



Modeling the Active and Idle Durations of Network Hosts

Soumyo Moitra
smoitra@sei.cmu.edu
2014



Copyright 2014 Carnegie Mellon University

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.

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 Conference on Applied Statistics in Defense and may not be used for any other purpose without the written consent of permission@sei.cmu.edu.

DM-0001581

Introduction

Important to understand network behavior of hosts

Durations active and idle

Patterns important for Situational Awareness

Baselining to detect anomalies

Decide whether a host should be in the inventory

Objectives of the Analysis

Distributions of the durations of active and idle times

Insights

Two metrics:

Probability of a host being active after a period of idleness

Conditional probability of a host becoming active within a time horizon
Given it has been idle for some time

Methodology

Flow data from the public domain

SiLK (CERT/SEI) and Unix Tools

Spreadsheets

Focus on web servers

References

- Bhattacharya, R. N. and Waymire, E. C. (2009) Stochastic Processes with Applications. SIAM.
- Brostrom, G. (2012) Event History Analysis with R. CRC Press.
- Crovella, M. and Krishnamurthy, B. (2006) Internet Measurement. John Wiley & Sons.
- Hayden, L. (2010) IT Security Metrics. McGraw Hill.
- Lawless, J. F. (2002) Statistical Models and methods for Lifetime Data. Wiley.
- Maindonald, J. and Braun, W. J. (2010) Data Analysis and Graphics using R: An Example-Based Approach. Cambridge University Press.
- Mills, M. (2011) Introducing Survival and Event History Analysis. Sage.
- Rausland, M. and Heyland, A. (2003) System Reliability Theory. Wiley.
- Ross, S. (2014) Applied Probability Models. Academic Press.
- Snyder, D. L. and Miller, M. I. (2011) Random Point Processes in Time and Space. Springer.

Analysis

Time series of network flows – out traffic

Time horizon = 23 hours

Time scale (bin size) = 1 hour

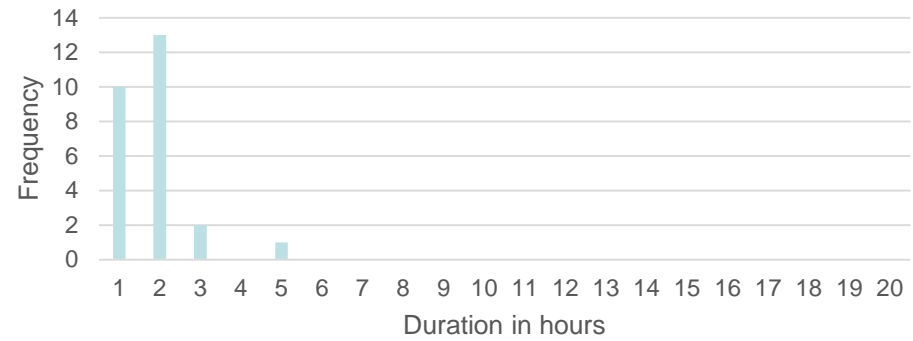
Convert volumes to a 0/1 series

Compute the durations of active and idle times

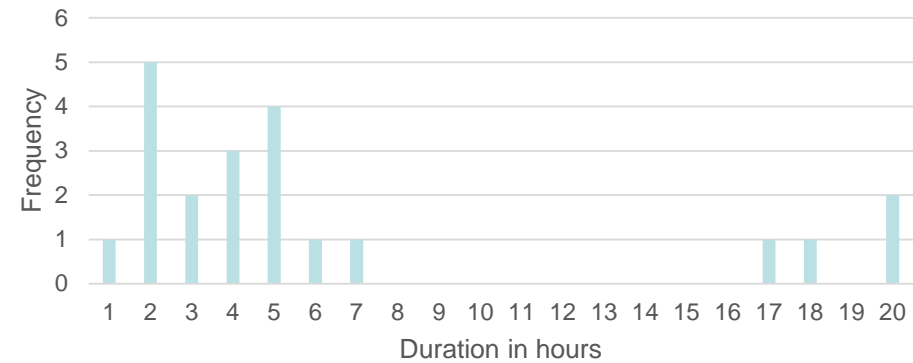
Plot the frequency distributions

Results

Distribution of active durations



Distribution of idle durations



Discussion

Active durations

Very compact (low variation – narrower than Poisson)

Mean = 1.8

Weibull?

Idle durations

Long tail or two populations

Issues with estimating the metrics

Censoring/Truncation problems

Future Work

Need much longer time series

Need to estimate the metrics with more data sets

Effects of varying the time scales and time horizons



Thank you!

Questions/comments?

