陽明交大
NYCU

Inferencing Cyber Attack Causal Relationship Using Cyber Threat Intelligence

Student:
Chih-Chien Cheng (Anson)

Advisor:
Professor Ying-Dar Lin

High Speed Network Lab
National Yang Ming Chiao Tung University, Taiwan

Outline

- Motivation
- Background
- Issues
  - Automated inference causal relationship with CTI
- Problem statements
  - Node expansion
  - Status evaluation
  - Result analysis
- Related work
  - Generate and use CTI
- Solution approach
  - Evaluate the current status
  - infer causal relationship

Motivation

- CTI
  - knowledge, skills and experience-based information
  - help mitigate potential attacks and harmful events

- Forensics with CTI
  - manual, time consuming
  - large amount of data

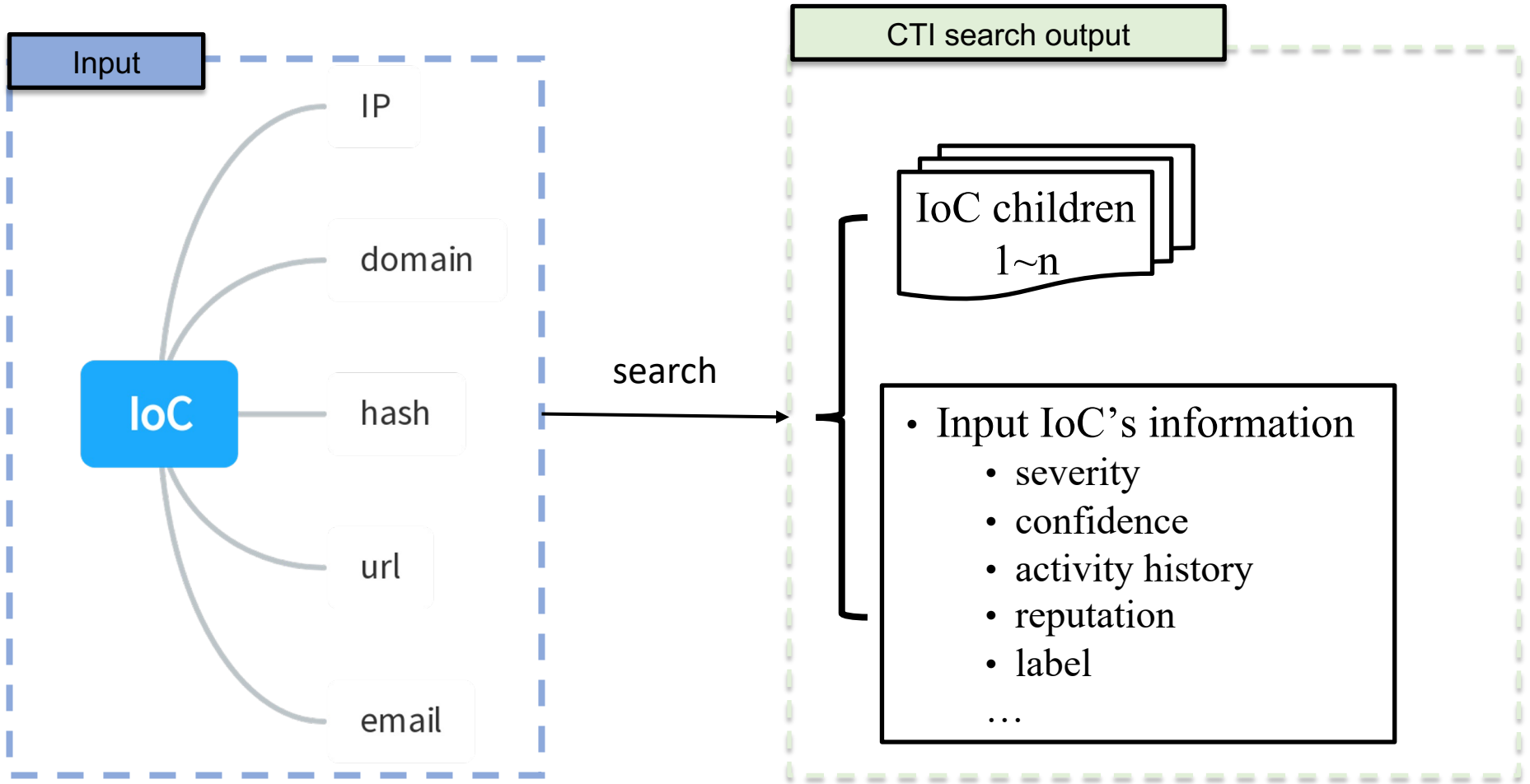
- Automatic use of CTI
  - Search, analyze and infer
  - Automation produces results

