Modeling Network Diversity for Evaluating the Robustness of Networks against Zero-Day Attacks

Lingyu Wang[1]

Mengyuan Zhang[1], Sushil Jajodia[2], Anoop Singhal[3], and Massimiliano Albanese[2]

[1] Concordia University, Canada
[2] George Mason University, USA
[3] National Institute of Standards and Technology, USA

ESORICS 2014

Outline

- Introduction
- Modeling Network Diversity
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

Outline

- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

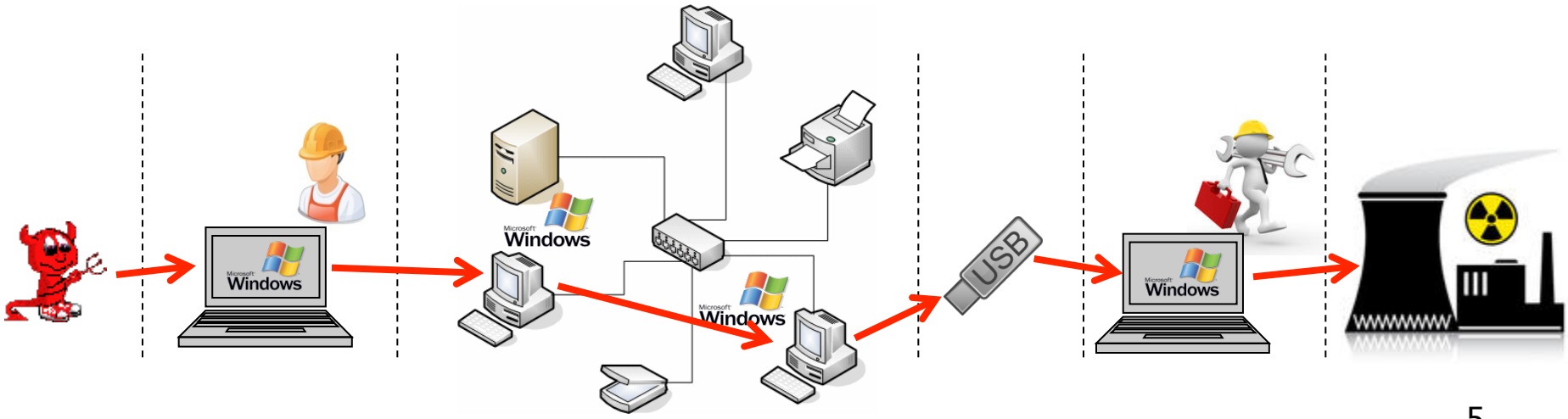
Why Worry about Zero-Day Attacks?

- A real threat to mission critical networks
- Governments and cybercriminals alike are stockpiling zero-day bugs[1]
  - The NSA spent more than $25 million a year to acquire software vulnerabilities - Edward Snowden
  - Private vendors provide at least 85 zero-day exploits on any given day of the year - Stefan Frei
  - E.g., Stuxnet exploits 4 different/complementary zero day vulnerabilities to infiltrate a SCADA network
- But what can we do about unknown attacks?

[1]http://krebsonsecurity.com/2013/12/how-many-zero-days-hit-you-today/

How Could Diversity Help?

- Stuxnet's attack strategy
  - 3rd party (e.g., contractor) → organization's network → machine with Siemens Step 7 → PLC
- The degree of software diversity along potential attack paths can be considered a good metric for the network's capability of resisting Stuxnet



Existing Work on Diversity

- Software diversity has long been regarded as a security mechanism for improving robustness
- Tolerating attacks as Byzantine faults by comparing outputs or behaviors of diverse variants
- Opportunistic or automatically generated diversity (e.g., via randomization) improves the practicality
- Many new applications for diversity
  - Moving target defense (MTD)
  - Resisting worms in sensor networks
  - Improving robustness of network routing

So Why Another Paper?

