



# Hunting New Malware Actors

Charles Nicholas  
nicholas@umbc.edu

Department of Computer Science and Electrical Engineering  
August 2016

This presentation is unclassified

# Summary

- We've spent the last several years working in the area of malware analysis.
- We noticed that some easy questions don't have easy answers:
  - Who wrote this malware?
  - What does it do?
  - Have we seen anything like it before?
- To address these questions, we need to appeal to *two* buzzwords: **cyber**, and **data science**

# Malware Analysis in the Large

- Arriving in a local malware analysis lab, I asked:
  - How much the malware collection is aimed at Windows/x86 vs. other targets?
  - Does the malware data form clusters?
  - What trends are being seen?
- Easy questions, but no easy answers!

# Comparing Malware Specimens

- If we have a suspected malware specimen, we want to know if anything like it has been seen before...easy question, no easy answer.
  - Document similarity (a concept from Information Retrieval) metrics do exist, but need to be adapted to malware
  - These static analysis methods can be frustrated by packing and encryption
  - For now, we restrict our discussion to static vs. dynamic malware analysis.

# Visualization!

- What if somebody was to “discover” a new play by Shakespeare?
- The new document might well stand out from the rest in a plot like [this](#)!  
(Created with a Python script, the anaconda and plotly packages, and character 4-grams. )

# Do Executable Binaries Form Clusters?

- Some large malware collections are available to malware researchers (which we should continue to encourage)
- VX Heaven is old, but it is labelled, and at 60+ GBs uncompressed, relatively big.
- Do Banking Trojans form a cluster? Or rootkits?
- Do the labels match the clusters? Let's [see](#)!

# Points to Ponder

- Such graphs are useful for spotting outliers, and maybe near-duplicates
- Outliers may or may not be interesting
  - It's okay to remove them, which may reveal more structure
- Not effective for in-depth analysis (except inasmuch as it helps direct your efforts)
- May be useful for spotting malware families

Then we were asked to focus on a particular type of malware!

# Beware of Exploit Kits!

- User is tricked into visiting an infected (but innocent) web site
- As a result of a few iframe redirects, such as `<iframe src=http://badGuys.Rus.ru>` the user's browser is sent to an EK web site
- A Javascript "landing page" is loaded,
- Which looks at user's box and tries some exploits
- If any succeed, word is sent back to EK operators, some of whom have grown wealthy



# Famous Exploit Kits

- Black Hole is still the most famous, but there are many others: Angler, Sweet Orange, Redkit, Nuclear, RIG, Fiesta, Magnitude, etc.
- Zscaler Blog [reported](#) on June 24 that Angler activity is decreasing, but others remain active.
- Note that just black-listing web sites is not sufficient, as they change all the time
- How can we tell when new EKs come out?

# Our Hypotheses


- If the Javascript code on landing pages corresponds to specific exploit kits, then similar scripts would be attributable to the same (or related versions of the same) exploit kit.
- Exploits may be updated from time to time, infrastructure maybe less often
- Scripts unlike any seen before may indicate a new, unknown exploit kit.
- This *landing page analysis* may let us better understand this form of malware.

# Concept of Operations

- *Seek out* malware domains (live external links)
  - [urlquery.net](http://urlquery.net) was useful to start
  - Or [malware-traffic-analysis.net](http://malware-traffic-analysis.net)
- Visit infected sites from a browser inside a VM
  - EK would often be cleaned up by time we visit ☹
- Capture and re-assemble packets
  - Virtualbox packet capture, Suricata
  - Analyze raw pcap data, or tcpick output, or extracted Javascript

## MALWARE-TRAFFIC-ANALYSIS.NET




 [RSS feed](#)

[About this blog](#)

[@malware\\_traffic on Twitter](#)