Background – CTI

- Has become an important issue for organizations[1]
- There are three overarching
  - Tactical: technical intelligence (IoC) which can be used to identify threat actors
  - Operational: details of the threat actors, including their tools, techniques and procedures
  - Strategic: intelligence about the overarching risks associated with cyber threats

- CTI application scenarios
  - Forensics and inference using CTI[5]

Background — CTI Unit

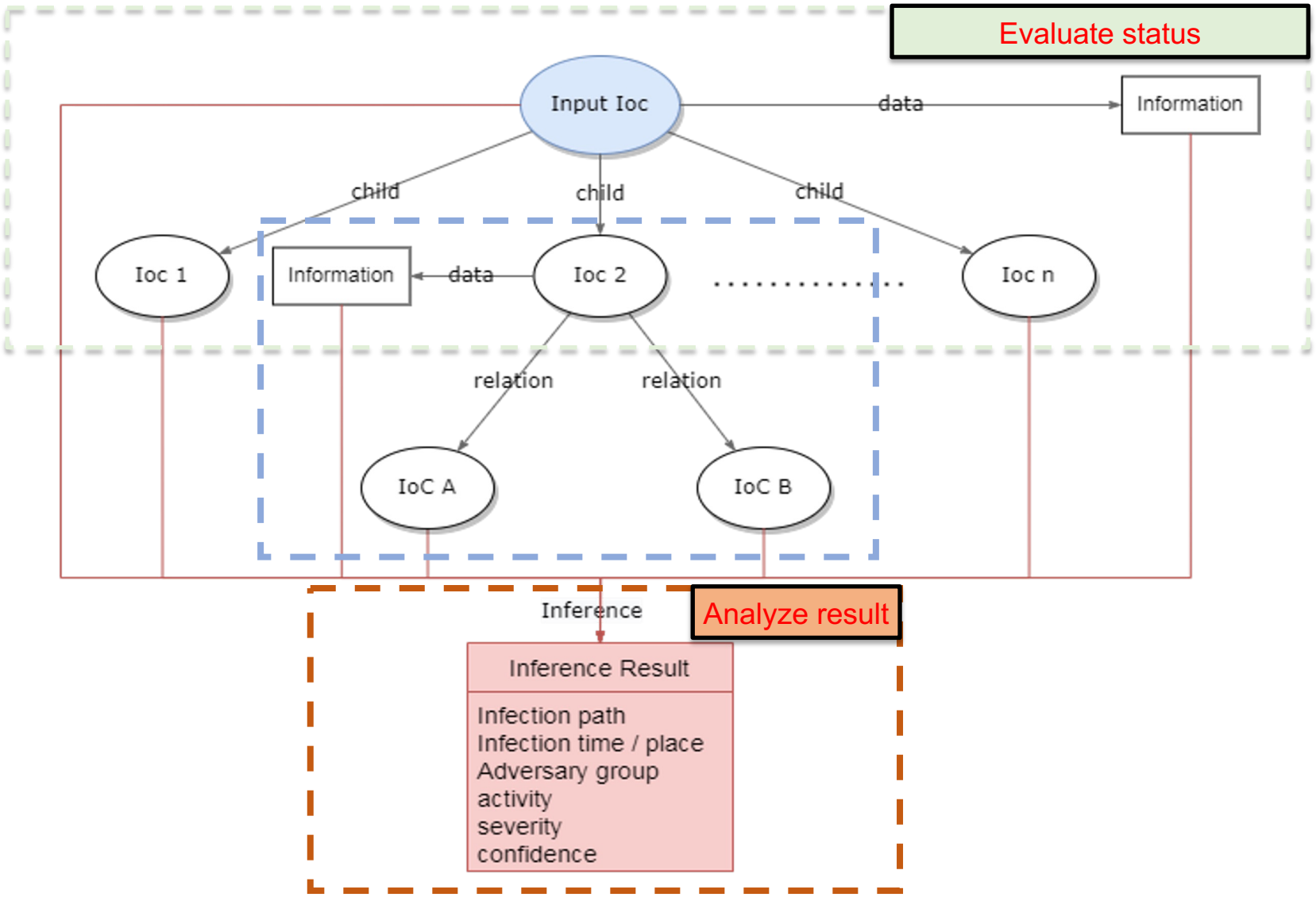


Issues – Automated inference causal relationship with CTI

- Basic vs. Expand steps
- Basic vs. Analyze result
- Basic vs. Inference
- Degree of automation
- Inference result

Approach	manual vs. automatic expand node	automatic analysis	manual vs. semi-automatic
basic	manual	No	manual
Expand steps	automatic	No	semi-automatic
Analyze result	manual	Yes	semi-automatic
Inference	automatic	Yes	automatic

Problem – Overview



Subproblem 0 – Node Expansion (development track)

- Subproblem
  - Input
    - target IoC to expand
  - Output
    - this node's children (other expandable IoCs)
    - information about target IoC
  - Objective
    - explore the profile of the target node
    - maintain all queried data on dynamic table
  - Constraint
    - None

Subproblem 1 – Status Evaluation

- Subproblem
  - Input
    - all queried data so far (current status)
  - Output
    - Determines the next node to be expanded
  - Objective
    - Analyze the results with the least number of expands
  - Constraint
    - None

Subproblem 2 – Result Analysis

- Subproblem
  - Input
    - all queried data so far (current status)
  - Output
    - Infection path
    - Infection time / location
    - Adversary group
    - Activity
    - Severity
    - Confidence
