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

Min Du, Feifei Li, Guineng Zheng, Vivek Srikumar
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_-320848348280074l142 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

```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0001_m_001575_0/part-01575. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$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

```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0001_m_001657_0/part-01657. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$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
......
```

System Event Log

Available practically on every computer system!

Background

System Event Log

Available practically on every computer system!

```
081111 083419 24621 INFO dfs.DataNode$DataXceiver: Receiving block blk_5214640714119373081 src:
/10.251.121.224:47915 dest: /10.251.121.224:50010
081111 083419 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock:
/user/root/rand7/_temporary/_task_200811101024_0002_m_001657_0/part-01657. blk_5214640714119373081
081111 083420 24633 INFO dfs.DataNode$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_0016_m_001470_0/part-01470. blk_-
3208483482800741142
081111 104233 26437 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-
3208483482800741142 terminating
......
```

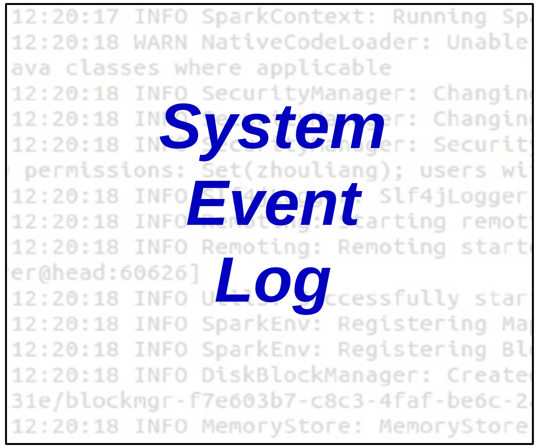
Automatic Analysis?

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$DataXceiver: Receiving block blk_-3208483482800741142 src:
/10.251.203.80:46[illegible]
081111 104136 27196 INFO dfs.DataNode$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

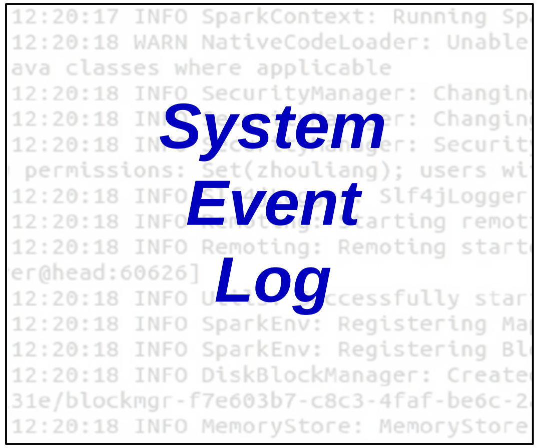


System
Event
Log

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 IN                  er: Changin
12:20:18 IN                  er: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO                Slf4jLogger
12:20:18 INFO                tarting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO U              ccessfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

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

Background



```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

Message type
Log key
……

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

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

Background

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

Message type
Log key
……

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

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

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 IN[illegible]er: Changin
12:20:18 IN[illegible]er: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO[illegible] Slf4jLogger
12:20:18 INFO[illegible]tarting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO U[illegible]ccessfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

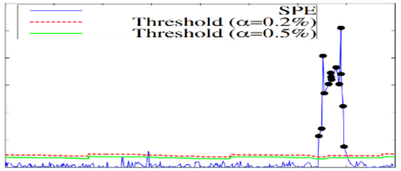
Message type

Log key

……

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

Anomaly Detection



LOG ANALYSIS

Background

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

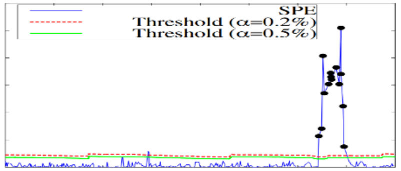
Message type

Log key

……

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

Anomaly Detection



LOG ANALYSIS

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

Background

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 IN                er: Changin
12:20:18 IN                er: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO            : Slf4jLogger
12:20:18 INFO            : Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO U            ccessfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

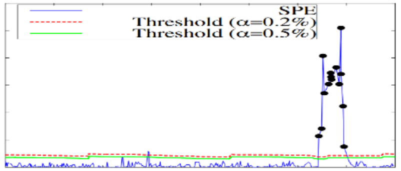
Message type

Log key

……

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

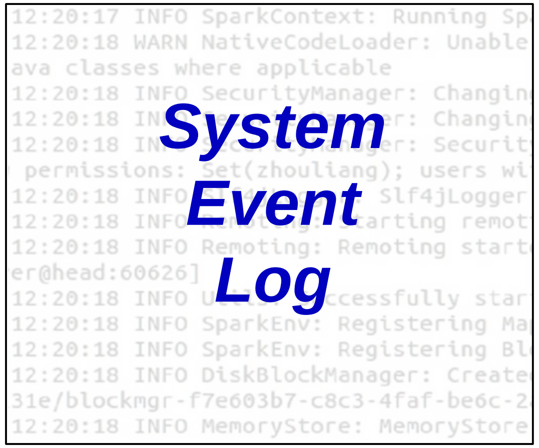
Anomaly Detection



LOG ANALYSIS

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

Background



```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

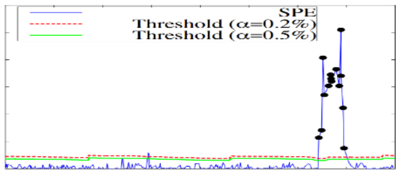
Message type

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

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

LOG PARSING

Structured Data

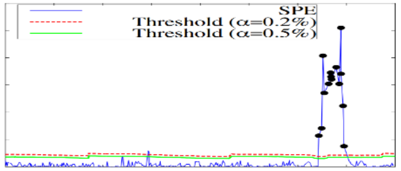
Message type

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

