Chain-of-Event: Interpretable Root Cause Analysis for Microservices through Automatically Learning Weighted Event Causal Graph

Zhenhe Yao[1], Changhua Pei[2], Wenxiao Chen, Hanzhang Wang, Liangfei Su, Huai Jiang, Zhe Xie, Xiaohui Nie, Dan Pei

1. Presenter. Email: yaozh20@mails.tsinghua.edu.cn
2. Corresponding Author

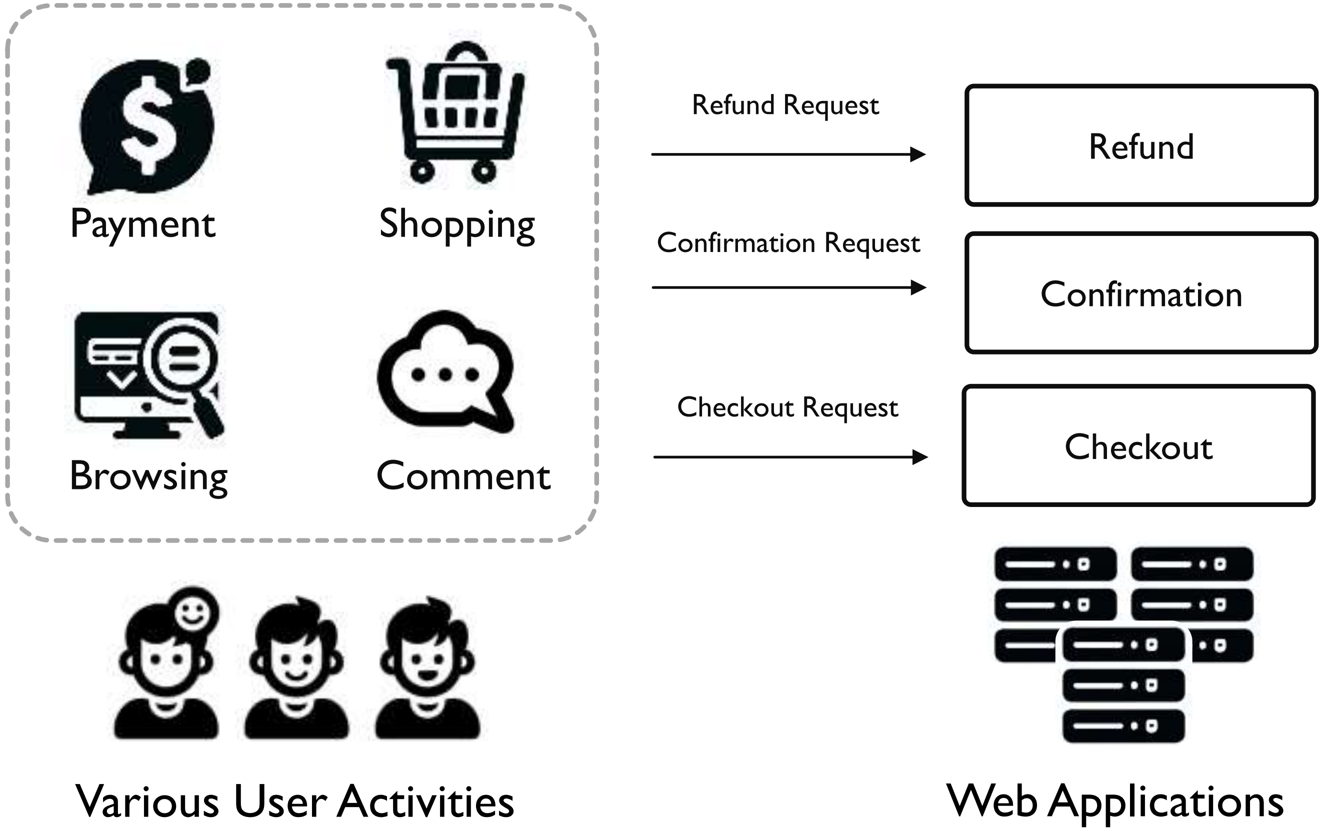
Outline

- Background
- Design
- Evaluation
- Conclusion

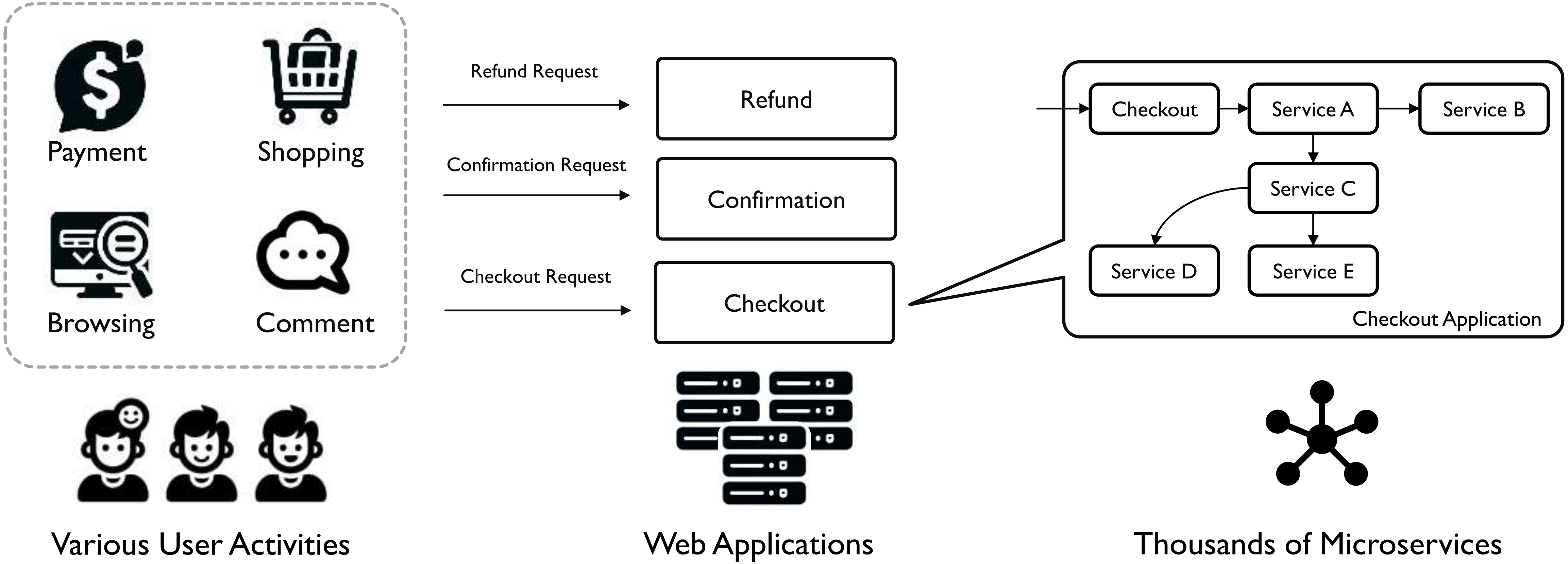
Microservice Architecture

Various User Activities

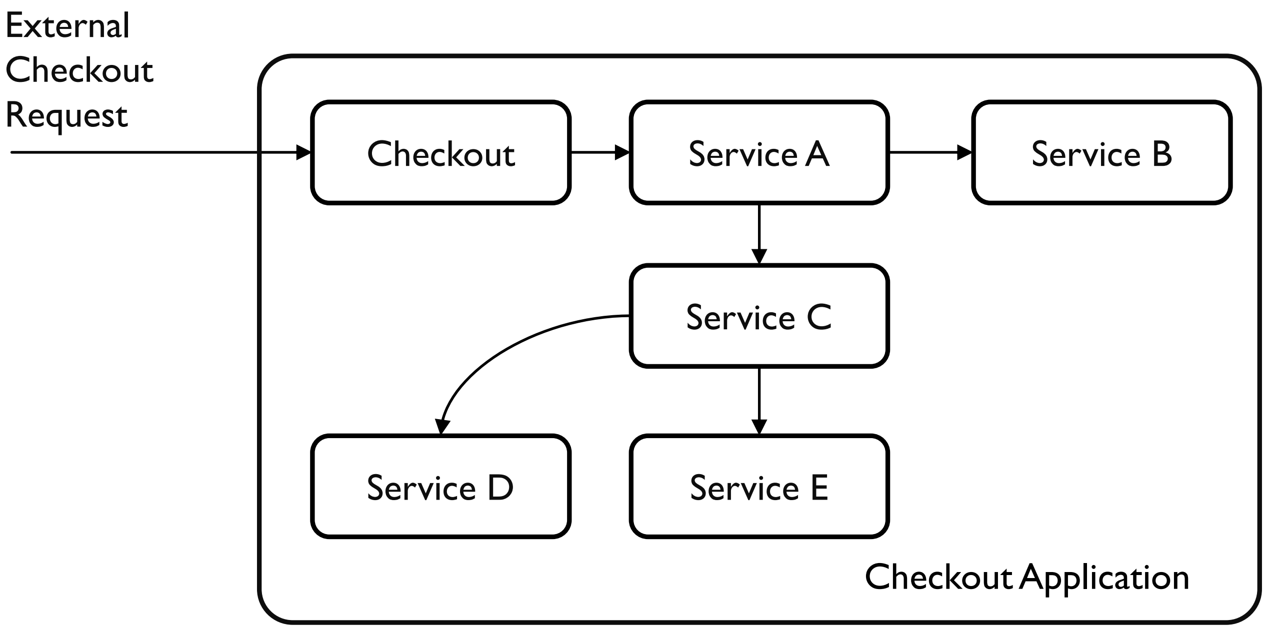
Microservice Architecture



Microservice Architecture



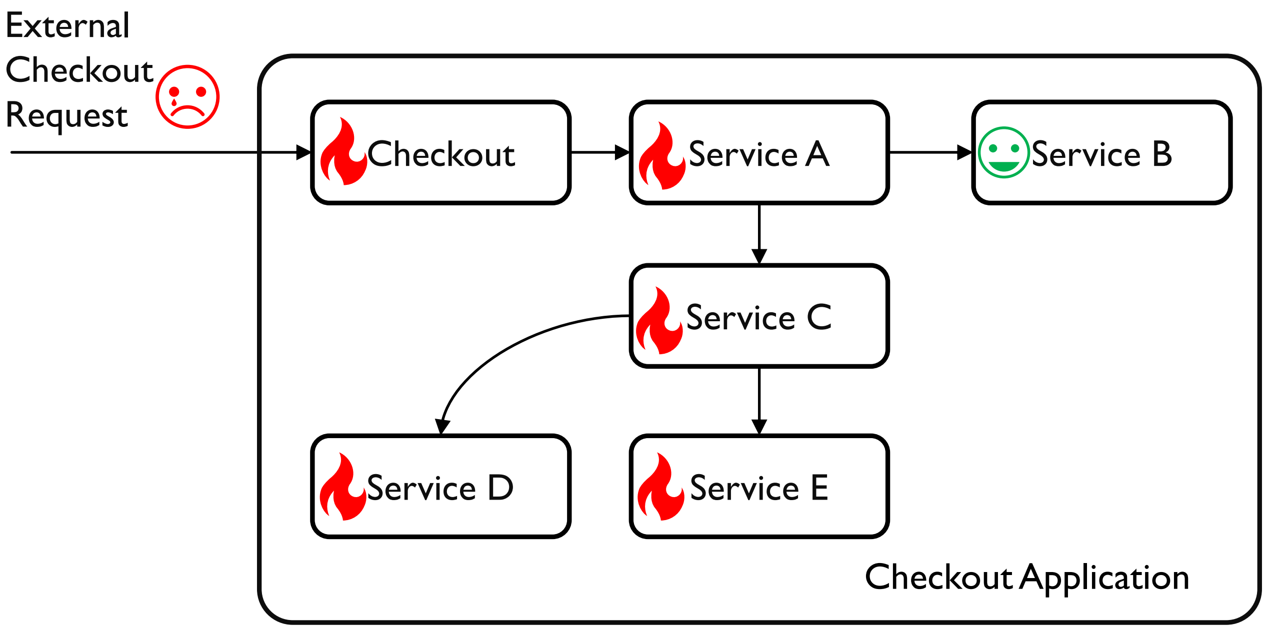
Microservice Reliability Maintenance



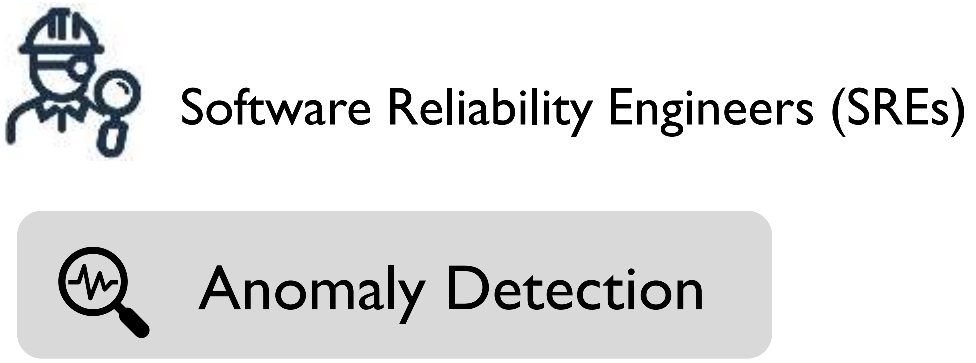
Checkout-related Microservices

Software Reliability Engineers (SREs)

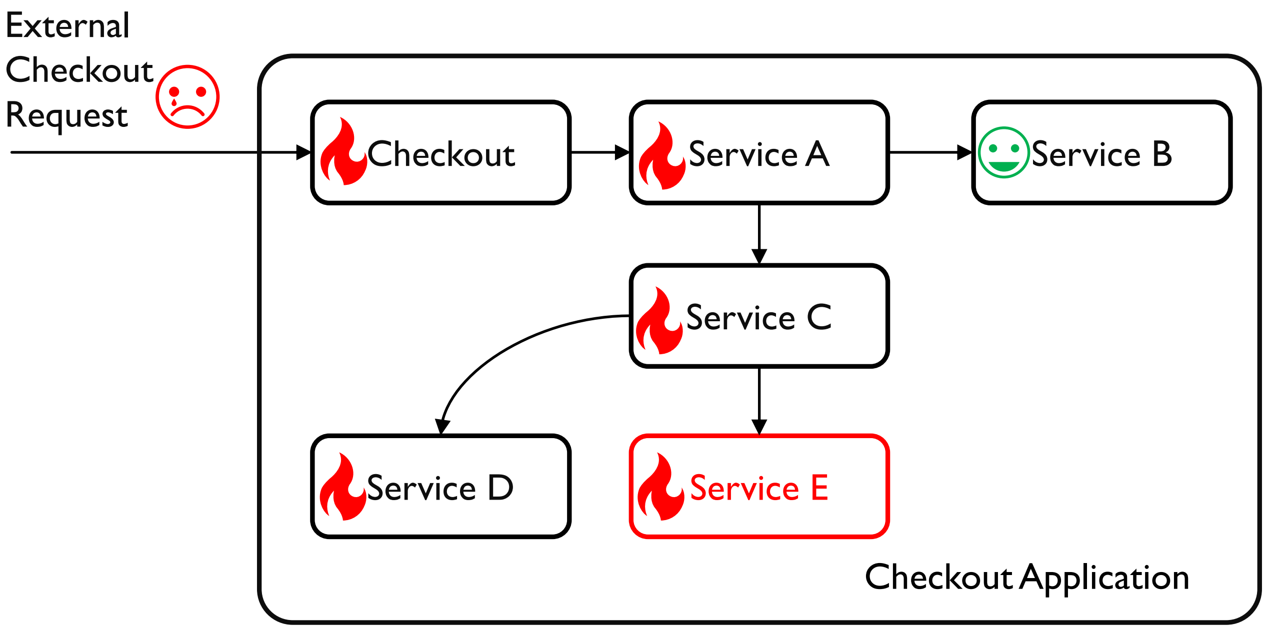
Microservice Reliability Maintenance



Checkout-related Microservices



Microservice Reliability Maintenance

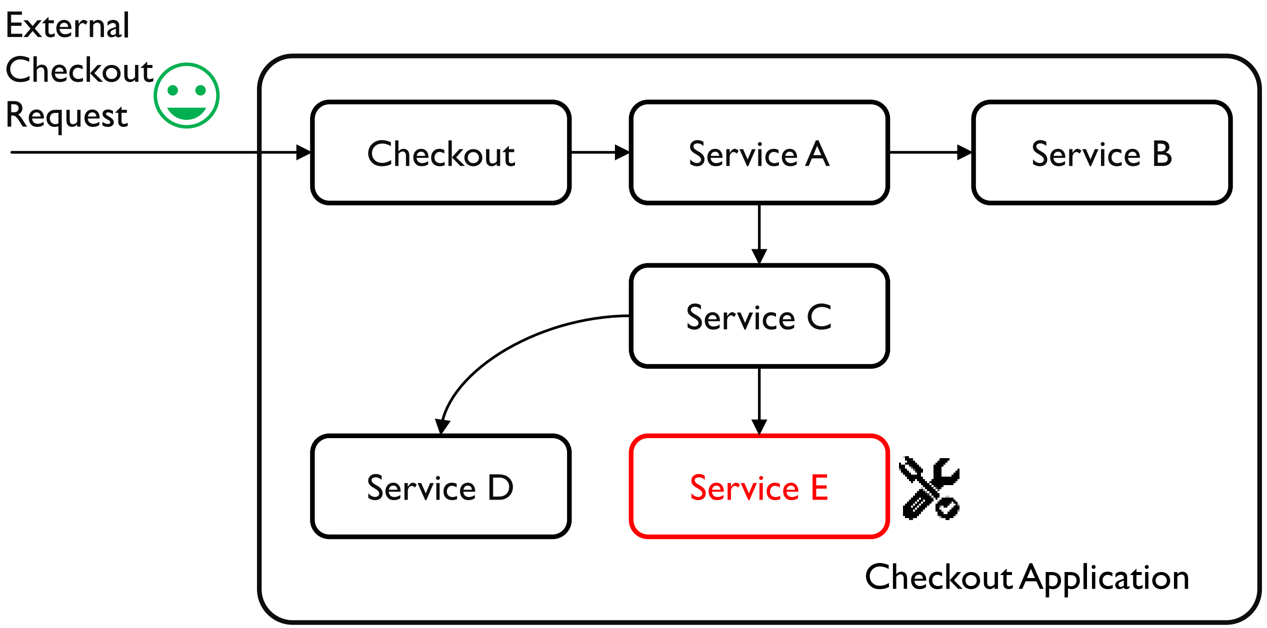


Checkout-related Microservices

Software Reliability Engineers (SREs)

- Anomaly Detection
- Root Cause Analysis

Microservice Reliability Maintenance

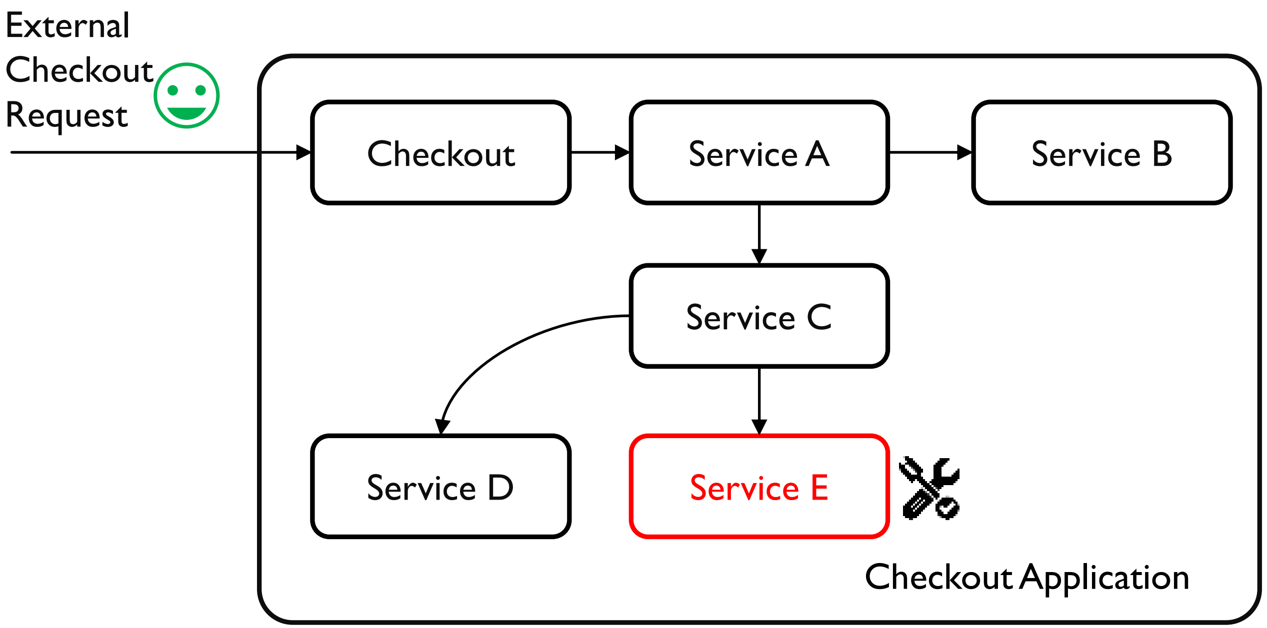


Checkout-related Microservices

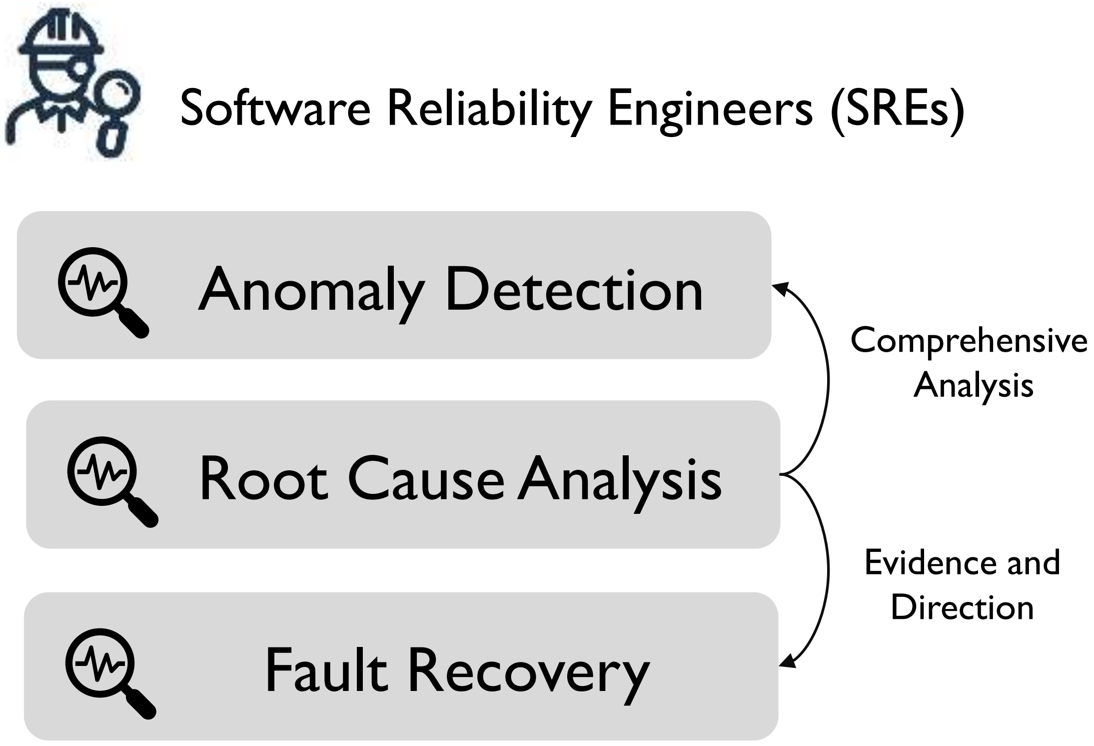
Software Reliability Engineers (SREs)

- Anomaly Detection
- Root Cause Analysis
- Fault Recovery

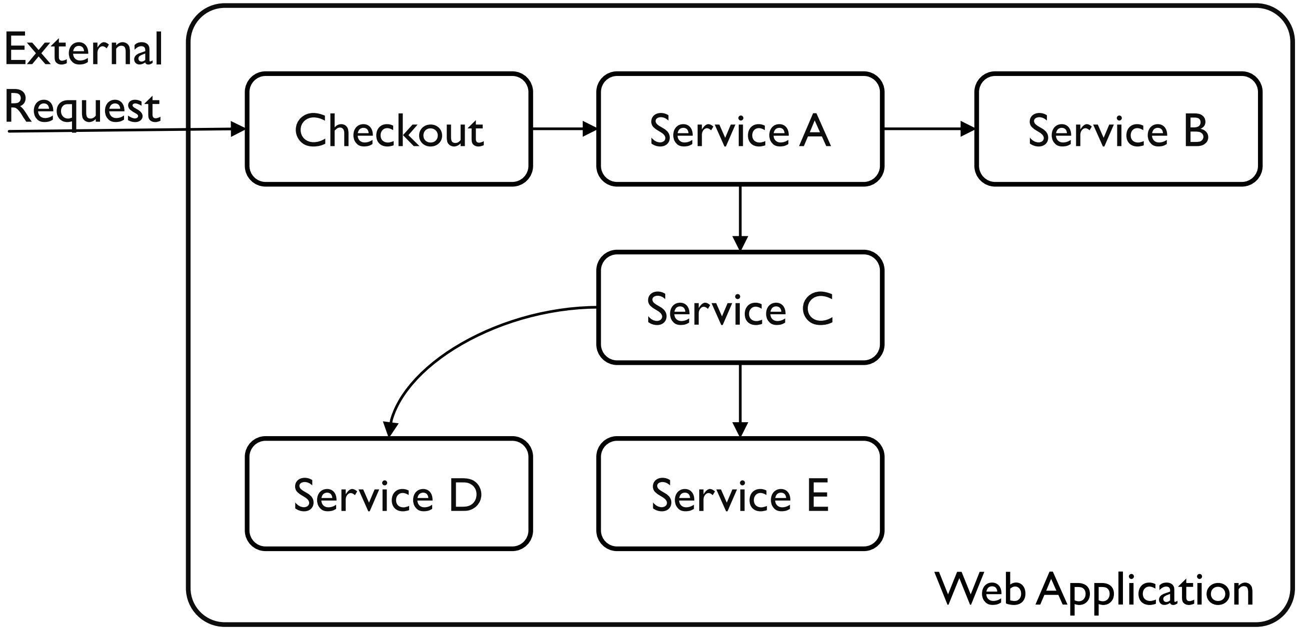
Microservice Reliability Maintenance



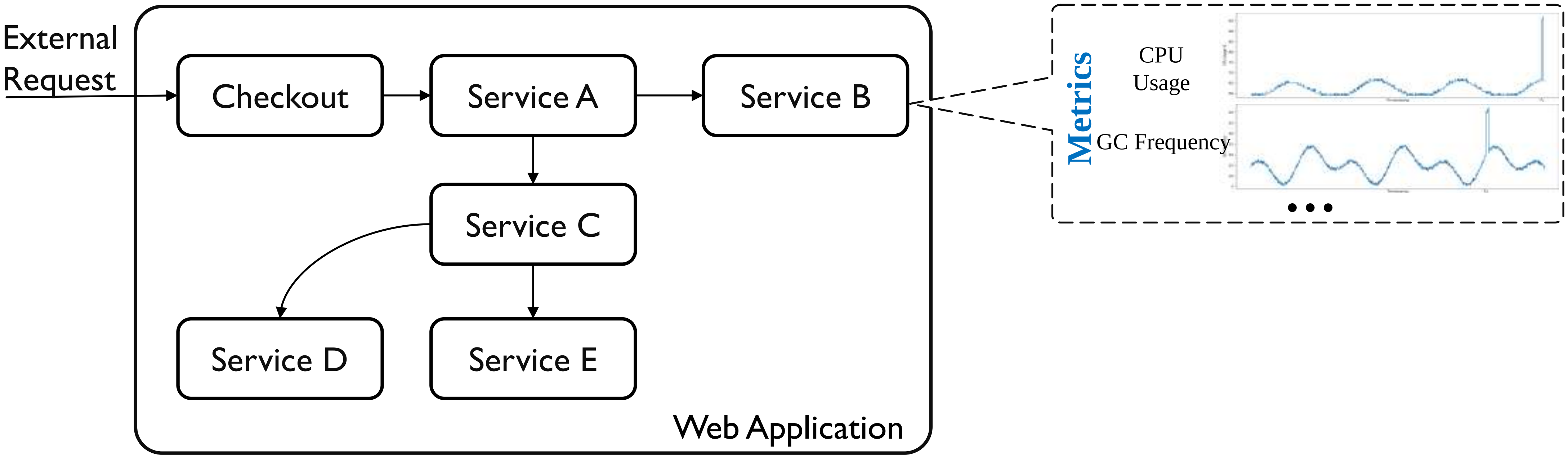
Checkout-related Microservices



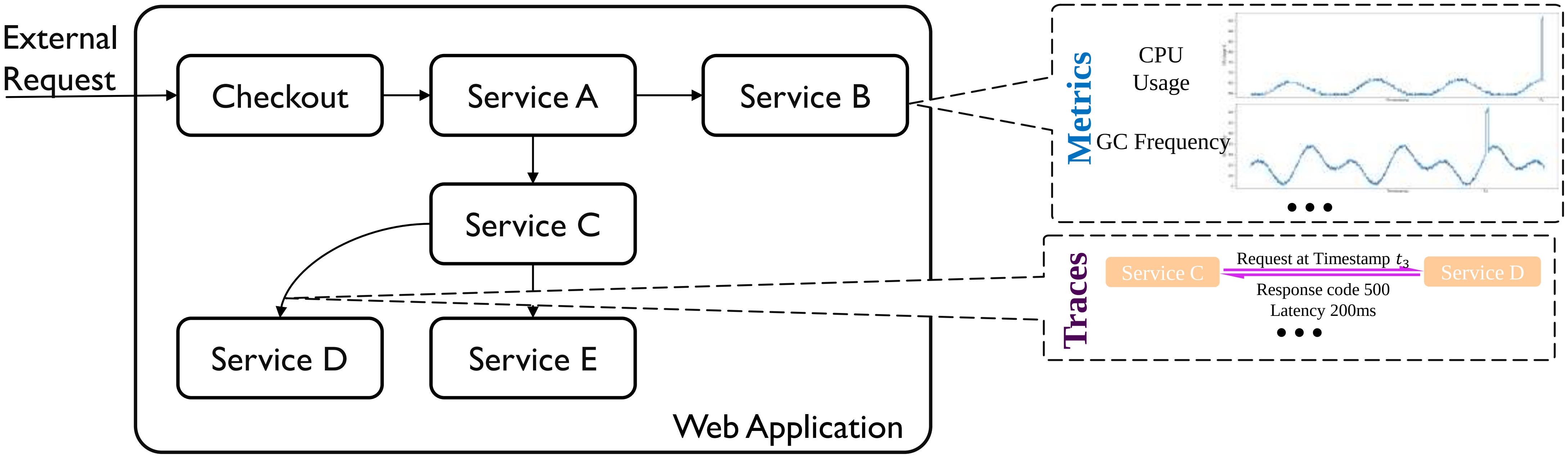
Multi-modal Monitoring Data in Microservice Architecture



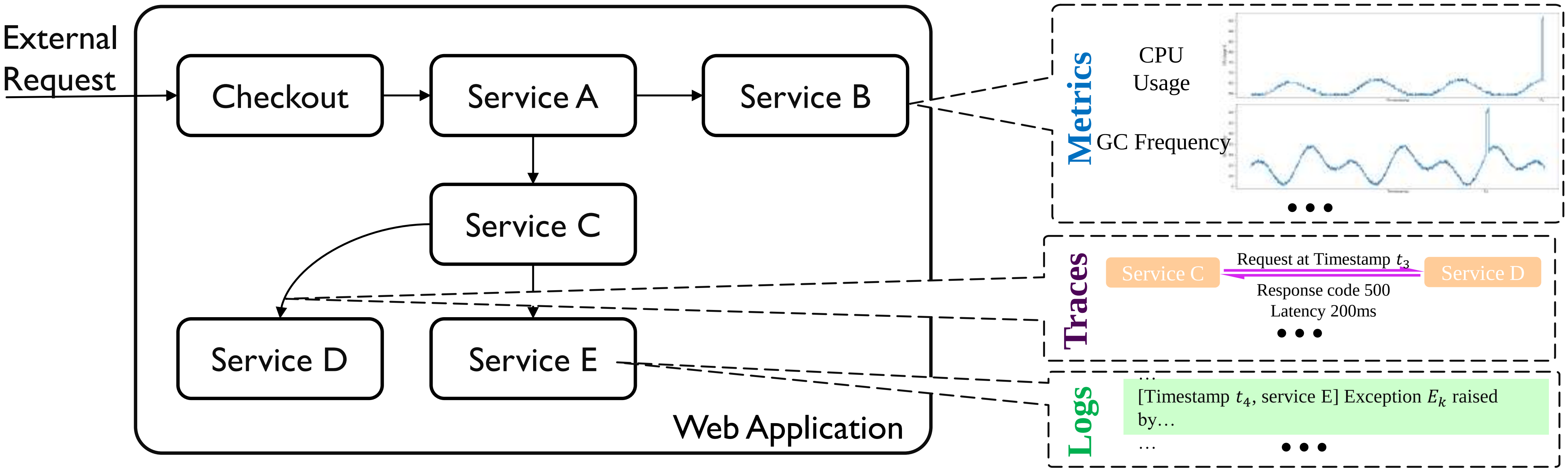
Multi-modal Monitoring Data in Microservice Architecture



