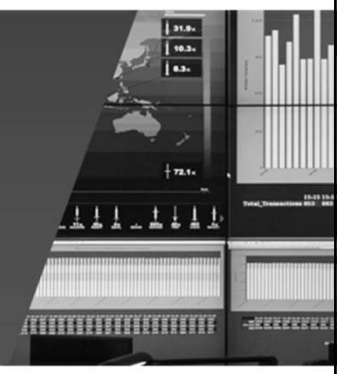
Threat Data Visualization

Steve Ginty
Brandon Dixon

July 18th 2013



Verisign Confidential

VERISIGN
iDefense®

Overview and Agenda

- Identify the problem
- Discuss the solution
- Walk through use targeted attack use case
- Identify caveats
- Conclusion

The Problem

- Attack's and the data associated with them are ever increasing
- Security organizations are now collecting more data than analyst can triage
- Multiple databases with separate UI's to query
  - Passive DNS data
  - WHOIS Data
  - Malware Repository
  - Open Source Intelligence
- How do you sift through terabytes of domains, IP addresses, malware strings, assembly code and network PCAPs to find that one indicator driving the next attack?

The Solution

- Leverage a data visualization product to actively query multiple data sources quickly and connect nodes.
  - Serves as analysis suite
  - Significantly expedites data processing
  - Focus analyst effort on information that matters
  - Many graph-based problems in security
- iDefense has chosen to use Paterva's Maltego software to act as a visualization.
  - Use open source transforms to expedite
  - Develop custom transforms to access internal and external systems
  - Automate repeatable processes through the use of machines

Targeted Attack Case Study

- iDefense identifies a malicious file using the following domain for C&C:
  - itsec.eicp.net
- Dynamic DNS domain resolving to over 800 IP addresses
- Over 500 Second level domain associations
- Manually sifting through all this data would take an extreme amount of analyst effort

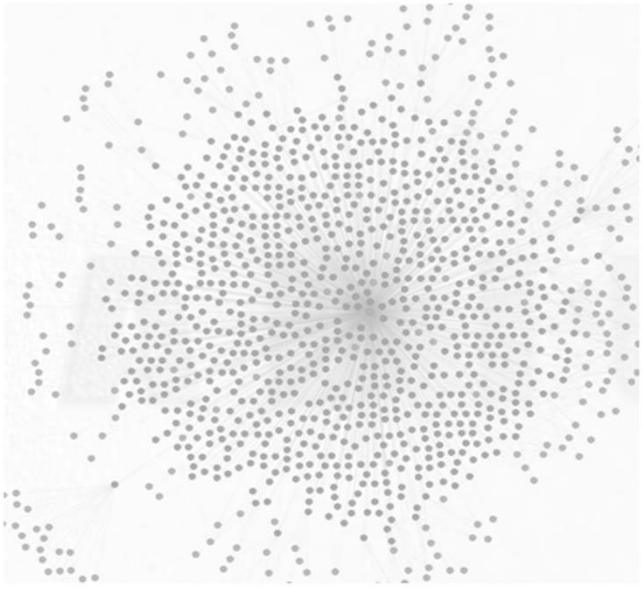
The old way

- Command Line Scripting
- Excel
- Text Files

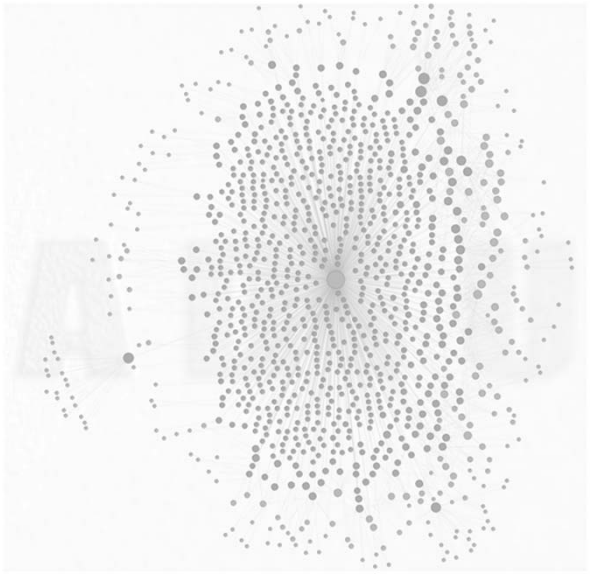
record	rrtype	first	last	count
0.0.0.0	A	2011-12-31 20:33:53	2013-04-16 18:40:09	50960
1.203.0.145	A	2012-05-26 18:16:35	2012-05-26 19:25:19	43
1.203.0.166	A	2013-01-26 06:47:36	2013-01-26 19:16:43	663
1.203.1.2	A	2012-05-12 18:31:07	2012-05-12 23:09:04	188
1.203.1.31	A	2012-05-16 07:31:01	2012-05-16 10:13:18	51
1.203.1.74	A	2012-06-11 21:14:16	2012-06-12 09:43:38	96
1.203.2.67	A	2012-05-25 17:19:17	2012-05-26 08:49:18	111
1.203.2.104	A	2012-05-23 21:21:46	2012-05-23 23:19:19	4
1.203.2.146	A	2012-06-12 21:20:05	2012-06-12 22:20:05	2
1.203.2.180	A	2012-06-08 10:19:11	2012-06-08 17:19:10	276
1.203.3.50	A	2012-06-13 04:50:04	2012-06-13 04:50:04	1
1.203.3.53	A	2012-03-21 08:48:49	2012-03-21 12:16:29	38
1.203.3.152	A	2013-01-18 04:22:16	2013-01-19 23:05:09	3344
1.203.7.121	A	2012-03-29 21:30:09	2012-03-29 22:42:41	23
1.203.7.234	A	2012-10-21 00:24:44	2012-10-21 08:00:55	15
1.203.9.204	A	2012-04-09 21:32:49	2012-04-09 23:05:03	33
1.203.10.5	A	2012-05-24 21:19:20	2012-05-24 23:19:17	33
1.203.10.99	A	2012-09-24 20:03:33	2012-09-25 01:40:07	10
1.203.10.220	A	2012-09-08 07:15:40	2012-09-09 16:35:19	1141
1.203.11.44	A	2012-07-30 04:00:01	2012-07-30 16:40:01	14

