Leveraging Vulnerability Prediction Models to Aid Cyber Security Planning

Raymond Chow
raymond.chow@lmco.com
Lockheed Martin

Shahram Sarkani, Ph.D.
sarkani@gwu.edu
George Washington University

Thomas Mazzuchi, D.Sc.
mazzu@gwu.edu
George Washington University

Abstract # 21431

Presentation Outline

- Background
- Purpose
- Methodology
- Results
- Discussion
- Conclusion
- References
- Q&A

Background

- Maintaining Cyber Security has remained a challenge despite new research and advances in technology.
  - Threats – Little to no control
  - Vulnerabilities – More control, but reactive rather than proactive
- Research attempts to forecast the time to next vulnerability using publicly available data has met with mixed results.[1]
- Research conclusions addressed applicability of predictive models but did not provide examples where these models can be practically applied.[2]

Purpose

- Instead of gauging security by next vulnerability occurrence, this study examines whether the level of effort required for cyber security can be tempered by projected increases or decreases in the number of future vulnerabilities.
  - Develop vulnerability prediction models for representative system
  - Predict whether the security budget will need to be increased or funds can be held in reserve based on trends
  - Utilize seasonality in vulnerabilities to guide scheduling of cyber security activities

Methodology

- Step 1: Define Representative System
  - Red Hat Enterprise Linux Workstation
  - Red Hat Enterprise Linux Server
  - Windows Server 2008
  - Windows 7
  - Internet Explorer
  - Microsoft SQL Server

Methodology (cont. 1)

- Step 2: Collect Vulnerability Data
  - Source – National Vulnerability Database (NVD)[3]
  - Aggregate Totals by Month
- Step 3: Time Series Analysis
  - Autoregressive Integrated Moving Average (ARIMA)[4]
  - Exponential Smoothing Models[5]

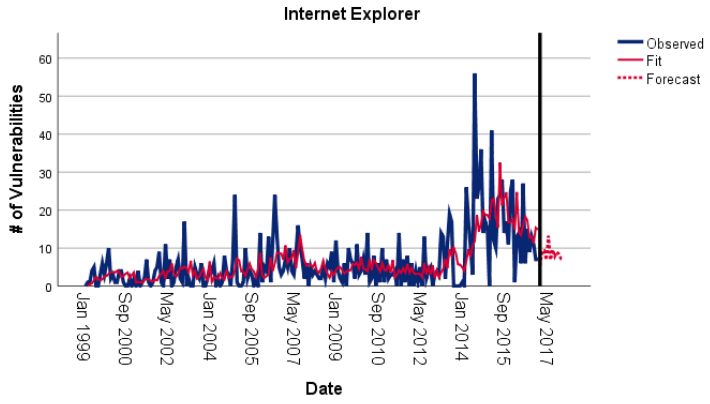
Methodology (cont. 2)

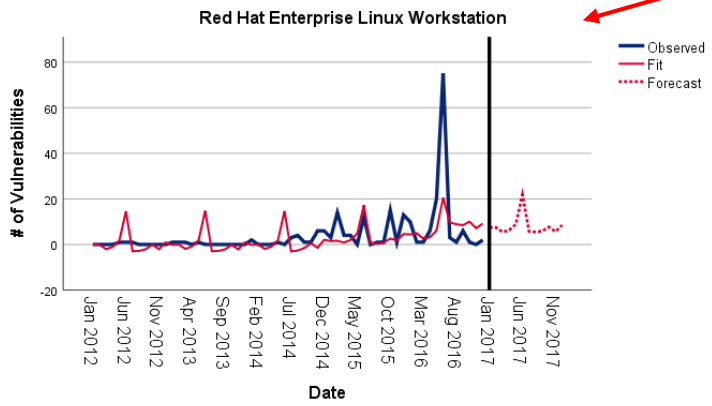
- Step 4: Utilize Predictions
  - Examine Predicted Vulnerability Totals to Guide Cyber Security Budget Allocation
  - Schedule Cyber Security Assessment and Resolution Activities to Minimize Vulnerability Persistence/Duration