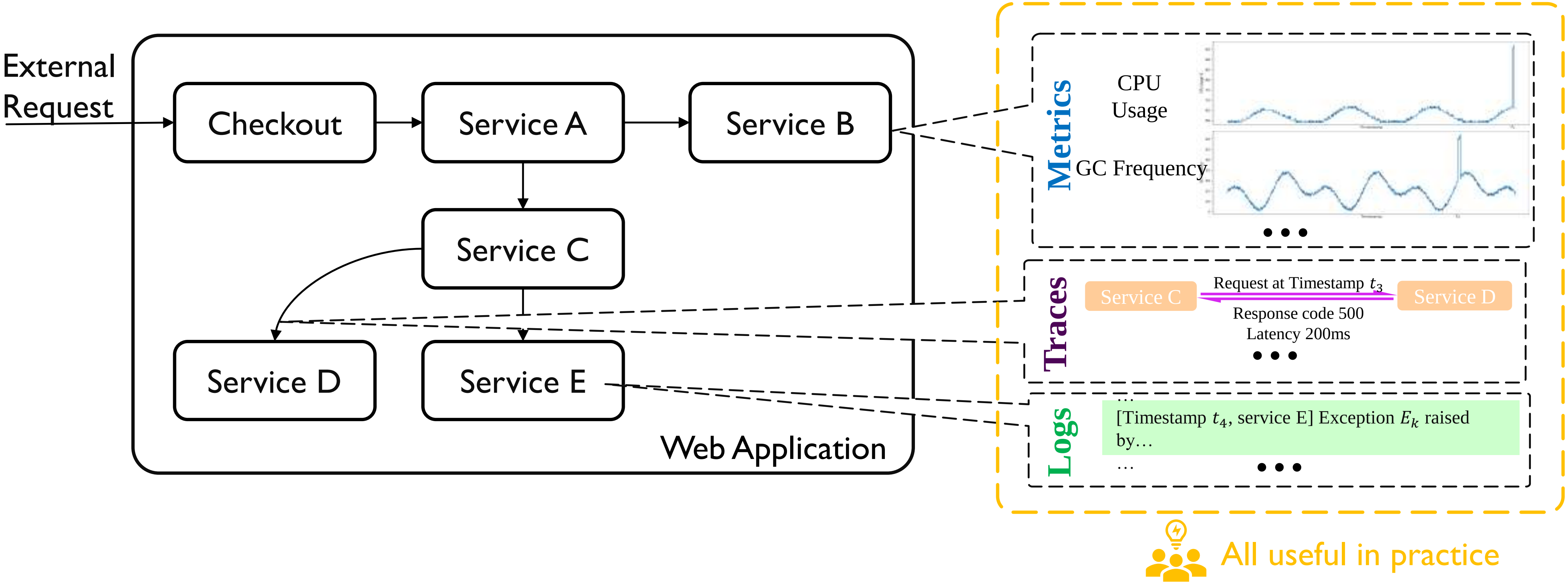
Multi-modal Monitoring Data in Microservice Architecture



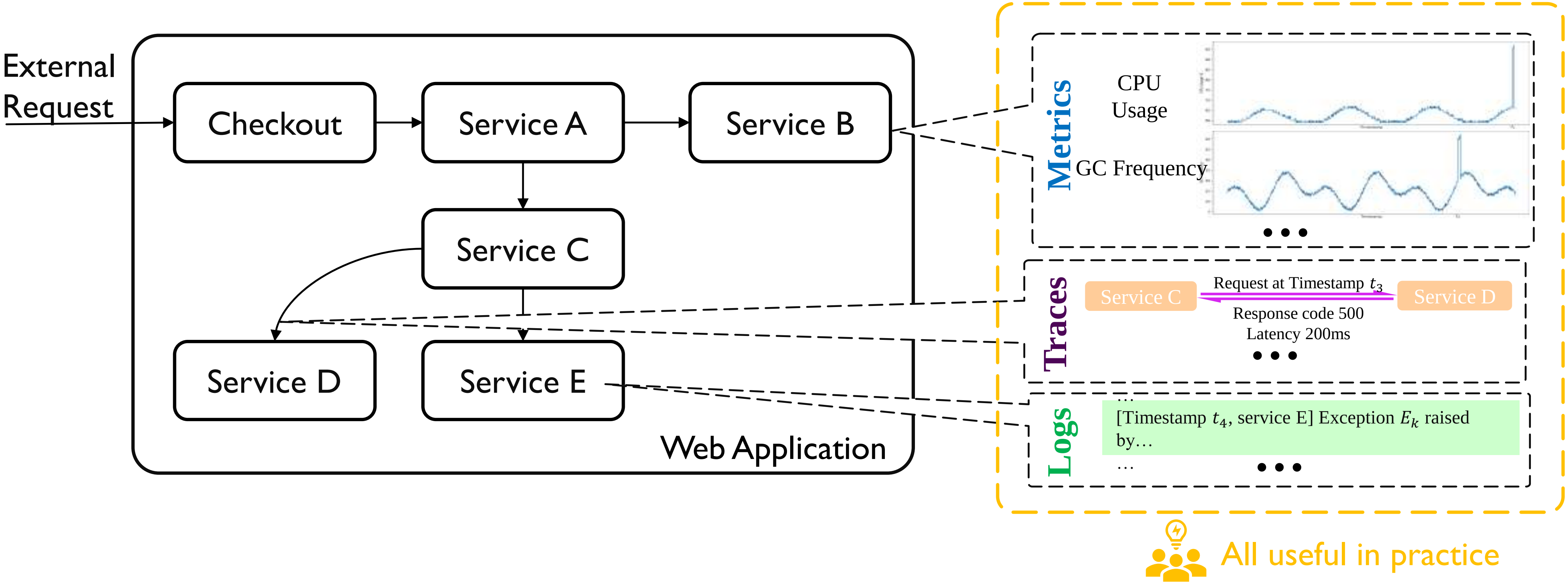
Multi-modal Monitoring Data in Microservice Architecture



Multi-modal Monitoring Data in Microservice Architecture



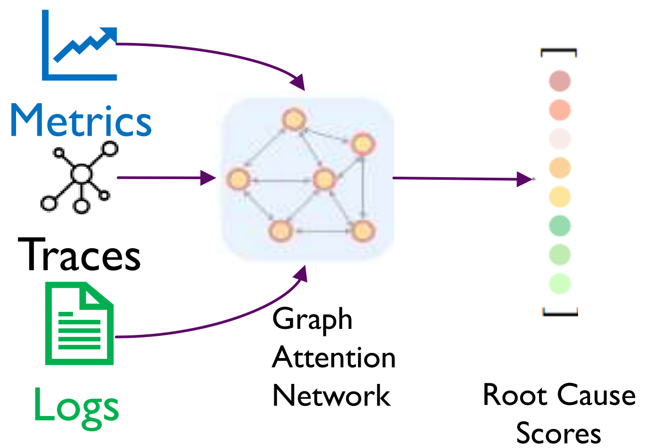
Multi-modal Monitoring Data in Microservice Architecture



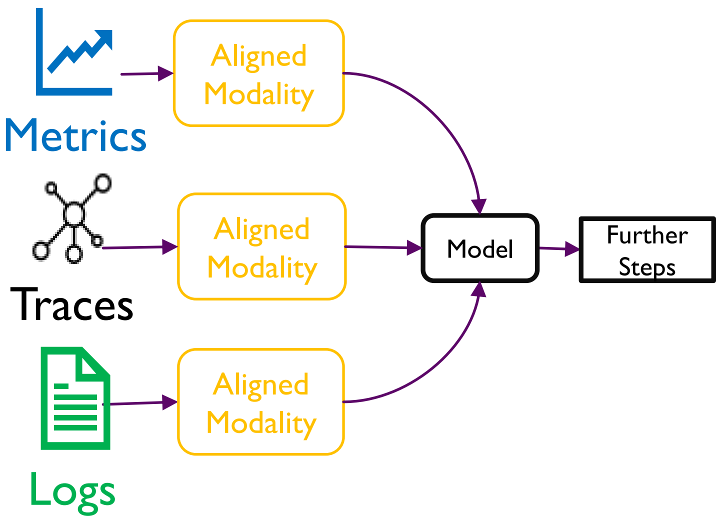
Challenge 1: How to integrate and analyze multi-modal data, leveraging information from various observation types?

Utilizing Multi-modal Monitoring Data

Black-box Fusion



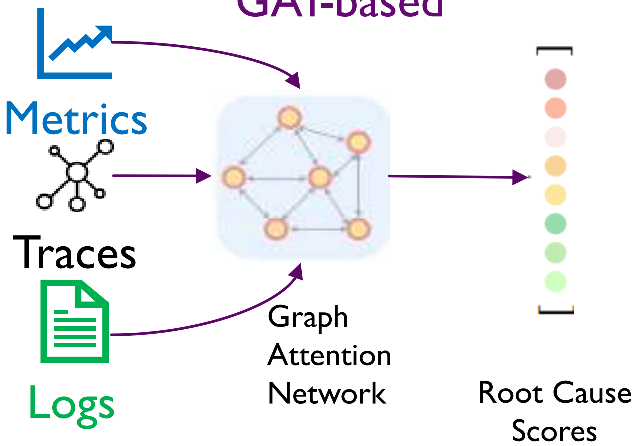
White-box Modality Alignment



Utilizing Multi-modal Monitoring Data

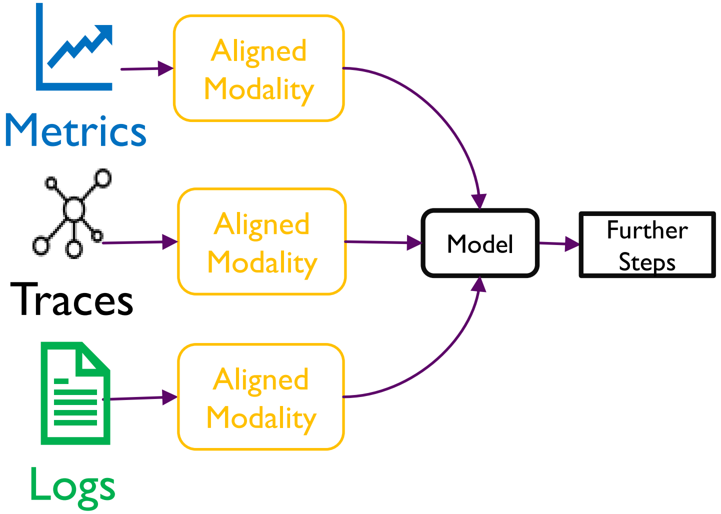
Black-box Fusion

GAT-based

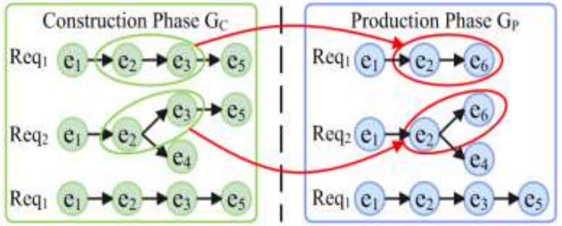


Eadro (ICSE' 23)

White-box Modality Alignment

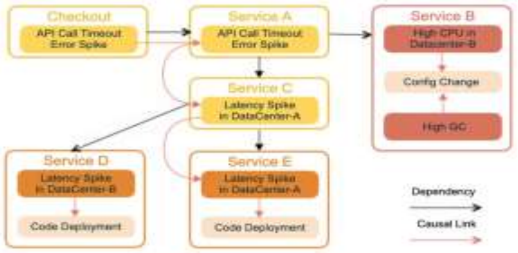


Event Pattern-based



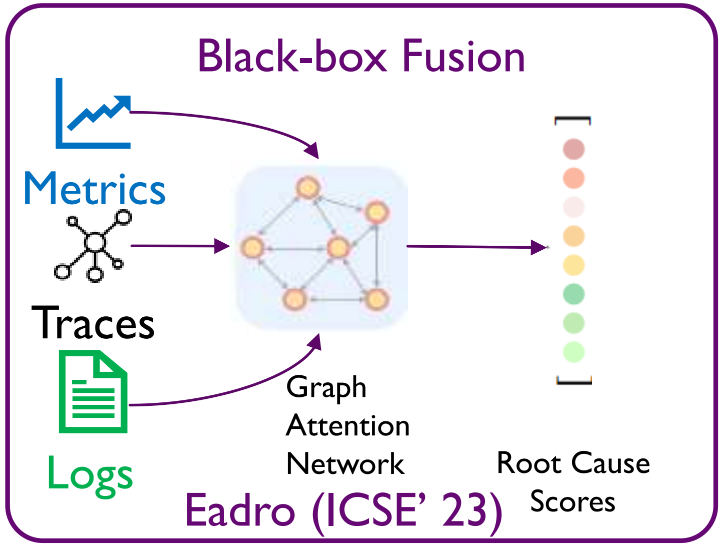
Nezha (ESEC/FSE' 23)

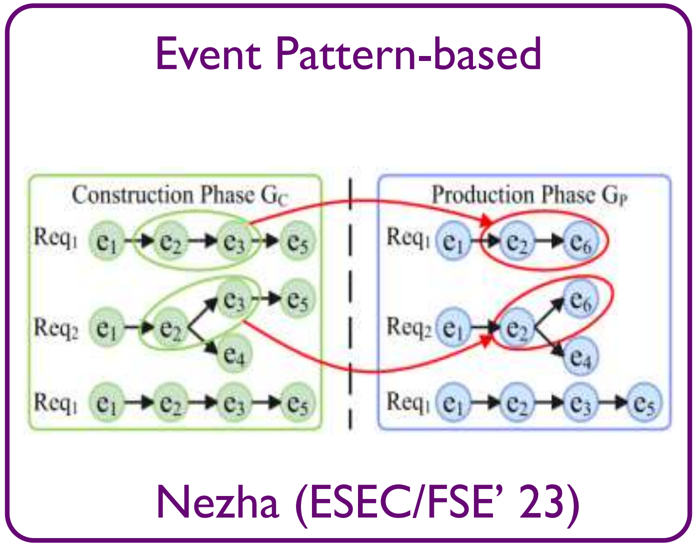
Event Causal Graph-based



Groot (ASE' 21)

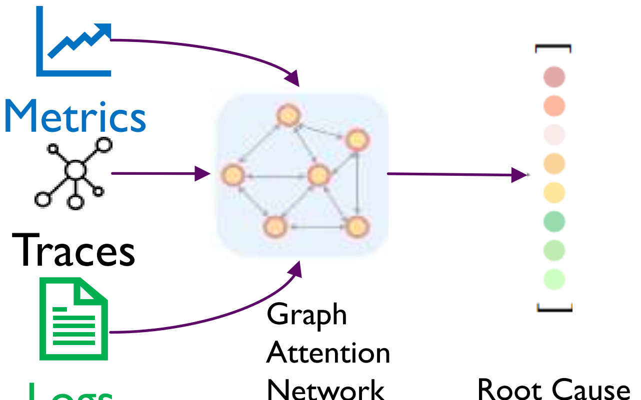
Utilizing Multi-modal Monitoring Data





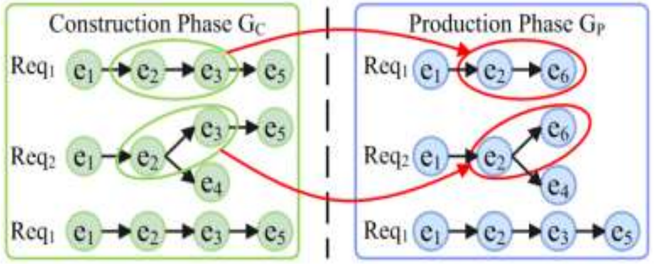
Utilizing Multi-modal Monitoring Data

Black-box Fusion



Eadro (ICSE' 23)

Event Pattern-based

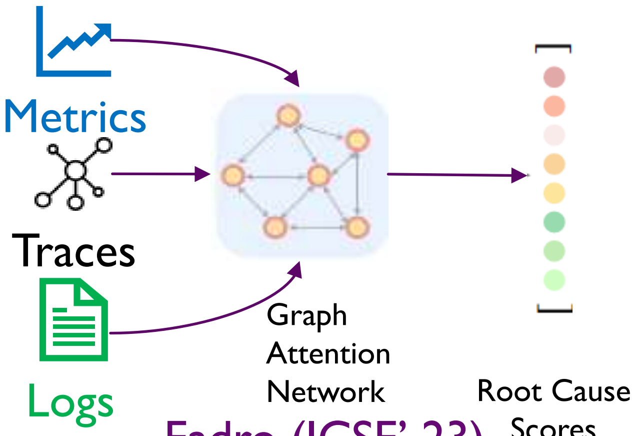


Nezha (ESEC/FSE' 23)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

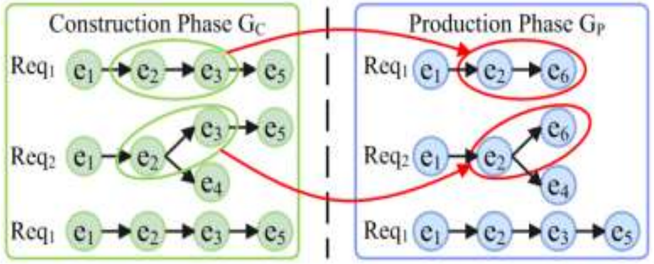
Utilizing Multi-modal Monitoring Data

Black-box Fusion



Eadro (ICSE' 23)

Event Pattern-based



Nezha (ESEC/FSE' 23)

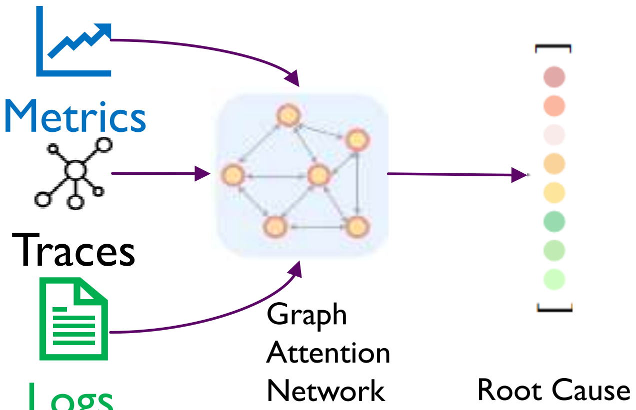
Interpretability

(parameters, outputs, and inference process)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

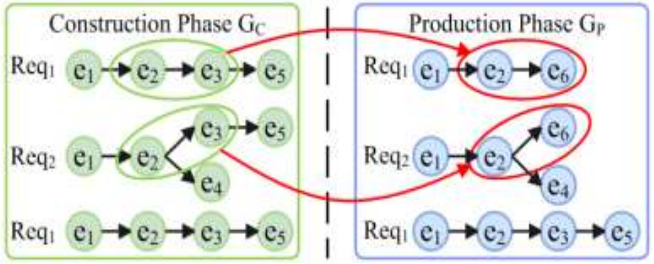
Utilizing Multi-modal Monitoring Data

Black-box Fusion



Eadro (ICSE’ 23)

Event Pattern-based



Nezha (ESEC/FSE’ 23)

Interpretability

(parameters, outputs, and inference process)

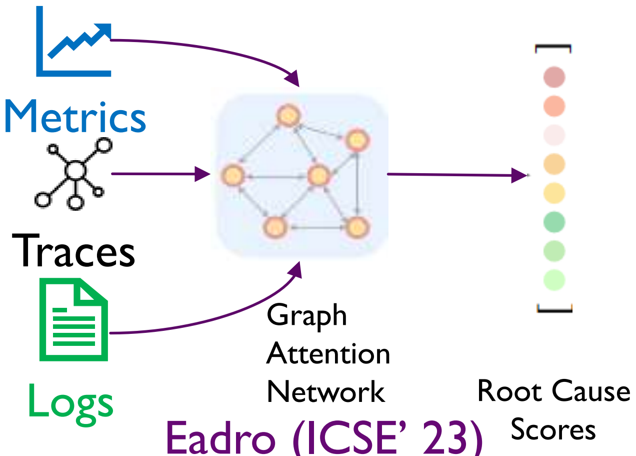
Human Knowledge Alignment

(optimize model with human insights)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

Utilizing Multi-modal Monitoring Data

Black-box Fusion

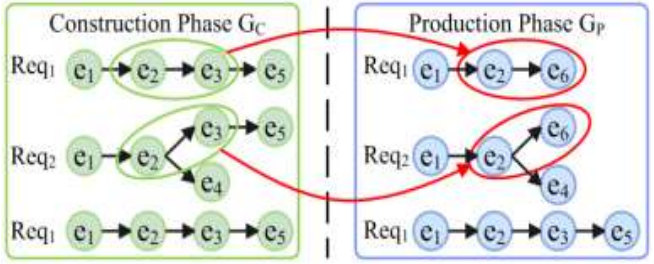


Eadro (ICSE' 23)

⚠ Blackbox Causality Learning

⚠ Parameters Hard to Understand

Event Pattern-based



Nezha (ESEC/FSE' 23)

⚠ Non-straightforward Output

Interpretability

(parameters, outputs, and inference process)

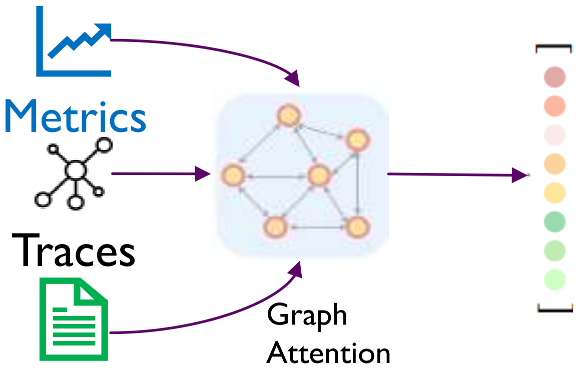
Human Knowledge Alignment

(optimize model with human insights)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

Utilizing Multi-modal Monitoring Data

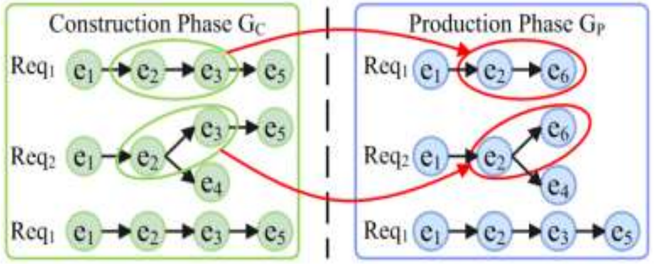
Black-box Fusion



Eadro (ICSE' 23)

- Blackbox Causality Learning
- Parameters Hard to Understand
- Requiring Deep-learning Background

Event Pattern-based



Nezha (ESEC/FSE' 23)

- Non-straightforward Output
- Hard to Integrate Human Knowledge

Interpretability

(parameters, outputs, and inference process)

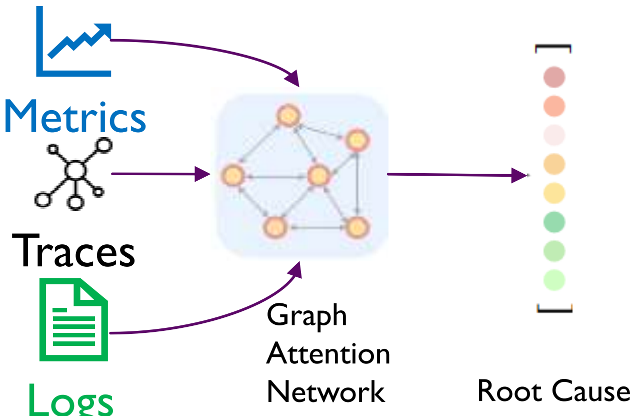
Human Knowledge Alignment

(optimize model with human insights)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

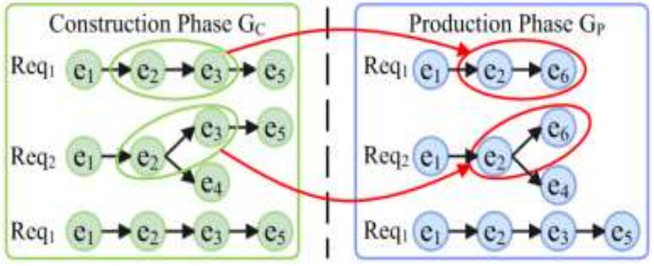
Utilizing Multi-modal Monitoring Data

Black-box Fusion



Eadro (ICSE' 23)

Event Pattern-based



Nezha (ESEC/FSE' 23)

- Blackbox Causality Learning
- Parameters Hard to Understand
- Requiring Deep-learning Background

- Non-straightforward Output
- Hard to Integrate Human Knowledge

Interpretability

(parameters, outputs, and inference process)

Human Knowledge Alignment

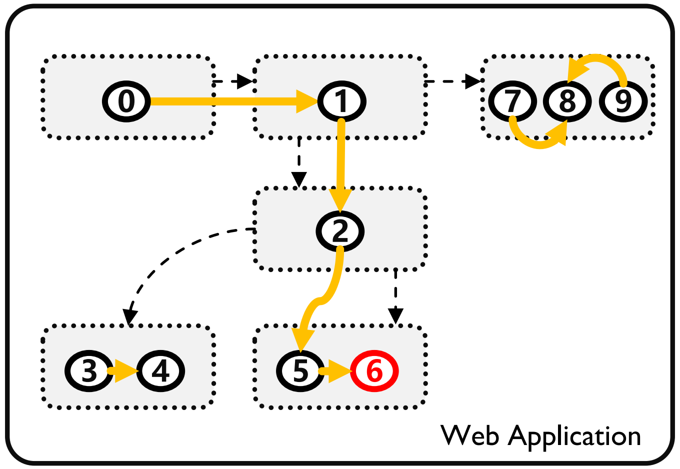
(optimize model with human insights)

How is this conclusion drawn?
What do the parameters mean?
What does the result mean?
Do these root causes align with my understanding?
How to optimize it with my insight of the system?

Challenge 2: Interpretability and Straightforward Alignment to Human Knowledge.

Utilizing Multi-modal Monitoring Data

Event Causal Graph-based

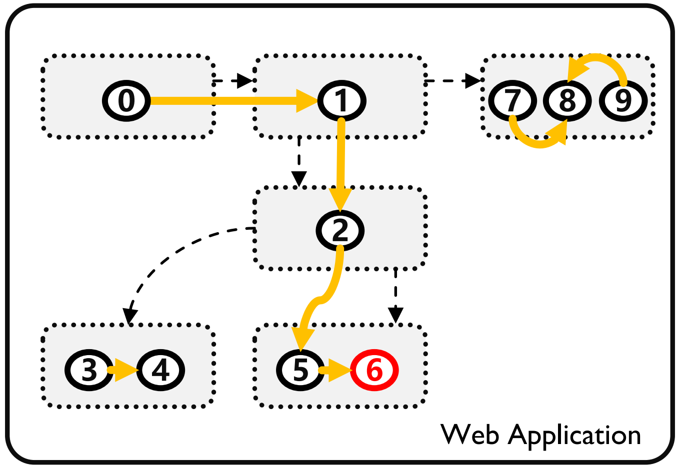


Groot (ASE' 21)

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Event causal link

Utilizing Multi-modal Monitoring Data

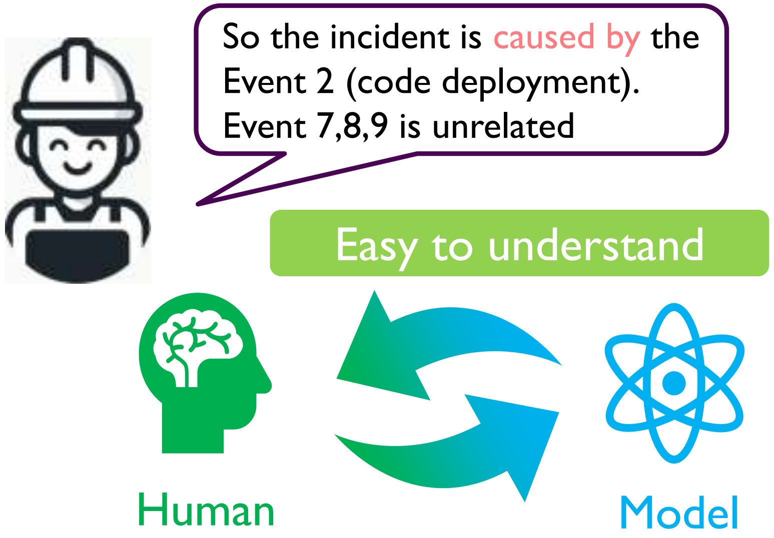
Event Causal Graph-based



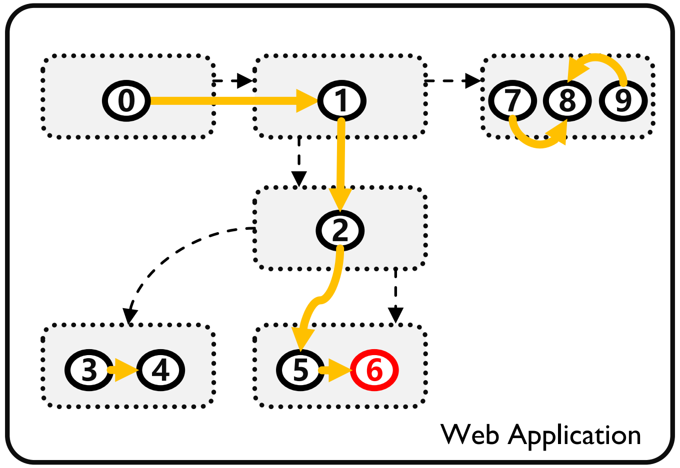
Groot (ASE' 21)

- 0 API Call Timeout Error Spike
- 1 API Call Timeout Error Spike
- 2 Latency Spike
- 3 Service-Client Error Spike
- 4 Code Deployment
- 5 DB Markdown
- 6 Code Deployment
- 7 High CPU Usage
- 8 Config Change
- 9 High GC
- Service
- Event causal link

Utilizing Multi-modal Monitoring Data



Event Causal Graph-based



Groot (ASE' 21)

- 0 API Call Timeout Error Spike
- 1 API Call Timeout Error Spike
- 2 Latency Spike
- 3 Service-Client Error Spike
- 4 Code Deployment
- 5 DB Markdown
- 6 Code Deployment
- 7 High CPU Usage
- 8 Config Change
- 9 High GC
- Service
- Event causal link

Utilizing Multi-modal Monitoring Data

So the incident is caused by the Event 2 (code deployment). Event 7,8,9 is unrelated

Easy to understand

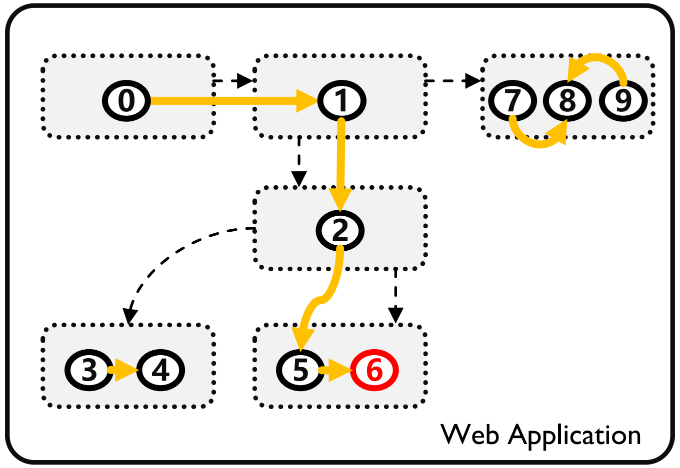
Human

Model

Easy to integrate human knowledge

I know the Event 5 is often caused by Event 6, let's add an edge!

Event Causal Graph-based



Groot (ASE' 21)

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Event causal link

Utilizing Multi-modal Monitoring Data

So the incident is caused by the Event 2 (code deployment). Event 7,8,9 is unrelated

Easy to understand

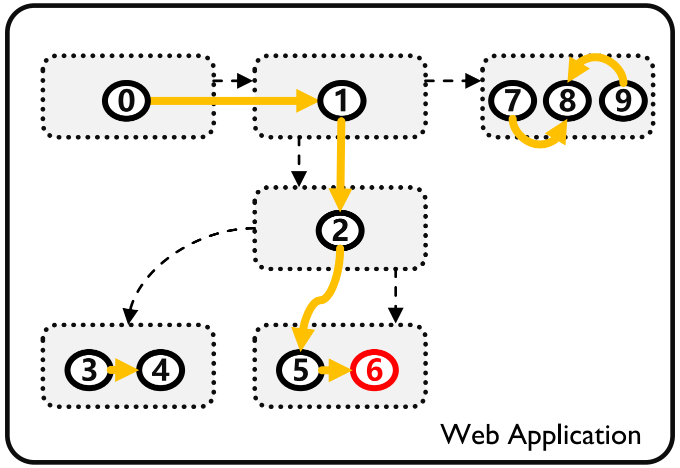
Human

Model

Easy to integrate human knowledge

I know the Event 5 is often caused by Event 6, let's add an edge!

Event Causal Graph-based



Groot (ASE' 21)

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Event causal link

Rulebook

Utilizing Multi-modal Monitoring Data

So the incident is caused by the Event 2 (code deployment). Event 7,8,9 is unrelated

Easy to understand

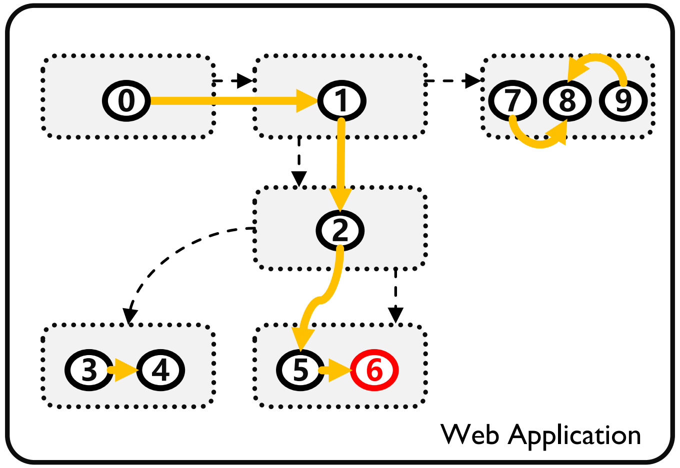
Human

Model

Easy to integrate human knowledge

I know the Event 5 is often caused by Event 6, let's add an edge!

Event Causal Graph-based



Groot (ASE' 21)

- 0 API Call Timeout Error Spike
- 1 API Call Timeout Error Spike
- 2 Latency Spike
- 3 Service-Client Error Spike
- 4 Code Deployment
- 5 DB Markdown
- 6 Code Deployment
- 7 High CPU Usage
- 8 Config Change
- 9 High GC
- Service
- Event causal link

Rulebook

Remediation Logs

Utilizing Multi-modal Monitoring Data

So the incident is caused by the Event 2 (code deployment). Event 7,8,9 is unrelated

Easy to understand

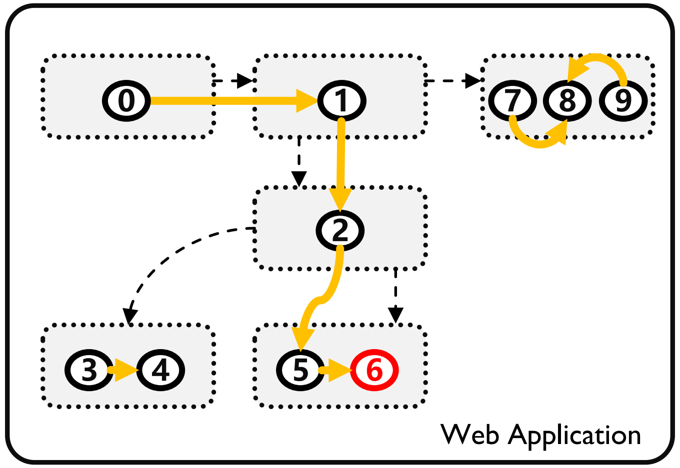
Human

Model

Easy to integrate human knowledge

I know the Event 5 is often caused by Event 6, let's add an edge!

