An Approach to Incorporating Uncertainty in Network Security Analysis

H. H. Nguyen, K. Palani, D. M. Nicol
University of Illinois at Urbana-Champaign

HoTSoS Symposium and Bootcamp, April '17

Background

- Increasing number of cyber-attacks per year
  - Many follow the **cyber kill chain** template[1]

- Today's computer networks are large, complex, and dynamic
  - Beyond the reasoning capability of human mind
  - Analyzable by computers -- given the appropriate models

- **Uncertainty** is an indispensable part of every model
  - Have to live with it
  - Reasoning about uncertainty is subtle but not impossible

1. M. Lee et al. *Analysis of the cyber attack on the Ukrainian power grid*. SANS ICS Report, 2016

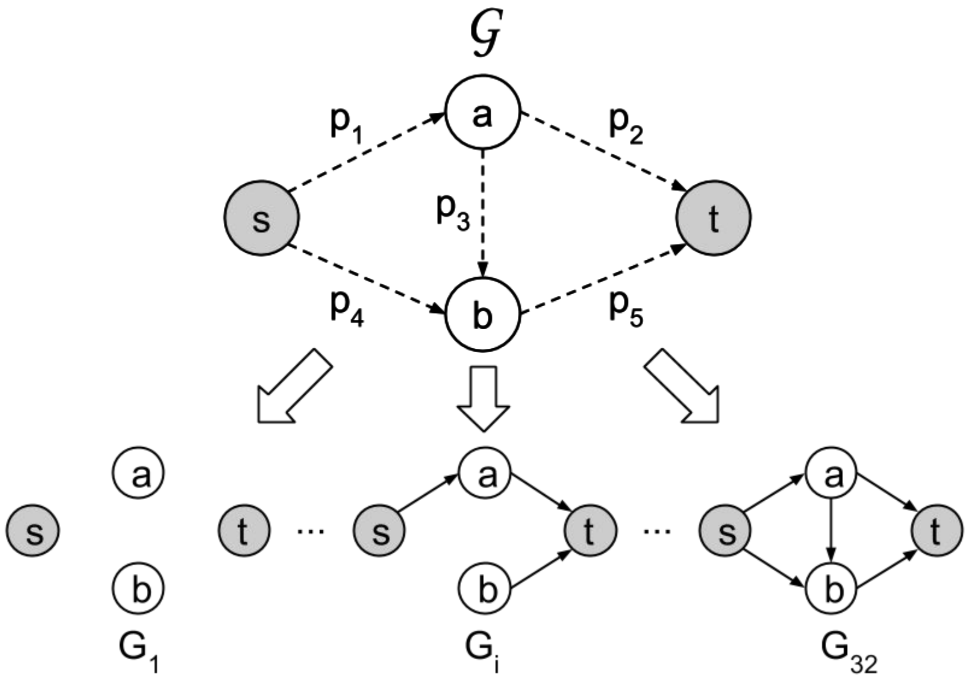
Our goals

1. To find good mathematical models that
   - Support reasoning about the risks of stepping-stone attacks against computer networks
   - In the presence of information uncertainty

2. To provide decision-support analysis tools to network defenders that are
   - Intuitive, easy to model, easy to interpret results
   - Computationally tractable

This talk: main theoretical results about **uncertain graphs**

Basics of uncertain graphs (1)



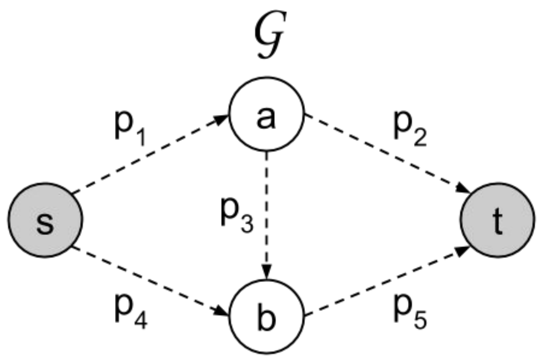
- $\mathcal{G}$ realizes into G with probability: $w_{G,\mathcal{G}} = \prod_{E_i \in E'} p_i \prod_{E_i \in E \setminus E'} (1 - p_i)$
- s reaches t in $\mathcal{G}$ with probability: $\mathcal{R}_{s,t}(\mathcal{G}) = \sum_{G \in \mathcal{G}} w_{G,\mathcal{G}} R_{s,t}(G)$

$$= p_1p_2 + p_4p_5 + p_1p_3p_5 - p_1p_2p_3p_5 - p_1p_2p_4p_5 - p_1p_3p_4p_5 + p_1p_2p_3p_4p_5$$