Google™ Custom Search



- **2014-10-10** -- Out for the next two weeks or so
- **2014-10-09** -- Magnitude EK from 178.32.82.137 - 3d9a766.0eec.bcf7e8.af992.1e705.5f8d3.f658a017o9.pressterminals.in
- **2014-10-08** -- Phishing email - Subject: Fw:Order Inquiry
- **2014-10-07** -- Phishing campaign - Subject: You have a voice message
- **2014-10-06** -- Sweet Orange EK from 8.28.175.75 - ray.whydoesmyeyetwitch.net:15106 & asquality.bastionwright.com:15106
- **2014-10-06** -- Rotator generates Angler EK on 5.135.230.183 - 7dws8yz0k2.sdiouvb.com
- **2014-10-05** -- Rig EK from 37.200.69.87 - contact.collegemotorsltd.com
- **2014-10-04** -- Rig EK and Upatre from phishing emails
- **2014-10-03** -- Phishing campaign - Incoming fax reports - fake HMRC tax notices
- **2014-10-03** -- Sweet Orange EK from 8.28.175.74 - b.epavers.com:17767 & k.epavers.com:17767
- **2014-10-02** -- Phishing email - Subject: Job in financial service
- **2014-10-02** -- Angler EK from 66.172.27.117 - asd.bingevomitsyndromesexy.net
- **2014-10-01** -- 32x32 gate leading to Angler EK on 66.172.27.117 - asd.crossheading.us
- **2014-10-01** -- Malware from fake IRS notification causes "CryptoWall 2.0" infection
- **2014-09-30** -- Phishing email - Subject: Requirement.
- **2014-09-30** -- Fiesta EK from 64.202.116.153 - affineairforce.us
- **2014-09-29** -- Nuclear EK delivers digitally-signed CryptoWall malware
- **2014-09-28** -- Styx EK from 162.244.33.39 - poolie.vvk49.com
- **2014-09-27** -- 32x32 gate to Angler EK on 66.172.12.231 - asd.branchiopodamericangentian.us
- **2014-09-26** -- Phishing campaign - Subject: Transaction not complete
- **2014-09-26** -- 32x32 gate to Angler EK on 162.248.243.78 - qwe.tributarykamarupan.us
- **2014-09-25** -- Sweet Orange EK from 8.28.175.67 - cdn.americasrapper.com:10016 & cdn5.blumaxmaterial.com:10016
- **2014-09-24** -- Fiesta EK from 104.28.6.73 - eoxsc.kulawyn.in
- **2014-09-24** -- Phishing campaign - Subject: Overdue Payment: 884272725375713
- **2014-09-23** -- Rig EK from 178.132.203.26 - mdif.boroughventuremenswear.com
- **2014-09-22** -- Phishing email - Subject: NatWest Statement
- **2014-09-22** -- Angler EK from 192.99.197.134 - asd.singularitymusculusintercostalis.us

## MALWARE-TRAFFIC-ANALYSIS.NET



### 2014-10-09 - MAGNITUDE EK FROM 178.32.82.137 - 3D9A766.0EEC.BCF7E8.AF992.1E705.5F8D3.F658A0I7O9.PRESSTERMINALS.IN

#### PCAP AND MALWARE:

- PCAP of the VM infection traffic: **2014-10-09-Magnitude-EK-traffic.pcap**
- ZIP file of the malware: **2014-10-09-Magnitude-EK-malware.zip**
- Malwr.com PCAP: **2014-10-09-Mangitude-EK-payload-1-of-6-malwr.com-analysis.pcap**
- Malwr.com PCAP: **2014-10-09-Mangitude-EK-payload-4-of-6-malwr.com-analysis.pcap**
- Malwr.com PCAP: **2014-10-09-Mangitude-EK-payload-5-of-6-malwr.com-analysis.pcap**
- Malwr.com PCAP: **2014-10-09-Mangitude-EK-payload-6-of-6-malwr.com-analysis.pcap**
- Malwr.com PCAP: **UpdateFlashPlayer\_811e7dfc.exe-malwr.com-analysis.pcap**

#### NOTES:

- Found this while checking through **Scumware.org** for interesting entries.
- It took me a few tries to get a full chain of infection traffic.