Event Causal Graph-based

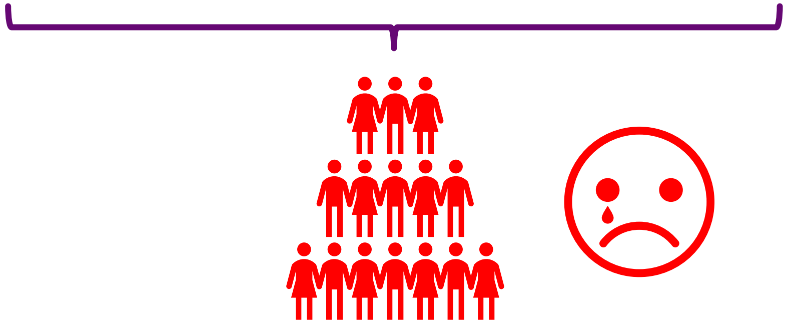


Groot (ASE' 21)

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Event causal link

Rulebook

Remediation Logs



Expert-labor Intensive

Utilizing Multi-modal Monitoring Data

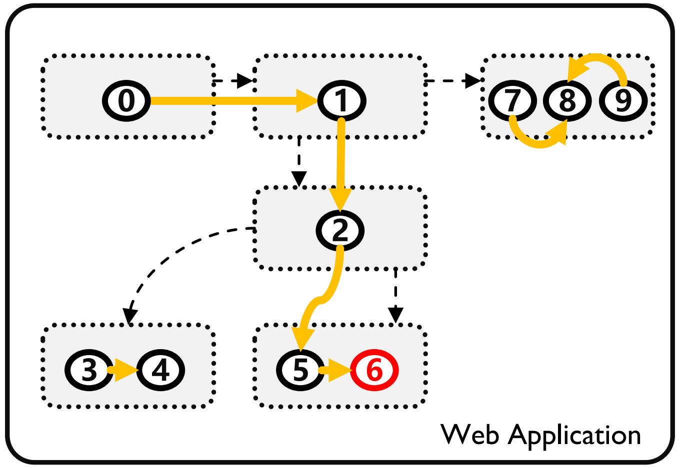
So the incident is caused by the Event 2 (code deployment). Event 7,8,9 is unrelated

Easy to understand

Easy to integrate human knowledge

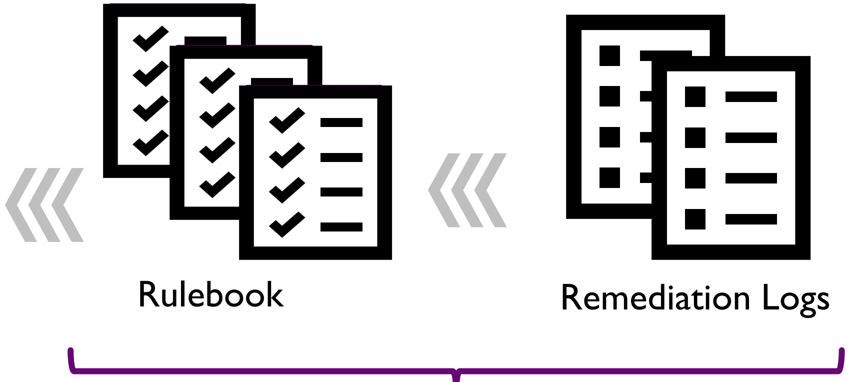
I know the Event 5 is often caused by Event 6, let's add an edge!

Event Causal Graph-based



Groot (ASE' 21)

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Event causal link



Expert-labor Intensive

Challenge 3: Automatic Causality Learning

Challenges

Multi-Modal Data Integration

- Leveraging information from various observation types, including metrics, traces, and logs

Interpretability and Straightforward Alignment to Human Knowledge

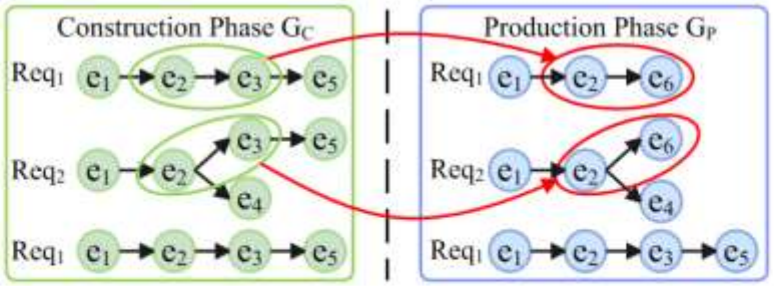
- Parameter structure should ideally have a clear physical meaning that aligns with the knowledge of SREs
- SREs without deep learning background can easily optimize the model based on their operational experience

Automatic Causality Learning

- Automatically learn causality in microservice systems, minimizing or eliminating the necessity for manual configuration.

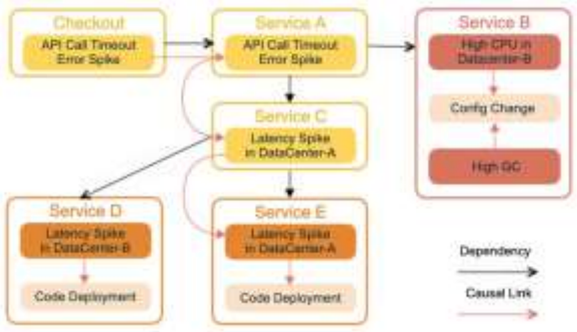
Event-based RCA

Event Pattern-based

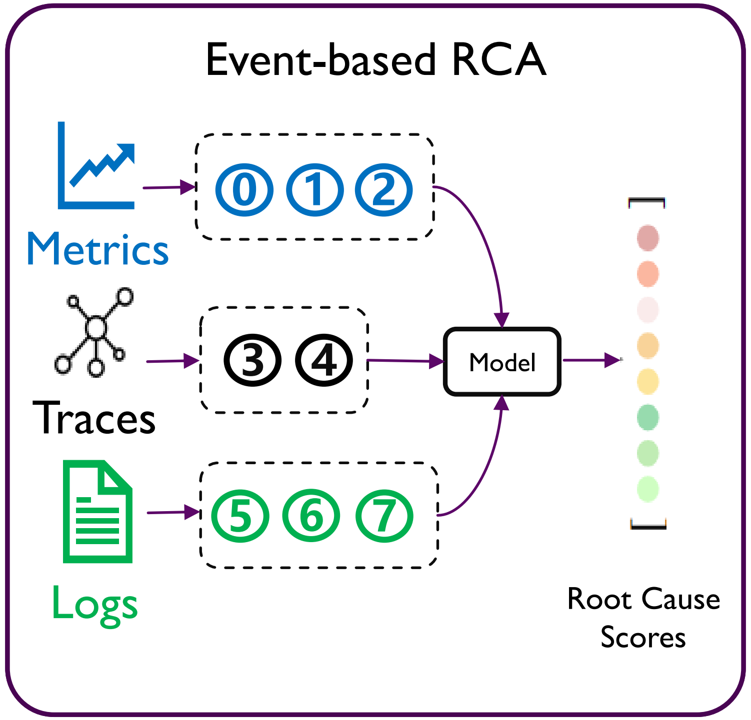


Nezha (ESEC/FSE' 23)

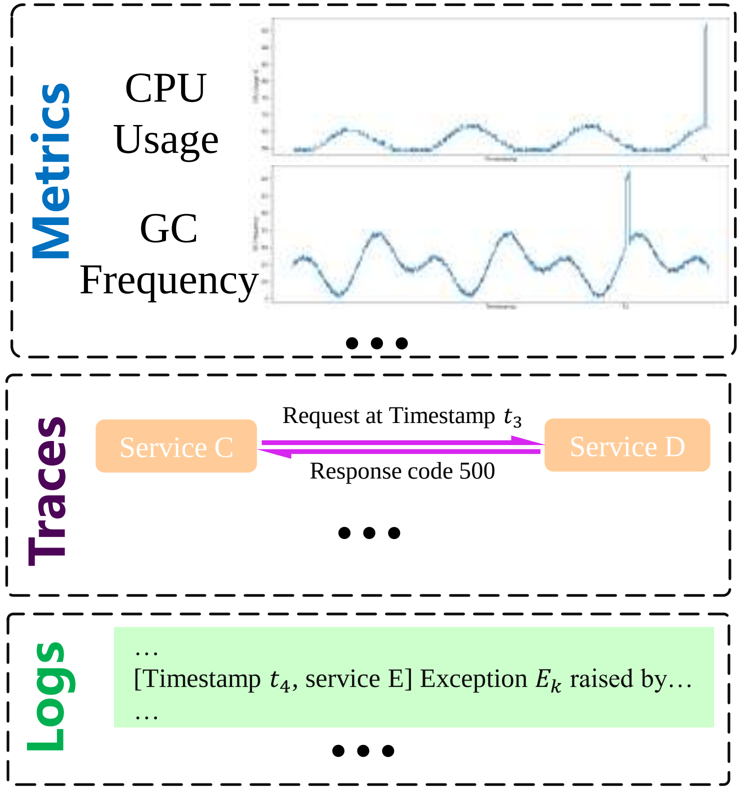
Event Causal Graph-based



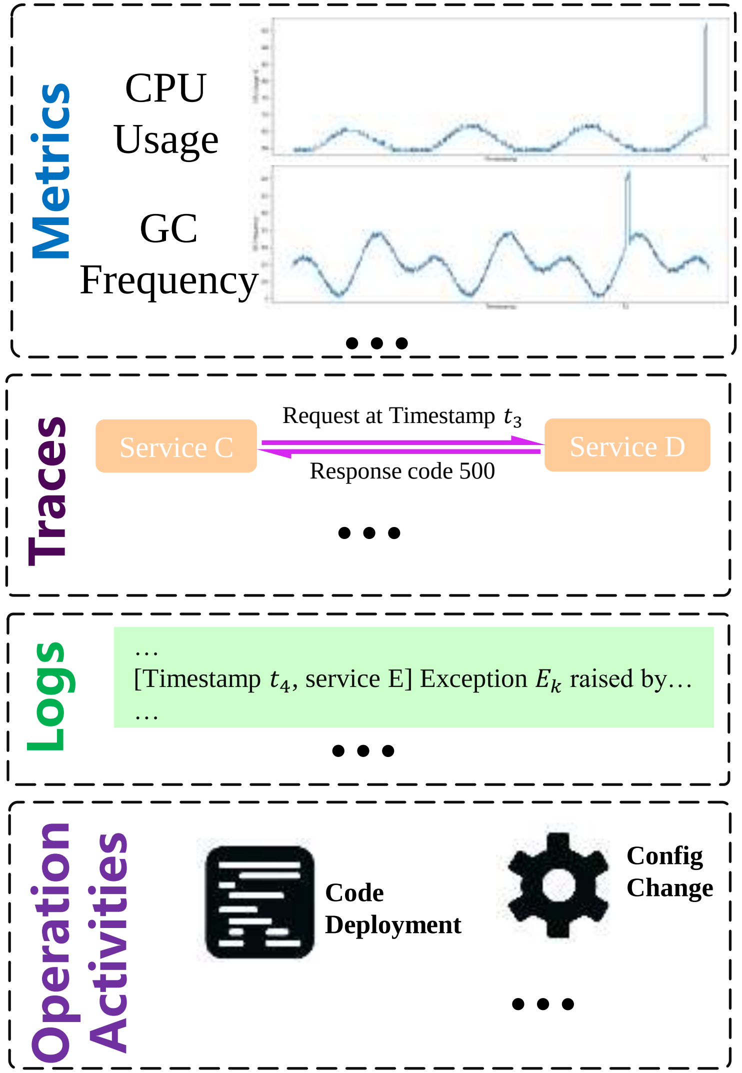
Groot (ASE' 21)



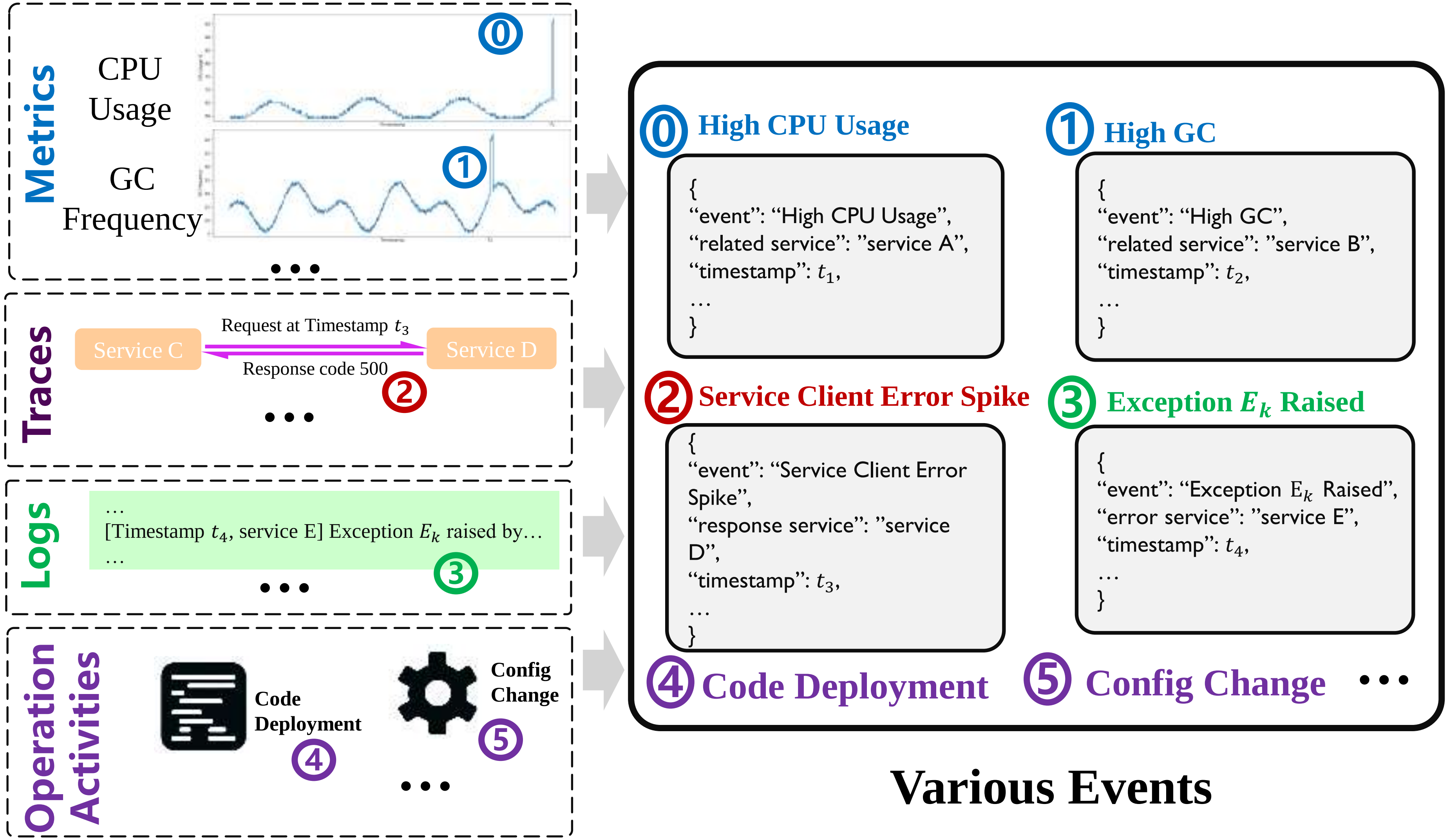
Multi-modal Monitoring Data into Events



Multi-modal Monitoring Data into Events

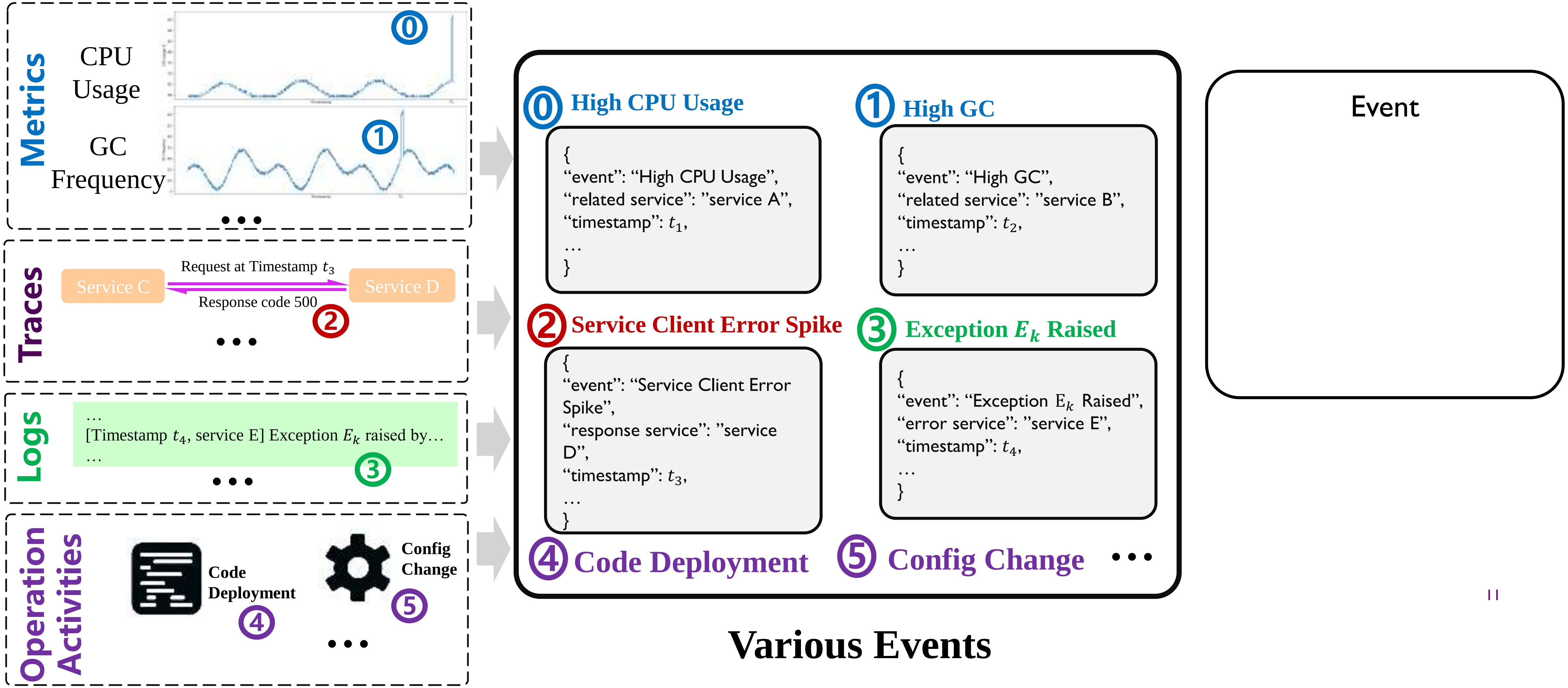


Multi-modal Monitoring Data into Events

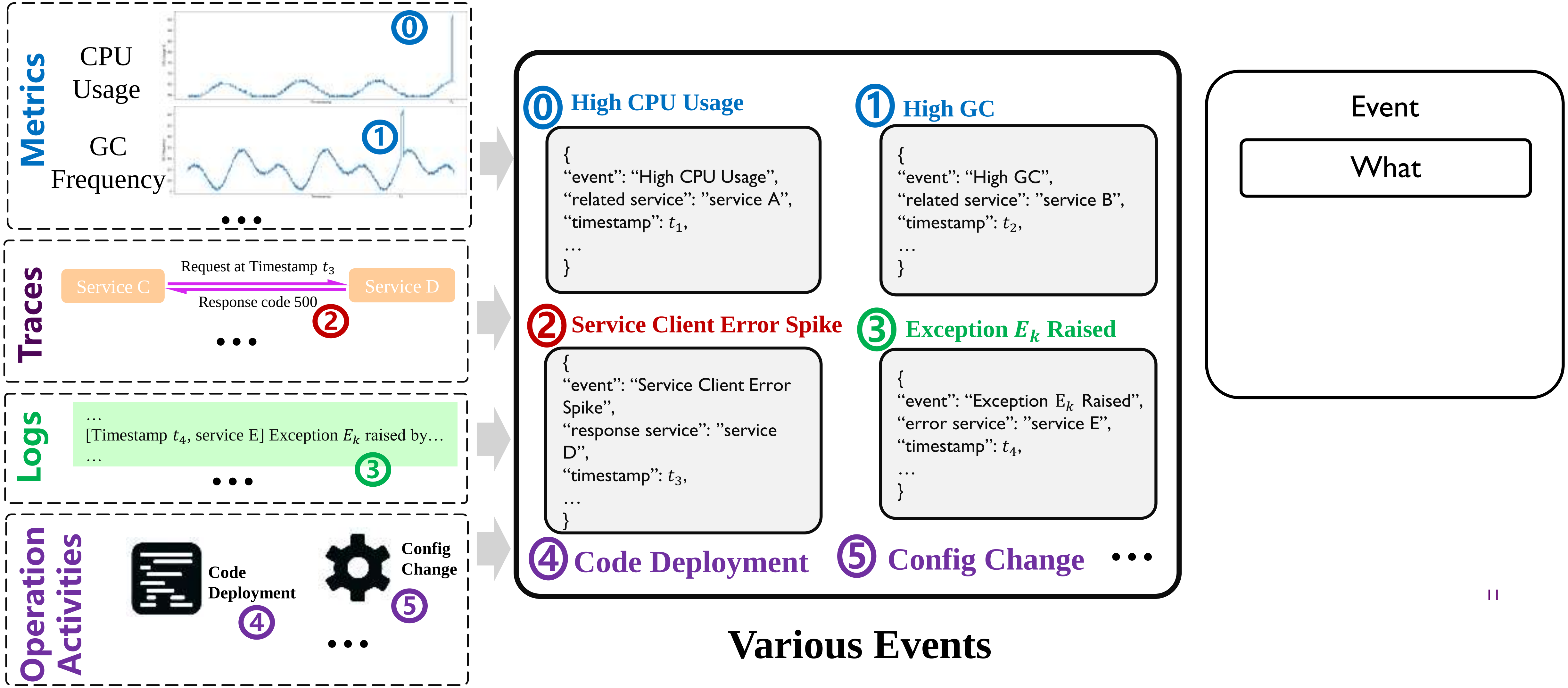


Various Events

Multi-modal Monitoring Data into Events

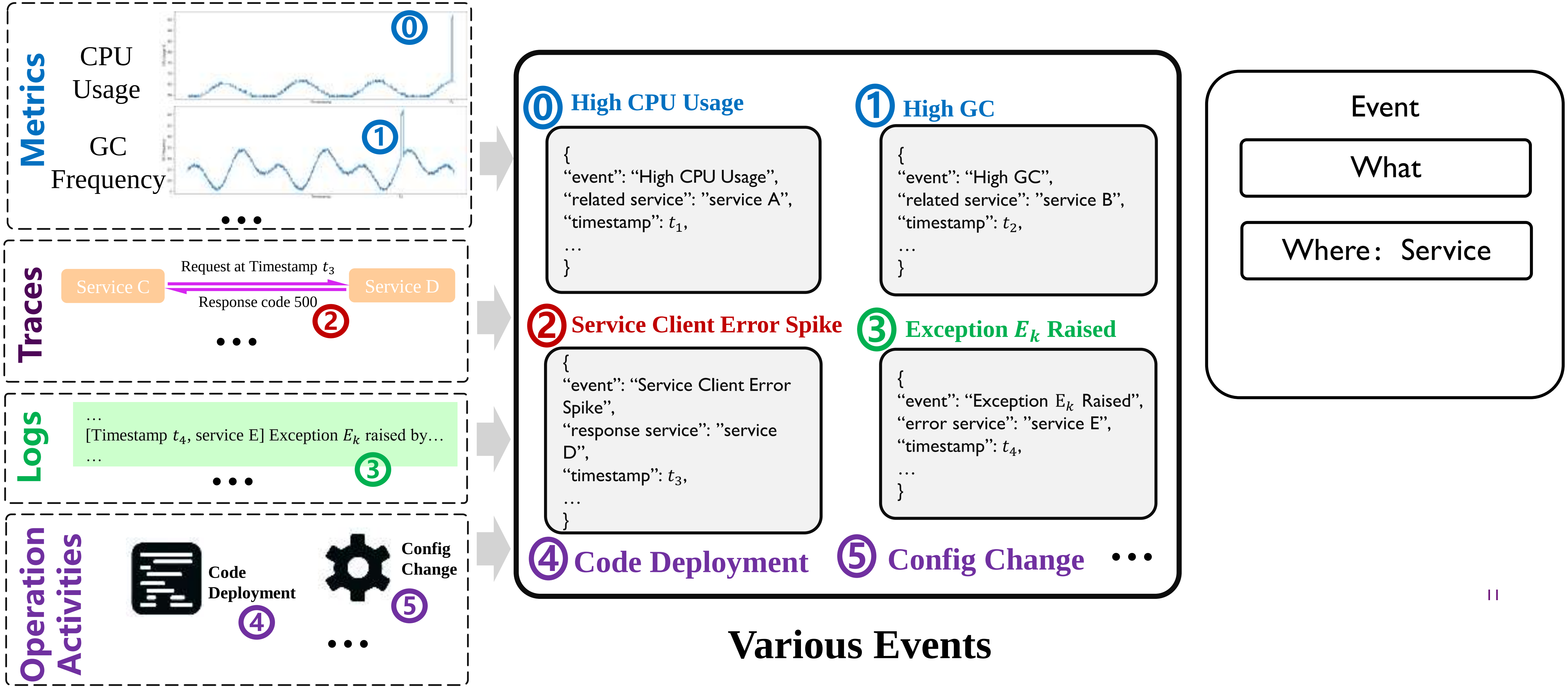


Multi-modal Monitoring Data into Events



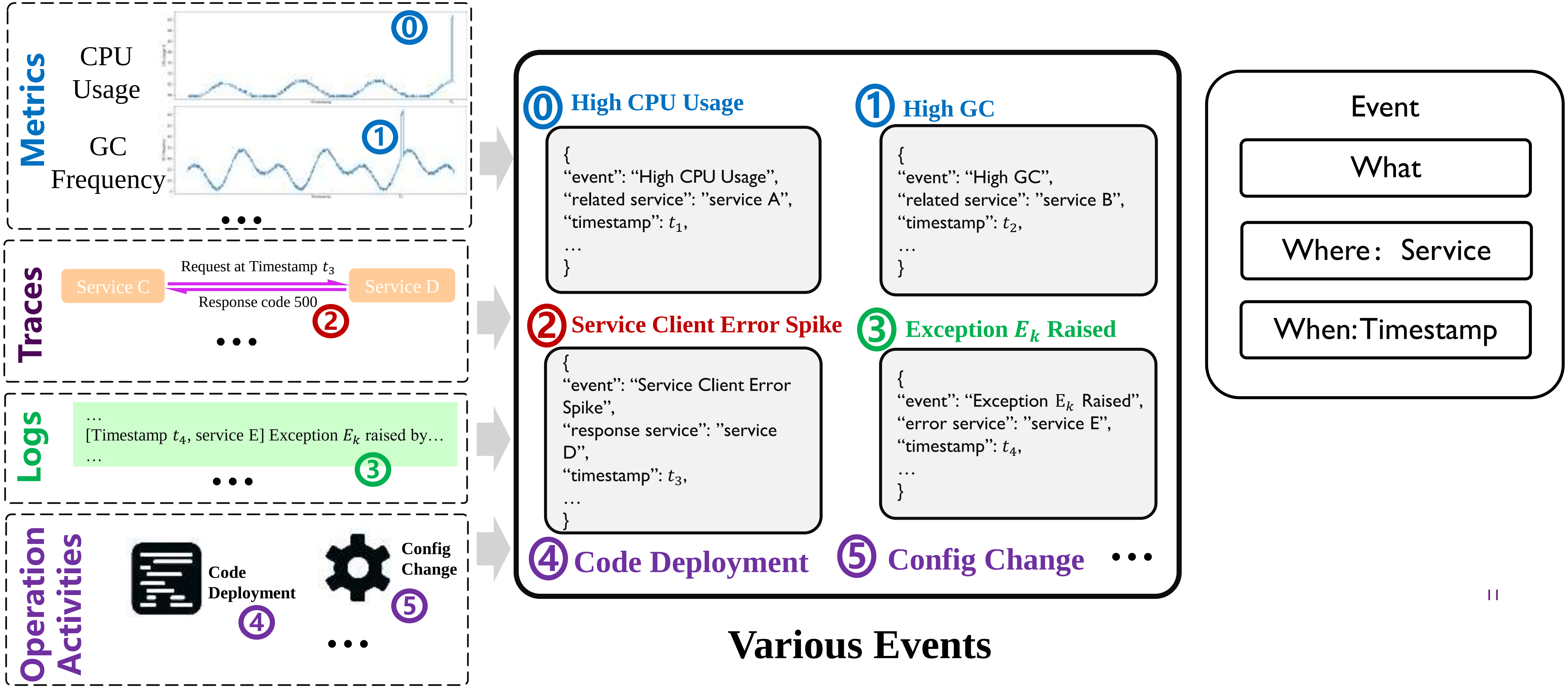
Various Events

Multi-modal Monitoring Data into Events



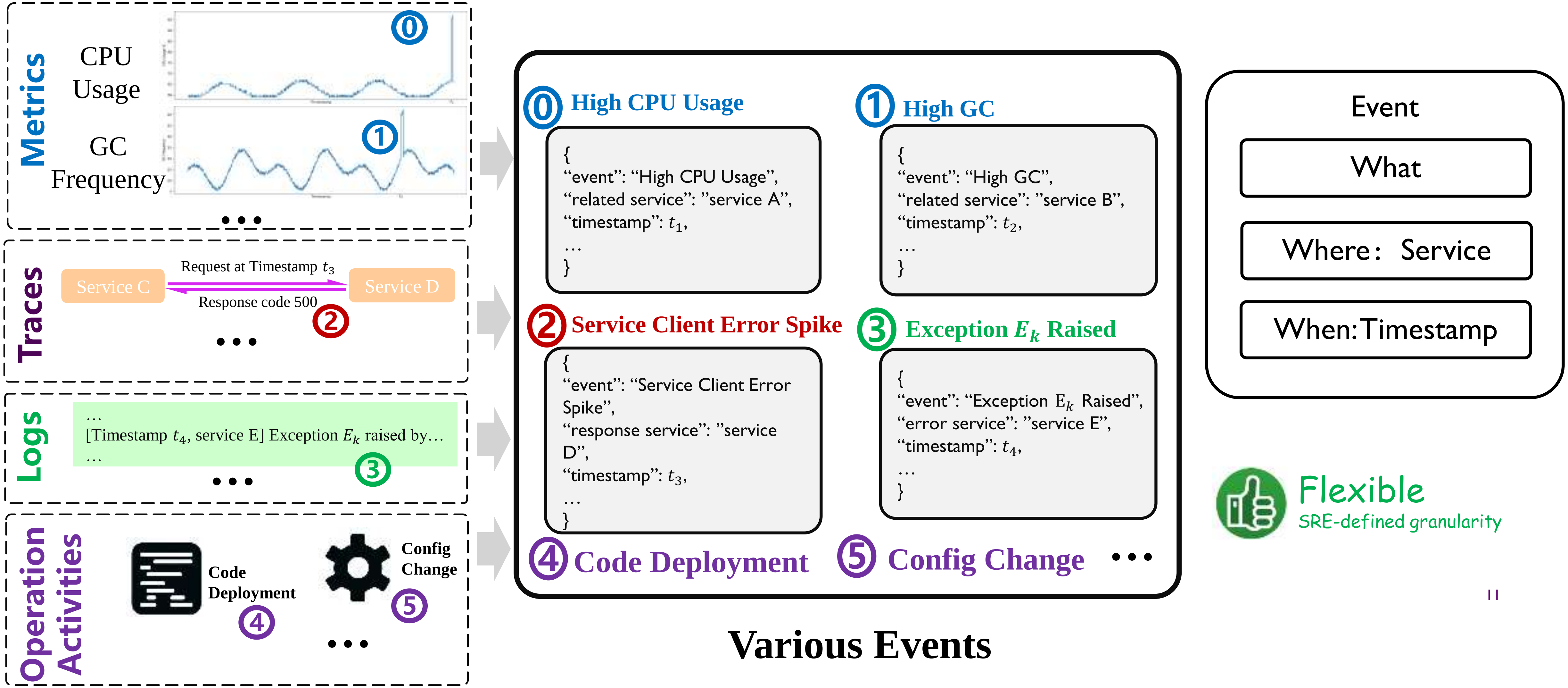
Various Events

Multi-modal Monitoring Data into Events



Various Events

Multi-modal Monitoring Data into Events



Problem Formulation

- Given the event collection

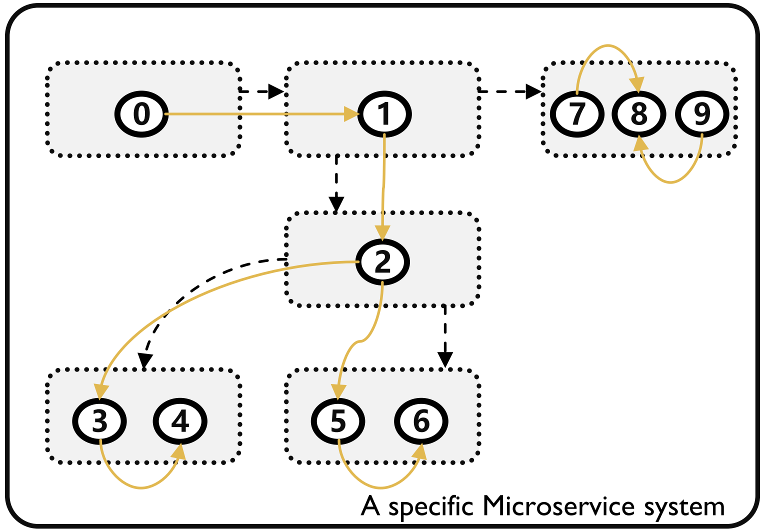
⓿ API Call Timeout Error Spike
❶ API Call Timeout Error Spike
❷ Latency Spike
❸ Service-Client Error Spike
❹ Code Deployment
❺ DB Markdown
❻ Code Deployment
❼ High CPU Usage
❽ Config Change
❾ High GC

