An Uncertain Graph-based Approach for Cyber-security Risk Assessment

Hoang Hai Nguyen (Frank), David M. Nicol
Information Trust Institute
University of Illinois at Urbana-Champaign

7th Annual
Hot Topics in the Science of Security
September 22-23, 2020 | Lawrence, Kansas

I. Background and Motivation

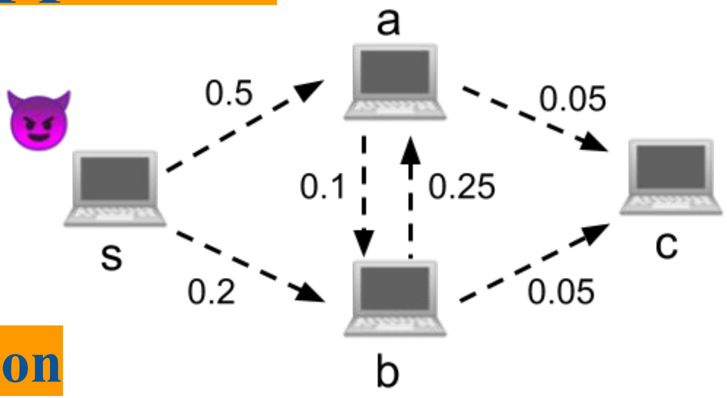
- Understanding the impacts of cyber-attacks allow business to compare the effectiveness of different defense solutions.

- Assessing the risk of a cyber-attacker who *gets access to the network, moves laterally, compromises critical assets, and causes damages* is challenging due to *uncertainty about the system's vulnerabilities and the attacker's ability to find and exploit them*.

- Quantification of losses to the network due to cyber-attacks must explicitly account for such uncertainty.

II. Modeling Approach

Figure 1:



(a) Attack propagation

Model the network as an *uncertain graph* **G = (V, E, p)** [1]

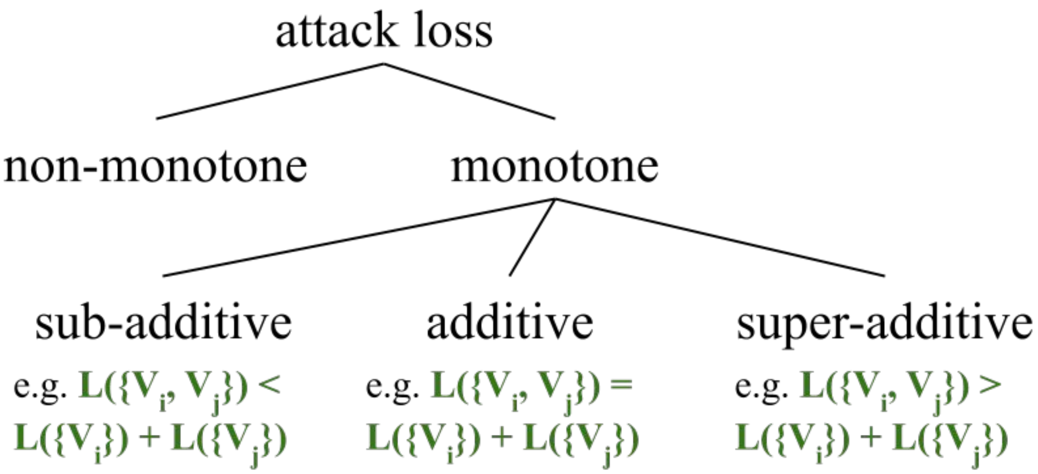
- $\mathbf{V = \{V_1, V_2, ..., V_n\}}$: hosts in the network
- $\mathbf{E = \{E_1, E_2, ..., E_m\}}$: links between hosts that allow attacks
- $\mathbf{p = (p_1, p_2, ..., p_m)}$ where $\mathbf{p_i}$ is the probability that $\mathbf{E_i}$ exists

Let $\mathbf{s \in V}$ be the starting point of the attack (Figure 1).

[1] Nguyen, H. H., Palani, K., and Nicol, D. M. *An approach to incorporating uncertainty in network security analysis*, HoTSoS (2017).