### CHAIN OF EVENTS

#### ASSOCIATED DOMAINS:

- 62.233.121.40 - **www.nottinghamshire-probation.org.uk** - Compromised website
- 37.9.53.90 - **stats.street-jeni.us** - Redirect
- 178.32.82.137 - **3d9a766.0eec.bcf7e8.af992.1e705.5f8d3.f658a0I7o9.pressterminals.in** - Magnitude EK

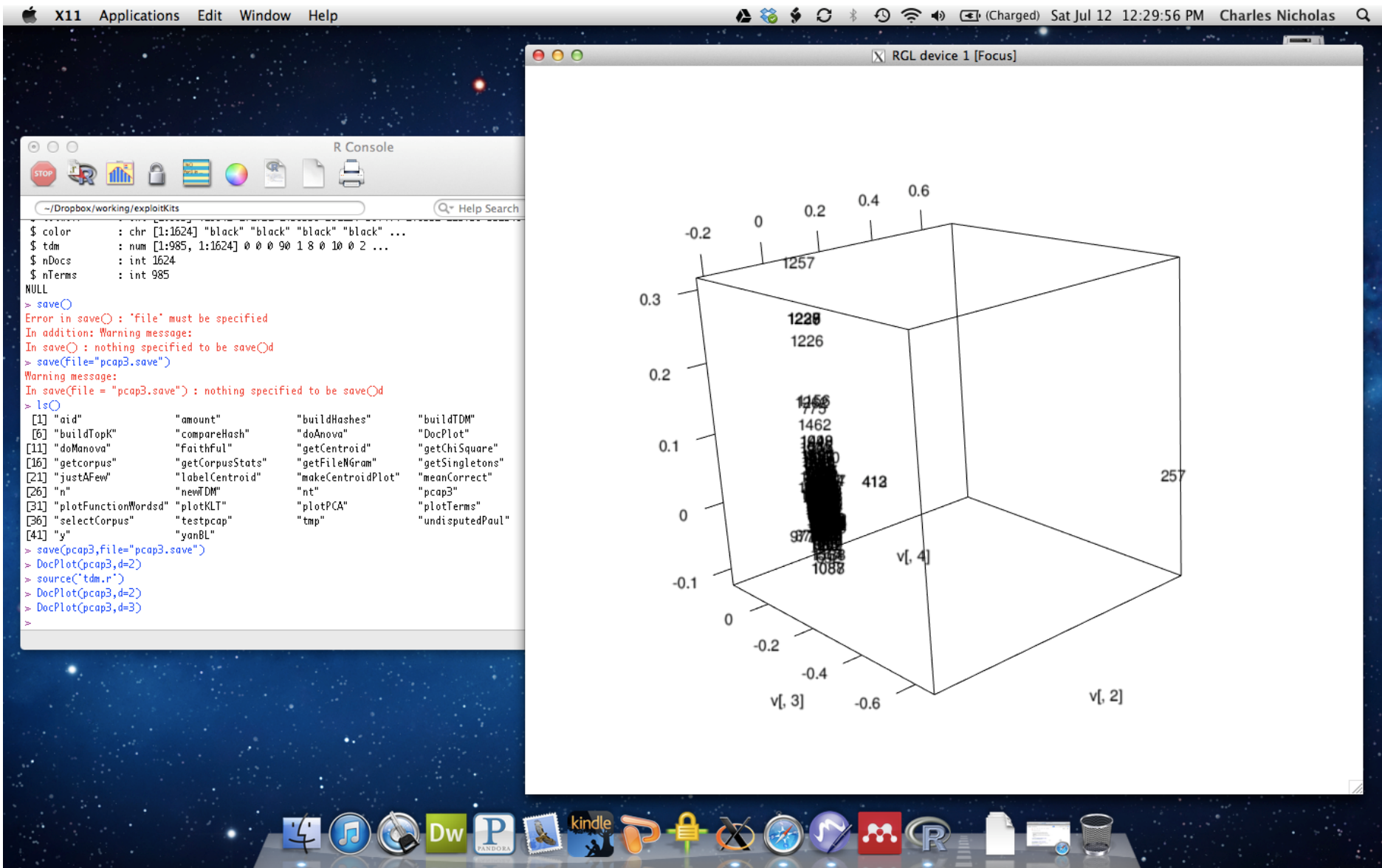
#### COMPROMISED WEBSITE AND REDIRECT CHAIN:

- 14:10:19 UTC - 192.168.204.145:49383 - 62.233.121.40:80 - **www.nottinghamshire-probation.org.uk** - GET /
- 14:10:10 UTC - 192.168.204.145:? - 62.233.121.40:80 - **www.nottinghamshire-probation.org.uk** - *[one of the .js files from the site]*
- 14:10:28 UTC - 192.168.204.145:49423 - 37.9.53.90:80 - **stats.street-jeni.us** - GET /show.php

#### MAGNITUDE EK:

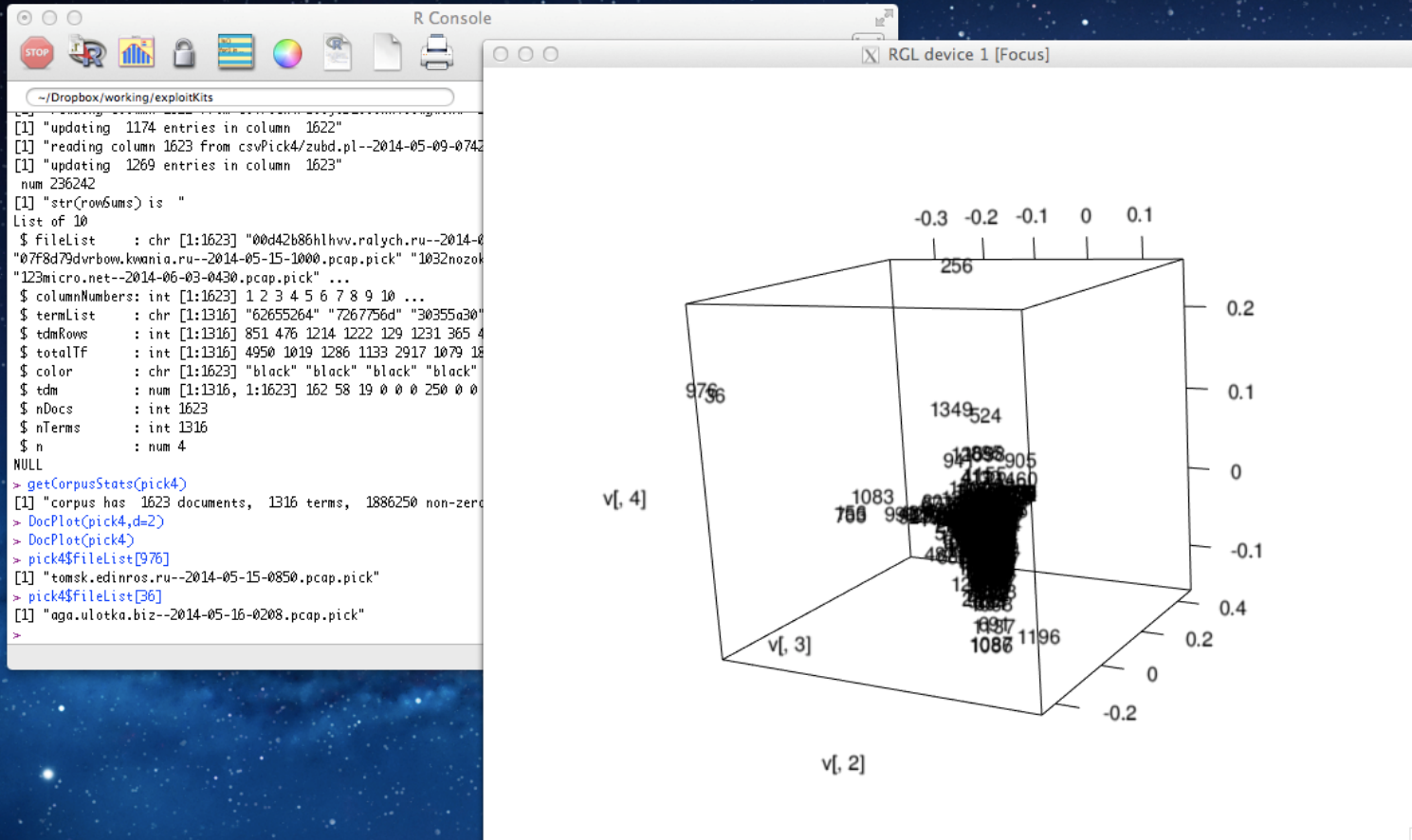
- 14:10:29 - 3d9a766.0eec.bcf7e8.af992.1e705.5f8d3.f658a0I7o9.pressterminals.in - GET /
- 14:10:34 - 3d9a766.0eec.bcf7e8.af992.1e705.5f8d3.f658a0I7o9.pressterminals.in - GET /a4adeaf2e3a08a34feeda27ad005fa91/5a85bcce264f195316838068dbb2b852
- 14:10:34 - 3d9a766.0eec.bcf7e8.af992.1e705.5f8d3.f658a0I7o9.pressterminals.in - GET /a4adeaf2e3a08a34feeda27ad005fa91/e8f096b5a0909917835416cb5c4c780f