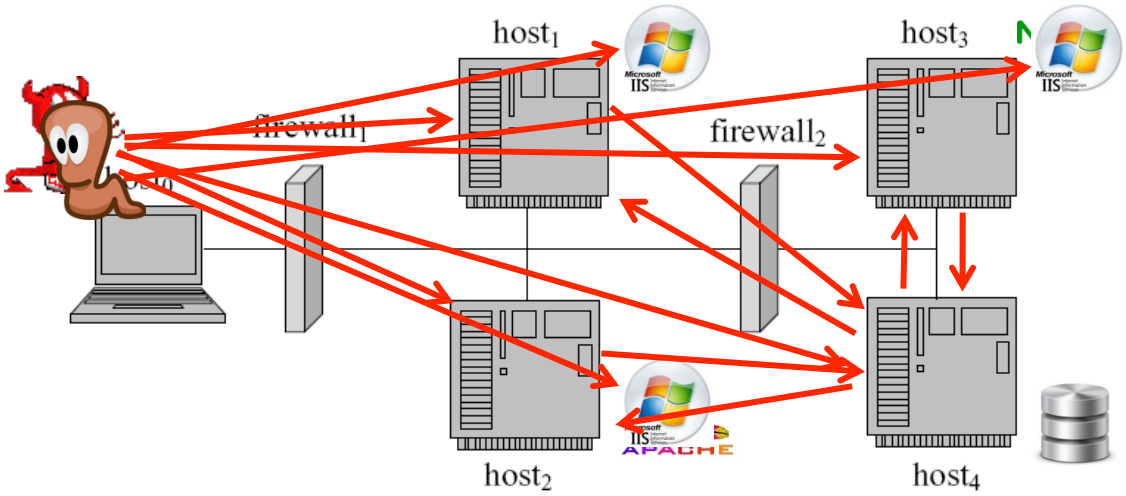
- At a higher abstraction level, as a global property of an entire network, *network diversity* and its impact on security has not been formally modeled
  - How to define the diversity metric function? count? how to count? ☾ ✶ ⦿ ■ ◆ ● ☾ ☽ ✖ ◆ ● ◈ ▣ × ⭘ □ ◈ ◘ ■ ✦
  - How to apply the metric function? on the network as a multiset of resources? what about "paths"?
  - On which path? On one or more paths?

- It depends on the use cases...

Example Use Cases

- Worm propagation
  - Count *might* be suffici
- Targeted attack (e.g.,
  - Least attack effort may
- MTD
  - The layers do not "combine" to increase attack effort

1. Intuitive notion of diversity can be misleading. A formal model is needed.
2. Different diversity metrics may be needed for different purposes.



Our Contribution

- We take the first step towards formally modeling network diversity as a security metric
  - We propose a network diversity function based on well known mathematical models of biodiversity in ecology
  - We design a network diversity metric based on the least attacking effort
  - We design a probabilistic network diversity metric to reflect the average attacking effort
  - We evaluate the metrics and algorithms through simulation
- The modeling effort helps understanding diversity and enables quantitative hardening approaches

Outline

- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

Bio-Diversity and Richness of Species

- A rich literature exists on biodiversity
  - Both theoretical studies and experiments confirm a positive relationship between biodiversity and the ecosystem's resistance to invasion and diseases
  - Many lessons may be borrowed, but we focus on metric functions and how they are applied
- *Richness* of species
  - The number of different species in an ecosystem
  - Problem: It ignores the relative abundance of each species

Effective Richness of Resources

- *Effective number*[1]
  - *Shannon-Wiener index* (Shannon entropy using natural logarithm) groups systems with same levels of diversity
  - The *effective number*, the exponential of this index, measures the number of equally-common species, even if in reality all species are not equally common.

  - We define the *effective richness of resources* as:

    (p_i is the relative frequency of resource i)

$$r(G) = \frac{1}{\prod_1^n p_i^{p_i}}$$

  - Problem: Assuming all resources are equally different

[1]M.O. Hill. Diversity and evenness: a unifying notation and its consequences. Ecology, 54(2):427–432, 1973.