(b) Attack impact

- Cyber-attacks may induce losses of various kinds including *direct financial losses* (e.g. system downtime) and *indirect losses* (e.g. loss of reputation).
- Define the *attack loss* function $\mathbf{L}: \mathbf{V} \rightarrow \mathbf{R}_{\geq 0}$ and consider $\mathbf{L}$ as *a function of the set of hosts in V that can be reached from* $\mathbf{s}$.
- Several types of $\mathbf{L}$: for some $\mathbf{V}_i, \mathbf{V}_j \in \mathbf{V}$



III. Cyber-security Risk Assessment

(a) Risk triplet [2]

- **Realization:** let $\mathcal{X} = \{0, 1\}^m$ and $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ the multivariate random variable where $X_i \sim \text{Bernoulli}(p_i)$ for $i = 1, 2, \ldots, m$ and $X_i = 1$ implies E_i exists. An element $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ in $\mathcal{X}$ is a realization of $\mathbf{X}$.

- **Probability:** given $\mathbf{x}$, assume X_i's are *mutually independent*

$$\Pr(\mathbf{X} = \mathbf{x}) = \Pi_i \, (x_i \, p_i + (1-x_i)\,(1-p_i))$$

(otherwise see technique in [1] for modeling edge correlations)

- **Impact:** given $\mathbf{x}$, define the directed graph $G(\mathbf{x}) = (V, E(\mathbf{x}))$ where $E(\mathbf{x}) = \{E_i \in E: x_i = 1\}$. Let $V_s(\mathbf{x}) \subseteq V$ containing all nodes in $G(\mathbf{x})$ that can be reached from s. The impact is simply defined as $L(\mathbf{x}) \equiv L(V_s(\mathbf{x}))$.

[2] Kaplan, S., and Garrick, B. J. *On the quantitative definition of risk*. Risk Analysis 1 (1981).

(b) Risk measure as security metric

- **Expected loss (EL):** $E(L(X)) = \Pi_{x \in X} L(x) \Pr(X = x)$

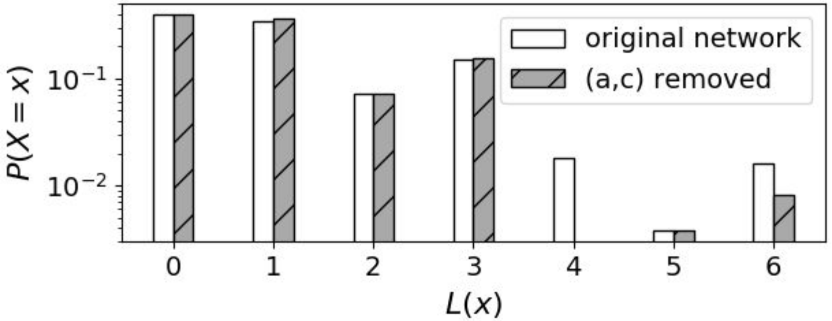
Example: assume additive loss with $L(\{s\}) = 0$, $L(\{a\}) = 1$, $L(\{b\}) = 2$, and $L(\{c\}) = 3$. The model in Figure 1 generates $2^6 = 64$ realizations with $L(X) \in \{0,1, \ldots, 6\}$ (Figure 2) and $E(L(X)) = 1.119$.

- **Loss tail probability (LTP):** let T be a selected threshold and $\mathcal{T} = \{x \in \mathcal{X} : L(x) > T\}$. The LTP is defined as

$$\Pr(L(X) > T) = \Pi_{x \in X} 1_{\{x \in \mathcal{T}\}} \Pr(X = x)$$

Example (cont.): using $T = 3$, we have $\Pr(L(X) > 3) = 0.038$.
Suppose the network access control is hardened and link (a,c) is removed as a result. The new EL remains relatively the same at **1.041** but the new LTP drops by 3x times to **0.012**.

Figure 2:



IV. Future Work

- Study different forms of the attack loss function.
- Study *computational techniques* for estimating the expected loss and loss tail probability (both are #P-complete [3]).
- Extend the model to capture *attacker's behaviors* and *interactions between the attacker and the defender*.
- Use the model to compute *premium for cyber-insurance*.

[3] Valiant, L. G. *The complexity of enumeration and reliability problems*. SIAM Journal on Computing 8, 3 (1979), 410–421.

Thank you!

Contact: hnguye11@illinois.edu