DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning

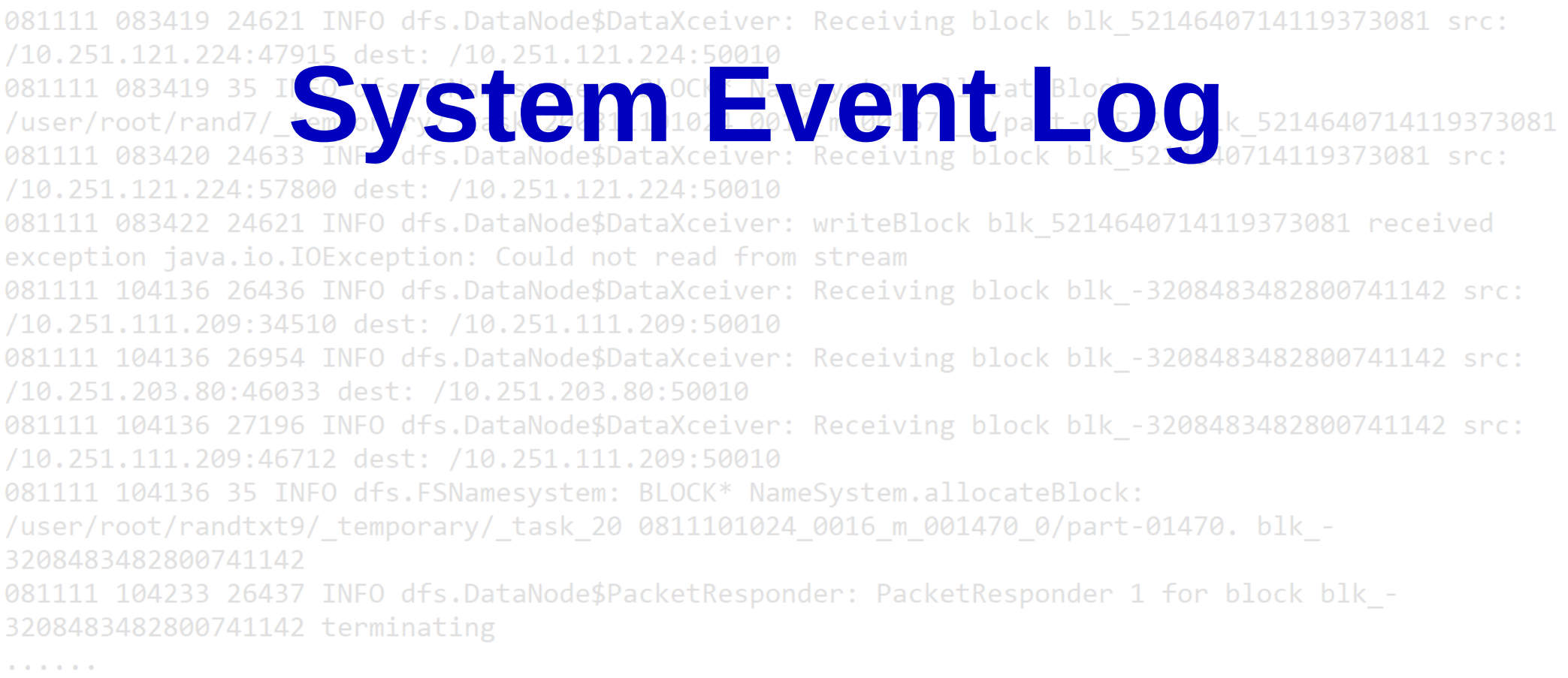
Min Du, Feifei Li, Guineng Zheng, Vivek Srikumar
University of Utah

THE UNIVERSITY OF UTAH

Background

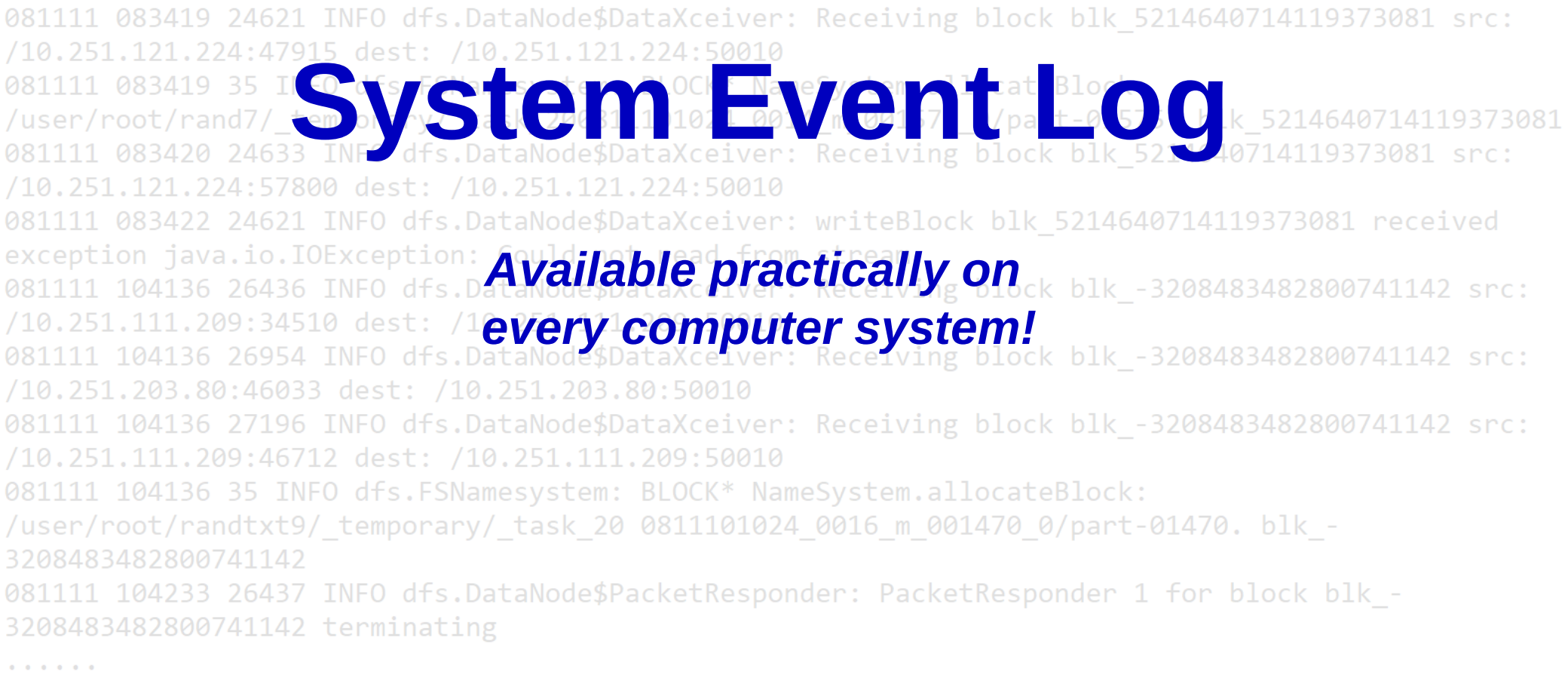
```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0014_m_001575_0/part-01575. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:57800 dest: /10.251.121.224:50010
081111 083422 24621 INFO dfs.DataNode$DataXceiver: writeBlock blk_5214640714119373081 received
exception java.io.IOException: Could not read from stream
081111 104136 26436 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:34510 dest: /10.251.111.209:50010
081111 104136 26954 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.203.80:46033 dest: /10.251.203.80:50010
081111 104136 27196 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:46712 dest: /10.251.111.209:50010
081111 104136 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/randtxt9/_temporary/_task_20 0811101024_0016_m_001470_0/part-01470. blk_-
3208483482800741142
081111 104233 26437 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-
3208483482800741142 terminating
......
```

Background



System Event Log

Background



System Event Log

Available practically on every computer system!

Background

081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0007_m_000007_0/part-00007. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:57800 dest: /10.251.121.224:50010
081111 083422 24621 INFO dfs.DataNode$DataXceiver: writeBlock blk_5214640714119373081 received
exception java.io.IOException: Could not read from stream
081111 104136 26436 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:34510 dest: /10.251.111.209:50010
081111 104136 26954 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.203.80:46033 dest: /10.251.203.80:50010
081111 104136 27196 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:46712 dest: /10.251.111.209:50010
081111 104136 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/randtxt9/_temporary/_task_200811101024_0045_m_001470_0/part-01470. blk_-
3208483482800741142
081111 104233 26437 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-
3208483482800741142 terminating
.....

System Event Log

Available practically on every computer system!

Automatic Analysis?

Background

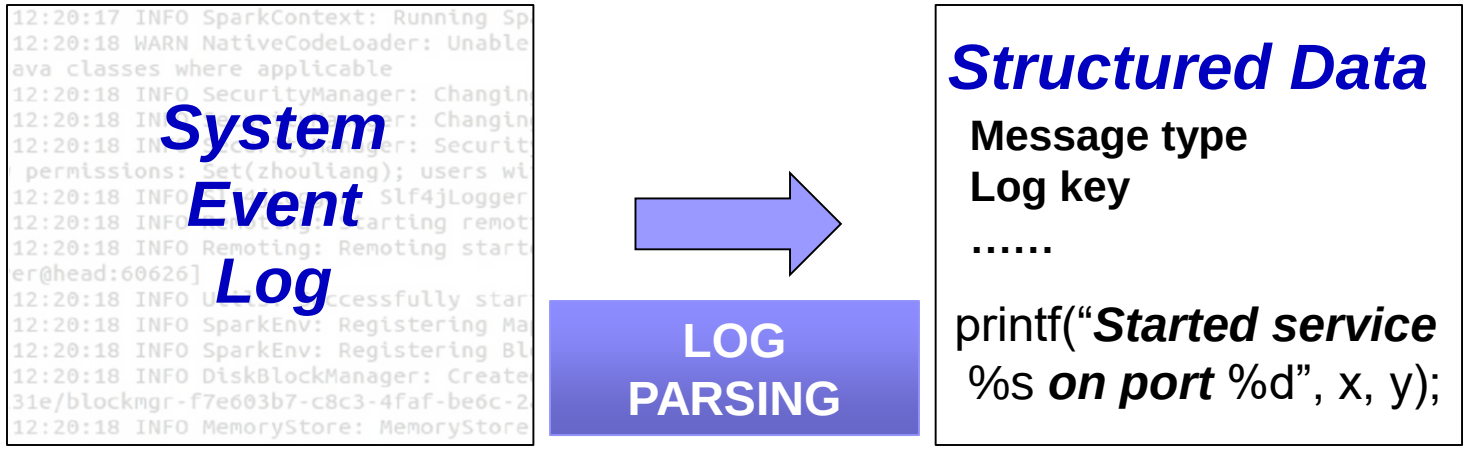
```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0014_m_001575_0/part-01575. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:57800 dest: /10.251.121.224:50010
081111 083422 24621 INFO dfs.DataNode$DataXceiver: writeBlock blk_5214640714119373081 received
exception java.io.IOException: Could not read from stream
081111 104136 26436 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:34510 dest: /10.251.111.209:50010
081111 104136 26954 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.203.80:40851 dest: /10.251.111.209:50010
081111 104136 27196 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.111.209:46712 dest: /10.251.111.209:50010
081111 104136 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/randtxt9/_temporary/_task_20 0811101024_0016_m_001470_0/part-01470. blk_-
3208483482800741142
081111 104233 26437 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-
3208483482800741142 terminating
......
```

Automatically detected anomaly

Background

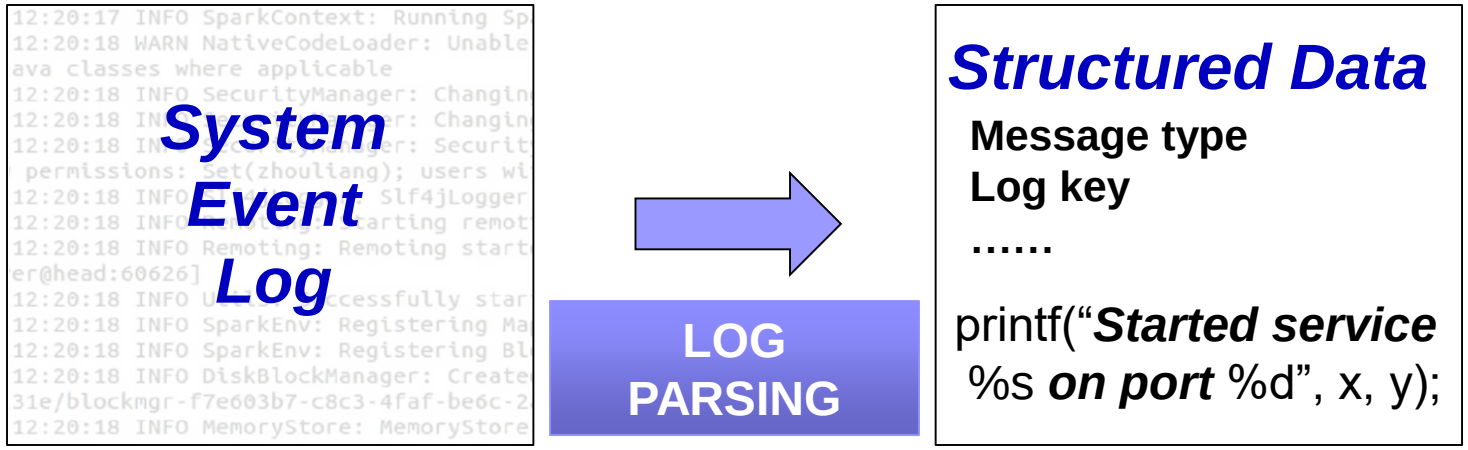
Started service A on port 80
Executor updated: app-1 is now LOADING
……

Background



Started service A on port 80
Executor updated: app-1 is now LOADING
……

Background



Started service A on port 80
Executor updated: app-1 is now LOADING
……

***Started service** * **on port** ** *(log key ID: 1)*
***Executor updated:** * **is now LOADING** *(log key ID: 2)*
……

Background

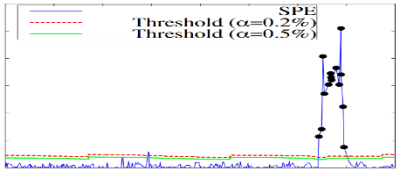
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

Background

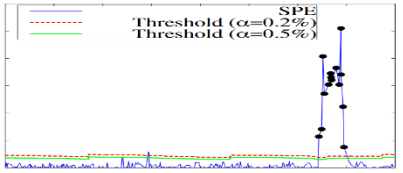
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

❑**Message count vector:**
Xu'SOSP09, Lou'ATC10, etc.

Background

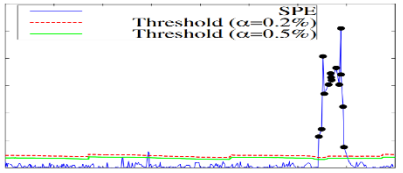
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

❑**Message count vector:**
Xu'SOSP09, Lou'ATC10, etc.
Problem: Offline batched processing

Background

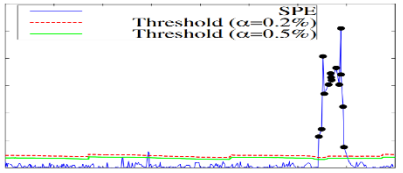
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

❑**Message count vector:**
Xu'SOSP09, Lou'ATC10, etc.
Problem: Offline batched processing

❑**Build workflow model:**
Lou'KDD10, Beschastnikh'ICSE14, Yu'ASPLOS16, etc.

Background

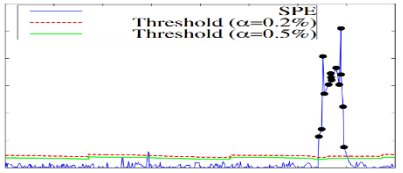
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

❑**Message count vector:**
Xu'SOSP09, Lou'ATC10, etc.
Problem: Offline batched processing

❑**Build workflow model:**
Lou'KDD10, Beschastnikh'ICSE14, Yu'ASPLOS16, etc.
Problem: Only for simple execution path anomalies

Background

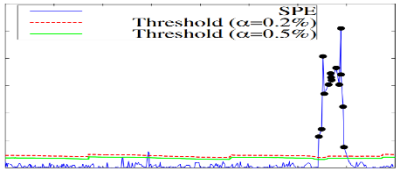
LOG PARSING

Structured Data

Message type
Log key
……

printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG ANALYSIS

Common problem: Only Log keys (Message types) are considered.

❑**Message count vector:**
Xu'SOSP09, Lou'ATC10, etc.
Problem: Offline batched processing

❑**Build workflow model:**
Lou'KDD10, Beschastnikh'ICSE14, Yu'ASPLOS16, etc.
Problem: Only for simple execution path anomalies

DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…

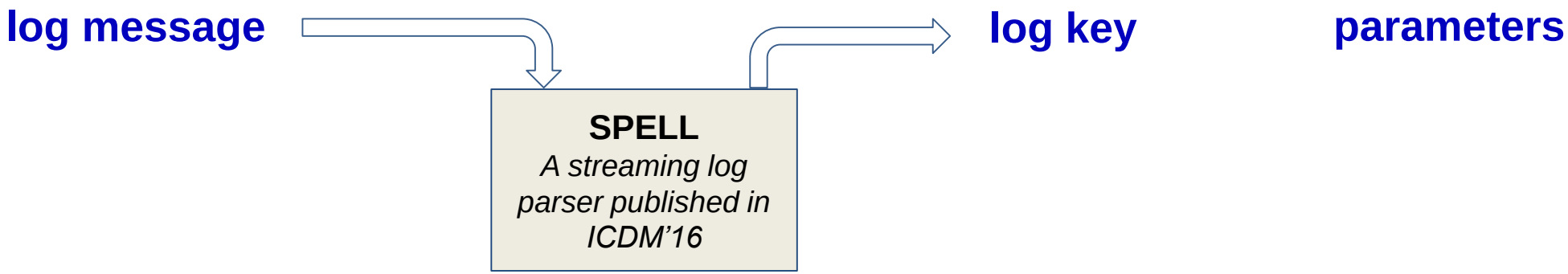
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…

SPELL
A streaming log parser published in ICDM'16

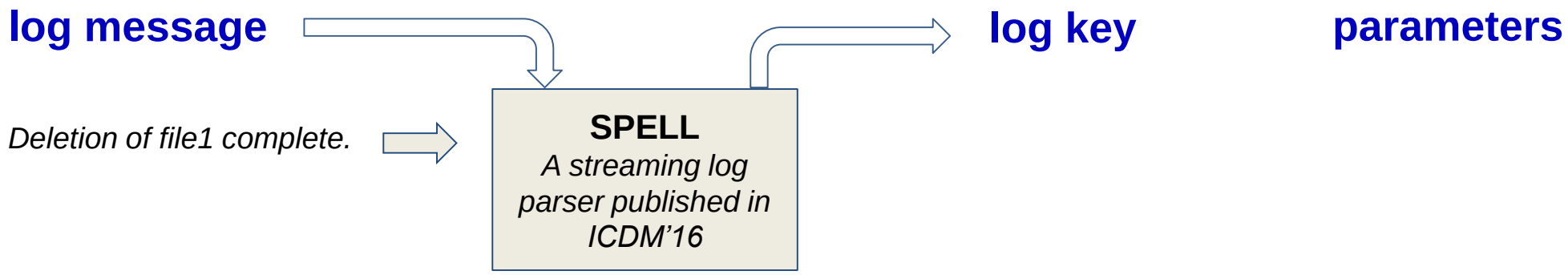
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



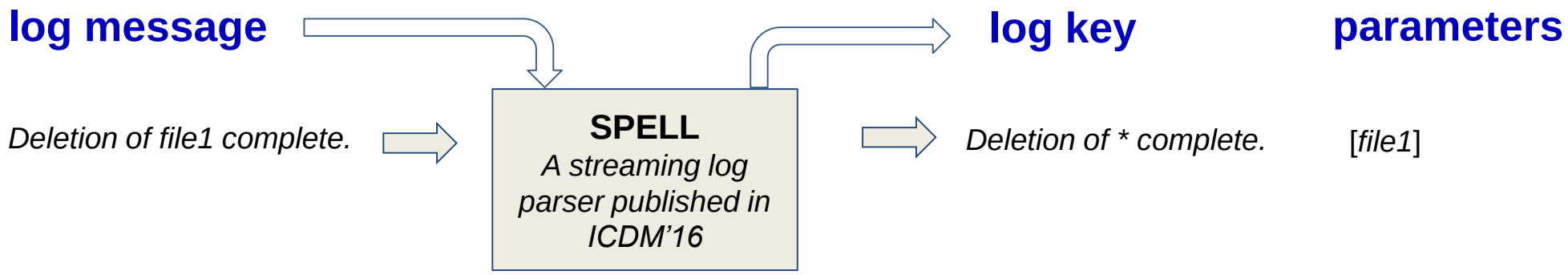
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



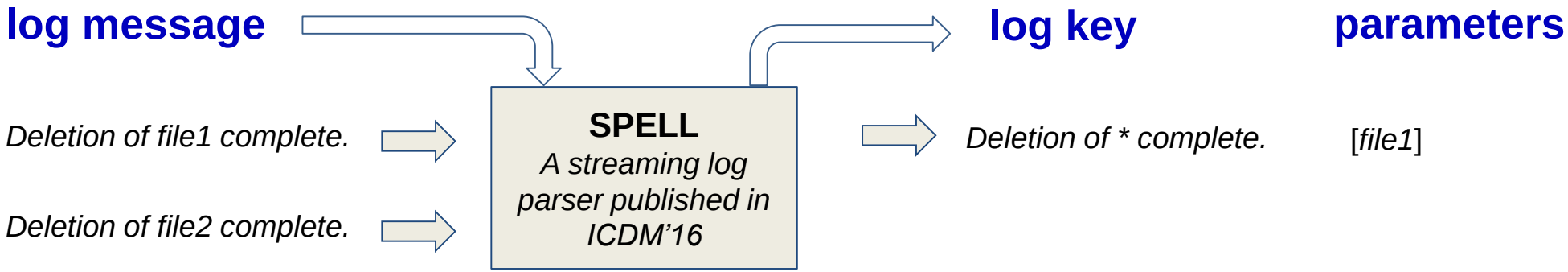
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



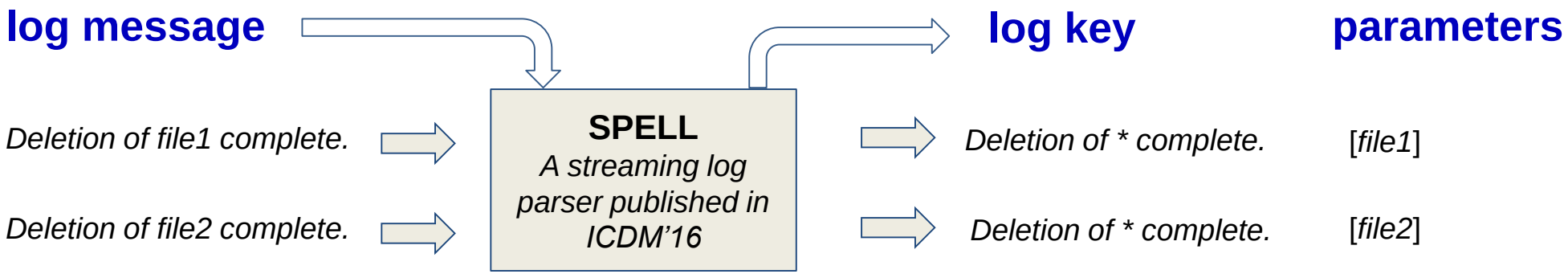
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



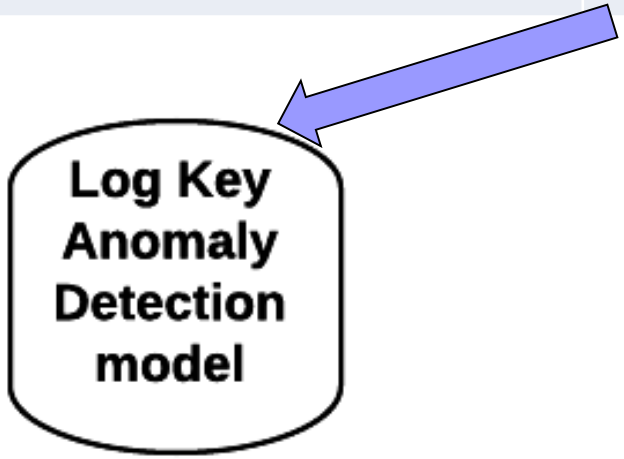
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



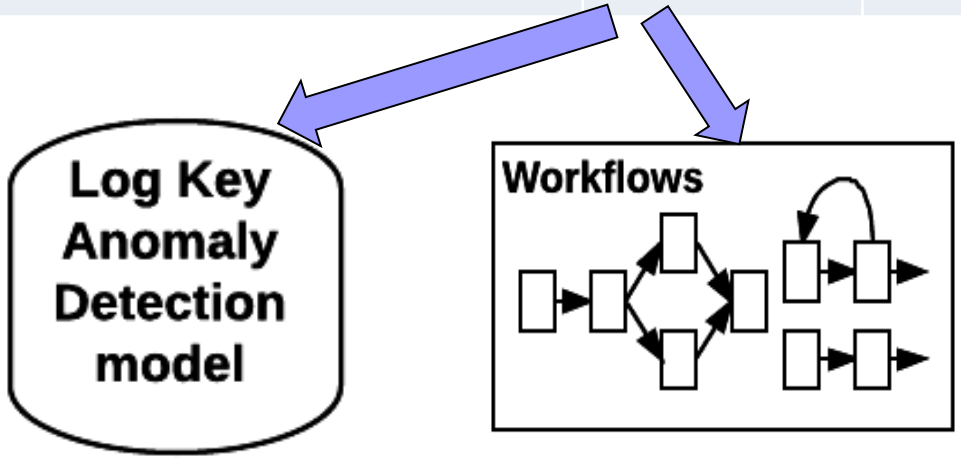
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



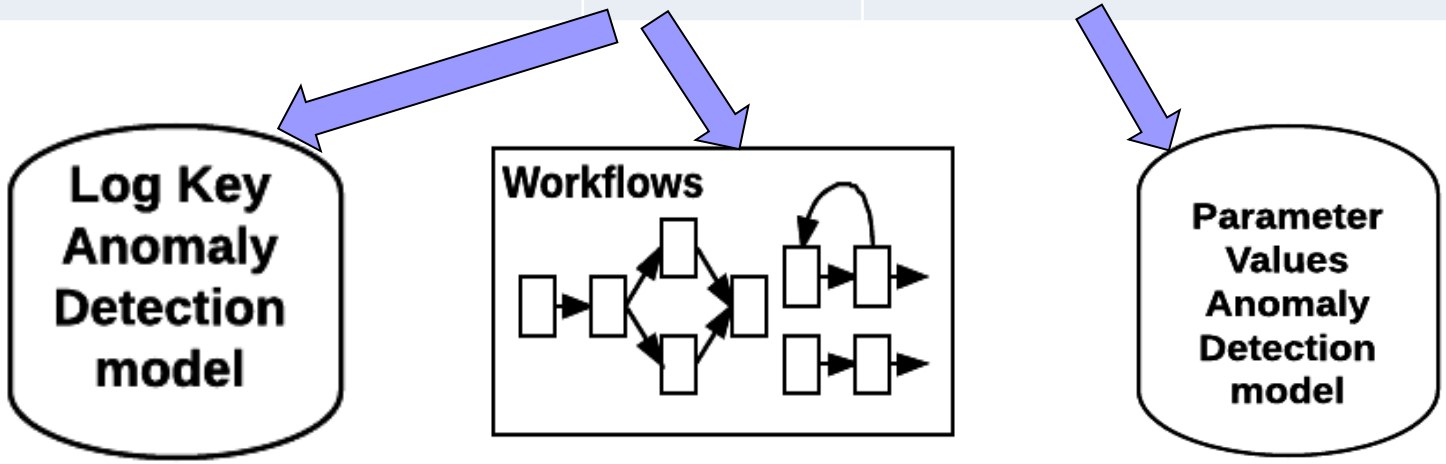
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



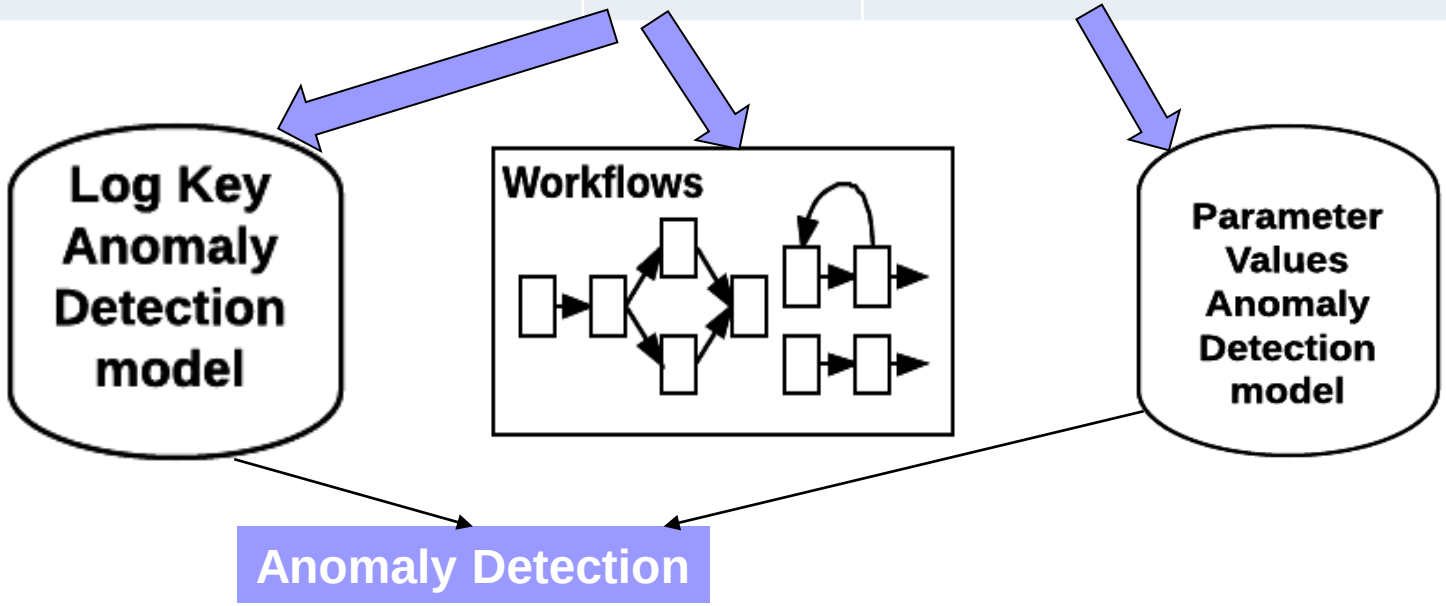
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



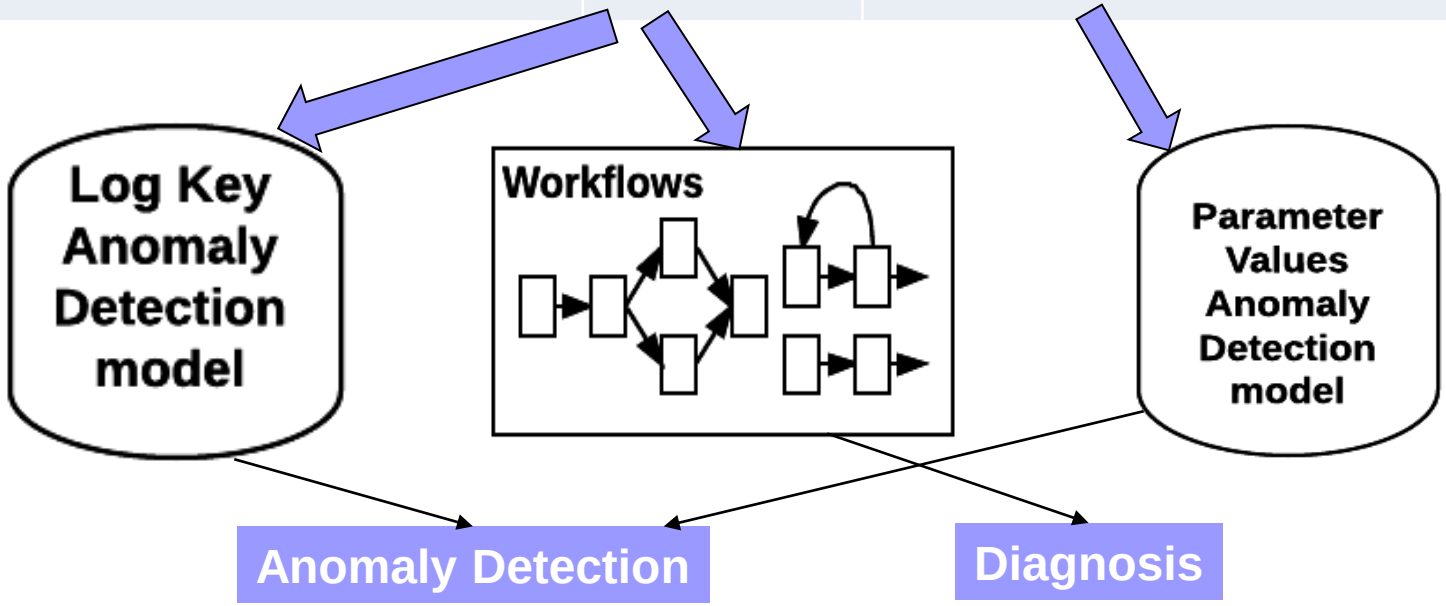
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