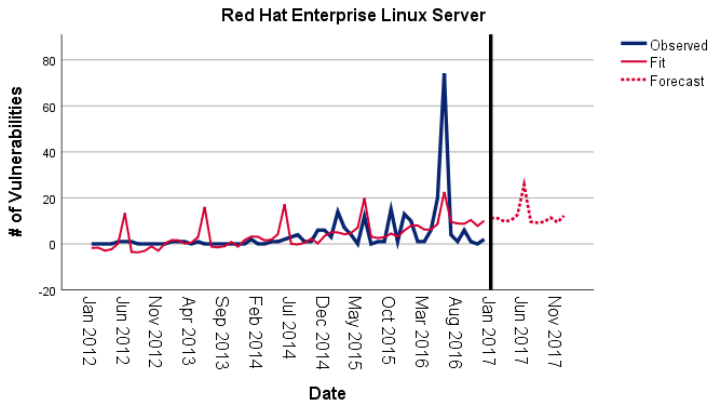
Leveraging Vulnerability Prediction Models to Aid Cyber Security Planning

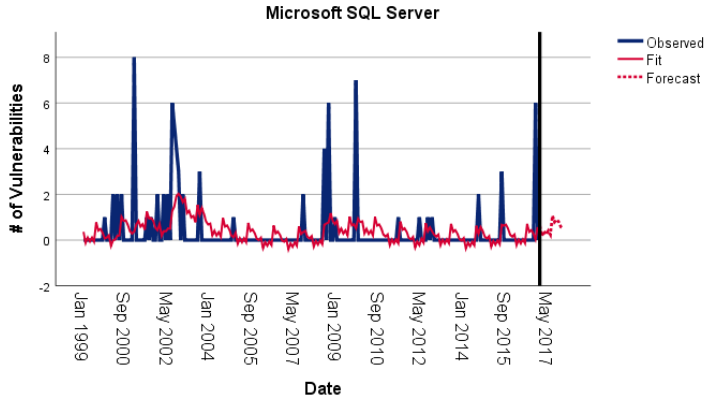
Results – Model Fitting

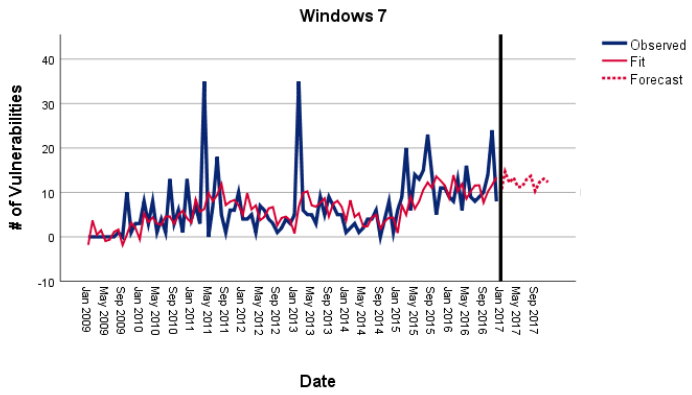
Note: Similar Vulnerabilities (Not an Error)

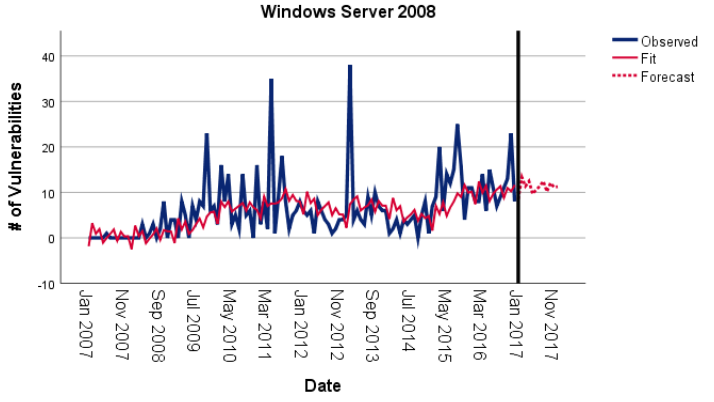












Results (cont. 1) – Statistical Values and Predictions

Software	Model	Stationary R^2	Normalized BIC	Ljung-Box[6]		
				Statistics	DF	Sig.
MS SQL Server	Simple Seasonal	0.732	0.396	13.724	16	0.619
RHEL Server	Simple Seasonal	0.444	4.521	9.121	16	0.908
RHEL Workstation	Simple Seasonal	0.454	4.558	7.651	16	0.959
Windows 7	Simple Seasonal	0.720	3.718	14.071	16	0.593
Windows Server 2008	Simple Seasonal	0.714	3.678	13.965	16	0.601
Internet Explorer	ARIMA(0,1,1)(1,0,0)	0.492	3.794	20.698	16	0.190
~~Internet Explorer~~	~~Simple Seasonal~~	~~0.631~~	~~3.727~~	~~37.699~~	~~16~~	~~0.002~~