```
12:20:17 INFO SparkContext: Running Sp
12:20:18 WARN NativeCodeLoader: Unable
ava classes where applicable
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Changin
12:20:18 INFO SecurityManager: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO Slf4jLogger: Slf4jLogger
12:20:18 INFO Remoting: Starting remot
12:20:18 INFO Remoting: Remoting start
er@head:60626]
12:20:18 INFO Utils: Successfully star
12:20:18 INFO SparkEnv: Registering Ma
12:20:18 INFO SparkEnv: Registering Bl
12:20:18 INFO DiskBlockManager: Create
31e/blockmgr-f7e603b7-c8c3-4faf-be6c-2
12:20:18 INFO MemoryStore: MemoryStore
```

System Event Log

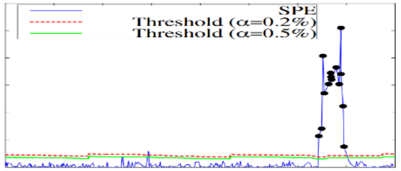
LOG PARSING

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

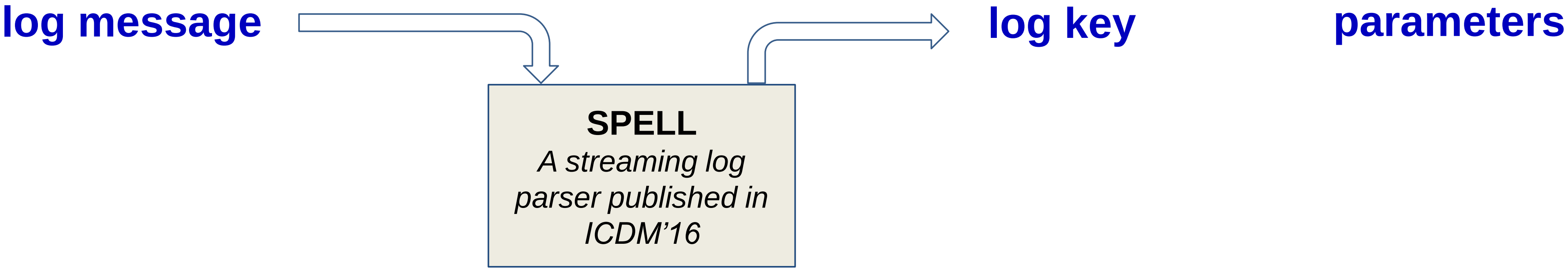
log message (log key underlined)	log key	parameter value vector
t_1 Deletion of file1 complete	k_1	[t_1 - t_0, file1]
t_2 Took 0.61 seconds to deallocate network …	k_2	[t_2 - t_1, 0.61]
t_3 VM Stopped (Lifecycle Event)	k_3	[t_3 - t_2]
…	…	…

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]
…	…	…



log message

SPELL

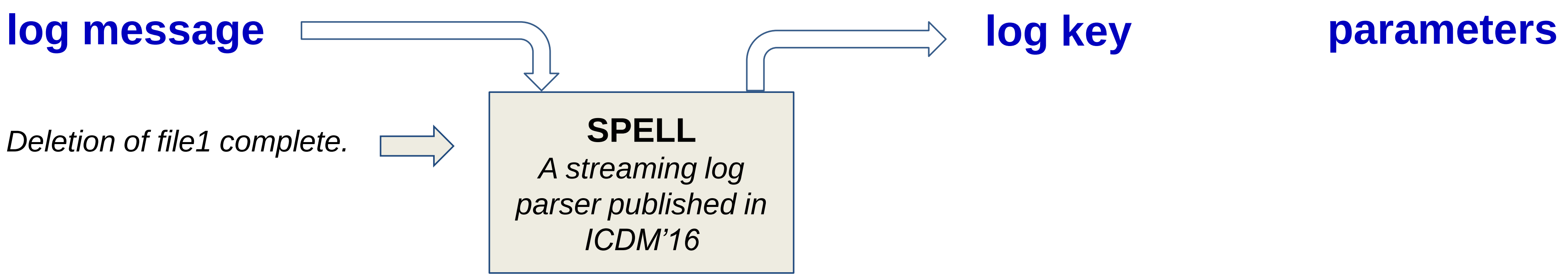
A streaming log parser published in ICDM'16

log key

parameters

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]
…	…	…



log message

log key

parameters

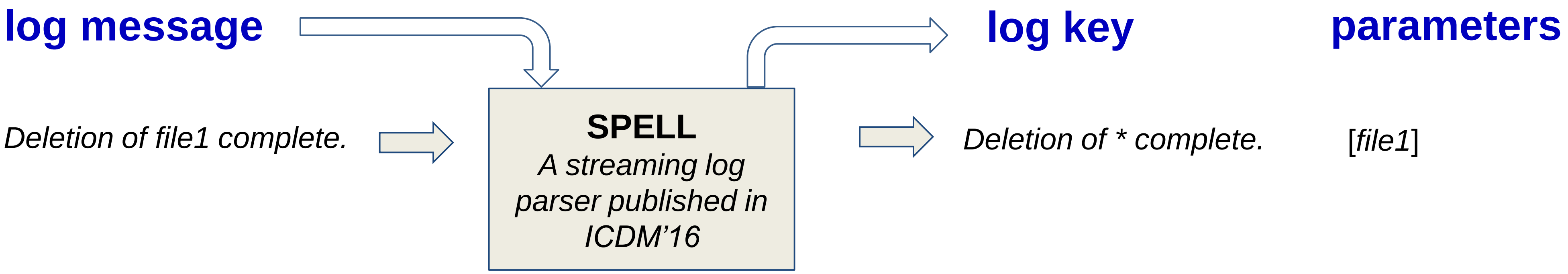
Deletion of file1 complete.

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]
…	…	…



log message

log key

parameters

Deletion of file1 complete.

SPELL

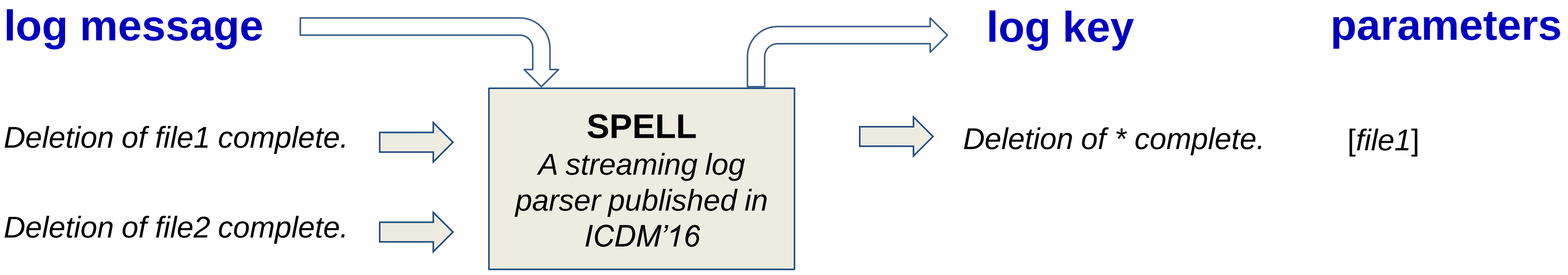
A streaming log parser published in ICDM’16

*Deletion of * complete.*

[*file1*]

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]
…	…	…



log message

Deletion of file1 complete.

Deletion of file2 complete.

SPELL

A streaming log parser published in ICDM'16

log key

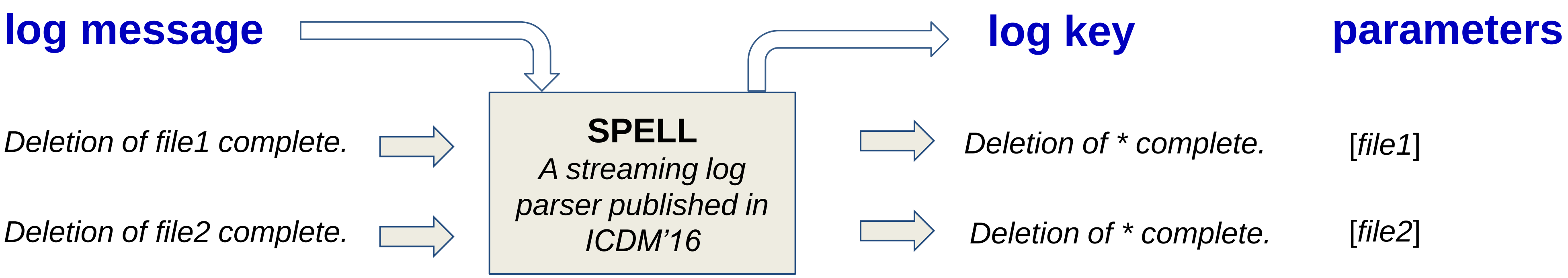
*Deletion of * complete.*

parameters

[*file1*]

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]