Problem Formulation

- Given the event collection

- Automatically learn the event causal graph

0 API Call Timeout Error Spike
1 API Call Timeout Error Spike
2 Latency Spike
3 Service-Client Error Spike
4 Code Deployment
5 DB Markdown
6 Code Deployment
7 High CPU Usage
8 Config Change
9 High GC
Service
Service Call



Problem Formulation

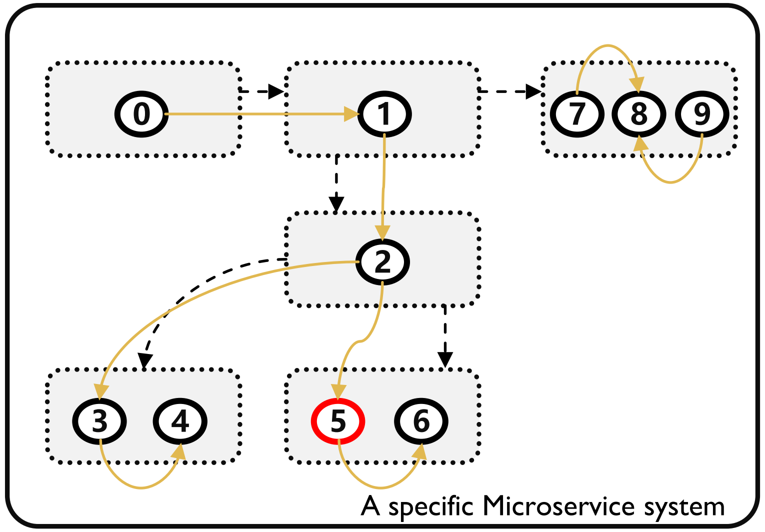
- Given the event collection

- Automatically learn the event causal graph
- Infer the real root cause event for a specific incident

0. API Call Timeout Error Spike
1. API Call Timeout Error Spike
2. Latency Spike
3. Service-Client Error Spike
4. Code Deployment
5. DB Markdown
6. Code Deployment
7. High CPU Usage
8. Config Change
9. High GC

Service

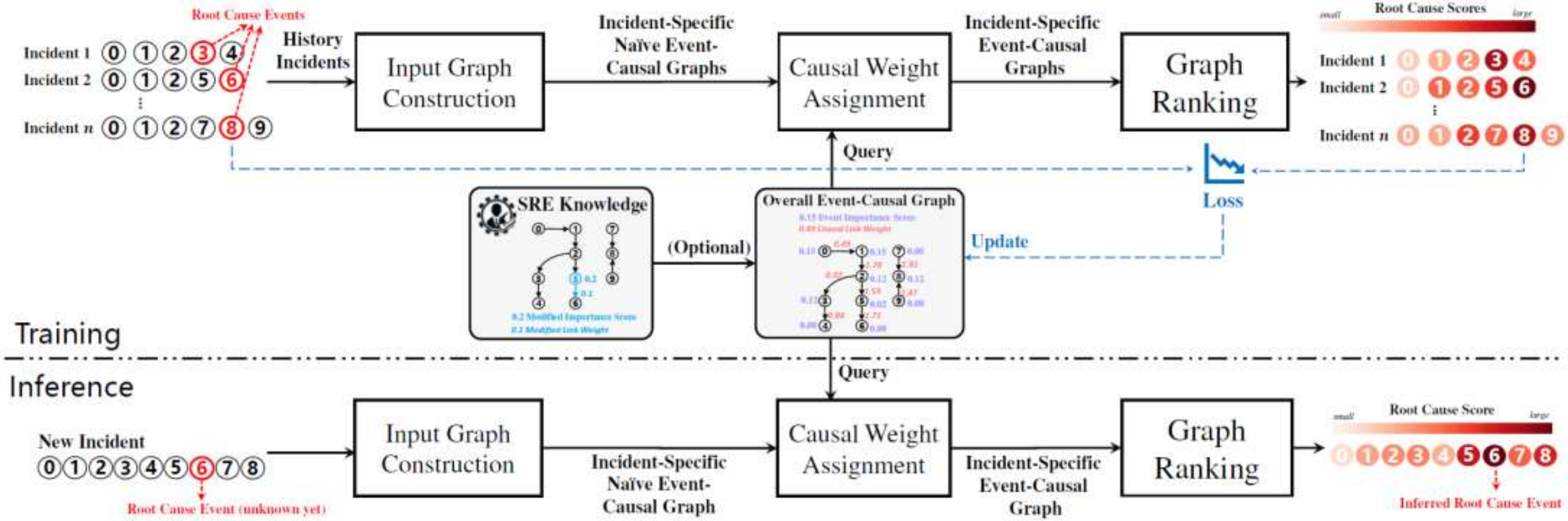
Service Call



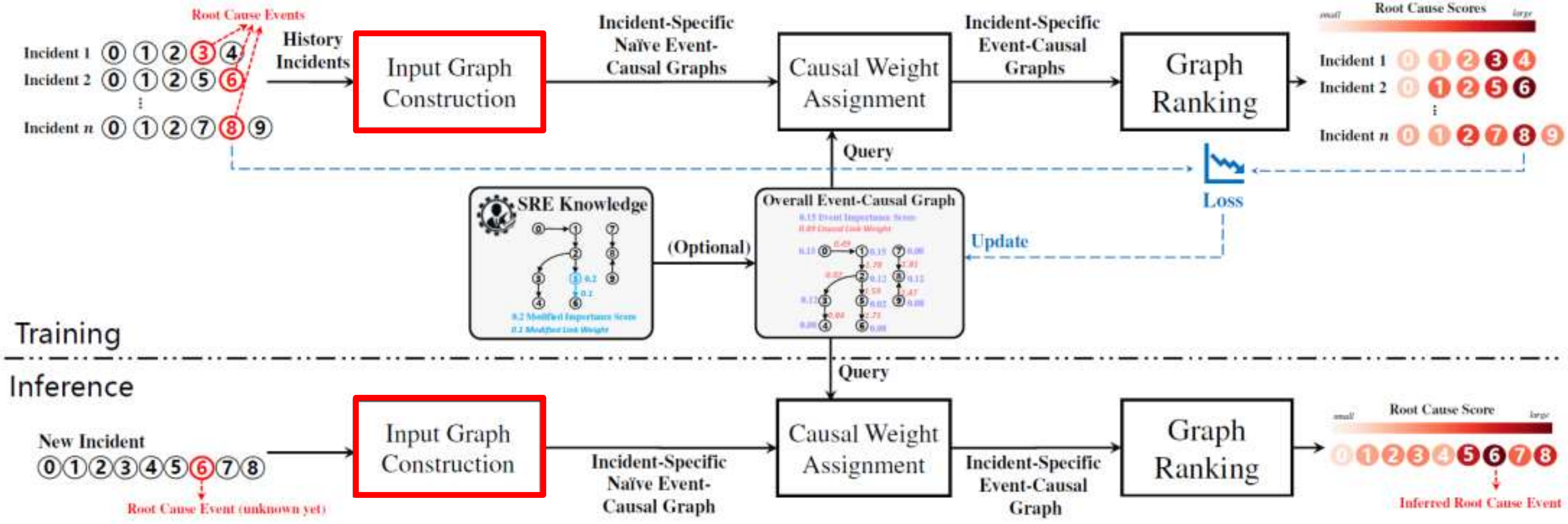
Outline

- Background
- Design
- Evaluation
- Conclusion

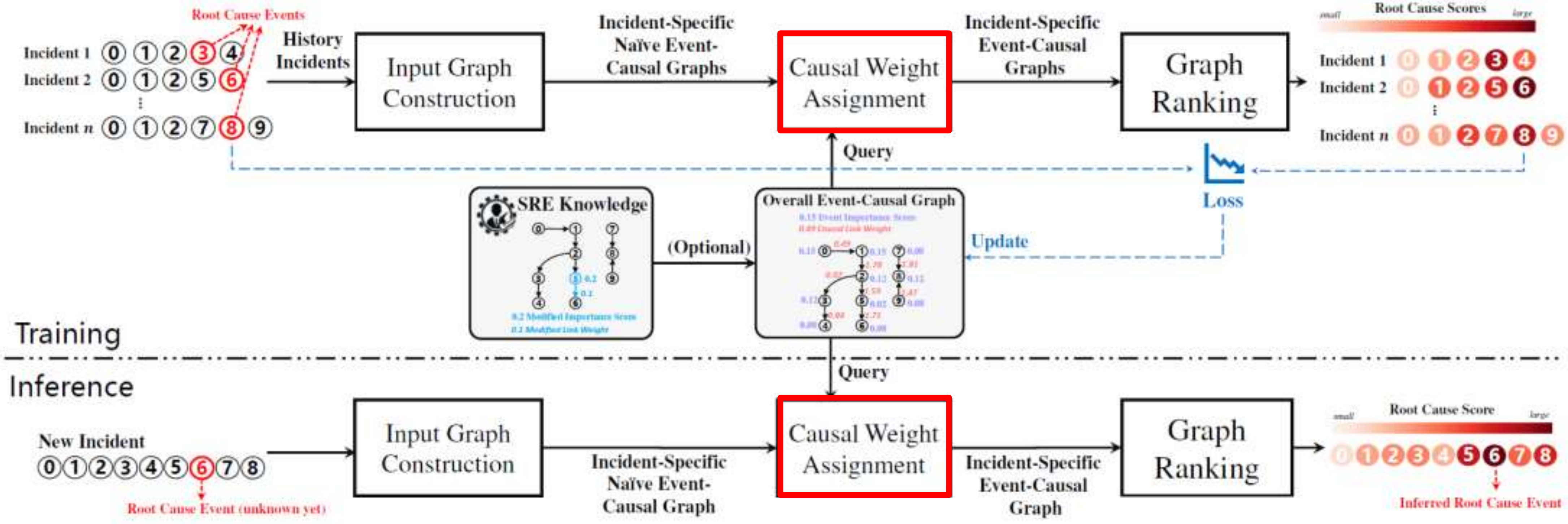
Chain-of-Event (CoE) Design Overview



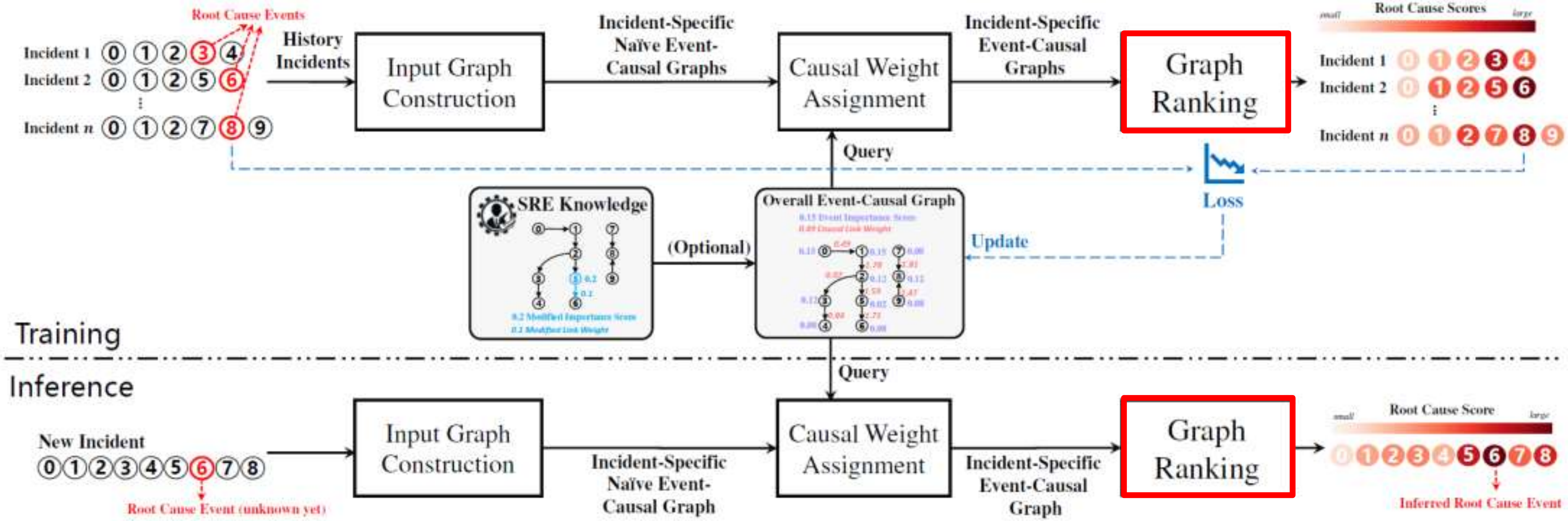
Chain-of-Event (CoE) Design Overview



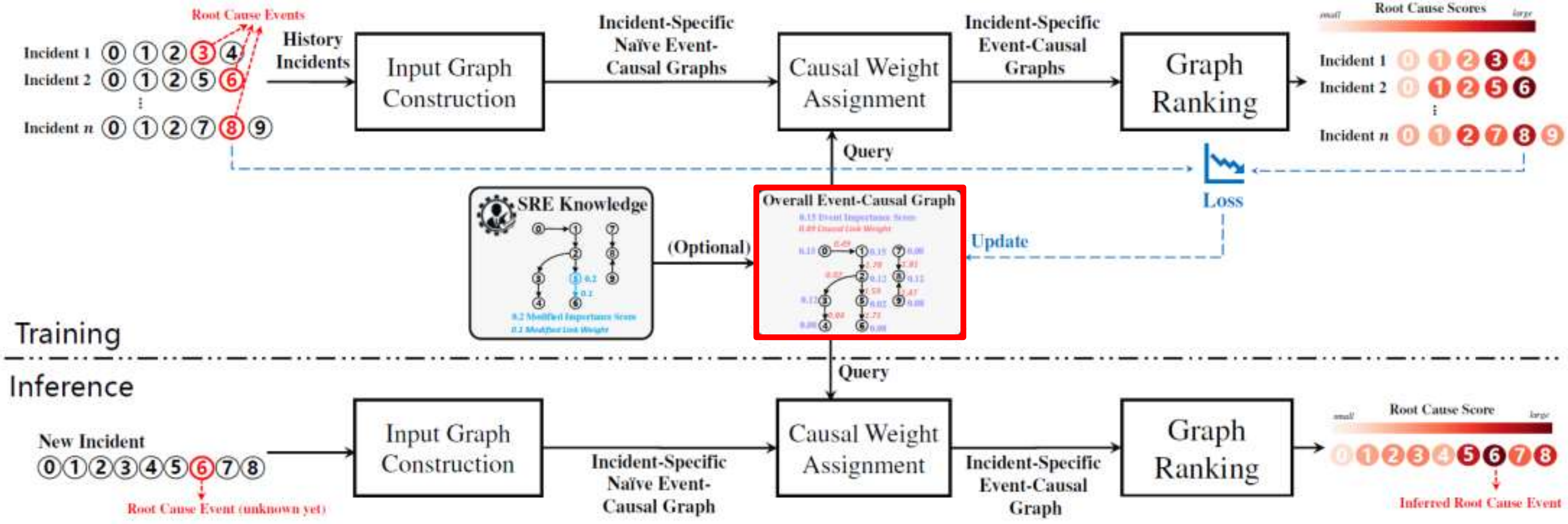
Chain-of-Event (CoE) Design Overview



Chain-of-Event (CoE) Design Overview



Chain-of-Event (CoE) Design Overview



Constructing Incident-specific Event-causal Graph

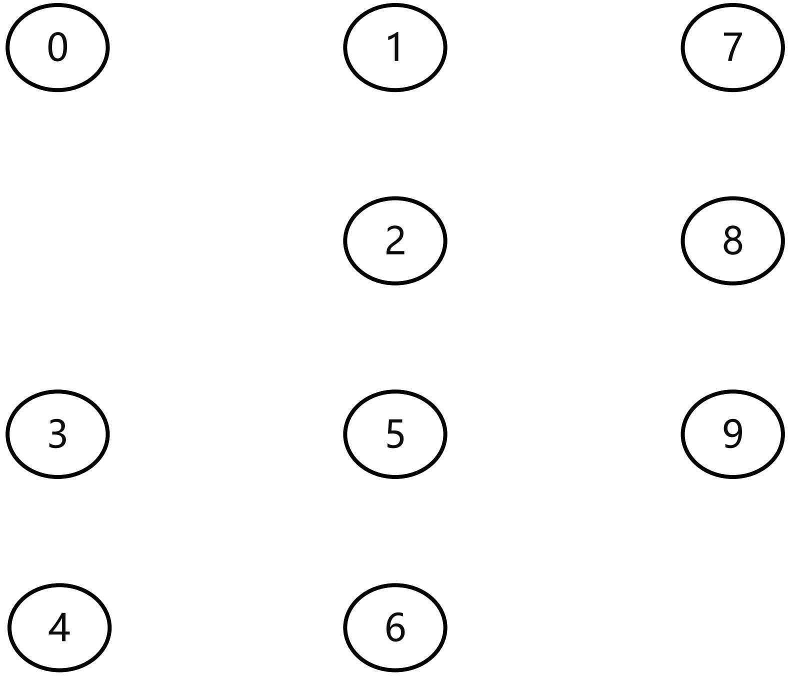
Constructing Incident-specific Event-causal Graph

Incident

Constructing Incident-specific Event-causal Graph

Incident

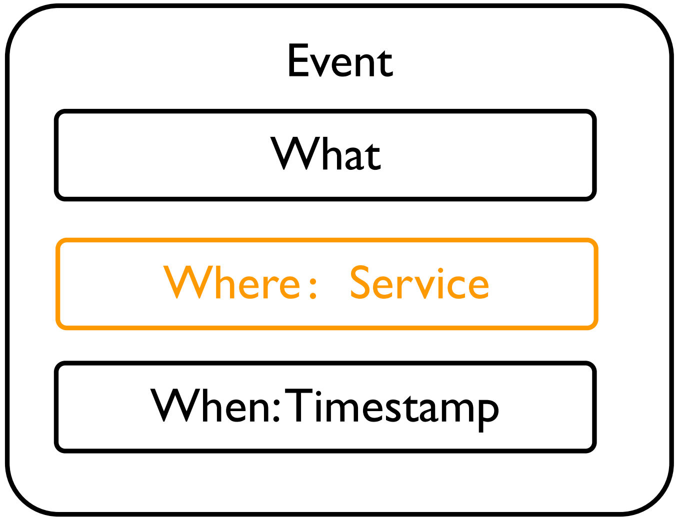
- A collection of events in a time window

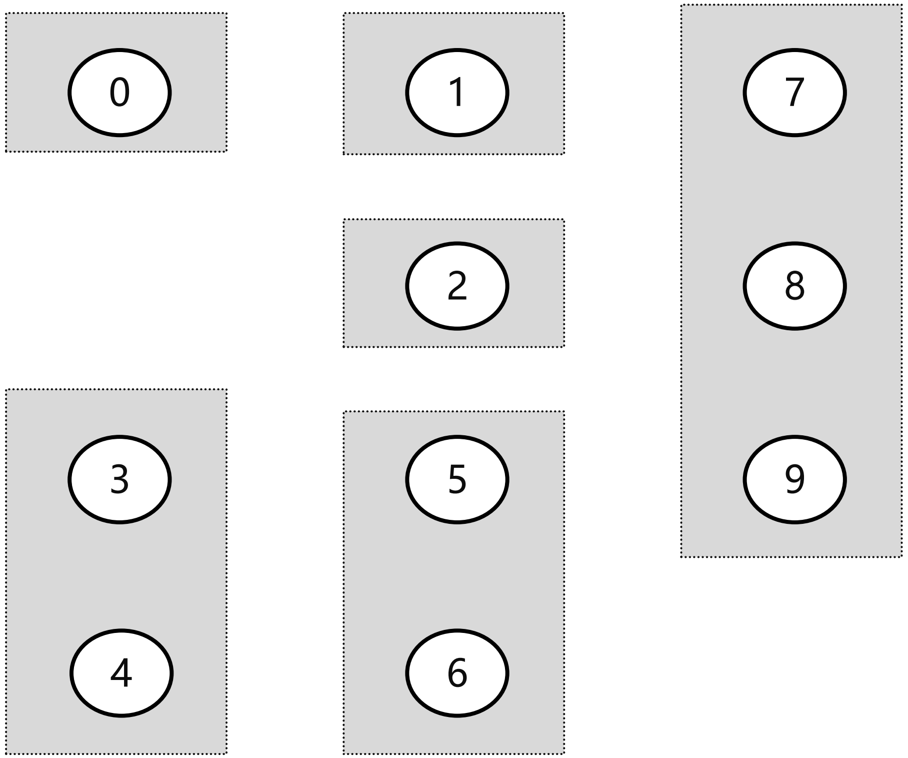


Constructing Incident-specific Event-causal Graph

Incident

- A collection of events in a time window
- Service recorded in the events



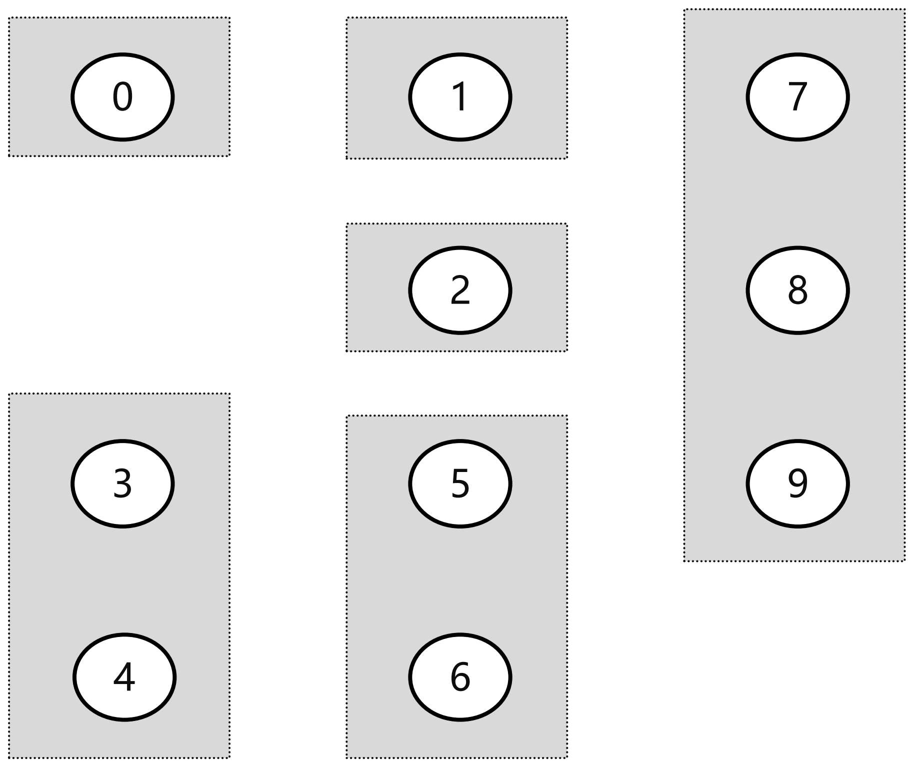


Constructing Incident-specific Event-causal Graph

Incident

- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph



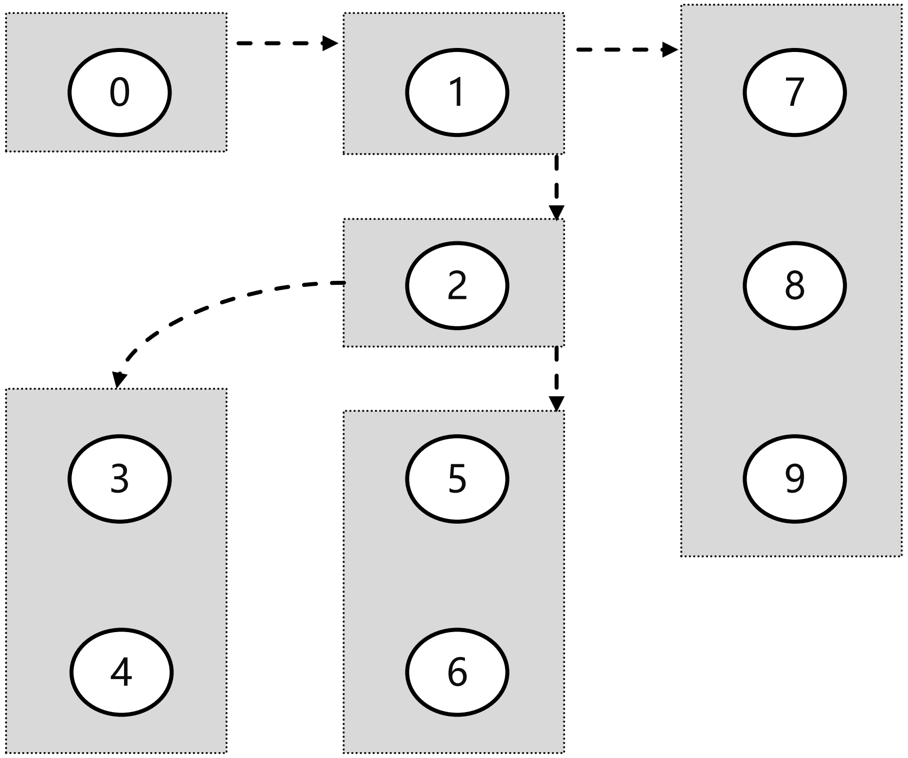
Constructing Incident-specific Event-causal Graph

Incident

- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship



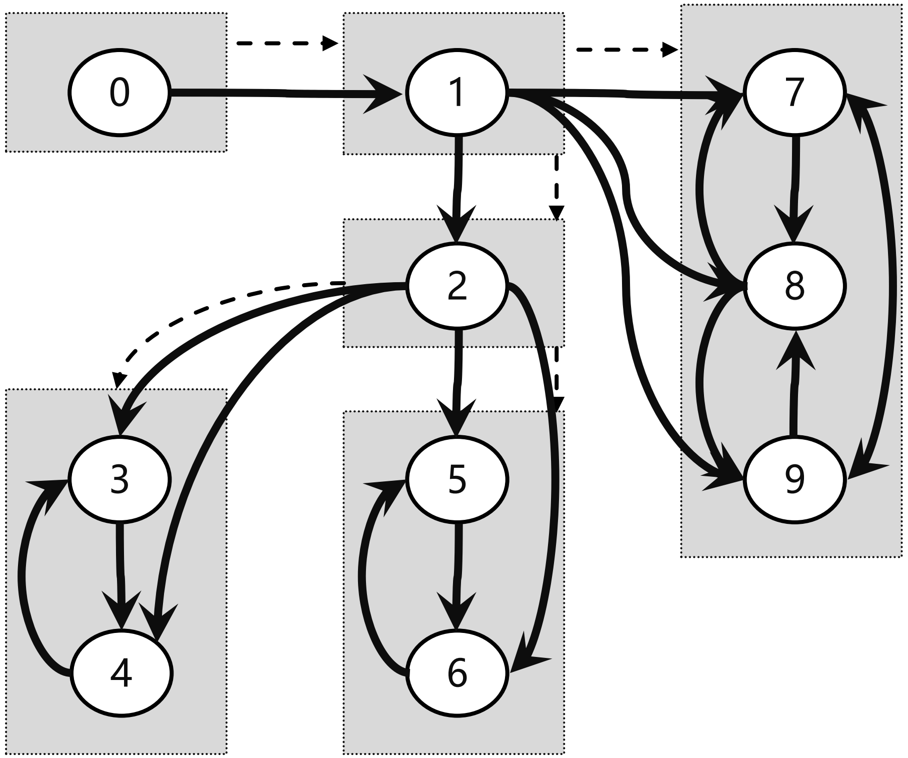
Constructing Incident-specific Event-causal Graph

Incident

- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event



Inter-service all connected along the service call direction

Intra-service bidirectional within each microservice

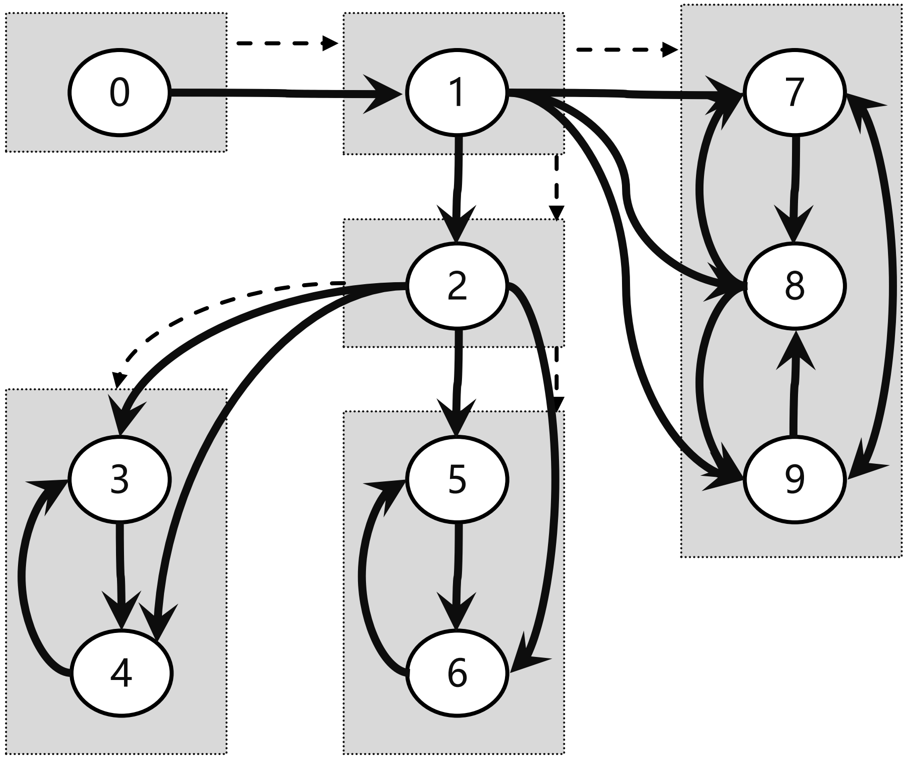
Constructing Incident-specific Event-causal Graph

Incident

- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event



Constructing Incident-specific Event-causal Graph

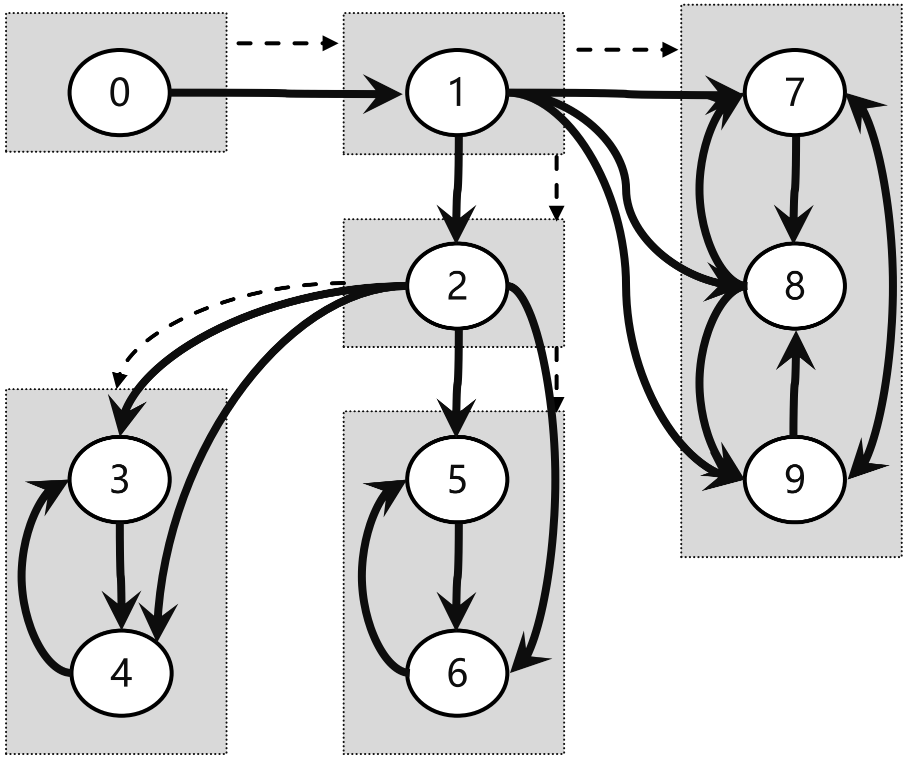
Incident

- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event

Incident-specific Event-causal Graph



Constructing Incident-specific Event-causal Graph

Incident

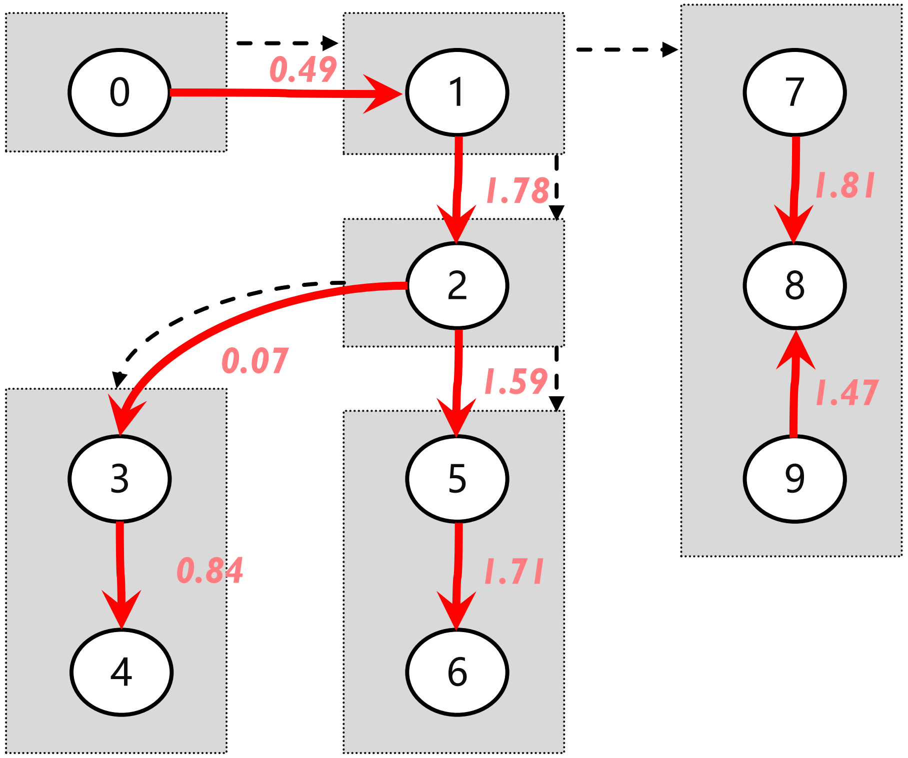
- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event

