



Inferring Patterns in Network Traffic: Time Scales and Variation

Soumyo Moitra
smoitra@sei.cmu.edu
INFORMS 2014
San Francisco



Report Documentation Page

*Form Approved
OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 21 OCT 2014	2. REPORT TYPE N/A	3. DATES COVERED	
4. TITLE AND SUBTITLE Inferring Patterns in Network Traffic: Time Scales and Variations		5a. CONTRACT NUMBER	
		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Moitra /Soumyo		5d. PROJECT NUMBER	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213		8. PERFORMING ORGANIZATION REPORT NUMBER	
		10. SPONSOR/MONITOR'S ACRONYM(S)	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
		12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited.	
13. SUPPLEMENTARY NOTES The original document contains color images.			
14. ABSTRACT			
15. SUBJECT TERMS			
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	
19a. NAME OF RESPONSIBLE PERSON			

This material is based upon work funded and supported by SEI Line Funding under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution.

The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013 and 252.227-7013 Alternate I.

This material was prepared for the exclusive use of INFORMS attendees and may not be used for any other purpose without the written consent of permission@sei.cmu.edu.

DM-0001793

Introduction

- A method and metrics for Situational Awareness
- SA → Monitoring trends and changes in traffic
- Analysis over time → Time series data analysis
- Metrics related to time series are key for SA

- Variations over time → Metrics for tracking
- Time windows and time scales are important to understand and interpret the metrics

Background

- Traffic patterns and variations important
 - Engineering and performance
 - SA for security – monitoring
- Important for anomaly detection
- Baselines for normalcy
- Thresholds | Inherent variations
- Alerts – Can be based on many metrics

- **Metrics based on variations in traffic**

References

Biersack, Callegari, and Matijasevic – Data Traffic Monitoring and Analysis

Box, Jenkins and Reinsel – Time Series Analysis: Forecasting and Control

Braun and Murdoch – A First Course in Statistical Programming with R

Brockwell and Davis – Time Series: Theory and Methods

Cowpertwait and Metcalfe - Introductory Time Series with R

Crovella and Krishnamurthy – Internet Measurement

Nucci and Papagiannaki – Design, Measurement and Management of

Large-Scale IP Networks

Park and Willinger – Self-Similar Network Traffic & Performance Evaluation

Shumway and Stoffer - Time Series Analysis and its Applications

Method of Analysis

- Analysis of flow data to investigate this issue
- Construct an initial time series | W and b
- Establish a time slot τ ($b < \tau < W$)
- Estimate the standard deviations within each τ
- Estimate the std. dev. of these std. dev.s **[H]**
- Compare this across varying bin sizes
- Vary time window (W)
- Compare τ -s across varying W | same bin size
- Metric can be tracked over time (successive W s)

Variance of the Variance

1	2	3	4	5	6	7	8	9	10	
τ_1					τ_2					
μ_1					μ_2					

Variance over time (traffic load)

Burstiness $\sim\sim$ Variance of the means

What about the variance of the variance?

\Rightarrow Heteroscedasticity

Estimating Heteroscedasticity [H]

1	2	3	4	5	6	7	8	9	10	
		τ_1					τ_2			
		μ_1					μ_2			
		σ_1					σ_2			

$$\sigma(\sigma_1, \sigma_2, \dots) = \mathbf{H}$$

Important to monitor \mathbf{H} as well.

Data and Design

- Analysis reported here was done with public domain data
- Two time windows (8 hours each)
- Two time scales ($b=4,8$ minutes)
- Analysis was done with SiLK and R
- Can be done with any flow data and scripts
- One set of comparisons shown
- A particular case of heteroscedasticity

Results

Table: Heteroscedasticity Estimates

(Overall standard deviation in parentheses)

($W1 = W2 = 8$ hours; $b1 = 4$ min, $b2 = 8$ min)

Time window	Time Scale = b1	Time Scale = b2
W1	13.57 MB (26.35 MB)	33.37 MB (50.10 MB)
W2	4.47 MB (10.66 MB)	7.99 MB (19.64 MB)

Conclusions

- An attack or intrusion usually implies some shift in traffic patterns
- One indicator of such shifts could be a change in the levels of heteroscedasticity
- This methodology has the potential to detect such attacks at an early stage
- Alert when H exceeds a threshold

Benefits

- This approach could detect attacks and intrusions that do not perturb the network traffic in other discernible ways
- Thus other techniques may not identify them early enough
- Early detection is important for effective mitigation

- This method also enhances SA by introducing a new metric to track traffic patterns

Future Work

Implications of changes H w. r. t. time scales?

Repeat the analysis: wide W & different networks

Predictions from attack/intrusion models $\langle H \rangle$

Test behavior of H with data with known attacks



Thank you!

Questions/comments?