…	…	…



log message

log key

parameters

Deletion of file1 complete.

Deletion of file2 complete.

SPELL

A streaming log parser published in ICDM'16

Deletion of * complete.

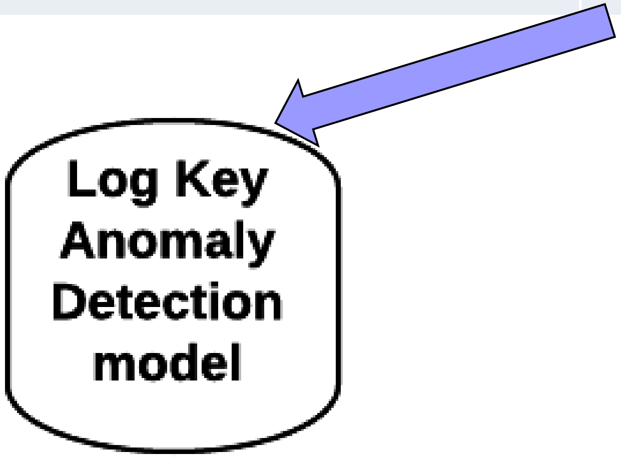
Deletion of * complete.

[file1]

[file2]

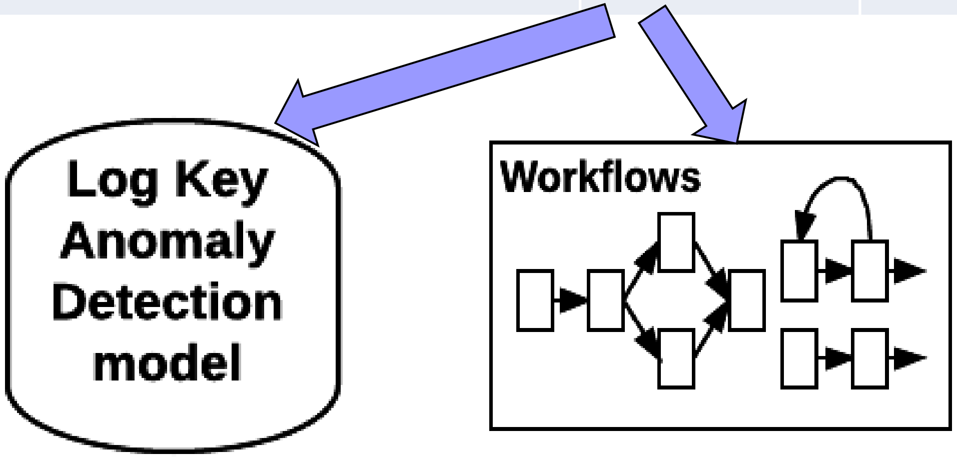
DeepLog

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]
…	…	…



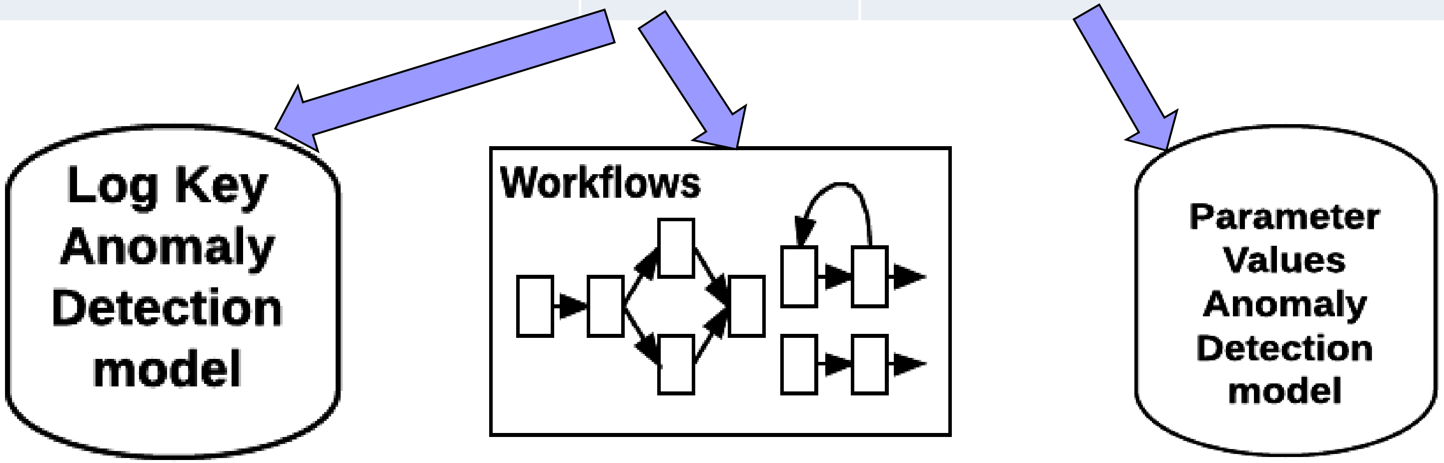
DeepLog

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]
…	…	…



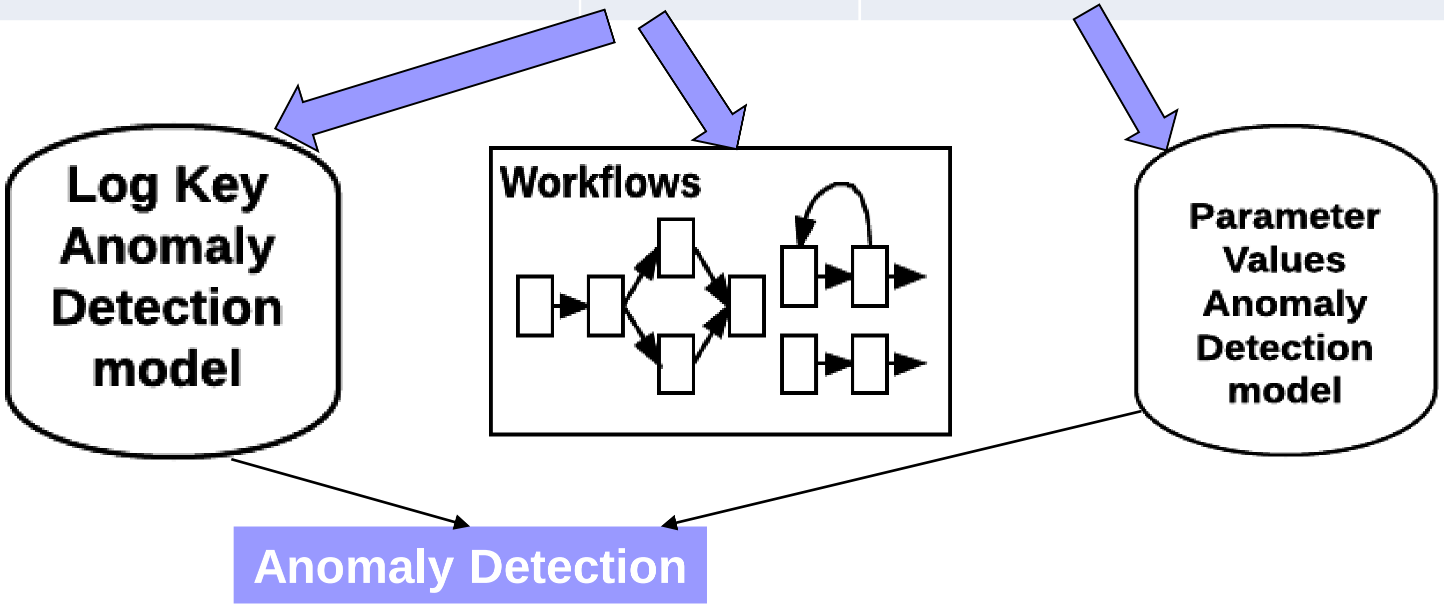
DeepLog

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]
…	…	…


Log Key Anomaly Detection model
Workflows
Parameter Values Anomaly Detection model

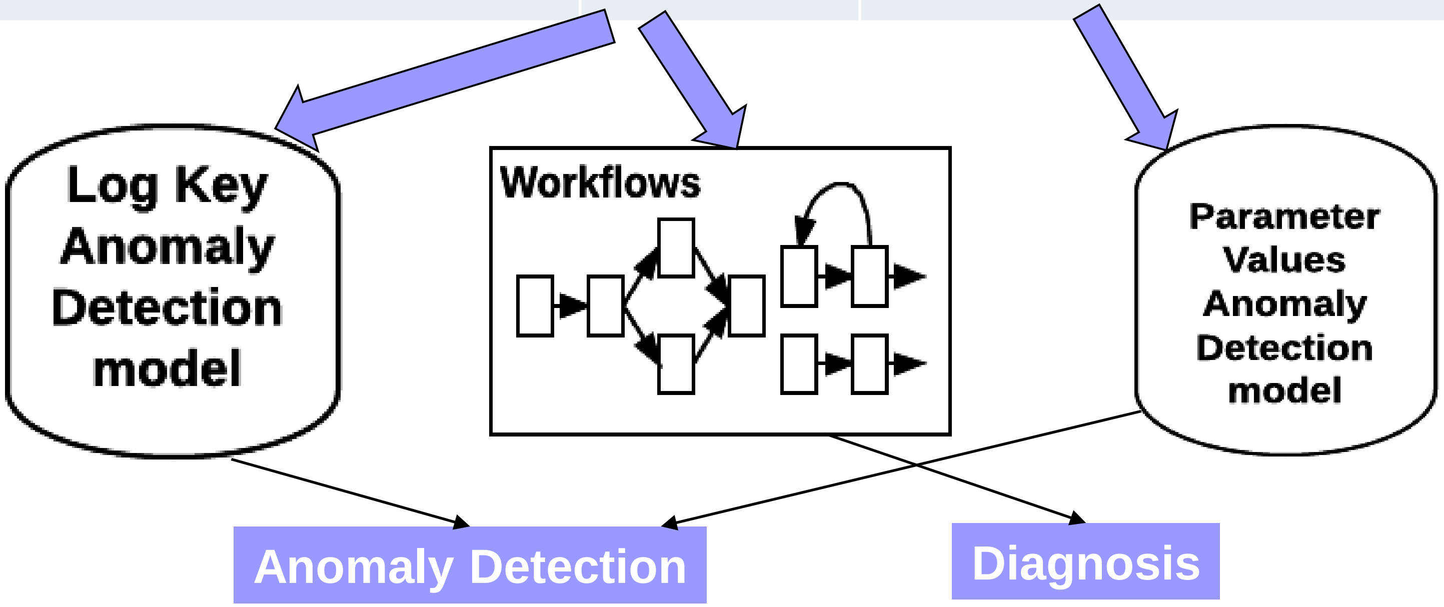
DeepLog

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]
…	…	…


Log Key Anomaly Detection model
Workflows
Parameter Values Anomaly Detection model
Anomaly Detection

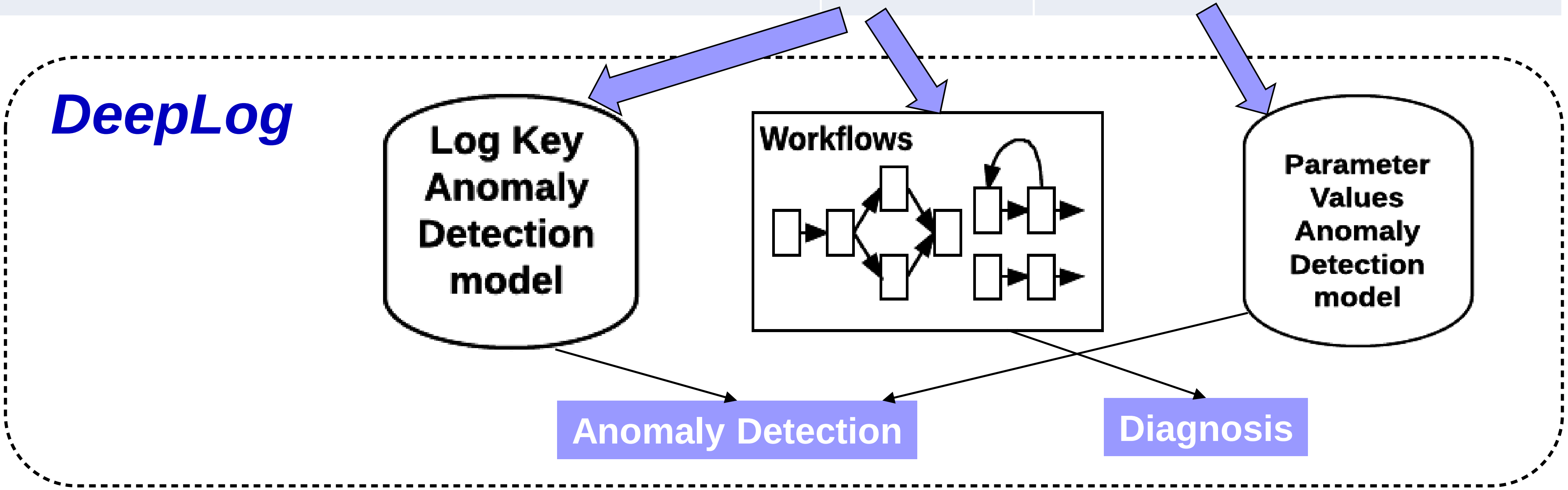
DeepLog

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]
…	…	…

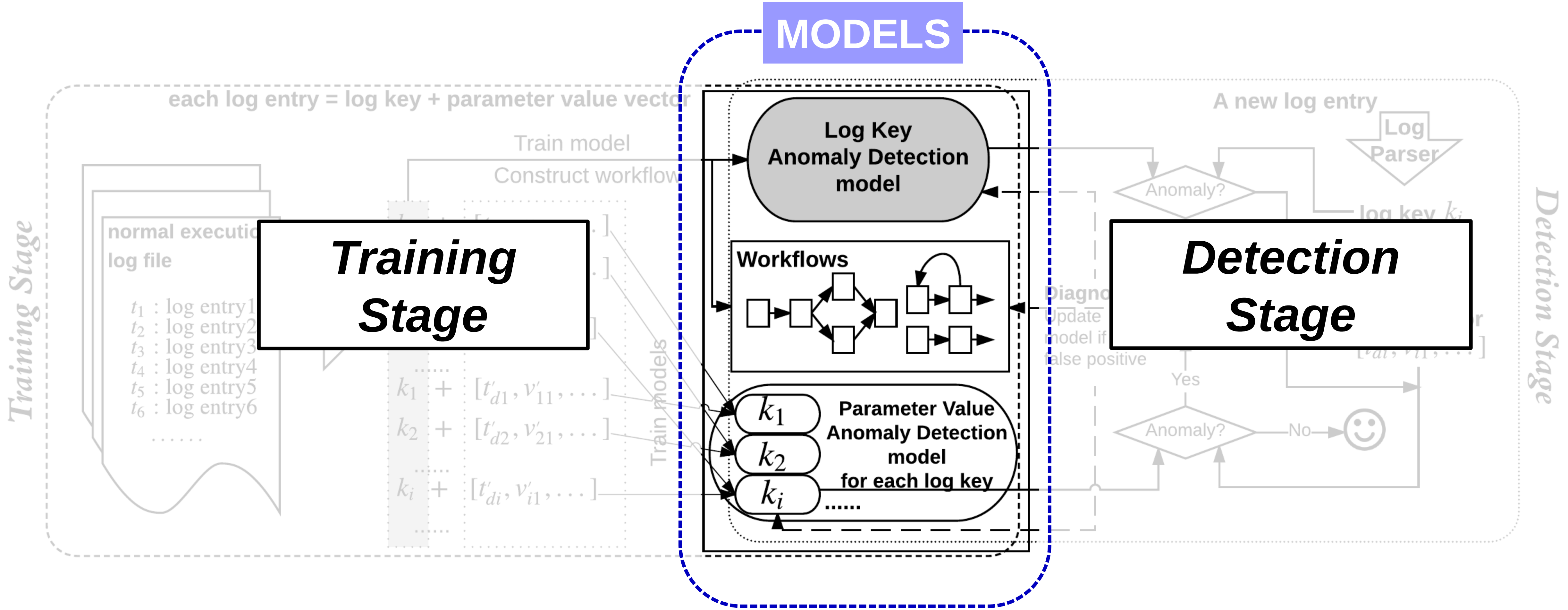

Log Key Anomaly Detection model
Workflows
Parameter Values Anomaly Detection model
Anomaly Detection
Diagnosis

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]
…	…	…

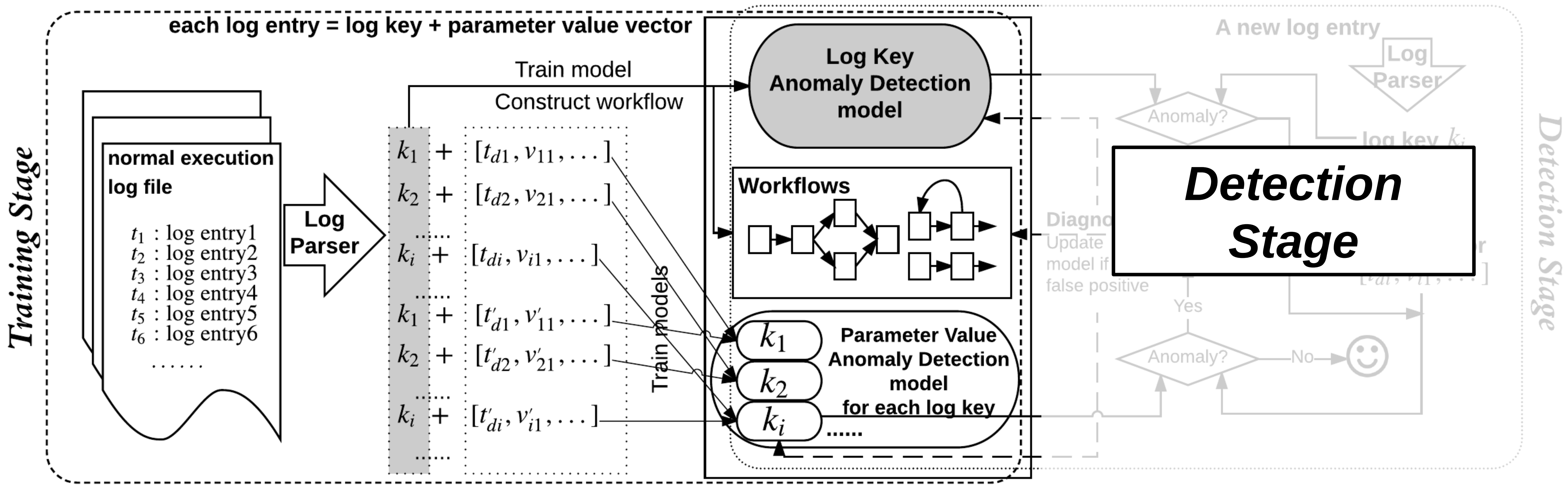

DeepLog
Log Key Anomaly Detection model
Workflows
Parameter Values Anomaly Detection model
Anomaly Detection
Diagnosis

DeepLog Architecture



DeepLog Architecture

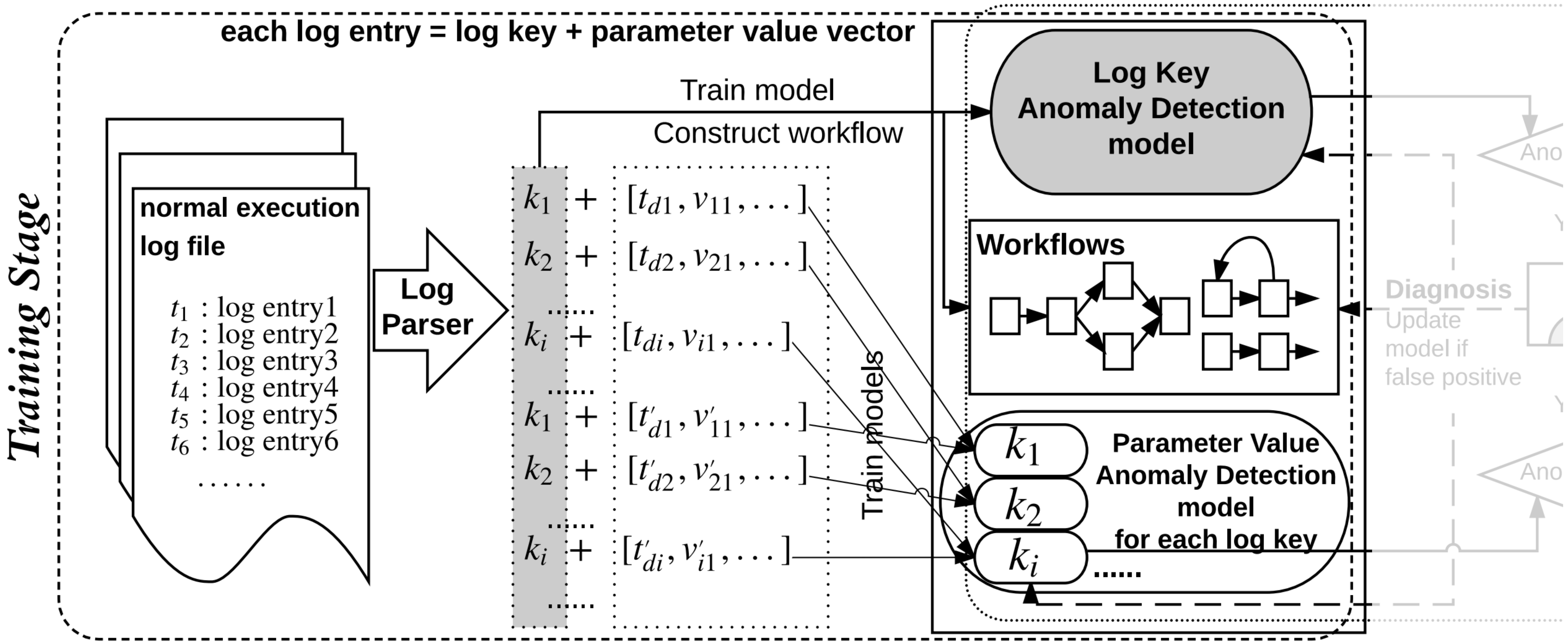
MODELS



Detection Stage

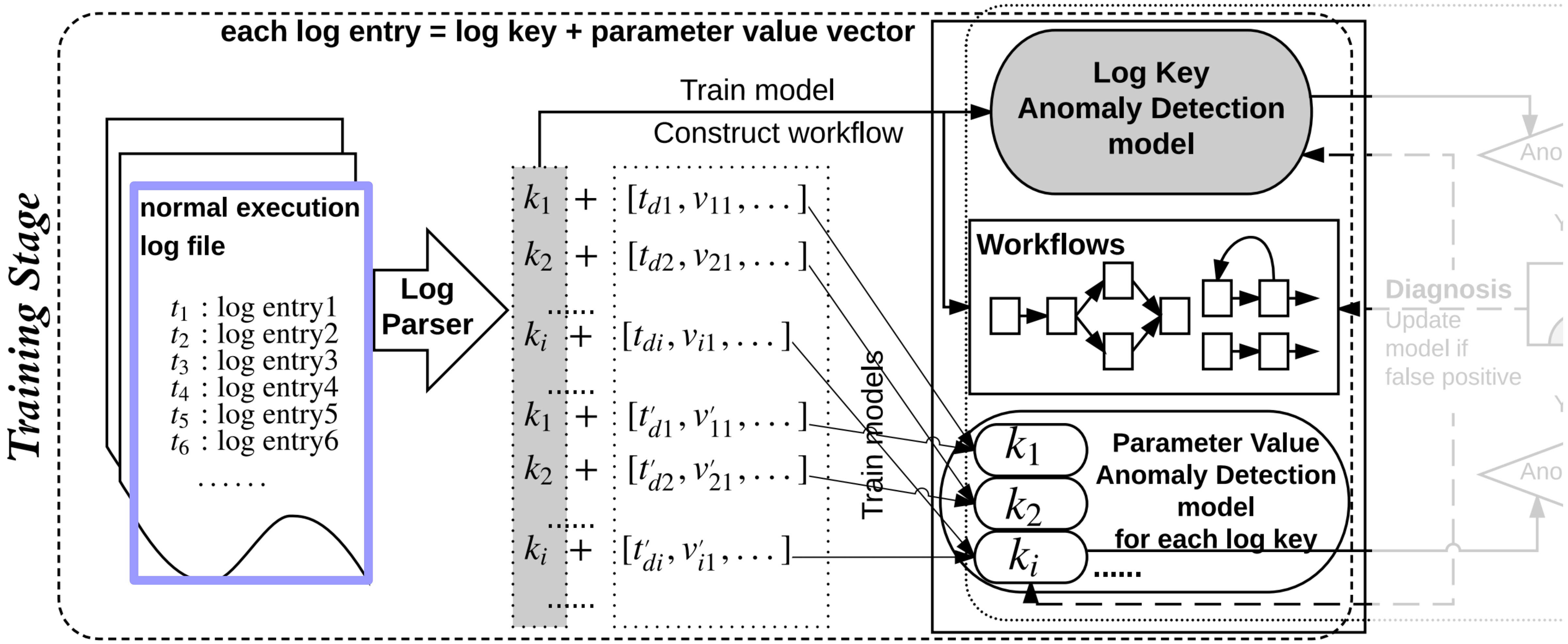
Training Stage
each log entry = log key + parameter value vector
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
......
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
......
k_i + $[t_{di}, v_{i1}, \ldots]$
......
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
......
k_i + $[t'_{di}, v'_{i1}, \ldots]$
......
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......
A new log entry