#P-complete

Basics of uncertain graphs (2)

- Known use of UGs:
  - Network reliability[1]
  - Protein-protein interactions[2]
  - Road networks with traffic jams[3]
  - And many others.



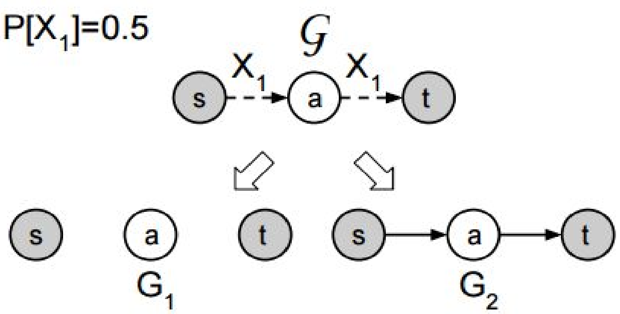
- UGs in security modeling:
  - s ~ compromised host and t ~ critical asset
  - $\{p_i\}$ ~ likelihoods that attacker can go from one host to another
  - $\mathcal{R}_{s,t}(\mathcal{G})$ ~ likelihood that attacker can reach the critical asset
  - Reachability metric gives **actionable insight** to network defenders
  -
- **Question 1:** How to capture correlation among edges in an UG?
- **Question 2:** What if we are unsure about the existence probabilities?

1. Valiant, L. G. *The Complexity of Enumeration and Reliability Problems*. SIAM Journal on Computing 8, 3 (1979)
2. Asthana, S., et al. *Predicting protein complex membership using probabilistic network reliability*. Genome Res. (2004)
3. Hua, M. et al. *Probabilistic Path Queries in Road Networks: Traffic Uncertainty Aware Path Selection*. In Proceedings of the 13th ACM International Conference on Extending Database Technology (2010)

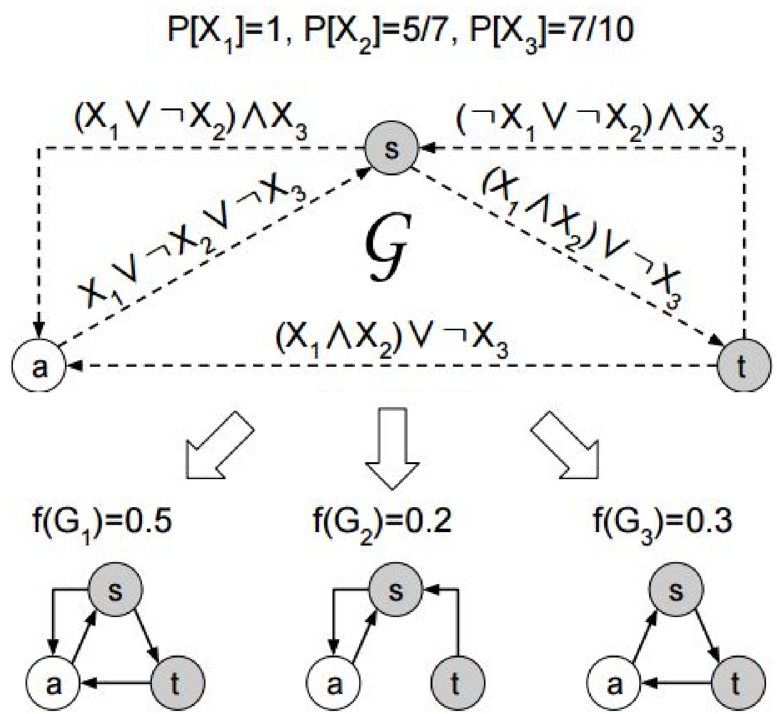
Correlation among edges

- **<u>Question 1:</u>** How to capture correlation among edges in an UG?
  - Associate edges with Boolean function of indicator random variables
  - We call them the **extended** UGs