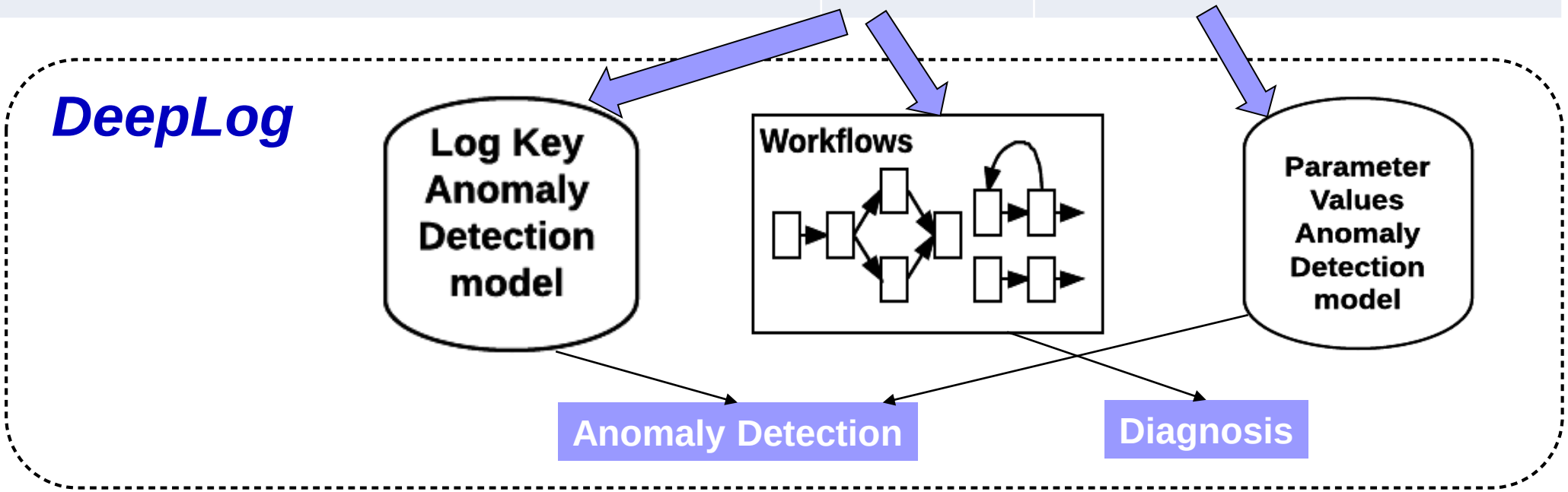
DeepLog

log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…

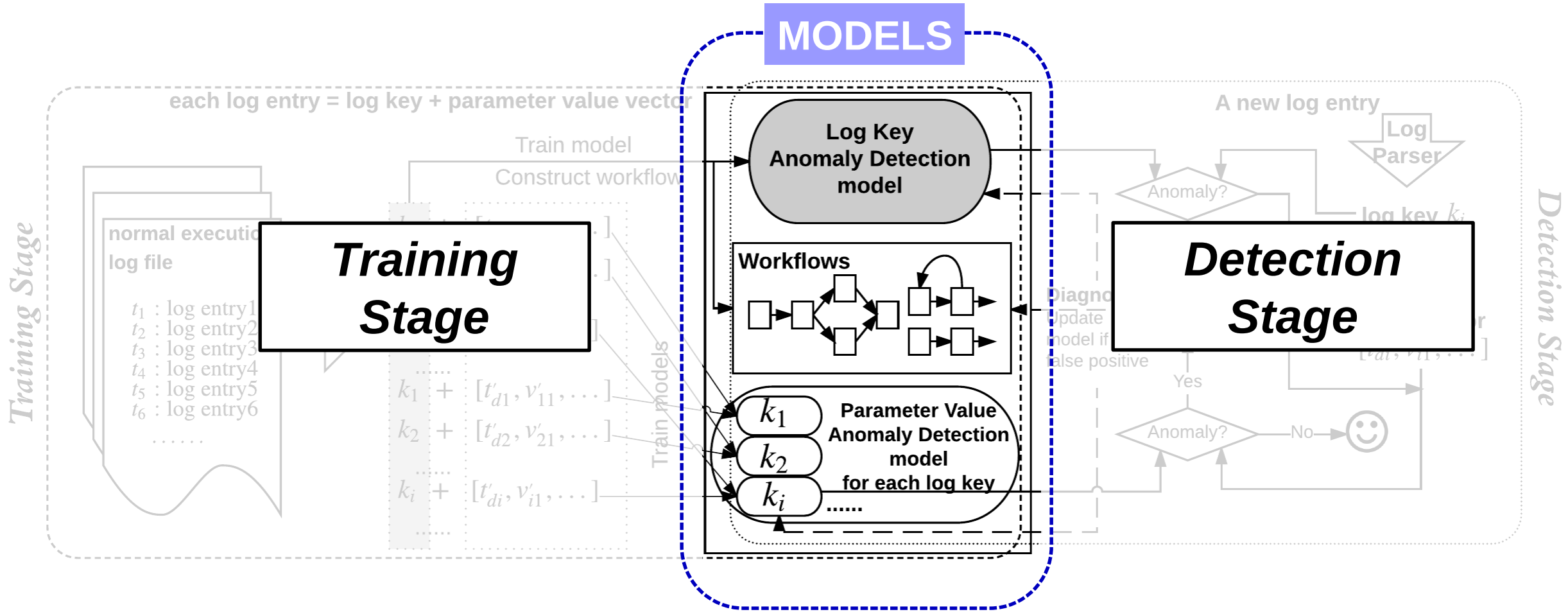


DeepLog

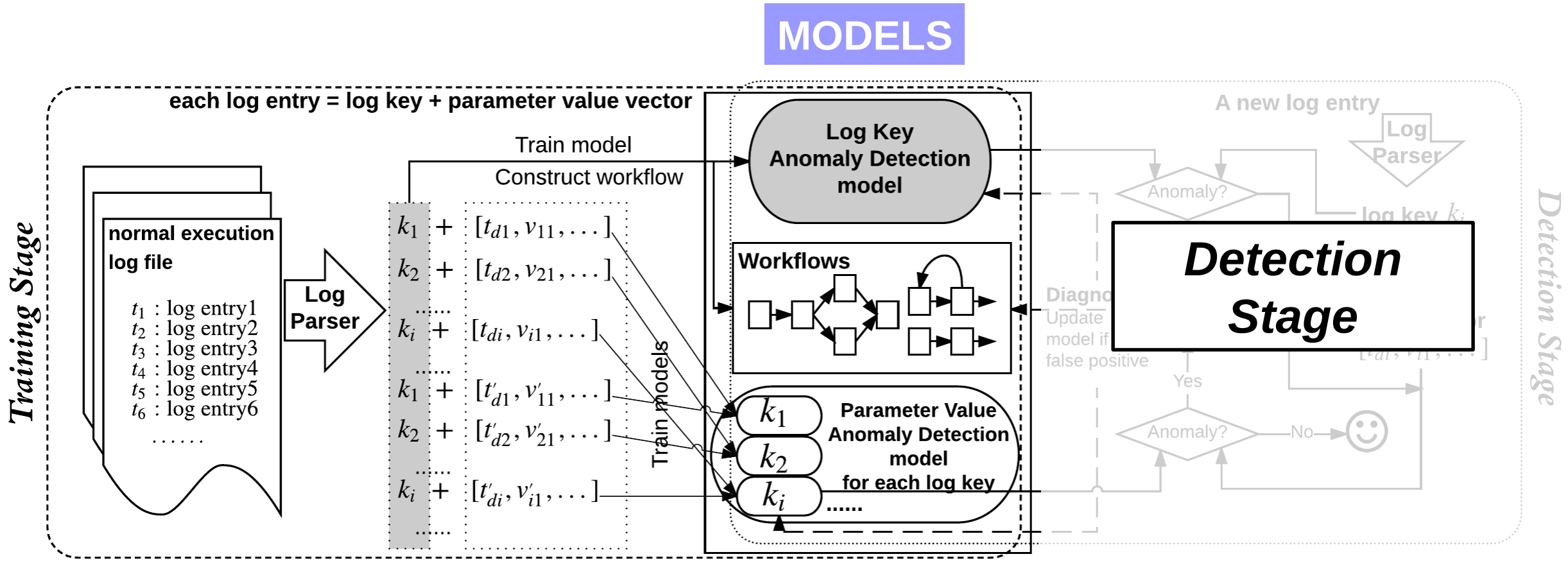
log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…



