

# Metrics to Characterize Temporal Patterns in Lifespans of Artifacts

Soumyo Moitra

Software Engineering Institute  
Carnegie Mellon University  
Pittsburgh, PA 15213



Software Engineering Institute

Carnegie Mellon University

© 2016 Carnegie Mellon University

[Distribution Statement A] This material has been approved  
for public release and unlimited distribution.

REV-03.18.2016.0

Copyright 2017 Carnegie Mellon University. All Rights Reserved.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

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.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

DM17-0420

# INTRODUCTION

This paper presents analysis of artifact lifespans

Based on observations over time

- Presence or absence of vulnerabilities seen
  - Network links up or down
  - Servers active or idle
- Seen or Unseen at  $t$  (1/0)

Lifecycle → Lifespan → Analysis

→ Stochastic Point Processes (Marked)

Illustrative examples presented ← *Simulated data*

# Metrics for Pattern & Anomaly Detection

The goal is to track metrics

- Baseline them
  - Establish thresholds

Alerts – Validation – Action

# Background

Here we focus on Life History-based metrics

Many approaches to analyzing life histories

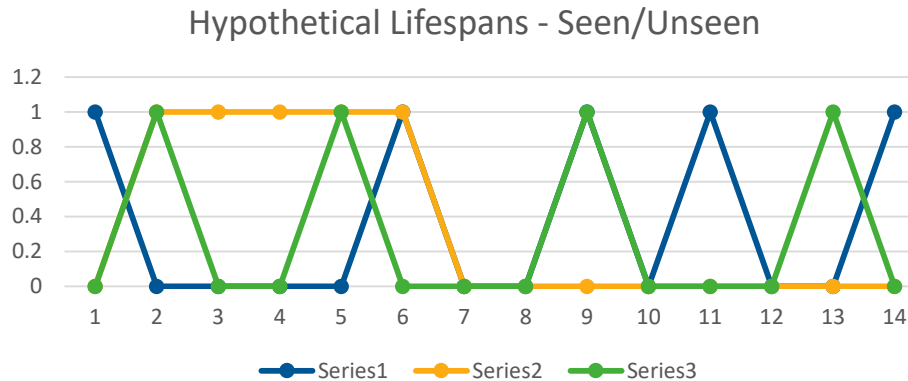
Not much analysis done for lifespans of artifacts

Development of appropriate metrics for lifespans that are not traditional lifecycles

Analyze time series constructed from temporal data (records)

# Modeling Challenges

## Not a traditional lifecycle



Something like a renewal process

Heterogeneous across the vulnerabilities

Need metrics to capture the lifespan features

Some existing metrics can be useful

Need some new concepts

# New Metrics for Lifespans/Point Processes

Existing: Lifetimes ( $\Pi$ ), Seen ( $\Psi$ ), Mean Times Seen ( $\langle T \rangle$ ), etc.

New:

- 1) Transilience  $\Phi$ : Count of 'seen - then unseen' sequences |  $W$
- 2) Sequacity  $\Xi$ : Count of seen consecutively |  $W$
- 3) Conformity  $\mathcal{C}$ : How close an artifact is to the median value of the metric across all the artifacts?

# Data, Methodology, Analysis

## Simulated Lifespans of Vulnerabilities

← Mentions over time (1 or 0) by day

\* 8 Lifespans & 14 days

Compute the metrics (Features) for each vulnerability

+ Functions of the metrics → New Features

9 Metrics or Features in all

# Results of the Analysis (Simulated Data)

PI	14	14	5	12	8	10	14	14
PSI	10	5	5	4	5	9	11	10
(PI-PSI)	4	9	0	8	3	1	3	4
PHI10	3	4	1	4	4	2	3	4
PI/(PI-PSI)	3.5	1.6	99	1.5	2.7	10	4.7	3.5
KSI	6	0	4	0	1	7	7	5
(KSI/PSI)	0.6	0	0.8	0	0.2	0.8	0.6	0.5
T	2.5	1	5	1	1.25	4.5	2.25	2
T-2.13	0.38	1.13	2.88	1.13	0.88	2.38	0.13	0.13
C	2.67	0.89	0.35	0.89	1.14	0.42	8.00	8.00

# Discussion of the Results

## Summary:

Independent in theory but correlated in real data

Different datasets will exhibit different correlations

Truncated data ( $W$ ) – Skewness in the distributions

## Potential Applications and Benefits:

Overall goal:

- Extract features of lifespans
- Understand patterns
- Cluster artifacts into similar groups
- Correlate patterns with particular malware

# Implications & Conclusions

These metrics help examine deeper temporal patterns:

Key to detecting subtle changes and surreptitious anomalies

Proposed 3 metrics that can be computed  
and tracked with relative ease

Based on stochastic point process models;  
all have intuitive interpretation

Properties match requirements to identify patterns

# Future Work

More data on lifespans: Baselineing and thresholds

Further validation of the metrics

Performance in detecting changes and anomalies in real data

Additional metrics to detect and track patterns

Implementation in information assurance analytics

*Thank you!*  
*Questions?*

Soumyo Moitra  
Senior Member of Technical Staff  
CERT/SEI/CMU  
Email: [smoitra@sei.cmu.edu](mailto:smoitra@sei.cmu.edu)