Log Parser
log key k_i
Anomaly?
Diagno
Update model if false positive
Yes
Anomaly?
No
Detection Stage

DeepLog Architecture

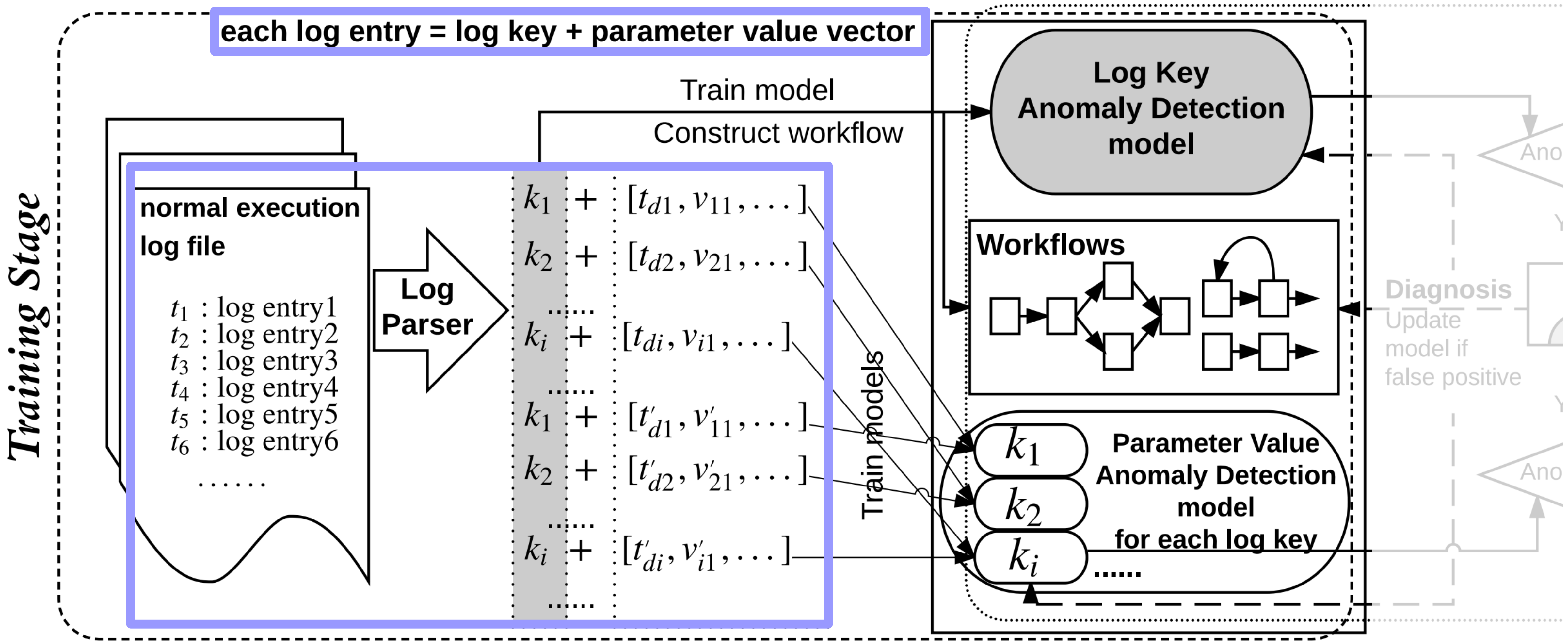


each log entry = log key + parameter value vector
Training Stage
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
......
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
......
k_i + $[t_{di}, v_{i1}, \ldots]$
......
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
......
k_i + $[t'_{di}, v'_{i1}, \ldots]$
......
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
Parameter Value Anomaly Detection model for each log key
k_1
k_2
k_i
......
Ano
Diagnosis
Update model if false positive
Ano

DeepLog Architecture

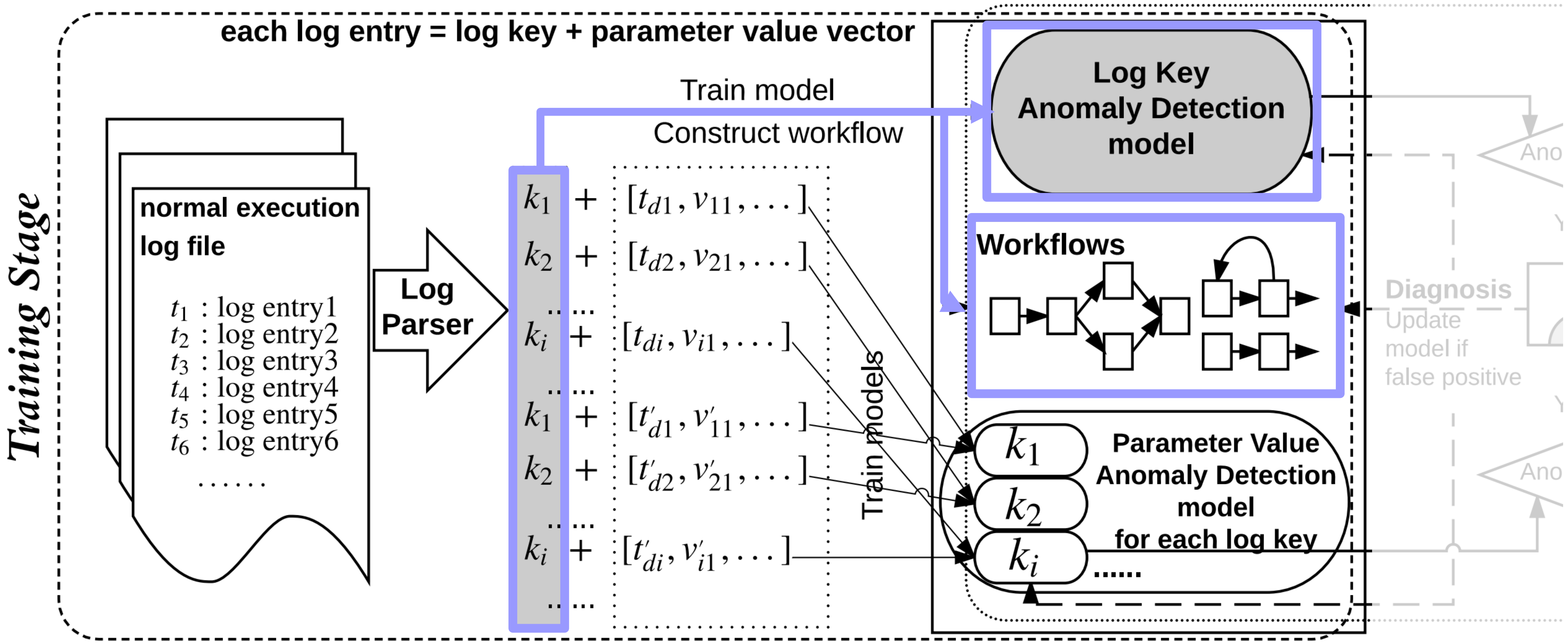

each log entry = log key + parameter value vector
Training Stage
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
......
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
......
k_i + $[t_{di}, v_{i1}, \ldots]$
......
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
......
k_i + $[t'_{di}, v'_{i1}, \ldots]$
......
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
Parameter Value Anomaly Detection model for each log key
k_1
k_2
k_i
......
Ano
Diagnosis
Update model if false positive
Ano

DeepLog Architecture

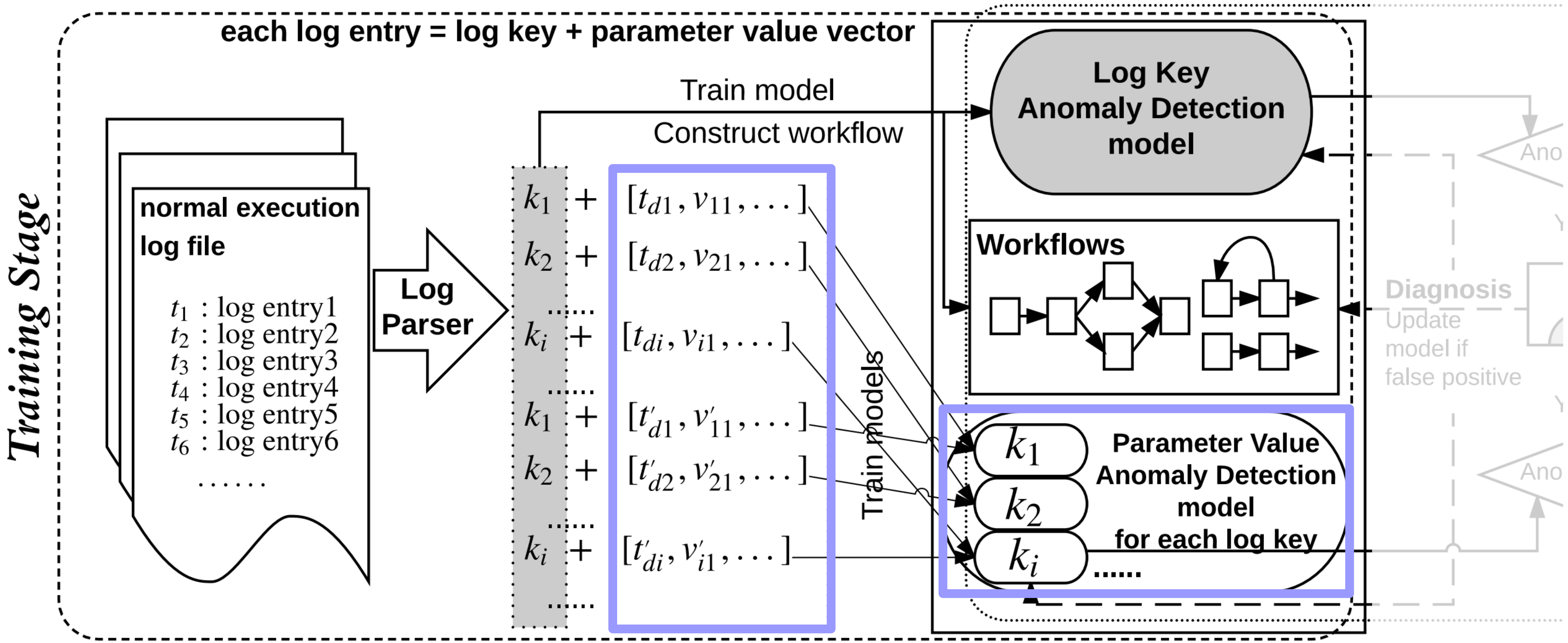


each log entry = log key + parameter value vector
Training Stage
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
……
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
k_i + $[t_{di}, v_{i1}, \ldots]$
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
k_i + $[t'_{di}, v'_{i1}, \ldots]$
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
Parameter Value Anomaly Detection model for each log key
k_1
k_2
k_i
......
Diagnosis
Update model if false positive
Ano
Ano

DeepLog Architecture

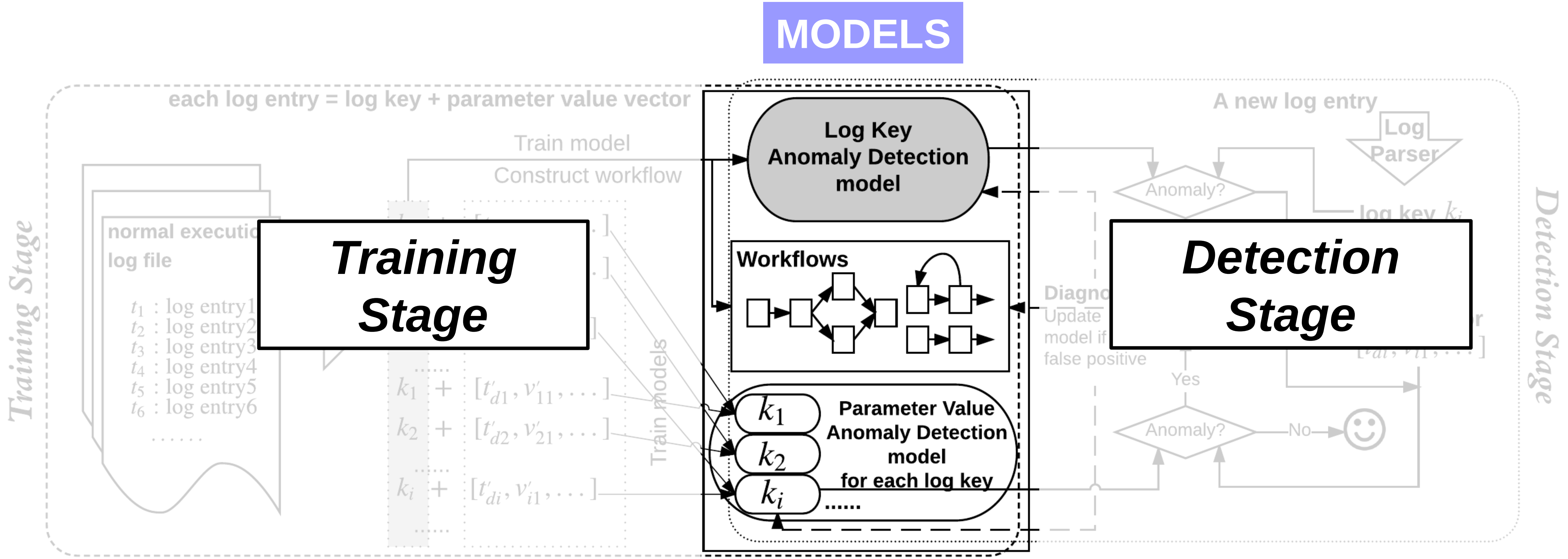

Training Stage
each log entry = log key + parameter value vector
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