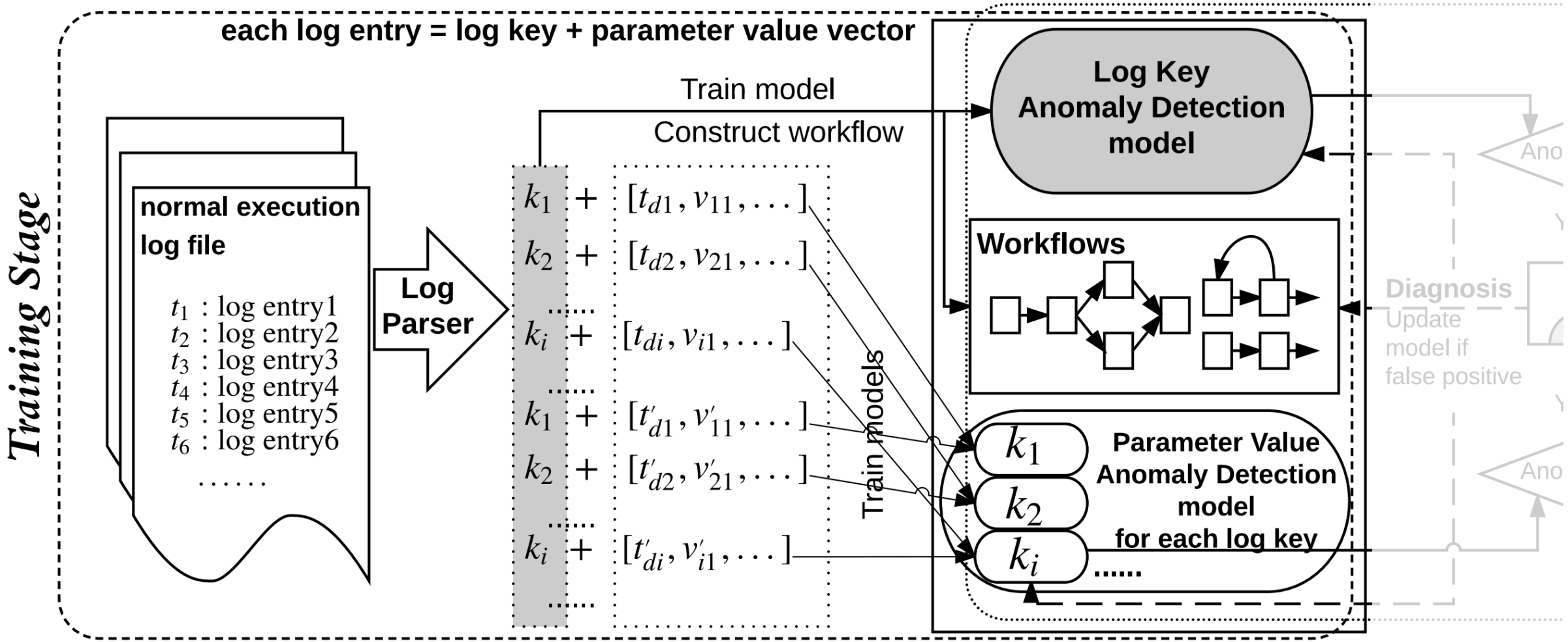
DeepLog Architecture



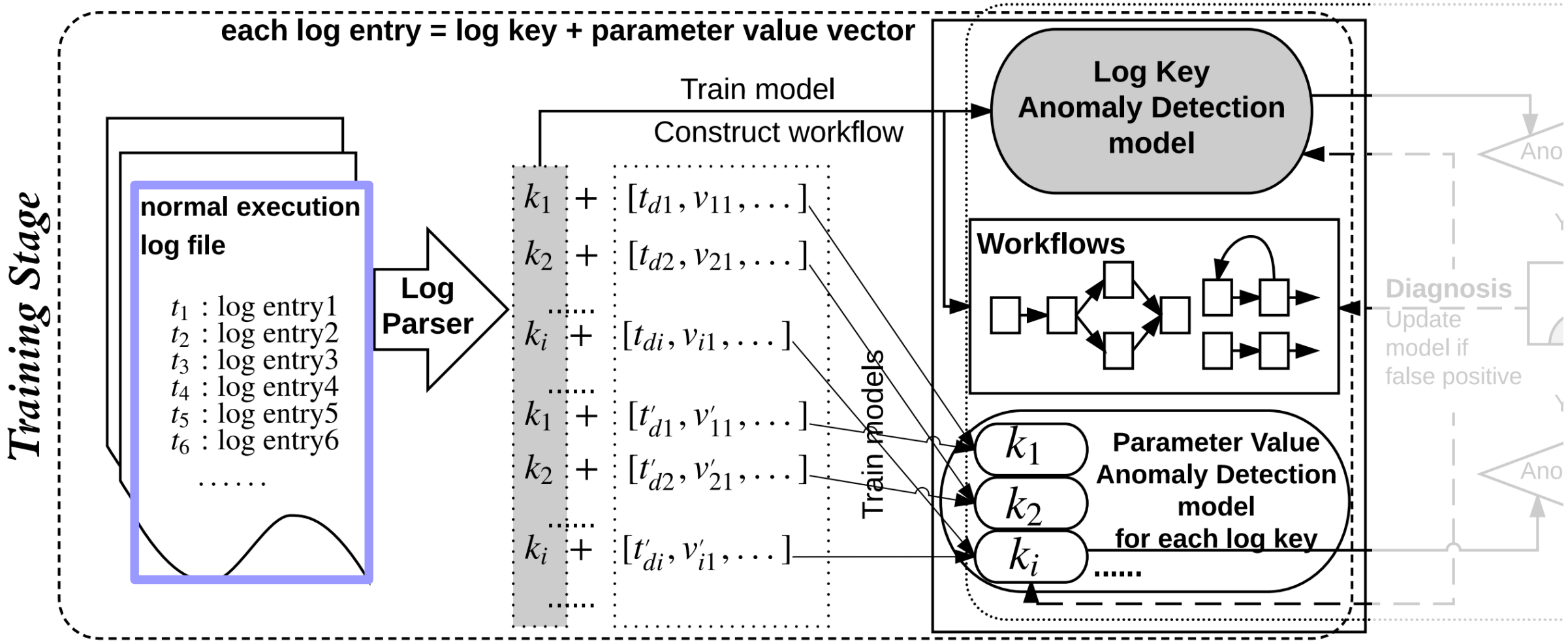
DeepLog Architecture



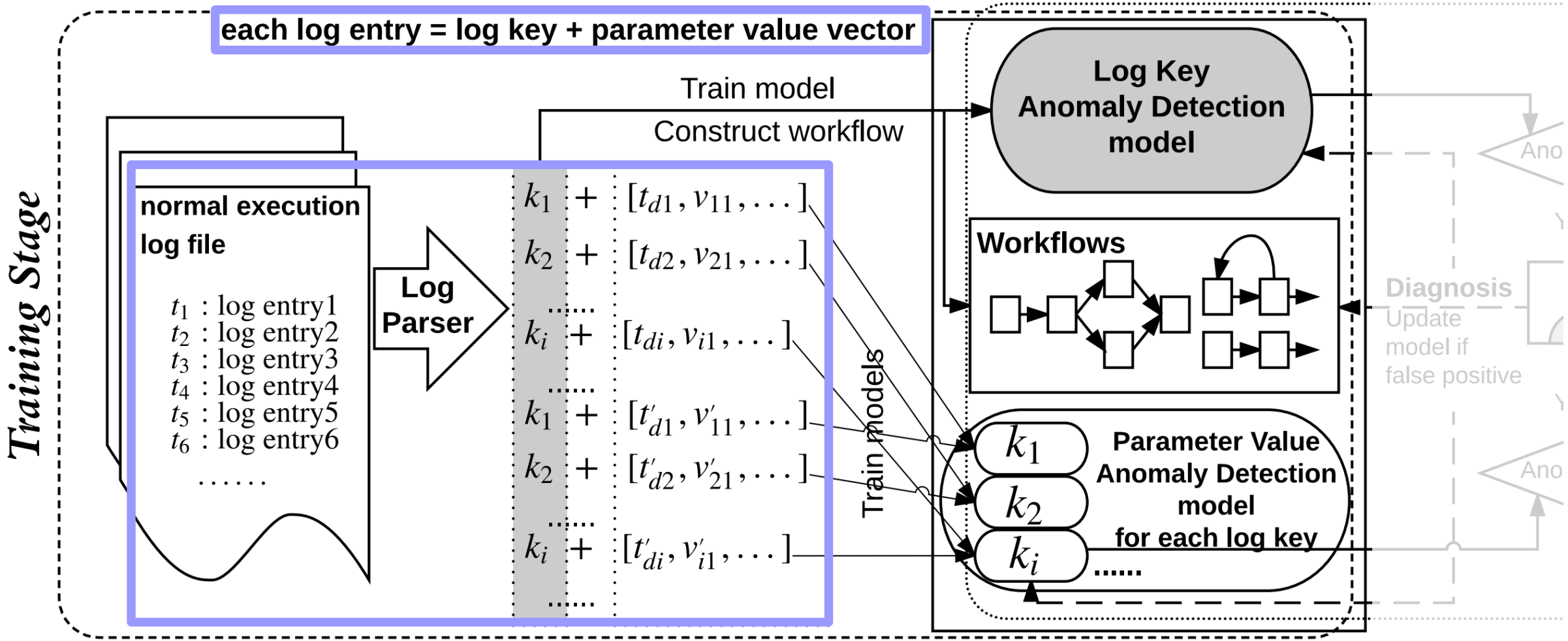
DeepLog Architecture



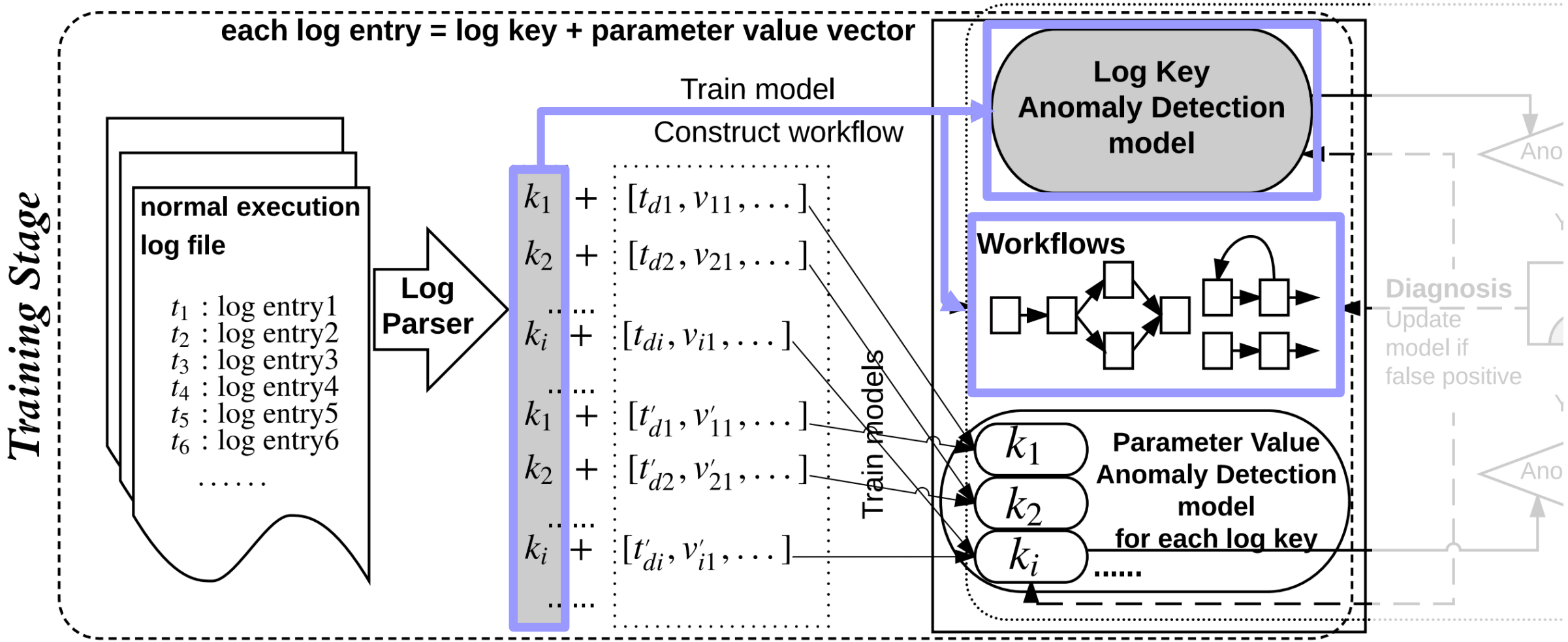
DeepLog Architecture



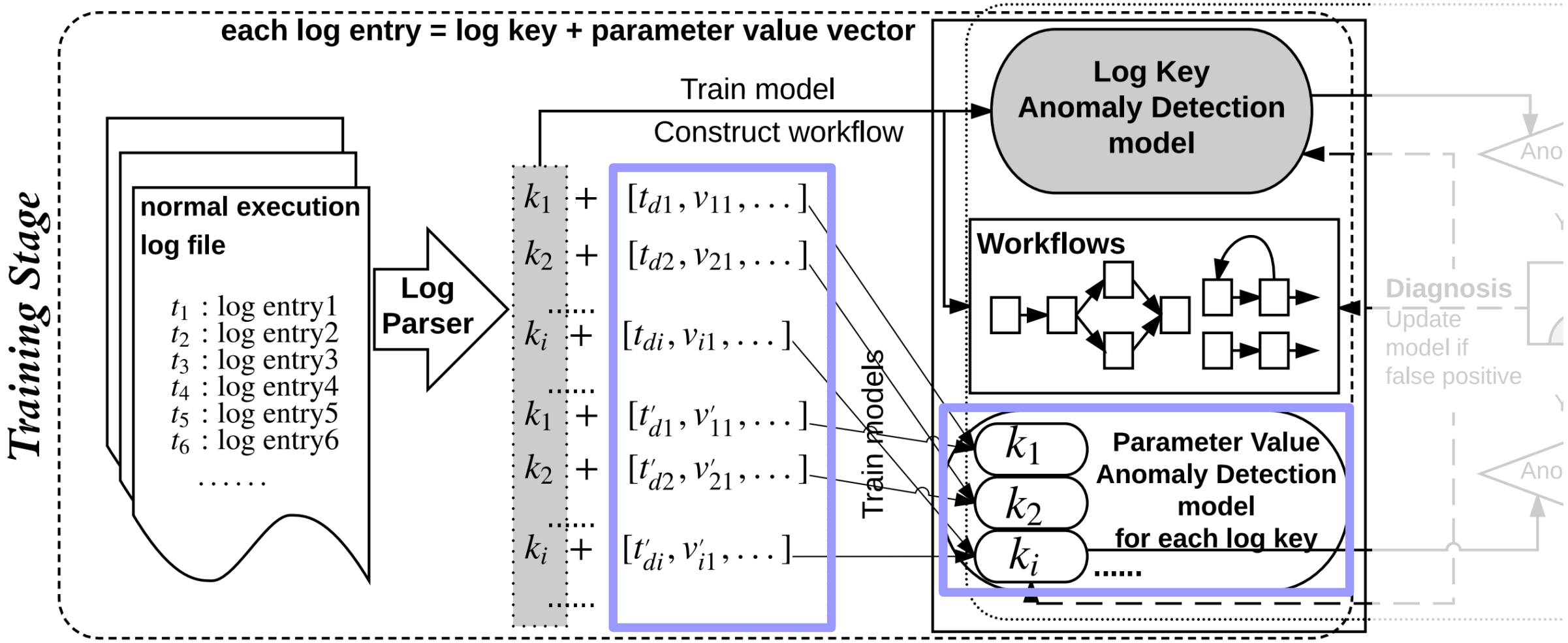
DeepLog Architecture



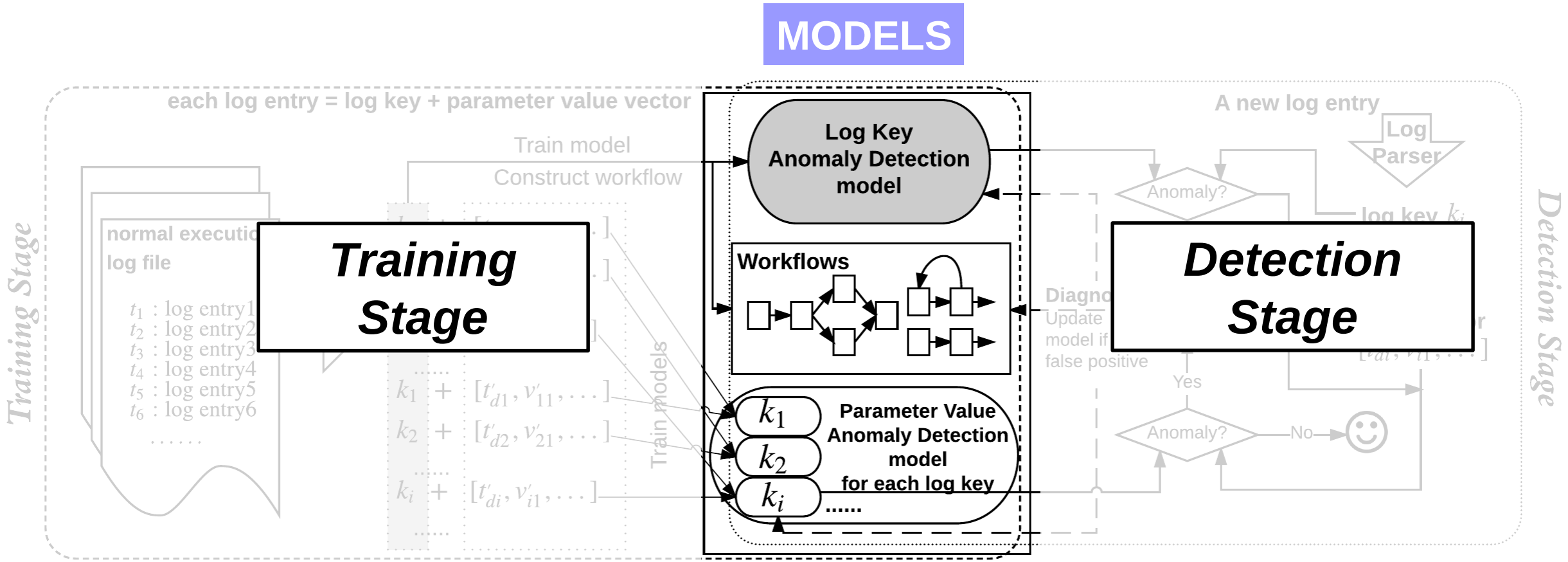
DeepLog Architecture



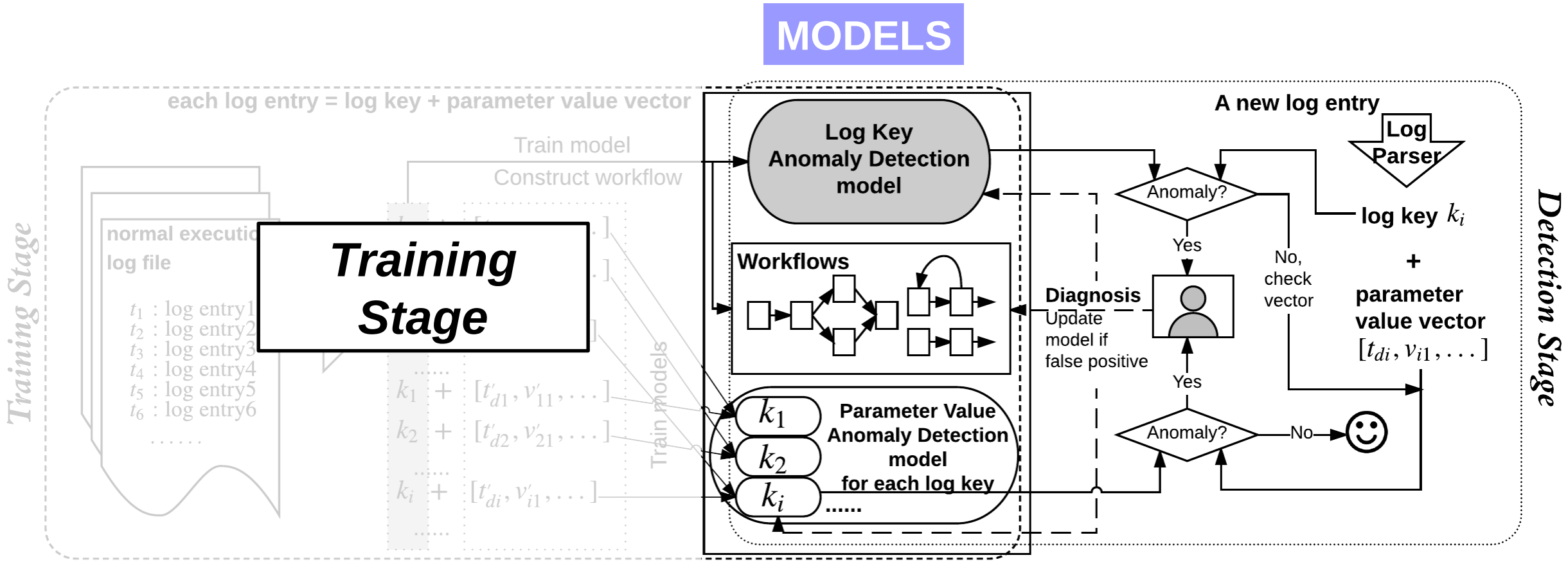
DeepLog Architecture



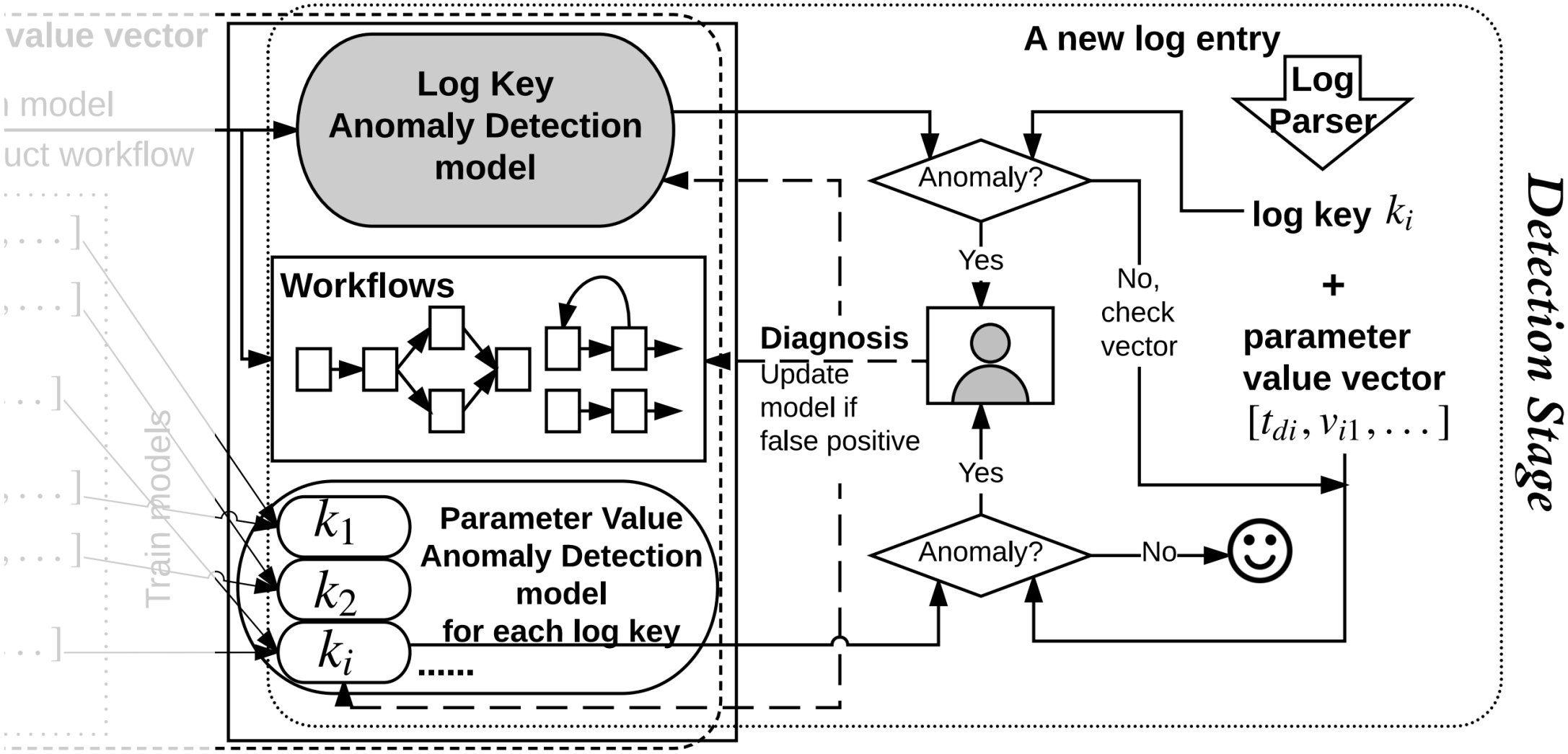
DeepLog Architecture



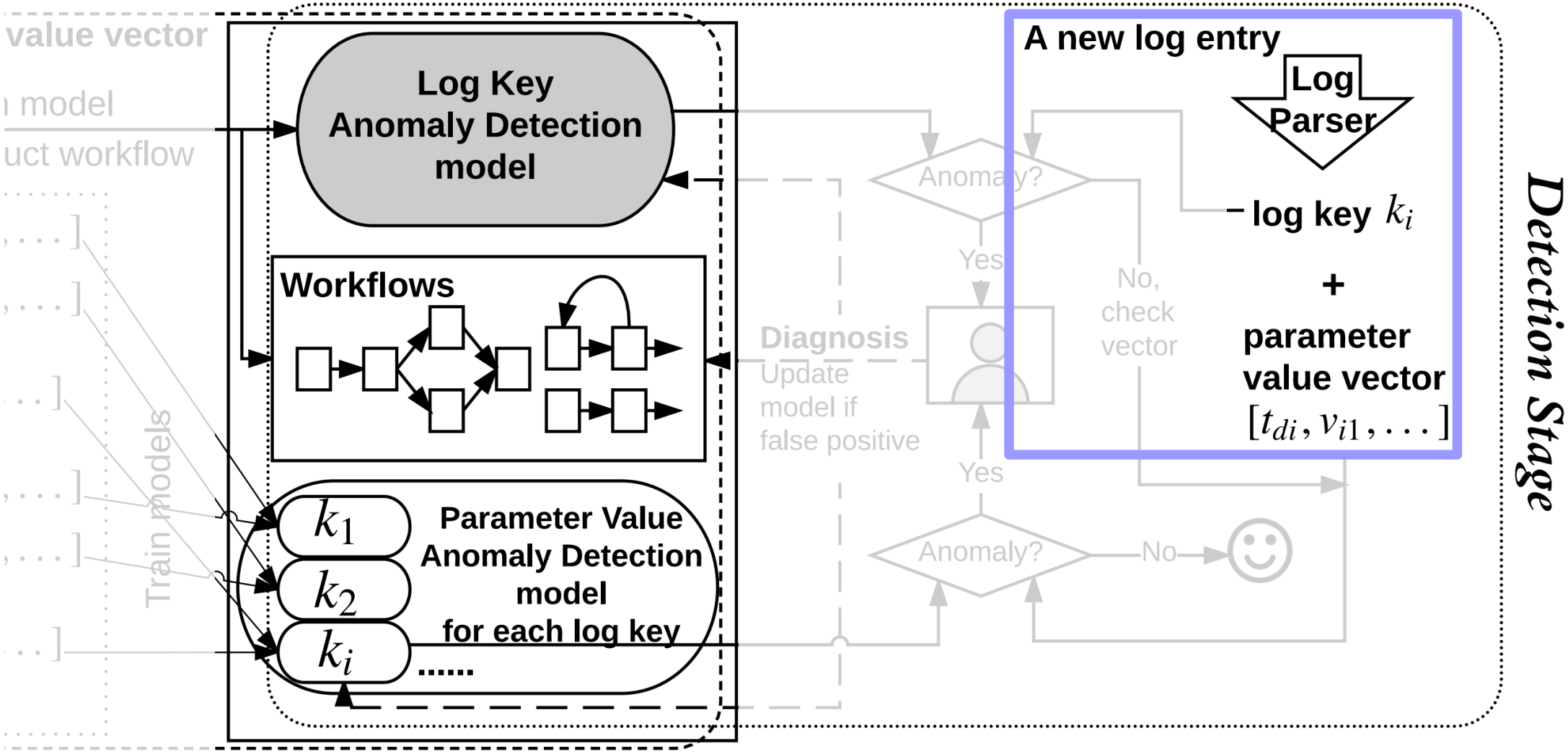
DeepLog Architecture



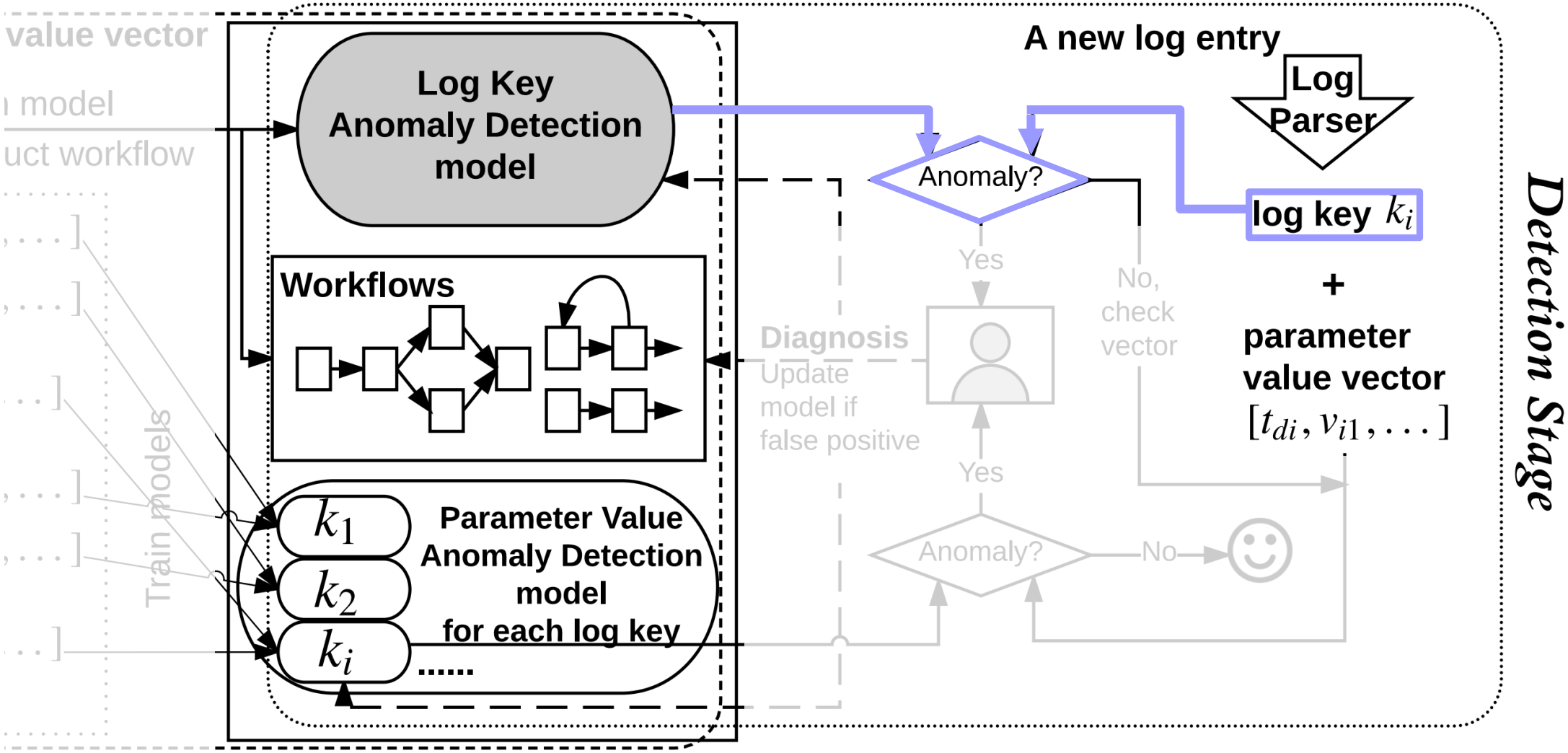
DeepLog Architecture



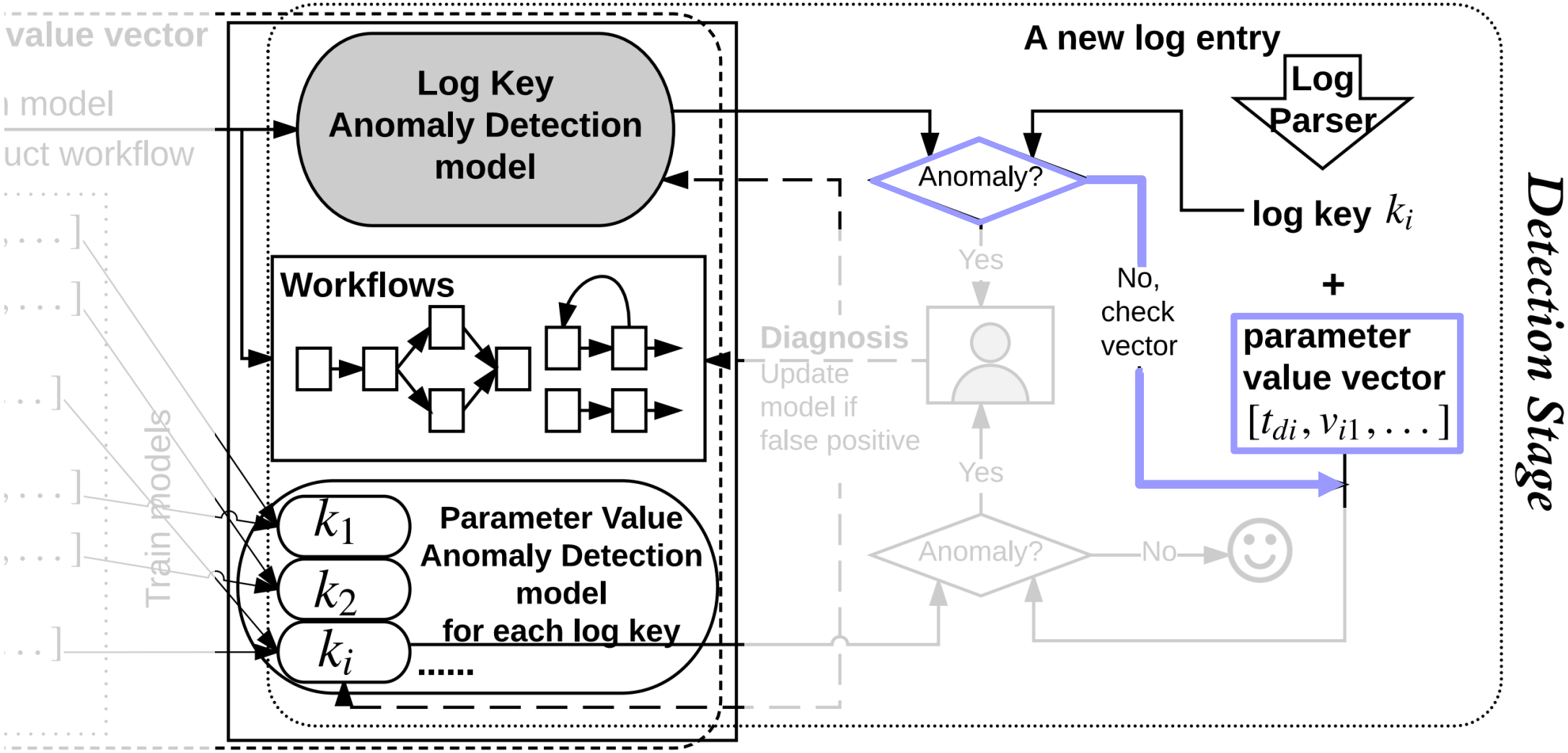
DeepLog Architecture



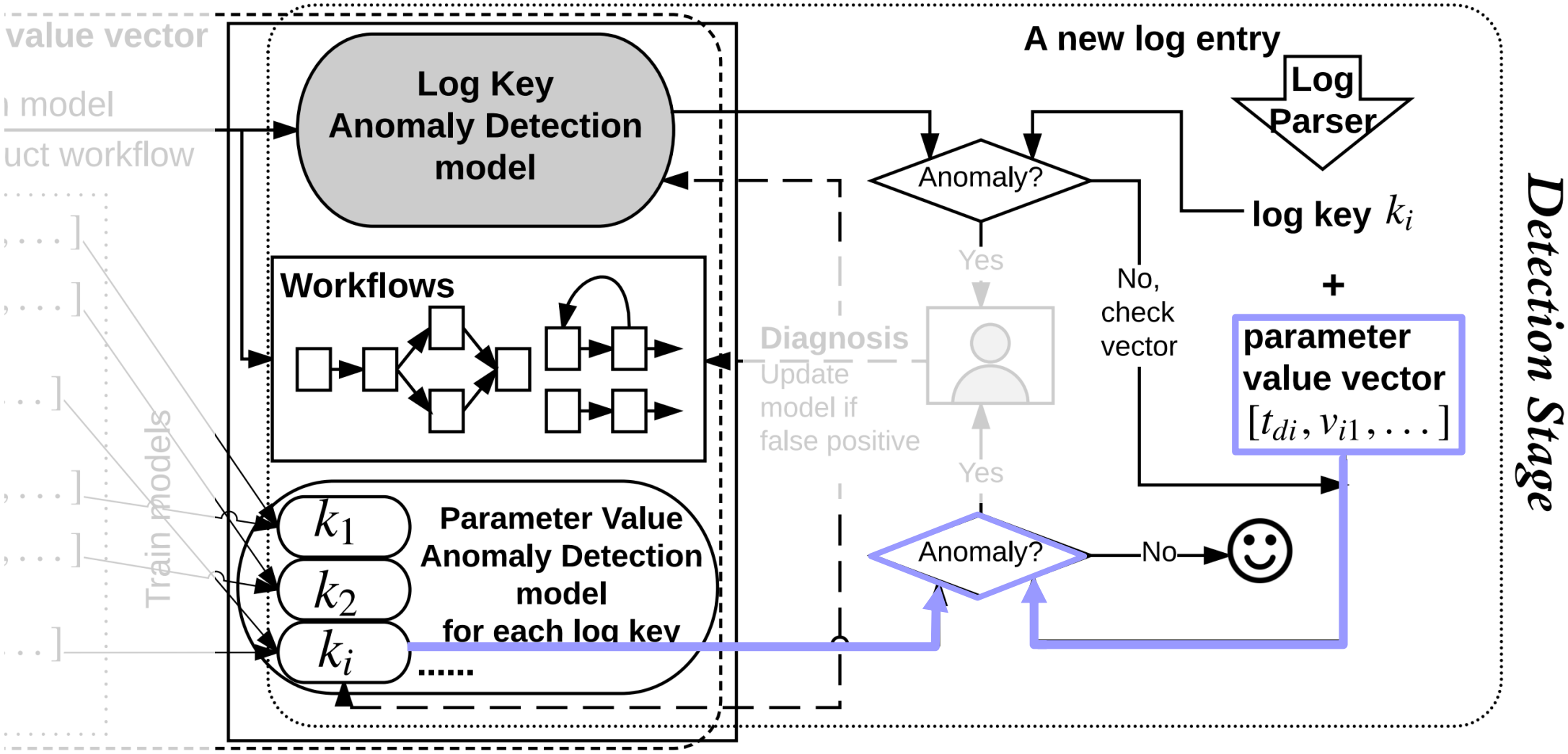
DeepLog Architecture



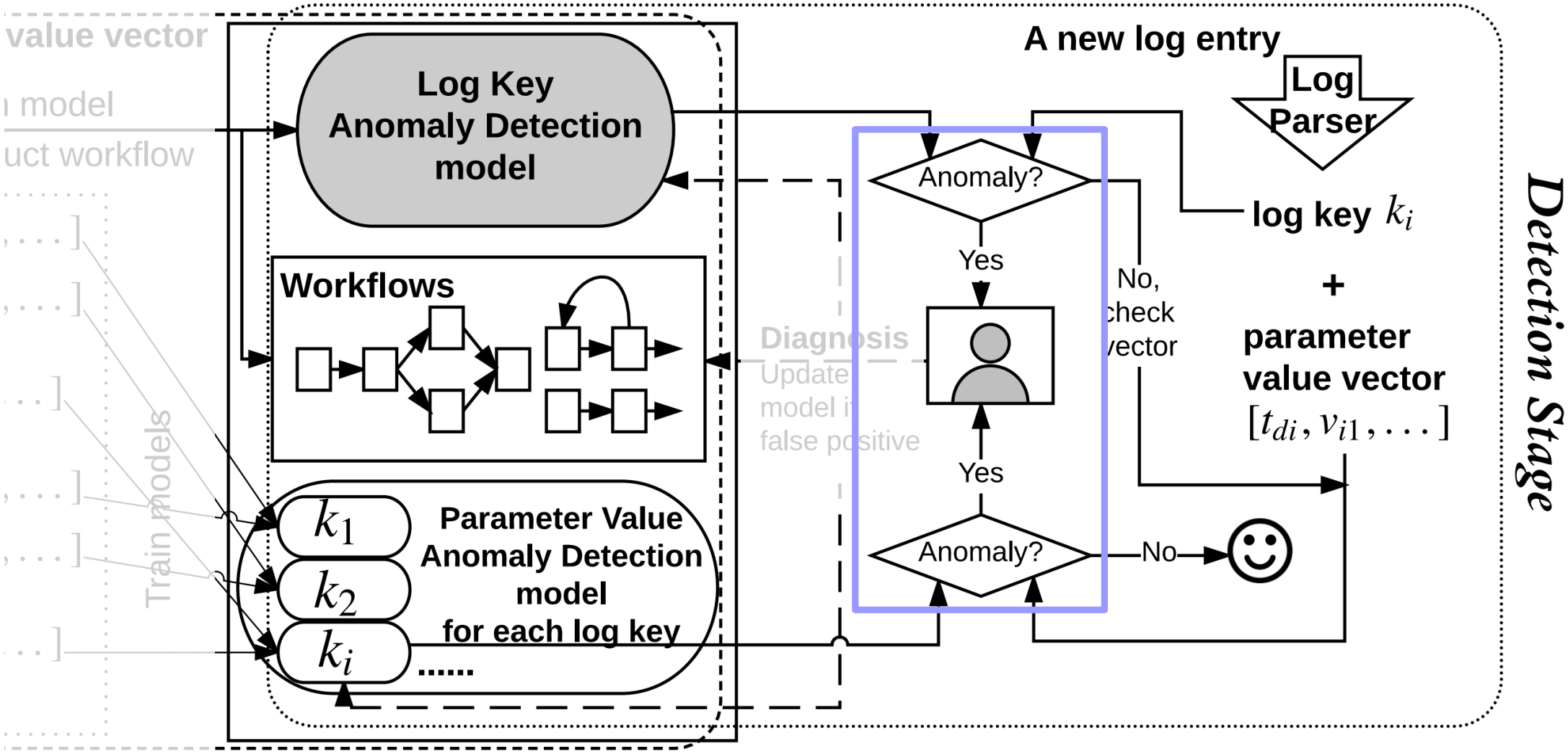
DeepLog Architecture



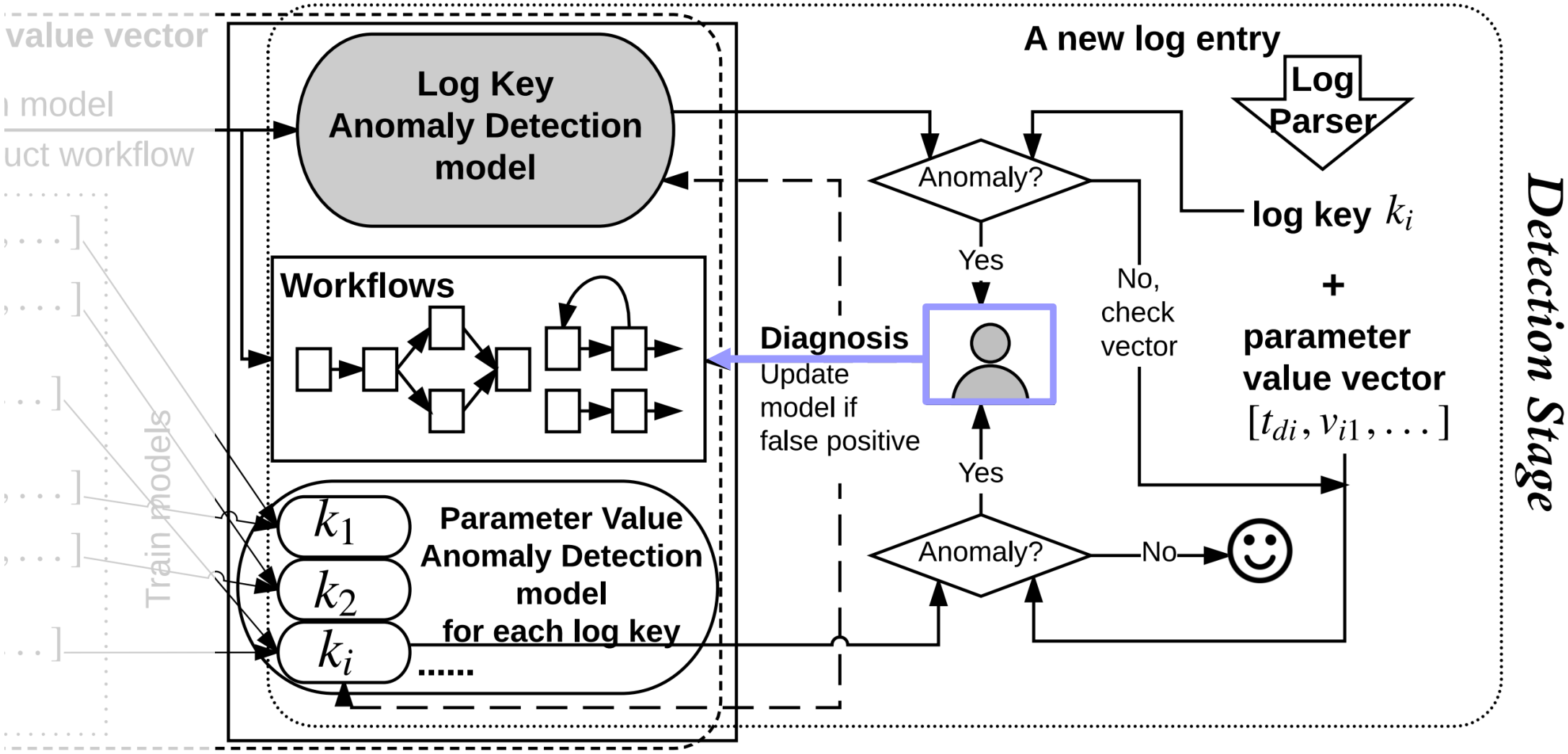
DeepLog Architecture



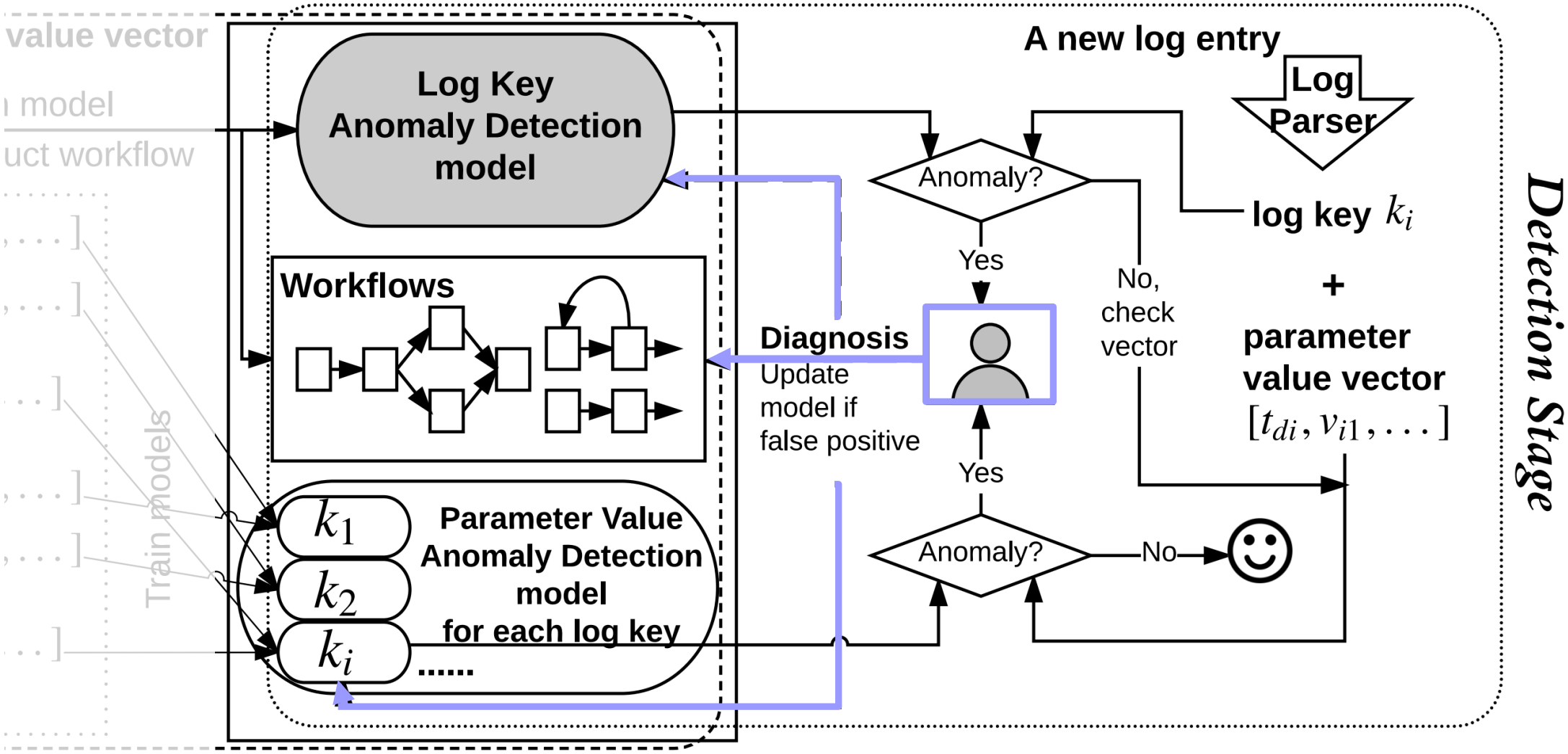
DeepLog Architecture



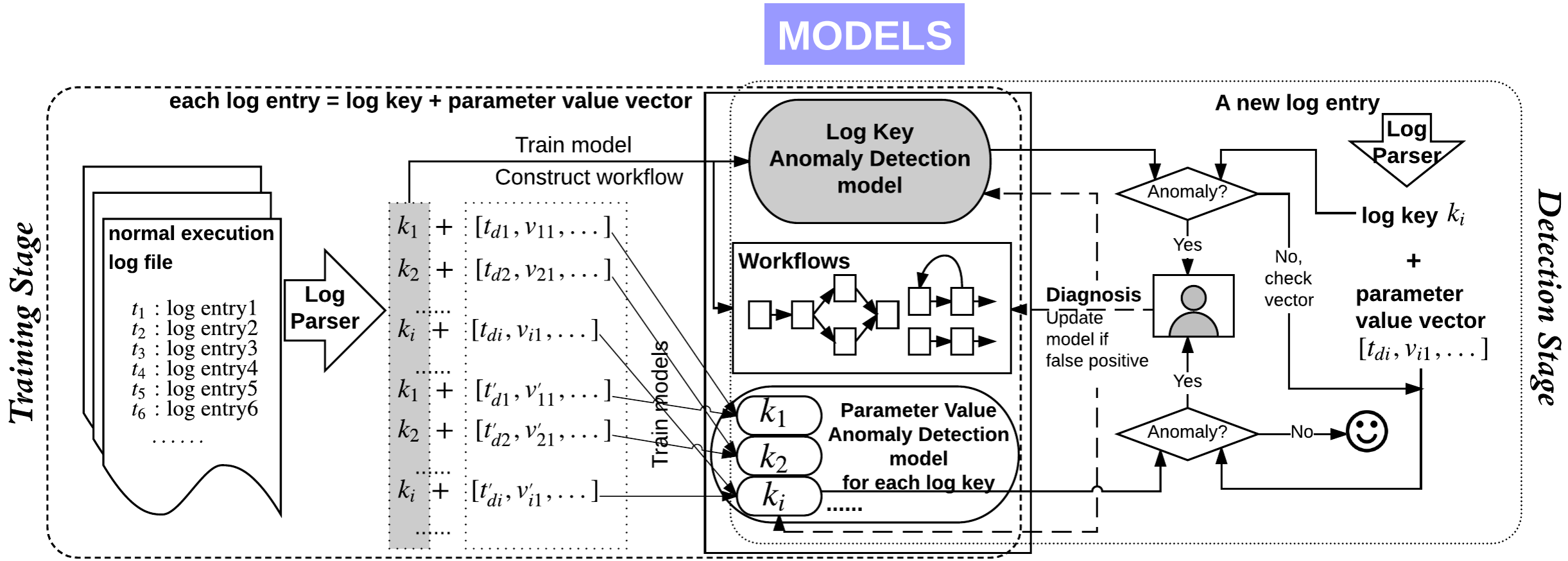
DeepLog Architecture



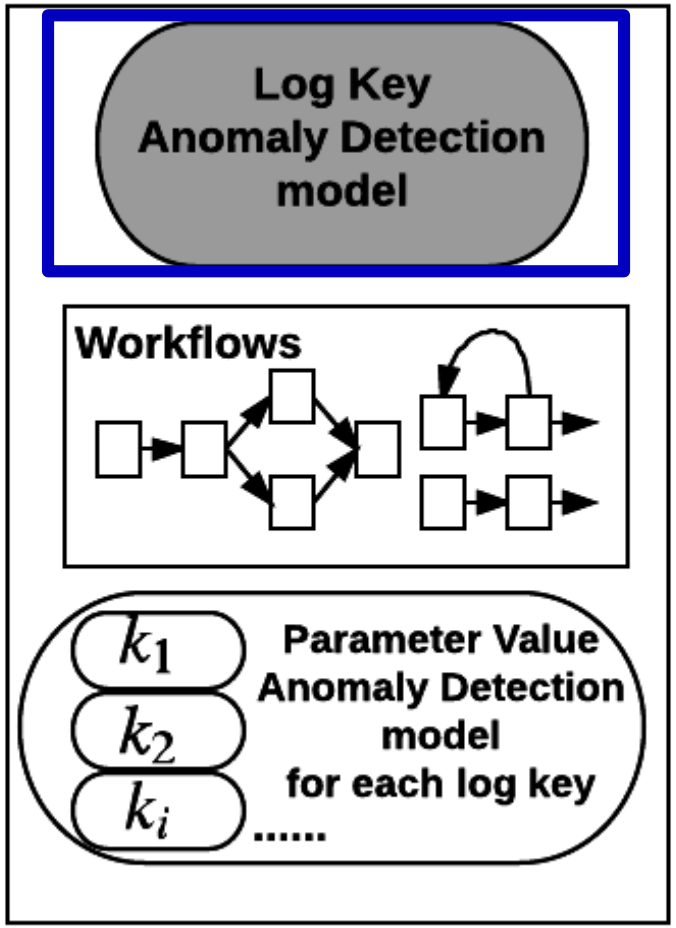
DeepLog Architecture



DeepLog Architecture



Log Key Anomaly Detection model

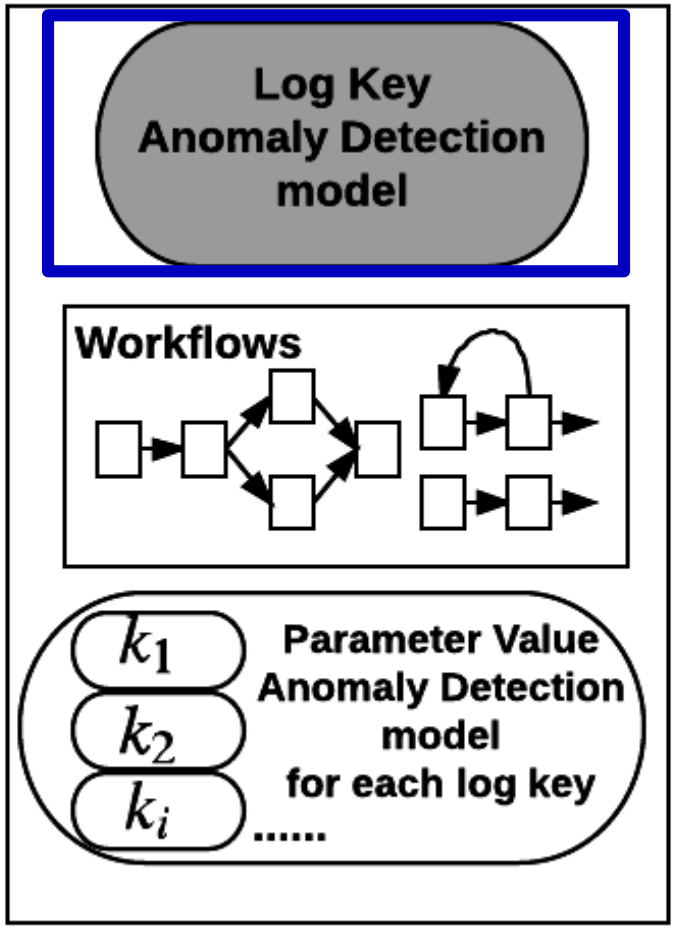


Example log key sequence:

25 18 54 57 18 56 … 25 18 54 57 56 18 …

- a rigorous set of logic and control flows
- a (*more structured*) natural language

Log Key Anomaly Detection model



Example log key sequence:

25 18 54 57 18 56 … 25 18 54 57 56 18 …

- a rigorous set of logic and control flows
- a (*more structured*) natural language

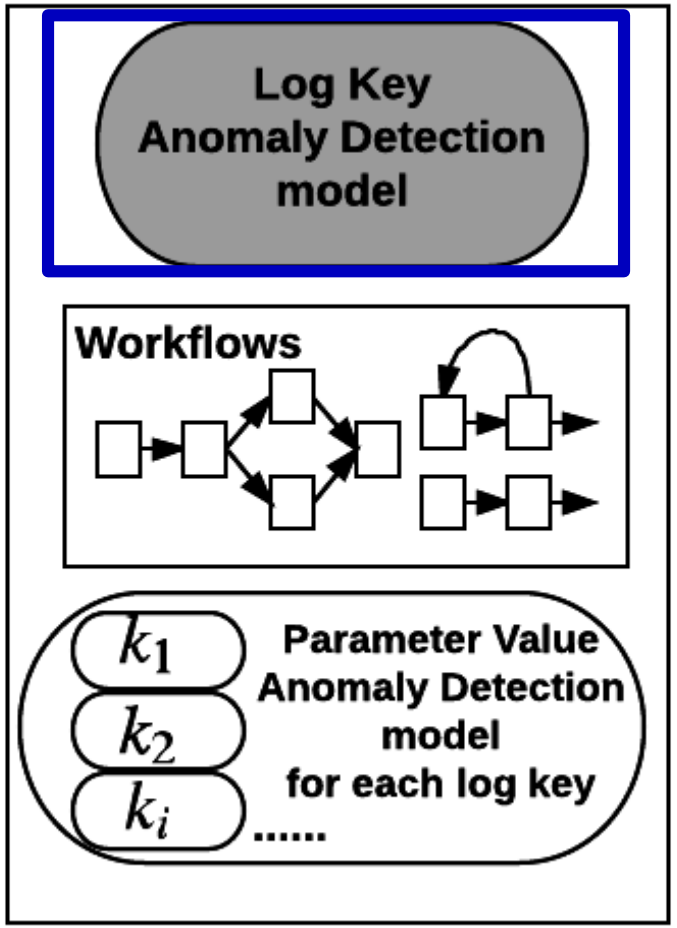
⇩

natural language modeling

⇩

multi-class classifier: *history sequence => next key to appear*

Log Key Anomaly Detection model



Example log key sequence:

25 18 54 57 18 56 … 25 18 54 57 56 18 …

- a rigorous set of logic and control flows
- a (*more structured*) natural language

⇩

natural language modeling

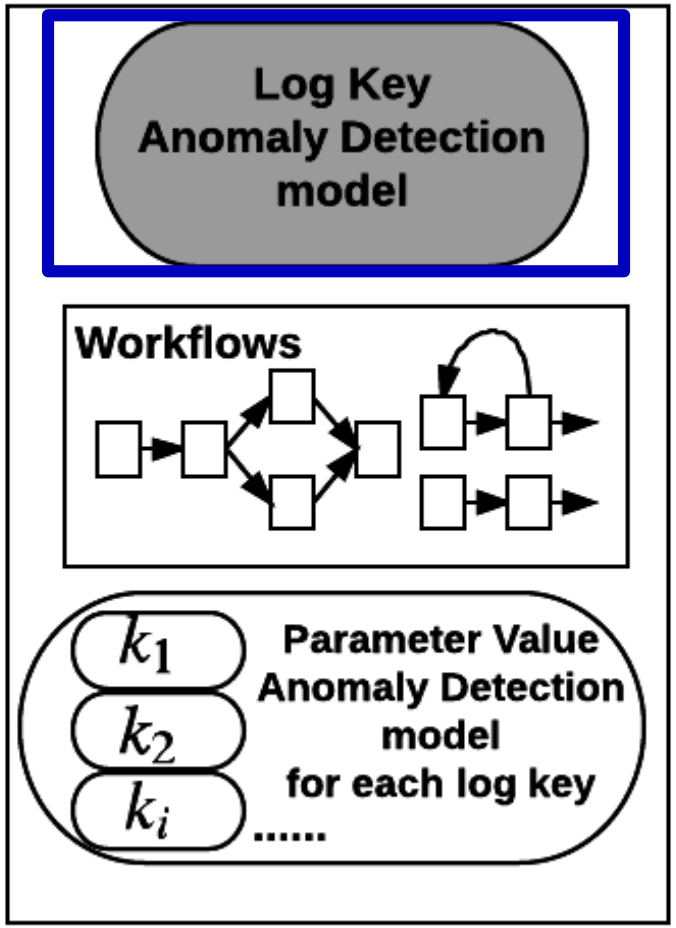
⇩

multi-class classifier: *history sequence => next key to appear*

⇩

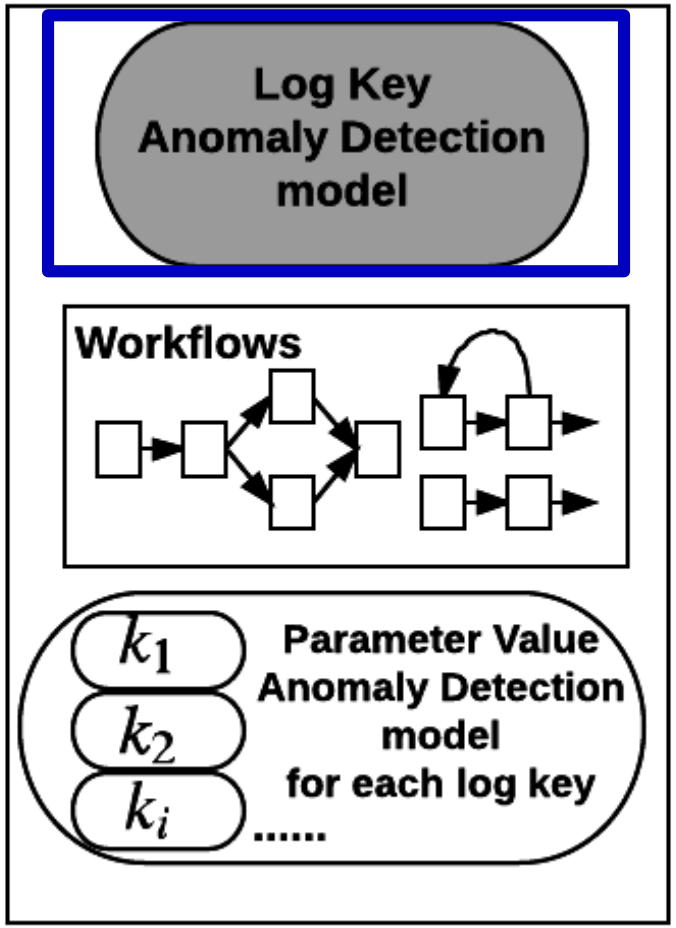
A log key is detected to be abnormal if it does not follow the prediction.

Log Key Anomaly Detection model

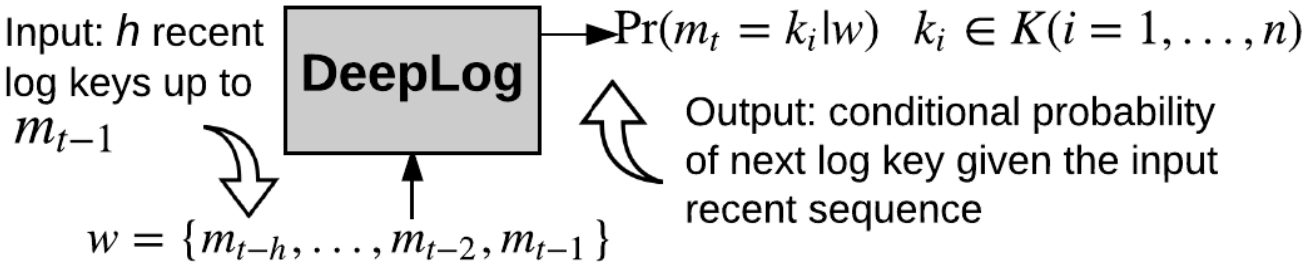


Use long short-term memory (LSTM) architecture

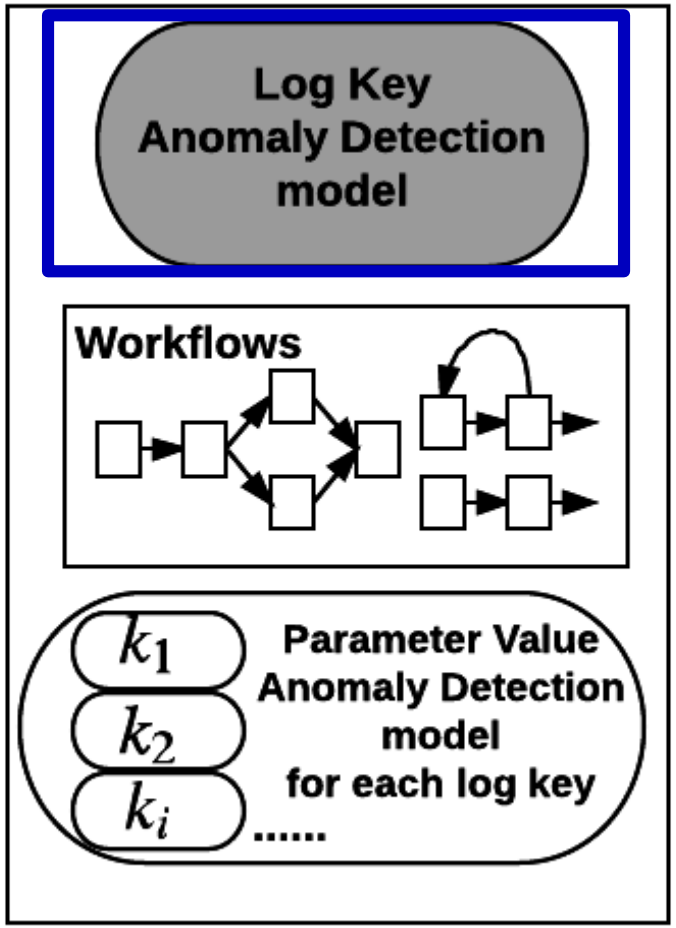
Log Key Anomaly Detection model



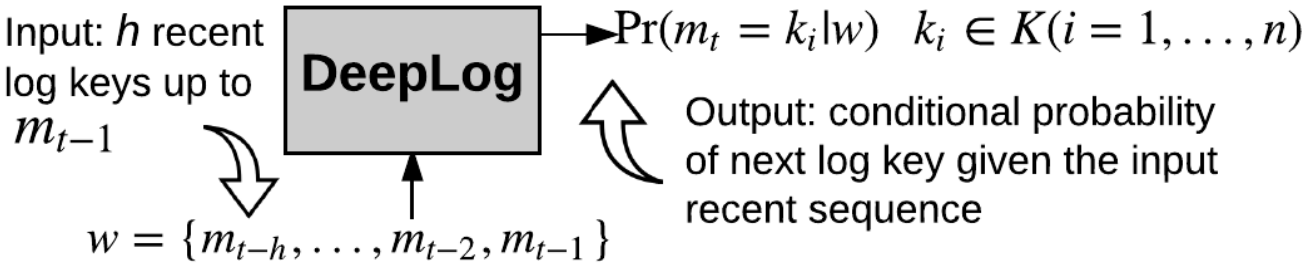
Use long short-term memory (LSTM) architecture



Log Key Anomaly Detection model



Use long short-term memory (LSTM) architecture

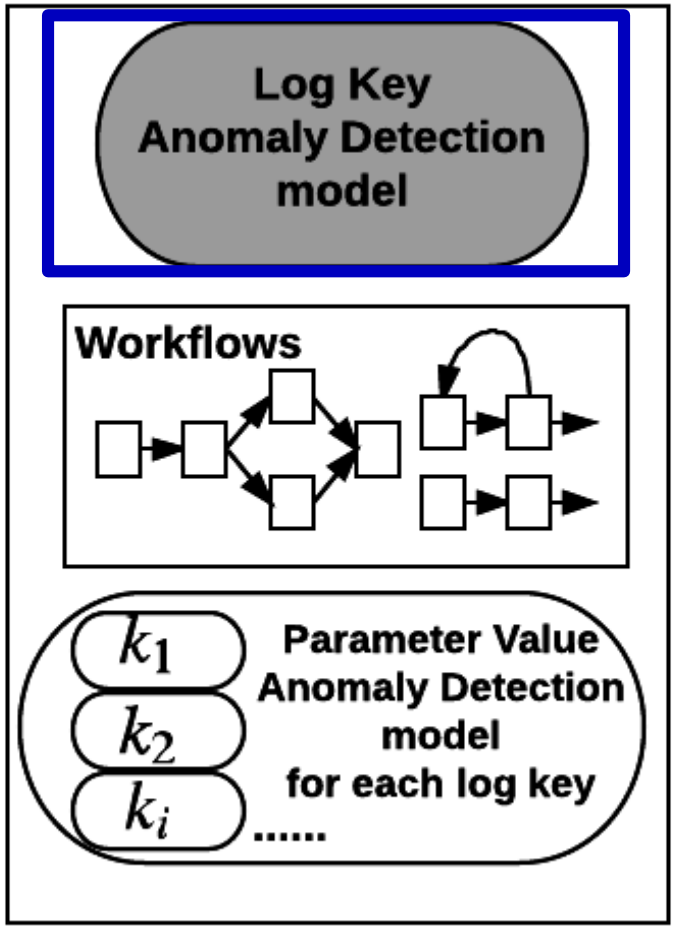


Training:

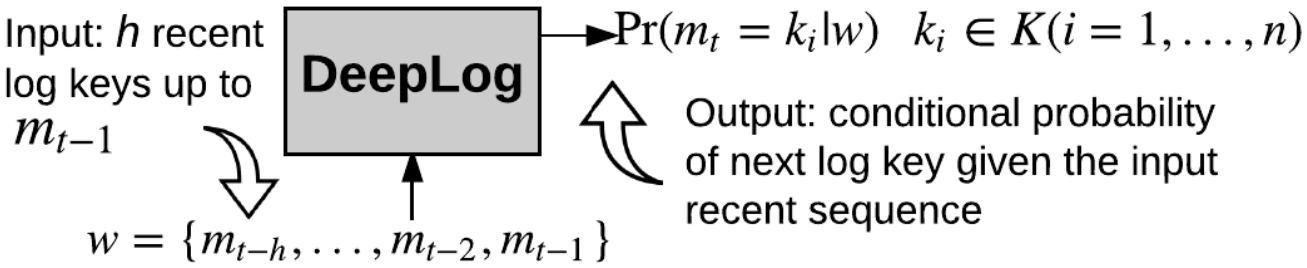
log key sequence:

h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



Use long short-term memory (LSTM) architecture

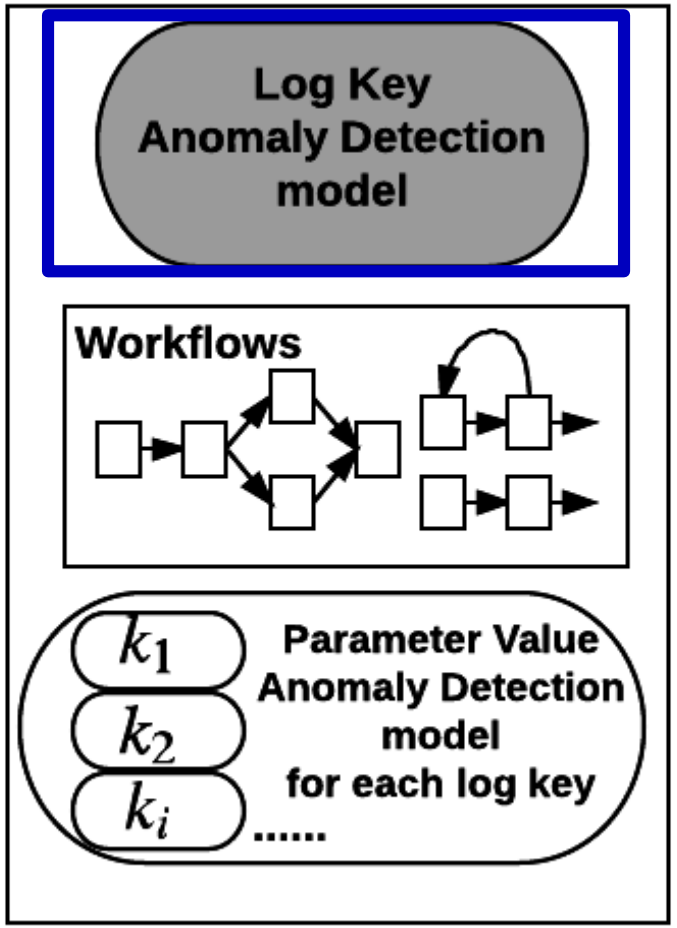


Training:

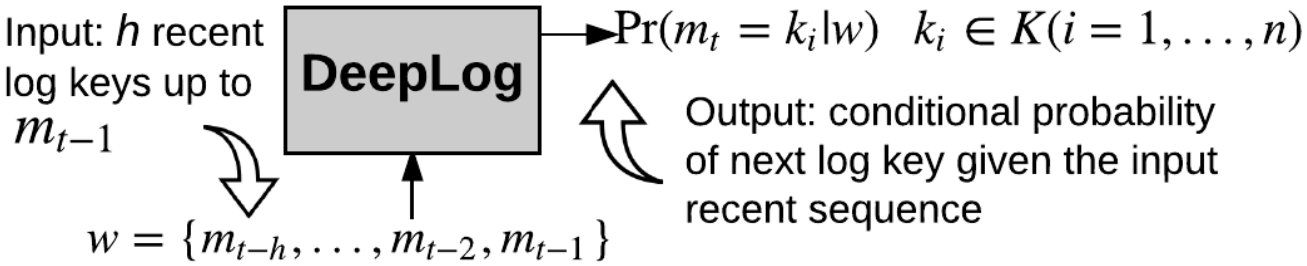
log key sequence:

h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



Use long short-term memory (LSTM) architecture

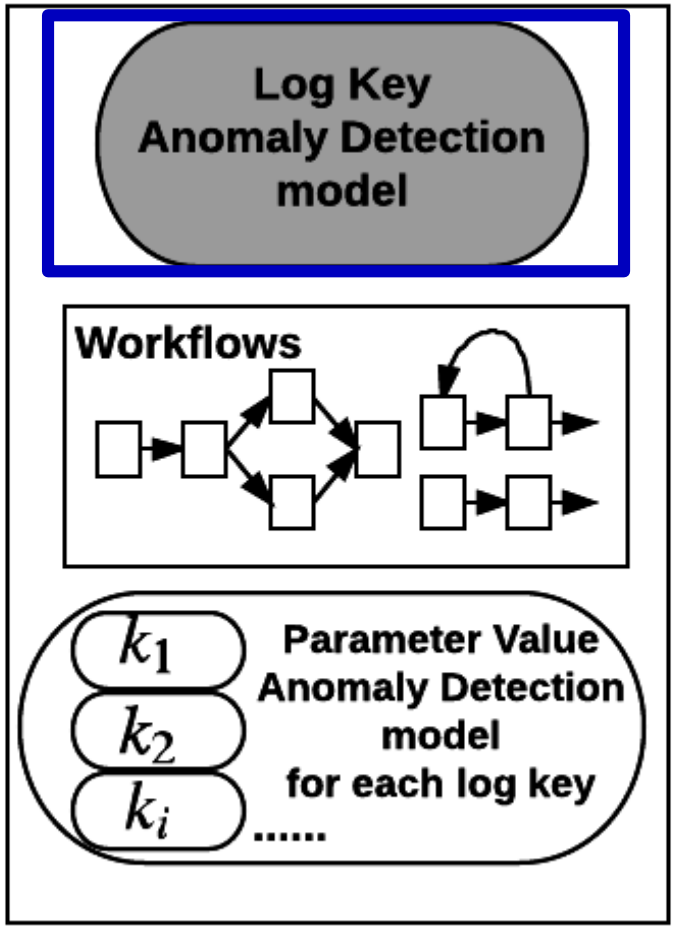


Training:

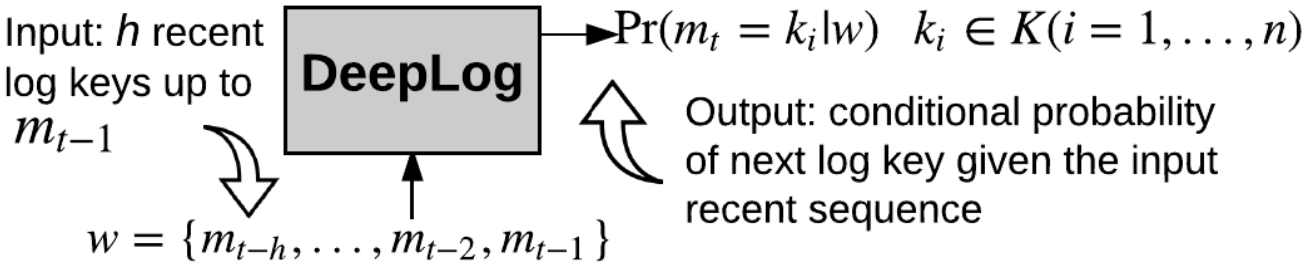
log key sequence:

h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



Use long short-term memory (LSTM) architecture

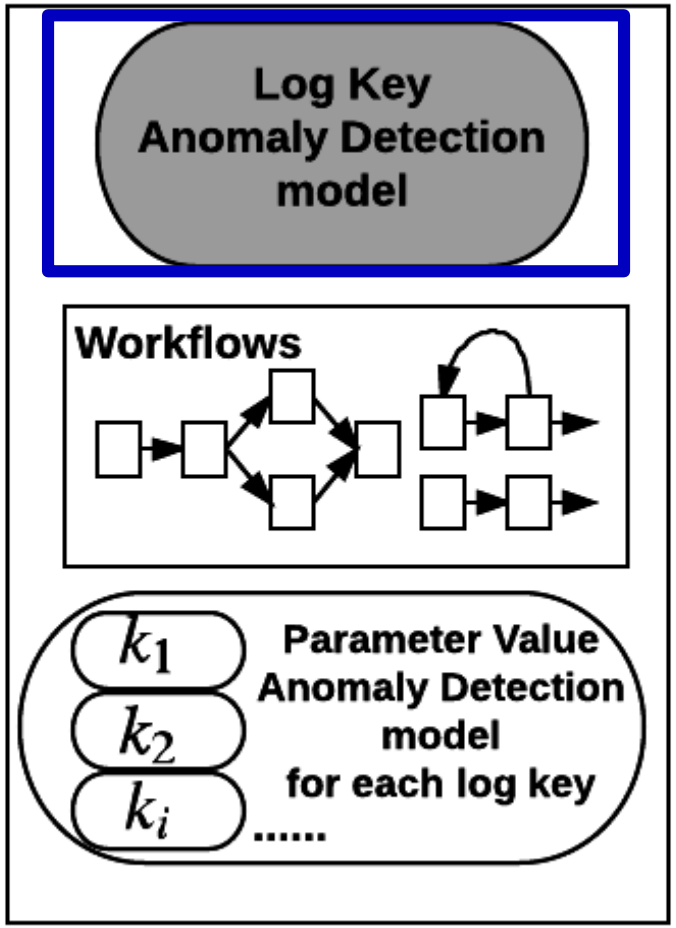


Training:

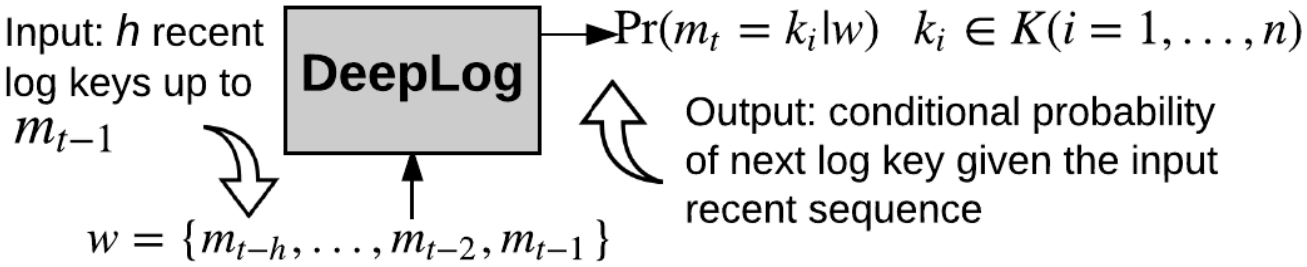
log key sequence:

h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



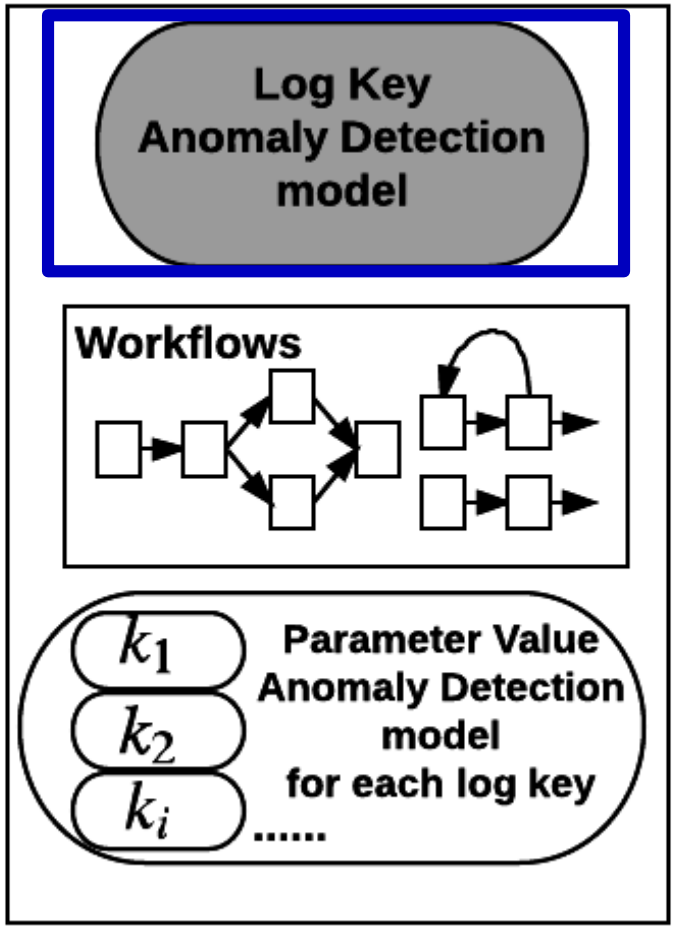
Use long short-term memory (LSTM) architecture

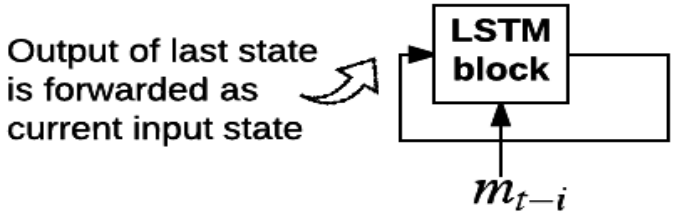


Detection:

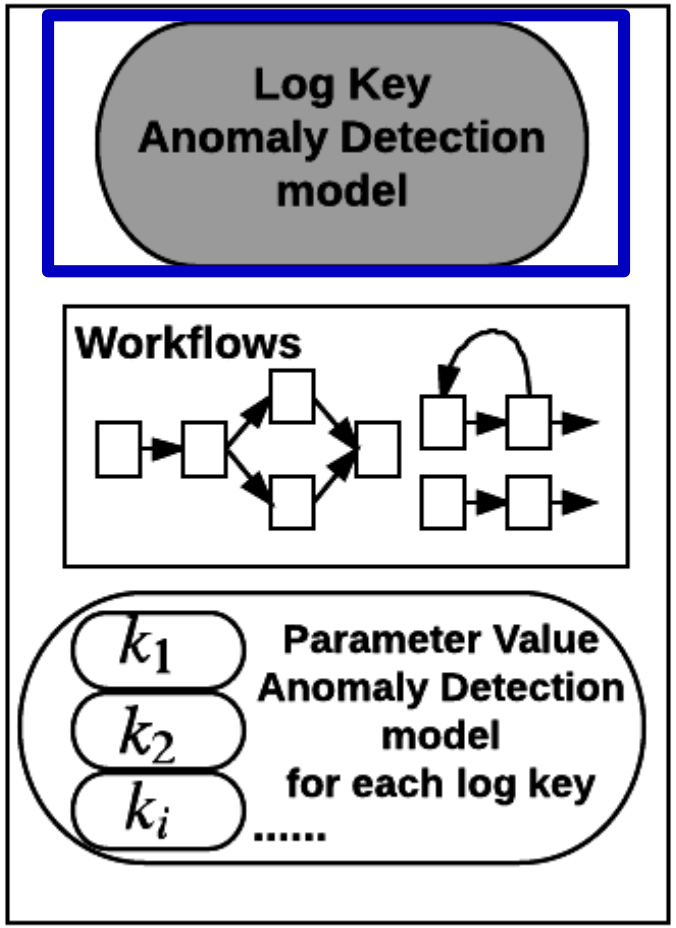
In detection stage, DeepLog checks if the actual next log key is among its top *g* probable predictions.