Similarity-Sensitive Effective Richness

- *Similarity-Sensitive Richness*[1]
  - Given a resource similarity function *z(.): [1;m] × [1;m] → [0; 1]* (with *z(i,i) = 1*), we define the effective richness of resources as:

$$r(G) = \frac{1}{\prod_1^n zp_i^{p_i}} \qquad \left(zp_i = \sum_{j=1}^{m} z(i,j)p_j \right)$$

- We can simply talk about "the number of distinct resource types" from now on, as if all resources are equally common and equally different

[1]T. Leinster and C.A Cobbold. Measuring diversity: the importance of species similarity. Ecology, 93(3):477–489, 2012.

Outline

- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

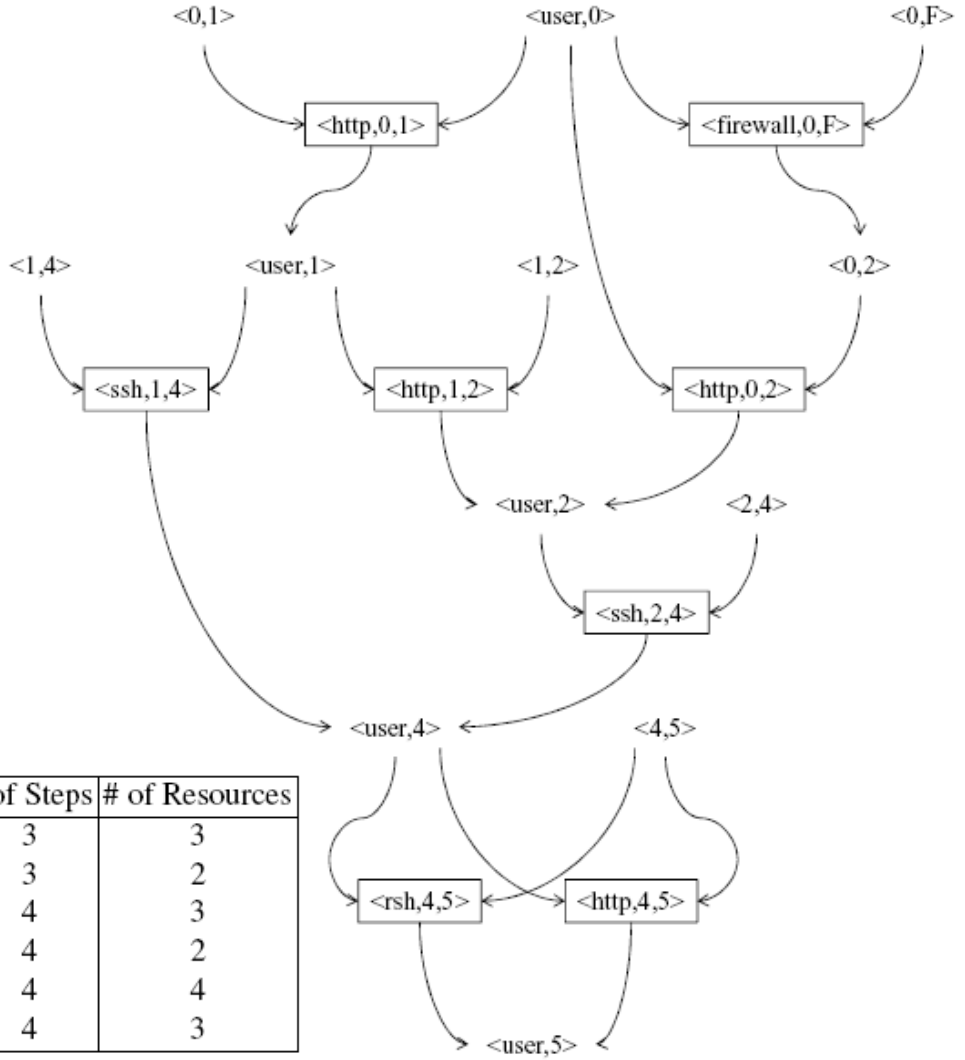
From Food Web to Resource Graph

- The second lesson from biodiversity
  - The effect of biodiversity on stability of an ecosystem critically depends on the interaction of different species inside a food web[1]
  - For food web, it is the feeding relationship (e.g., disease in one species affect those who feed on it)
  - For networks, it is the causal relationships (e.g., hacking one service may lead to accesses to others)
  - Resource graph
  - Syntactically equivalent to attack graph
  - Models causal relationships between network resources (instead of known vulnerabilities)

[1]K.S. McCann. The diversity-stability debate. Nature, 405:228–233, 2000.