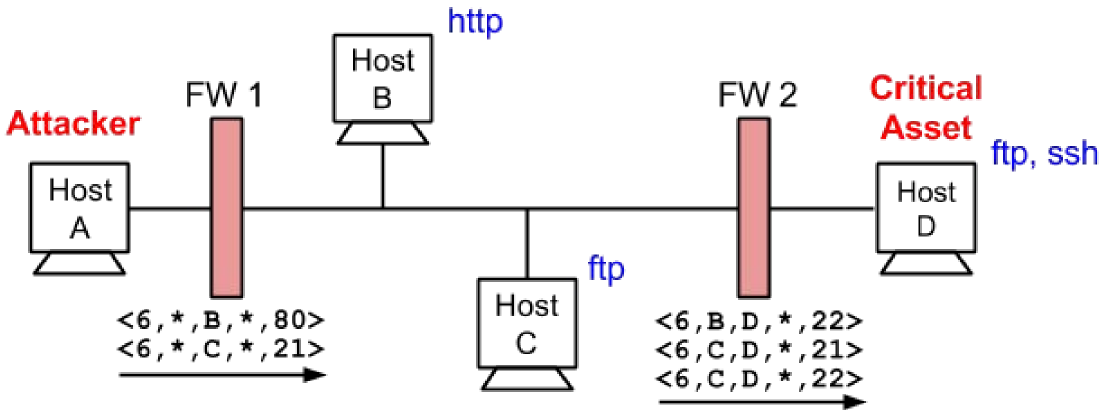
<u>Example 1:</u>



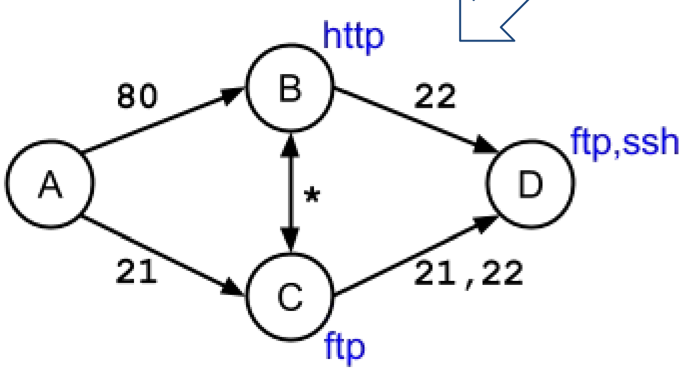
<u>Example 2:</u>



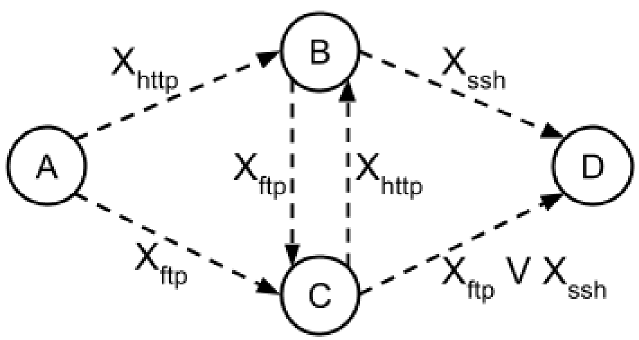
Example



A simple network



Flow graph



Extended uncertain graph

Expressiveness

1. Do we gain anything from using Boolean functions?

   Yes. Extended UGs are **more expressive** than basic UGs.

   (proof by giving an example)

2. If so, then how expressive are extended UGs?

   They can describe **any joint distribution** of edge existence probabilities.