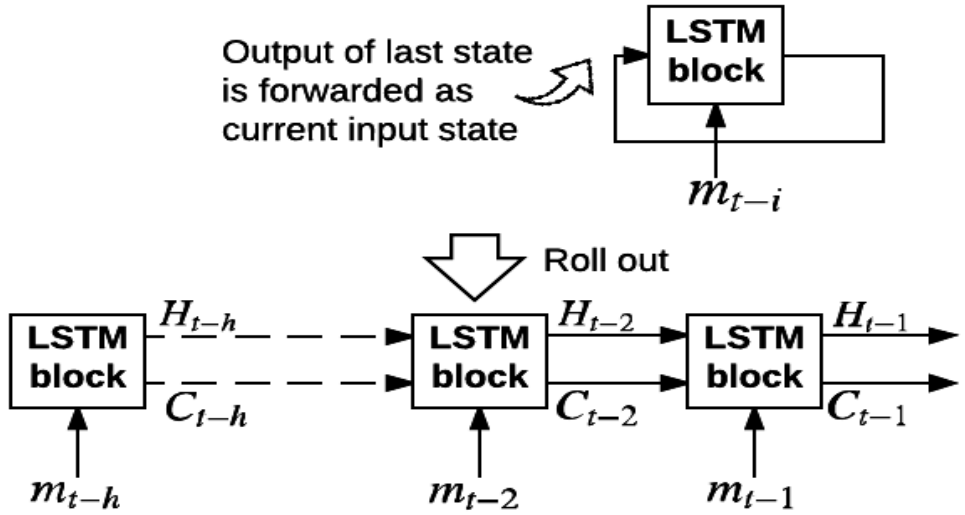
Log Key Anomaly Detection model



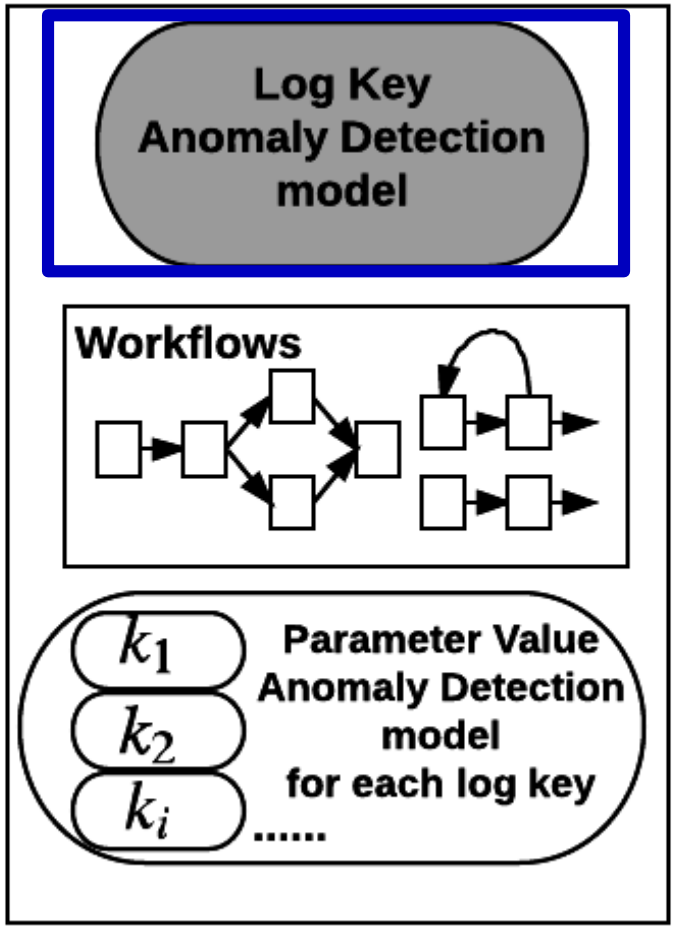


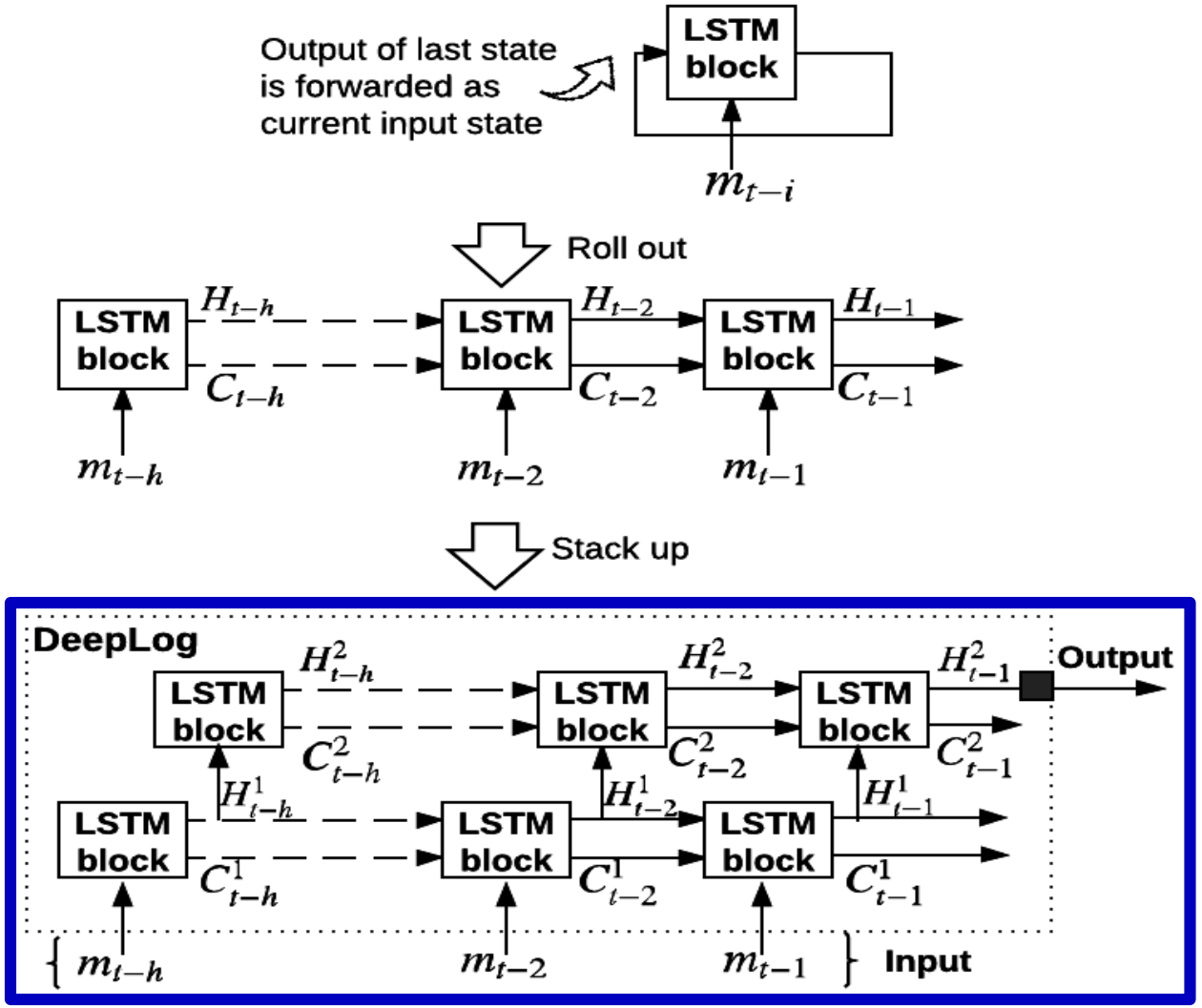
Log Key Anomaly Detection model



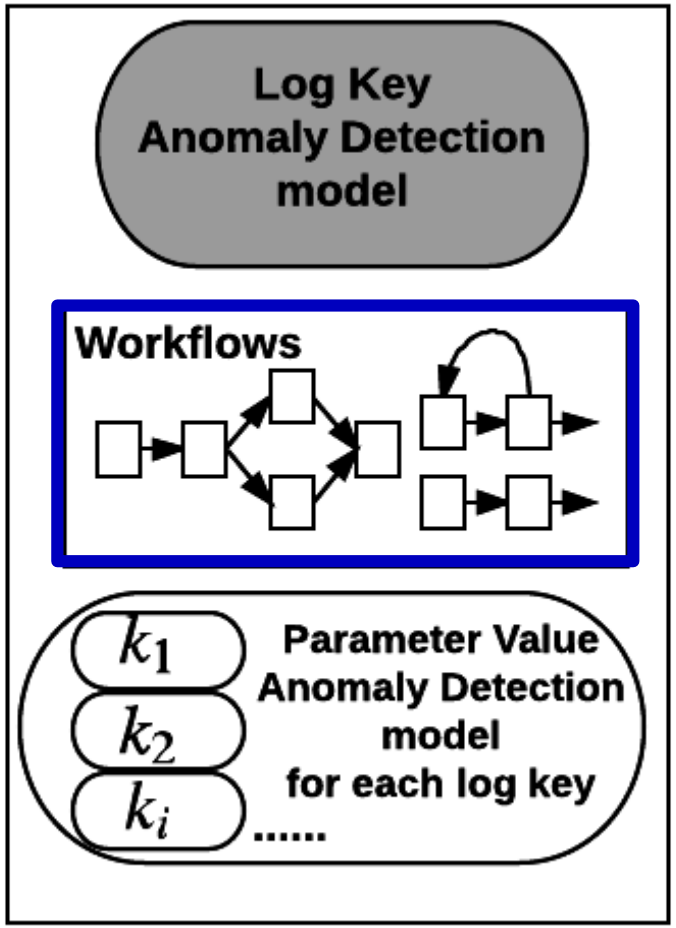


Log Key Anomaly Detection model





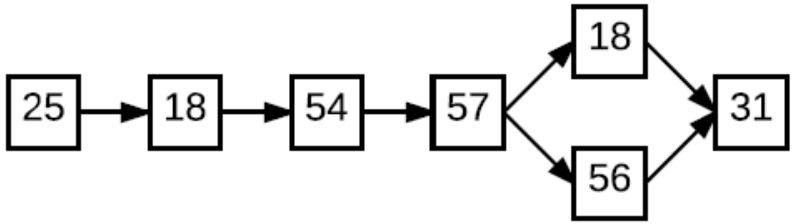
Workflow Construction



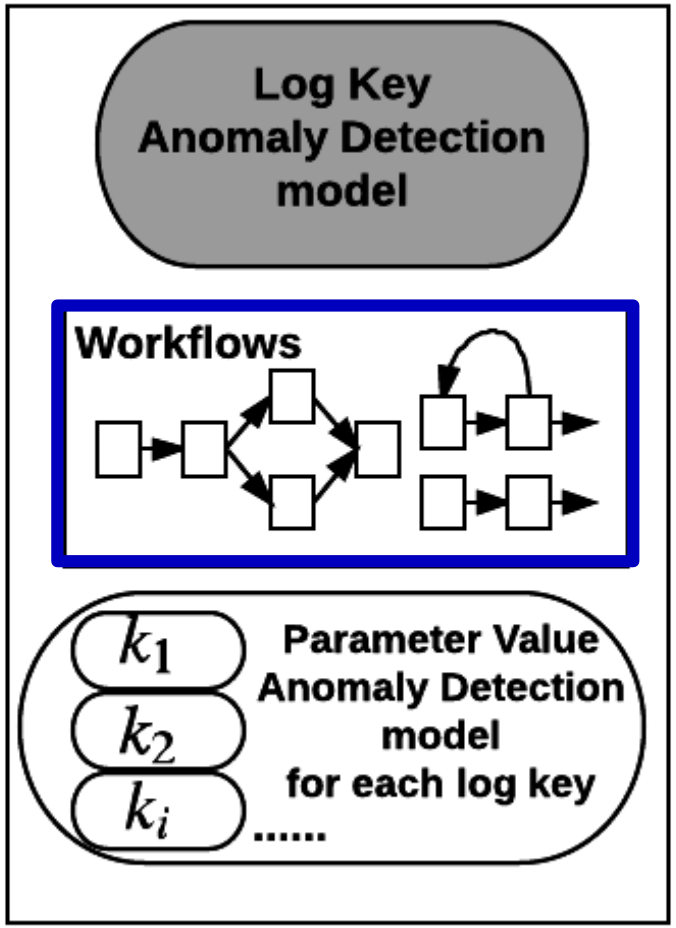
Input: log key sequence

25 18 54 57 18 56 … 25 18 54 57 56 18 …

Output:



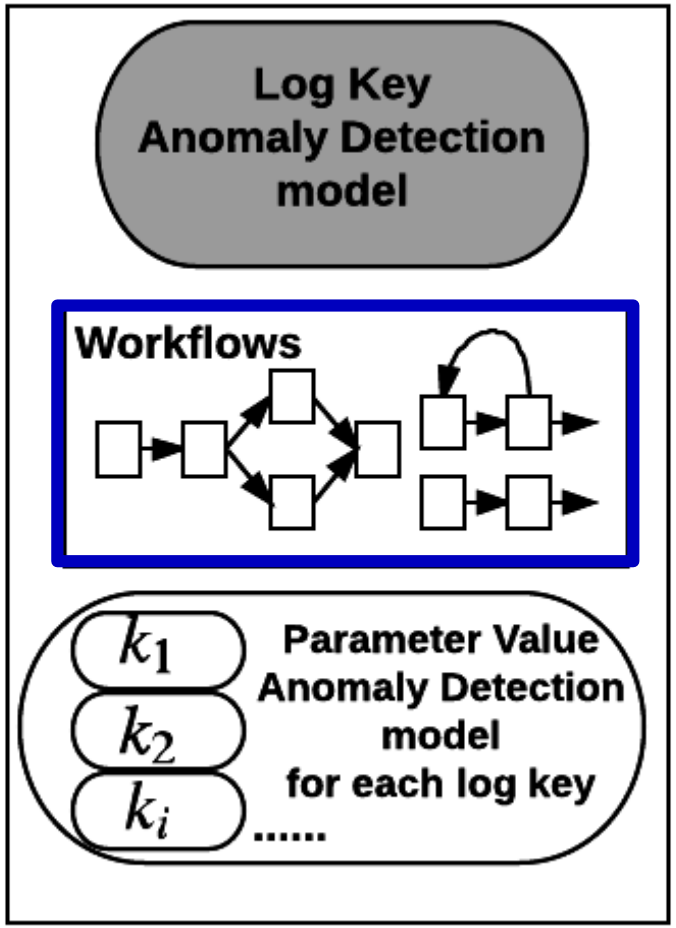
Workflow Construction



Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

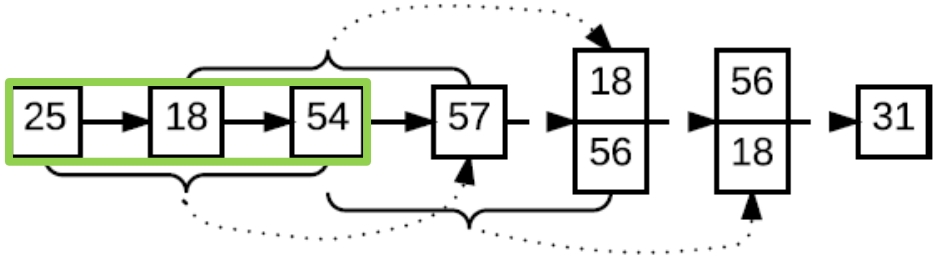
Workflow Construction



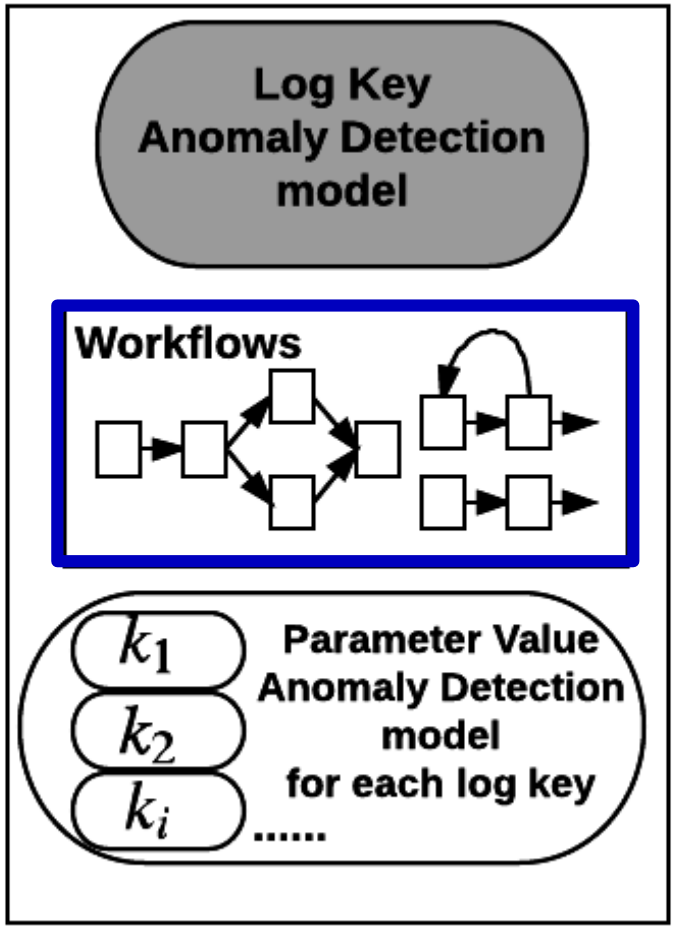
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



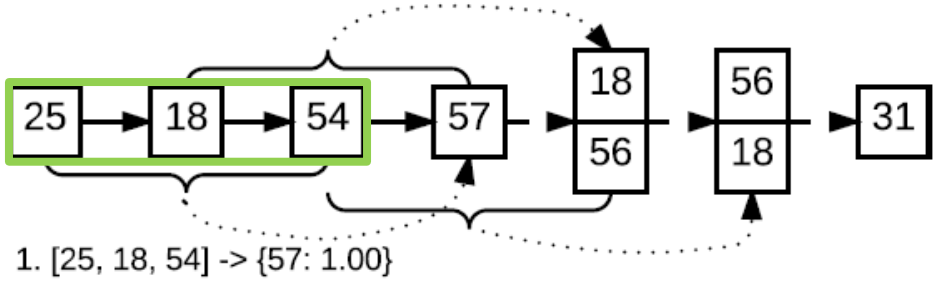
Workflow Construction



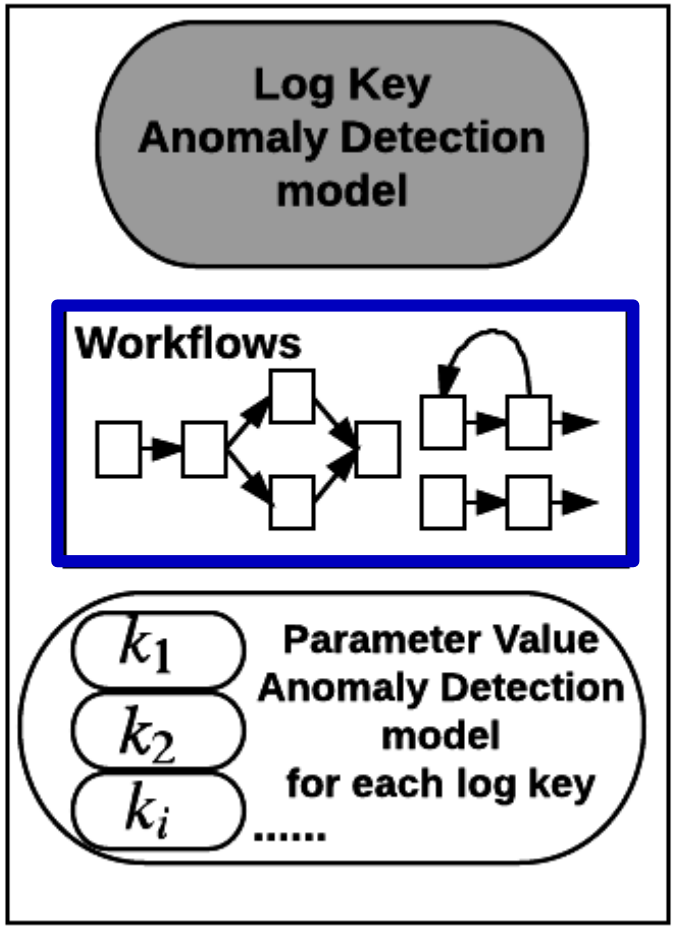
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



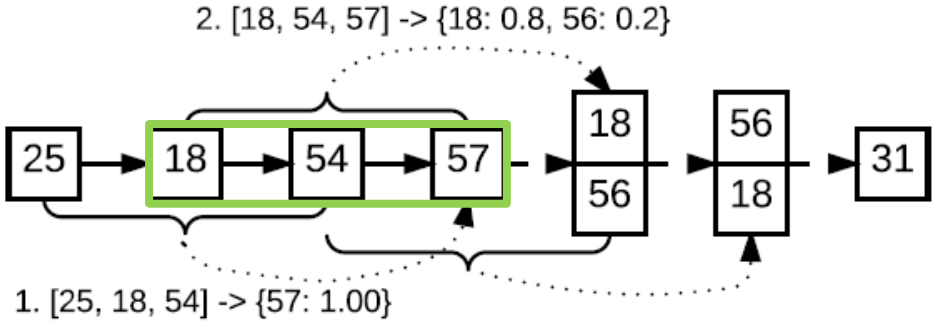
Workflow Construction



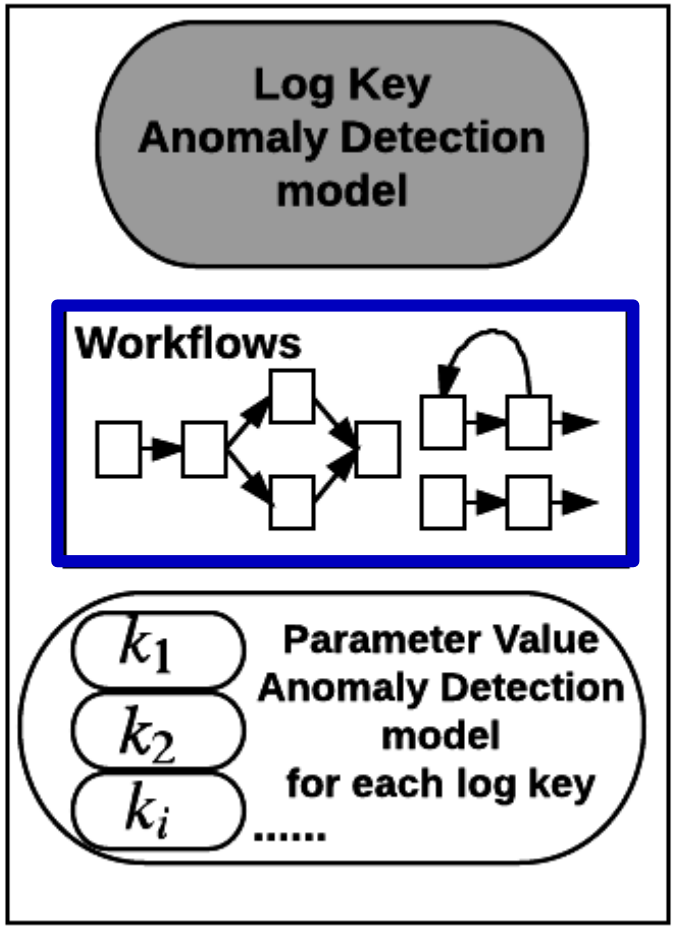
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



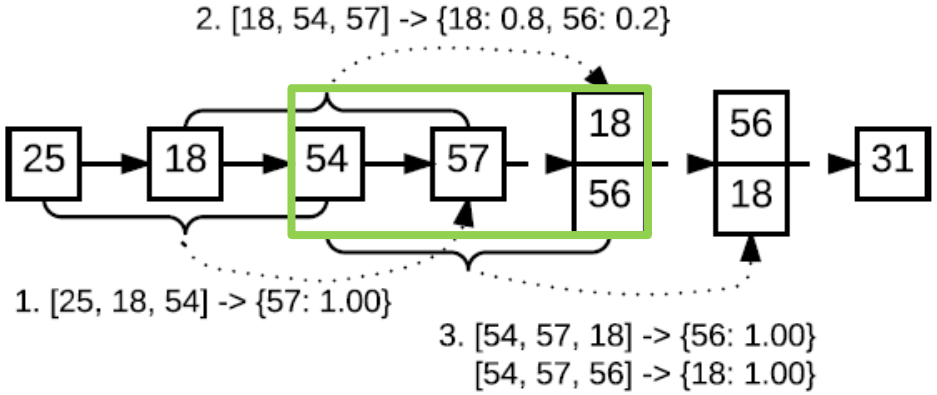
Workflow Construction



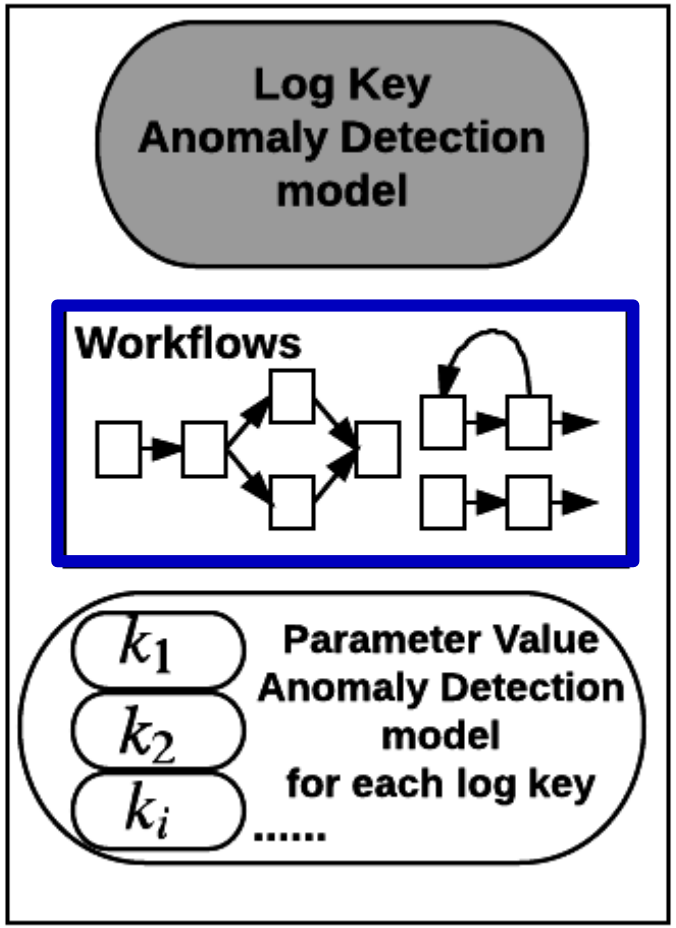
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



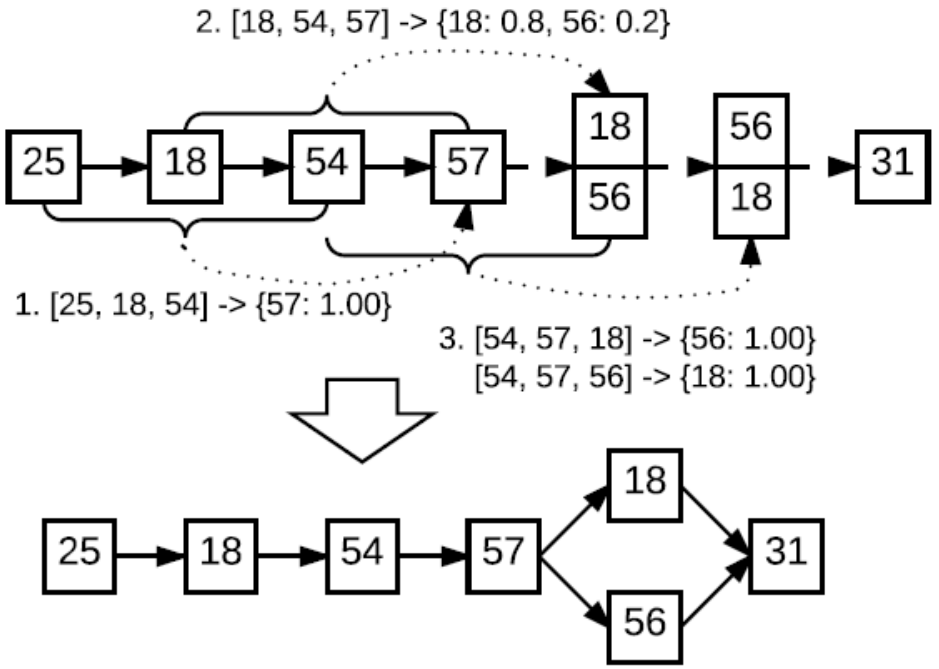
Workflow Construction



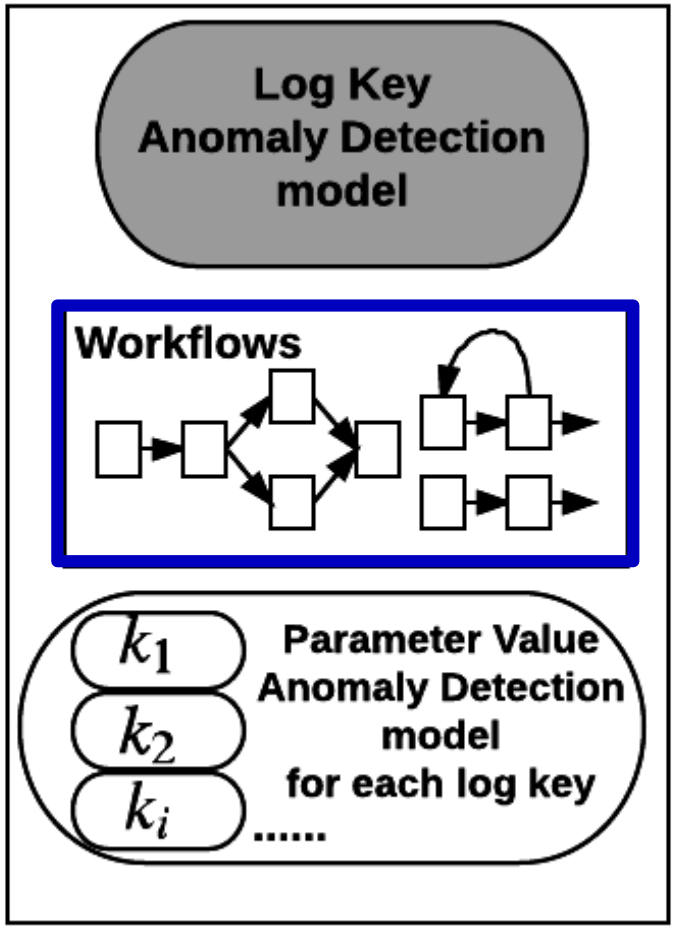
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:

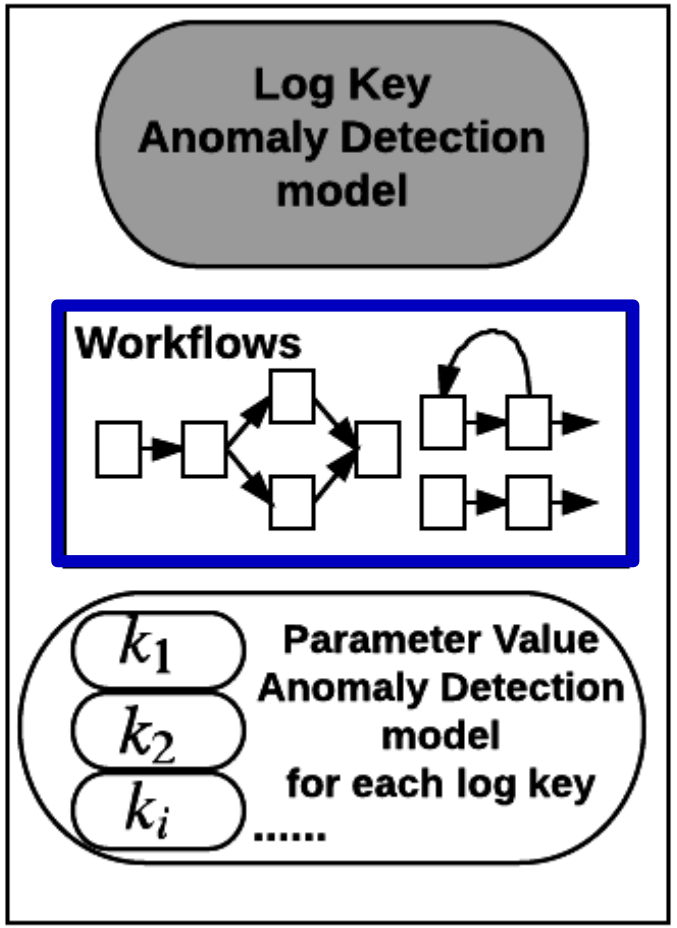


Workflow Construction



Method 2: A density-based clustering approach

Workflow Construction



Method 2: A density-based clustering approach

Co-occurrence matrix of log keys (k_i, k_j) within distance d

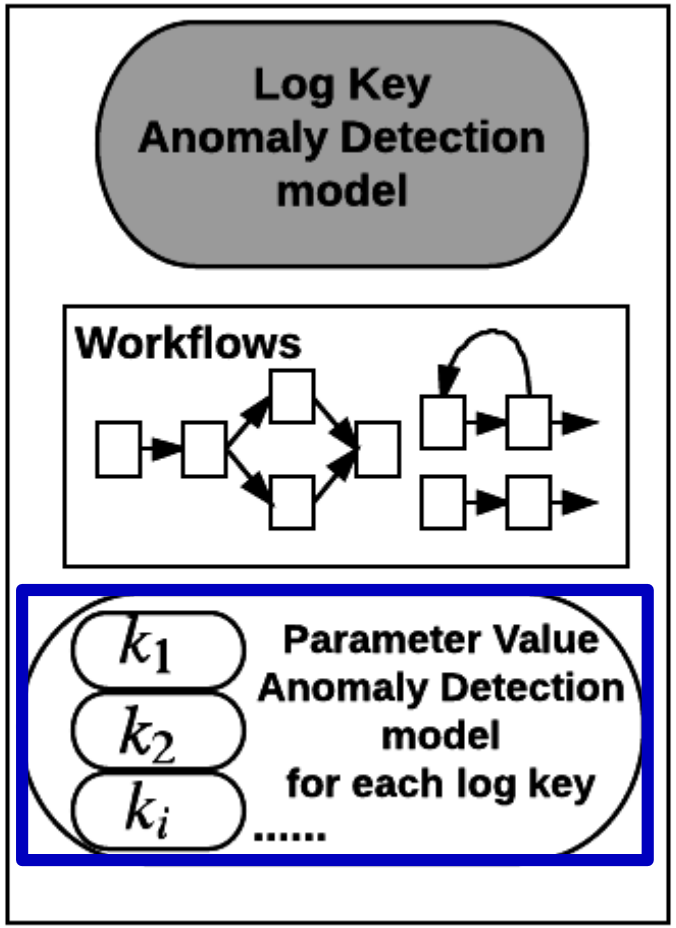
	k_1	...	k_j	...	k_n
k_1	$p_d(1, 1)$		$p_d(1, j)$		
...					
k_i	$p_d(i, 1)$		$p_d(i, j) = \frac{f_d(k_i, k_j)}{d \cdot f(k_i)}$		
...					
k_n	$p_d(n, 1)$		$p_d(n, j)$		

$f_d(k_i, k_j)$: the frequency of (k_i, k_j) appearing together within distance d

$f(k_i)$: the frequency of k_i in the input sequence

$p_d(i, j)$: the probability of (k_i, k_j) appearing together within distance d

Parameter Value Anomaly Detection model



Example:

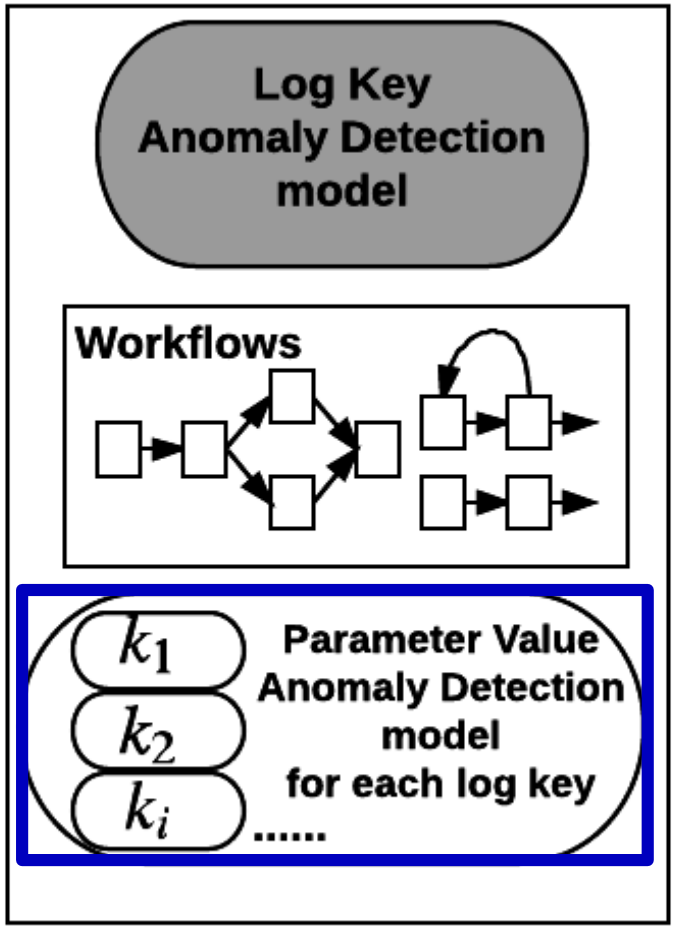
Log messages of a particular log key:

t_2: $Took\ 0.61\ seconds\ to\ deallocate\ network$ ...

t'_2: $Took\ 1.1\ seconds\ to\ deallocate\ network$ ...

....

Parameter Value Anomaly Detection model



Example:

Log messages of a particular log key:

t_2: $Took\ 0.61\ seconds\ to\ deallocate\ network$ ...

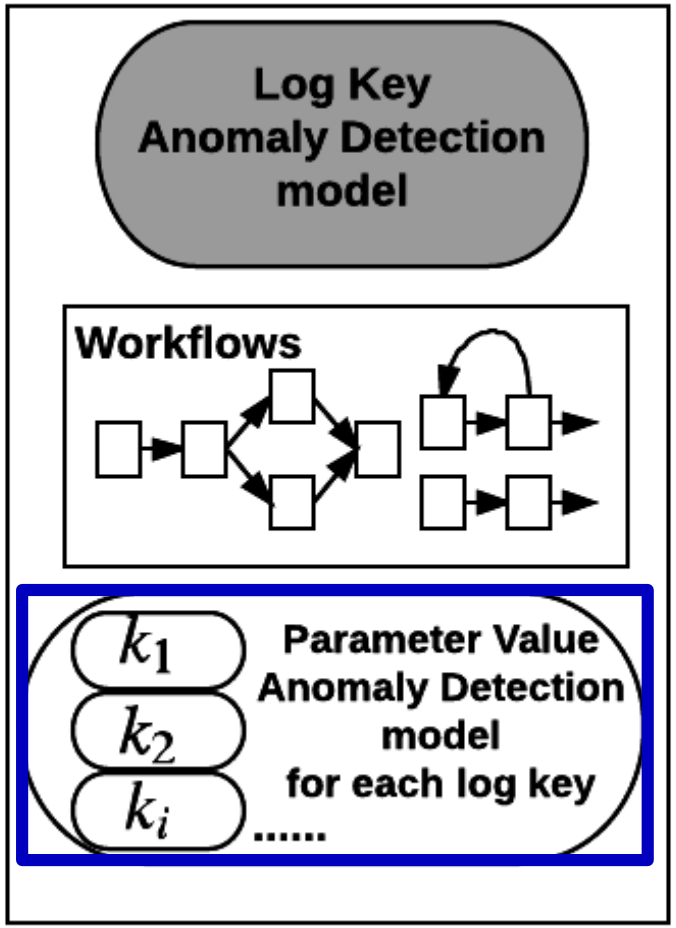
t'_2: $Took\ 1.1\ seconds\ to\ deallocate\ network$ ...

....

Parameter value vectors overtime:

[t_2- t_1, 0.61], [t'_2- t'_1, 1.1], ...

Parameter Value Anomaly Detection model



Example:

Log messages of a particular log key:

t_2: *Took* 0.61 *seconds to deallocate network* ...

t'_2: *Took* 1.1 *seconds to deallocate network* ...

....

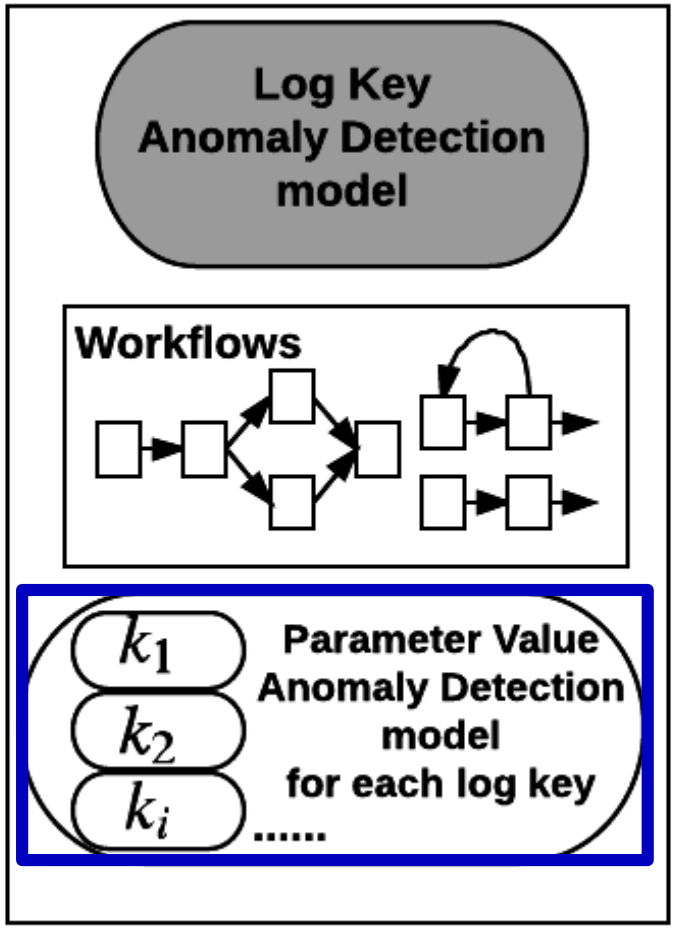
Parameter value vectors overtime:

[t_2- t_1, 0.61], [t'_2- t'_1, 1.1], ...

The time difference between current log and preceding log (they might have different log keys)

Multi-variate time series data anomaly detection problem!

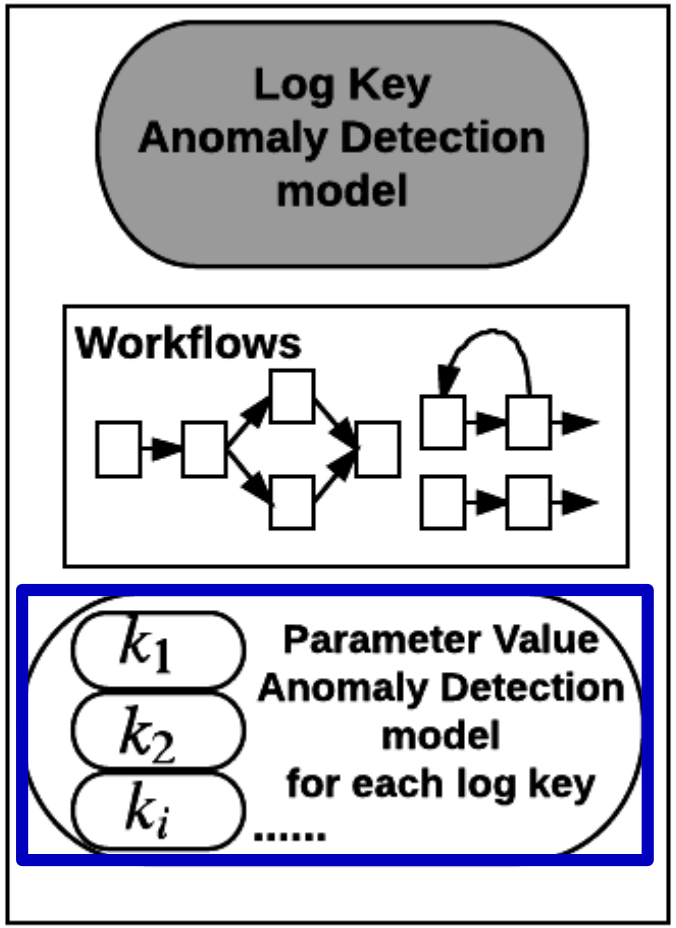
Parameter Value Anomaly Detection model



Multi-variate time series data anomaly detection problem

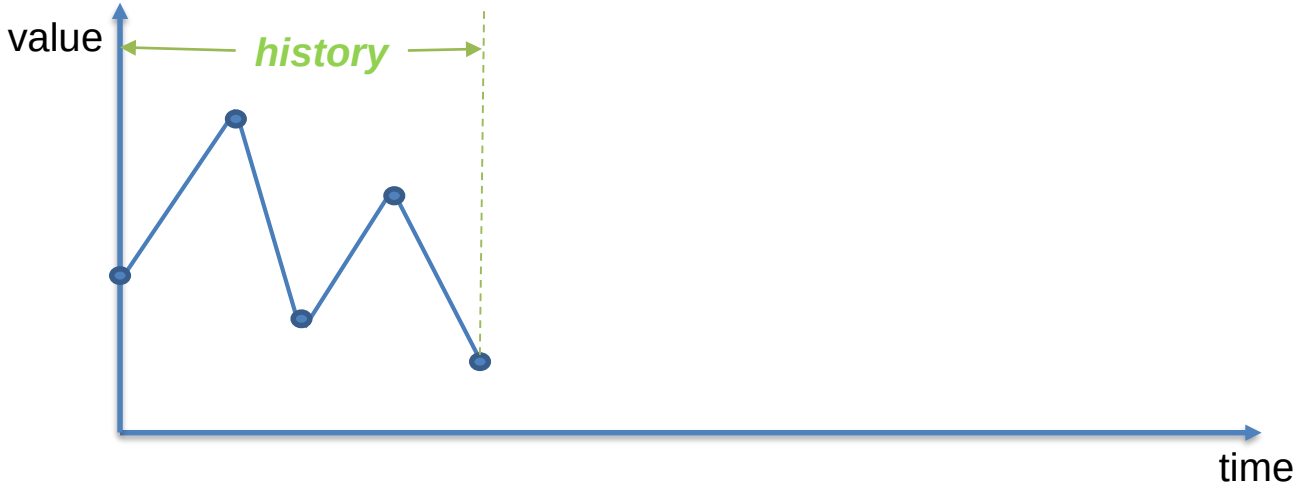
- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