Software	Prediction												Total (2017)
	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17	
MS SQL Server	1	0	0	0	0	0	1	1	1	1	1	1	7
RHEL Server	7	7	6	6	9	22	6	5	6	8	6	8	96
RHEL Workstation	7	7	5	6	9	22	6	5	6	8	6	8	95
Windows 7	10	15	12	13	11	11	13	14	10	12	13	12	146
Windows Server 2008	8	13	11	12	10	11	11	12	10	12	11	11	132
Internet Explorer	5	9	9	7	13	7	10	8	9	8	7	7	99
Total (2017)	38	51	43	44	52	73	47	45	42	49	44	47	575

Results (cont. 2) – Budget Considerations

Software	Vulnerabilities		
	Total (2016)	Predicted Total (2017)	% Change
MS SQL Server	6	7	16.66666667
RHEL Server	126	96	-23.80952381
RHEL Workstation	126	95	-24.6031746
Windows 7	134	146	8.955223881
Windows Server 2008	133	132	-0.751879699
Internet Explorer	129	99	-23.25581395
Overall Total	654	575	-12.0795107

- Lower # of Vulnerabilities Expected in 2017
- No Budget Modification Suggested
  - Small Predicted Change (~12%)

Results (cont. 3) – Scheduling Resolution Activities

Areas of Consideration	Prediction											
	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	38	51	43	44	52	73	47	45	42	49	44	47
Potential to be Addressed	0	38	89	132	176	228	301	348	393	435	484	528
Persistant Vulnerability Value	0	0	38	127	259	435	663	964	1312	1705	2140	2624

- Scheduling Goals:
  - Maximize Vulnerabilities Addressed
  - Minimize Persistent Vulnerability
  - Minimize Leftovers for Next Year

Leftovers on Jan-18 if No Action	
Open Vulnerabilities	575
Persistent Vulnerability Value	3152

Results (cont. 3) – Scheduling Resolution Activities

Areas of Consideration	Prediction											
	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	38	51	43	44	52	73	47	45	42	49	44	47
Potential to be Addressed	0	38	89	132	176	228	301	348	393	435	484	528
Persistent Vulnerability Value	0	0	38	127	259	435	663	964	1312	1705	2140	2624

Scheduled Activities in June

Areas of Consideration	Prediction					
	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	47	45	42	49	44	47
Potential to be Addressed	73	120	165	207	256	300
Persistent Vulnerability Value	0	0	47	139	273	456

Leftovers on Jan-18	
Open Vulnerabilities	347
Persistent Vulnerability Value	683

Scheduled Activities in July

Areas of Consideration	Prediction				
	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	45	42	49	44	47
Potential to be Addressed	47	92	134	183	227
Persistent Vulnerability Value	0	0	45	132	268

Leftovers on Jan-18	
Open Vulnerabilities	274
Persistent Vulnerability Value	448

Results (cont. 4) – Model Accuracy vs Actuals

Software	Prediction												Totals (Predicted)	Totals (Actuals)	SMAPE[7]
	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17			
MS SQL Server	1	0	0	0	0	0	1	1	1	1	1	1	7	1	
MS SQL Server (Actuals)	0	0	0	0	0	0	0	1	0	0	0	0			N/A
RHEL Server	7	7	6	6	9	22	6	5	6	8	6	8	96	51	
RHEL Server (Actuals)	4	0	0	8	0	6	5	1	0	17	3	7			108.4098
RHEL Workstation	7	7	5	6	9	22	6	5	6	8	6	8	95	40	
RHEL Workstation (Actuals)	4	0	0	8	0	5	5	1	1	7	2	7			102.5068
Windows 7	10	15	12	13	11	11	13	14	10	12	13	12	146	229	
Windows 7 (Actuals)	1	1	57	8	26	50	22	9	23	20	10	2			93.18991
Windows Server 2008	8	13	11	12	10	11	11	12	10	12	11	11	132	243	
Windows Server 2008 (Actuals)	1	1	60	15	27	51	22	10	25	19	10	2			89.32234
Internet Explorer	5	9	9	7	13	7	10	8	9	8	7	7	99	79	
Internet Explorer (Actuals)	0	1	11	3	6	7	7	7	7	5	12	13			63.84142
Total (Predicted)	38	51	43	44	52	73	47	45	42	49	44	47	575	643	
Total (Actuals)	10	3	128	42	59	119	61	29	56	68	37	31			53.96408

