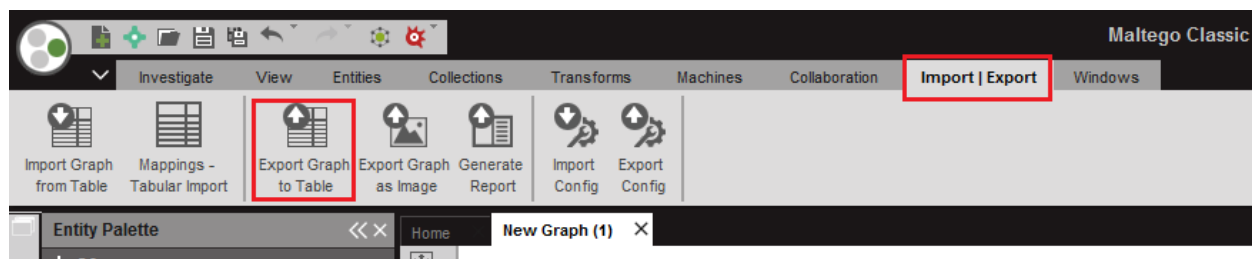


Figure 46 Export to CSV

The CSV file can be used to search the network for additional machines with matching indicators. Blocks can be put in on domains that haven't been used yet by the same actor.

## PNG

It is possible to export the graph as a PNG image. The image is zoom-able, but any custom coloring used on links is lost.

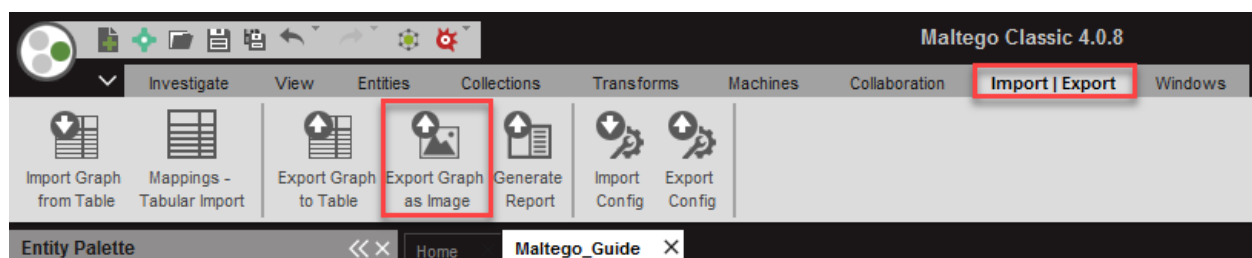


Figure 47 Export Graph to Image

The next image is of the Graph, after it was exported as shown by Photos in Windows 10. This same image scaled slightly was used as the image on the cover of this document. The background is actually clear, and Photos has a back background. It's best to not use this if you don't need too.