Parameter Value Anomaly Detection model

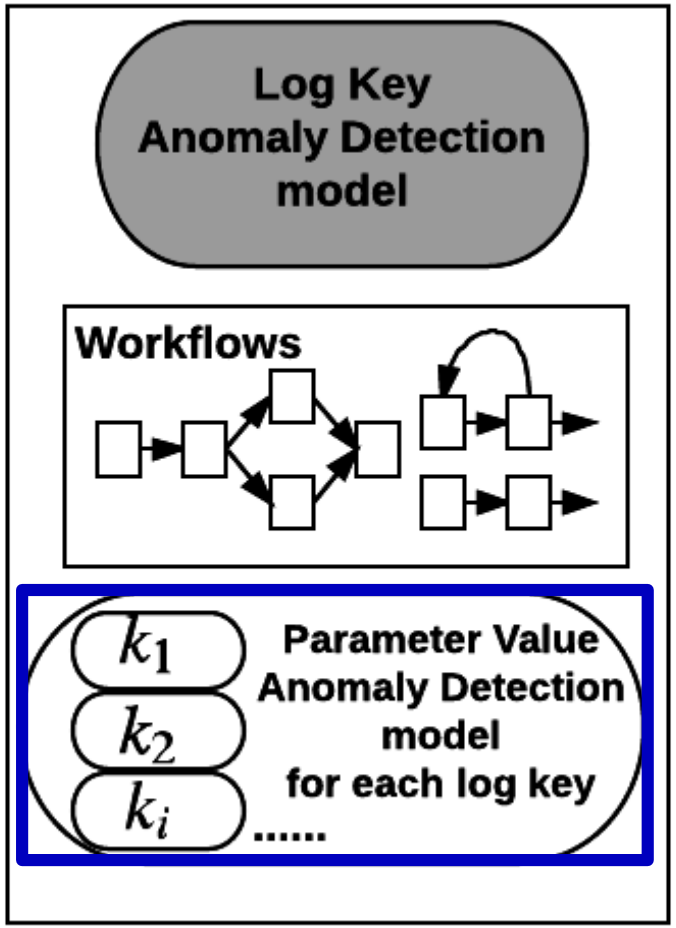


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

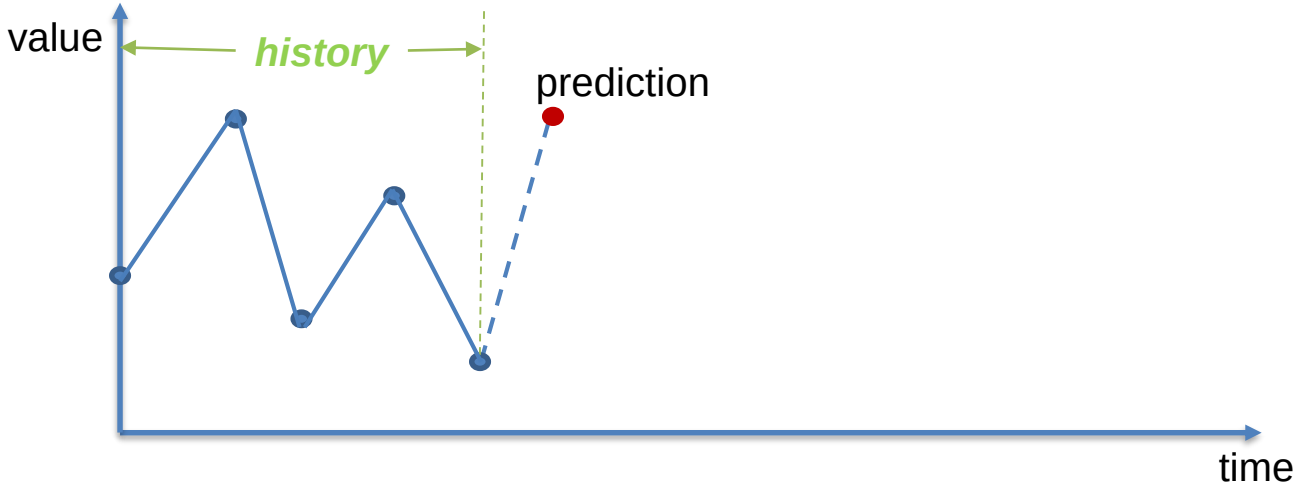


Parameter Value Anomaly Detection model

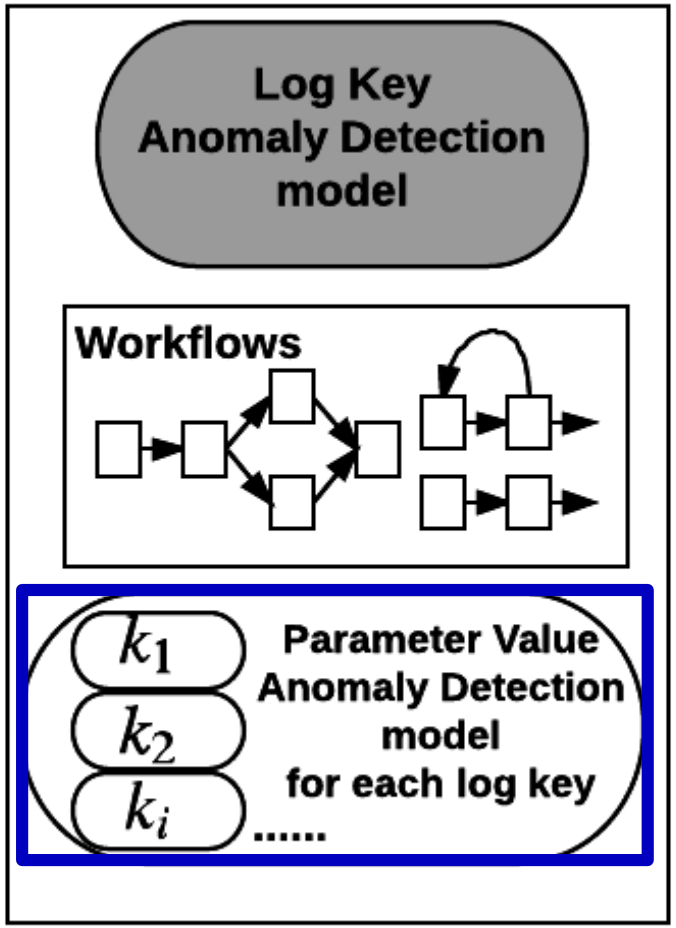


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

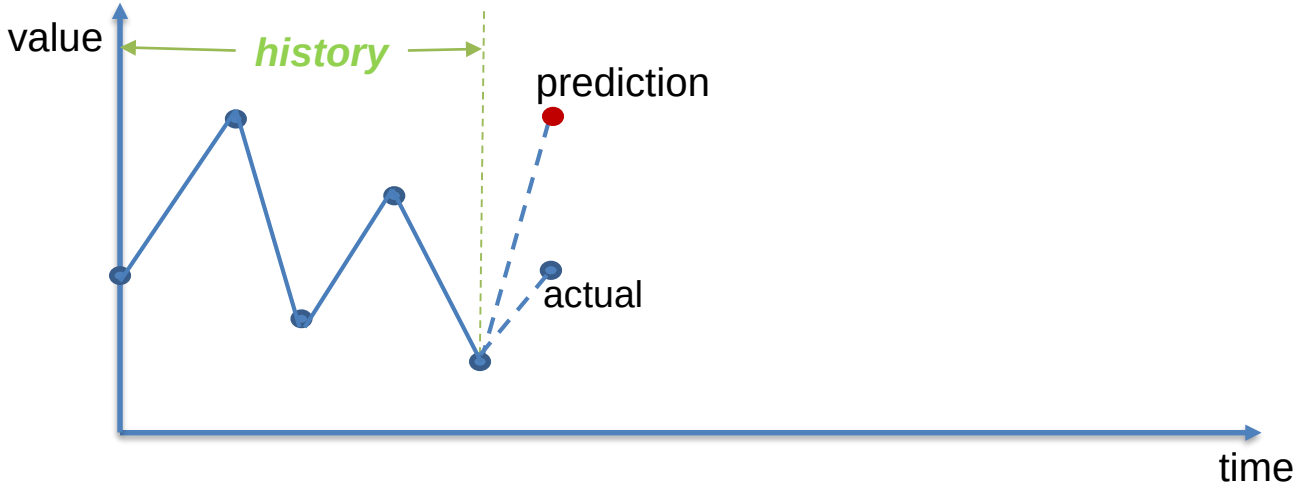


Parameter Value Anomaly Detection model

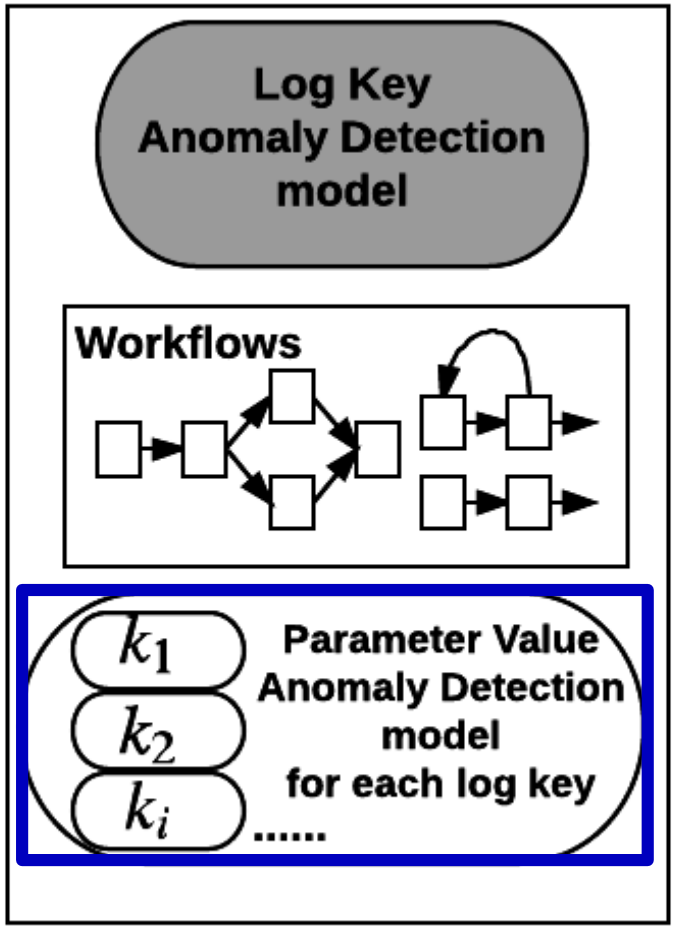


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

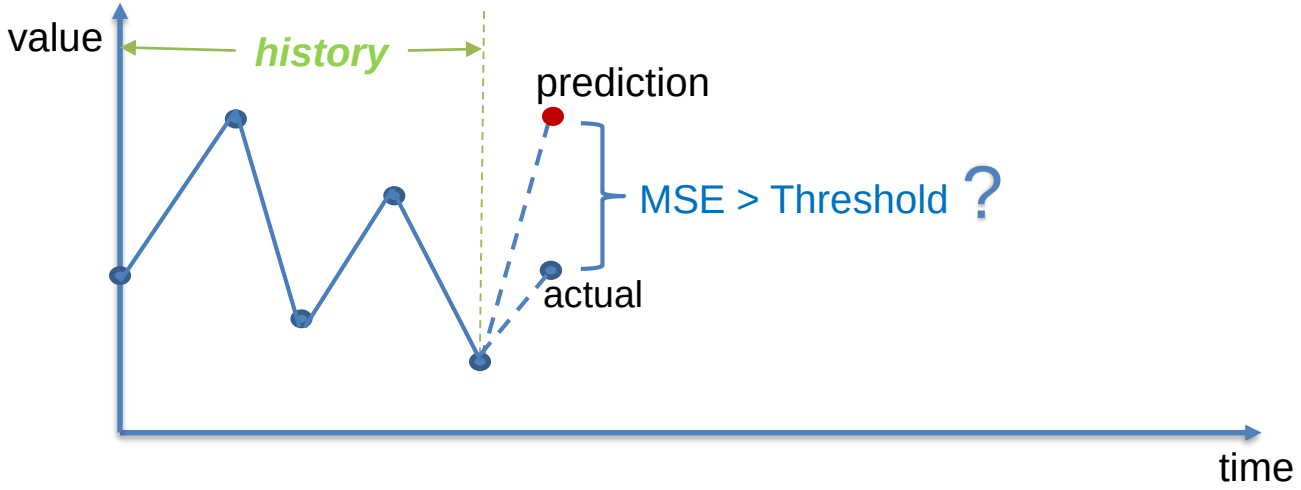


Parameter Value Anomaly Detection model

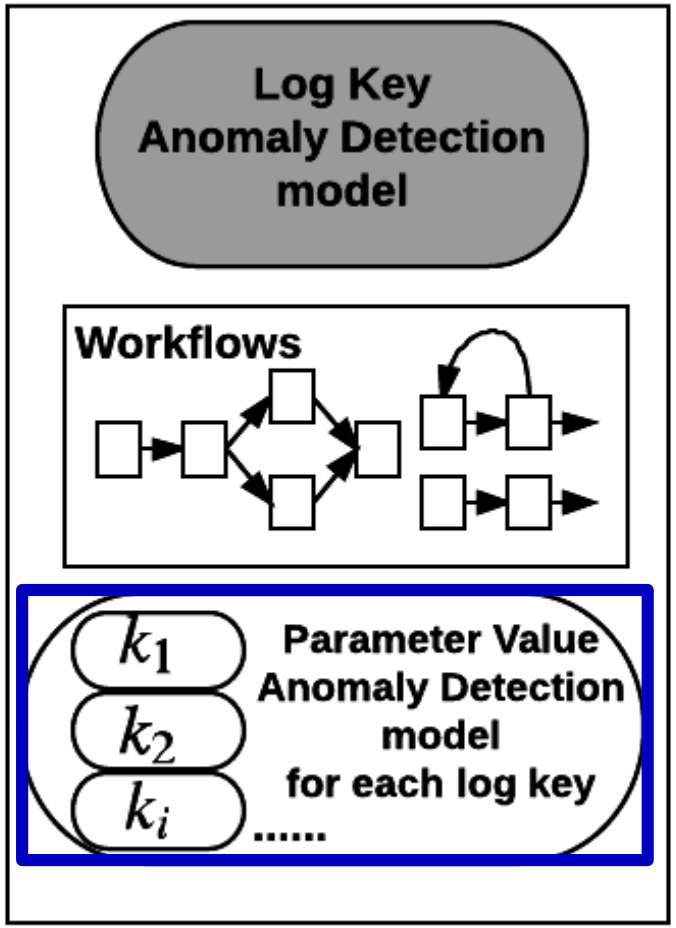


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

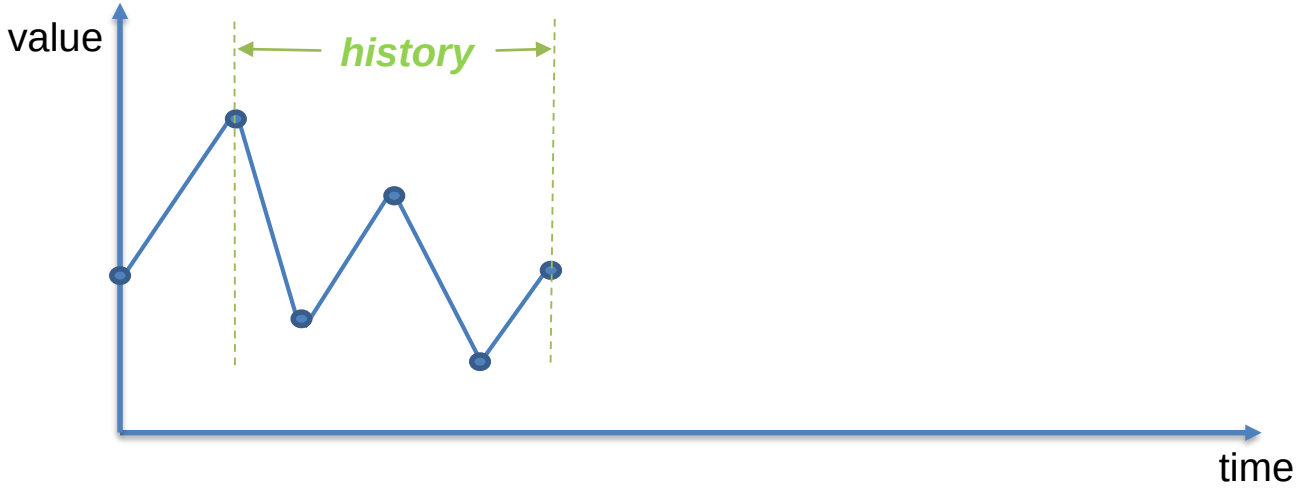


Parameter Value Anomaly Detection model

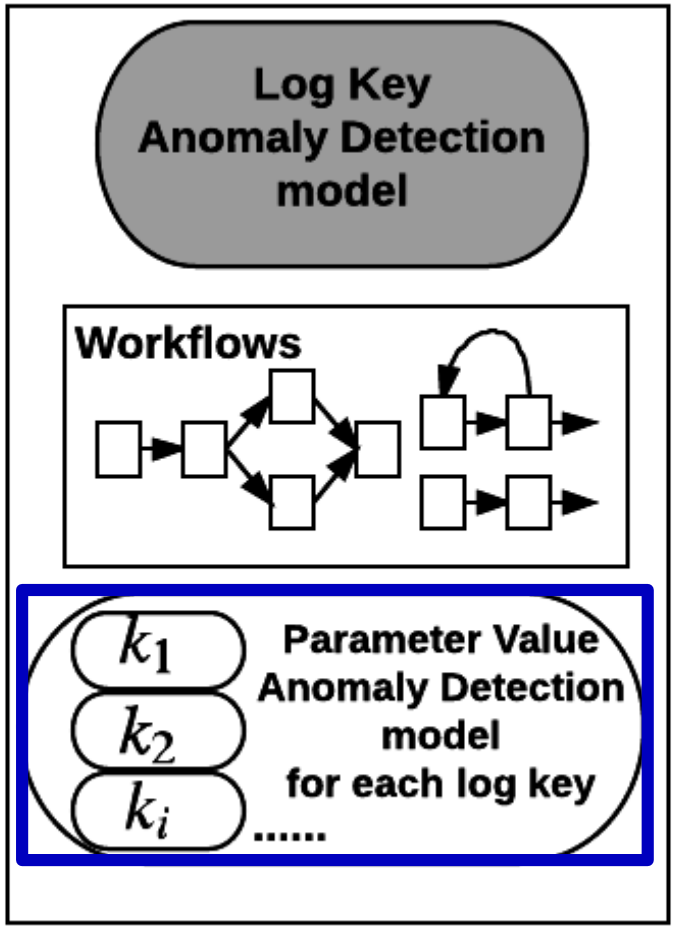


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

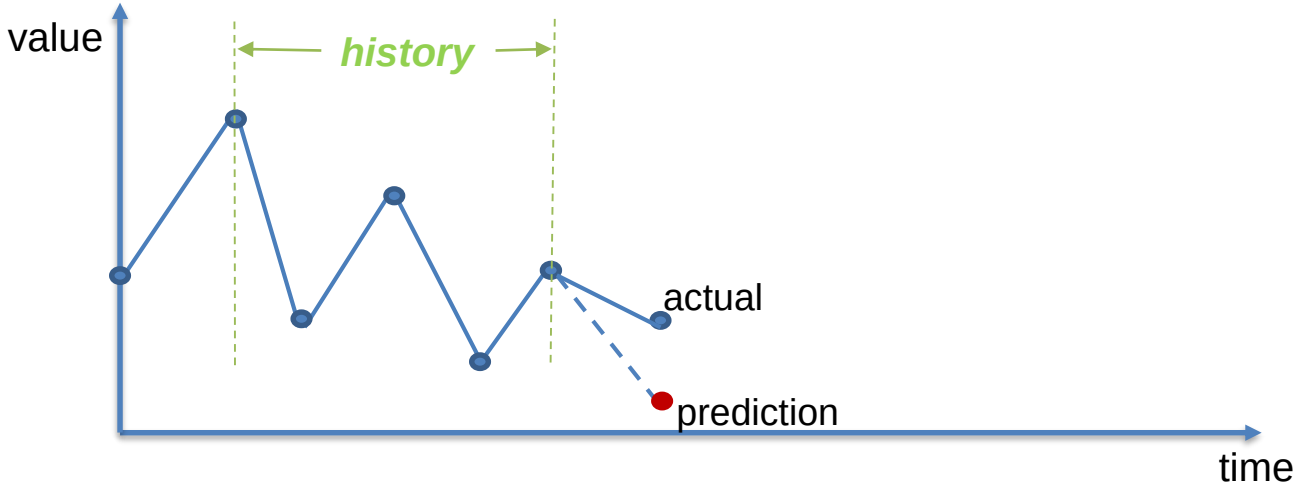


Parameter Value Anomaly Detection model

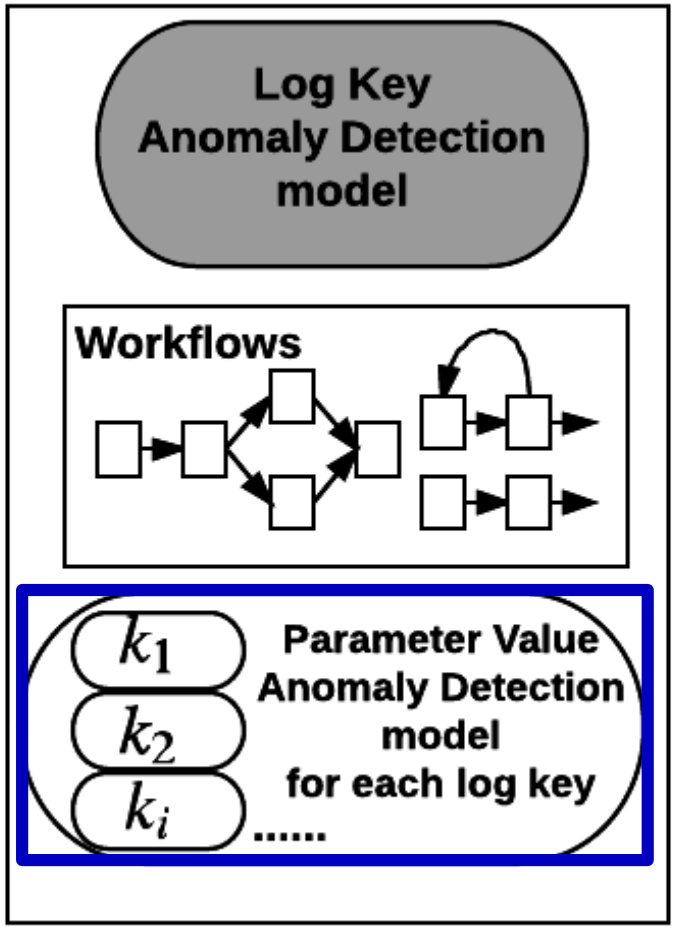


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

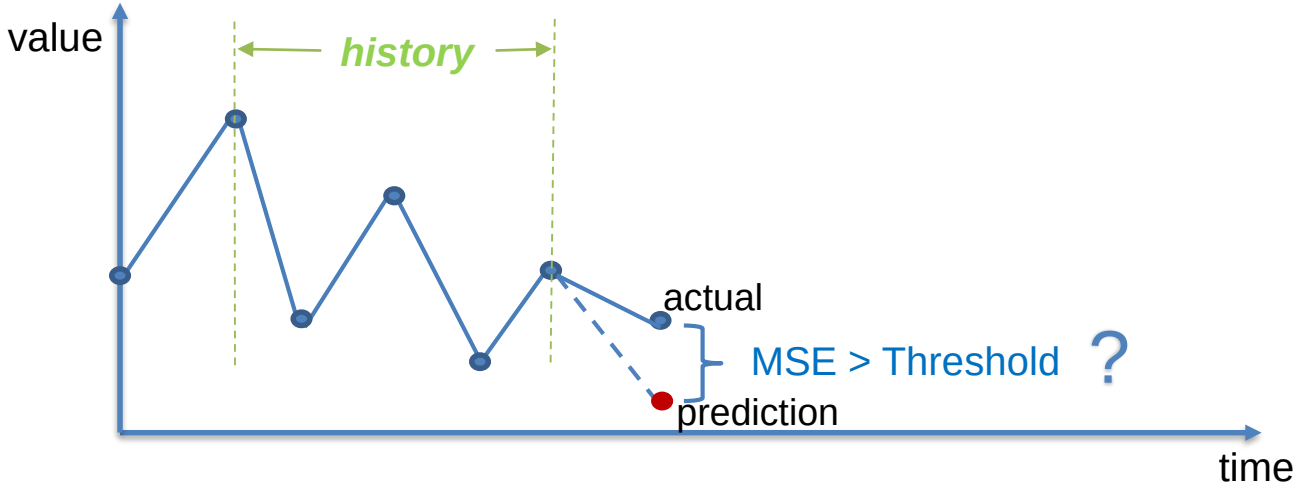


Parameter Value Anomaly Detection model

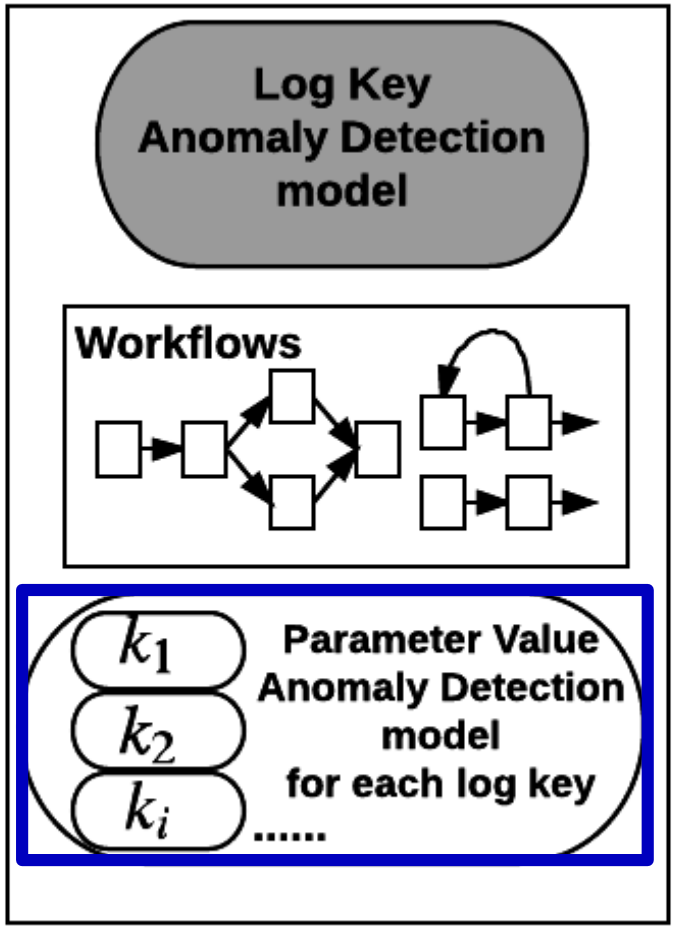


Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

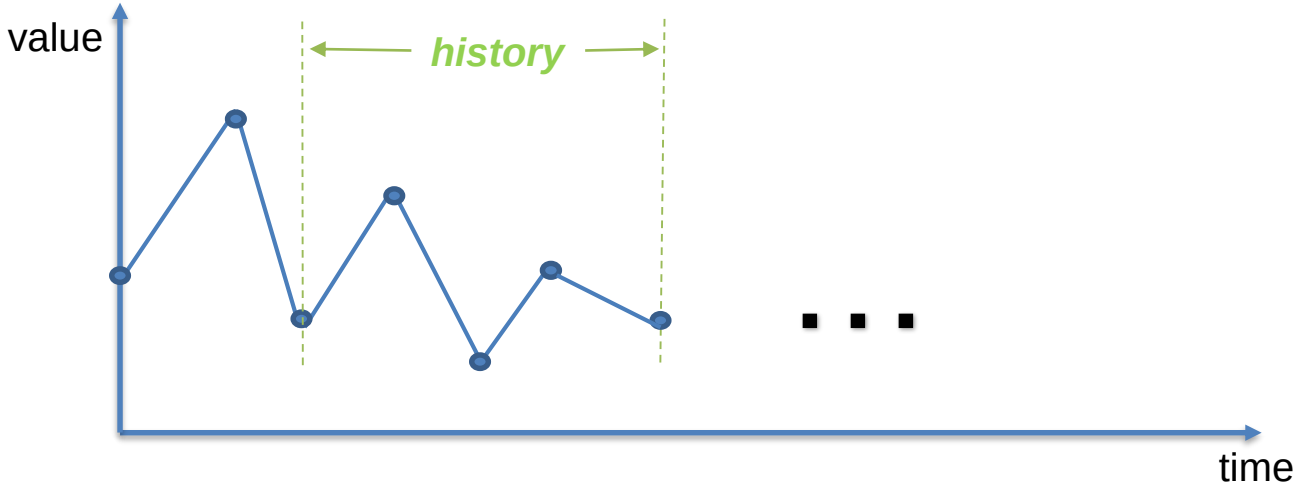


Parameter Value Anomaly Detection model



Multi-variate time series data anomaly detection problem

- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.



LSTM model online update

Q: How to handle false positive?

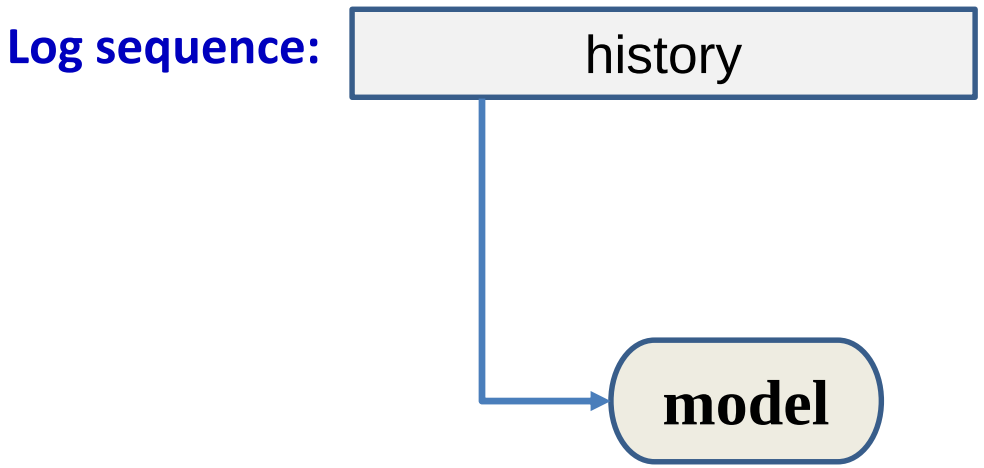
LSTM model online update

Q: How to handle false positive?

Log sequence: history

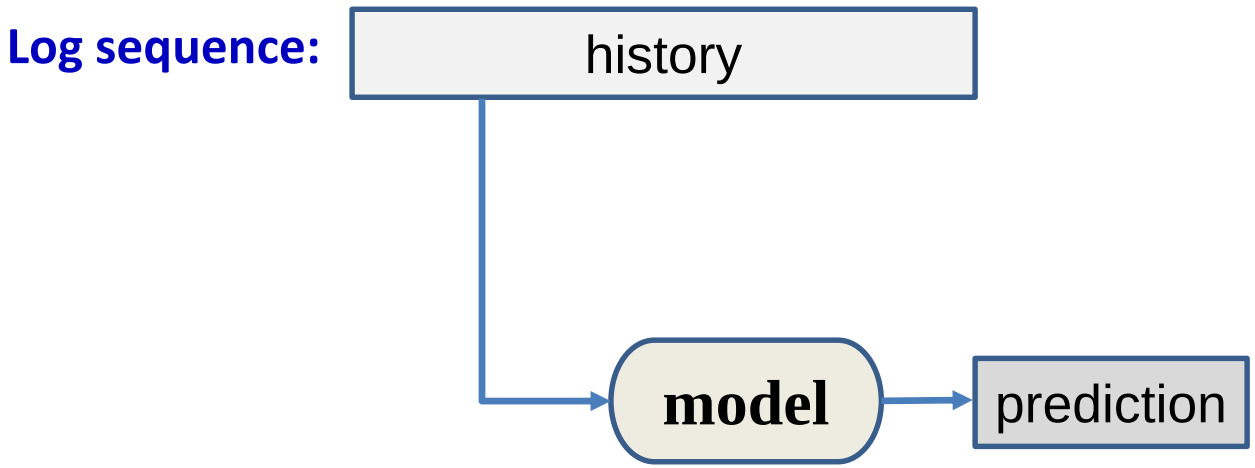
LSTM model online update

Q: How to handle false positive?



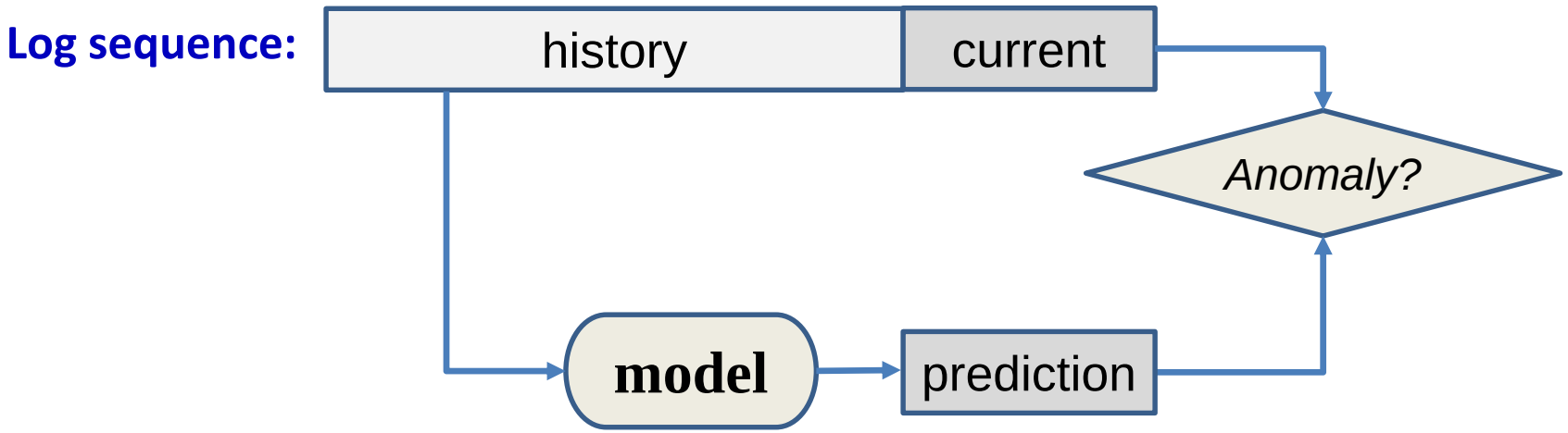
LSTM model online update

Q: How to handle false positive?



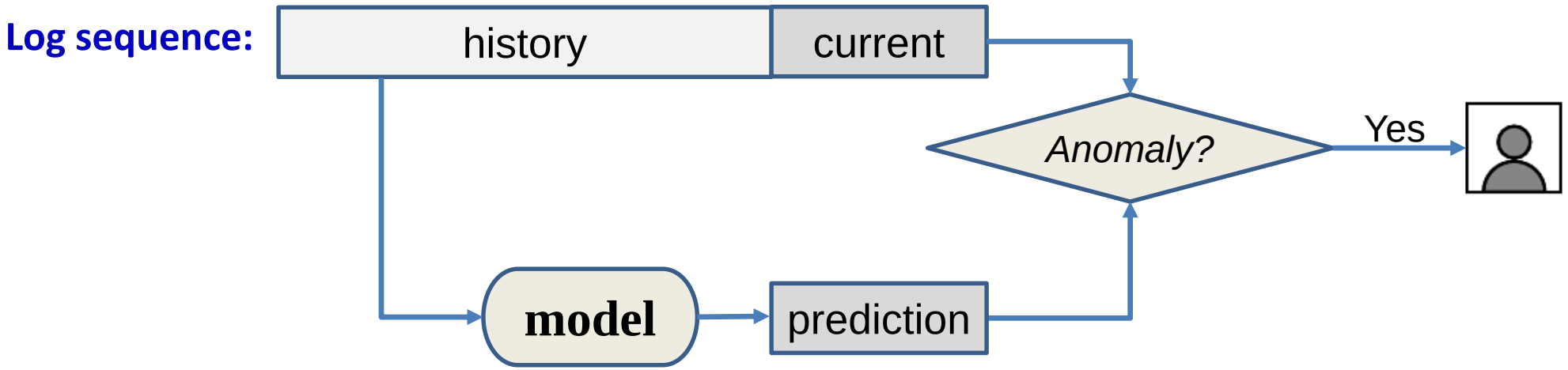
LSTM model online update

Q: How to handle false positive?



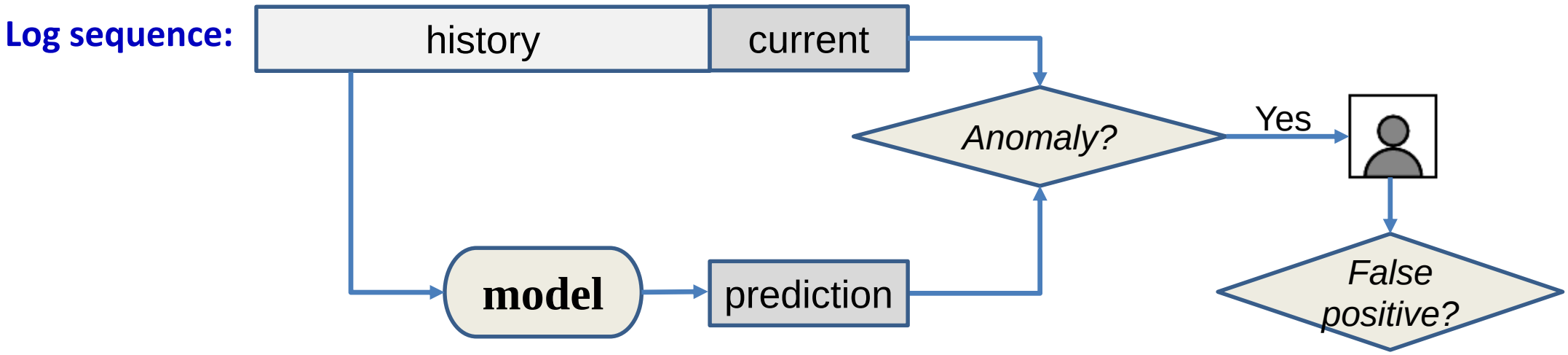
LSTM model online update

Q: How to handle false positive?



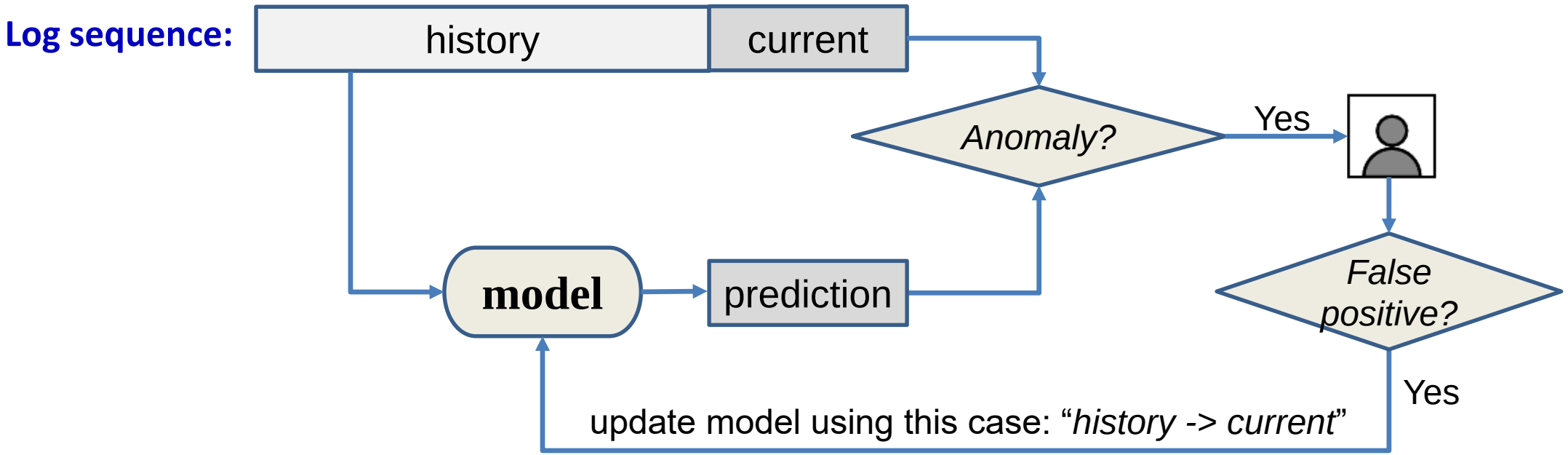
LSTM model online update

Q: How to handle false positive?



LSTM model online update

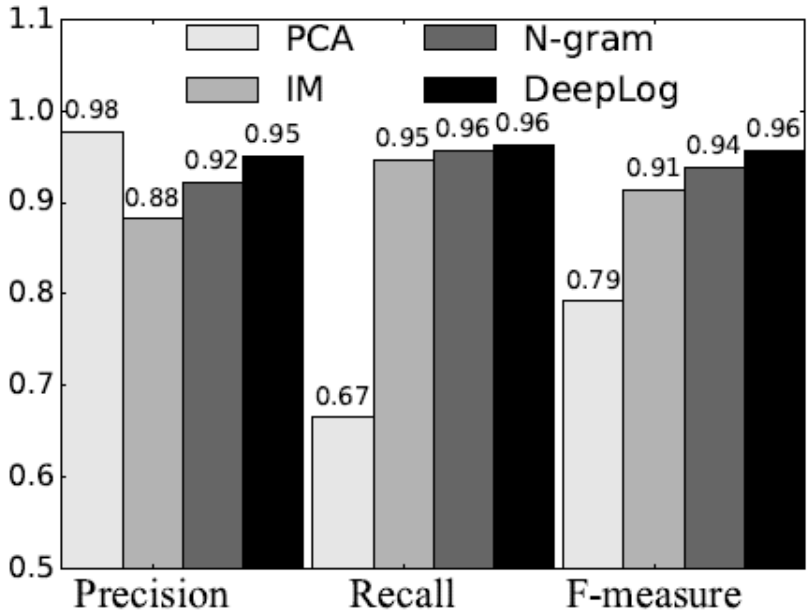
Q: How to handle false positive?



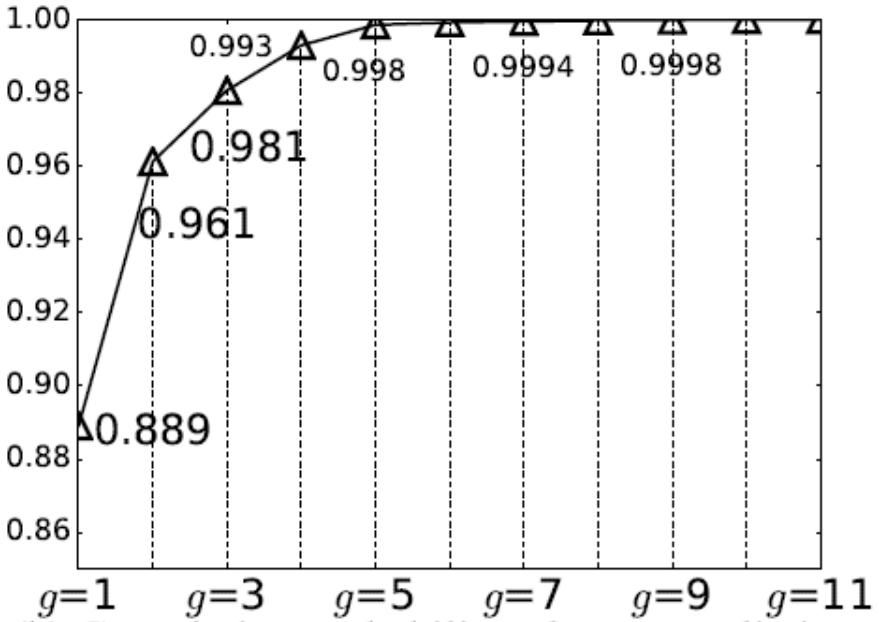
adjusts the weights to minimize the error between model output and actual observed values from the false positive cases.

Evaluation – log key anomaly detection

Up is good



(a) Accuracy on HDFS.



(b) Cumulative probability of top g predictions.

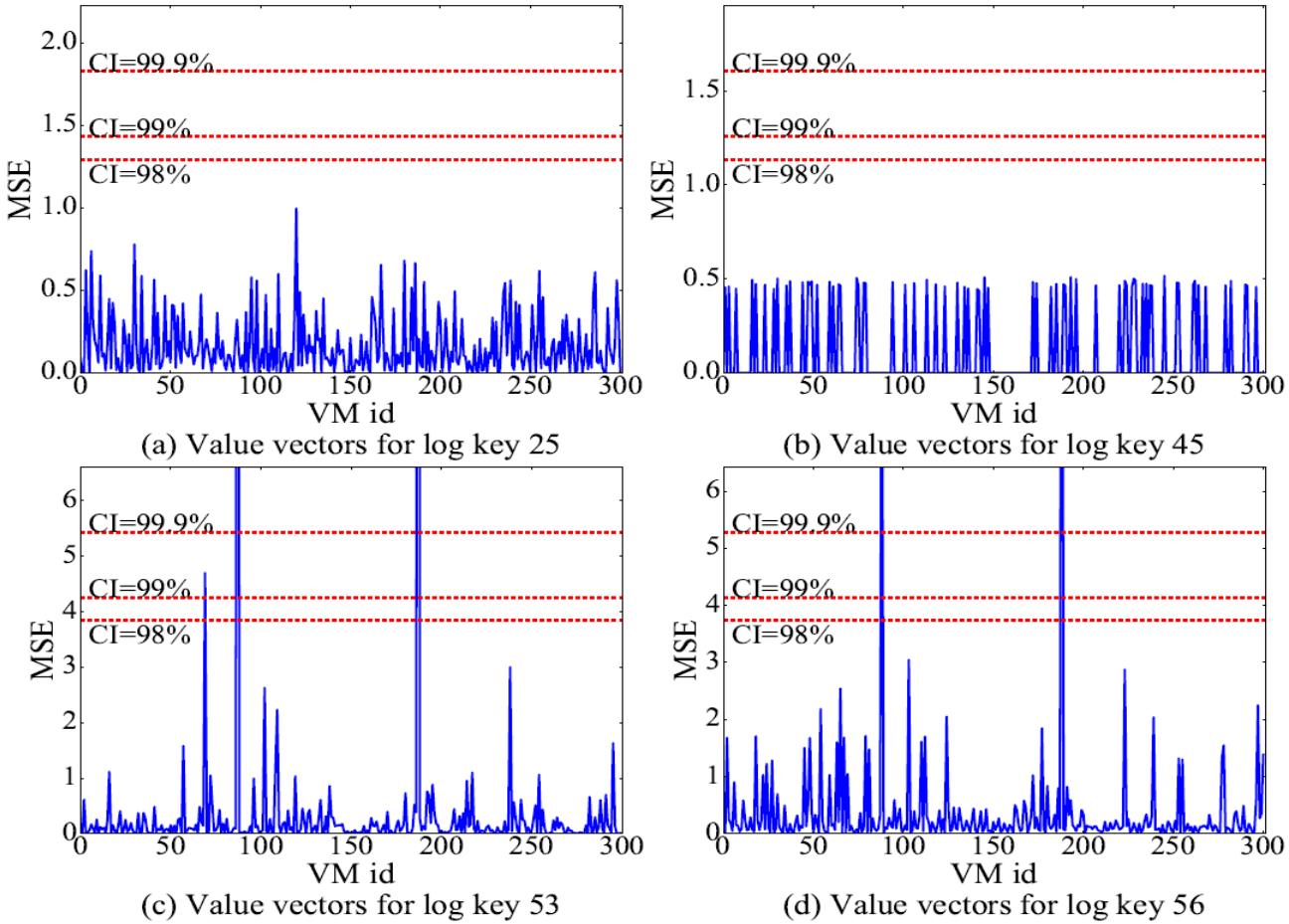
Evaluation results on HDFS log data [1].