Data Visualization – First Level Connections

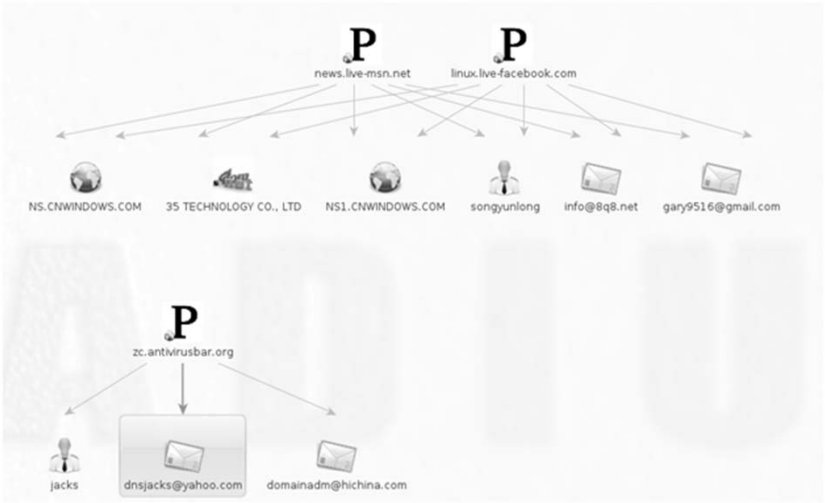
Data Visualization – Second Level Connections



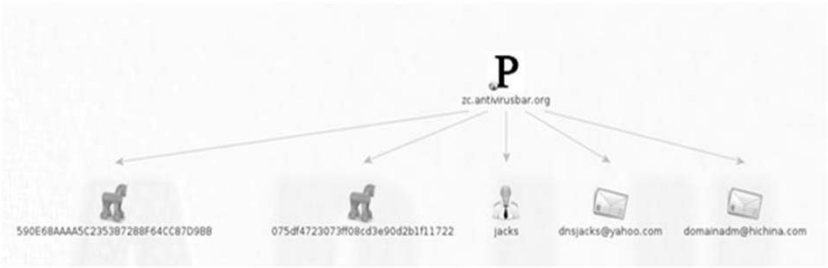
Data Visualization – Weighted Graph



New Indicators of Interest Identified



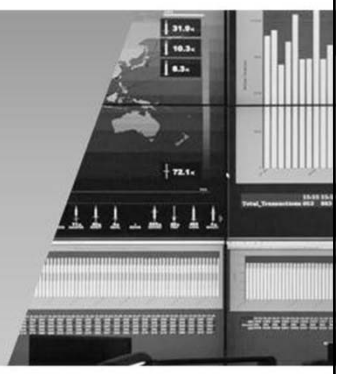
New Malware Samples Identified



Problems to Overcome

- Not a 100% solution
  - But significantly better than before
- Platform Limitations
  - Smaller user base
  - Collaborative research is an issue
  - There is no undo button – this is a bitch
  - Slow update process to client
  - Lacks ability to pass-the-graph
- The representation of time in Passive DNS is a hurdle

Live Demo



VERISIGN

Need for More

- Not enough companies doing visualization of security data
- Platforms need to easily integrate into existing databases and solutions
- Analysts need to have a way to save the graph connections outside of a platform
- Data should be queried in a graph-based fashion to obtain connections without visuals
- Time and other factors need to be accounted for when plotting data

Conclusions

- Visualization is not perfect, but speeds up analyst workflow if done properly
- Visual platforms are able to take multiple databases filled with potentially overlapping data and show connections
- Analyst is still and will always be required to validate decisions and identify the most useful information