# Analysis of Raw PCAP





# Analysis of tcpick Output

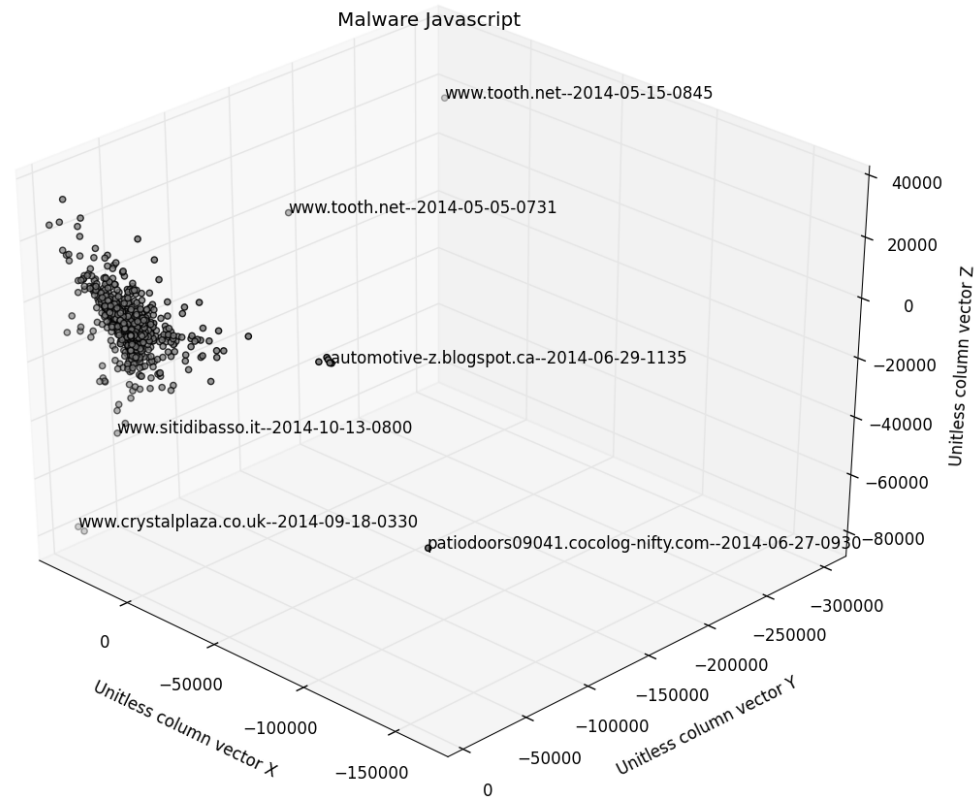


# Discussion

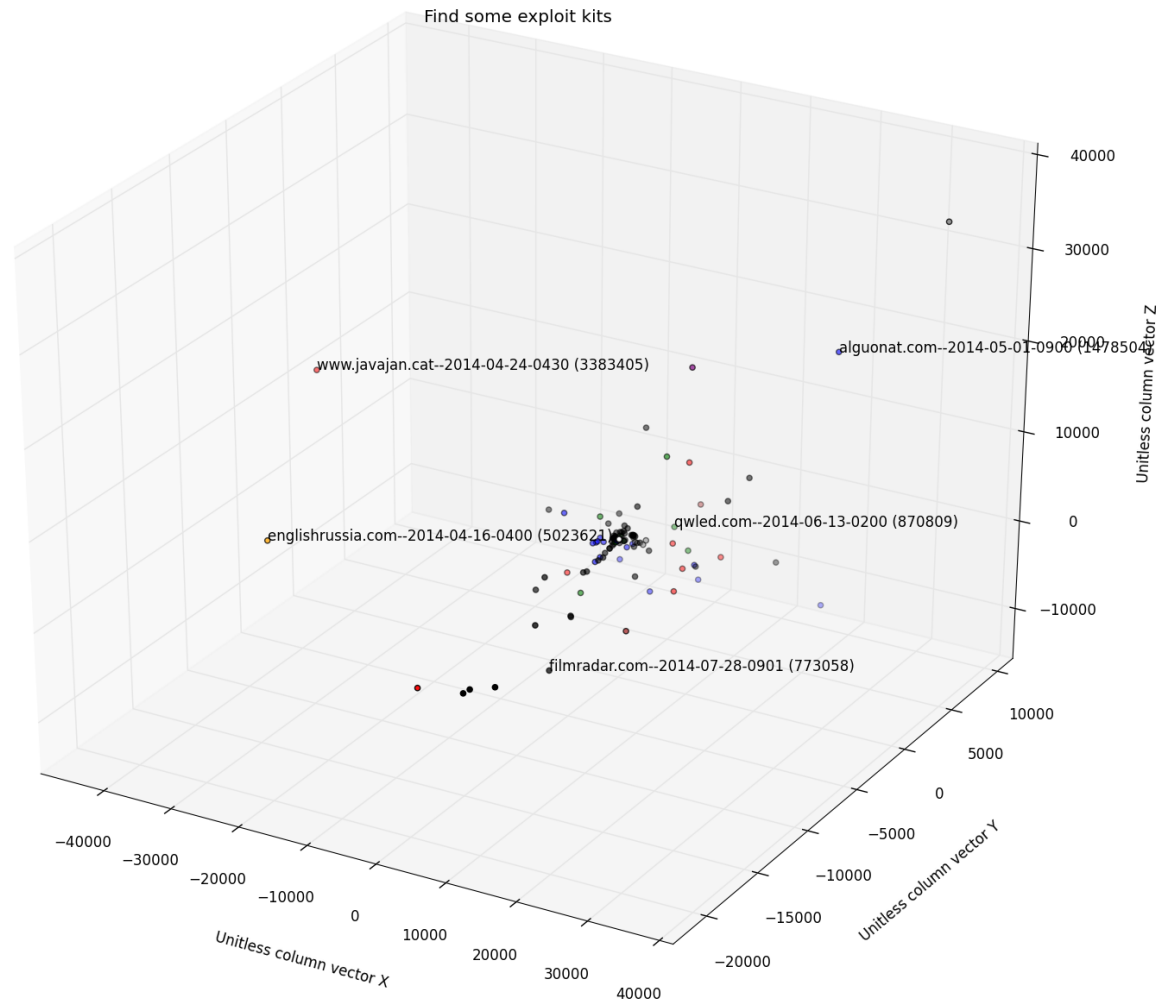
- Analysis of raw pcap data might be used to spot outliers, or near duplicates.
- Analysis of tcpick output with 4-grams shows two almost duplicates, among other phenomena
- Suricata extracts specific files, including HTML with embedded Javascript, from specific sites and timestamps, as shown



