1ST BN WOCS (RTI)
169th REGT, Camp Nett, Niantic CT

Detecting SUNBURST Activity with Shannon Entropy

WOC Wong

Class 21-001

Overall Classification
for this
Information Brief
is:

UNCLASSIFIED

Agenda

- Background on SUNBURST malware.
- Design Considerations.
- Basics of Shannon Entropy.
- Applying Shannon Entropy as the Detection Methodology.

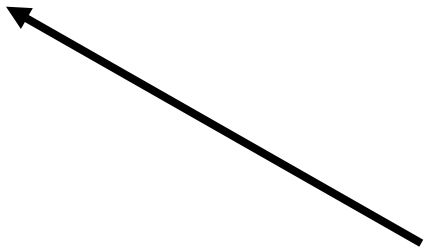
Background

- Malware identified in DEC 2020.
- Infrastructure established around JUL 2018.
  - Based on open-source WHOIS records
- Supply Chain Attack.
  - Threat actor injected malicious backdoor (SUNBURST) into Solarwinds Orion software updates.
- Global Attack Surface.
  - 18,000+ organizations affected.

Background

- SUNBURST C2 routinely used complex prefixes (KC: C2/Stage-1):

6a57jk2ba1d9keg15cbg[.]appsync-api[.]eu-west-1[.]avsvmcloud[.]com



Prefix contains encrypted information.

- Premise: C2 behavior & naming schema are predictable.
- Hypothesis: If behavior is exploitable, then detection is possible.

Design Considerations

Problems	Solutions
Heterogeneous Networks	Export Logs as Text Files
Text Files	Structured Data
Time & Scalability	Python Script
Compatibility	Python is Cross-Platform
Find Evil C2 Activity	Shannon Entropy

Basics of Shannon Entropy

- Measurement of information in text.

- Shannon Entropy characterizes:
  - Chaos (Uncertainty)
  - Randomness
  - Amount of Information

- Study of Information Theory.
  - Machine Learning (ML)
  - Artificial Intelligence (AI)

(source: MarketWatch)

Basics of Shannon Entropy (Formula)

- 2 Components:
  - Probability Table
  - Analysis Sample

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log_b \mathrm{P}(x_i)$$

(source: TowardsDataScience)