......
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
......
k_i + $[t_{di}, v_{i1}, \ldots]$
......
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
......
k_i + $[t'_{di}, v'_{i1}, \ldots]$
......
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......
Ano
Diagnosis
Update model if false positive
Ano

DeepLog Architecture



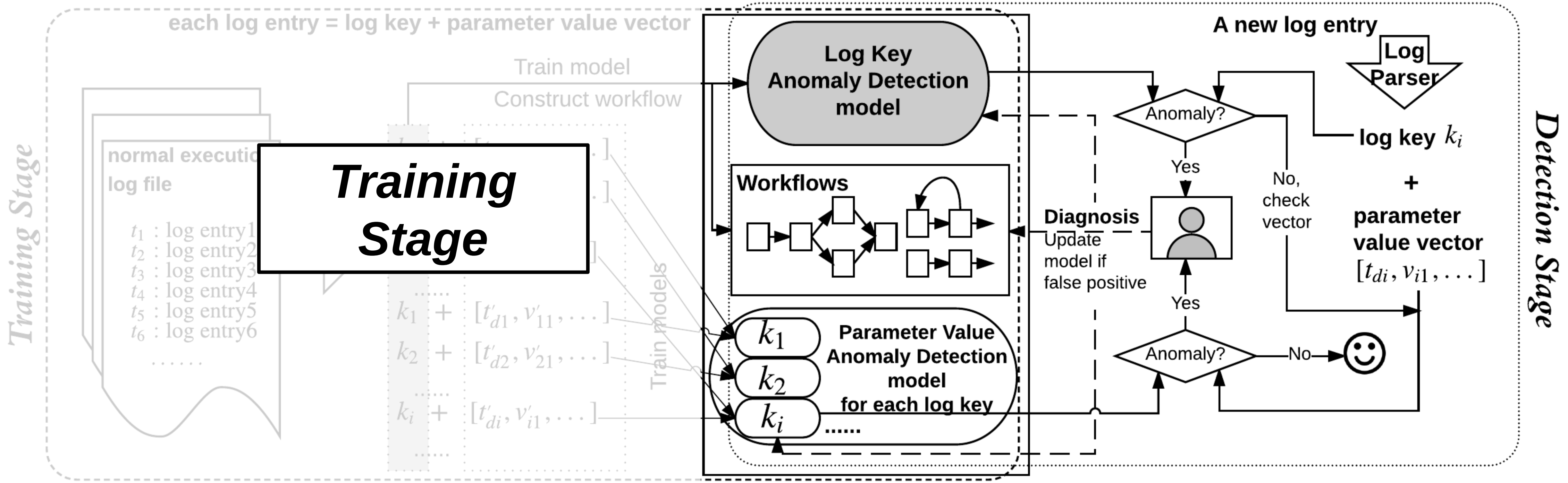
each log entry = log key + parameter value vector
Training Stage
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
……
Log Parser
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
k_i + $[t_{di}, v_{i1}, \ldots]$
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
k_i + $[t'_{di}, v'_{i1}, \ldots]$
Train model
Construct workflow
Train models
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......
Diagnosis
Update model if false positive
Ano
Ano

DeepLog Architecture

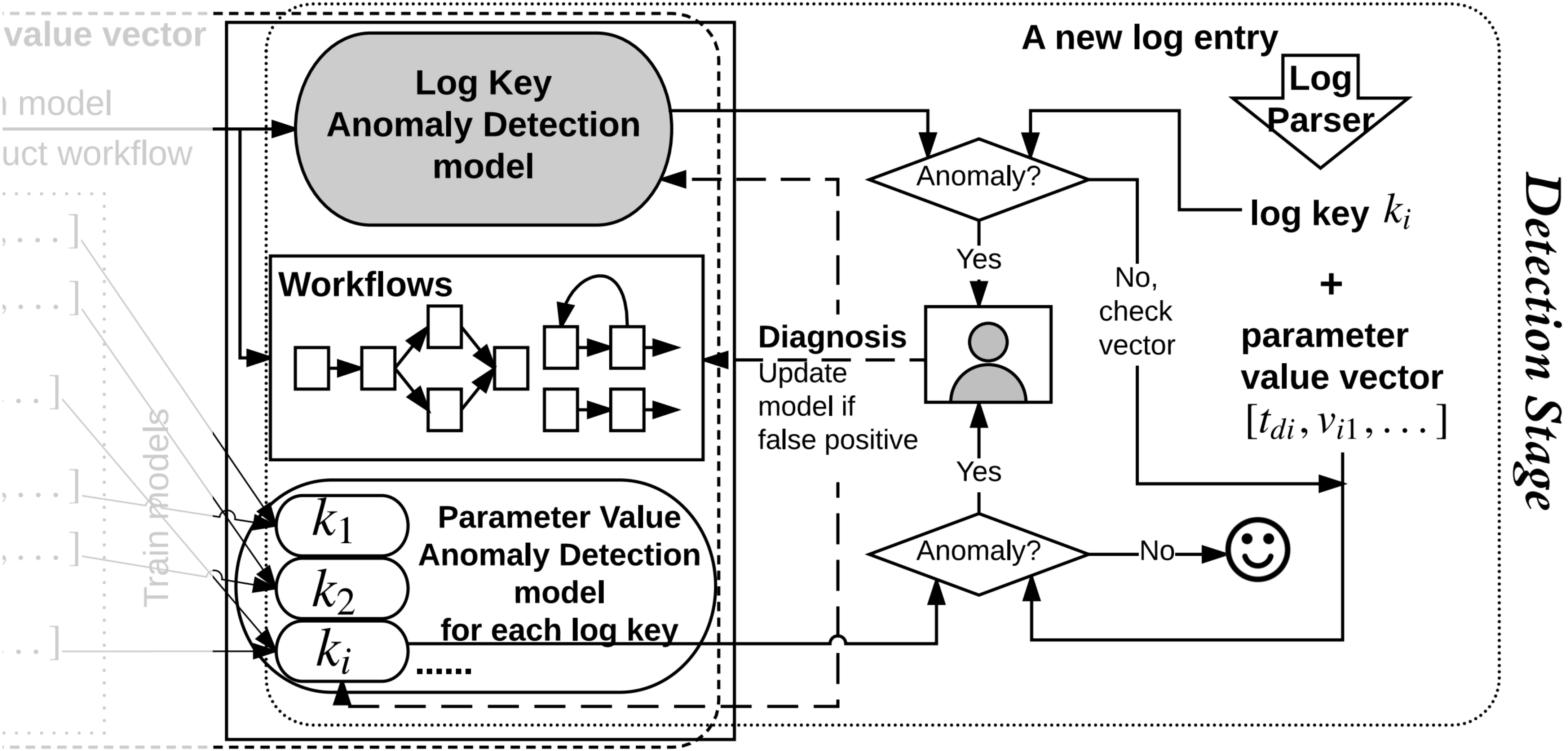

MODELS
Training Stage
Detection Stage
each log entry = log key + parameter value vector
Train model
Construct workflow
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
k_i + $[t'_{di}, v'_{i1}, \ldots]$
Train models
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......
A new log entry
Log Parser
Anomaly?
log key k_i
Diagnosis
Update model if false positive
Yes
Anomaly?
No
☺

DeepLog Architecture

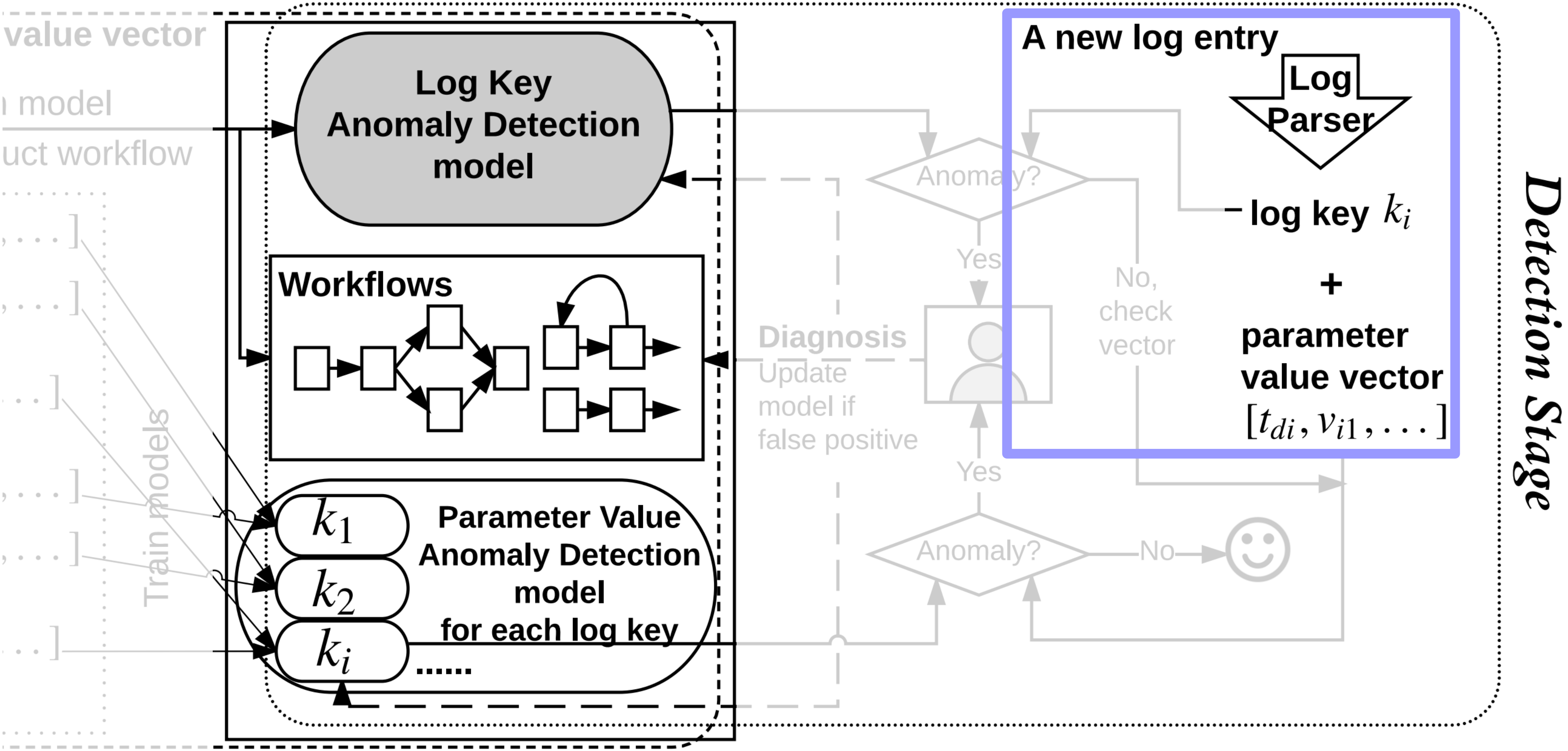
MODELS



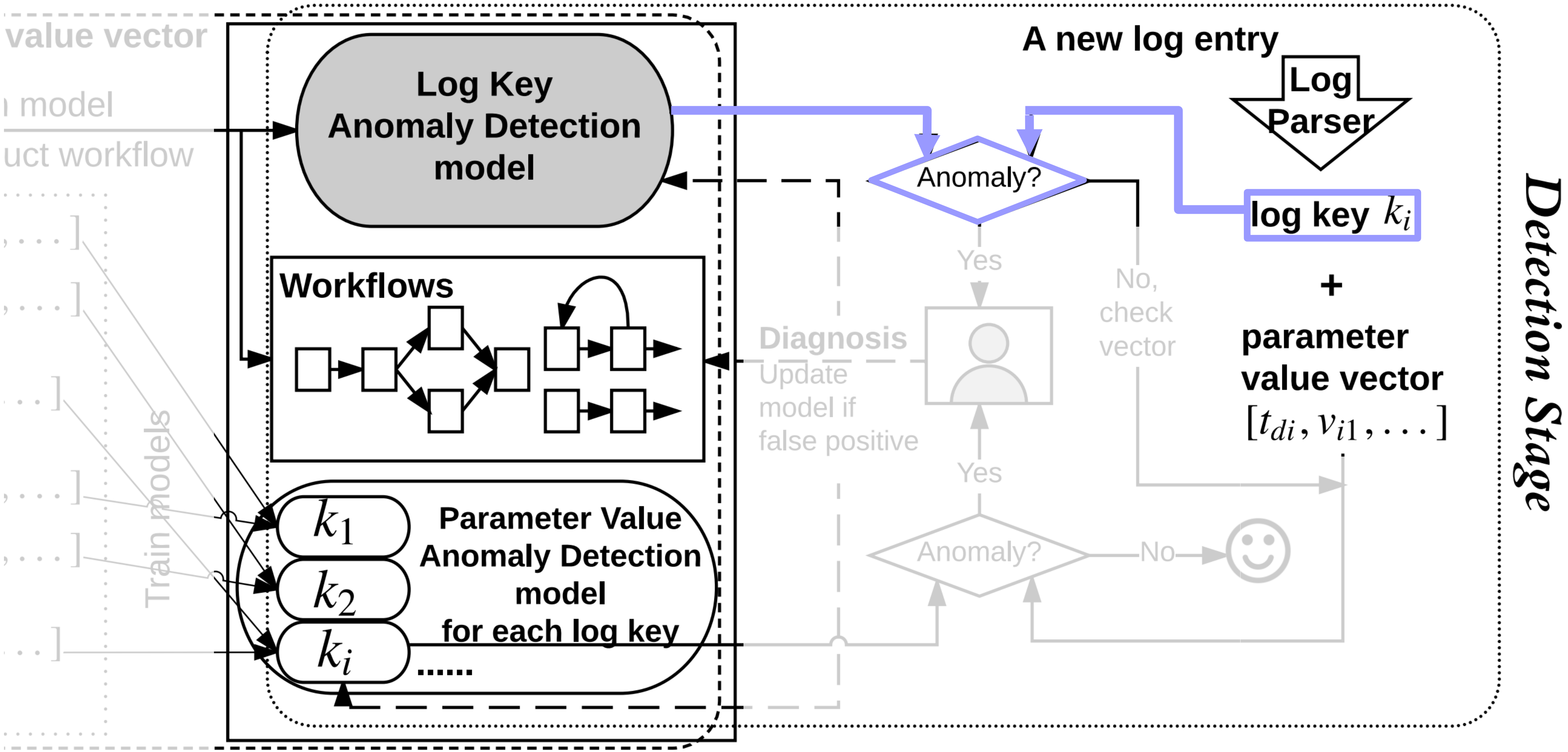
DeepLog Architecture



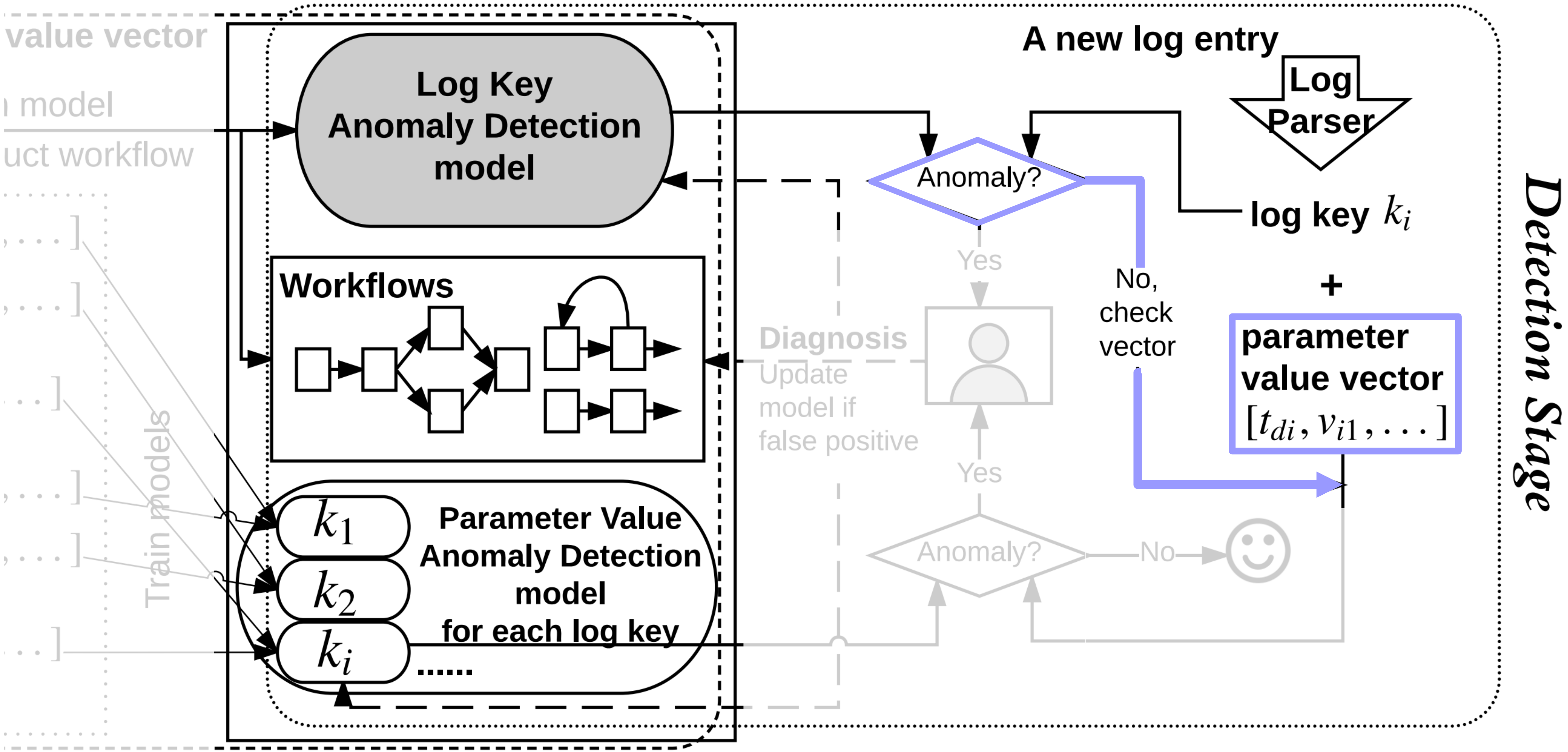
DeepLog Architecture



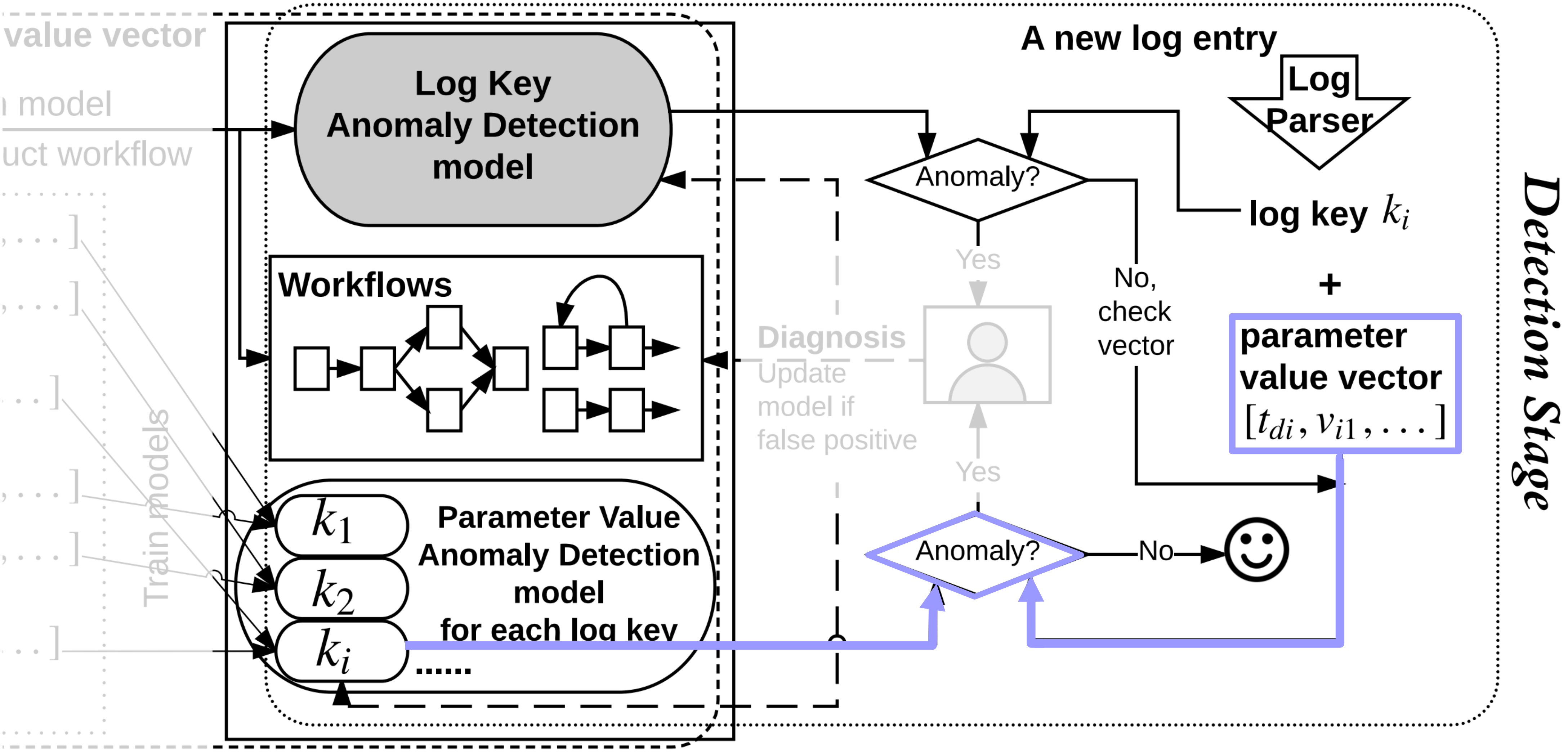
DeepLog Architecture



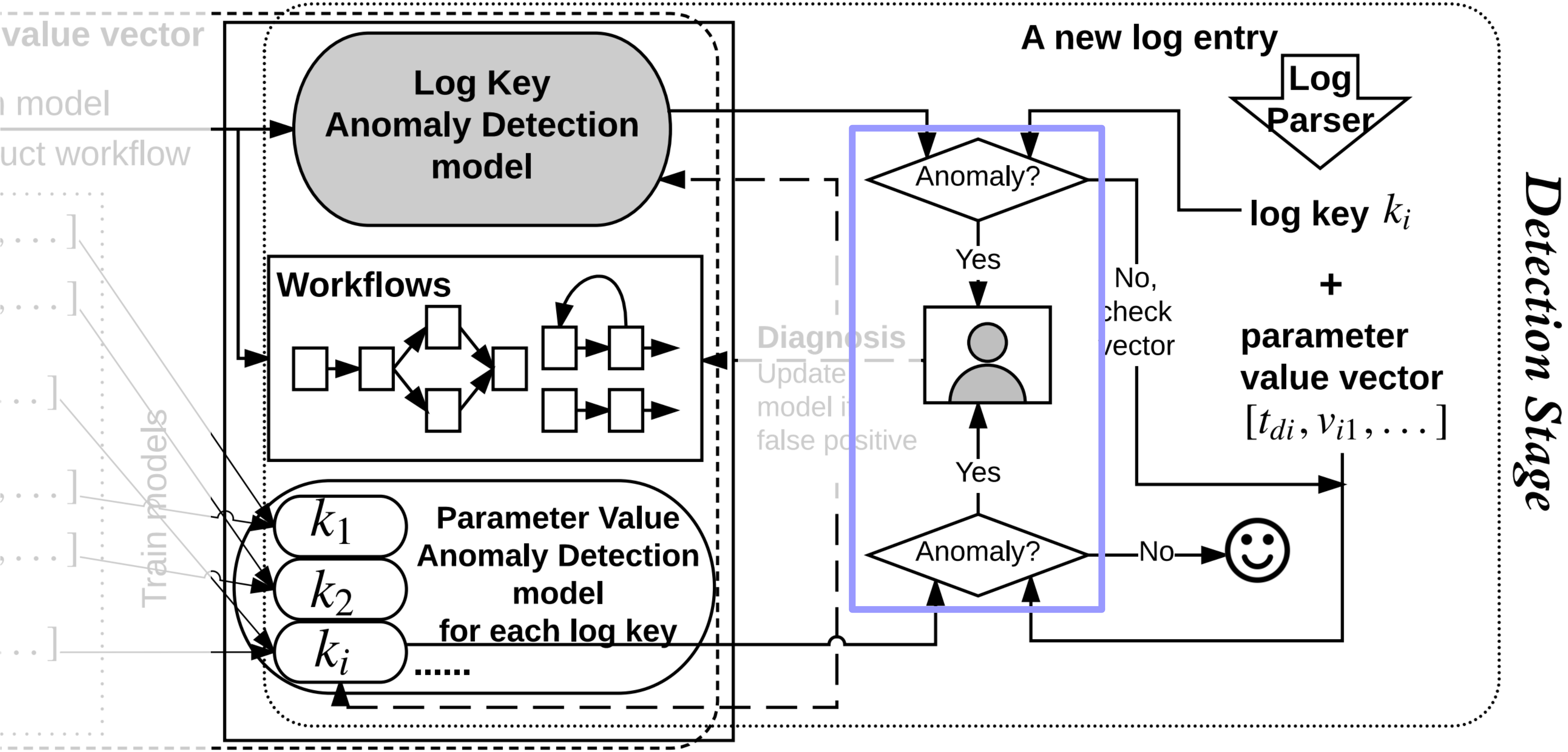
DeepLog Architecture



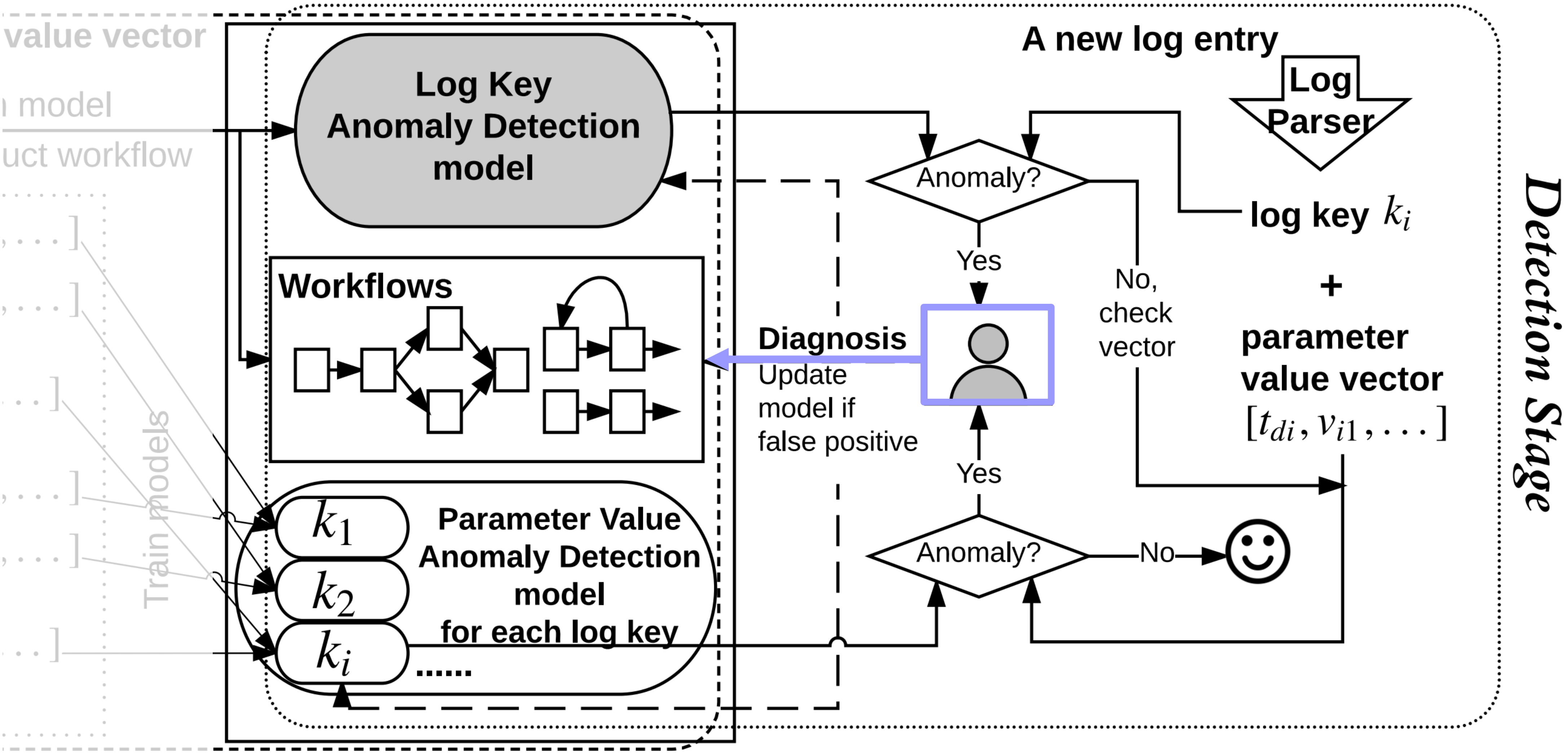
DeepLog Architecture



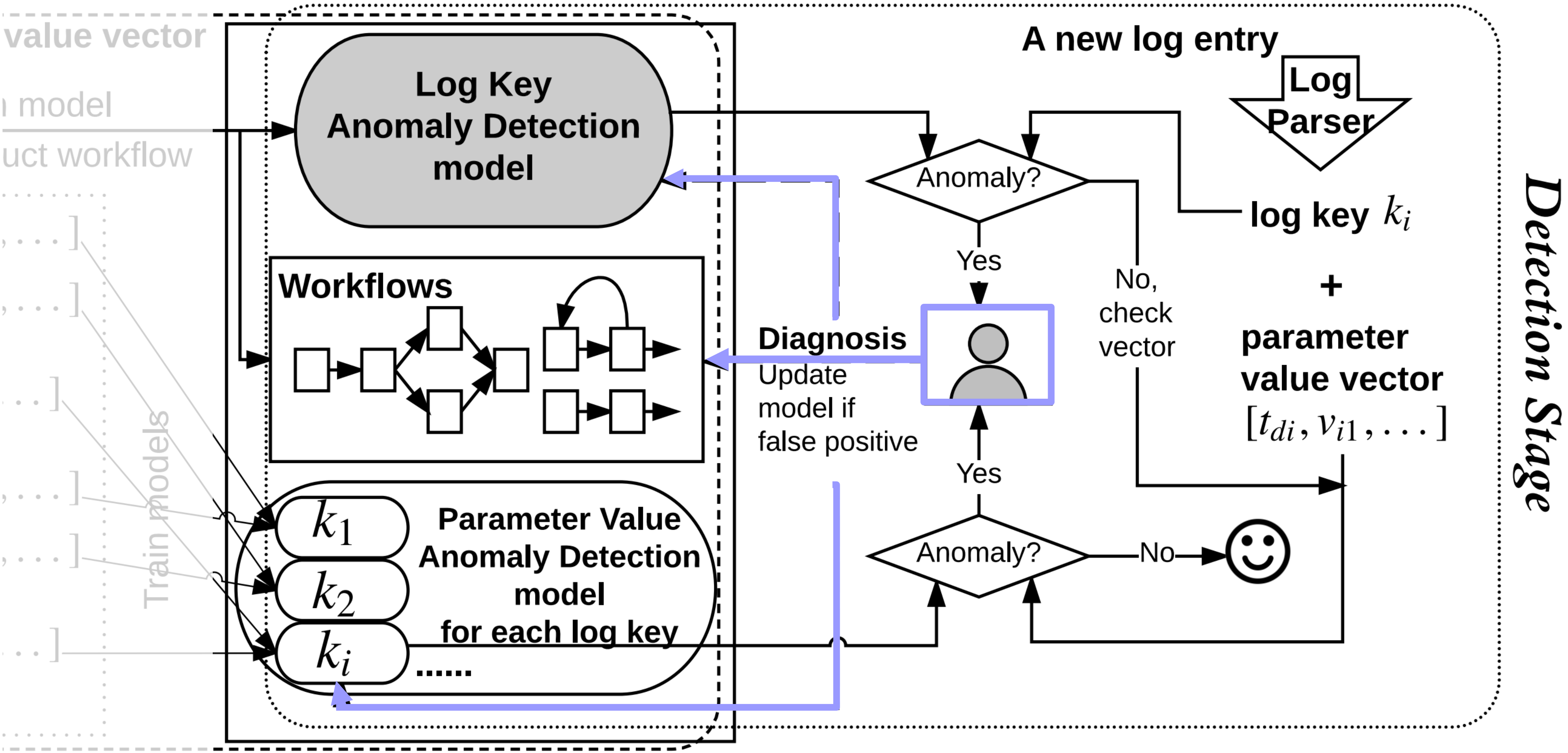
DeepLog Architecture



DeepLog Architecture

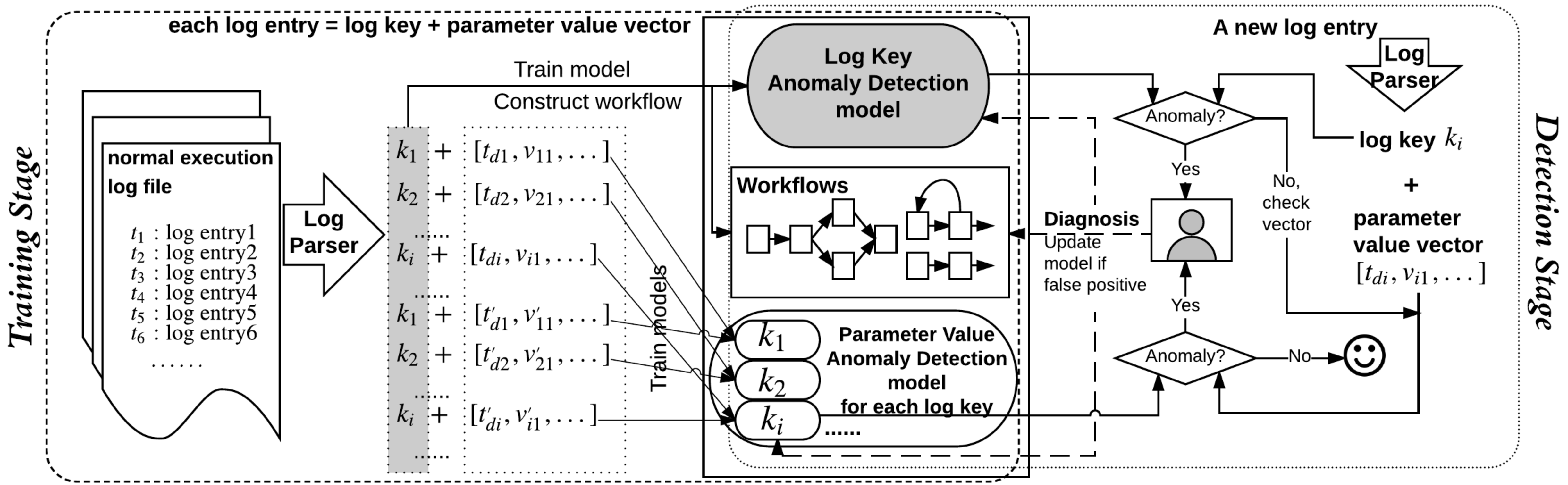


DeepLog Architecture



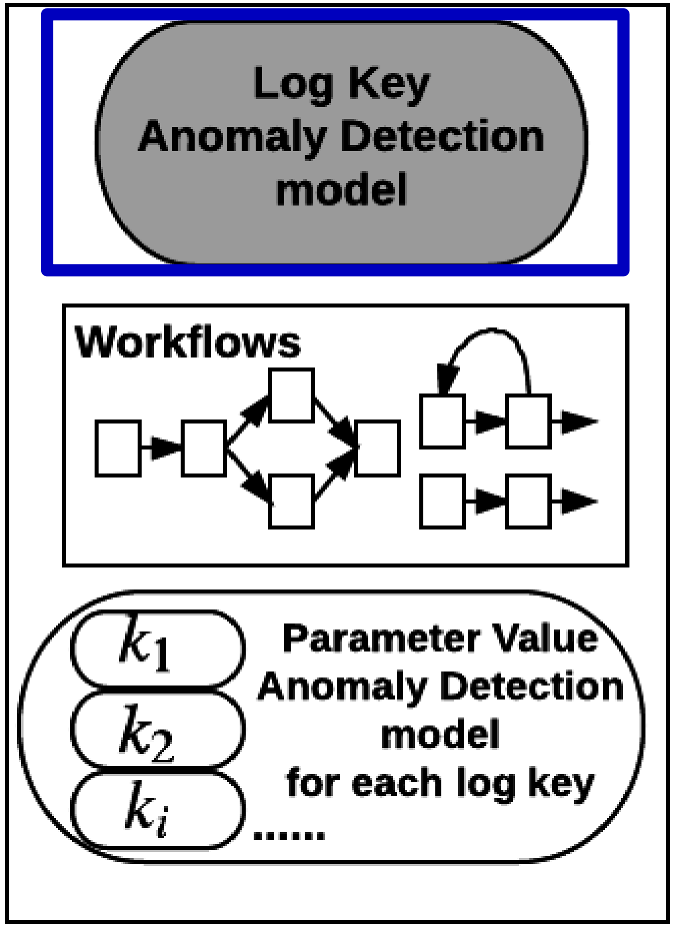
DeepLog Architecture

MODELS



each log entry = log key + parameter value vector