Resource Graph

- Vertices: zero day exploits of resources, their pre- and post-conditions
- Edges: AND between pre-conditions, OR between exploits
- On which path should we apply the diversity metric function (i.e., the number of distinct resource types)?



Attack Path	# of Steps	# of Resources
1. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	3	3
2. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle http, 4, 5\rangle$	3	2
3. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	3
4. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	2
5. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	4
6. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	3

Selecting the Right Path(s)

- Intuitively, it should be the "shortest" path
  - 1 or 2, minimum # of steps? But 4 may take less effort than 1!
  - 2 or 4, minimum # of resources? But they both have 2 resources, so which one to choose, 2 or 4?
  - 4, minimizing (#steps/#of resources)? But what if there's a path with 9 steps and 3 resources? 1/3<2/4, but it clearly does not represent the least attack effort!

Attack Path	# of Steps	# of Resources
1. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	3	3
2. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle http, 4, 5\rangle$	3	2
3. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	3
4. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	2
5. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	4
6. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	3

Network Diversity in Least Attack Effort

- We define network diversity as:
  - (minimum # of resources on any path)/(minimum # of steps on any path)
  - Note: These may or may not be the same path! e.g., in this case: = 2 (path 2, 4) / 3 (path 1, 2)
  - The numerator 2 denotes the network's current level of diversity and the denominator 3 maximum potential (# of resources can never be greater than # of steps)

Attack Path	# of Steps	# of Resources
1. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	3	3
2. $\langle http, 0, 1\rangle \rightarrow \langle ssh, 1, 4\rangle \rightarrow \langle http, 4, 5\rangle$	3	2
3. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	3
4. $\langle http, 0, 1\rangle \rightarrow \langle http, 1, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	2
5. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle rsh, 4, 5\rangle$	4	4
6. $\langle firewall, 0, F\rangle \rightarrow \langle http, 0, 2\rangle \rightarrow \langle ssh, 2, 4\rangle \rightarrow \langle http, 4, 5\rangle$	4	3

Complexity and Heuristic Algorithm

- Determining the network diversity is NP-hard
- Heuristic algorithm
  - Only keep a limited number of local optima at each step

Procedure *Heuristic_Diversity*
Input: Resource graph $G(E \cup C, R_r \cup R_i)$, goal condition c_g, parameter k
Output: d_2
Method:

1. **For** each $e \in E$ and $c \in C \setminus C_I$
2. **Mark** e and c as unprocessed
3. **For** each $c \in C_I$
4. **Mark** c as processed
5. **Let** $\sigma(c) = \sigma'(c) = \phi$
6. **While** $(\exists e \in E)(e$ is unprocessed) and $(\forall c \in C)((c, e) \in R_r \Rightarrow c$ is processed)
7. **Let** $\{c \in C : (c, e) \in R_r\} = \{c_1, c_2, \ldots, c_n\}$
8. **Let** $\sigma(e) = ShortestK(\{q_1 \cup q_2 \cup \ldots \cup q_n \cup \{e\} : q_i \in \sigma(c_i), 1 \leq i \leq n\}, k)$
9. **Let** $\sigma'(e) = ShortestK'(\{q_1 \cup q_2 \cup \ldots \cup q_n \cup \{e\} : q_i \in \sigma(c_i), 1 \leq i \leq n\}, k)$
10. **Mark** e as processed
11. **For** each c s.t. $(e, c) \in R_i$
12. **If** $(\forall e' \in E)((e', c) \in R_i \Rightarrow e'$ is processed) **Then**
13. **Let** $\sigma(c) = ShortestK(\bigcup_{e' \text{ s.t. } (e',c) \in R_i} \sigma(e'), k)$
14. **Let** $\sigma'(c) = ShortestK'(\bigcup_{e' \text{ s.t. } (e',c) \in R_i} \sigma(e'), k)$
15. **Mark** c as processed
16. **Return** $\frac{min_{q \in seq(c_g)} |R(q)|}{min_{q' \in seq(c_g)} |q'|}$