(over a million log entries with labeled anomalies)

[1] PCA (SOSP'09), IM (UsenixATC'10), N-gram (baseline language model)

Evaluation – parameter value anomaly detection

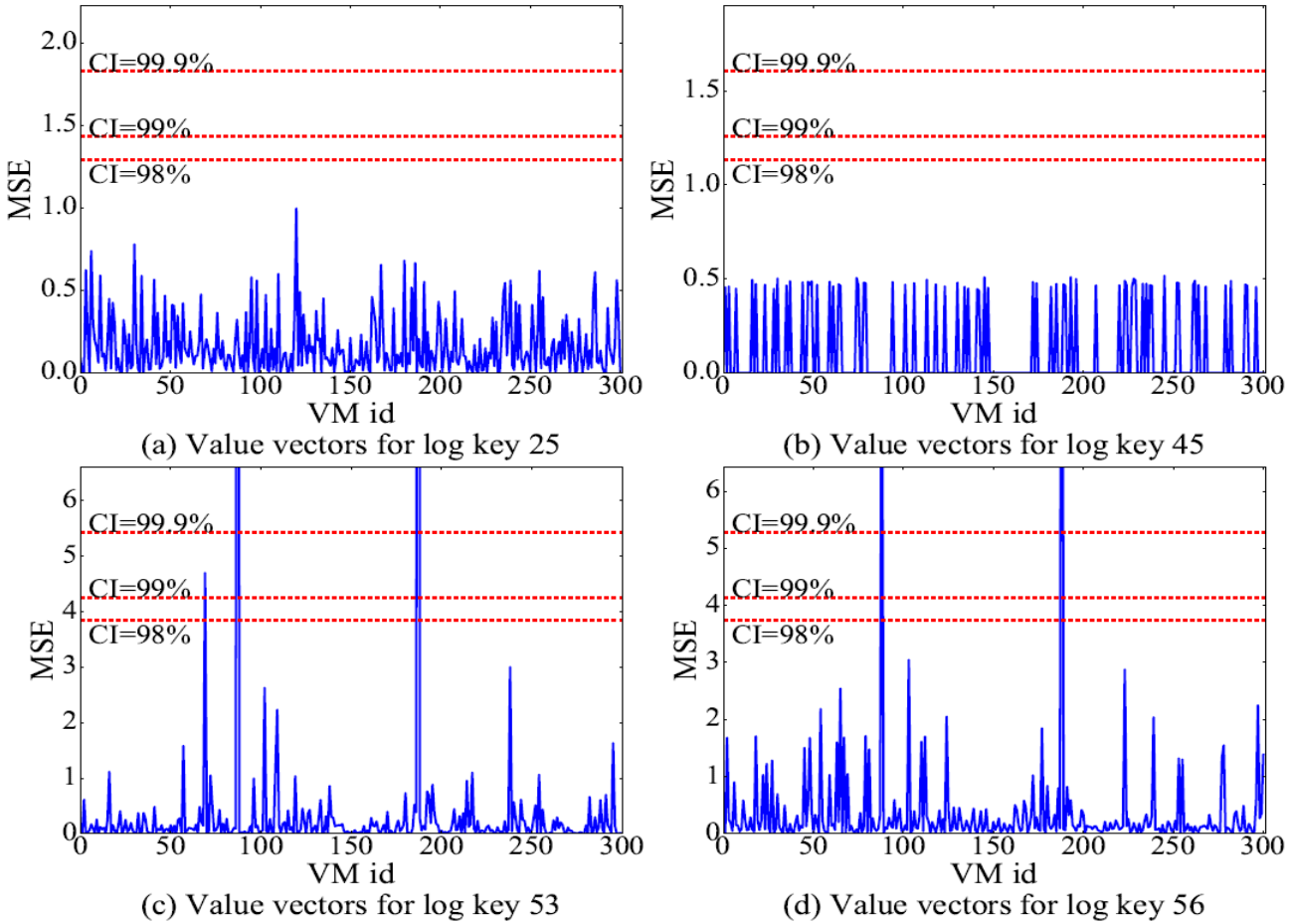
MSE:
mean square error



Evaluation results on OpenStack cloud log with different confidence intervals (CIs)

Evaluation – parameter value anomaly detection

MSE:
mean square error

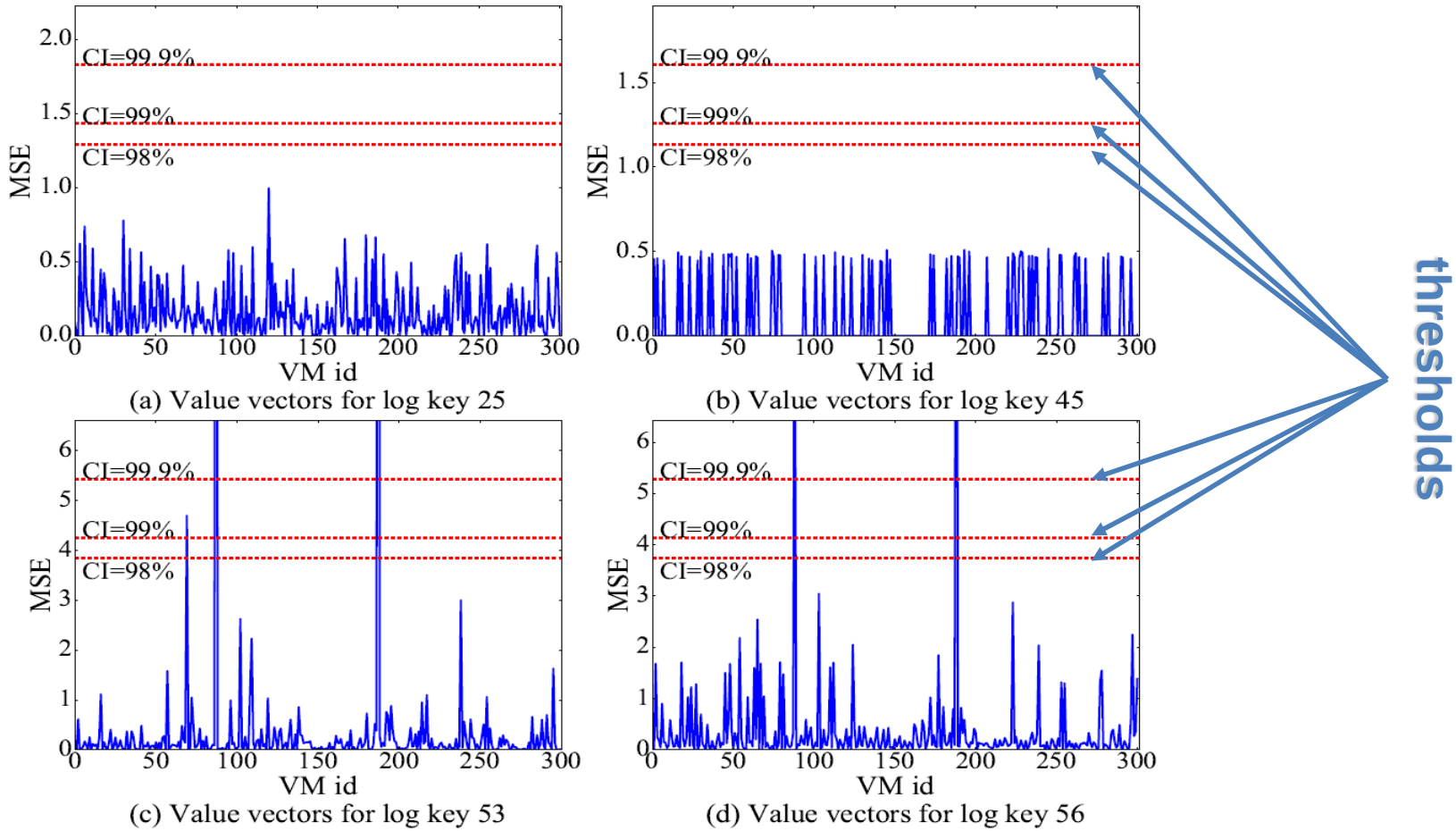


Evaluation results on OpenStack cloud log with different confidence intervals (CIs)

generated on CloudLab;
VM creation/deletion operations;
injected performance anomalies.

Evaluation – parameter value anomaly detection

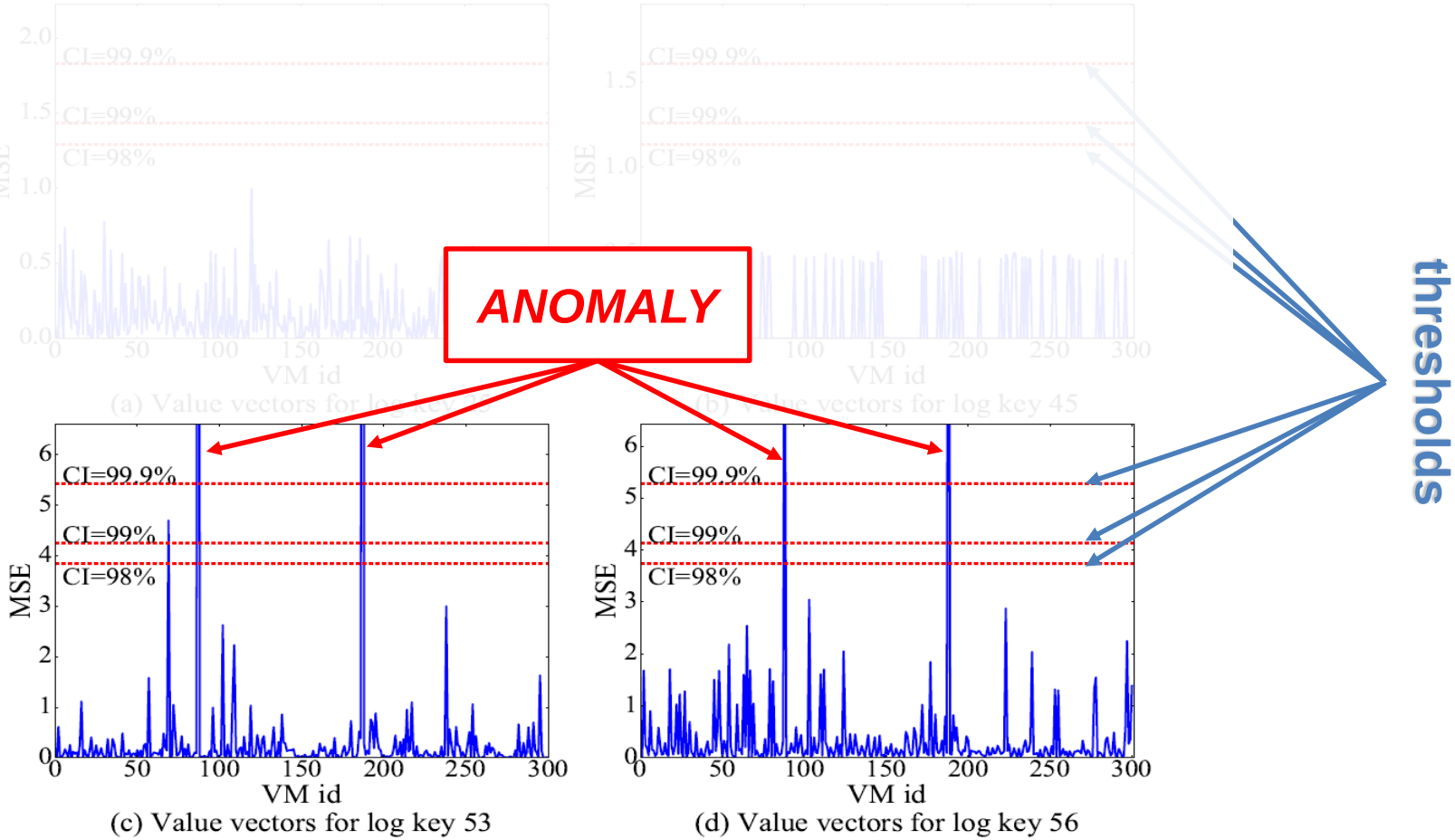
MSE:
mean square error



Evaluation results on OpenStack cloud log with different confidence intervals (CIs)

Evaluation – parameter value anomaly detection

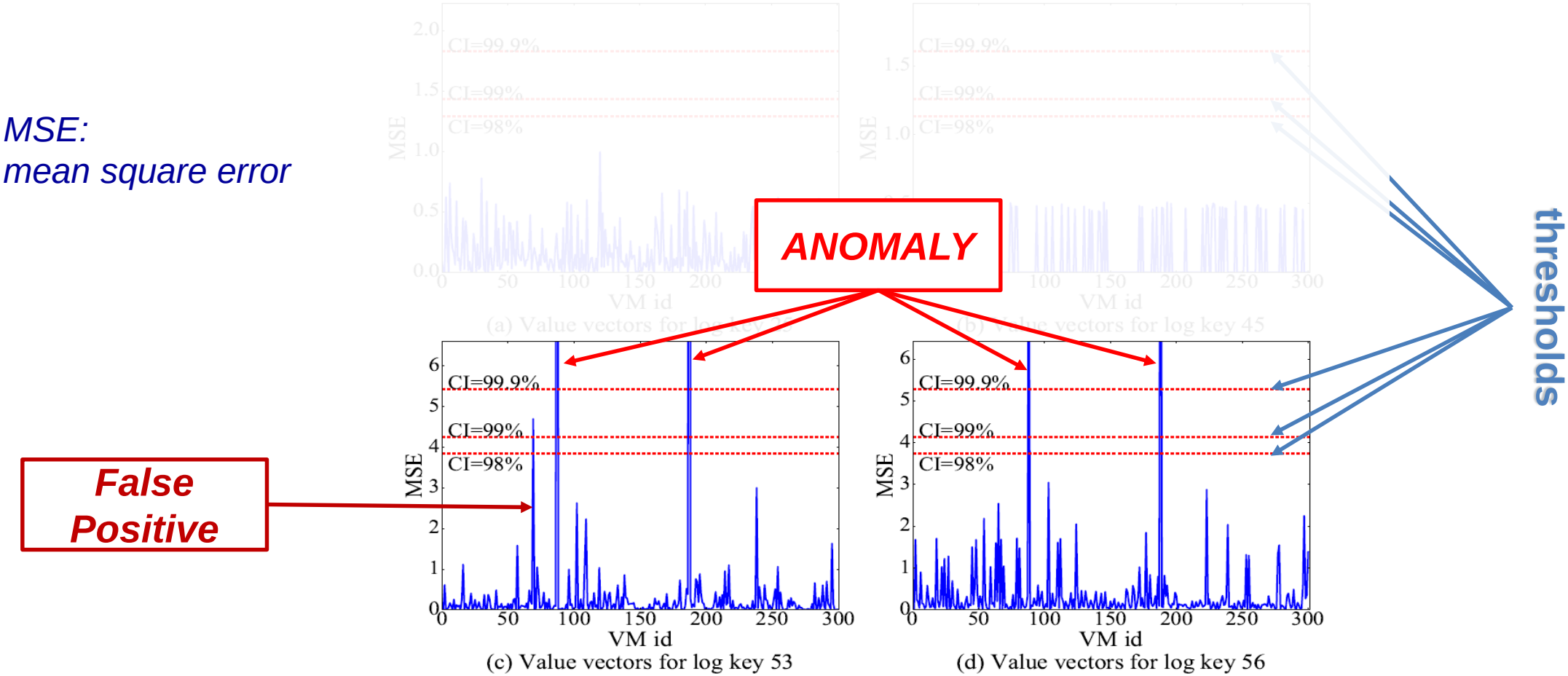
MSE:
mean square error



Evaluation results on OpenStack cloud log with different confidence intervals (CIs)

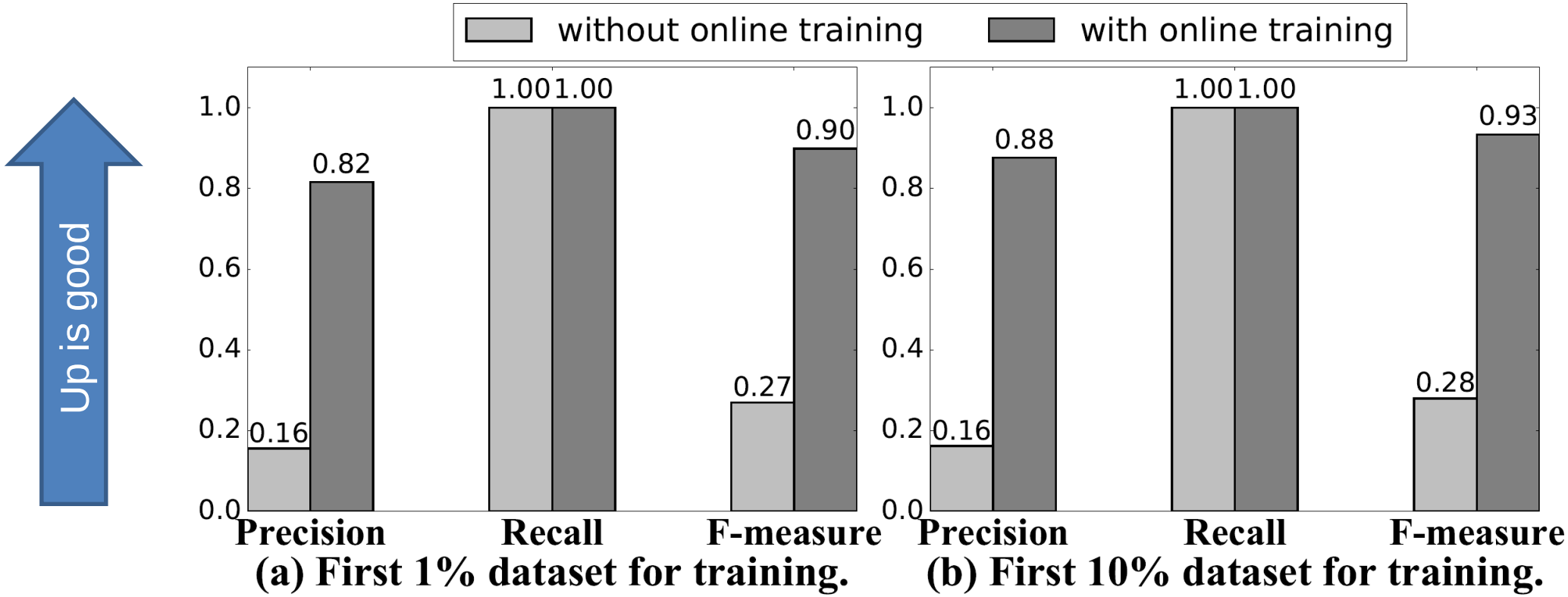
Evaluation – parameter value anomaly detection

MSE:
mean square error



Evaluation results on OpenStack cloud log with different confidence intervals (CIs)

Evaluation – LSTM model online update

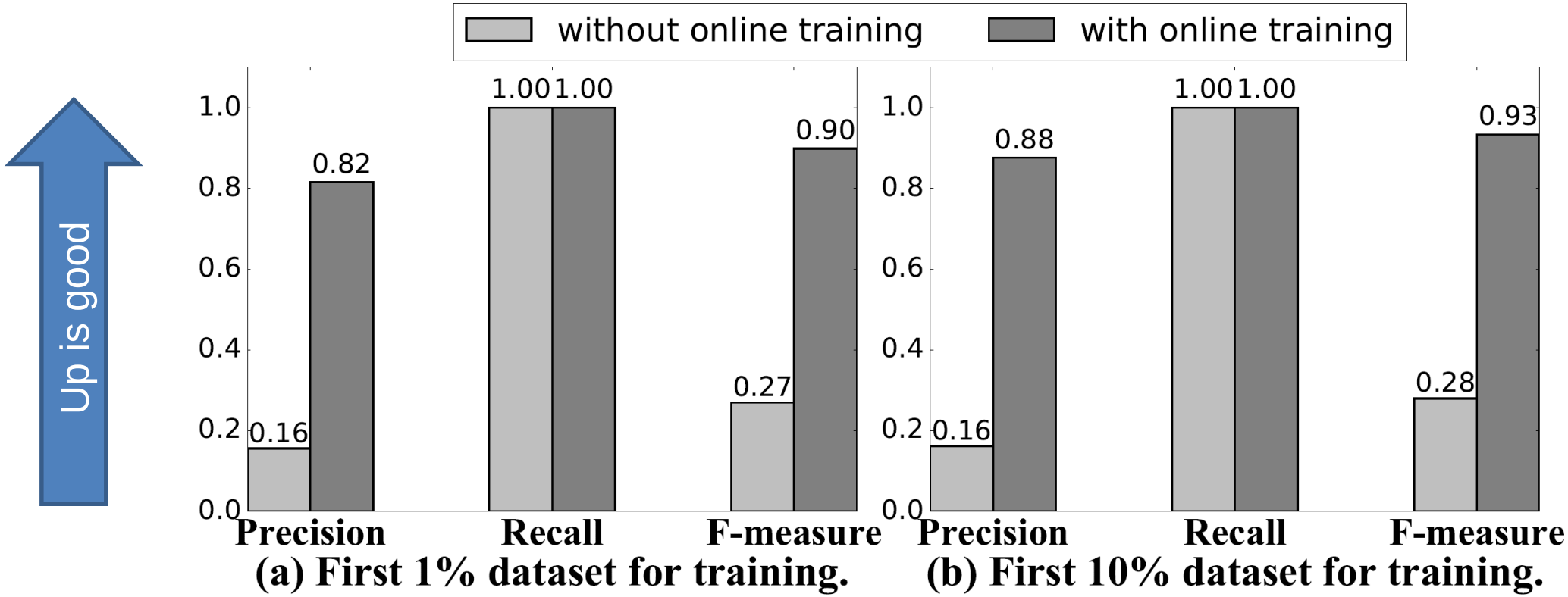


(a) First 1% dataset for training.

(b) First 10% dataset for training.

Evaluation on Blue Gene/L log,
with and without online model update.

Evaluation – LSTM model online update



(a) First 1% dataset for training.

(b) First 10% dataset for training.

Evaluation on Blue Gene/L log, with and without online model update.

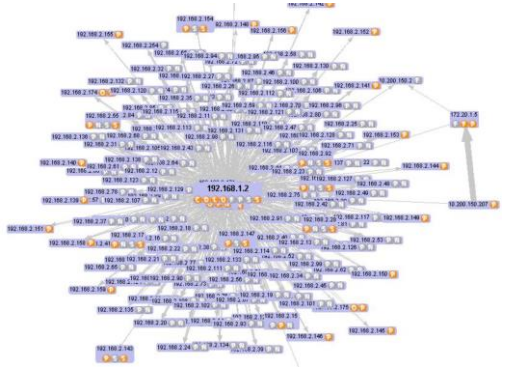
HPC log with labeled anomalies; Available at https://www.usenix.org/cfdr-data

Evaluation – case study: network security log

Dataset: IEEE VAST Challenge 2011

(Mini Challenge 2 – Computer Networking Operations)

The dataset contains firewall log, IDS log, etc.

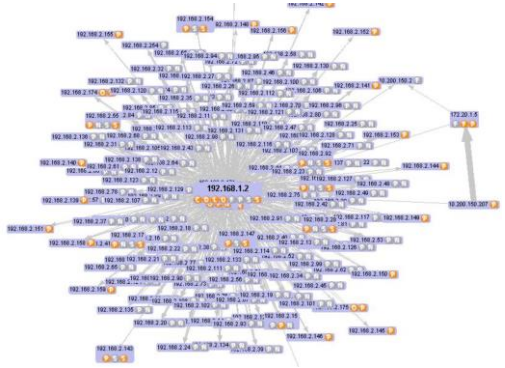


Evaluation – case study: network security log

Dataset: IEEE VAST Challenge 2011

(Mini Challenge 2 – Computer Networking Operations)

The dataset contains firewall log, IDS log, etc.



suspicious activity	detected?
Day 1: Denial of Service attack	Yes, log key anomaly in IDS log
Day 1: port scan	Yes, log key anomaly in IDS log
Day 2: port scan 1	Yes, log key anomaly in IDS log
Day 2: port scan 2	Yes, log key anomaly in IDS log
Day 2: socially engineered attack	Yes, log key anomaly in firewall log
Day 3: undocumented IP address	No

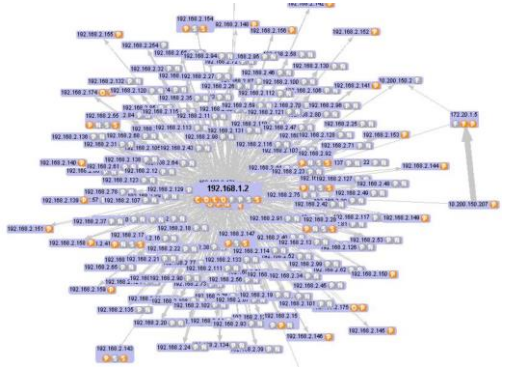
Detection results.

Evaluation – case study: network security log

Dataset: IEEE VAST Challenge 2011

(Mini Challenge 2 – Computer Networking Operations)

The dataset contains firewall log, IDS log, etc.

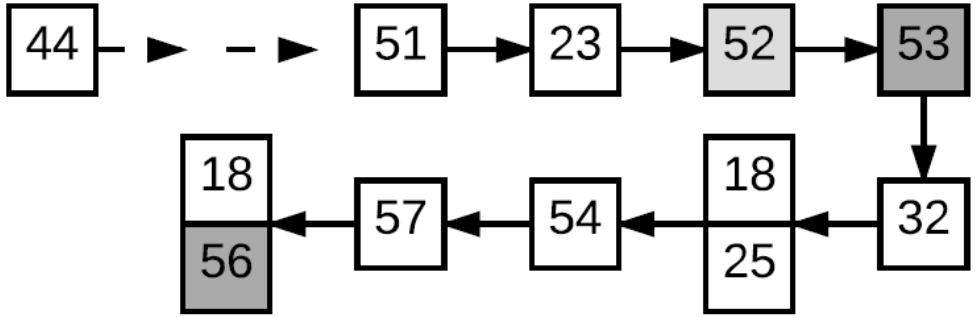


suspicious activity	detected?
Day 1: Denial of Service attack	Yes, log key anomaly in IDS log
Day 1: port scan	Yes, log key anomaly in IDS log
Day 2: port scan 1	Yes, log key anomaly in IDS log
Day 2: port scan 2	Yes, log key anomaly in IDS log
Day 2: socially engineered attack	Yes, log key anomaly in firewall log
Day 3: undocumented IP address	No

Detection results.

Could be fixed with prior knowledge of "documented IP"

Evaluation – workflow construction



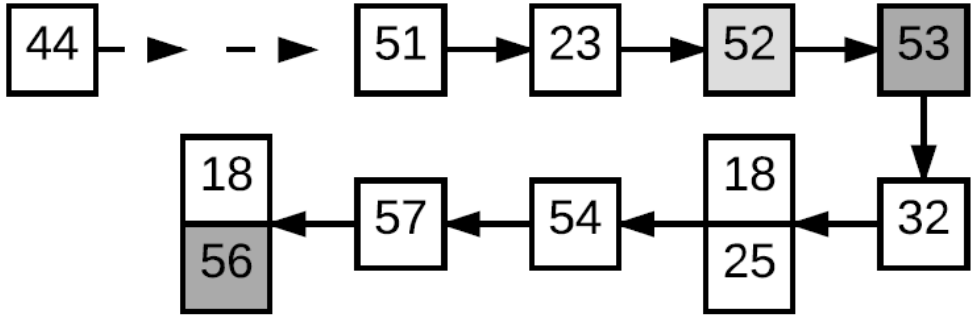
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTP/1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (Lifecycle Event)
.....
56: instance: * Took * seconds to build instance

Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?



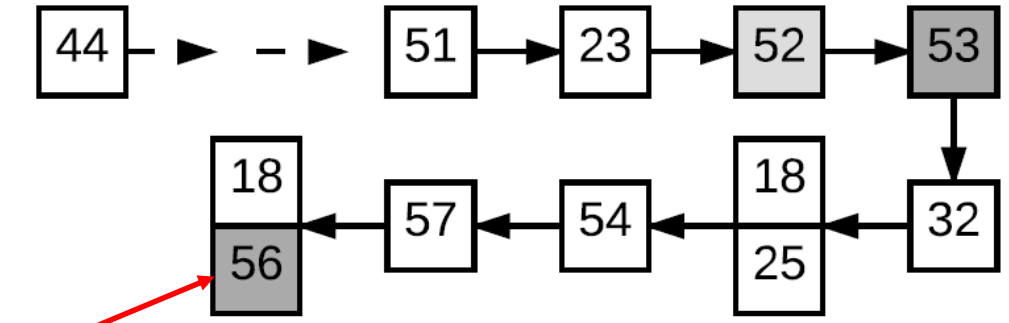
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTP/1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (Lifecycle Event)
.....
56: instance: * Took * seconds to build instance

Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?



44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTP/1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (Lifecycle Event)
.....
56: instance: * Took * seconds to build instance

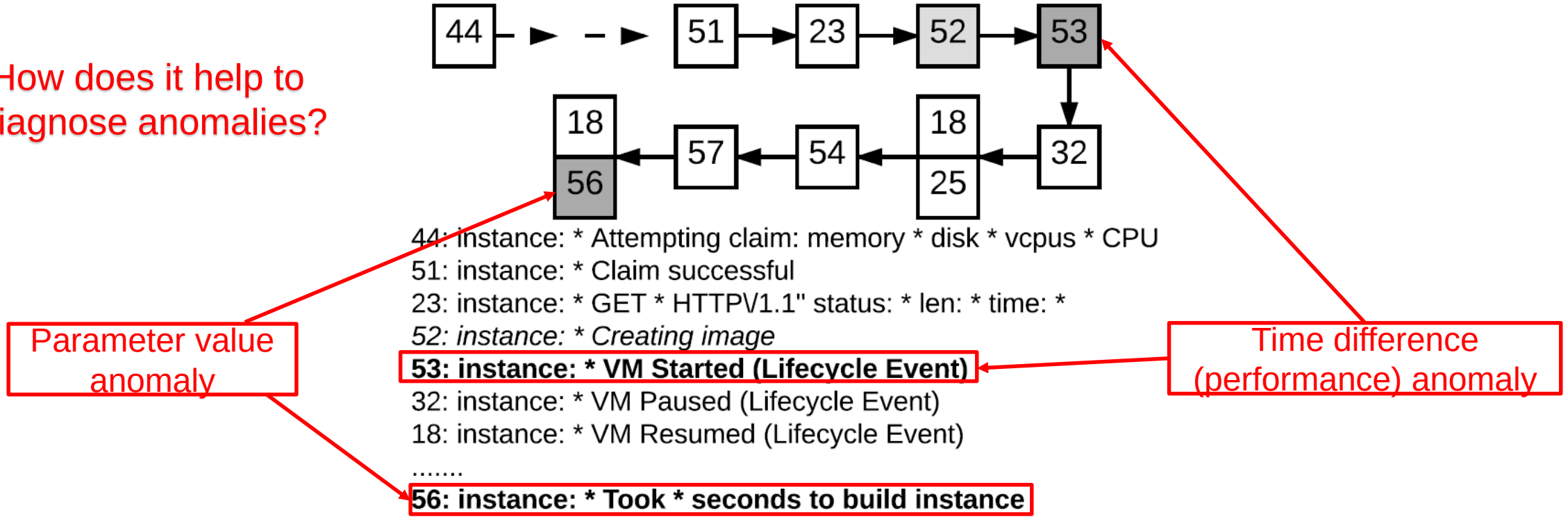
Parameter value anomaly

Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?

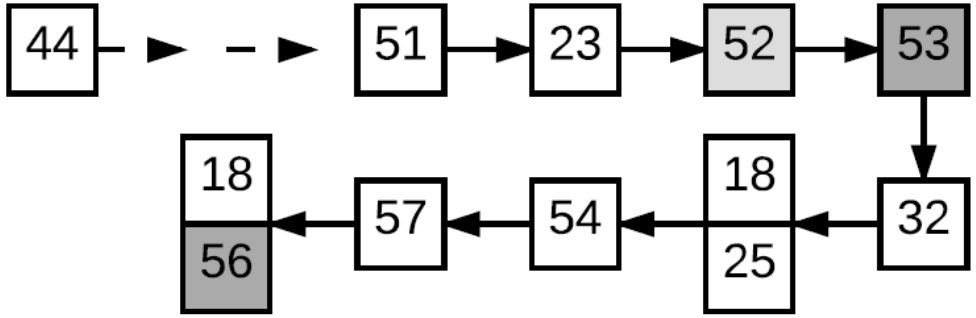


Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?



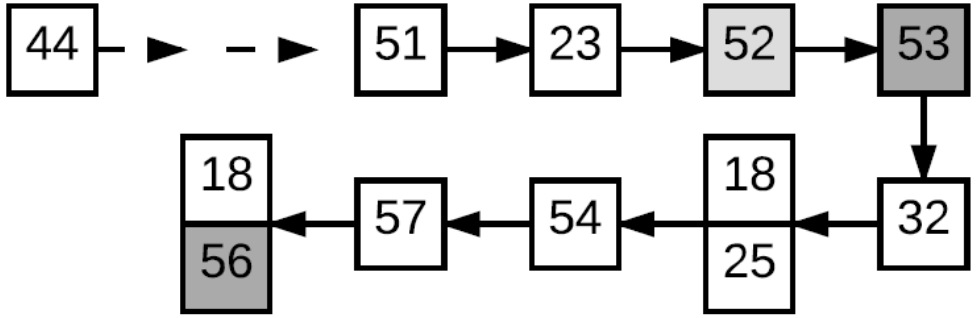
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTP/1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (Lifecycle Event)
.....
56: instance: * Took * seconds to build instance

Identified anomaly:
Instance took too long to build because of the transition from 52 -> 53

Constructed workflow of *VM Creation*.
(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?



44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTP/1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (Lifecycle Event)
.....
56: instance: * Took * seconds to build instance

Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Identified anomaly:

Instance took too long to build because of the transition from 52 -> 53

Injected anomaly:

During VM creation, network speed from controller to compute node is throttled.

Summary

DeepLog

- A realtime system log anomaly detection framework.
- LSTM is used to model system execution paths and log parameter values.
- Workflow models are built to help anomaly diagnosis.
- It supports online model update.

Thank you!

Min Du
mind@cs.utah.edu

Feifei Li
lifeifei@cs.utah.edu