SAND2011-1604P

Statistical Modeling of Malware to detect New Threats

Julie Ard

EEC 274 Winter 2011 Final Project

11 Mar 2011

Motivation

- Stuxnet went undetected for six months
  - Propagated via physical media and vulnerable hosts
  - Selectively infected only the hosts it wanted, stayed "under the radar"
- "The world's first precision-guided cyber munition"
- Expected that it will influence future emerging threats
  - AV tools detect signature and polymorphic variants
  - What about the next Stuxnet?

Problem Statement

- Two main detection categories
  - Signature Scanning
  - Anomaly Detection
- Can statistical models detect a malicious file not included in the original data set?

Related Work

- ESET and Symantec have performed detailed analyses of known Stuxnet variants
- 32 files collected from Offensive Computing
- Detection focus is on AV signature scanning

Approach/Methodology

- Use known malicious and benign software behavioral data to derive coefficients of chosen behavioral variables
- Validate models on randomly selected test data not included in the training set
- Test them on Stuxnet data
  - Not present in either the training or test set

Models

- Sans constant coefficient
  - $b_1*X_1 + b_2*X_2 + \ldots + b_{10}*X_{10} = Y$
- With constant coefficient
  - $b_0 + b_1*X_1 + b_2*X_2 + \ldots + b_{11}*X_{11} = Y$

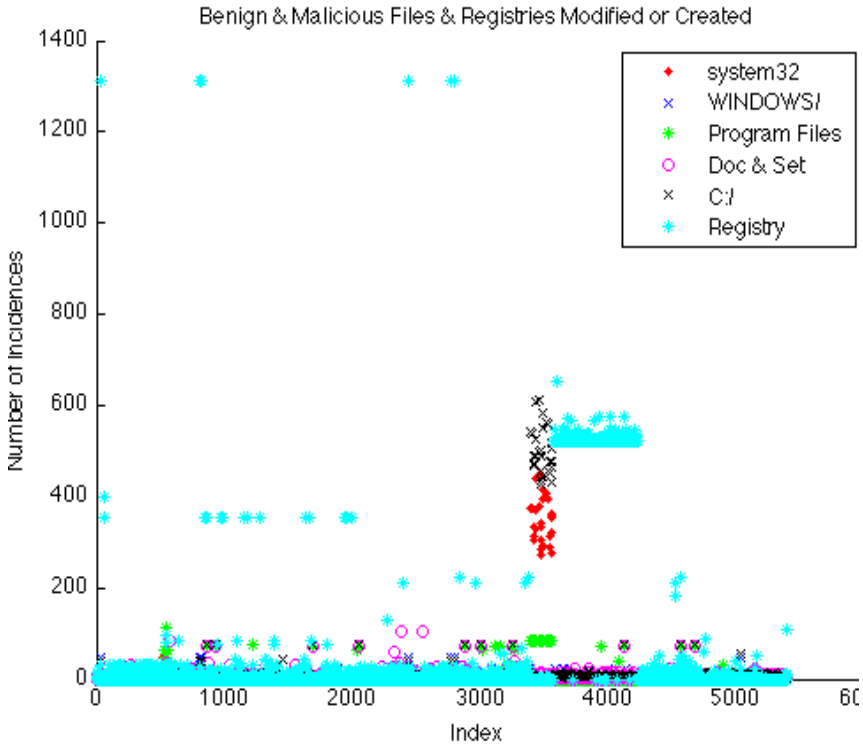
Model ID	Description	Distribution	Constant Coeff
MLR	Multilinear Regression	n/a	No
GLMN	Generalized Linear Model	Normal	No
GLMB	Generalized Linear Model	Binomial	No
GLMP	Generalized Linear Model	Poisson	No
GLMNC	Generalized Linear Model	Normal	Yes
GLMBC	Generalized Linear Model	Binomial	Yes
GLMPC	Generalized Linear Model	Poisson	Yes