Incident-specific Event-causal Graph

- Weighted event causal links
  (eliminating false links with zero weights)



Constructing Incident-specific Event-causal Graph

Incident

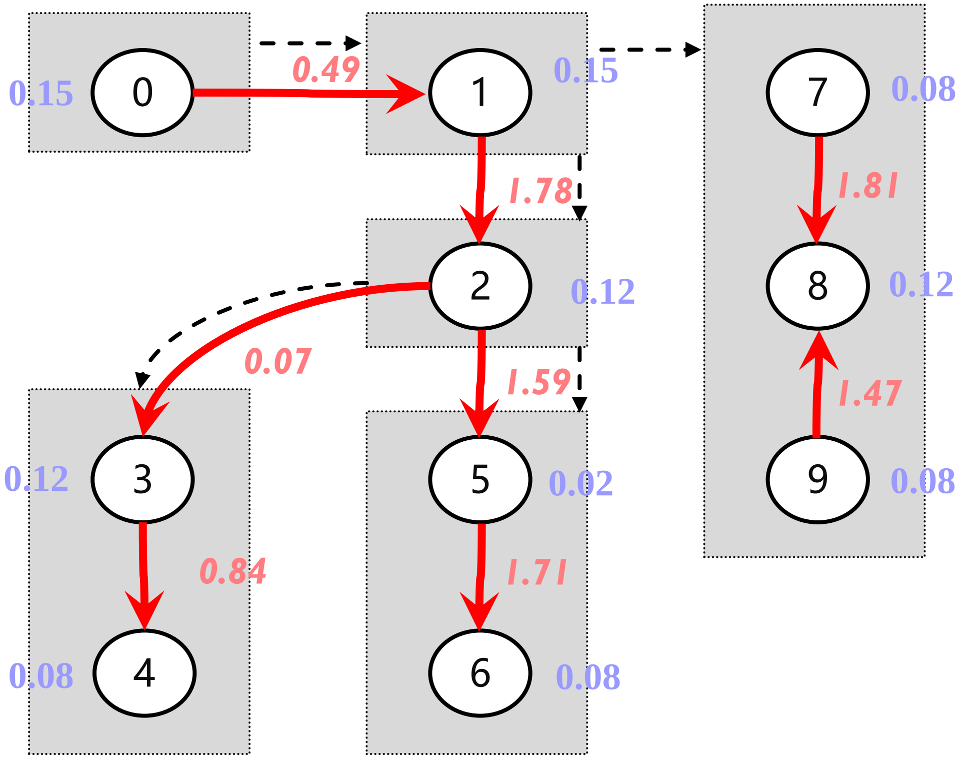
- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event

Incident-specific Event-causal Graph

- Weighted event causal links
  (eliminating false links with zero weights)
- Weighted event importance scores



Constructing Incident-specific Event-causal Graph

Incident

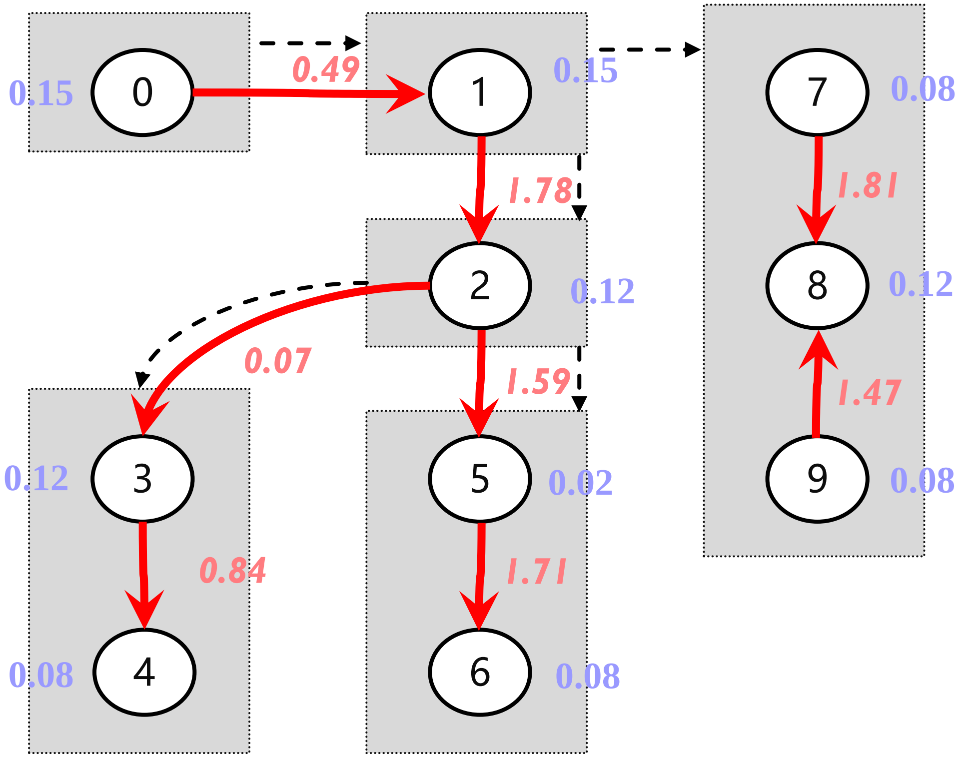
- A collection of events in a time window
- Service recorded in the events

Incident-specific Naïve Event-causal Graph

- Service call relationship
- All possible event causal links (unweighted)
  causal link: result event → cause event

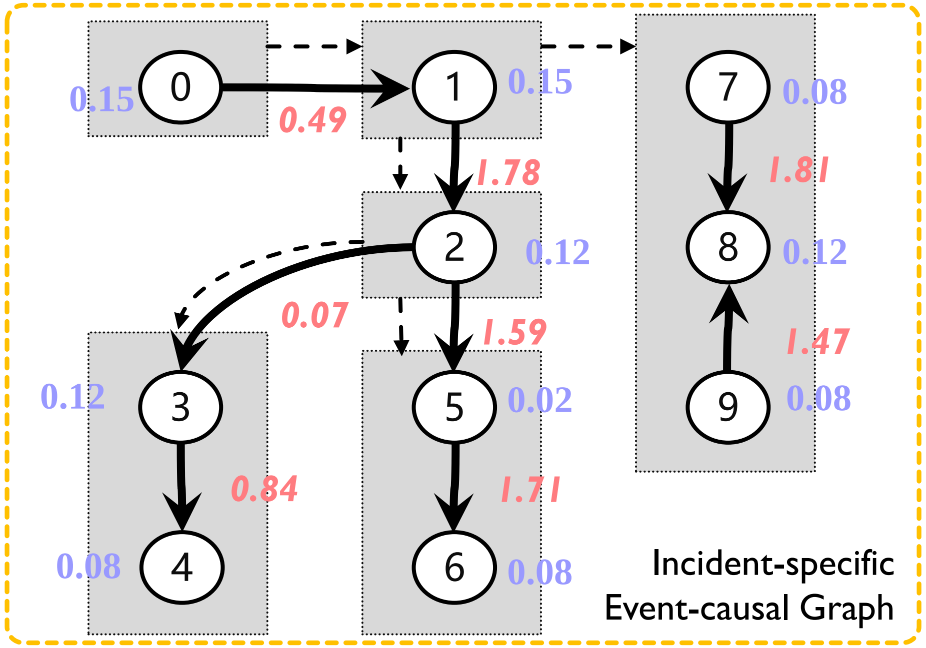
Incident-specific Event-causal Graph

- Weighted event causal links
  (eliminating false links with zero weights)
- Weighted event importance scores

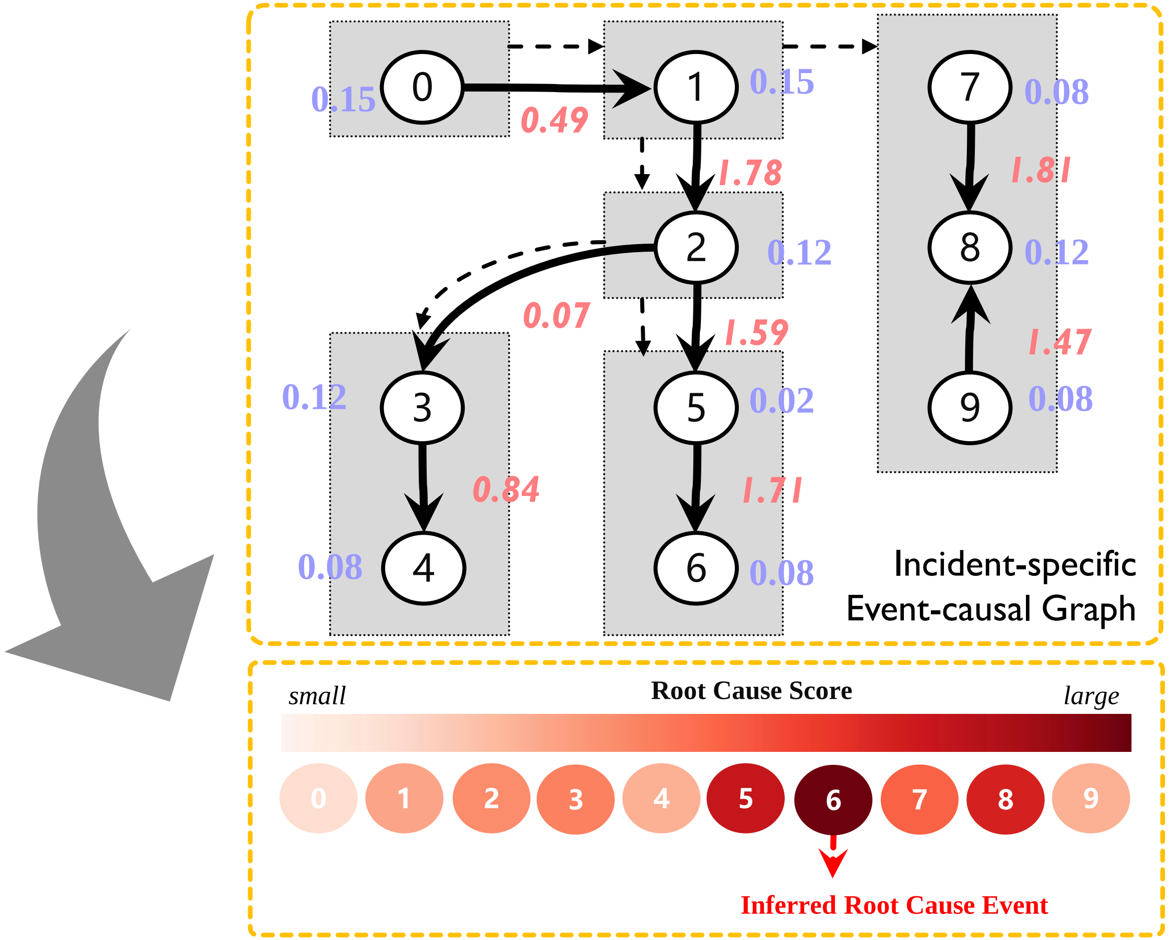


Incident-specific Event-casual Graph
(Subgraph of Overall Event-causal Graph)

Graph Ranking on Incident-specific Event-causal Graph

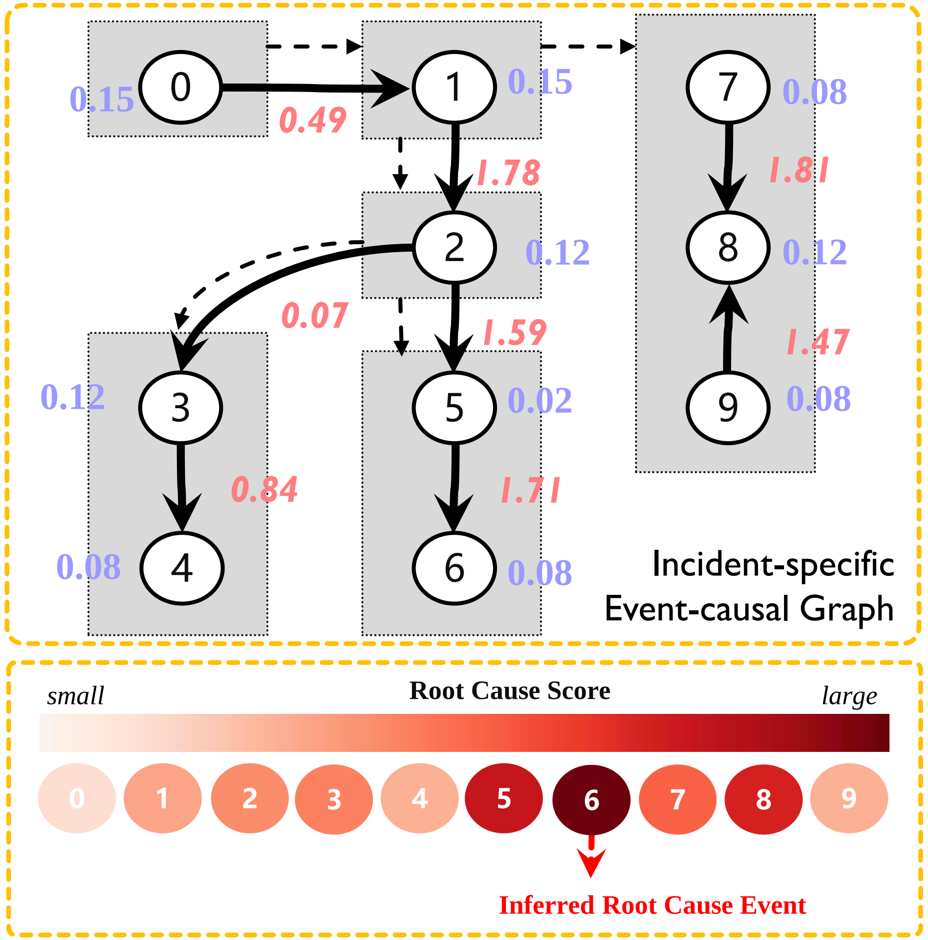


Graph Ranking on Incident-specific Event-causal Graph



Graph Ranking on Incident-specific Event-causal Graph

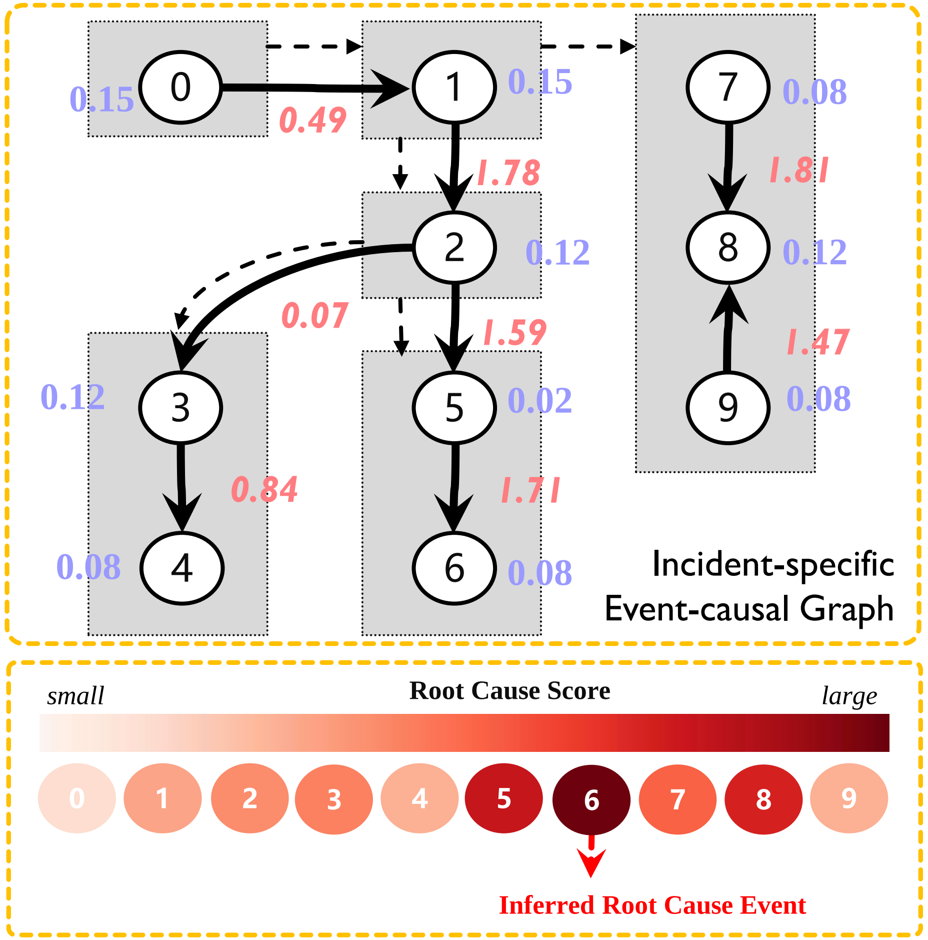
Graph Ranking



Graph Ranking on Incident-specific Event-causal Graph

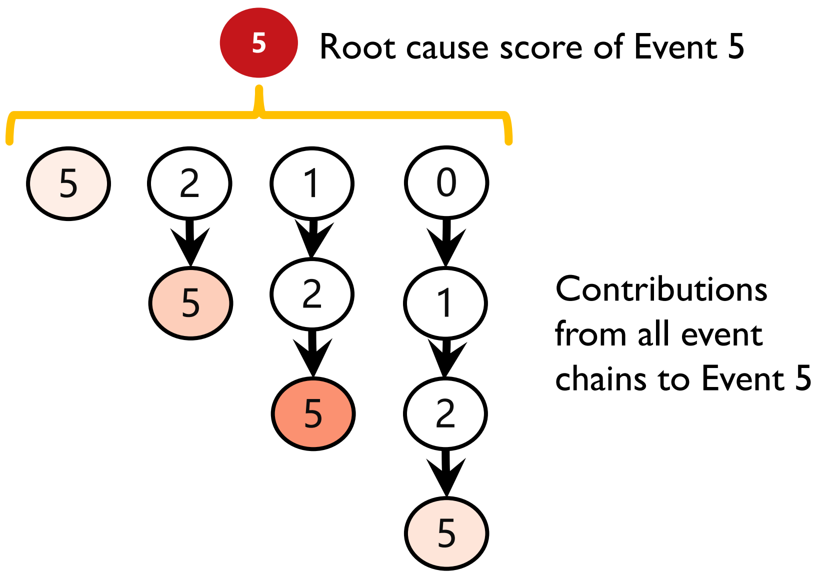
Graph Ranking

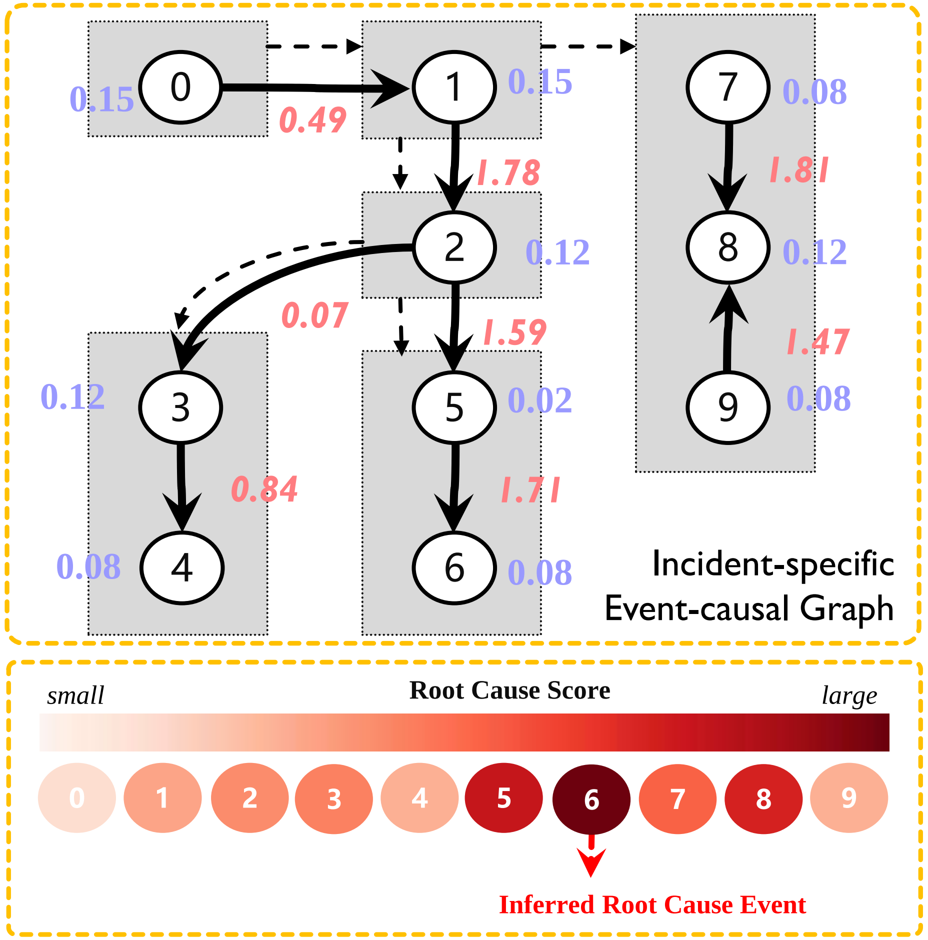
5 Root cause score of Event 5



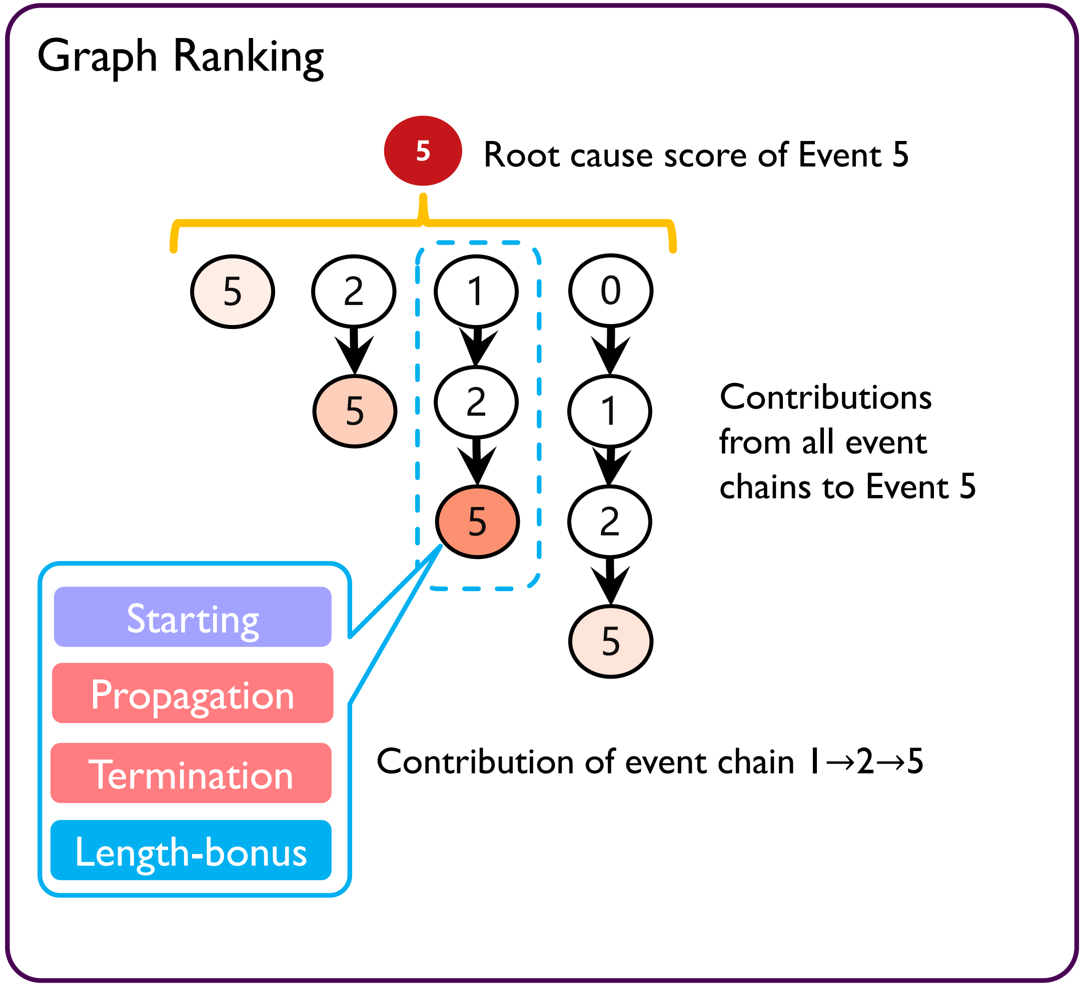
Graph Ranking on Incident-specific Event-causal Graph

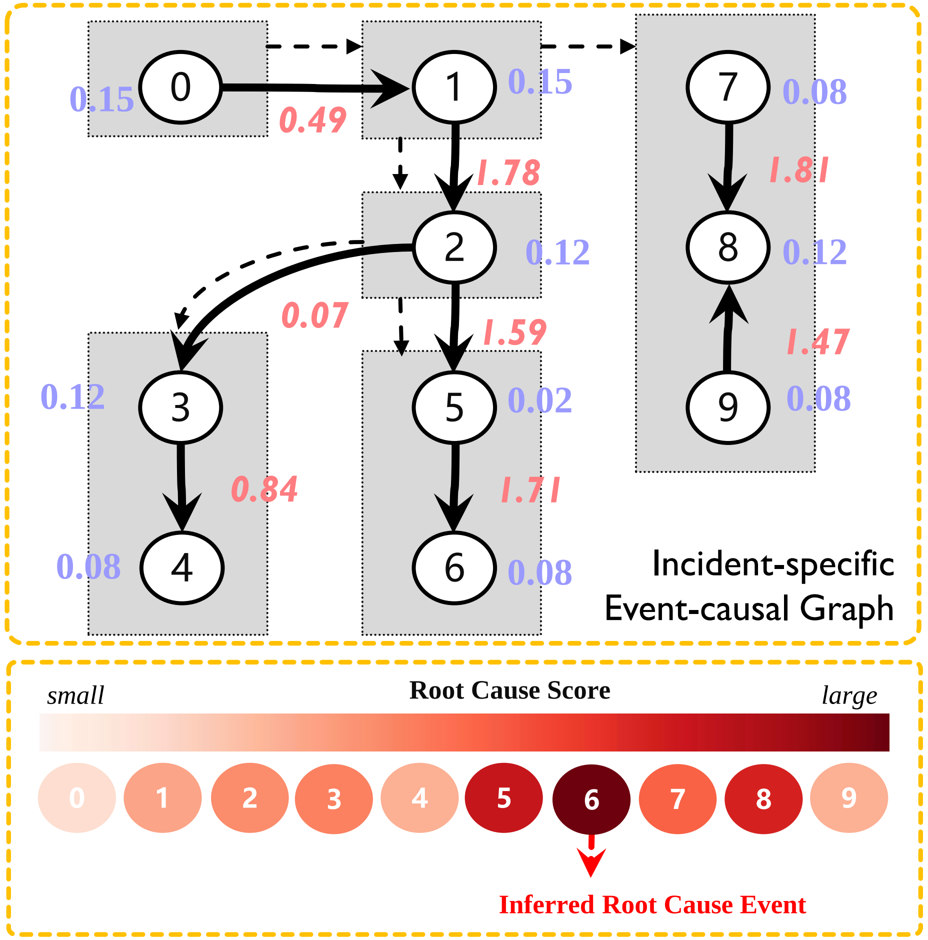
Graph Ranking



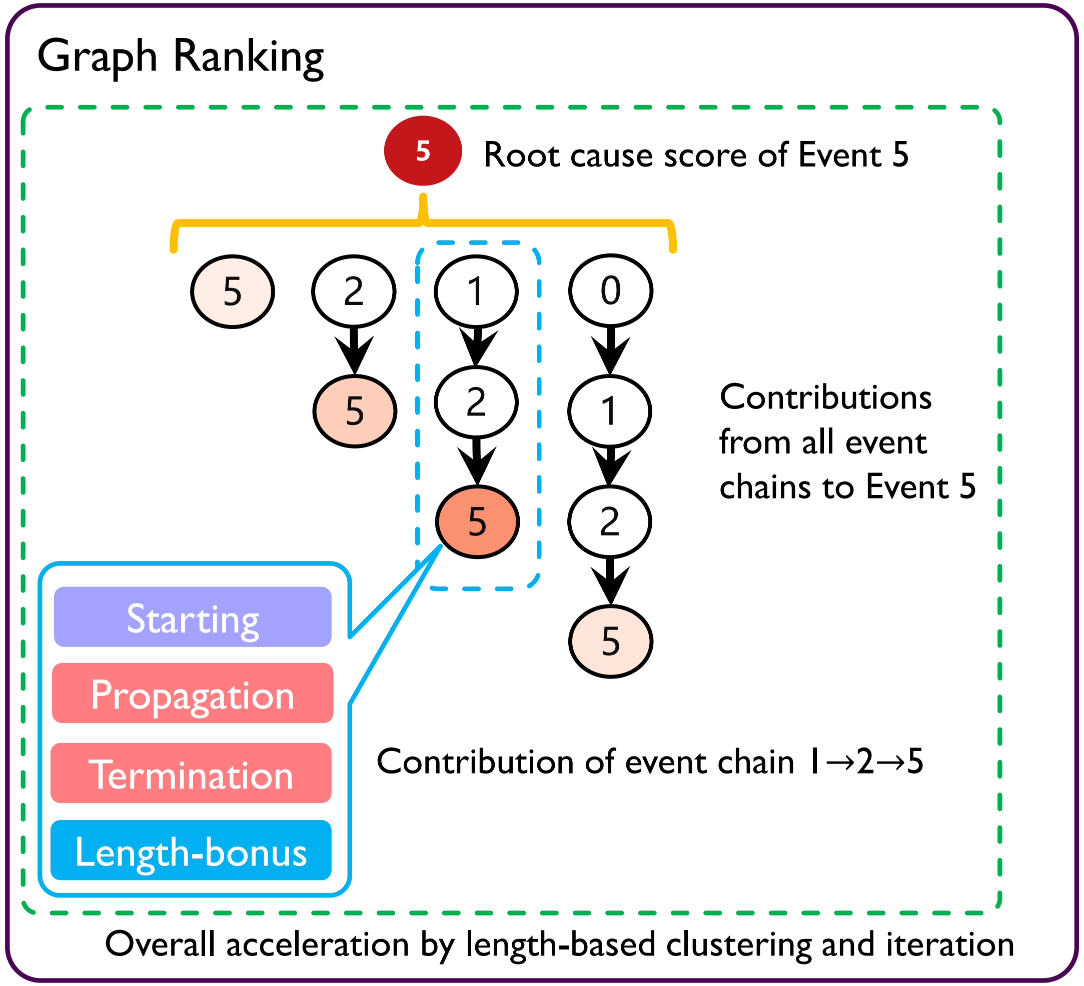


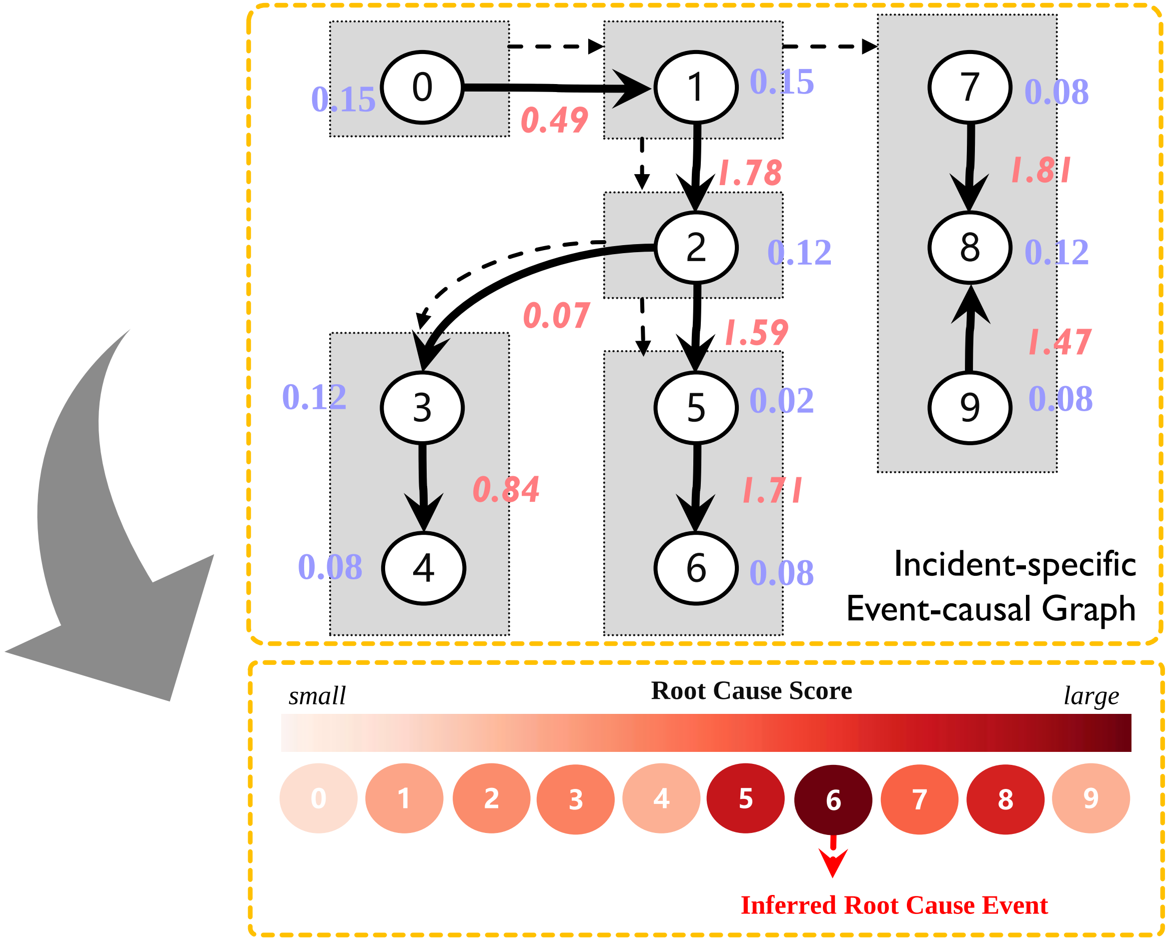
Graph Ranking on Incident-specific Event-causal Graph