Outline

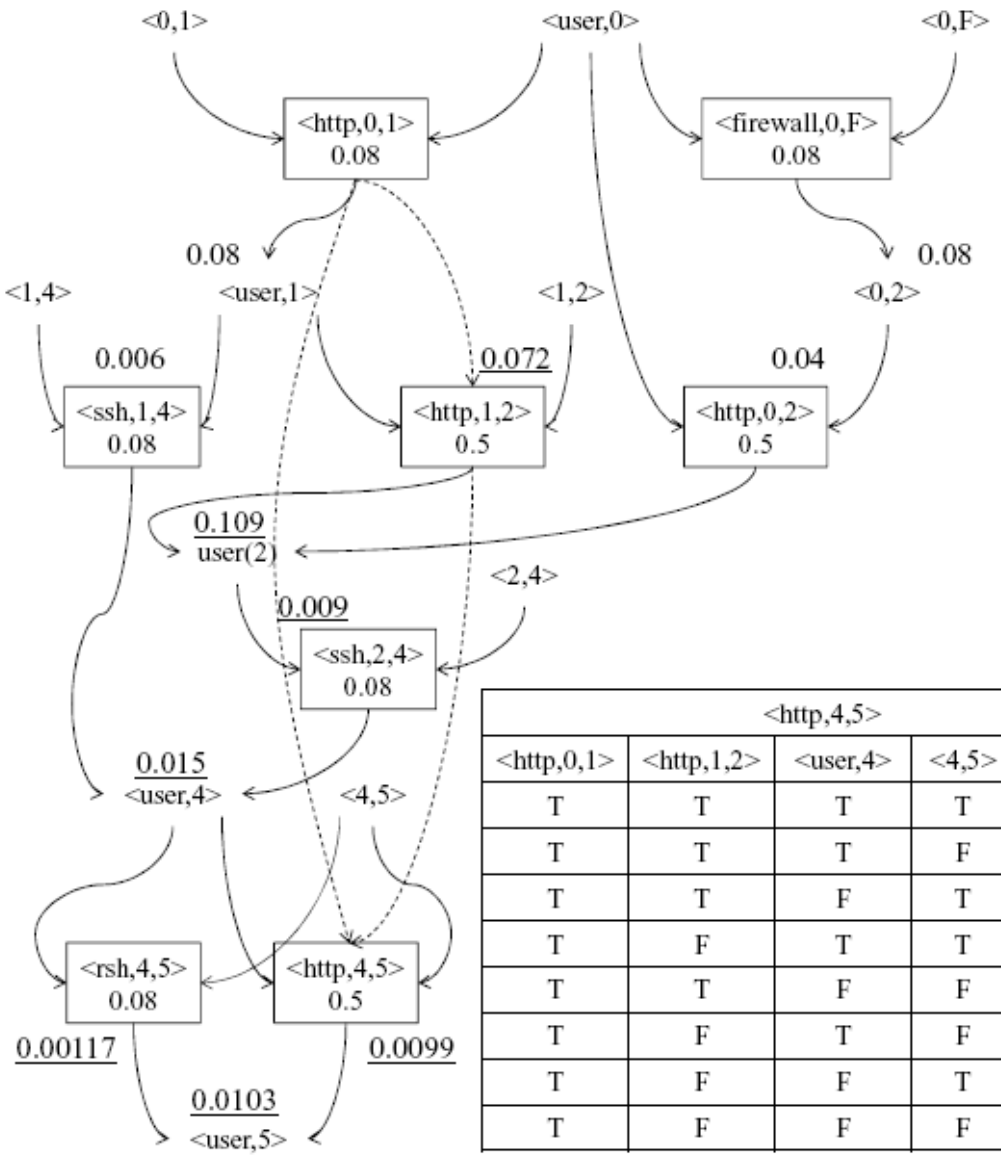
- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

Network Diversity in Average Effort

- The least attacking effort-based metric only provides a partial picture of the threat
- We now define a probabilistic network diversity metric based on the average attacking effort
- Informally, as p_1/p_2, where
  - p_1 is the probability an attacker can compromise the given asset now, and
  - p_2 the probability he/she can still compromise it if all the resources were to be made different (i.e., every resource type would appear at most once)