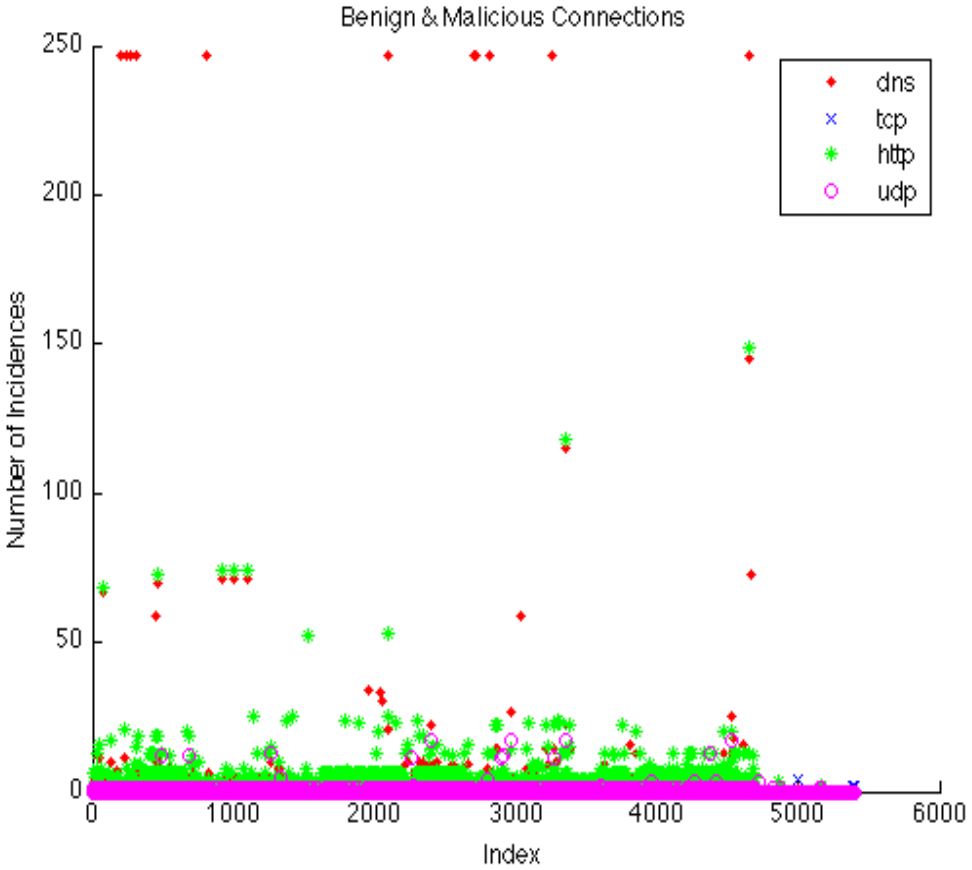
Variables

1. # of files created or modified in the C:/WINDOWS directory (excluding system32)
2. # of files created or modified in the C:/WINDOWS/system32 directory
3. # of files created or modified in the C:/Program Files directory
4. # of files created or modified in the C:/Documents and Settings directory
5. # of files created or modified in the root C:/ directory
6. # of registries read, created, or modified