Training Stage
normal execution log file
t_1 : log entry1
t_2 : log entry2
t_3 : log entry3
t_4 : log entry4
t_5 : log entry5
t_6 : log entry6
......
Log Parser
Train model
Construct workflow
k_1 + $[t_{d1}, v_{11}, \ldots]$
k_2 + $[t_{d2}, v_{21}, \ldots]$
......
k_i + $[t_{di}, v_{i1}, \ldots]$
......
k_1 + $[t'_{d1}, v'_{11}, \ldots]$
k_2 + $[t'_{d2}, v'_{21}, \ldots]$
......
k_i + $[t'_{di}, v'_{i1}, \ldots]$
......
Train models
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
......
Parameter Value Anomaly Detection model for each log key
Diagnosis
Update model if false positive
A new log entry
Log Parser
log key k_i
+
parameter value vector
$[t_{di}, v_{i1}, \ldots]$
Anomaly?
Yes
No, check vector
Yes
Anomaly?
No
☺
Detection Stage

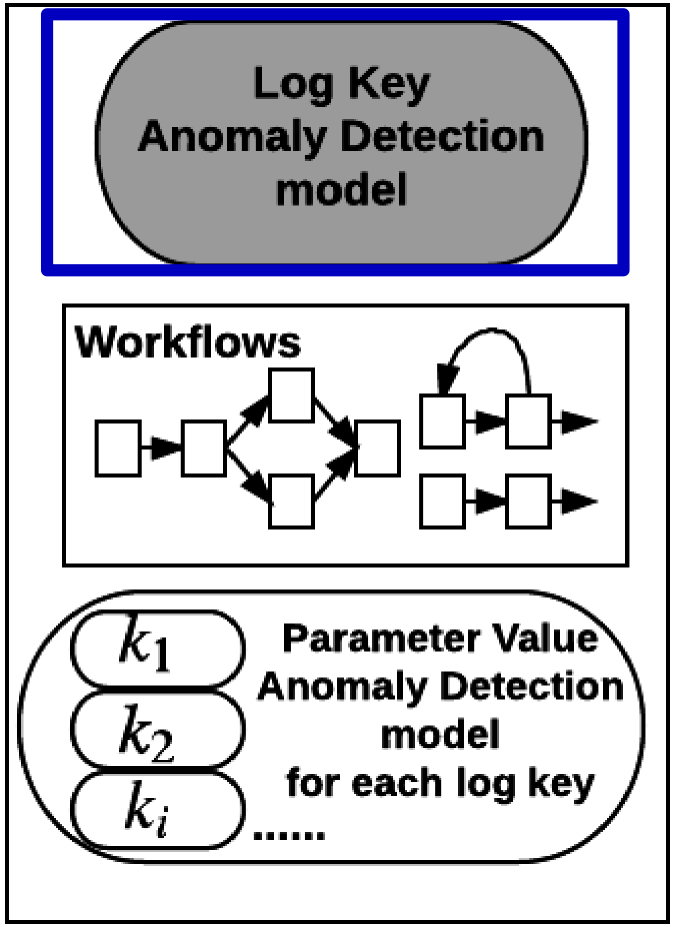
Log Key Anomaly Detection model



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

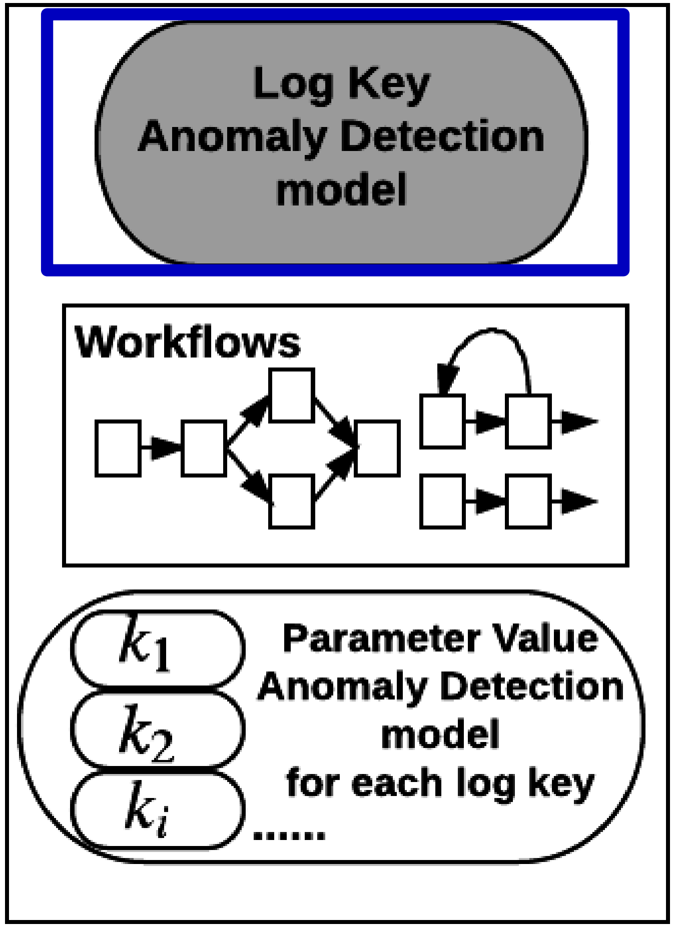
⇩

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

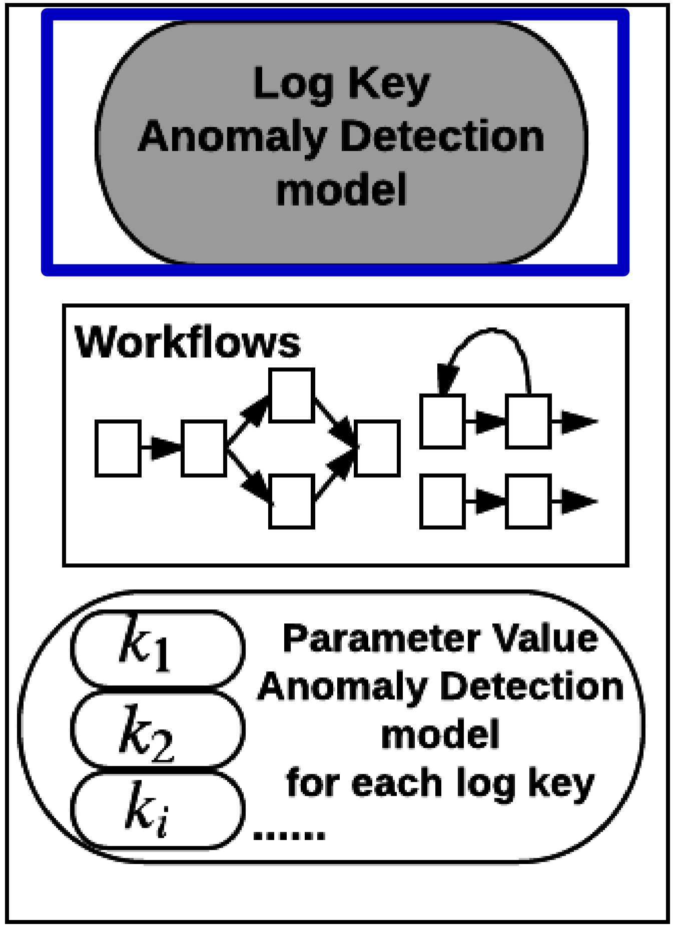
⇩

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

⇩

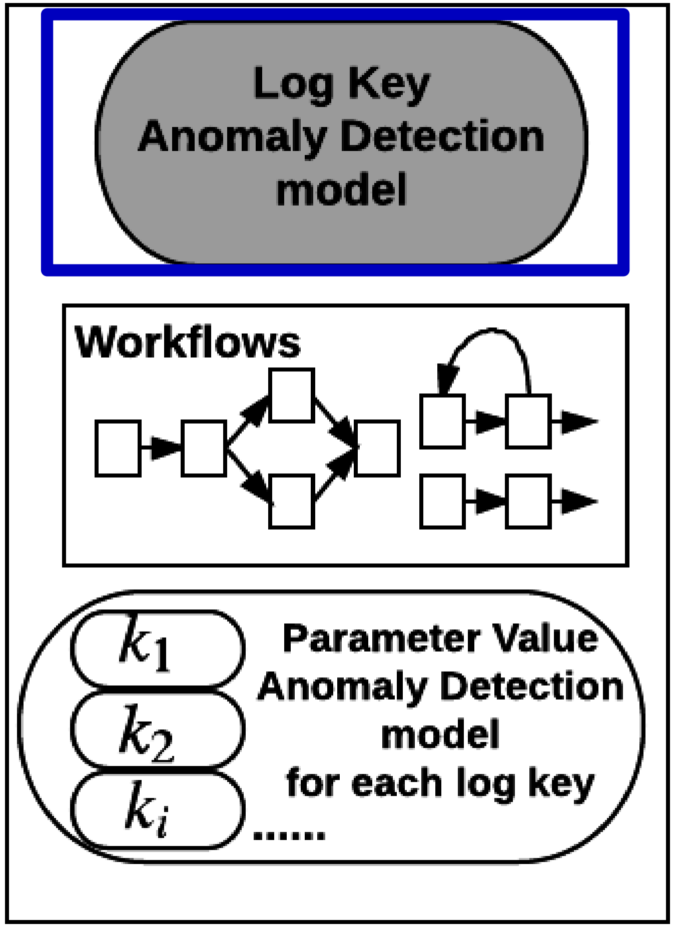
A log key is detected to be abnormal if it does not follow the prediction.

Log Key Anomaly Detection model

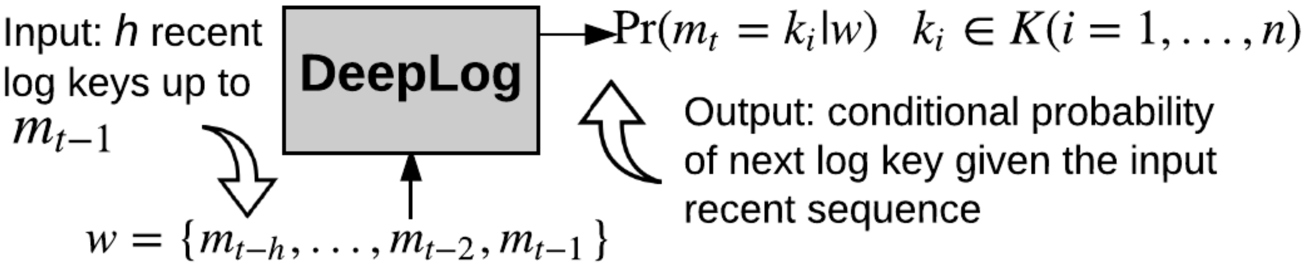


Use long short-term memory (LSTM) architecture

Log Key Anomaly Detection model



Use long short-term memory (LSTM) architecture



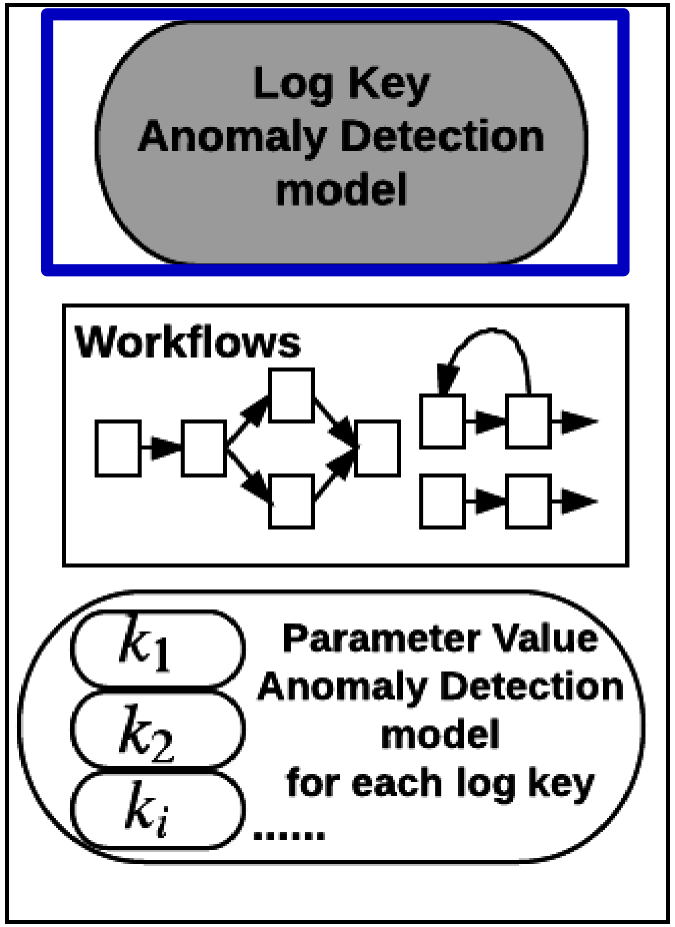
Input: *h* recent log keys up to m_{t-1}

$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

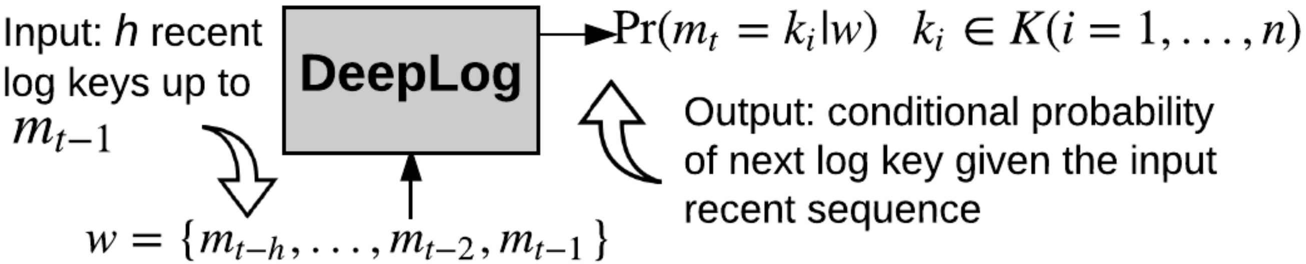
$\Pr(m_t = k_i | w) \quad k_i \in K (i = 1, \ldots, n)$

Output: conditional probability of next log key given the input recent sequence

Log Key Anomaly Detection model


Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

Use long short-term memory (LSTM) architecture

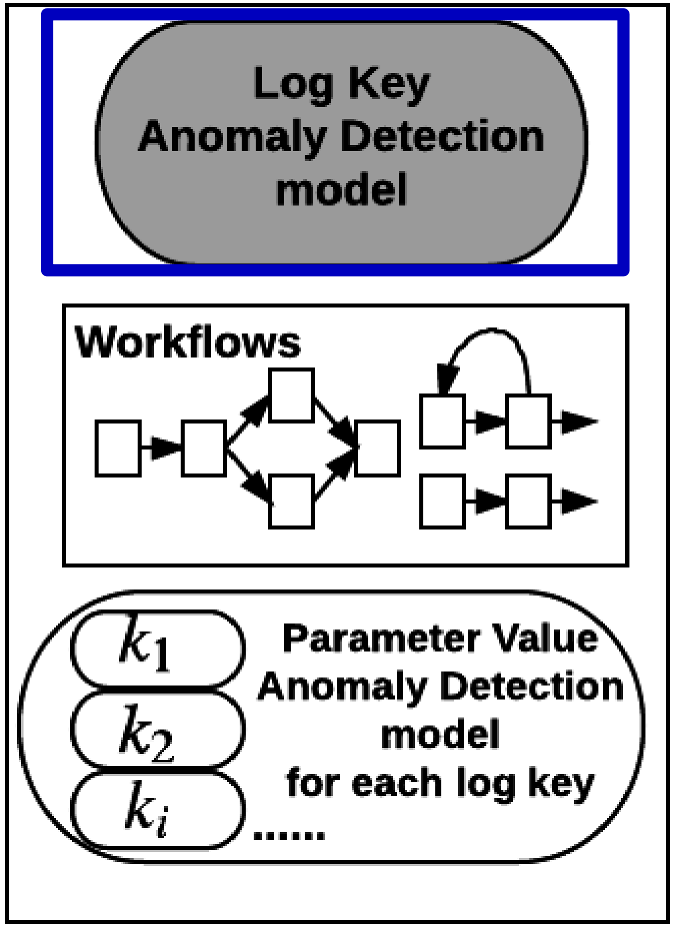

Input: h recent log keys up to m_{t-1}
DeepLog
