How to Assign Weights to Different Factors in Vulnerability Analysis: Towards a Justification of a Heuristic Technique

Beverly Rivera, Irbis Gallegos, and Vladik Kreinovich

University of Texas at El Paso
El Paso, TX 79968, USA
barivera@miners.utep.edu, irbisg@utep.edu,
vladik@utep.edu

1. Need for Vulnerability Analysis

- Many important systems are vulnerable – to a storm, to a terrorist attack, to hackers' attack, etc.
- We need to protect them.
- Usually, there are many different ways to protect the same system.
- It is desirable to select the protection scheme with the largest degree of protection within the given budget.
- The corresponding analysis of different vulnerability aspects is known as *vulnerability analysis*.

2. Vulnerability Analysis: Reminder

- There are many different aspects of vulnerability.
- Usually, it is known how to gauge the vulnerability v_i of each aspect i.
- Thus, each alternative can be characterized by the corresponding vulnerability values $(v_1, \ldots, v_n)$.
- To compare alternatives, we need to combine the values v_i into a single index $v = f(v_1, \ldots, v_n)$.
- If one of the vulnerabilities v_i increases, then the overall vulnerability index v must also increase.
- Thus, $f(v_1, \ldots, v_n)$ must be increasing in each v_i.
- Usually, vulnerabilities v_i are reasonably small.
- Thus, we can expand $f(v_1, \ldots, v_n)$ in Taylor series in v_i and keep only linear terms: $v = c_0 + \sum\limits_{i=1}^{n} c_i \cdot v_i$.

3. Vulnerability Analysis (cont-d)

- Comparison does not change if we subtract the same constant c_0 from all the combined values:

$$v < v' \Leftrightarrow v - c_0 < v' - c_0.$$

- So, we can safely assume $c_0 = 0$ and $v = \sum\limits_{i=1}^{n} c_i \cdot v_i$.
- Similarly, comparison does not change if we re-scale all the values, e.g., divide them by $\sum\limits_{i=1}^{n} c_i$.
- This is equivalent to considering a new (re-scaled) combined function $f(v_1, \ldots, v_n) = \sum\limits_{i=1}^{n} w_i \cdot v_i$ with $\sum\limits_{i=1}^{n} w_i = 1$.
- The important challenge is how to compute the corresponding weights w_i.

4. How to Find Weights? Heuristic Solution

- For each aspect i, we know the frequency f_i with which this aspect is mentioned in the corr. requirements.
- Sometimes, this is the only information that we have.
- Then, it is reasonable to determine w_i based on f_i, i.e., to take $w_i = F(f_i)$ for some function $F(f)$.
- The following empirical idea works well: take $w_i = c \cdot f_i$.
- A big problem is that this idea does not have a solid theoretical explanation.
- In this talk, we provide a possible theoretical explanation for this empirically successful idea.

5. Towards a Theoretical Explanation

- The more frequently the aspect is mentioned, the more important it is: $f_i > f_j \Rightarrow w_i = F(f_i) > F(f_j) = w_j$.
- So, $F(f)$ must be *increasing*.
- For every combination of frequencies $f_1, \ldots, f_n$ for which $\sum\limits_{i=1}^{n} f_i = 1$, the resulting weights must add up to 1:

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} F(f_i) = 1.$$

- **Proposition.** *Let* $F : [0,1] \to [0,1]$ *be an increasing f-n for which* $\sum\limits_{i=1}^{n} f_i = 1$ *implies* $\sum\limits_{i=1}^{n} F(f_i) = 1$. *Then,*

$$F(x) = x.$$

- This justifies the empirically successful heuristic idea.

6. Towards a More General Approach

- So far, we assumed that the i-th weight w_i depends only on the i-th frequency f_i.
- Alternatively, we can normalize the "pre-weights" $F(f_i)$ so that they add up to one: $w_i = \dfrac{F(f_i)}{\sum\limits_{k=1}^{n} F(f_k)}$.
- In this more general approach, how to select $F(f)$?
- *Example:* we have four aspects, each mentioned n_i times, then $f_i = \dfrac{n_i}{n_1 + n_2 + n_3 + n_4}$.
- For some problems, the fourth aspect is irrelevant, so $v_4 = 0$ and $v = w_1 \cdot v_1 + w_2 \cdot v_2 + w_3 \cdot v_3$.
- On the other hand, since the 4th aspect is irrelevant, it makes sense to only consider n_1, n_2, and n_3:

$$f'_i = \frac{n_i}{n_1 + n_2 + n_3}.$$

7. General Approach (cont-d)

- Based on the new frequencies f'_i, we can compute the new weights w'_i and

$$v' = w'_1 \cdot v_1 + w'_2 \cdot v_2 + w'_3 \cdot v_3.$$

- Whether we use v or v', the selection should be the same.
- To make sure that the selections are the same, we must guarantee that $\dfrac{w'_i}{w'_j} = \dfrac{w_i}{w_j}$.
- The new frequencies f'_i can be obtained from the previous ones by multiplying by the same constant:

$$f'_i = \frac{n_i}{n_1 + n_2 + n_3} = \frac{n_1 + n_2 + n_3 + n_4}{n_1 + n_2 + n_3} \cdot \frac{n_i}{n_1 + n_2 + n_3 + n_4} = k \cdot f_i.$$