  - Objective
    - get inferences from the queried data
    - automate the inference process
  - Constraint
    - None

Related work – Generate and use CTI

paper	Input	Output	Objective			Method
	Structured CTI		Extract Information	Automation	efficiency	
[6]	X	threat action & TTPs	O	O	X	· context aware analytics · NLP · IR (Information Retrieval)
[7]	O	· standard CTI · detected threat	X	O	X	· automated threat detection tools · antivirus software
[8]	X	analyze result	O	X	X	· Open-CyKG : CTI KG
[9]	O	Maximum Entropy Model	X	O	X	· automatically label text · by leveraging related, domain-specific, structured data
[10]	X	automation CTI service platform	X	O	O	machine learning-based integrated framework
Ours	O	· causal relationship · confidence score	O	O	O	· search contacted IoCs · Infer from all queried data

Solution Approach

There are two **sections** in this solution:

- Evaluate the current status
  - Find the next node to expand
    - more than two input IoCs
    - AI solution
  - infer causal relationship

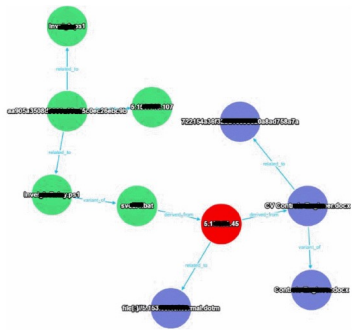
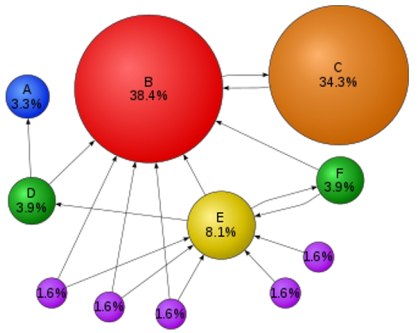
Sub-Solution 1 – Status Evaluation Method

- Find the next node to expand
  - more than two input IoCs
    - bidirectional depth limit search
  - AI solution
    - Use the expanded data as dataset
    - Train the model to output predicted high-value nodes

Algorithm	AI	Off-line learning	On-line learning
Calculation	-	-	-
Off-line learning	V	V	-
On-line learning	V	-	V
Off-line + On-line learning	V	V	V

Sub-Solution 2 – Result Analysis Method

- Infer causal relationship
  - Find the intersection of attack types
  - PageRank calculation
  - Betweenness centrality calculation
- The above method is currently used to find important nodes in graph.