$\Pr(m_t = k_i | w) \quad k_i \in K(i = 1, \ldots, n)$
Output: conditional probability of next log key given the input recent sequence
$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

Training:

log key sequence:

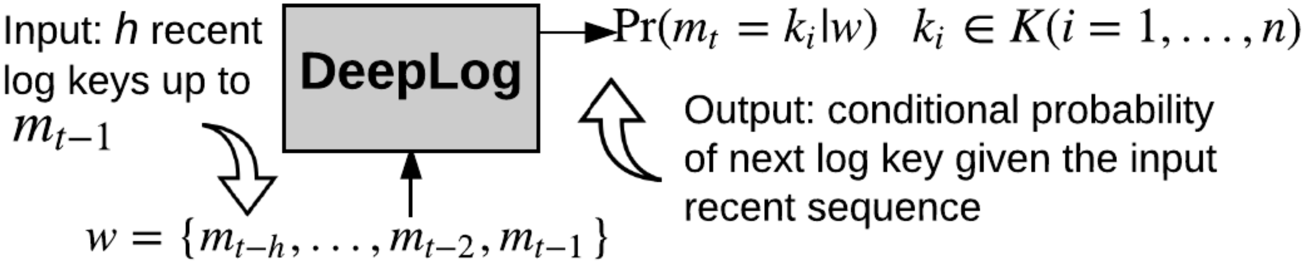
h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

Use long short-term memory (LSTM) architecture



Input: *h* recent log keys up to m_{t-1}

DeepLog

$\Pr(m_t = k_i | w) \quad k_i \in K (i = 1, \ldots, n)$

Output: conditional probability of next log key given the input recent sequence

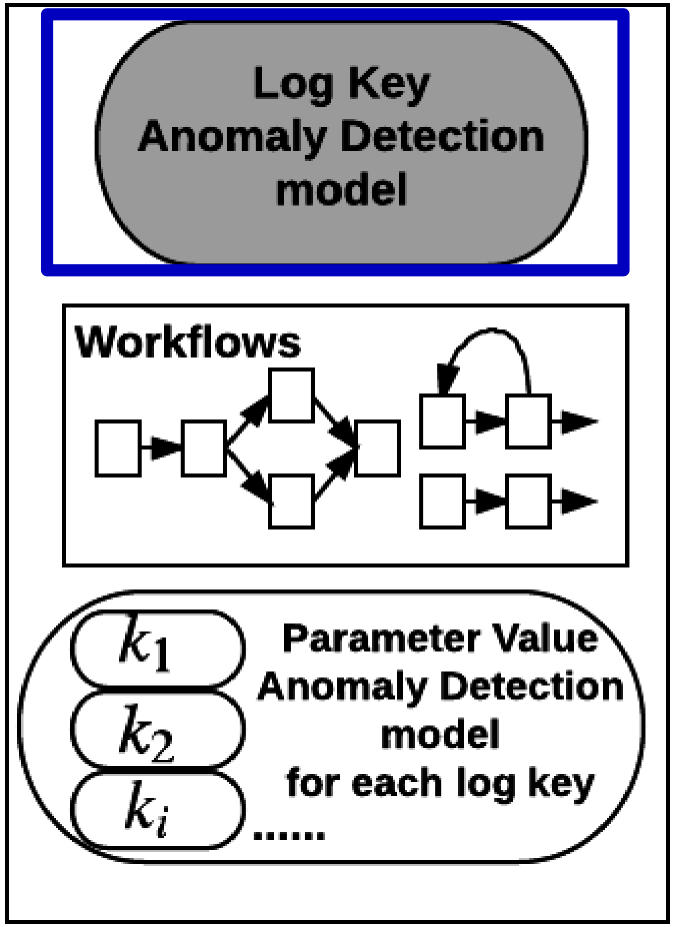
$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

Training:

log key sequence:

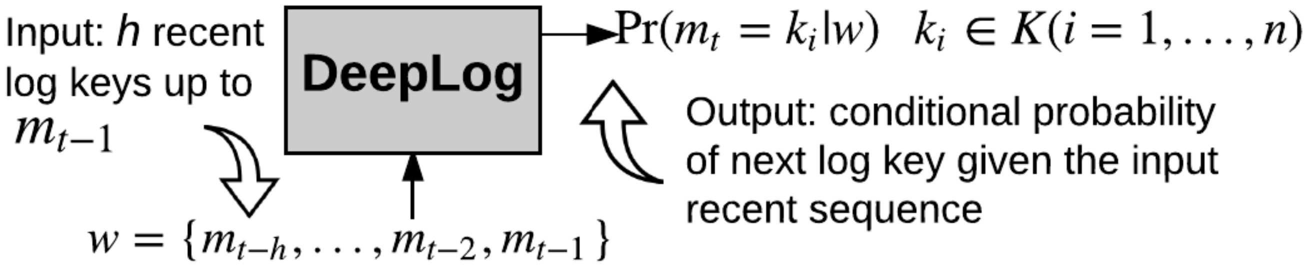
h=3 25 18 54 57 18 56 … 25 18 54 57 56 18 …

Log Key Anomaly Detection model



Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

Use long short-term memory (LSTM) architecture



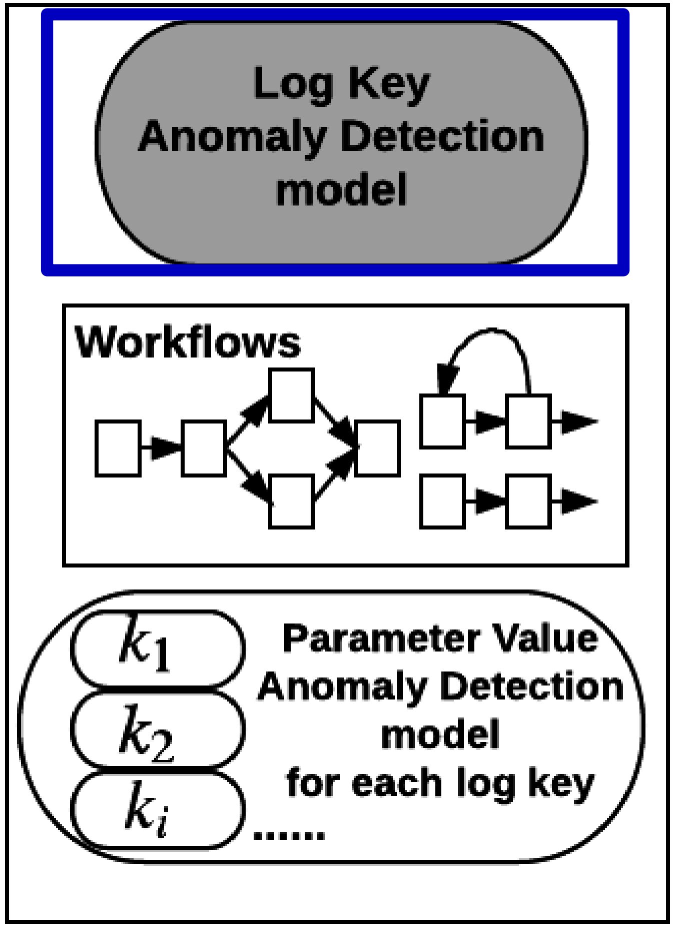
Input: *h* recent log keys up to m_{t-1}
DeepLog
$\Pr(m_t = k_i | w) \quad k_i \in K (i = 1, \ldots, n)$
Output: conditional probability of next log key given the input recent sequence
$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

Training:

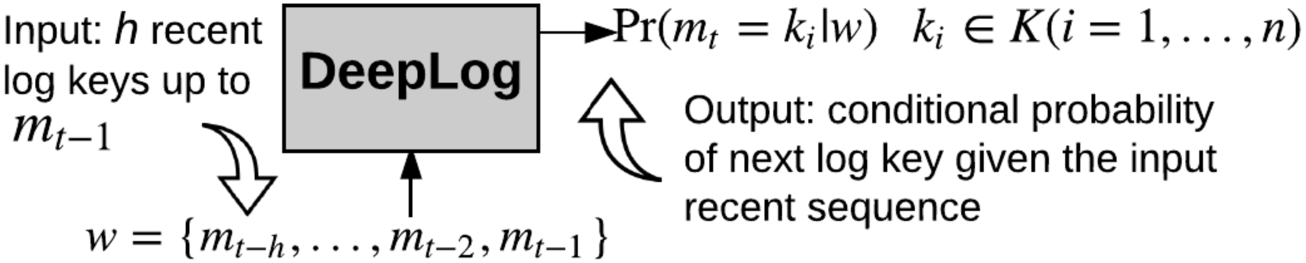
log key sequence:

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

Log Key Anomaly Detection model


Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

Use long short-term memory (LSTM) architecture

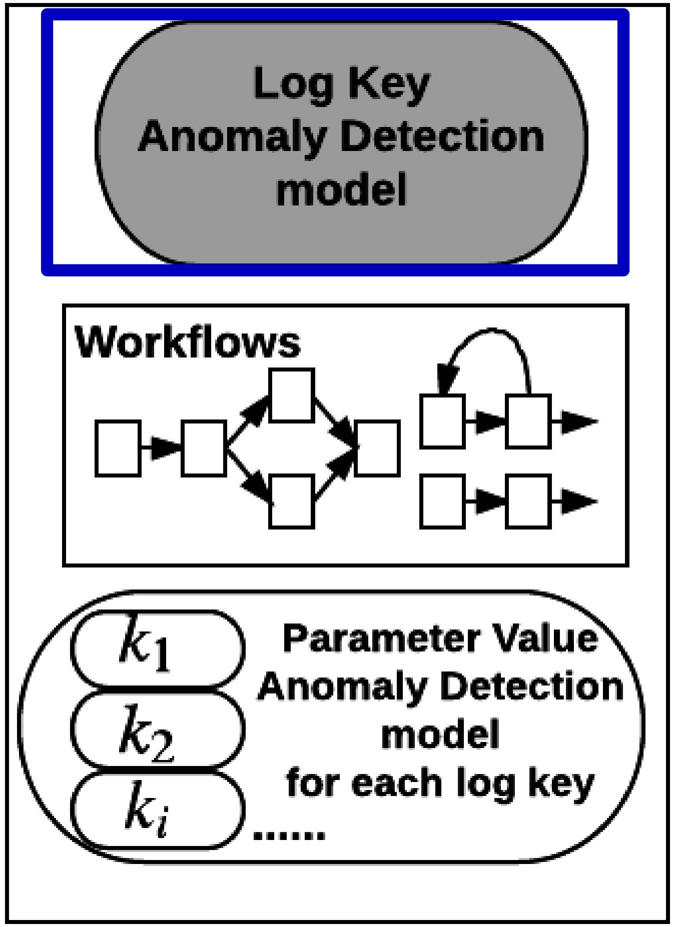

Input: h recent log keys up to m_{t-1}
DeepLog
$\Pr(m_t = k_i | w) \quad k_i \in K (i = 1, \ldots, n)$
Output: conditional probability of next log key given the input recent sequence
$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

Training:

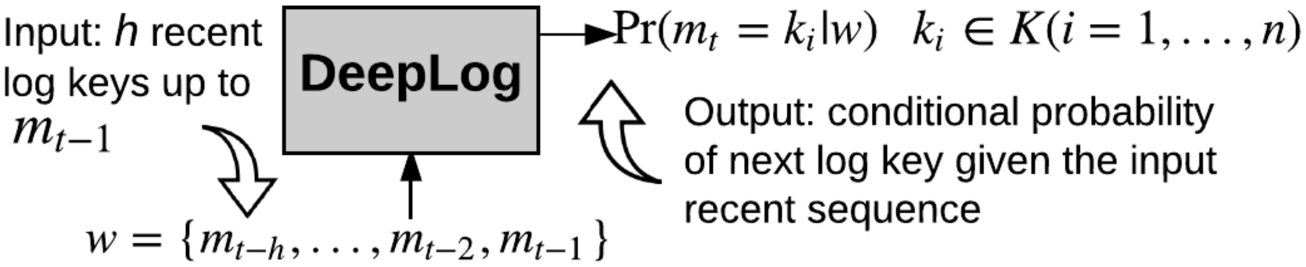
log key sequence:

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

Log Key Anomaly Detection model



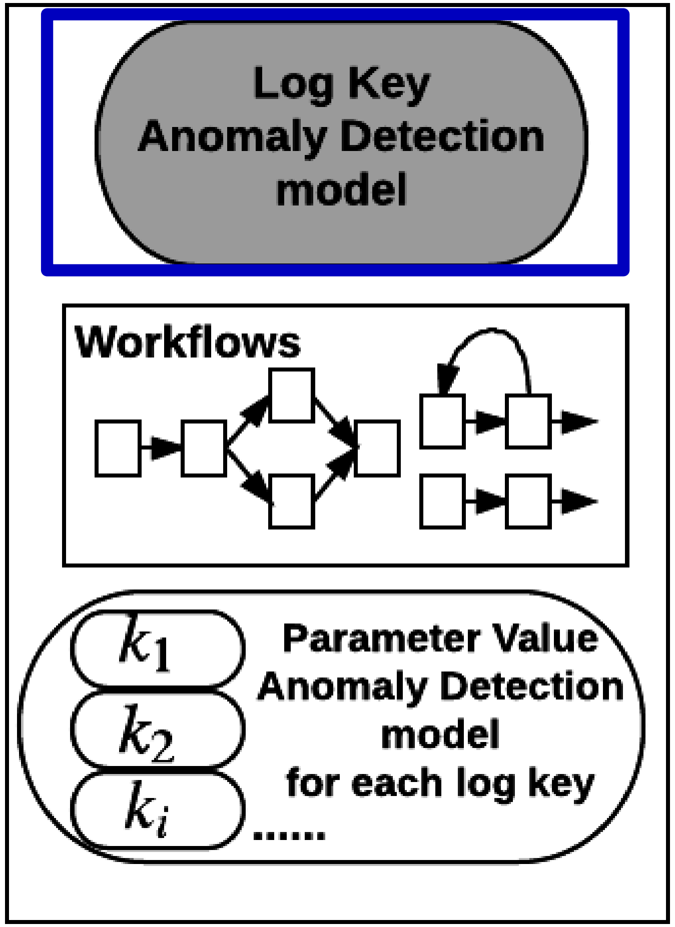
Use long short-term memory (LSTM) architecture