An Example

- Number in smaller font: probability of an exploit given all pre-conditions are true
- Number in larger font: probability of reaching this node
- Dotted lines/ underlined numbers: given probabilities of reusing an exploit
- Network diversity = *0.007/0.0103*



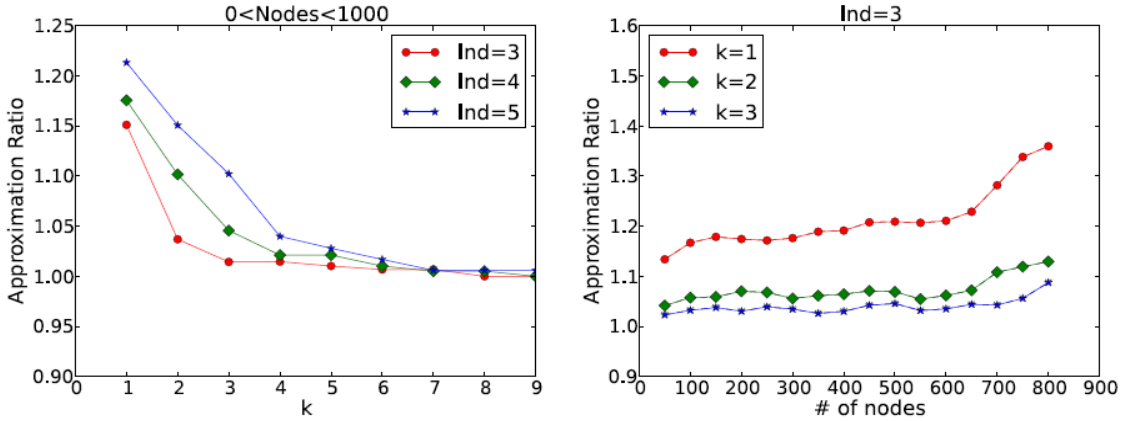
<http,4,5>					
<http,0,1>	<http,1,2>	<user,4>	<4,5>	T	F
T	T	T	T	0.9	0.1
T	T	T	F	0	1
T	T	F	T	0	1
T	F	T	T	0.9	0.1
T	T	F	F	0	1
T	F	T	F	0	1
T	F	F	T	0	1
T	F	F	F	0	1

Outline

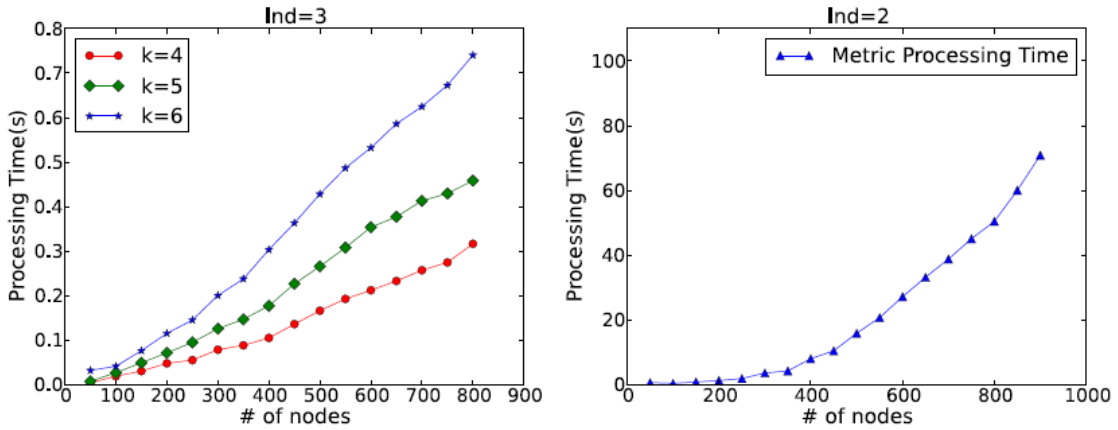
- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

