



Analysis of Some Time-Series Metrics for Network Monitoring

Soumyo Moitra smoitra@cert.org
FloCon 2014



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 FloCon 2014 attendees and may not be used for any other purpose without the written consent of permission@sei.cmu.edu.

Carnegie Mellon®, CERT® and FloCon® are registered marks of Carnegie Mellon University.

DM-0000654



Software Engineering Institute

Carnegie Mellon

Introduction

- A method and metrics for Situational Awareness
- SA → Monitoring trends and changes in traffic
- Analysis over time → Time series models
- Metrics related to time series are key for SA
- Correlations over time → Autocorrelation Function
- Time window and time scale are important to understand the ACF

Background

- The ACF shows how one observation in time is related to other observations at other points in time
- The ACF and most metrics related to time series are dependent on the time window (W) and the time scale (b) over which they are computed
- Therefore W and b are important for interpreting T-S metrics
- Identify short-term & long-term dependencies
- Important for anomaly detection

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 and Performance Evaluation
- Shumway and Stoffer - Time Series Analysis and its Applications

Method of Analysis

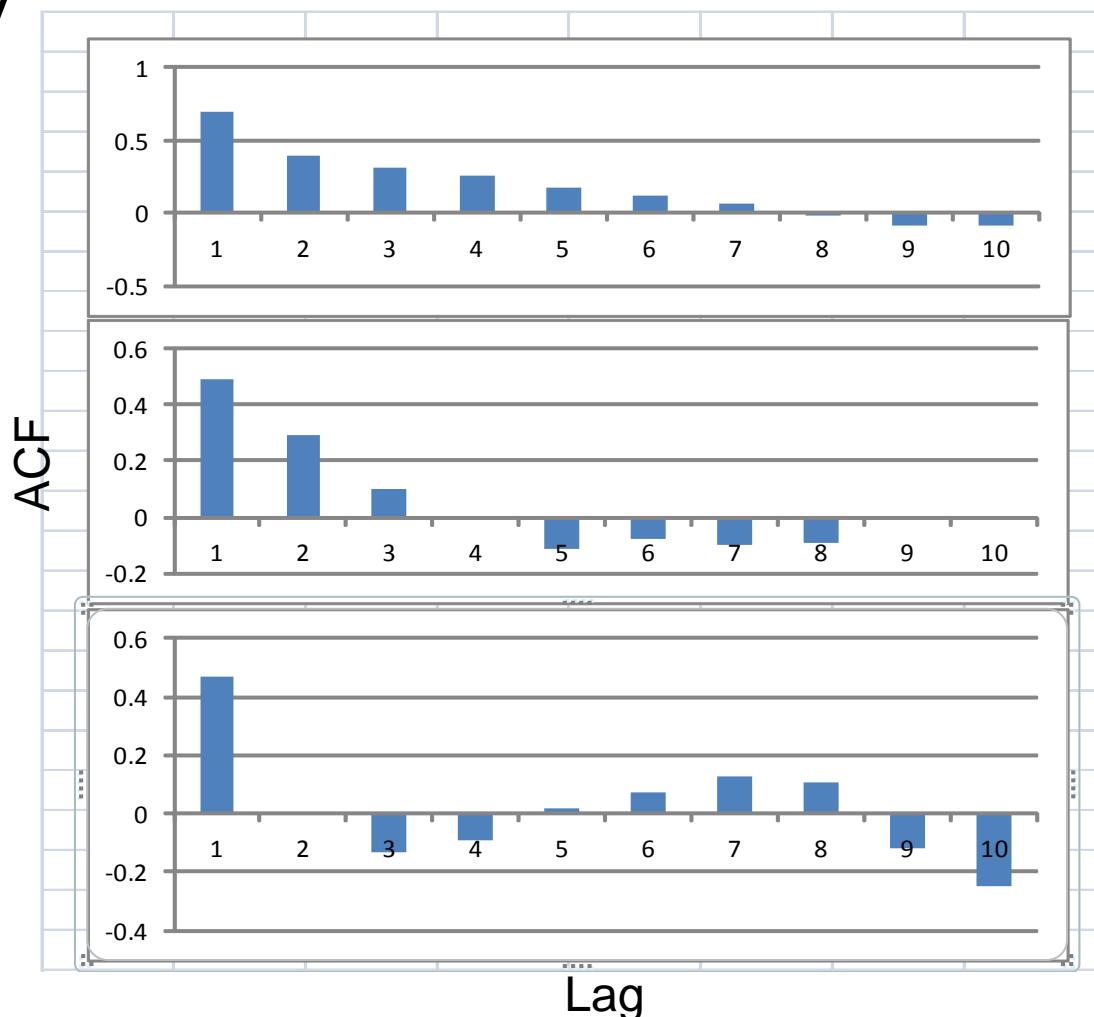
- Analysis of flow data to investigate this issue
- Construct an initial time series | W and b
- Estimate the autocorrelation function for this
- Vary the time scale (bin size) and estimate the ACF for each new time series
- Compare the ACFs across varying bin sizes
- Develop a metric to quantify the differences
- Vary time window (W)
- Compare ACFs across varying W |same bin size
- Metric can be tracked over time (successive Ws)

Data and Design

- Analysis reported here was done with publicly available data
- Three time windows (8 hours each)
- Three time scales ($b=4,8,16$ minutes)
- Analysis was done with SiLK and R
- Can be done with any flow data and scripts
- One set of comparisons shown (10 lags)
- One comparison of ACFs from two Ws
- Metric to investigate differences in ACFs:
 - = Sum of absolute differences

Results

Autocorrelations by
Time Scale –
Lags one to ten.



Discussion

- ACF1 (bin size = 4min.) -> 0 at lag 8; low negative values after that till lag 17.
- ACF2 (bin size = 8 min.) -> sharper decrease
 - -> 0 at lag 4; then approximately cyclical
 - Less long-term effect
- ACF3 (bin size = 16 min.) -> 0 at lag 2 [$\sim \text{MA}(1)$]
- ACFs across 2 time windows (bin size = 4min.)
 - Sum of absolute differences = 1
 - with mean = .1 (less than std. err.) >> **Stable**

Conclusions

- An attack or intrusion usually implies some shift in traffic patterns
- One indicator of such shifts could be a change from a stable long-term dependency to a short-term dependency
- This methodology has the potential to detect such attacks at an early stage

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 allows us to distinguish between short-term and long-term dependencies within traffic patterns
 - This distinction is important for selecting the appropriate techniques for further analyzing network traffic
 - E.G. Short term → Traditional Poisson/Erlang Models
 - E.G. Long term → Complicated Self-Similar Models

Future Work

Implications of changes in the ACF wrt time scales

Predictions from attack/intrusion models

Alternative metrics to quantify differences in ACFs

Repeat the analysis: wide W & different networks

Test methodology with data with known attacks



Thank you!