- Thus, the requirement takes the form $\dfrac{F(k \cdot f_i)}{F(k \cdot f_j)} = \dfrac{F(f_i)}{F(f_j)}$.

8. General Approach: Main Result

- **Proposition.** *For an increasing f-n $F : [0,1] \to [0,1]$:*

$$\frac{F(k \cdot f_i)}{F(k \cdot f_j)} = \frac{F(f_i)}{F(f_j)} \text{ for all } k, f_i, f_j \Leftrightarrow F(f) = C \cdot f^\alpha \text{ for } \alpha > 0.$$

- So, we should take $F(f) = C \cdot f^\alpha$.
- *Discussion:*
  - The previous case corresponds to $\alpha = 1$.
  - If we multiply all the values $F(f_i)$ by a constant C, then the resulting weights do not change.
  - Thus, from the viewpoint of application to vulnerability, it is sufficient to consider only functions

$$F(f) = f^\alpha.$$

9. Possible Probabilistic Interpretation of $w_i = f_i$

- Let us assume that the actual weights of two aspects are w_1 and $w_2 = 1 - w_1$.
- Let us also assume that vulnerabilities v_i are independent identically distributed random variables.
- A document mentions the 1st aspect if this aspect is more important (i.e., $w_1 \cdot v_1 > w_2 \cdot v_2$), so:

$$f_1 = P(w_1 \cdot v_1 > w_2 \cdot v_2).$$

- In a reasonable situation when both vulnerabilities are exponentially distributed, we have

$$w_1 = P(w_1 \cdot v_1 > w_2 \cdot v_2), \text{ i.e., } w_i = f_i.$$

10. Acknowledgments

- This work was supported by
  - the University of Texas at El Paso Regional Cyber and Energy Security Center (RCES)
  - supported by the City of El Paso's Planning and Economic Development division.
- This work was also supported in part by the National Science Foundation grants
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721.

11. Appendix: Proof of the First Result

- We require that $\sum\limits_{i=1}^{n} f_i = 1$ implies $\sum\limits_{i=1}^{n} F(f_i) = 1$.
- We want to prove that $F(f) = f$ for all f.
- For $n = 1$ and $f_1 = 1$, we get $F(f_1) = F(1) = 1$.
- For $f_1 = 0$ and $f_2 = 1$, we get $F(0) + F(1) = 1$ hence $F(0) = 1 - F(1) = 1 - 1 = 0$.
- For every $m \geq 2$, for $f_1 = \ldots = f_m = \dfrac{1}{m}$, we get $\sum\limits_{i=1}^{m} F(f_i) = m \cdot F\left(\dfrac{1}{m}\right) = 1$, hence $F\left(\dfrac{1}{m}\right) = \dfrac{1}{m}$.
- For every $k \leq m$, for $f_1 = \dfrac{k}{m}$ and $f_2 = \ldots = f_{m-k+1} = \dfrac{1}{m}$, we get $F\left(\dfrac{k}{m}\right) + (m-k) \cdot F\left(\dfrac{1}{m}\right) = 1$, hence

$$F\left(\frac{k}{m}\right) = 1 - (m-k) \cdot F\left(\frac{1}{m}\right) = 1 - (m-k) \cdot \frac{1}{m} = \frac{k}{m}.$$

12. Proof (cont-d)

- We have proved that $F\left(\frac{k}{m}\right) = \frac{k}{m}$ for any rational number $\frac{k}{m}$.
- Any real number f can be approximated by rational numbers: $\frac{k}{m} \le f < \frac{k+1}{m}$.
- When $m \to \infty$, we have $\frac{k}{m} \to f$ and $\frac{k+1}{m} \to f$.
- Due to monotonicity,

$$\frac{k}{m} = F\left(\frac{k}{m}\right) \le F(f) < F\left(\frac{k+1}{m}\right) = \frac{k+1}{m}.$$

- In the limit $m \to \infty$, we conclude that $F(f) = f$ for any real number f. Q.E.D.