# Plotting of Packets

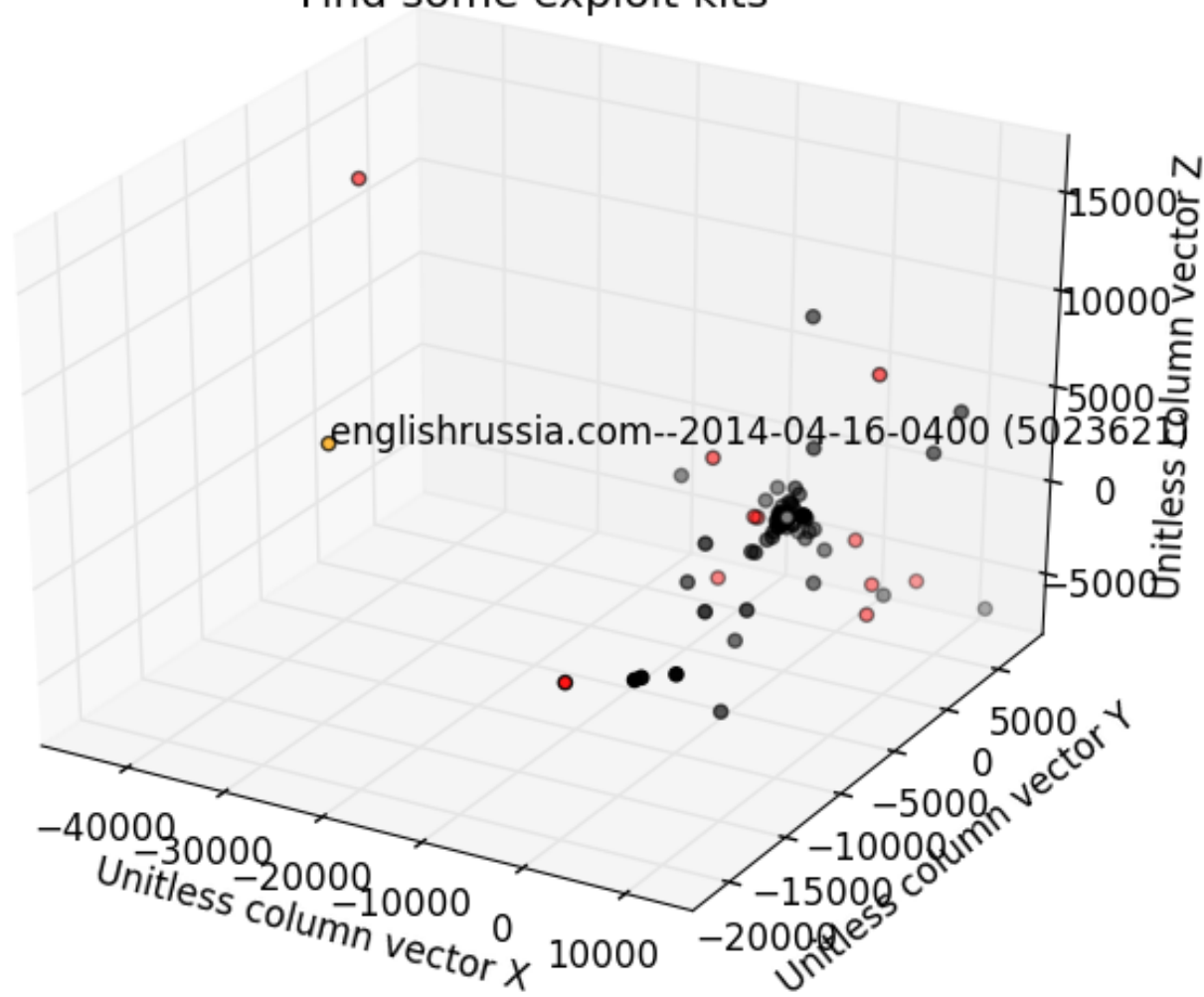


# Javascript with Kit Labels



# Sweet Orange, Redkit, and Blackhole

Find some exploit kits



# ~~Original~~ Revised Research Results

- We built a system to seek out EKs, visit them, and record the packets
- We parsed the Javascript files
  - we can spot unusual specimens
  - not yet able to associate Javascript syntax with specific EKs (n-grams may not be up to this task)
- Many EKs out there, but only a few are popular at a given time

# Discussion

- Lots of exploit kits exist, but only a few are popular
- Outliers can provide insight, but discard them to drill deeper
- Too many n-grams carry too little information
  - This is old news
  - There seems to be lots of Blackhole activity
  - Perhaps other EK activity is being drowned out
  - So select better (or fewer) features

# Limitations

- This approach is naïve, in the sense of using little or no domain knowledge
  - File structure information? No
  - N-grams → machine instructions? No
  - N-grams → Javascript language constructs? No
  - Specifying n-grams “of importance”? No
  - Similarity to “known” specimens? Some
  - Collection-related metadata? No
  - Knowledge of known actors? No

# Machine Learning

- We've shown examples of unsupervised ML
- Supervised ML, especially deep learning, is an area of ongoing research
  - Neural nets, especially convolutional neural nets
    - Can we build neural nets to distinguish malware from benign is a related problem
  - Long short-term memories, or LSTMs
    - To discern what functions do, not just opcodes
  - Quality data sets for training is important!

# Mix of Students

- Malware analysis courses have been a good source of students
  - Both basic and advanced
  - Half day version of first course available [here](#)