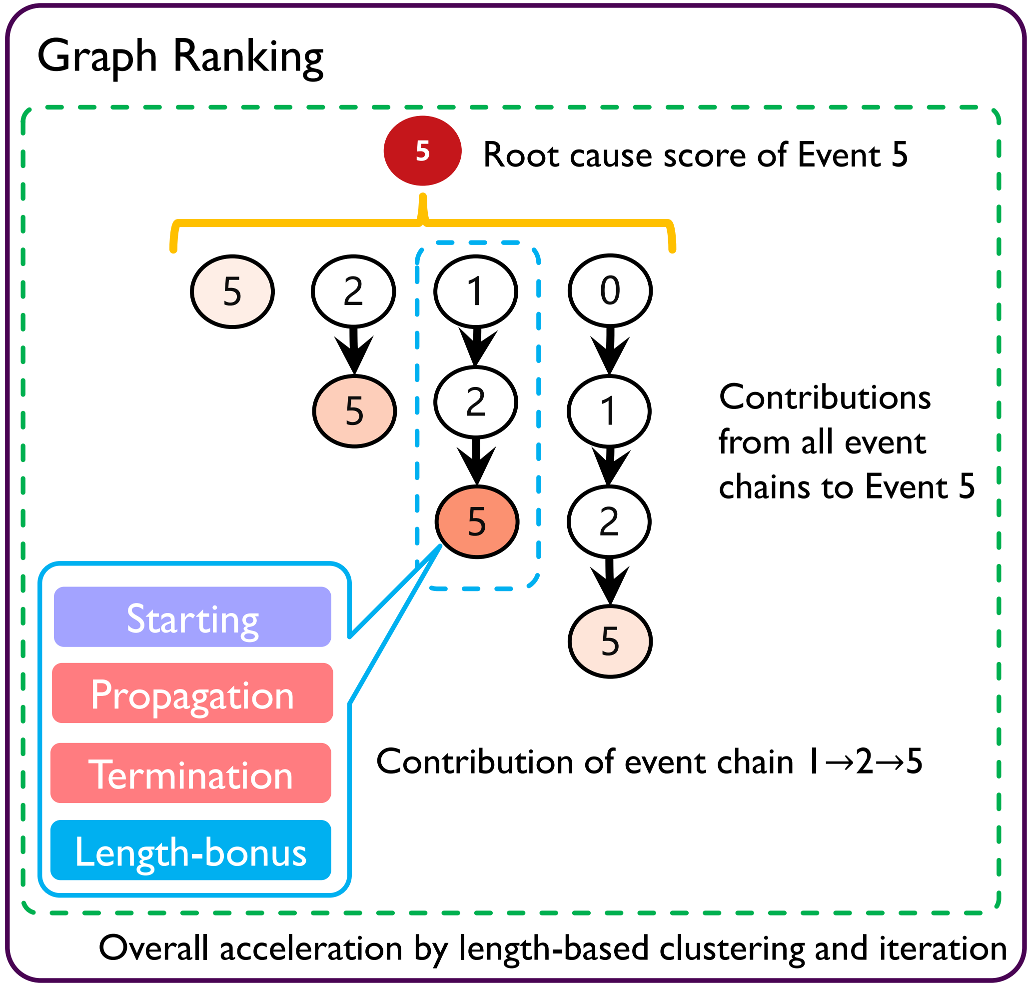


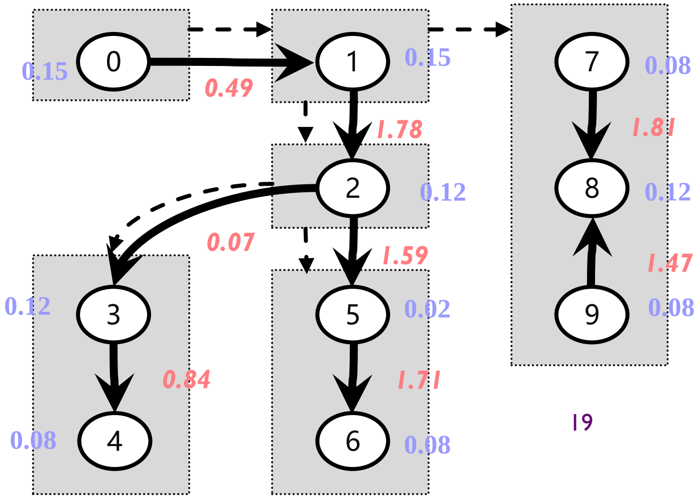
Graph Ranking on Incident-specific Event-causal Graph





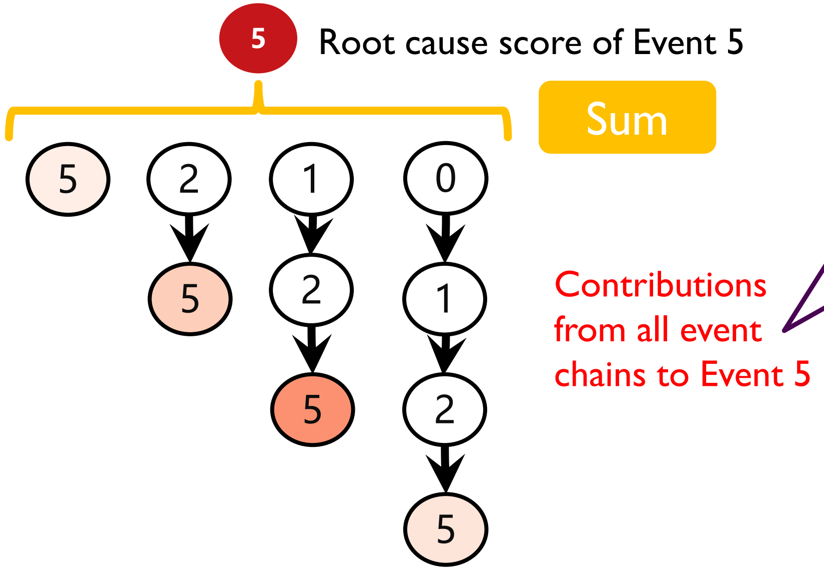
Graph Ranking in Detail

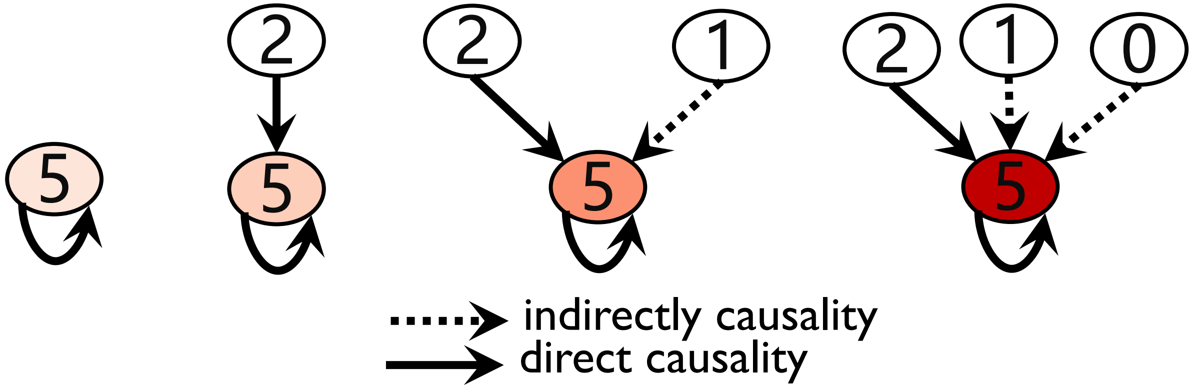




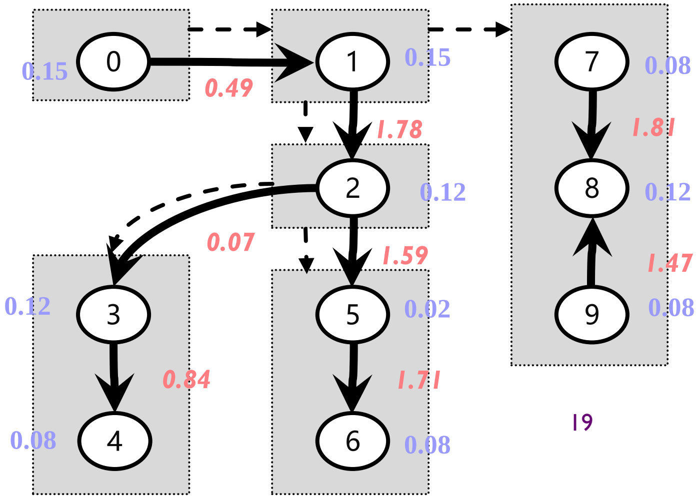
Graph Ranking in Detail

Graph Ranking



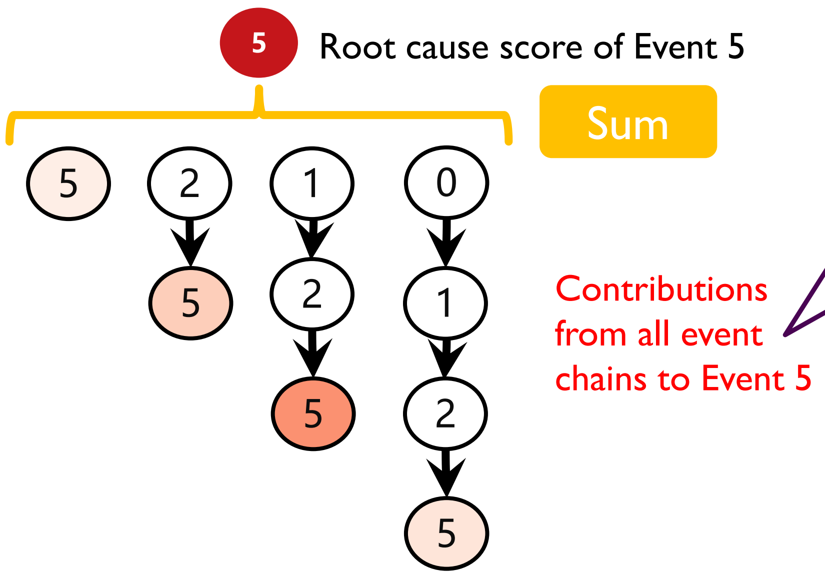


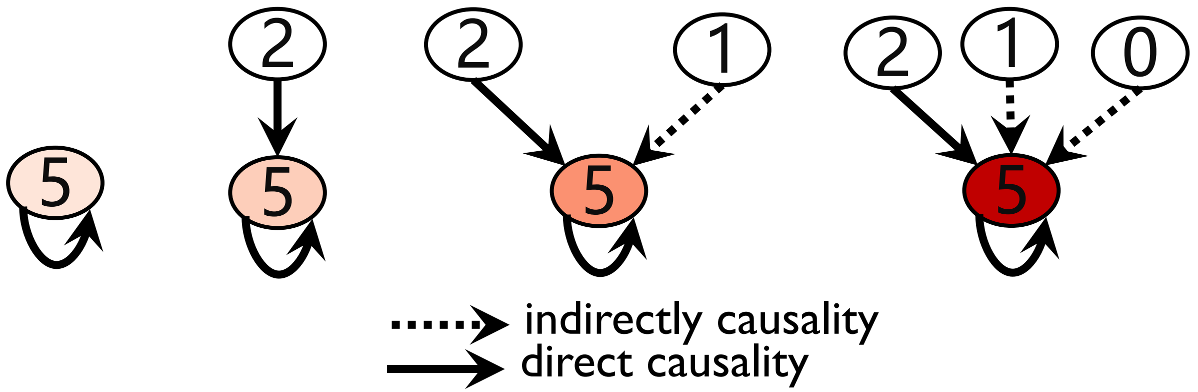
An event's suspiciousness increases as more related events are found in the incident



Graph Ranking in Detail

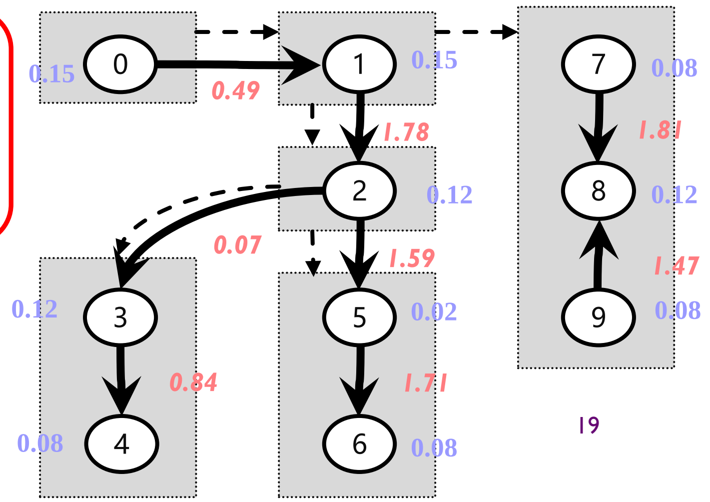
Graph Ranking





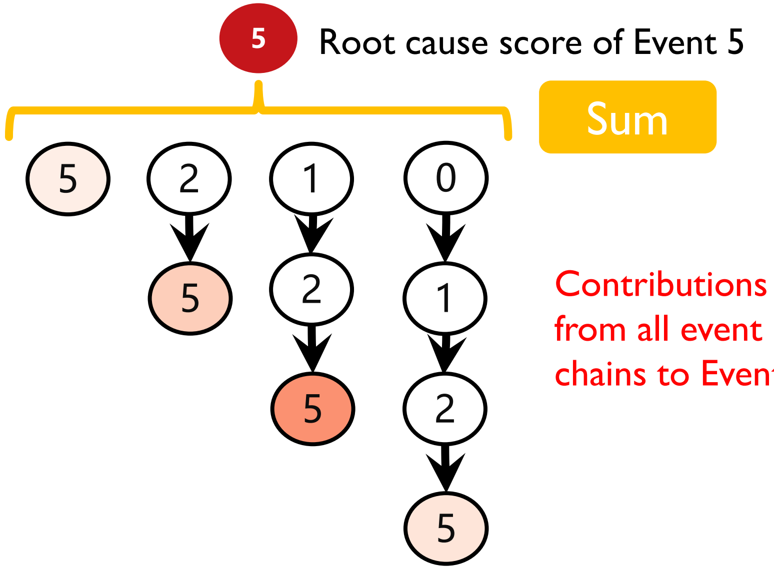
An event's suspiciousness increases as more related events are found in the incident

The existence of Event 5, 2, 1, 0 in this incident contributes a root cause score for Event 5

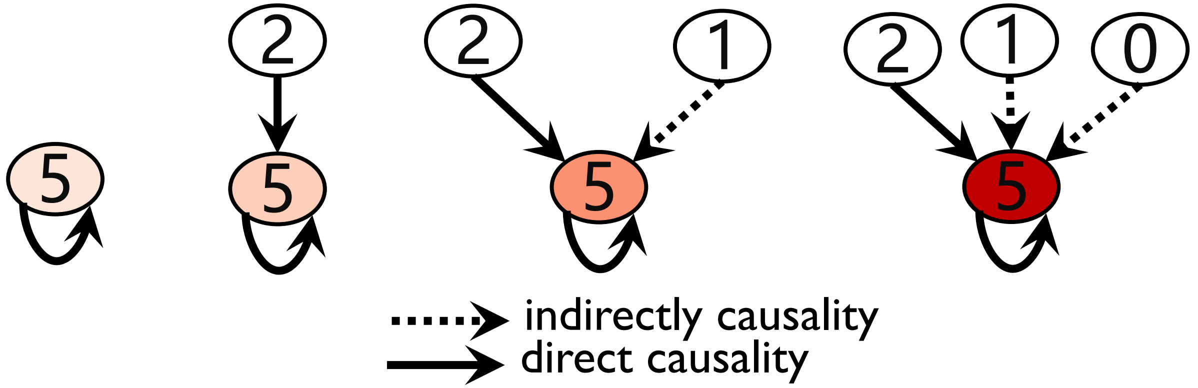


Graph Ranking in Detail

Graph Ranking

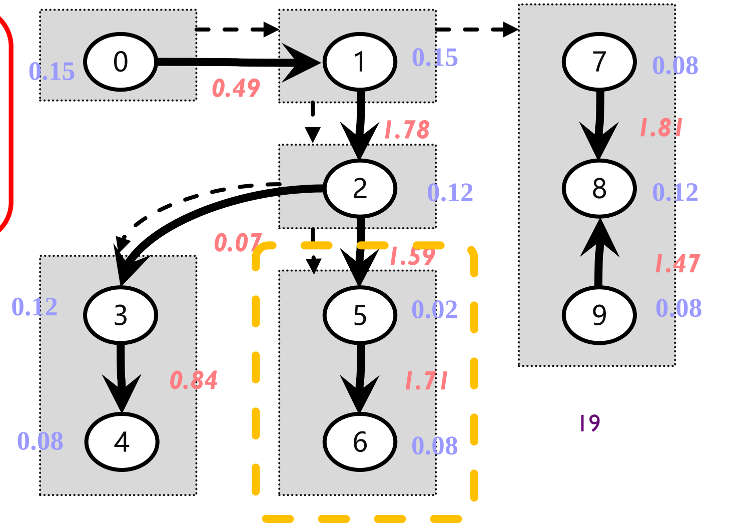


Contributions from all event chains to Event 5



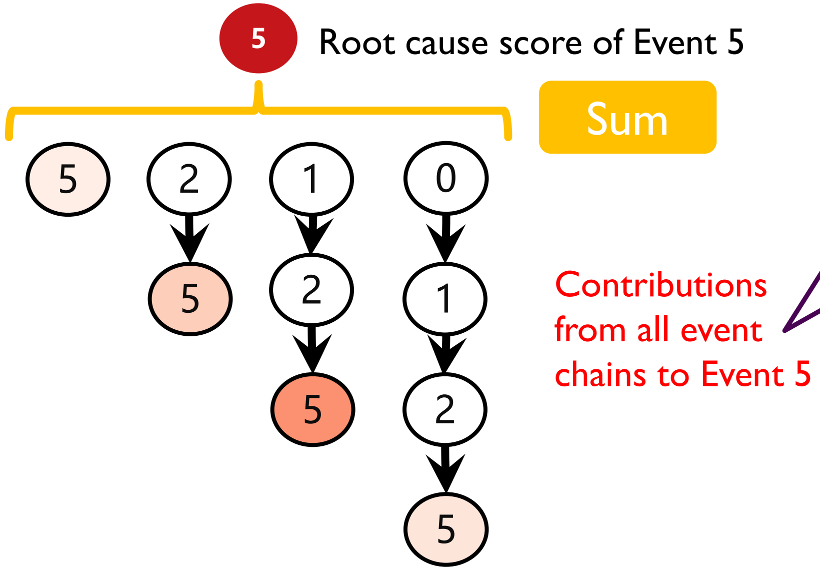
An event's suspiciousness increases as more related events are found in the incident

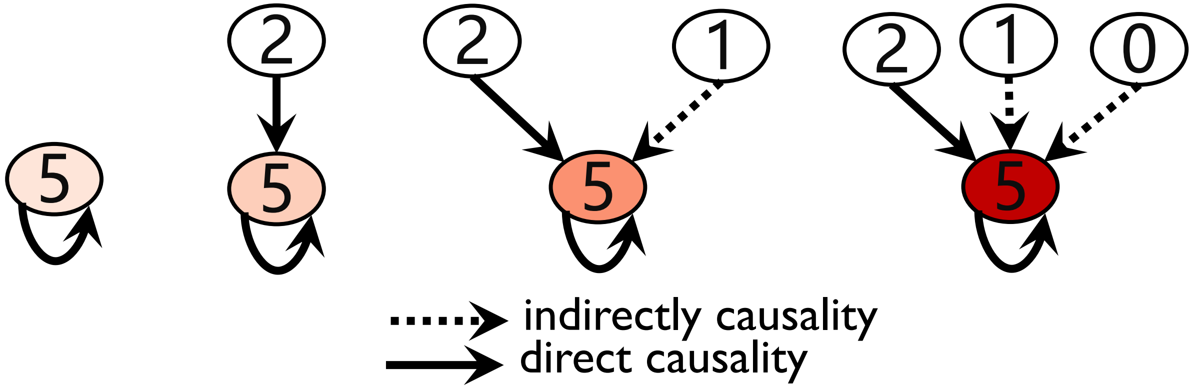
The existence of Event 5, 2, 1, 0 in this incident contributes a root cause score for Event 5



Graph Ranking in Detail

Graph Ranking

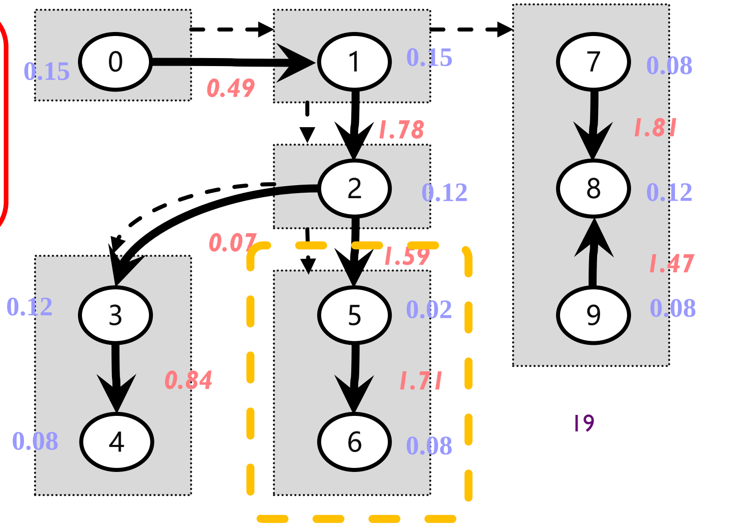




An event's suspiciousness increases as more related events are found in the incident

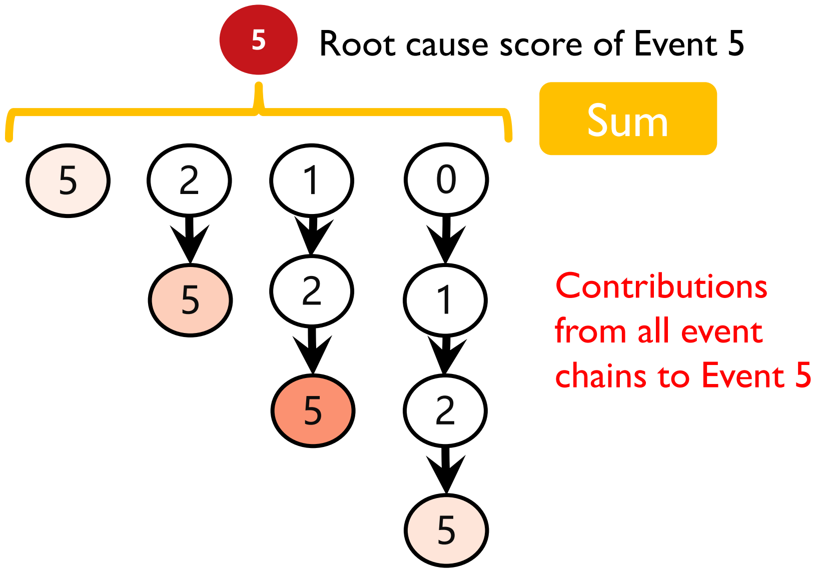
The existence of Event 5, 2, 1, 0 in this incident contributes a root cause score for Event 5

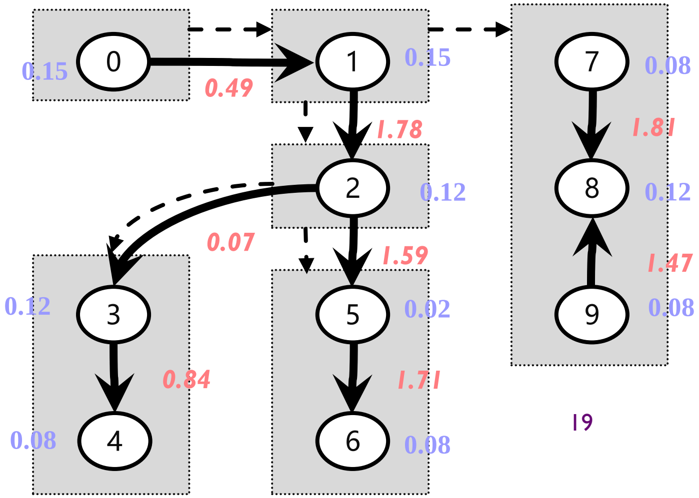
Event chains with broken (false) edges has zero contributions (e.g., 6→5)



Graph Ranking in Detail

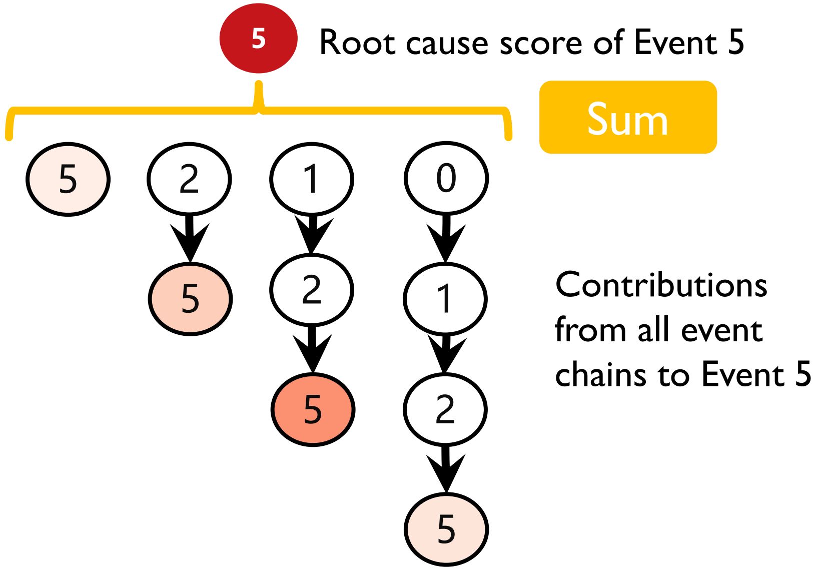
Graph Ranking

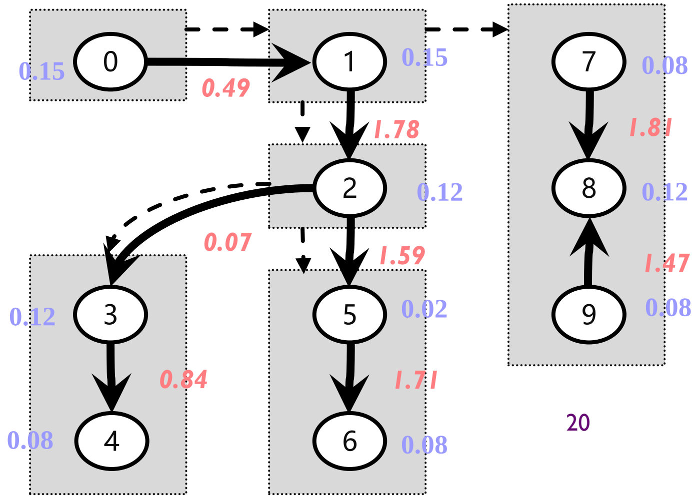




Graph Ranking in Detail

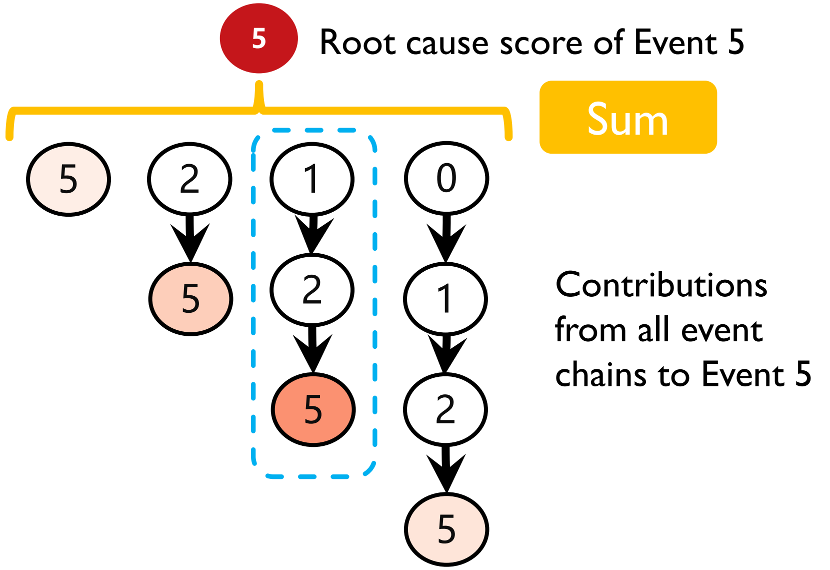
Graph Ranking

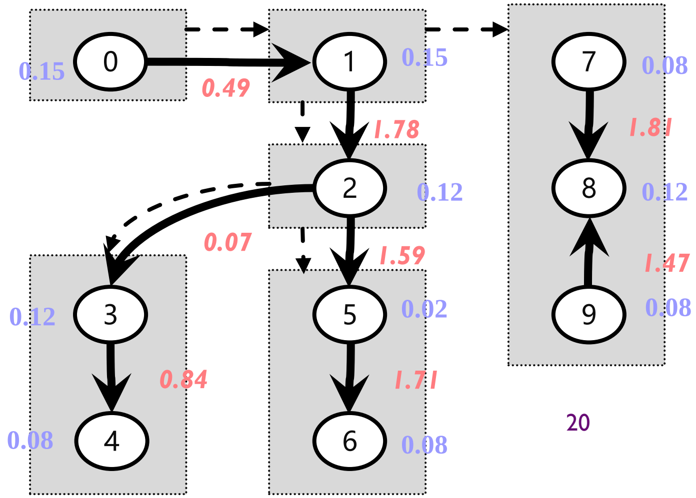




Graph Ranking in Detail

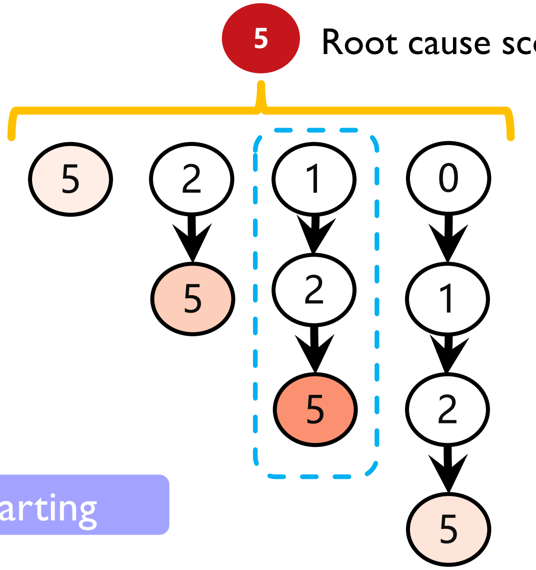
Graph Ranking





Graph Ranking in Detail

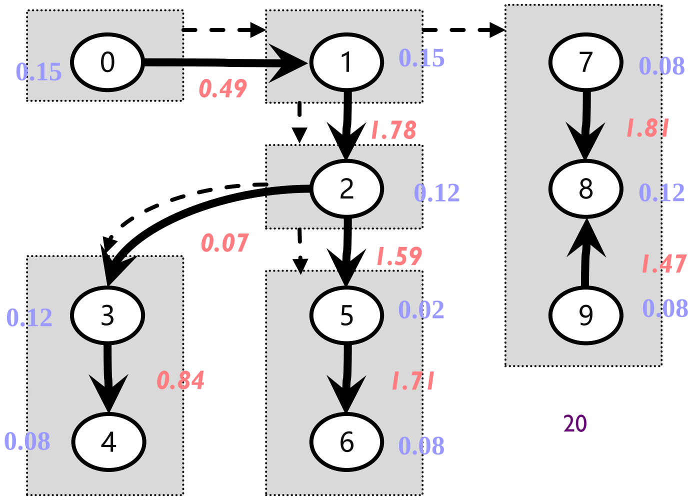
Graph Ranking



Starting

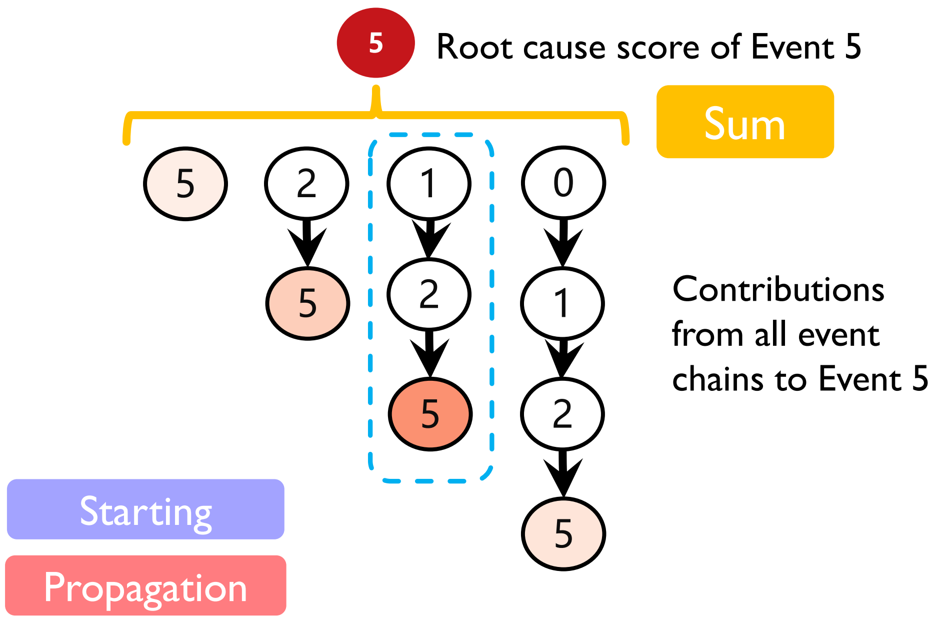
Starting

The importance score of starting event



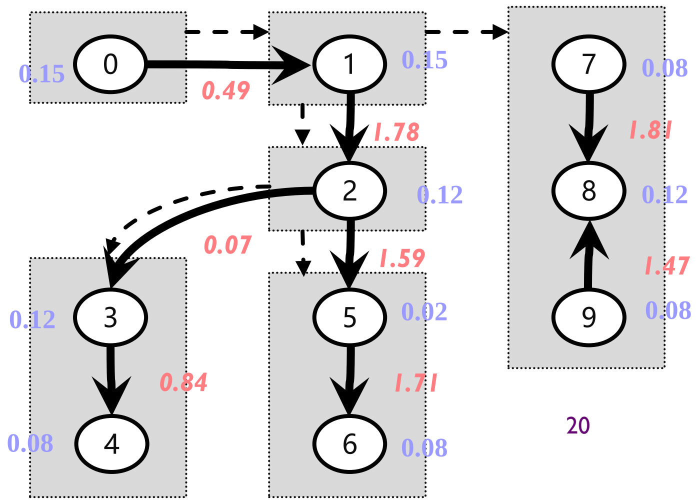
Graph Ranking in Detail

Graph Ranking



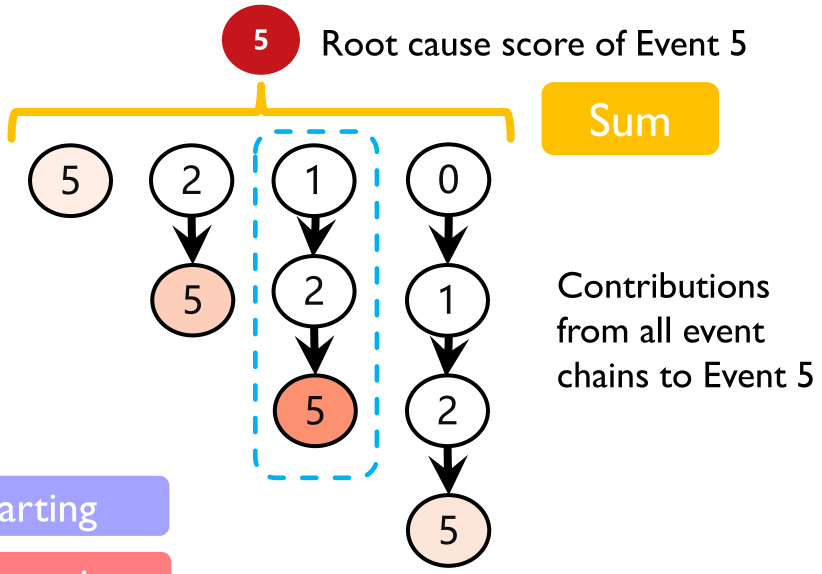
Starting — The importance score of starting event