Simulation Results

- Accuracy/performance of the heuristic algorithm



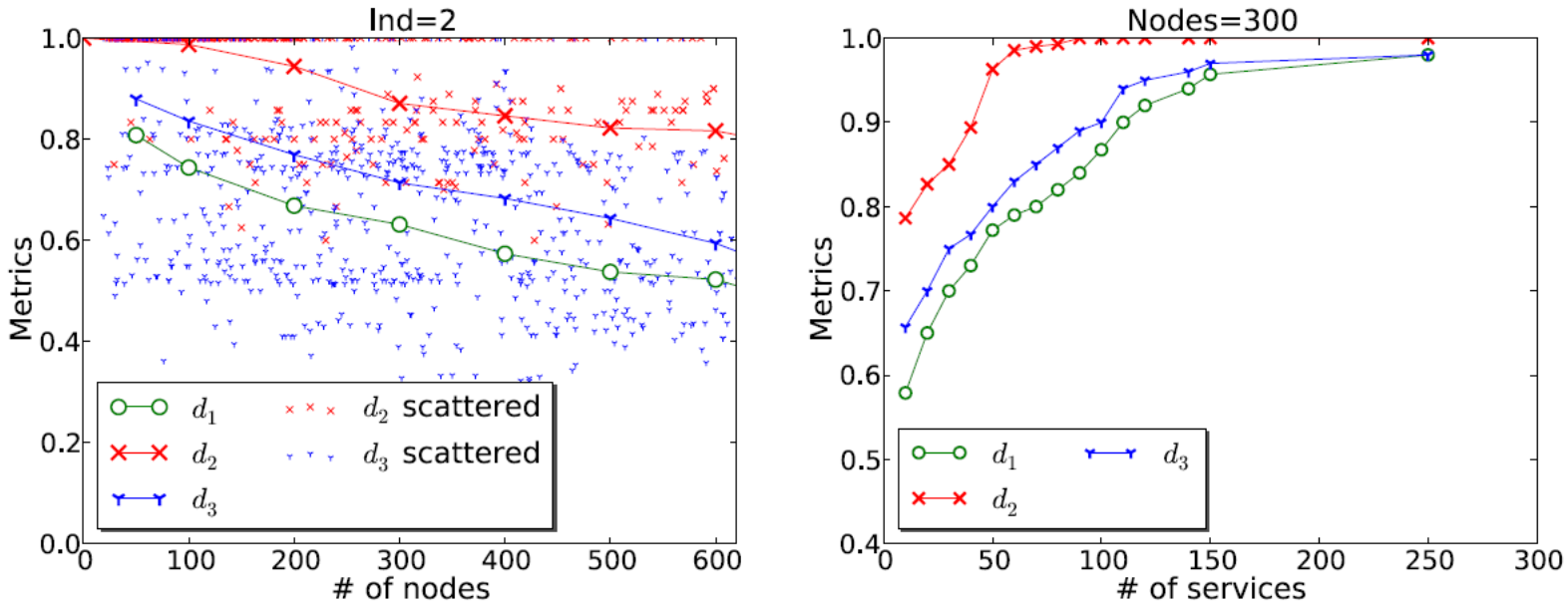
Approximation Ratio in k under Different In-degrees (Left) and in Graph Size under Different k (Right)



Processing Time for Computing d_2 in Graph Size under Different k (Left) and Processing Time for Computing d_3

Simulation Results Cont'd

- Comparison of the metrics



Comparison of Metrics (Left) and the Effect of Increasing Diversity (Right)

Outline

- Introduction
- Network Diversity Metrics
  - Biodiversity-Inspired Metric
  - Least Attacking Effort-Based Metric
  - Probabilistic Metric
- Simulation
- Conclusion

Conclusion

- We have taken a first step towards modeling network diversity, by proposing
  - a biodiversity-inspired metric
  - a least attacking effort-based metric
  - a probabilistic metric
- Limitations and future work
  - Depending on the availability and accuracy of inputs, e.g., resources, their relationships, and similarity
  - Simulations are based on random inputs
  - Not considering the cost and impact of diversity
  - Not considering the likelihood of attacks on resources
  - Future work will also apply other biodiversity results

Q & A

Thank You!

Contact Author: Lingyu Wang (wang@ciise.concordia.ca)