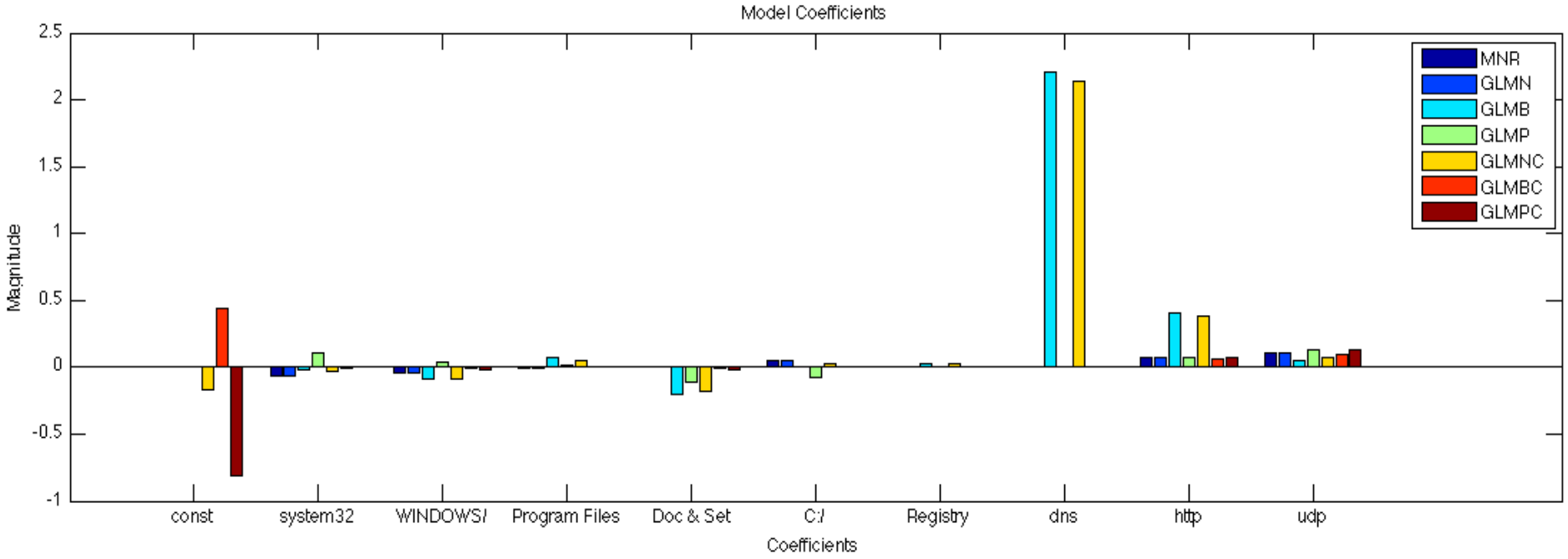
7. # of DNS queries
8. # of tcp connections
9. # of http connections
10. # of udp connections

Raw Data


Benign & Malicious Files & Registries Modified or Created
system32
WINDOWS/
Program Files
Doc & Set
C:/
Registry
Number of Incidences
Index
0
200
400
600
800
1000
1200
1400
1000
2000
3000
4000
5000


Benign & Malicious Connections
dns
tcp
http
udp
Number of Incidences
Index
0
50
100
150
200
250
1000
2000
3000
4000
5000
6000