Propagation — The joint probability of propagation
(Event 1 is caused by Event 2, while Event 2 is caused by Event 5)



Graph Ranking in Detail

Graph Ranking



Starting

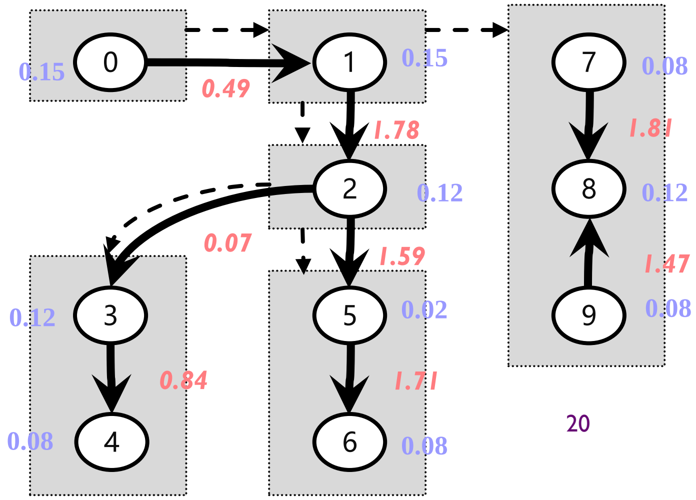
Propagation

Termination

Starting The importance score of starting event

Propagation The joint probability of propagation
(Event 1 is caused by Event 2, while Event 2 is caused by Event 5)

Termination The probability of termination at last node
(Event 5 is not caused by other events)

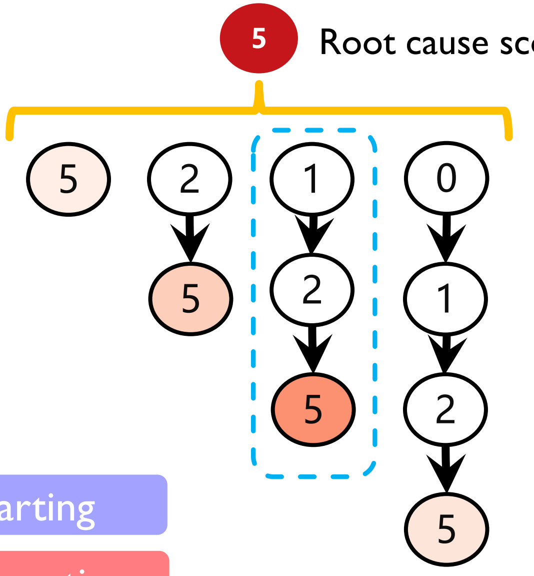


Graph Ranking in Detail

Graph Ranking

Root cause score of Event 5

Sum



Contributions from all event chains to Event 5

Starting

Propagation

Termination

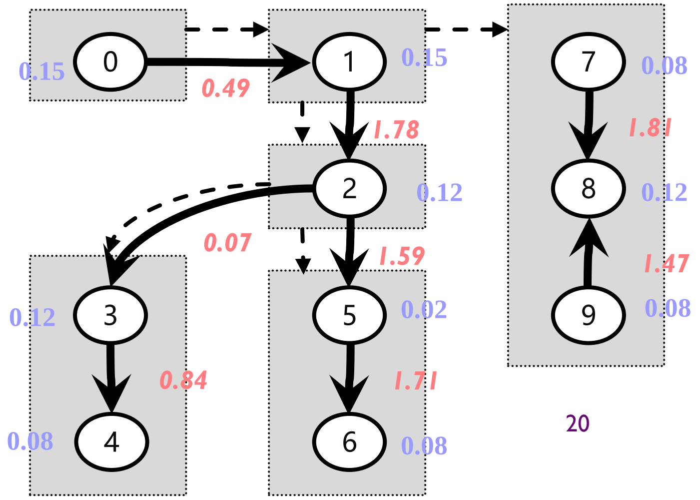
Length-bonus

Starting — The importance score of starting event

Propagation — The joint probability of propagation
(Event 1 is caused by Event 2, while Event 2 is caused by Event 5)

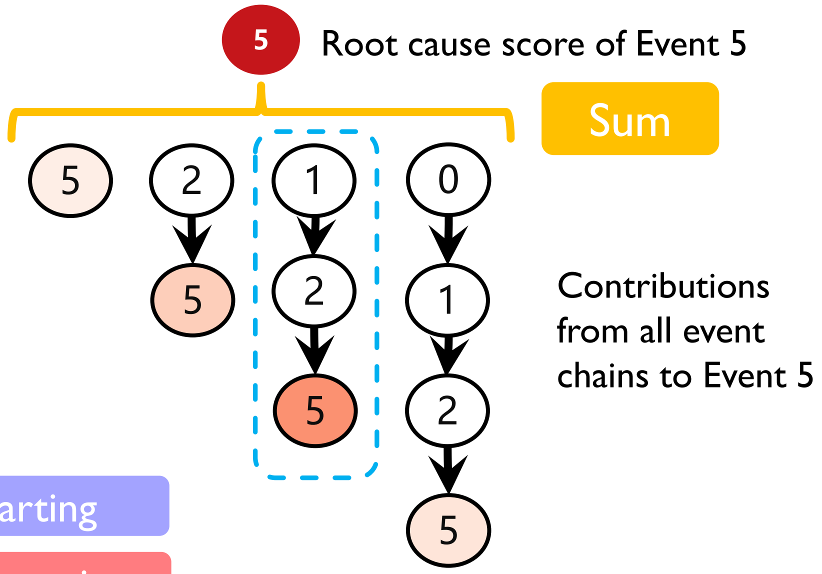
Termination — The probability of termination at last node
(Event 5 is not caused by other events)

Length-bonus — Avoid eliminating long chains



Graph Ranking in Detail

Graph Ranking



Starting

Propagation

Termination

Length-bonus

Starting The importance score of starting event $S(v_1)$

Propagation The joint probability of propagation

(Event 1 is caused by Event 2, while Event 2 is caused by Event 5)

$$p(v_2|v_1)p(v_5|v_2)$$

Termination The probability of termination at last node

(Event 5 is not caused by other events) $1 - p(v_6|v_5)$

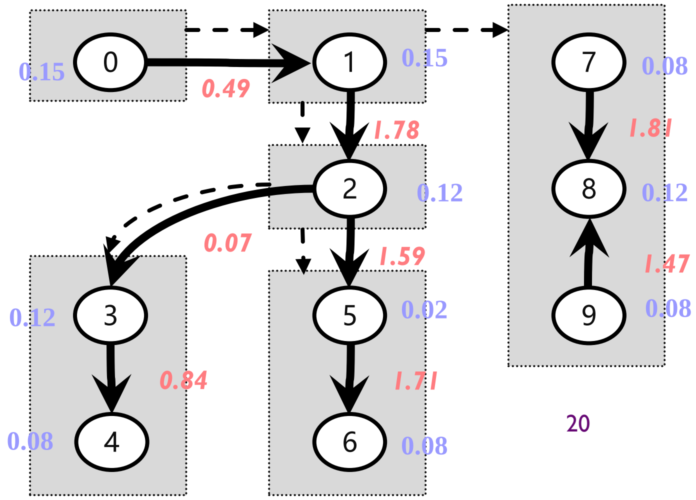
Length-bonus Avoid eliminating long chains

$$\mathrm{LB}(3)$$

$$p(v_i|v_j) = \frac{E[e_{(j,i)}]}{\sum_t E[e_{(t,i)}] + k}$$

$$k = \alpha * mean(E[e_{(j,i)}])$$

$$\mathrm{LB}(i) = \min(1, 0.01 * 2^{i+1})$$

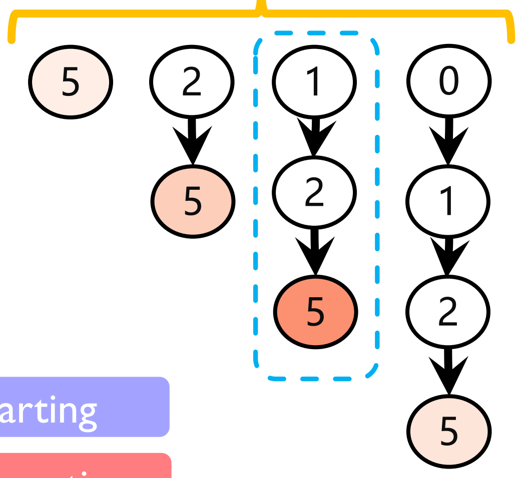


Graph Ranking in Detail

Graph Ranking

Root cause score of Event 5

Sum



Contributions from all event chains to Event 5

Starting

Propagation

Termination

Length-bonus

Product

Starting The importance score of starting event $S(v_1)$

Propagation The joint probability of propagation
(Event 1 is caused by Event 2, while Event 2 is caused by Event 5)
$p(v_2|v_1)p(v_5|v_2)$

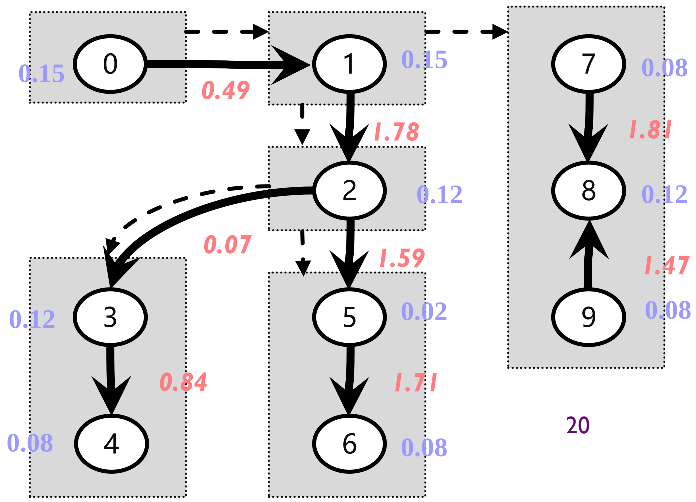
Termination The probability of termination at last node
(Event 5 is not caused by other events) $1-p(v_6|v_5)$

Length-bonus Avoid eliminating long chains $\mathrm{LB}(3)$

$$p(v_i|v_j) = \frac{E[e_{(j,i)}]}{\sum_t E[e_{(t,i)}] + k}$$

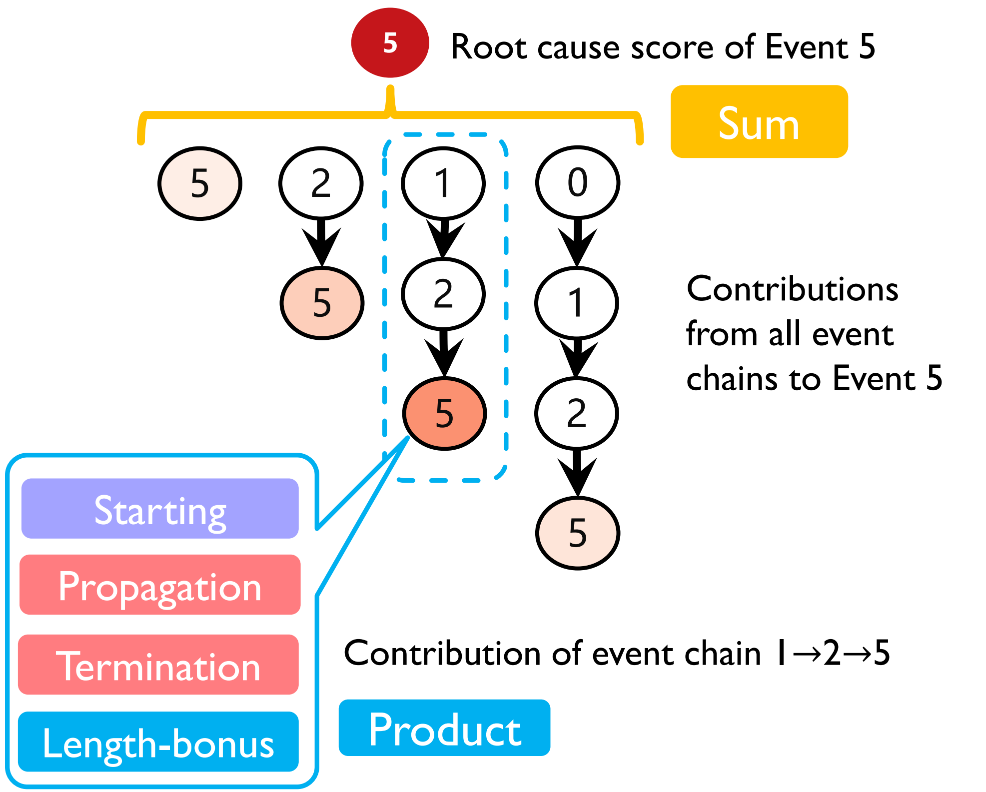
$$k = \alpha * mean(E[e_{(j,i)}])$$

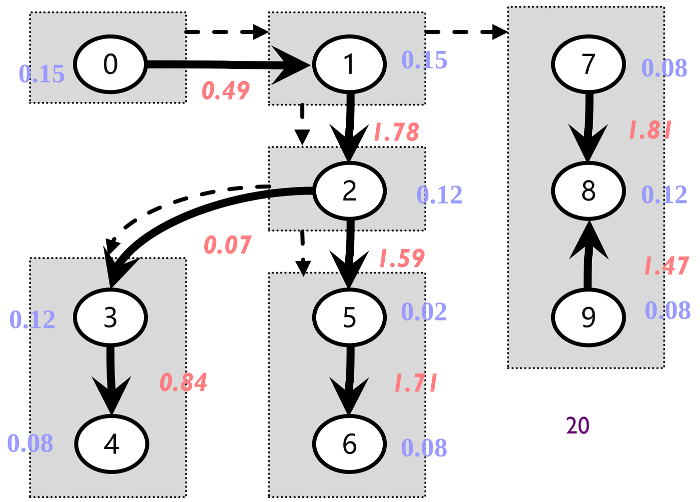
$$\mathrm{LB}(i) = \min(1, 0.01 * 2^{i+1})$$



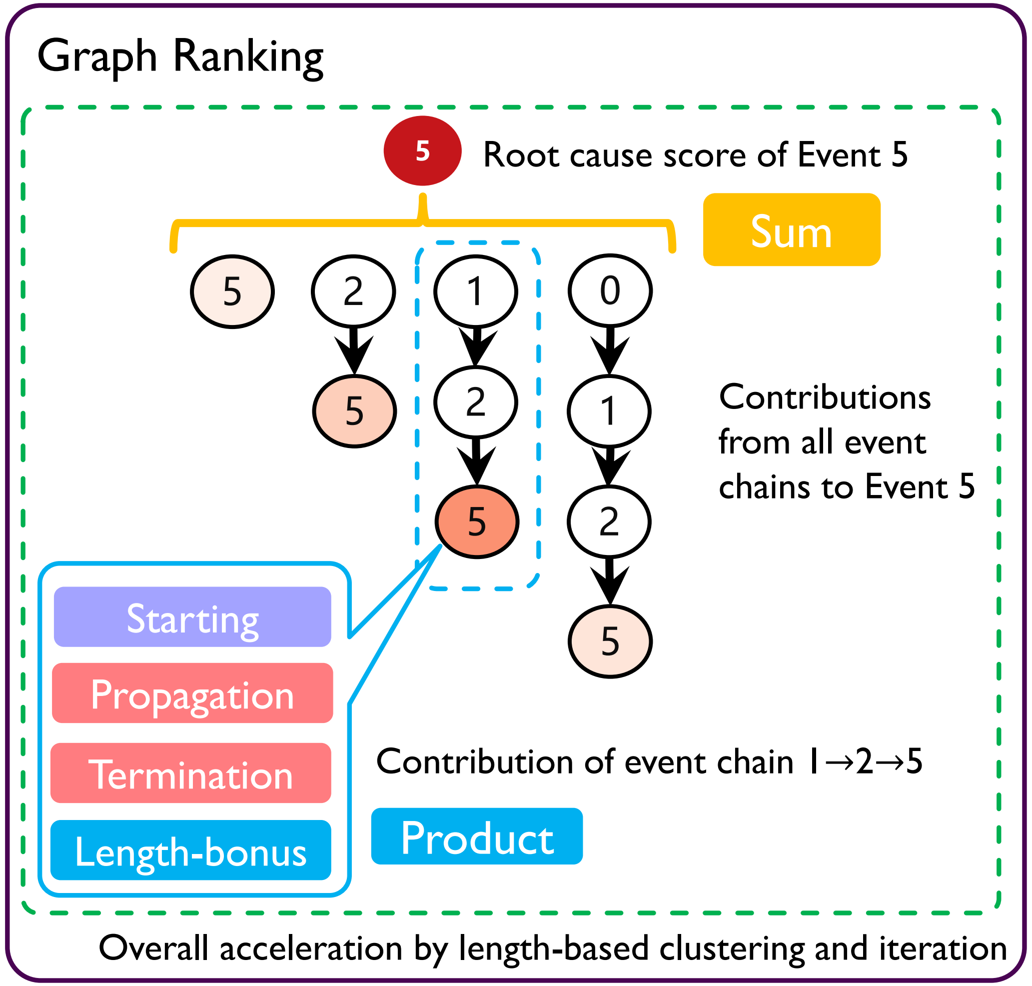
Graph Ranking in Detail

Graph Ranking

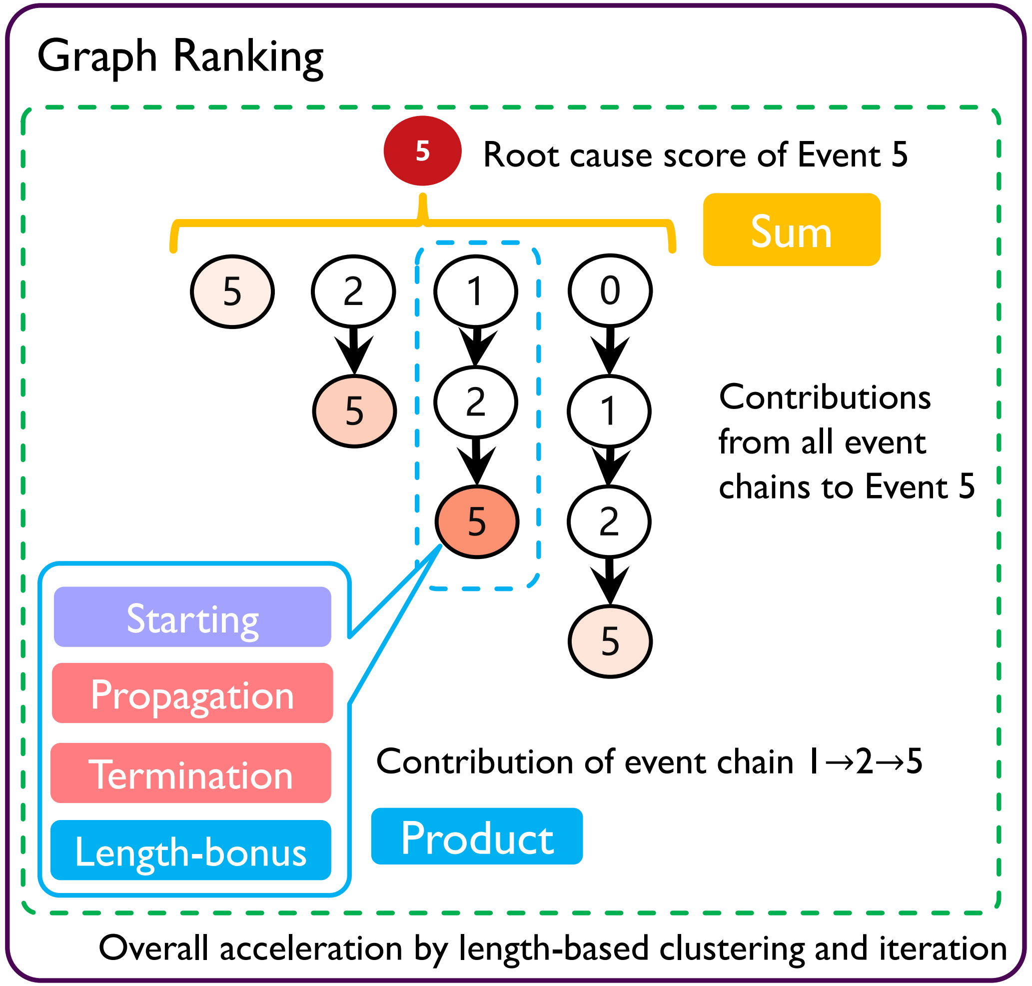




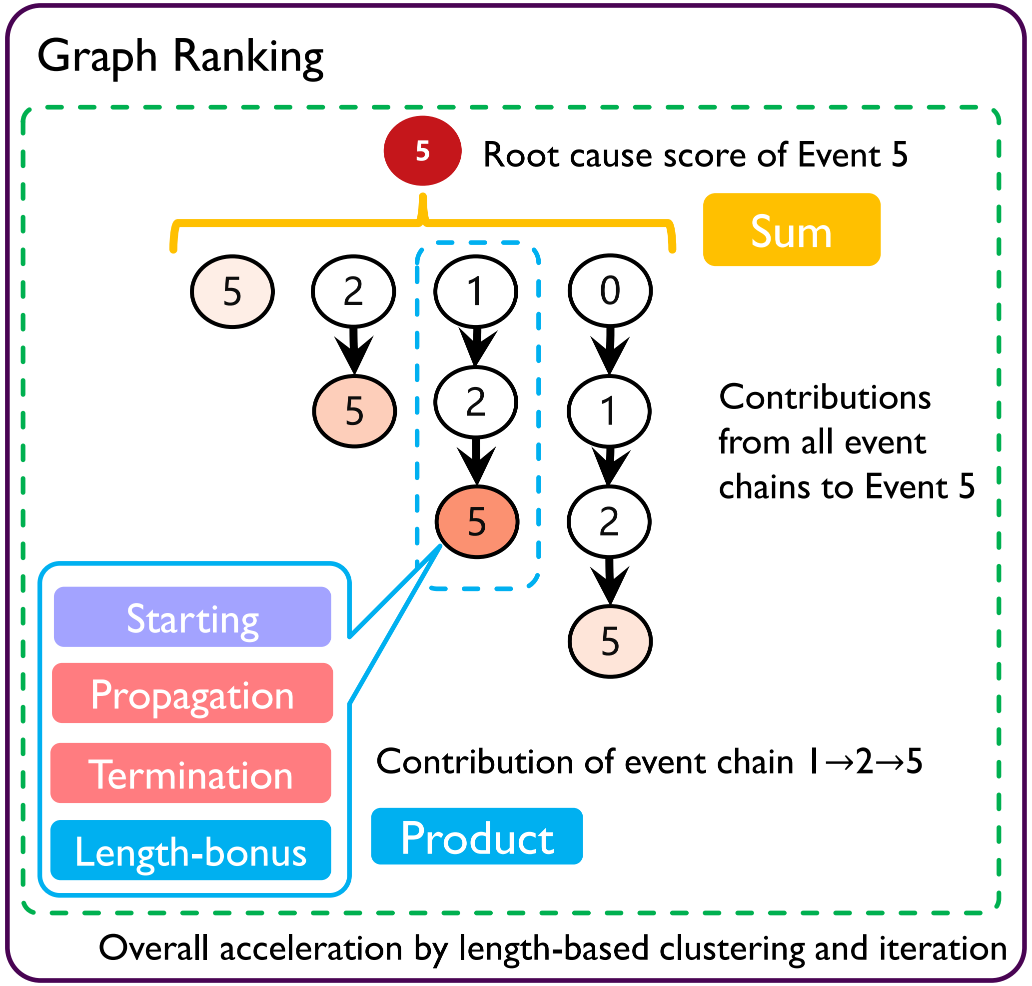
Graph Ranking in Detail



Graph Ranking in Detail

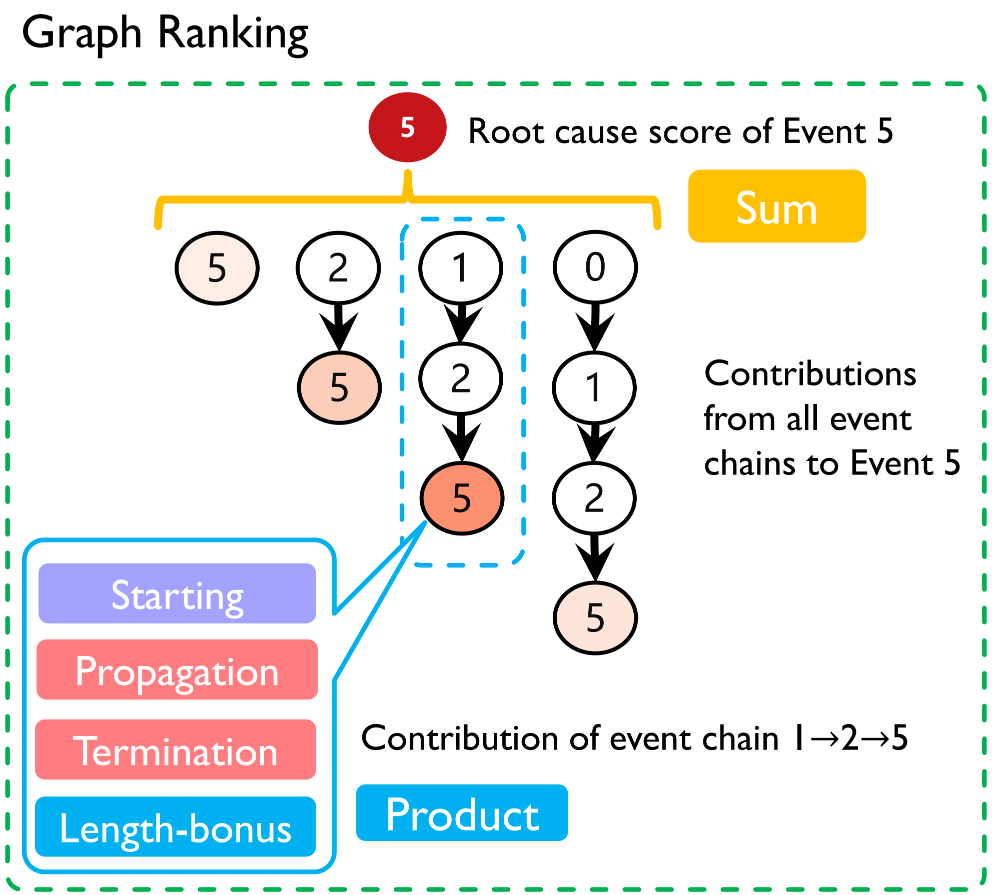


Graph Ranking in Detail



(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$

Graph Ranking in Detail



Overall acceleration by length-based clustering and iteration

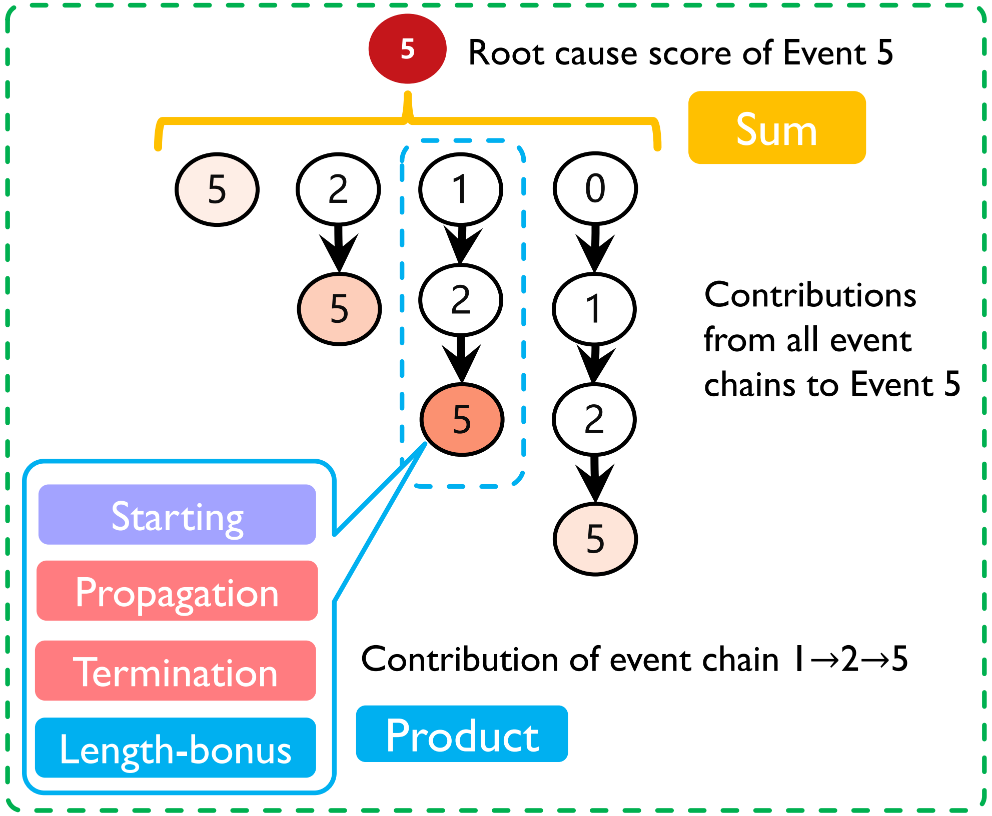
(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$

(Optimization) For an event v, assume an n-length chain l_i is $v_{d_1^{(i)}}, v_{d_2^{(i)}}, \dots, v_{d_n^{(i)}}, v_{d_n^{(i)}}$ is v, then all the chain contributions to v is:

$$score(v) = \sum_{l_i} S(v_{d_1^{(i)}}) \prod_j p(v_{d_{j+1}^{(i)}} | v_{d_j^{(i)}}) \; Term(v) LB(n)$$

Graph Ranking in Detail

Graph Ranking



Overall acceleration by length-based clustering and iteration

(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$ ☹

(Optimization) For an event v, assume an n-length chain l_i is $v_{d_1^{(i)}}, v_{d_2^{(i)}}, \ldots, v_{d_n^{(i)}}, v_{d_n^{(i)}}$ is v, then all the chain contributions to v is:

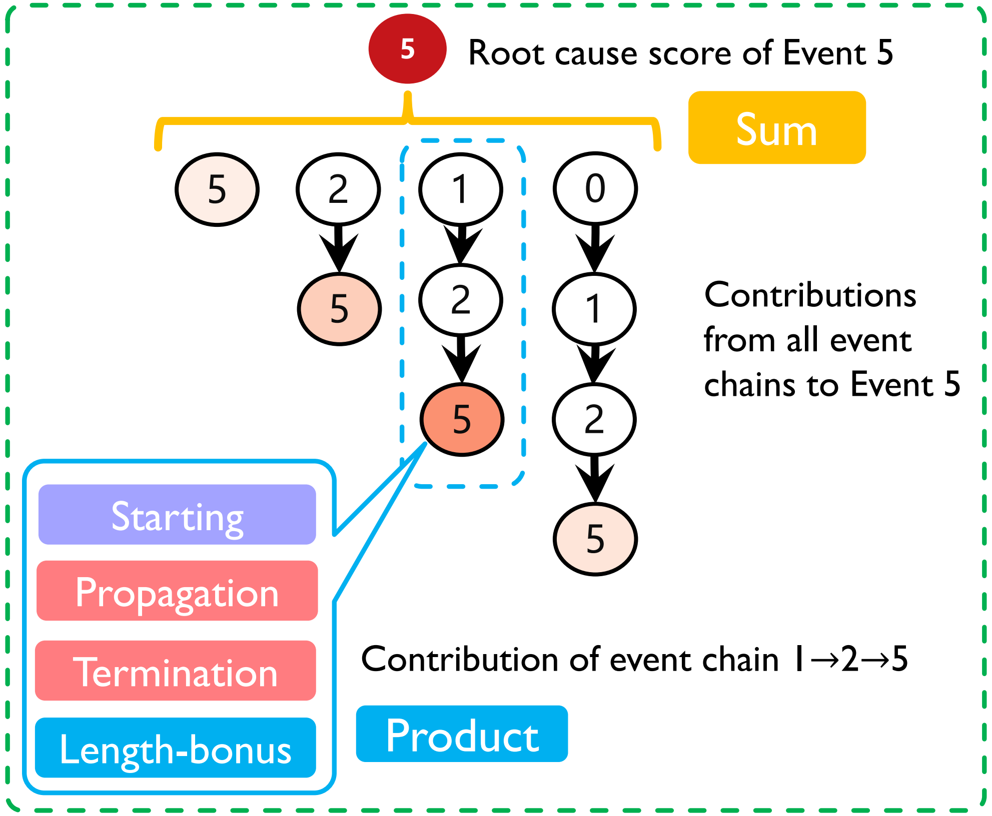
$$score(v) = \sum_{n} \underbrace{\sum_{l_i} S\left(v_{d_1^{(i)}}\right) \prod_j p\left(v_{d_{j+1}^{(i)}} \middle| v_{d_j^{(i)}}\right)}_{Q^n[v]} Term(v) LB(n)$$