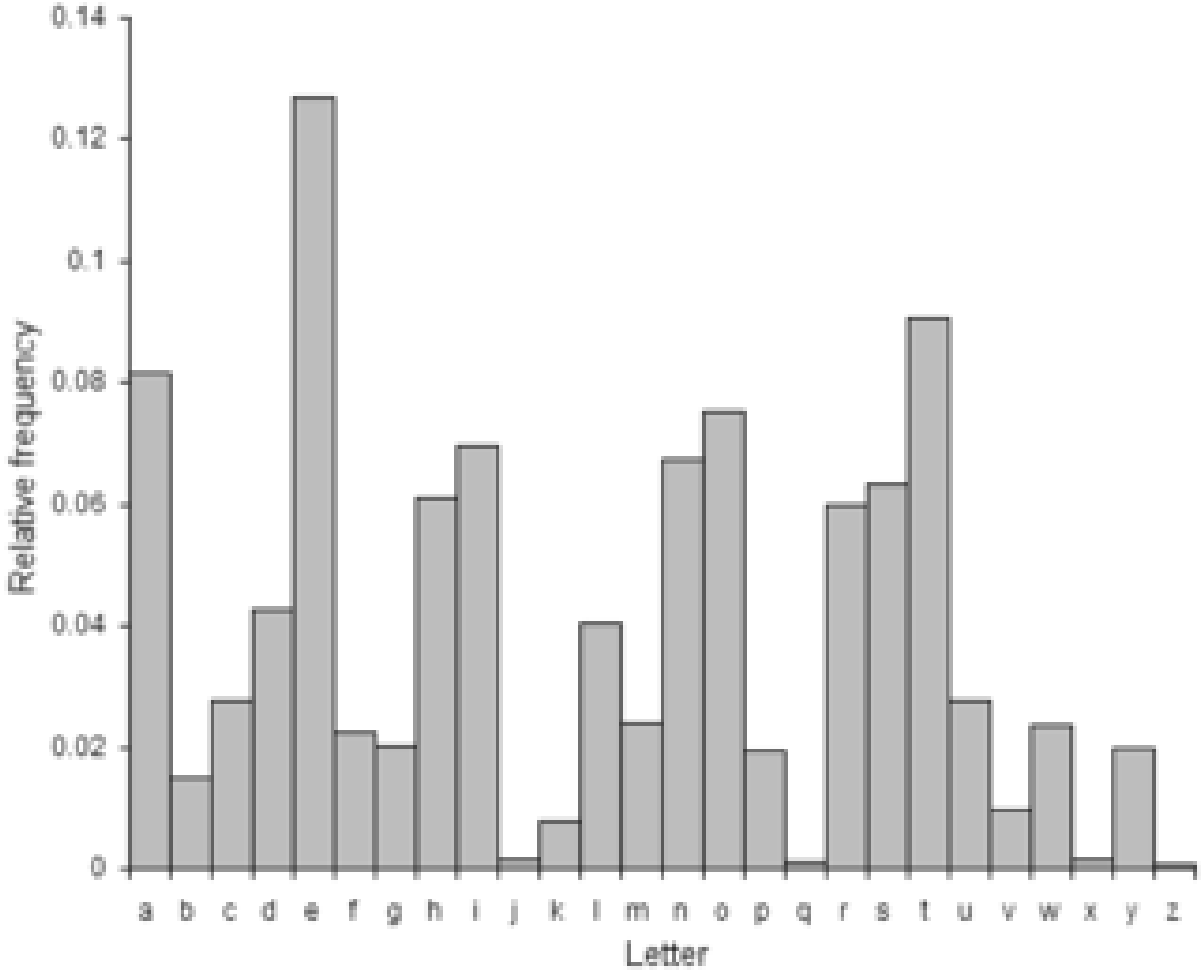
- $P(x_i)$: Probability of character
- H(X): “String” to analyze

```
# Iterate through each character of the URL.
for char in subdomain_prefix:
    if '.' in char:
        pass
    else:
        # Calculating entropy Log2 (because using binary values).
        p = freq_dict[char]
        this_entropy += p * math.log2(1/p)
```

Shannon Entropy (Probability)

- Assumptions:
  - Representative of population.
  - Data changes over time.
  - Data changes with location.
  - Frequency → Probability

- Probability Source Dataset:
  - English Dictionary (non-IT)
  - Alexa Top-1 Million (US)
  - Cisco Umbrella (US)
  - Majestic Million (UK)



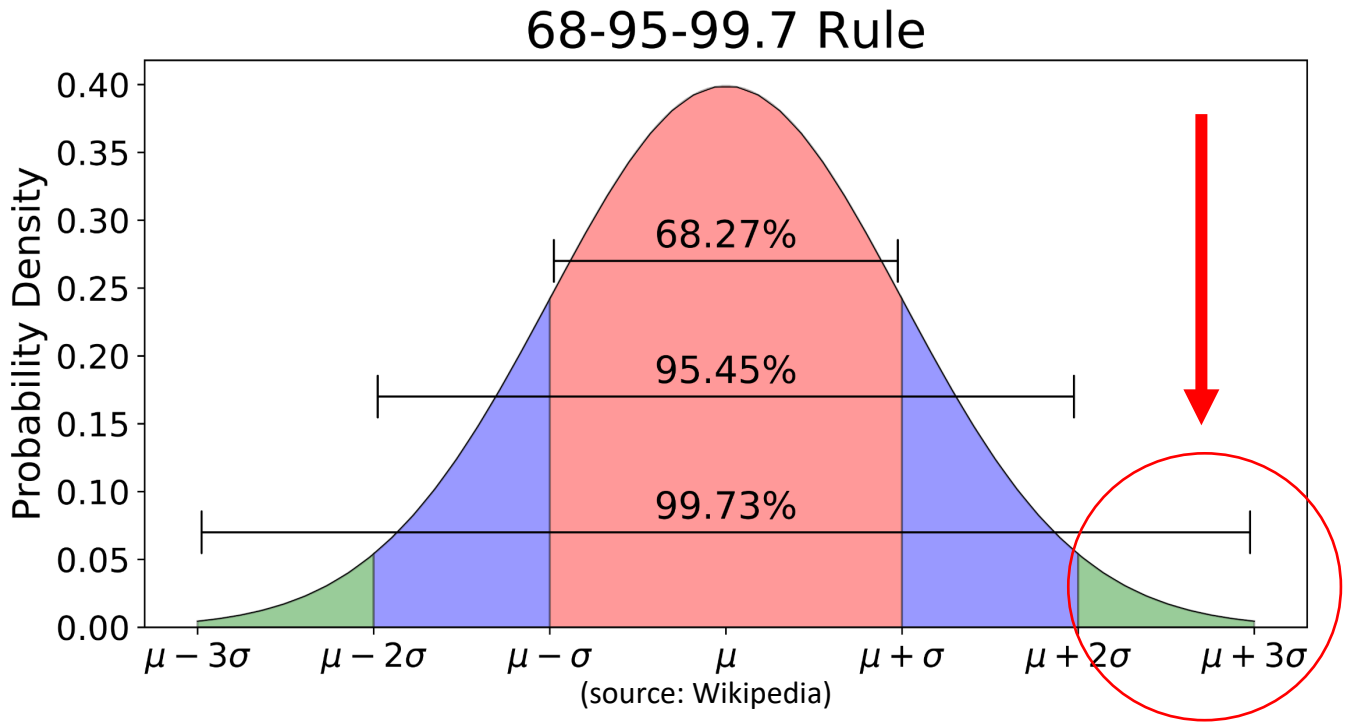
(source: Wikipedia)