Results (cont. 5) – Budget Considerations Revisited

Software	Vulnerabilities		
	Total (2016)	Predicted Total (2017)	Actual (2017)
MS SQL Server	6	7	1
RHEL Server	126	96	51
RHEL Workstation	126	95	40
Windows 7	134	146	229
Windows Server 2008	133	132	243
Internet Explorer	129	99	79
Overall Total	654	575	643

- ✓ **Prediction of Lower # of Vulnerabilities vs 2016**
- ✓ **Budget Recommendation - Unchanged**

Results (cont. 6) – Scheduling Revisited

Areas of Consideration	Actuals											
	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	10	3	128	42	59	119	61	29	56	68	37	31
Potential to be Addressed	0	10	13	141	183	242	361	422	451	507	575	612
Persistent Vulnerability Value	0	0	10	23	164	347	589	950	1372	1823	2330	2905

Scheduled Activities in June

Areas of Consideration	Actuals					
	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	61	29	56	68	37	31
Potential to be Addressed	119	180	209	265	333	370
Persistent Vulnerability Value	0	0	61	151	297	511

Leftovers on Jan-18	
Open Vulnerabilities	401
Persistent Vulnerability Value	762

Scheduled Activities in July

Areas of Consideration	Actuals				
	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17
Vulnerabilities Discovered	29	56	68	37	31
Potential to be Addressed	61	90	146	214	251
Persistent Vulnerability Value	0	0	29	114	267

Leftovers on Jan-18	
Open Vulnerabilities	282
Persistent Vulnerability Value	457

✓ **July Remains the Optimal Month to Conduct Activities**

Discussion

- NVD Data Accuracy
  - Entries using different names (e.g. IE vs Internet Explorer)
  - Entries without clear software version
- Predictive Models Accuracy
  - Budget changes should be driven by significant trends with excess funds held in reserve
- Current Methods are Computationally Intensive

Conclusion

- Vulnerability Prediction Models can be Leveraged as a Planning Aid for Cyber Security Activities
  - Proactive allocation of resources
  - Balance between addressing maximum number of vulnerabilities and minimizing persistent vulnerabilities (while also minimizing future impact)
- Potential Future Research
  - Test applicability to other representative systems
  - Improve prediction accuracy

References

1. Zhang, Su, Xinming Ou, and Doina Caragea. 2015. "Predicting Cyber Risks through National Vulnerability Database." Information Security Journal: A Global Perspective 24 (4-6): 194-206. doi:10.1080/19393555.2015.1111961.
2. Roumani, Yaman, Yazan F. Roumani, and Joseph K. Nwankpa. 2015. "Time Series Modeling of Vulnerabilities." Computers & Security 51: 32-40. doi:10.1016/j.cose.2015.03.003.
3. National Vulnerability Database (NVD). "NVD Data Feeds." 2018. https://nvd.nist.gov/vuln/data-feeds (accessed August 15, 2018)
4. Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, and Greta M. Ljung. 2015. *Time Series Analysis : Forecasting and Control*. Wiley Series in Probability and Statistics. Fifth edition ed. New York: John Wiley & Sons, Incorporated.
5. Holt, Charles C. 2004. "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages." *International Journal of Forecasting* 20 (1): 5-10. doi:10.1016/j.ijforecast.2003.09.015.
6. G. M. Ljung and G. E. P. Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (2): 297-303. doi:10.1093/biomet/65.2.297.
7. Armstrong, Jon Scott. 1985. *Long-Range Forecasting*. 2. ed. ed. New York [u.a.]: Wiley.

Q&A

Raymond Chow
858-267-5349
Lockheed Martin
raymond.chow@lmco.com