Applying Security Metrics for Quality Control and Situational Awareness

Jan Kohlrausch / Eugene A Brin
DFN-CERT
Introduction and Motivation

• A large quantity of technical threat intelligence feeds is available
• Threat intelligence platforms share technical threat intelligence data: MISP, ACDC “Central Clearing House (CCH)”
• But:
  - How to measure and assure data quality?
  - How to achieve an overview of the data?
  - Does this data contribute to strategical threat intelligence (situational awareness)?
Security Metrics: What is a „good“ Metric?

• Quantification of data characteristics
  - Number of incidents per month
  - Number of IDS alerts per day

• SMART or foolish?
  - Measurement should be well-defined
  - Measurement should contribute achieving a specific aim:
    ⇒ Quality control
    ⇒ Situational awareness
Security Metrics: Classes of Metrics

- Performance vs effectiveness (Marika Chauvin and Toni Gidwani, FIRST TI Symposium, March 2019, London):
  - **Performance**: Reasonable to maintain technical systems and develop software. Easy to measure
  - **Effectiveness**: Indication if a purpose has been accomplished (e.g. number of incidents). Harder to measure, but usually more expressive!
Security Metrics: Classes of Metrics

- Classification by use cases:
  - Quality metric: Assessment of data quality (effectiveness)
  - Operational metric: Gain insight into data properties (contributes to situational awareness)
  - Malware metric: Metrics focusing on Malware
Security Metrics: Types of Metrics

- Divided by methods of quantification:
  - **Counter**: Counting number of events (e.g. number of submitted reports)
    - **Uniqueness**: Counting unique items in the data set (e.g. IP addresses)
  - **Histogram**: Grouping data into bins (e.g. for real numbers and time spans)
  - **Statistics**: Measuring statistical properties of the data
Quality Metrics: 6 Dimensions

- **Accuracy**: Is the information correct?
- **Uniqueness**: Are duplicates in the data set?
- **Timeliness**: The time span between detection and submission
- **Consistency**: Do different partitions of the data have similar properties?
- **Completeness**: Are all submitted reports in the data set or are any reports missing?
- **Validity**: Are syntax and structure of reports correct?
Quality Metrics: Accuracy

• Hard if not impossible to measure:
  − Often data itself does not contain relevant information about correctness: e.g. Proxy / NAT gateways

• Mitigation:
  − Focus on “low hanging fruits”: IP addresses from private address spaces or Bogons.
  − Interaction of participating sites: “sightings” in MISP
Quality Metrics: Timeliness

• Time spans (e.g. difference between detection and submission) can be quantified as follows:
  - Histogram: E.g. one-hour bins
  - Statistical values: mean time (average) and standard deviation

• An acceptable delay depends on use case:
  - Incident handling
  - Blocking of attacks → fast reaction required
Quality Metrics: Completeness

• Completeness on a data set is hard to guarantee:
  − Often no central instance can measure completeness

• What we can do:
  − Measurement of numbers of reports to find significant gaps
  − Is there a significant difference between the expected and measured number of reports/events?
  − Give participating sites feedback pertaining to their data submissions
Quality Metrics: Uniqueness

• Not easy to define:
  - Identical events or duplicate features?

• Use case:
  - Count reports (e.g. DDoS): duplicates might be valuable
  - Incident reporting: rather avoid duplicates
Metrics for situational awareness

• Spot new emerging threats (*strategic* threat intelligence): Internet worms, IoC botnets, large scale attacks

• Is a baseline in the data?

• Are there:
  - Outliers?
  - Anomalies?
  - Change points?
Operational Metrics: Unique IP Addresses

- Number of unique IP addresses being submitted in a specific time span
- Special challenge: dynamically assigned IP addresses
- Contribution to situational awareness:
  - Reasonable to assume a “base line” if the number of events is sufficiently large
  - Significant increase over time points to large scale incident (e.g. new Internet worm or IoC Botnet)
  - May point to an incident in a network of a participating site
Operational Metrics: Novelty and Intersection of IP Addresses

- Number of unique IP addresses that are not present in the last time slice
- Special challenge: dynamically assigned IP addresses
- Contribution to situational awareness:
  - Number of newly compromised or suspicious systems (novelty)
  - Time span a system is compromised (intersection): Indication for incident handling effectiveness
Operational Metrics: Other important features

• Number of IP addresses (unique, intersection, and novelty) per ASN
  - Be aware of „political“ issues (e.g. worst ASN)
• Number of connections targeting TCP/UDP ports
• Port TCP/3389:
  - Emerging Windows RDP worm?
• Specific metrics (e.g. Sources targeting tcp/3389) on demand?
  - Number of metrics may explode
What comes next?

• Malware metrics: normalization of naming required

• Automation of baselining (consistency) and anomaly detection

• Test of statistical approach based on ARIMA
Thanks for your attention!

Questions?