- Mixture of graduates and undergraduates
  - Inspired three PhD theses, in progress
  - Many MS theses and writing projects on this and related topics e.g. Android malware
- Cyberdawgs!



# Lessons Learned

- Finding good quality data is important, and half the battle
- Students need to know cyber *and* data science
- Machine learning was more important to this problem than we expected
- An excuse for CKN to learn Python!

# Selected Publications and Presentations

- Charles Nicholas, Robert Brandon, Joshua Domangue, Andrew Hallemeyer, Peter Olsen, Alison Pfannenstien and John Seymour, “The Exploit Kit Club”, Malware Technical Exchange Meeting, July 22-24, 2014, Albuquerque, NM. (poster session)
- John Seymour and Charles Nicholas, “Overgeneralization in Feature Selection for Classification of Malware”, Malware Technical Exchange Meeting, June 17-19, 2015, Boston, MA. (poster session)

# continued

- Digant Bhagya Channegowda, "Exploratory Analysis of Exploit Kit Javascript", M.S. Thesis, July 2015.
- Charles Nicholas, Brian Hillsley, Robert Brandon, Digant Channegowda, Tobechukwu Ezekwenna, Andrew Hallameyer, Jacob Kogan, Cameron Lee, Edward Mukasey, Peter Olsen, Alison Pfannenstien, John Seymour, Payal Singh, Brendan Stryker, "From AlphaPack to Zuponcic: a Survey of Exploit Kits", Malware Technical Exchange Meeting, June 17-19, 2015, Boston, MA. (poster session)
- John Seymour and Charles Nicholas, ""Quantum" Classification of Malware", presented at DefCon 23, Las Vegas, NV, August 2015.

# continued

- Charles Nicholas and Robert Brandon, “Document Engineering Issues in Malware Analysis”, ACM SIGWEB International Symposium on Document Engineering, Lausanne, Switzerland, September 8, 2015. (tutorial session)
- Revised version in preparation for 2016