Shannon Entropy (Log Analysis)

Observed Website	Shannon Entropy
www.google.com	0.431693137
www.icloud.com	0.431593137
static.nytimes.com	1.388405646
kinesis.us-east-1.amazonaws.com	1.5508556971591
nr2ia9qfa349b0q2oi60bou6iuir02rn.appsync-api.us-east-1.avsvmcloud[.]com	4.4128208776066

Method of Detection (1-Variable Test)

- Search for SUNBURST prefixes or assumed high entropy values.
- More than 2-standard deviations above average (top ~2%).



Method of Detection

Defining our norm.

```
### Entropy Test ###
Shannon_Entropy > 3.9907182572681723 is statistically significant!

### ENTROPY ###
Avg Entropy is: 1.627903, StdDev is: 1.181408
```

Identify top 2% of websites with very high entropy.

```
Shannon_Entropy: 4.3189148782803555; Suspect_FQDN: firebaseremoteconfig.googleapis.com
Shannon_Entropy: 4.3404967271871353; Suspect_FQDN: onestopdataanalysis.com
Shannon_Entropy: 4.3648502038196213; Suspect_FQDN: crashlyticsreports-pa.googleapis.com
Shannon_Entropy: 4.3748372396711073; Suspect_FQDN: zn42v6draxyafsjmv-homedepot.siteintercept.qualtrics.com
Shannon_Entropy: 4.4128208776066069; Suspect_FQDN: nr2ia9qfa349b0q2oi60bou6iuir02rn.appsync-api.us-east-1.avsvmcloud.com
Shannon_Entropy: 4.4336435117567277; Suspect_FQDN: mobileappcommunicator.auth.microsoft.com
```

Summary

- Background on SUNBURST malware.
- Design Considerations.
- Basics of Shannon Entropy.
  - Formula
  - Building Probability Tables from Character Frequencies.
- Applying Shannon Entropy as the Detection Methodology.
  - 1-Variable Statistical Test

Reference

Splunk Shannon Entropy

- https://www.splunk.com/en_us/blog/tips-and-tricks/when-entropy-meets-shannon.html

SANS Mark Baggett - Tool RedCanary used to analyze Alexa Top 1M

- https://github.com/markbaggett/freq

RedCanary - Blog where Probability scores come from

- https://redcanary.com/blog/threat-hunting-entropy/

Alexa's Top 1M Domains - Data Corpus used by RedCanary

- https://www.alexa.com/topsites

DGA Detector

- https://github.com/exp0se/dga_detector

Shannon Entropy - Formula

- https://towardsdatascience.com/the-intuition-behind-shannons-entropy-e74820fe9800