What we mean:

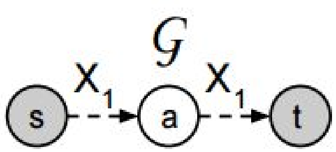
- V ~ the set of vertices; Γ_V ~ the set of directed graphs with vertex set V.
- Define a mapping f: $\Gamma_V \rightarrow R$ such that:
   a. $f(G_i) \geq 0, \ \forall G_i \in \Gamma$
   b. $\sum_{G_i \in \Gamma_V} f(G_i) = 1$
- Then every mapping f has an equivalent extended UG.

  (proof by showing an iterative construction)

Probability bounds

- **Question 2:** What if we are unsure about the existence probabilities?
  - Use bounds for input probabilities
  - The output reachability $\mathcal{R}_{s,t}(\mathcal{G})$ is also represented by a bound

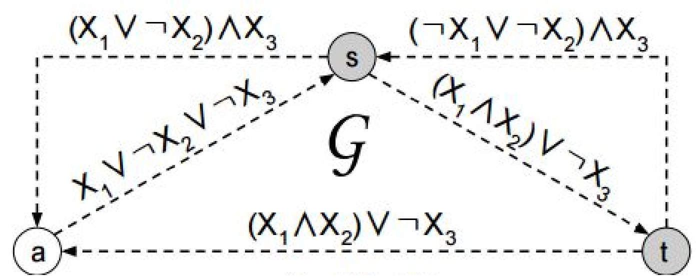
Example 1:



$P[X_1] = .5$ ➡ $\mathcal{R}_{s,t}(\mathcal{G}) = .5$

$P[X_1] \in [.7,.9]$ ➡ $\mathcal{R}_{s,t}(\mathcal{G}) \in [.7,.9]$

Example 2:



$P[X_1] = 1$, $P[X_2] = 5/7$, $P[X_3] = .7$ ➡ $\mathcal{R}_{s,t}(\mathcal{G}) = .8$ (from previous example)

$P[X_1] = P[X_2] = P[X_3] \in [.5,.8]$ ➡ $\mathcal{R}_{s,t}(\mathcal{G}) \in$ [?,?]

≡ uncertainty analysis

- **Follow-up question:** Can we compute the bound of $\mathcal{R}_{s,t}(\mathcal{G})$ **efficiently**?

➔ Yes, but have to rely on metric-specific property: **monotonicity**

Monotonicity of reachability

- Deterministic graphs:
  - Adding an edge to the graph does not decrease its reachability status (same logic for removing an edge).

- Monotone UGs:
  - Extended UG where Boolean functions assigned to edges only use AND and OR logic operators (**strict** subset of extended UG).
  - Main result for monotone UGs:
    - min input probabilities ➡ min $\mathcal{R}_{s,t}(\mathcal{G})$
    - max input probabilities ➡ max $\mathcal{R}_{s,t}(\mathcal{G})$
  - Weird situations arise when the NOT logic operator is used.

Moving forward

- UGs only model uncertainty about the networks
  - Generalized UGs can model **uncertain knowledge about attacker**
    - How hard to traverse a link?
    - What if the same vulnerability is encountered again?
  - But are difficult to analyze (ongoing research)

- Sensitivity analysis
  - Gives actionable insight to network defenders (e.g. what are the top 5 vulnerabilities to fix?)
  - Is key to the **model development process** (together with uncertainty analysis)
  - But technical details are largely unavailable

- Case studies:
  - Model large-scale and real-world systems
  - Perform scenario analysis, e.g. what if SSL is broken (again)?
  - **Defense with a fixed budget**

Conclusion

- UGs can be used to model structural uncertainty in computer networks; reachability of UGs nicely translates to a security metric.
- Traditional UGs do not model correlation among edges whereas extended UGs can; moreover, they are maximally expressive.
- Edge existence probabilities can be represented using bounds; obtaining the bound for reachability (i.e. uncertainty analysis) is easy for the class of monotone UGs.
- There are many other interesting research questions we can ask regarding generalizing and analyzing extended UGs.

Thank you!

Contact:
hnguye11@illinois.edu
palani2@illinois.edu

Find the paper at http://dl.acm.org/citation.cfm?id=3055308