Algorithm	intersection of attack types	PageRank	Betweenness
Description			$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$

Reference

[1] 2021 SANS Cyber Threat Intelligence (CTI) Survey Rebekah Brown and Robert M. Lee, SANS Institute 2021
https://assets.contentstack.io/v3/assets/blt36c2e63521272fdc/blt43a990b140efaa96/6112a525f0c97e39497dc96d/40080.pdf

[2] TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data, Jun Zhao, Qiben Yan, Jianxin Li, Minglai Shao, Zuti He, Bo Li, Computers & Security Volume 95, August 2020, 101867

https://www.ciencedirect.com/science/article/pii/S0167404820301395

[3] A Framework for Cyber Threat Intelligence Extraction from Raw Log Data, Max Landauer; Florian Skopik; Markus Wurzenberger; Wolfgang Hotwagner; Andreas Rauber, 2019 IEEE International Conference on Big Data (Big Data)

https://ieeexplore.ieee.org/abstract/document/9006328

[4] A Supervised Machine Learning Based Approach for Automatically Extracting High-Level Threat Intelligence from Unstructured Sources, Yumna Ghazi; Zahid Anwar; Rafia Mumtaz; Shahzad Saleem; Ali Tahir, 2018 International Conference on Frontiers of Information Technology (FIT)

https://ieeexplore.ieee.org/abstract/document/8616979

Reference

[5] Ontology-based Cyber Risk Monitoring Using Cyber Threat Intelligence, Yazid Merah, Tayeb Kenaza, ARES 2021: The 16th International Conference on Availability, Reliability and Security

https://dl.acm.org/doi/abs/10.1145/3465481.3470024

[6] TTPDrill: Automatic and Accurate Extraction of Threat Actions from Unstructured Text of CTI Sources, Ghaith Husari, Ehab Al-Shaer, Mohiuddin Ahmed, Bill Chu, Xi Niu, ACSAC 2017: Proceedings of the 33rd Annual Computer Security Applications ConferenceDecember 2017 Pages 103–115

[7] Toward Automated Cyber Defense with Secure Sharing of Structured Cyber Threat Intelligence, Md. Farhan Haque & Ram Krishnan , Information Systems Frontiers volume 23, pages883–896 (2021)

[8] Open-CyKG: An Open Cyber Threat Intelligence Knowledge Graph, Injy Sarhanab, Marco Spruitbcd, Knowledge-Based Systems, Volume 233, 5 December 2021, 107524

Reference

[9] Automatic Labeling for Entity Extraction in Cyber Security, Bridges, Robert A; Jones, Corinne L; Iannacone, Michael D; Testa, Kelly M; Goodall, John R, Conference: 2014 ASE International Conference on Cyber Security, Stanford, CA, USA, 20140527, 20140331

[10] inTIME: A Machine Learning-Based Framework for Gathering and Leveraging Web Data to Cyber-Threat Intelligence, Paris Koloveas, Sofia Alevizopoulou, Christos

[11] Enabling Efficient Cyber Threat Hunting With Cyber Threat Intelligence, Peng Gao; Fei Shao; Xiaoyuan Liu; Xusheng Xiao; Zheng Qin; Fengyuan Xu, 2021 IEEE 37th International Conference on Data Engineering (ICDE)

[12] Generating Fake Cyber Threat Intelligence Using Transformer-Based Models, Priyanka Ranade; Aritran Piplai; Sudip Mittal; Anupam Joshi, 2021 International Joint

陽明交大
NYCU

High Speed Network Lab
National Yang-Ming Chiao-Tung University, Taiwan