*Figure 48 Zoomed all the way out*



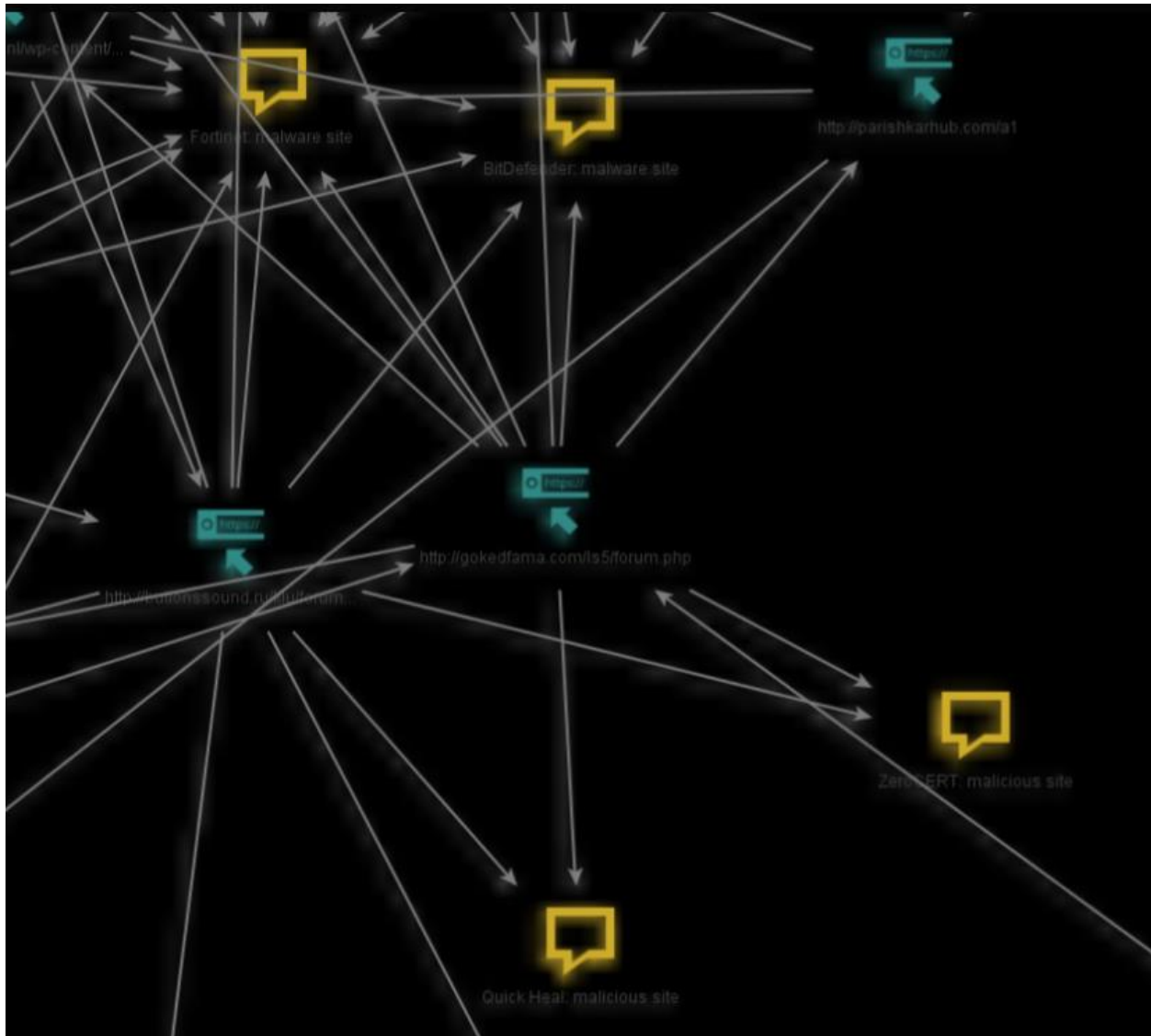


Figure 49 Zoomed to the Second image in section 7

## PDF

The last option is to create a PDF report based on the information in the graph.