Partial contribution of all the n-length chain

$$\boldsymbol{Q}^{n+1} = \boldsymbol{Q}^n \boldsymbol{P}$$

Graph Ranking in Detail

Graph Ranking



Overall acceleration by length-based clustering and iteration

(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$

(Optimization) For an event v, assume an n-length chain l_i is $v_{d_1^{(i)}}, v_{d_2^{(i)}}, \dots, v_{d_n^{(i)}}, v_{d_n^{(i)}}$ is v, then all the chain contributions to v is:

$$score(v) = \underbrace{\sum_{l_i} S\left(v_{d_1^{(i)}}\right) \prod_j p\left(v_{d_{j+1}^{(i)}} \middle| v_{d_j^{(i)}}\right)}_{\sum_n \quad Q^n[v]} Term(v) LB(n)$$

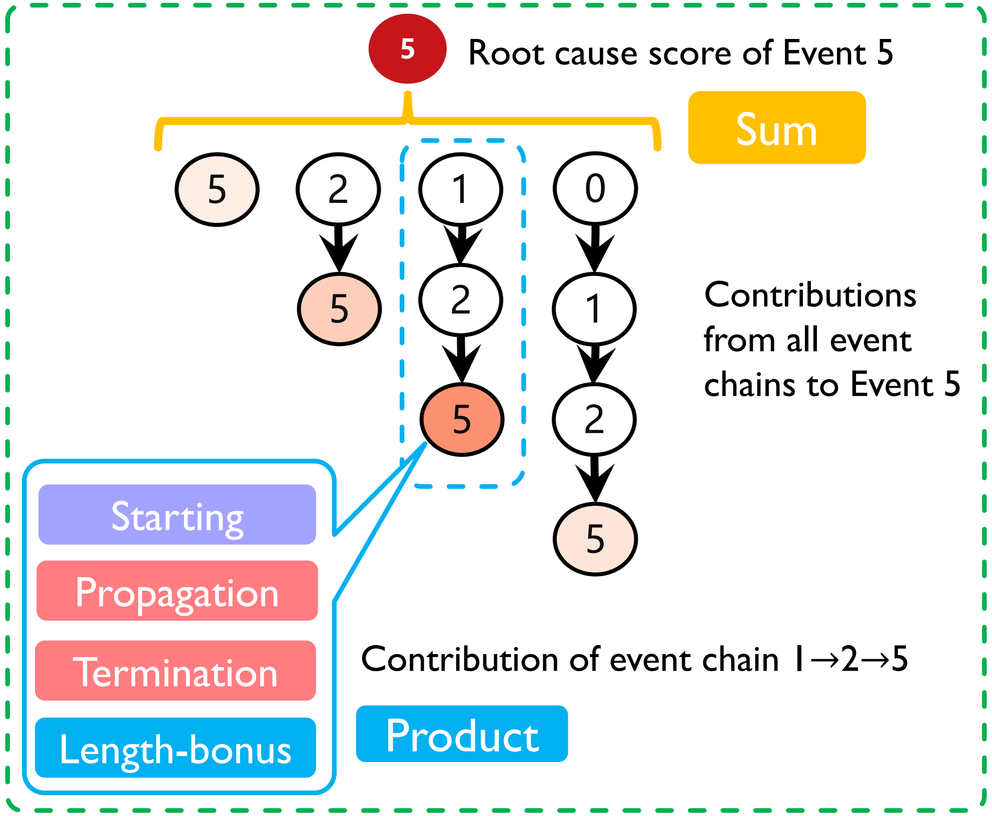
Partial contribution of all the n-length chain

$$\boldsymbol{Q}^{n+1} = \boldsymbol{Q}^n \boldsymbol{P}$$

Complexity is $O(mT)$

Graph Ranking in Detail

Graph Ranking



Overall acceleration by length-based clustering and iteration

(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$

(Optimization) For an event v, assume an n-length chain l_i is $v_{d_1^{(i)}}, v_{d_2^{(i)}}, \dots, v_{d_n^{(i)}}, v_{d_n^{(i)}}$ is v, then all the chain contributions to v is:

$$score(v) = \sum_{l_i} S\left(v_{d_1^{(i)}}\right) \prod_j p\left(v_{d_{j+1}^{(i)}} \middle| v_{d_j^{(i)}}\right) Term(v) LB(n)$$

$$\sum_n \qquad Q^n[v]$$

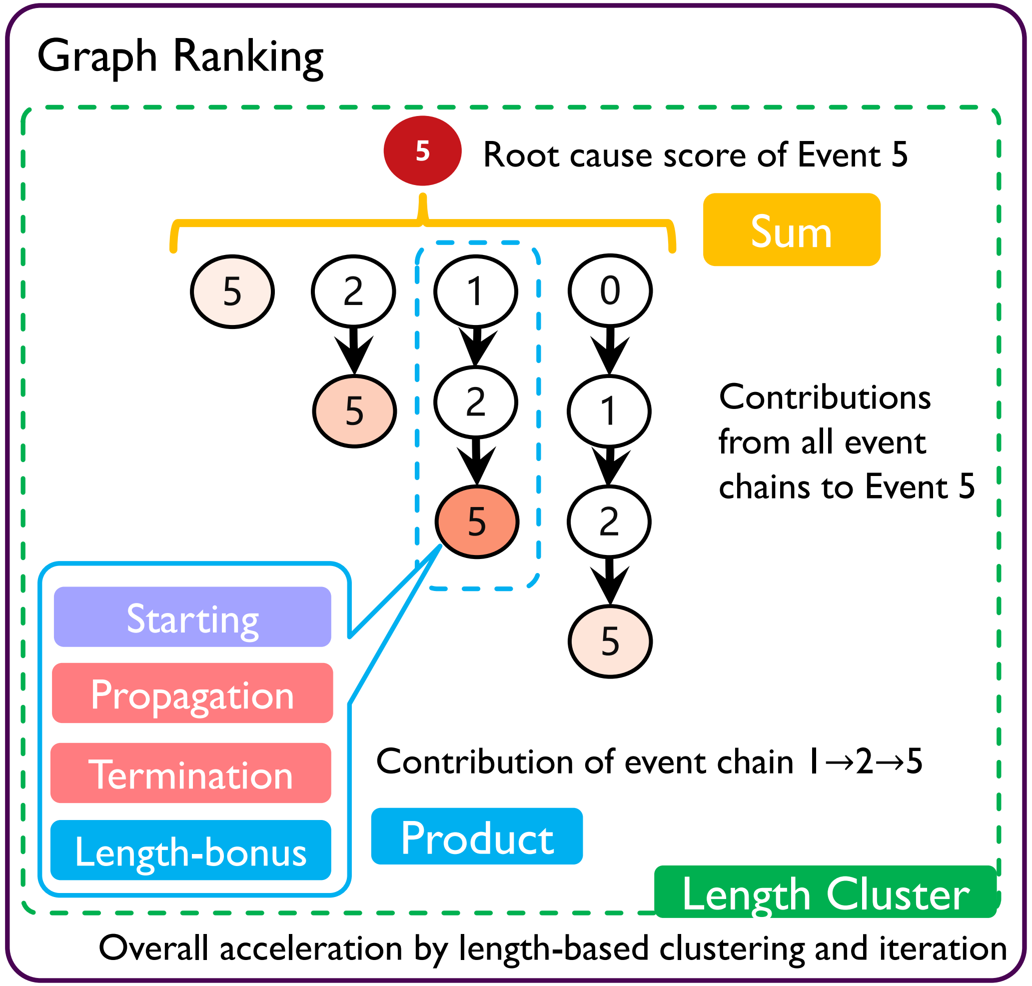
Partial contribution of all the n-length chain

$$\boldsymbol{Q}^{n+1} = \boldsymbol{Q}^n \boldsymbol{P}$$

Complexity is $O(mT)$

In practice, the number of iteration is not infinite. If the upper limit is T, then the approximation error upper bound of $score(v)$ is $\left(\frac{1}{1+\alpha}\right)^T$

Graph Ranking in Detail



(Non-optimized) In an event-causal graph with m causal links, if the maximum event chain length is T, the complexity of enumerating all possible event chains is $O(m^T)$

(Optimization) For an event v, assume an n-length chain l_i is $v_{d_1^{(i)}}, v_{d_2^{(i)}}, \ldots, v_{d_n^{(i)}}, v_{d_n^{(i)}}$ is v, then all the chain contributions to v is:

$$score(v) = \sum_{n} \underbrace{\sum_{l_i} S\left(v_{d_1^{(i)}}\right) \prod_{j} p\left(v_{d_{j+1}^{(i)}} \middle| v_{d_j^{(i)}}\right)}_{Q^n[v]} Term(v) LB(n)$$

Partial contribution of all the n-length chain

$$\boldsymbol{Q}^{n+1} = \boldsymbol{Q}^n \boldsymbol{P}$$

Complexity is $O(mT)$

In practice, the number of iteration is not infinite. If the upper limit is T, then the approximation error upper bound of $score(v)$ is $\left(\frac{1}{1+\alpha}\right)^T$

Outline

- Background
- Design
- Evaluation
- Conclusion

Experiment Setup

Dataset Construction

- Collected from three eBay datacenters with over 5k services in three data centers
- 800k monitoring signals transformed into events (46 signals per service)
- Extracted events from 16 basic sources, collected during 16 months
- Two datasets: Business Dataset (170 incidents) & Service Dataset (782 incidents)
  - Split half by half as train/test set.
  - Labels from historical remediation logs.

Type	Event	Detection Method
Monitoring Data	High GC (Overhead)	Rule-based
	High CPU Usage	Rule-based
	Out of Memory	Rule-based
	LB Connection Stacking	Statistical Model
	Latency Spike	Statistical Model
	TPS Spike	Statistical Model
	Database Anomaly	ML Model
	Business Metric Anomaly	ML Model
	WebAPI Error	Statistical Model
	Internal Error	Statistical Model
	ServiceClient Error	Statistical Model
	Bad Host	ML Model
	Hystrix Circuit Break	De Facto
Human Activity	Code Deployment	De Facto
	Configuration Change	De Facto
	External URL	De Facto

#Service	#Users	#Signals	#Signal/svc	#Incident		#Month
				Business	Service	
5k	185m	800k	46	170	782	16

Experiment Setup

Metric

- Top-1 and Top-3 accuracy

Experiment Environment and Implementation Efficiency

- Intel(R) Core(TM) i9-9980HK CPU, an 11GB GTX1080Ti GPU, and 32GB memory.
- About 45 minutes to train the CoE with all the incidents in each dataset.
- Minimal storage cost at just 52.06KB
- Performance similar to Groot

Model	Service (per incident)		Business (per incident)	
	ExecTime	#event	ExecTime	#event
CoE(Ours)	2.16s	14.70	4.06s	18.73
Groot	3.16s	14.70	2.98s	18.73

Accuracy Evaluation

	Model	MEG	Service Top-1	Service Top-3	Business Top-1	Business Top-3
Baselines	PageRank		16.1%	25.3%	1.2%	1.8%
	GraphSAGE		62.2%	78.1%	81.1%	93.7%
	GAT		12.2%	47.6%	60.5%	79.2%
	GCN		29.3%	57.3%	69.2%	85.3%
	Groot w/o *MEG*		17.1%	48.8%	23.2%	45.5%
	Groot	✓	74%	92%	81%	96%
Ours	CoE with *MEG*	✓	78.1%	93.9%	78.7%	95%
	CoE		**79.3%**	**98.8%**	**85.3%**	**96.6%**

MEG: manual event-causal graph

Accuracy Evaluation

	Model	MEG	Service		Business	
			Top-1	Top-3	Top-1	Top-3
Baselines	PageRank		16.1%	25.3%	1.2%	1.8%
	GraphSAGE		62.2%	78.1%	81.1%	93.7%
	GAT		12.2%	47.6%	60.5%	79.2%
	GCN		29.3%	57.3%	69.2%	85.3%
	Groot w/o *MEG*		17.1%	48.8%	23.2%	45.5%
	Groot	✓	74%	92%	81%	96%
Ours	CoE with *MEG*	✓	78.1%	93.9%	78.7%	95%
	CoE		**79.3%**	**98.8%**	**85.3%**	**96.6%**

MEG: manual event-causal graph

Accuracy Evaluation

	Model	MEG	Service		Business	
			Top-1	Top-3	Top-1	Top-3
Baselines	PageRank		16.1%	25.3%	1.2%	1.8%
	GraphSAGE		62.2%	78.1%	81.1%	93.7%
	GAT		12.2%	47.6%	60.5%	79.2%
	GCN		29.3%	57.3%	69.2%	85.3%
	Groot w/o *MEG*		17.1%	48.8%	23.2%	45.5%
	Groot	✓	74%	92%	81%	96%
Ours	CoE with *MEG*	✓	78.1%	93.9%	78.7%	95%
	CoE		**79.3%**	**98.8%**	**85.3%**	**96.6%**

MEG: manual event-causal graph

Compared with baseline methods, CoE is indeed effective in detecting the root cause events

Accuracy Evaluation

	Model	MEG	Service		Business	
			Top-1	Top-3	Top-1	Top-3
Baselines	PageRank		16.1%	25.3%	1.2%	1.8%
	GraphSAGE		62.2%	78.1%	81.1%	93.7%
	GAT		12.2%	47.6%	60.5%	79.2%
	GCN		29.3%	57.3%	69.2%	85.3%
	Groot w/o *MEG*		17.1%	48.8%	23.2%	45.5%
	Groot	✓	74%	92%	81%	96%
Ours	CoE with *MEG*	✓	78.1%	93.9%	78.7%	95%
	CoE		**79.3%**	**98.8%**	**85.3%**	**96.6%**

MEG: manual event-causal graph

Compared with baseline methods, CoE is indeed effective in detecting the root cause events

Accuracy Evaluation

	Model	MEG	Service Top-1	Service Top-3	Business Top-1	Business Top-3
Baselines	PageRank		16.1%	25.3%	1.2%	1.8%
	GraphSAGE		62.2%	78.1%	81.1%	93.7%
	GAT		12.2%	47.6%	60.5%	79.2%
	GCN		29.3%	57.3%	69.2%	85.3%
	Groot w/o *MEG*		17.1%	48.8%	23.2%	45.5%
	Groot	✓	74%	92%	81%	96%
Ours	CoE with *MEG*	✓	78.1%	93.9%	78.7%	95%
	CoE		**79.3%**	**98.8%**	**85.3%**	**96.6%**

MEG: manual event-causal graph

Groot seriously depends on the manual rulebook

The learned weighted rulebook in CoE is more precise than the manual rulebook

Compared with baseline methods, CoE is indeed effective in detecting the root cause events

Ablation Study

- Length bonus, $LB(n)$
- Out-edge bonus (termination term), $Term(v)$
- Event Importance S
- Inter-service causal link weights R_d
- Intra-service causal link weights R_s

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o length bonus	79.3%	96.3%	83.2%	96.1%
CoE w/o out-edge bonus	75.6%	96.3%	83.4%	95.3%
CoE w/o both bonus terms	75.6%	93.9%	83.4%	95.3%

Evaluating the two bonus term

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o learning S	75.6%	96.3%	84.5%	96.1%
CoE w/o learning R_d	78.1%	97.6%	84.5%	95.8%
CoE w/o learning R_s	51.2%	86.6%	84.5%	95.5%
CoE w/o bonus terms	75.6%	93.9%	83.4s%	95.3%

Removing/freezing a certain component

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
Naive CoE	31.7%	64.6%	71.7%	90.3%
+bonus terms	33.0%	67.1%	72.1%	90.5%
+learn S	51.2%	81.7%	82.1%	95%
+learn R_d	51.2%	86.6%	84.5%	95.5%
+learn R_s	**79.3%**	**98.8%**	**85.3%**	**96.6%**

Adding components one by one

Ablation Study

- Length bonus, $LB(n)$
- Out-edge bonus (termination term), $Term(v)$
- Event Importance S
- Inter-service causal link weights R_d
- Intra-service causal link weights R_s

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o length bonus	79.3%	96.3%	83.2%	96.1%
CoE w/o out-edge bonus	75.6%	96.3%	83.4%	95.3%
CoE w/o both bonus terms	75.6%	93.9%	83.4%	95.3%

Evaluating the two bonus term

All the components has its impact, while S is the most significant

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o learning S	75.6%	96.3%	84.5%	96.1%
CoE w/o learning R_d	78.1%	97.6%	84.5%	95.8%
CoE w/o learning R_s	51.2%	86.6%	84.5%	95.5%
CoE w/o bonus terms	75.6%	93.9%	83.4s%	95.3%

Removing/freezing a certain component

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
Naive CoE	31.7%	64.6%	71.7%	90.3%
+bonus terms	33.0%	67.1%	72.1%	90.5%
+learn S	51.2%	81.7%	82.1%	95%
+learn R_d	51.2%	86.6%	84.5%	95.5%
+learn R_s	**79.3%**	**98.8%**	**85.3%**	**96.6%**

Adding components one by one

Ablation Study

- Length bonus, $LB(n)$
- Out-edge bonus (termination term), $Term(v)$
- Event Importance S
- Inter-service causal link weights R_d
- Intra-service causal link weights R_s

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o length bonus	79.3%	96.3%	83.2%	96.1%
CoE w/o out-edge bonus	75.6%	96.3%	83.4%	95.3%
CoE w/o both bonus terms	75.6%	93.9%	83.4%	95.3%

Evaluating the two bonus term

All the components has its impact, while S is the most significant

The terms has biased impact on different dataset (e.g., R_s in the Service Dataset)

Incident typically associated with a single service

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
CoE	**79.3%**	**98.8%**	**85.3%**	**96.6%**
CoE w/o learning S	75.6%	96.3%	84.5%	96.1%
CoE w/o learning R_d	78.1%	97.6%	84.5%	95.8%
CoE w/o learning R_s	51.2%	86.6%	84.5%	95.5%
CoE w/o bonus terms	75.6%	93.9%	83.4s%	95.3%

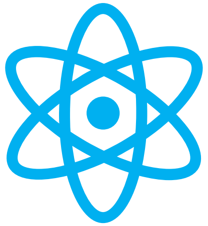
Removing/freezing a certain component

	Service		Business	
	Top 1	Top 3	Top 1	Top 3
Naive CoE	31.7%	64.6%	71.7%	90.3%
+bonus terms	33.0%	67.1%	72.1%	90.5%
+learn S	51.2%	81.7%	82.1%	95%
+learn R_d	51.2%	86.6%	84.5%	95.5%
+learn R_s	**79.3%**	**98.8%**	**85.3%**	**96.6%**

Adding components one by one

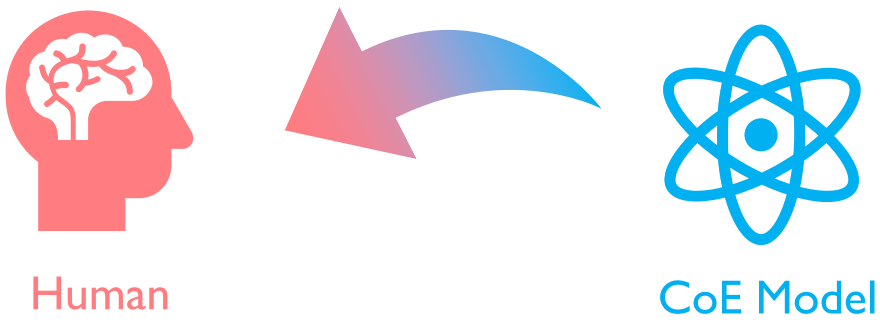
Discussion: How does CoE Align with Human Knowledge

Human

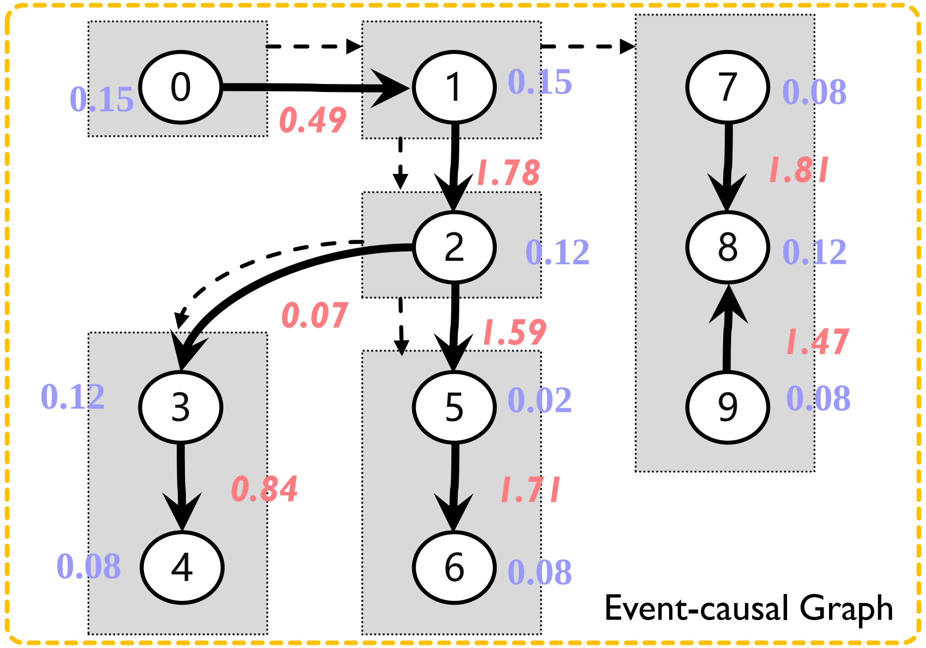


CoE Model

Discussion: How does CoE Align with Human Knowledge



Discussion: How does CoE Align with Human Knowledge



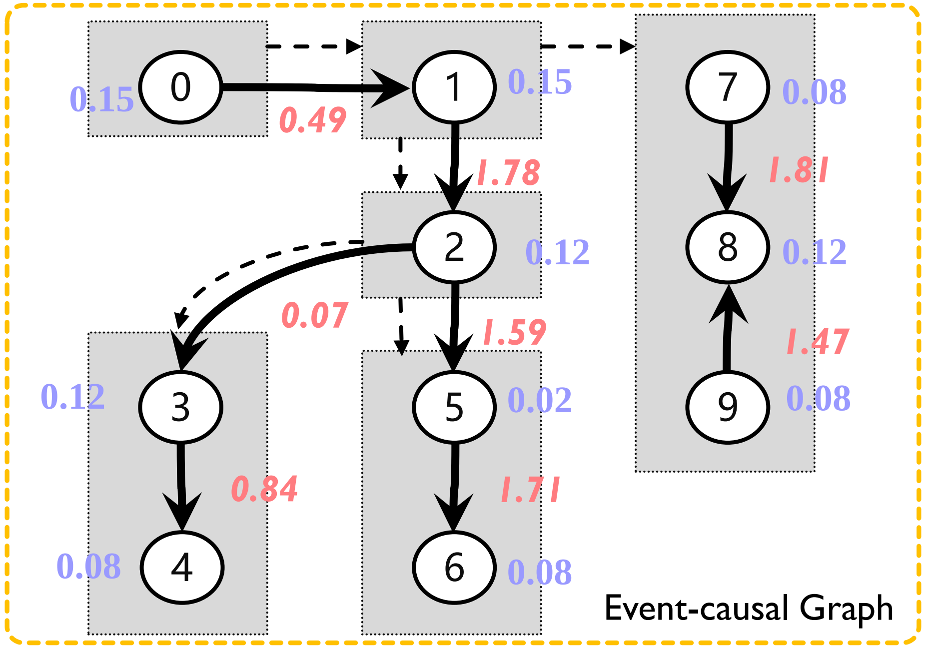
Discussion: How does CoE Align with Human Knowledge

Input Events of user-defined granularity

Parameters Clear physical meanings
Easy to understand

Inference process Align with SREs’ daily diagnosis process

Few hyper-parameters No need for deep learning background



Discussion: How does CoE Align with Human Knowledge

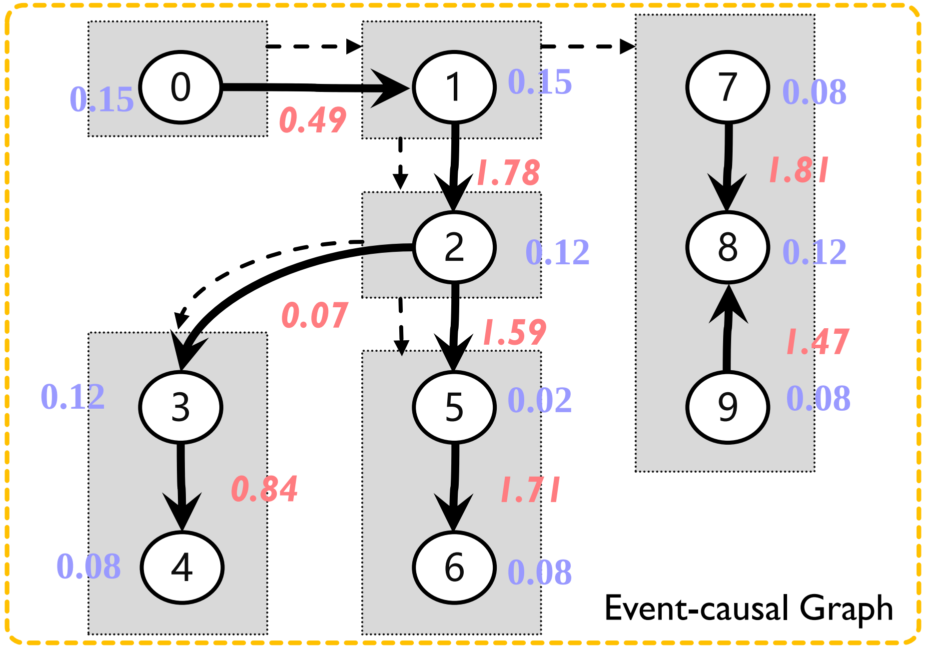
Input Events of user-defined granularity

Parameters Clear physical meanings
Easy to understand

Inference process Align with SREs' daily diagnosis process

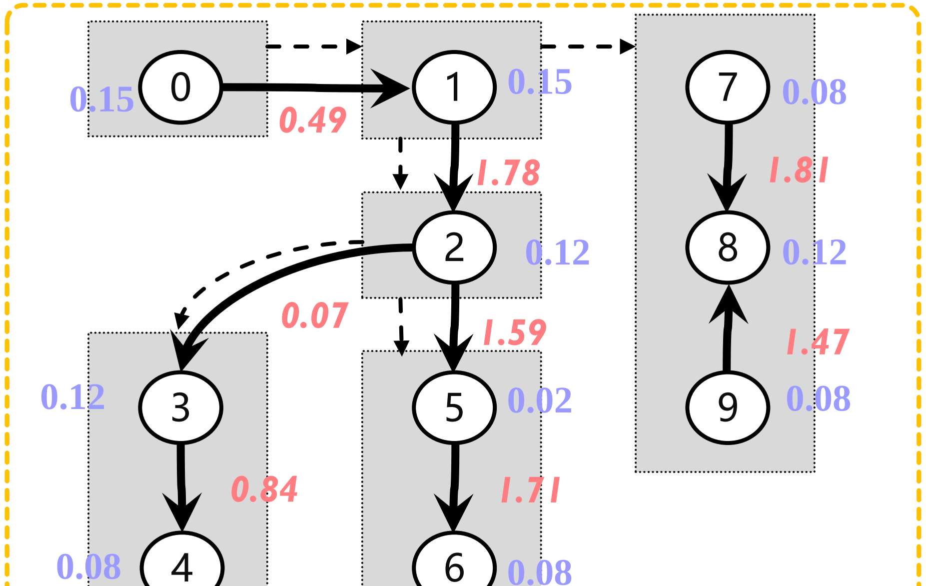
Few hyper-parameters No need for deep learning background

event 1 is caused by event 2
event 5 is more likely to cause event 2 than event 3
event 0 more important than event 5



Discussion: How does CoE Align with Human Knowledge

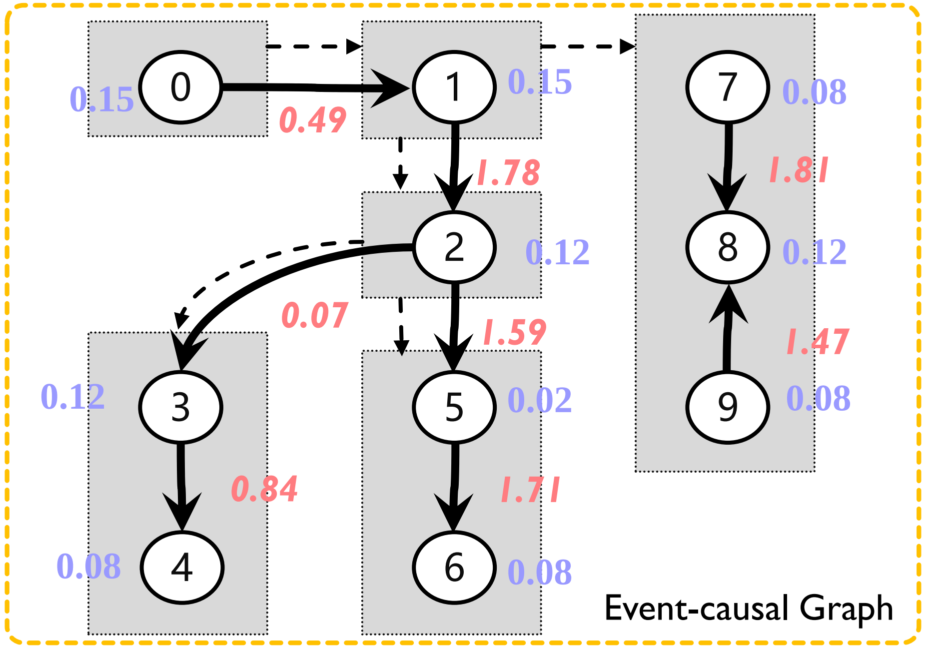
Input	Events of user-defined granularity
Parameters	Clear physical meanings Easy to understand
Inference process	Align with SREs' daily diagnosis process
Few hyper-parameters	No need for deep learning background



Event-causal Graph

Discussion: How does CoE Align with Human Knowledge

Input	Events of user-defined granularity
Parameters	Clear physical meanings Easy to understand
Inference process	Align with SREs' daily diagnosis process
Few hyper-parameters	No need for deep learning background

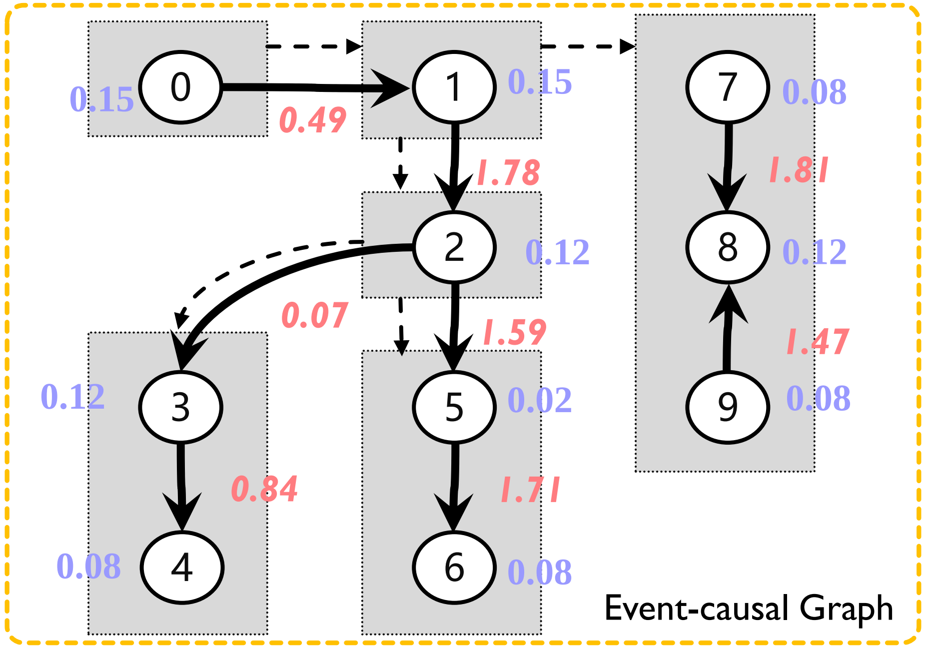


Discussion: How does CoE Align with Human Knowledge

Input	Events of user-defined granularity
Parameters	Clear physical meanings Easy to understand
Inference process	Align with SREs' daily diagnosis process
Few hyper-parameters	No need for deep learning background

Clear physical meanings of parameters

Easy to edit without further explanation



Outline

- Background
- Design
- Evaluation
- Conclusion

Conclusion

Chain-of-Event (CoE)

- Event-based RCA algorithm utilizing multi-modal monitoring data
- Interpretable parameter design aligning with human knowledge
- Automatic learning event-causal graph
- High accuracy of root cause analysis evaluated with real-world dataset

Key Designs of CoE

- Incident-specific and overall event-causal graph
- Graph ranking with event chains
- Chain-length-based acceleration
- Proved effectiveness of the key components in ablation study

Open source code

- https://github.com/NetManAIOps/Chain-of-Event

清華大學
Tsinghua University

中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

ebay

Thank You !

Chain-of-Event: Interpretable Root Cause Analysis for Microservices through Automatically Learning Weighted Event Causal Graph

Paper: https://doi.org/10.1145/3663529.3663827

Code: https://github.com/NetManAIOps/Chain-of-Event