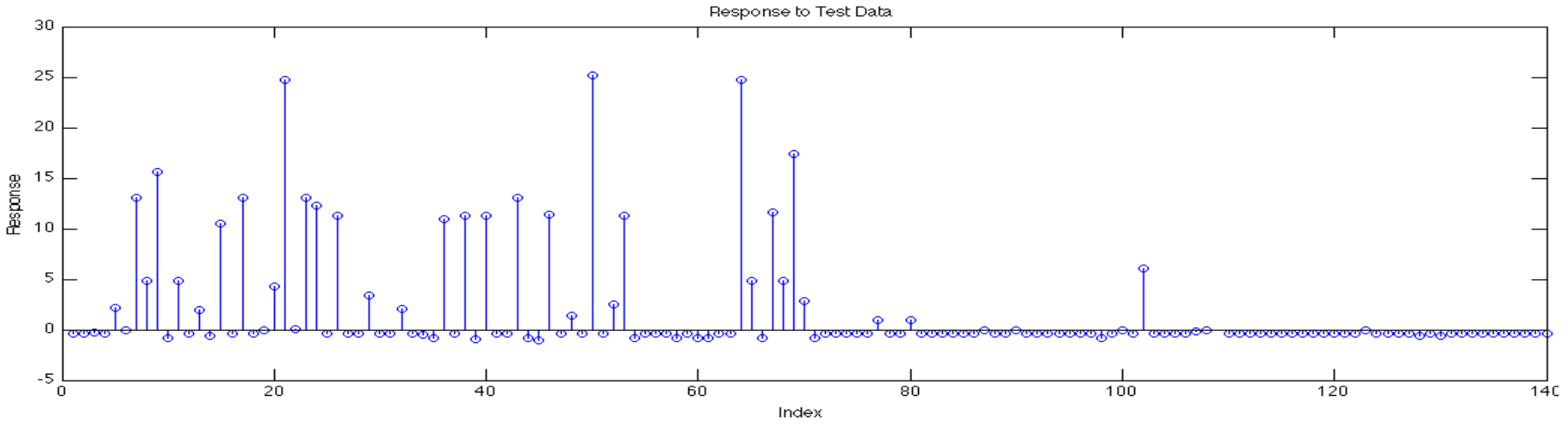
Results: Coefficients



- Negative coefficients associated with benign data (training score=0)
  - Constant, file activity, and TCP (not pictured)
- Positive coefficients imply malicious behavior
  - DNS, HTTP, UDP

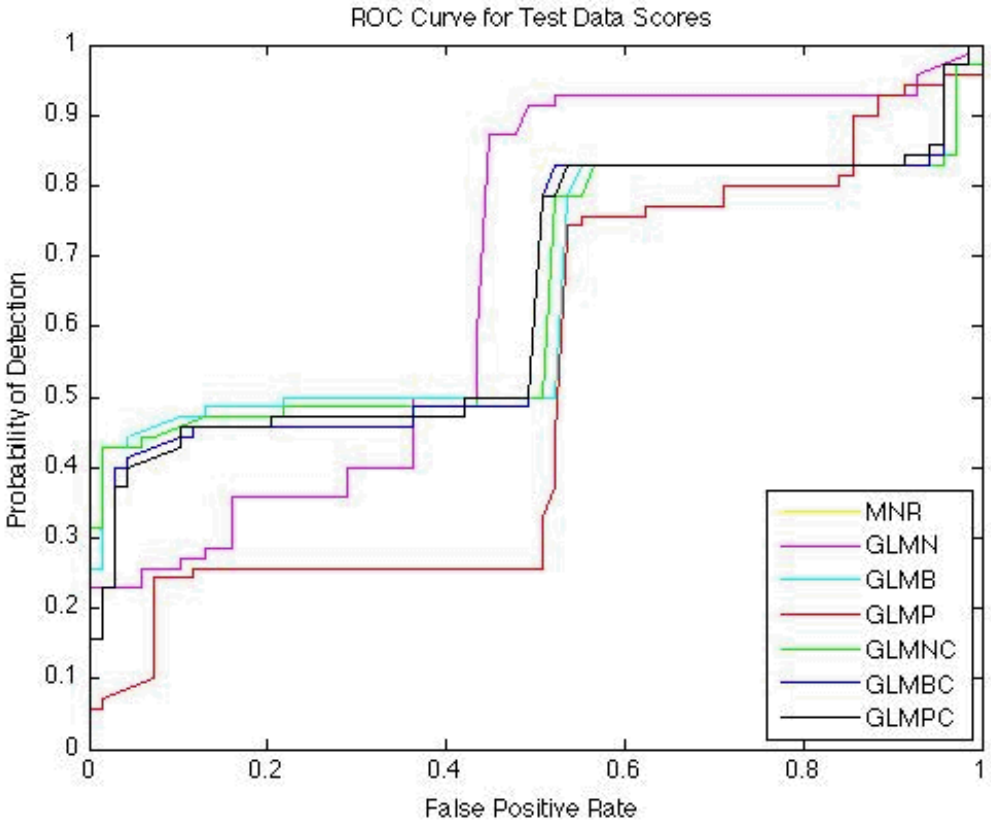
Results: Response to Training Data

- GLMB response pictured
- First half (1:70) malicious data
- Second half (71:140) benign data