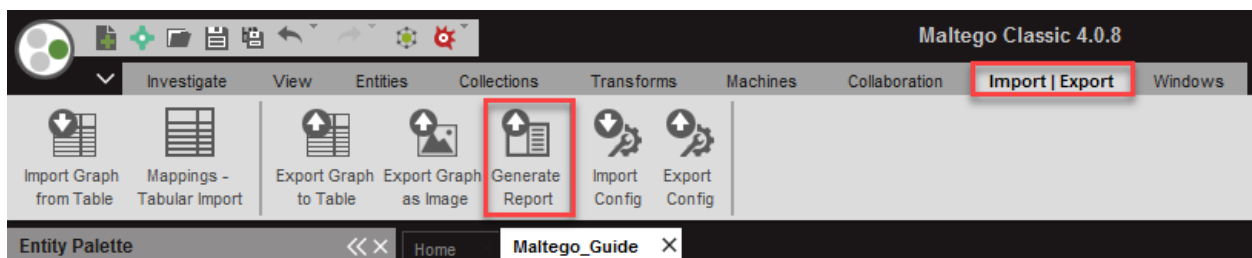


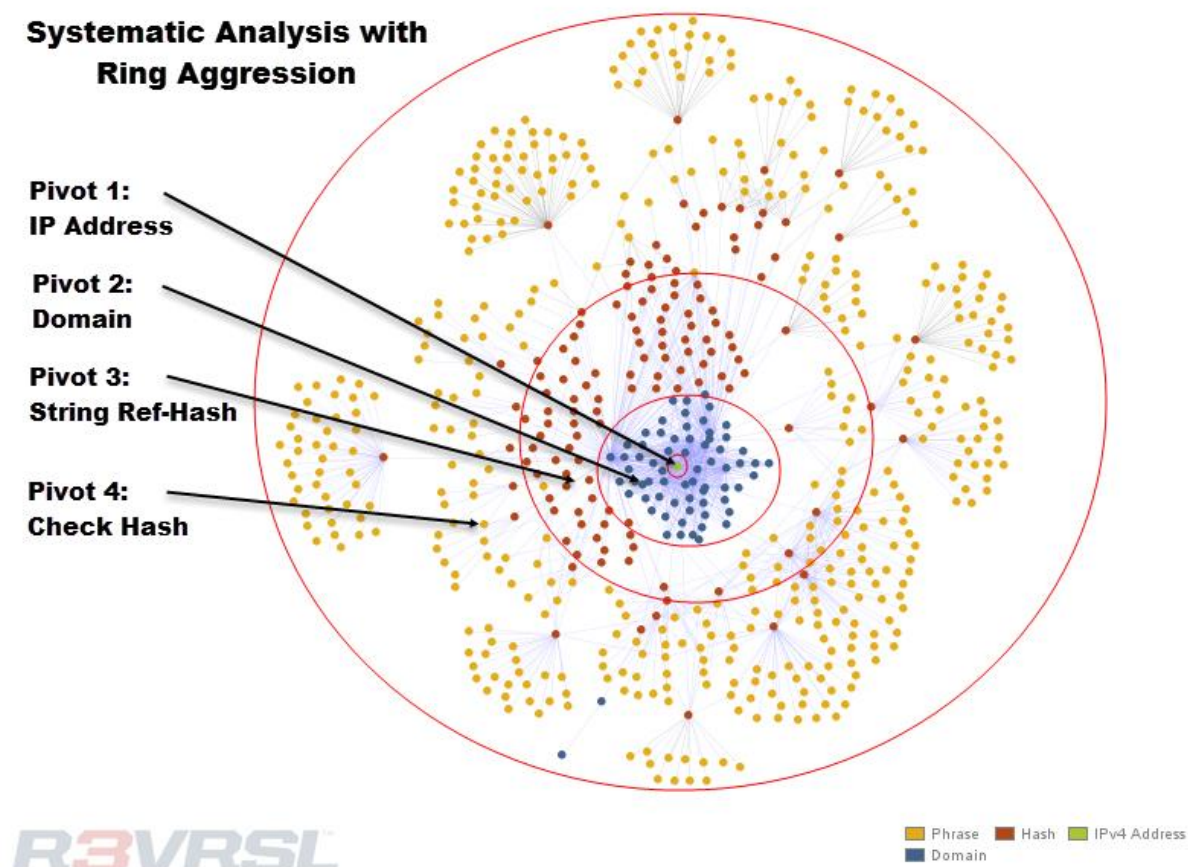
Figure 50 Generate Report

The report created starts with a copy of the Graph, with colors on links, followed by the Top Entities, and a breakdown of each connection on the graph. The final report for this graph was 137 pages of data.

## Appendix 1: Ring Aggression-Systematic Analysis-Pivots

The blog posts by Brad and J didn't fit in to this very well. So it was broken out into its own section.

The data used for the ring aggression technique can come from multiple sources. These sources include open source information, and internal logs. Each transition in this type of analysis is dependent upon the previous pivot and the scoping of the investigation. The analysis will take precedence with what data to include in the next step of the analysis. Each step is its own set of analysis, combined to make a final product.



For example, pivot 1, would be pivoting solely on an atomic indicator such as an IP address will result in the resolution of domains associated with that IP address over a period of time. The historical analysis of an IP pivot will provide information on domains over the required time scale, in this example 1 year. As the IP address belongs to a particular subnet or Autonomous System Number (ASN), it makes sense to block not only the IP, but also the subnet, ASN, peering networks or Country if it will not impact business functions.

The next analysis would be to identify the domains and the number of domains that appear to be or have been hostile. This would be the natural pivot, pivot 2. These domains would be resolved to identify a percentage of domains that are indeed hostile versus benign.

After the domains have been resolved to the IP address, then pivot 3 would be of Domains to hash strings. This will also have a historical attributes for malicious binaries that are/or have resided on the domains.

Once all malicious nodes from pivot 3 are assessed, then a pivot into malicious references. This is pivot 4. This can show variants, timeline, and detection ratios.

This is used, along with the Business' risk acceptance, can be used to create a risk model. The model is used to create the policy and processes on how deep to block, and which indicators to block.

Atomic indicators are notoriously difficult to gain insight and context into and by having a successful, repeatable process one can gather enough data to provide information into the depth and breadth of the indicator and thus unearth actionable intelligence.