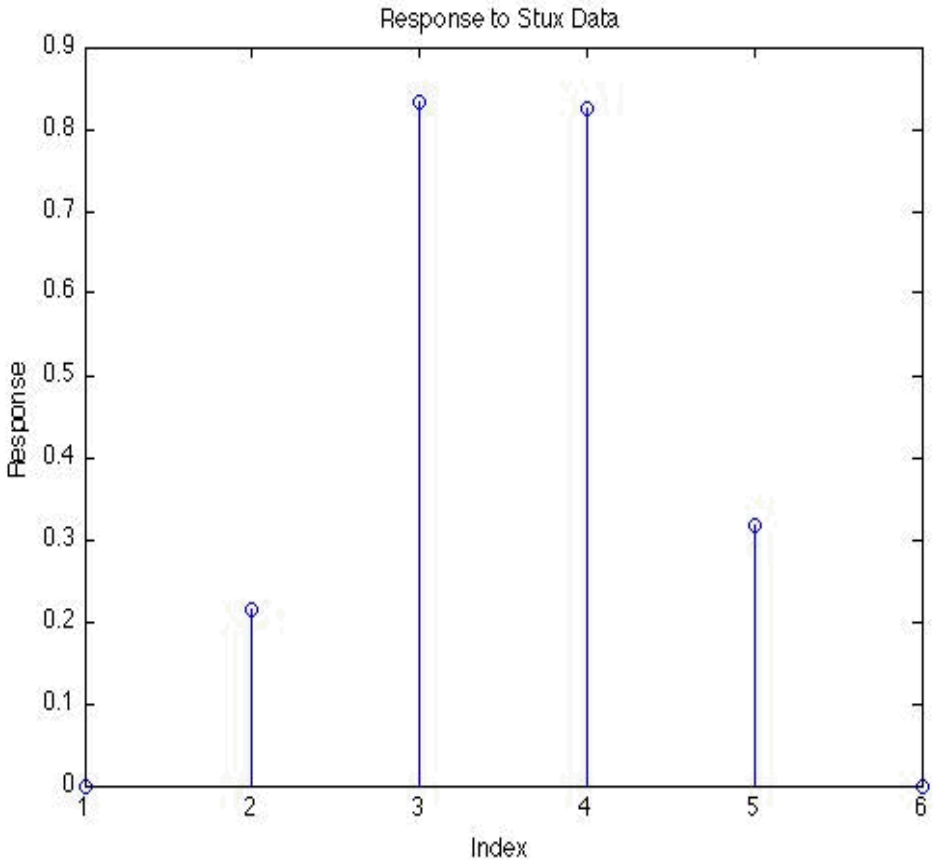
Results: ROC

- The GLMN/MNR models would perform the best with a simple threshold filter
  - 91% Pd
  - 48% FAR



Results: Stuxnet

- The GLMN and MLR models scored all 4 as benign
- All other models scored #2 &3 as malicious but #1 & 4 as benign



GLMNC Response to Stux

Conclusions

- Small influence of Registry variable consistent across all models
  - Stuxnet creates or modifies 21 on average
  - Known malware creates or modifies 3
  - Known benign software creates or modifies 1
- High influence of network activity correctly classified 2 of 4 as malicious

Conclusions

Stux	MLR	GLMN	GLMB	GLMP	GLMNC	GLMBC	GLMPC
1	B	B	B	B	B	B	B
2	B	B	M	M	M	M	M
3	B	B	M	M	M	M	M
4	B	B	B	B	B	B	B

- Benign data issues
  - Analysis not comparable with that of malware
  - Benign files ask permission and require user interaction; malware does not

Future Work

- More variables
- String analysis
- Variable-length data
- More model types (multivariate, higher order)
- Assignment of training scores
- Malicious data classification
- Conditional probabilities (events)

References

- PortableApps.com - Portable software for USB, portable and cloud drives. Rare Ideas, LLC. 6 3 2011 <http://portableapps.com>.
- Van Randwyk, Jamie, et al. "Farm: An automated malware analysis environment." 42nd Annual IEEE International Carnahan Conference on Security Technology. IEEE ICCST, 2008. 321-325.
- Matrosov, Aleksandr, et al. Stuxnet Under the Microscope. ESET. Revision 1.3.1. <http://www.eset.com>.
- Falliere, Nicolas, et al. W32.Stuxnet Dossier. Symantec Security Response. Version 1.4 (February 2011).