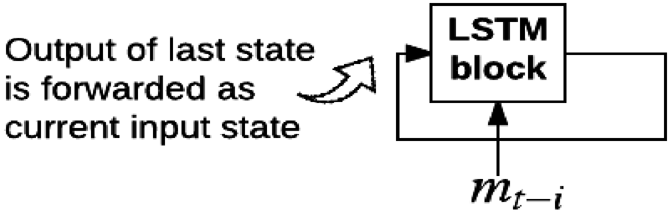

Input: *h* recent
log keys up to
m_{t-1}
DeepLog
$\Pr(m_t = k_i|w) \quad k_i \in K(i = 1, \ldots, n)$
Output: conditional probability
of next log key given the input
recent sequence
$w = \{m_{t-h}, \ldots, m_{t-2}, m_{t-1}\}$

Detection:

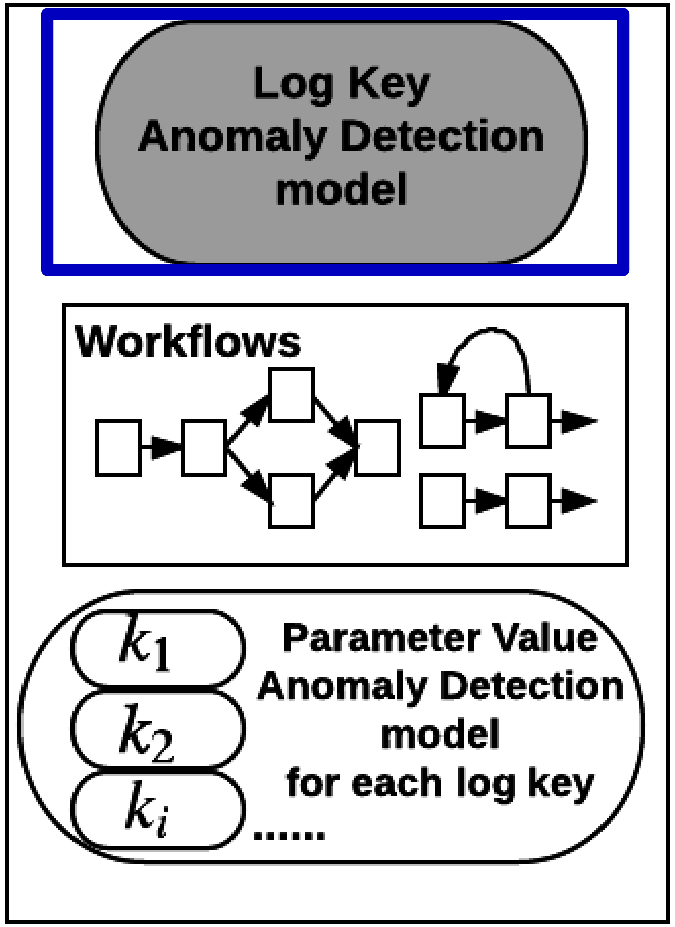
In detection stage, DeepLog checks if the actual next log key is among its top *g* probable predictions.

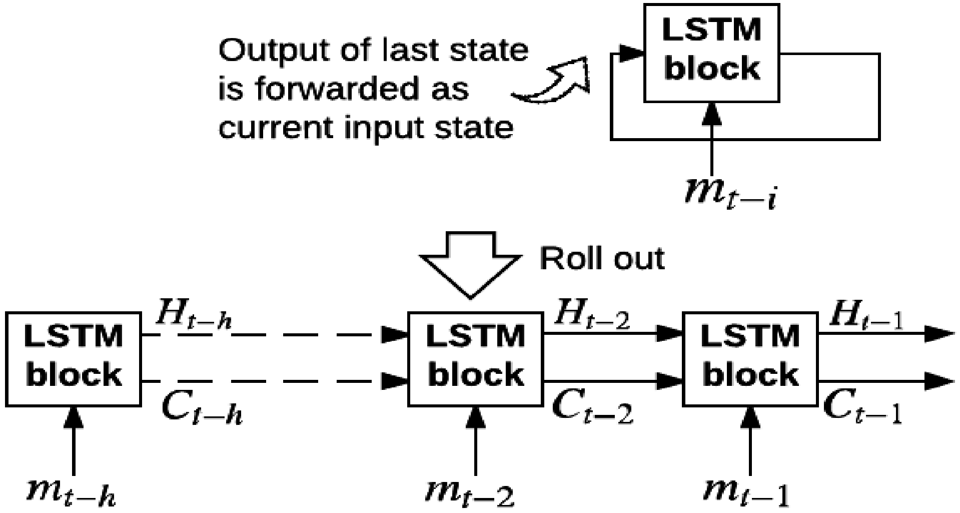
Log Key Anomaly Detection model



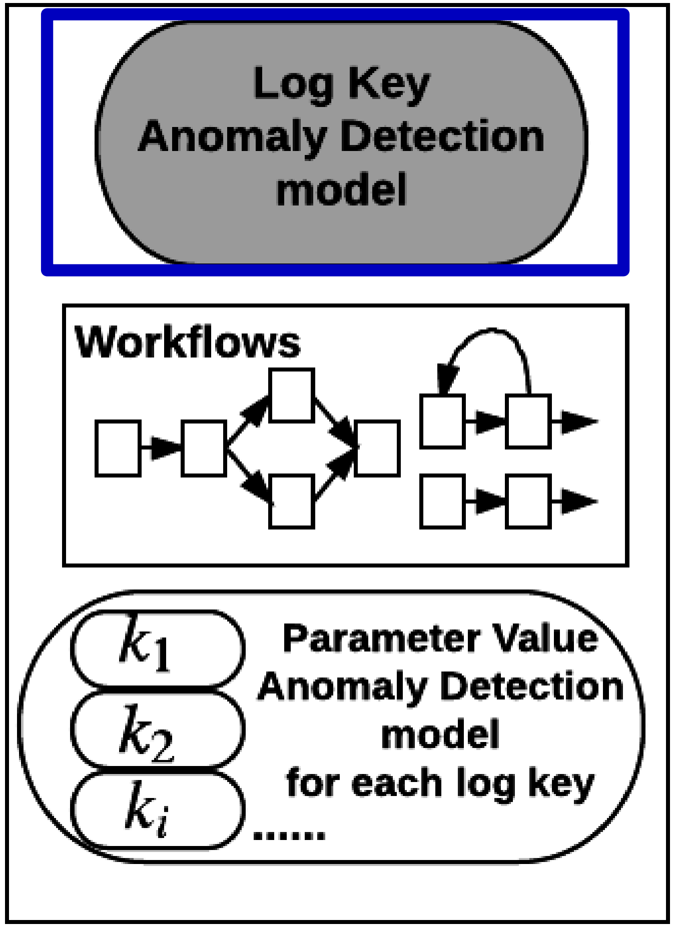


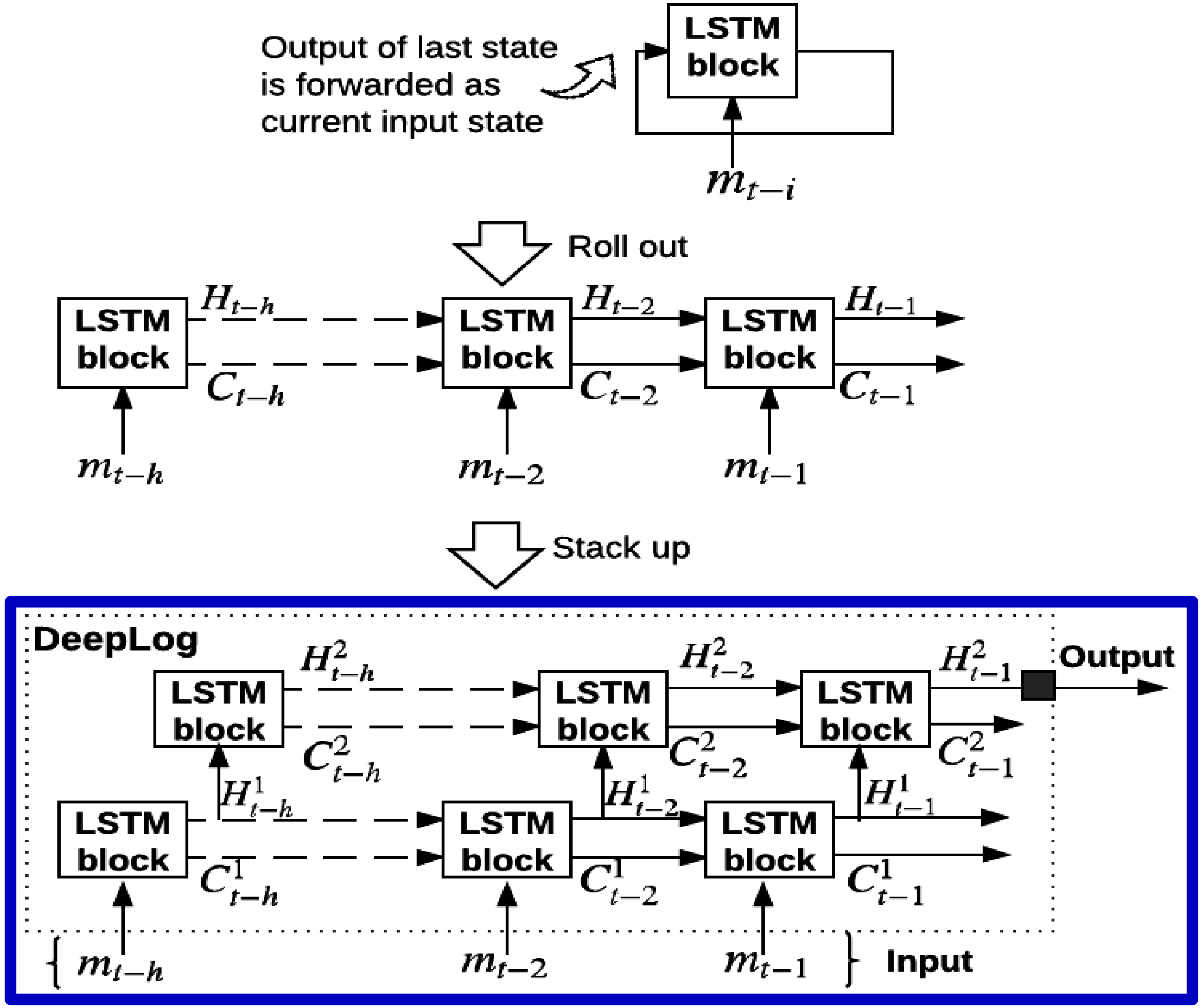
Log Key Anomaly Detection model


Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

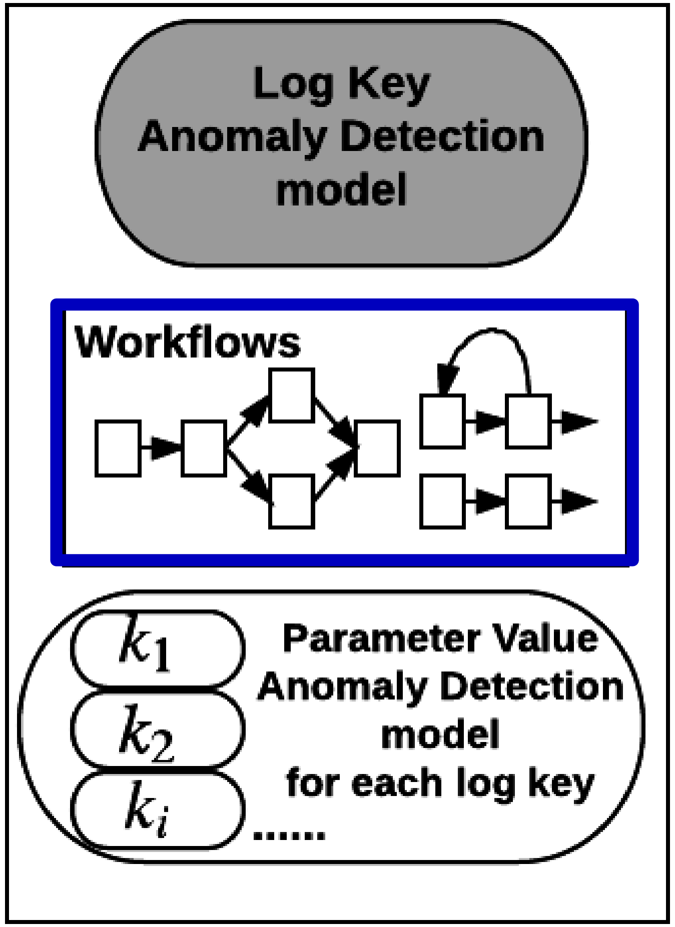


Log Key Anomaly Detection model





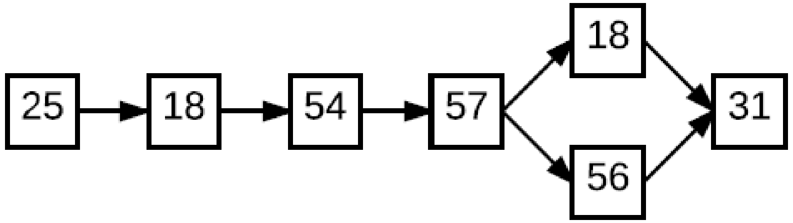
Workflow Construction



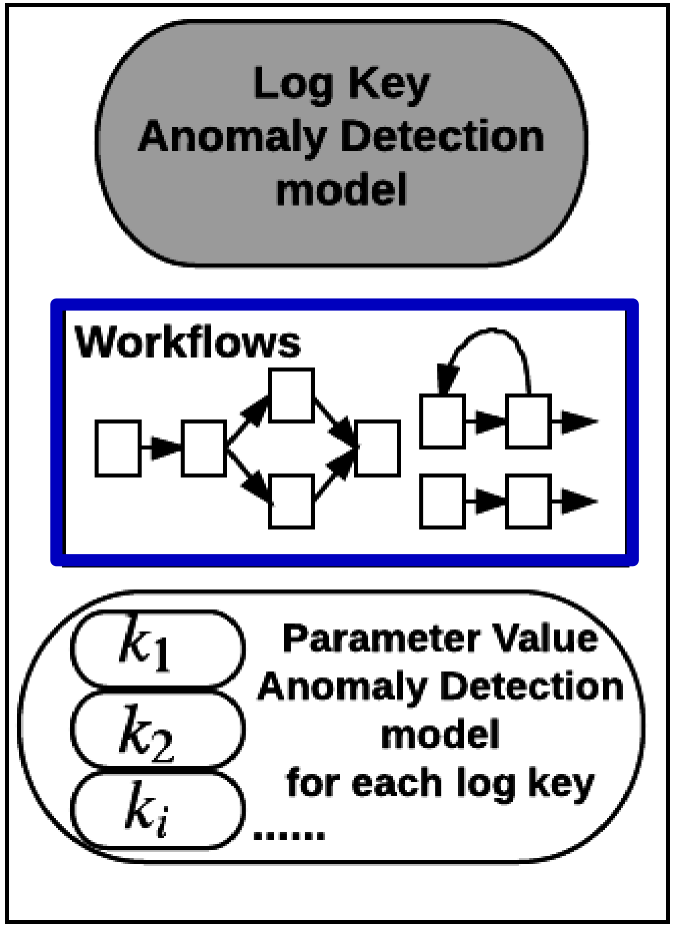
Input: log key sequence

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

Output:



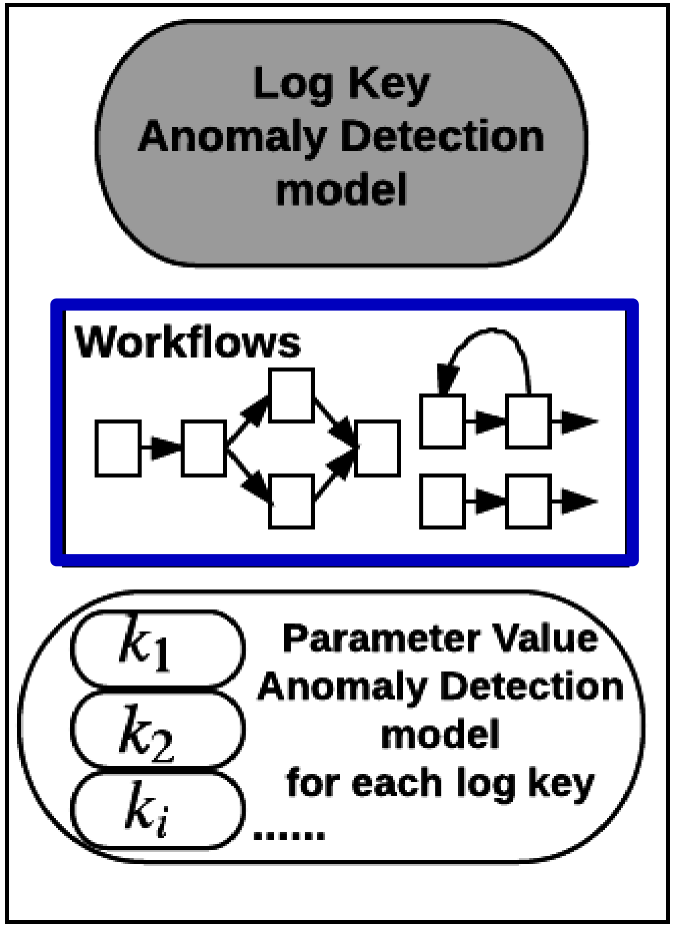
Workflow Construction



Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

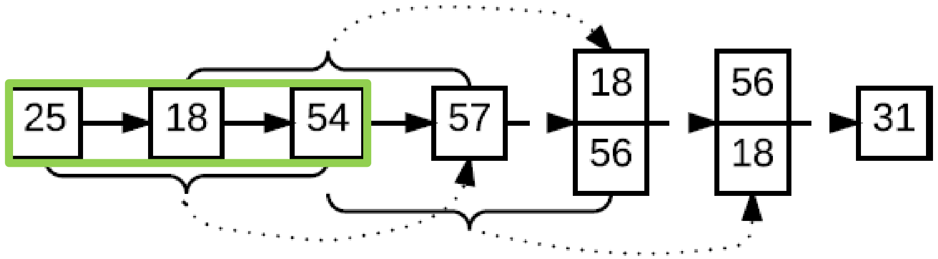
Workflow Construction



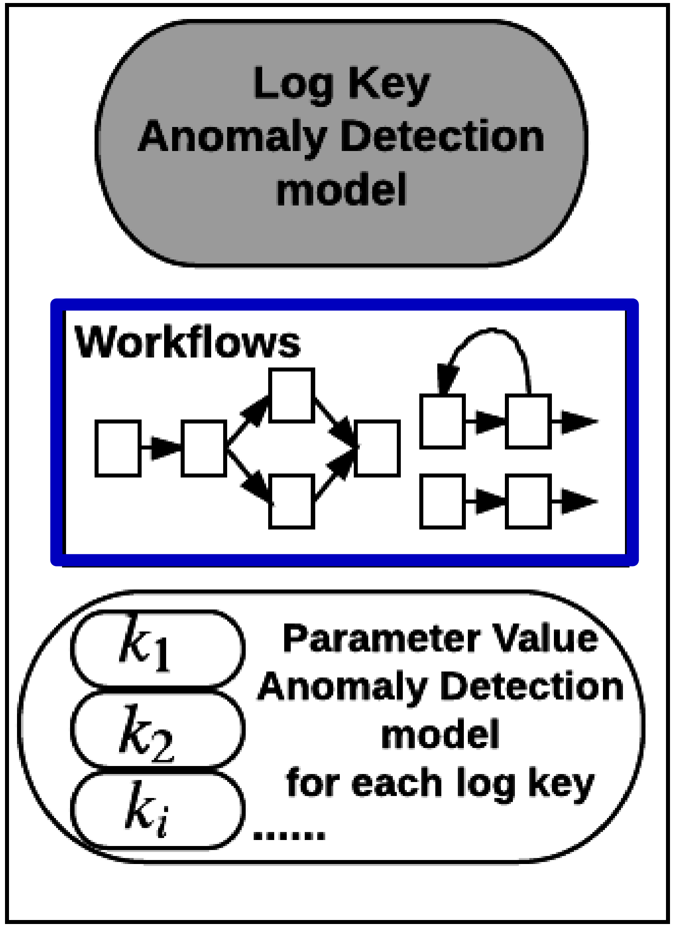
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



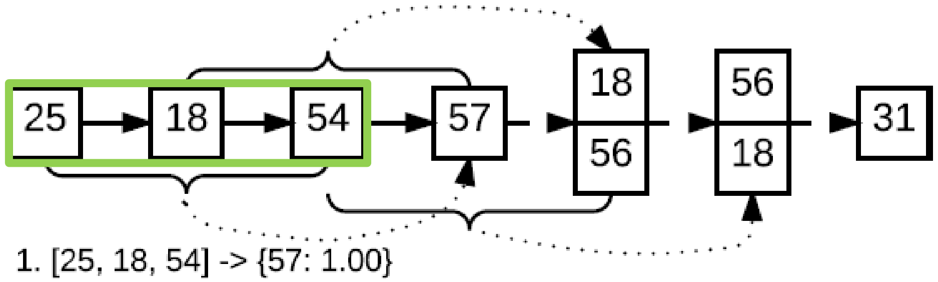
Workflow Construction



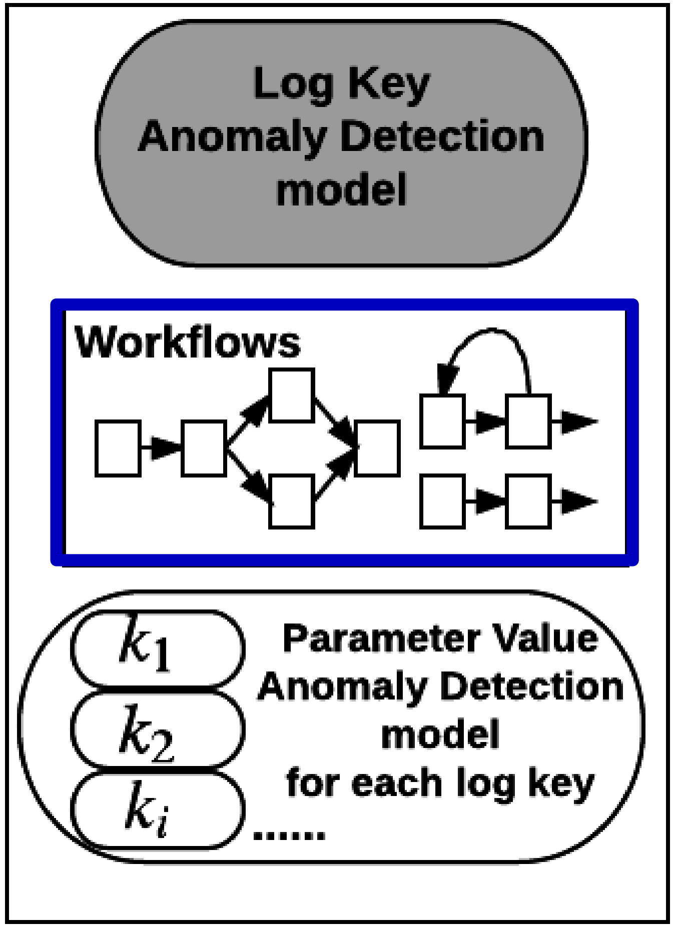
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



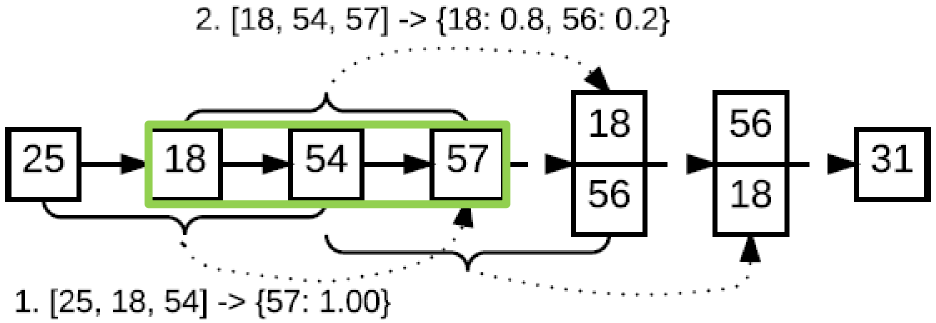
Workflow Construction



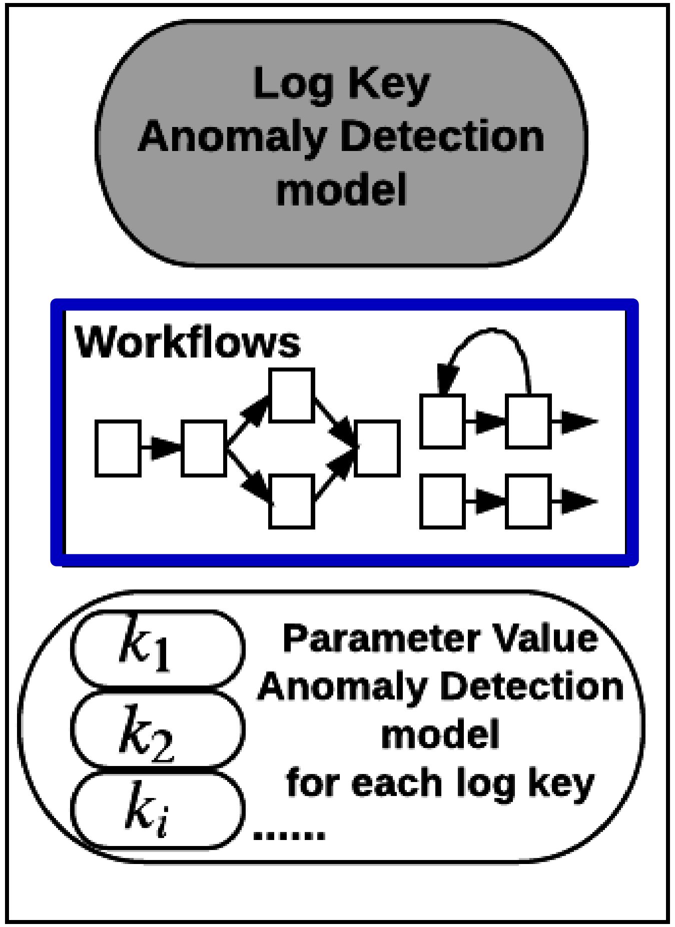
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



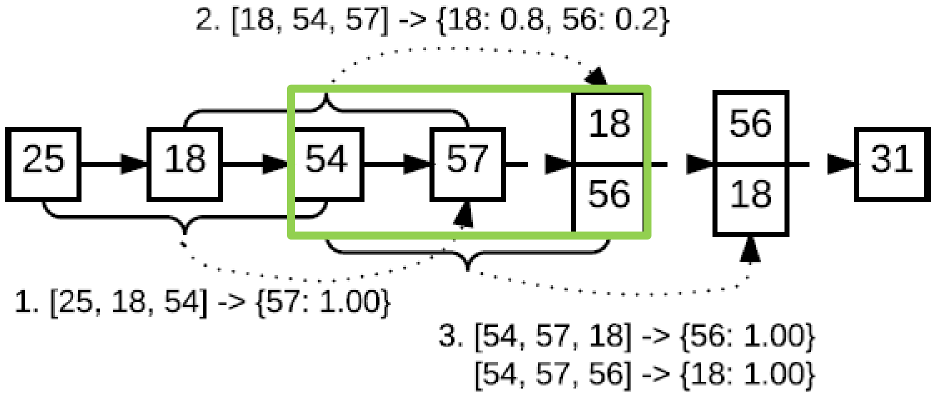
Workflow Construction



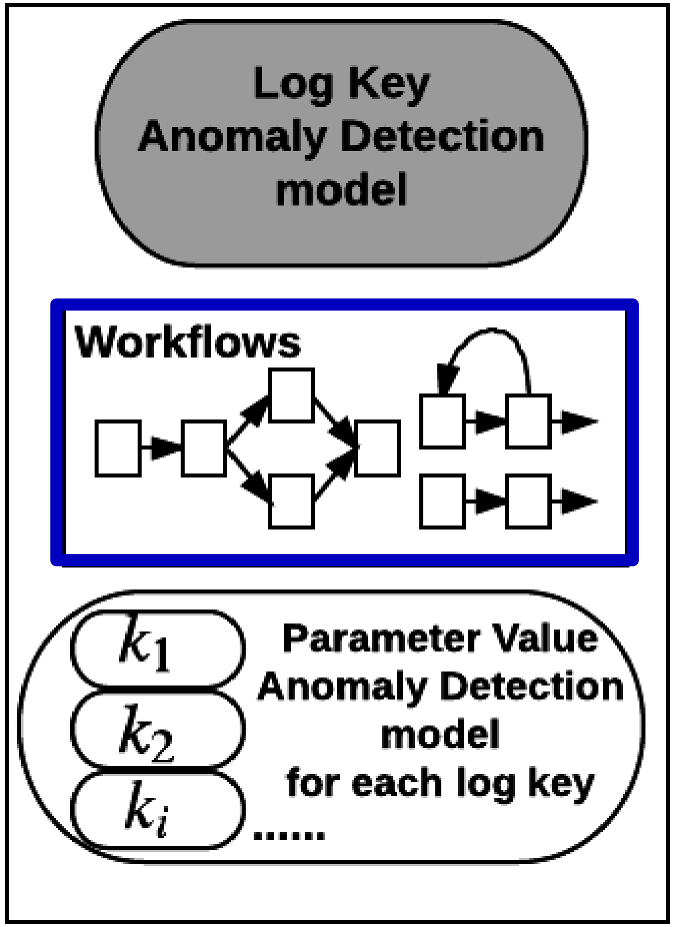
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:



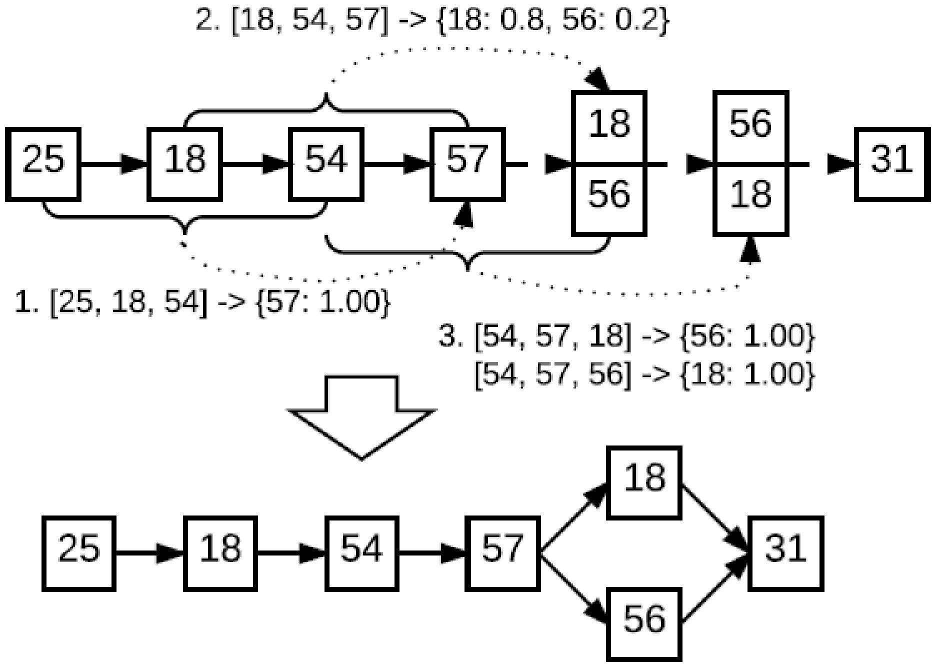
Workflow Construction



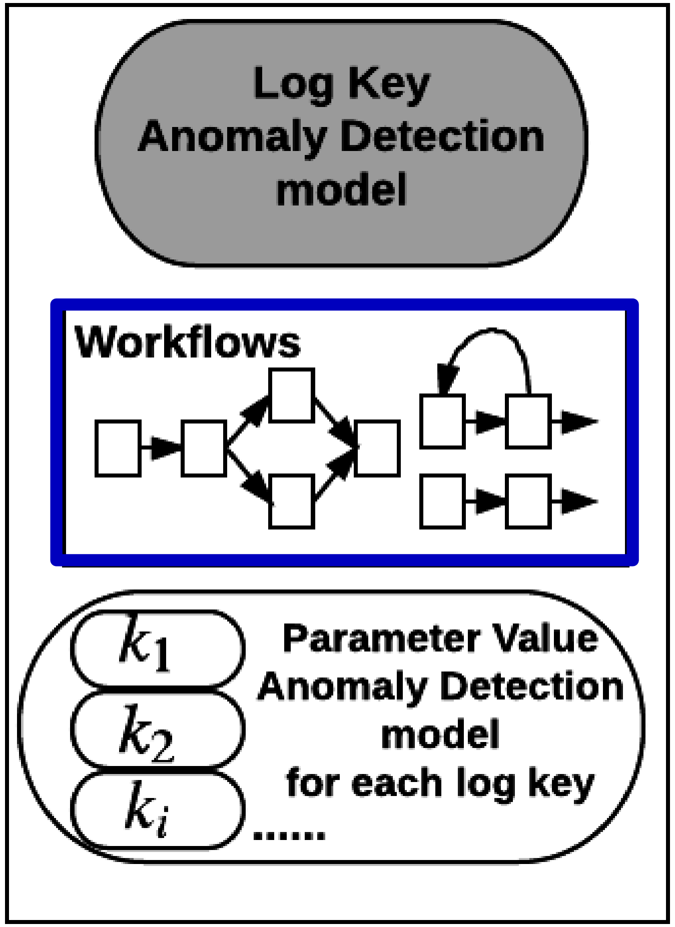
Method 1: Using Log Key Anomaly Detection model

--- LSTM prediction probabilities

An example of concurrency detection:

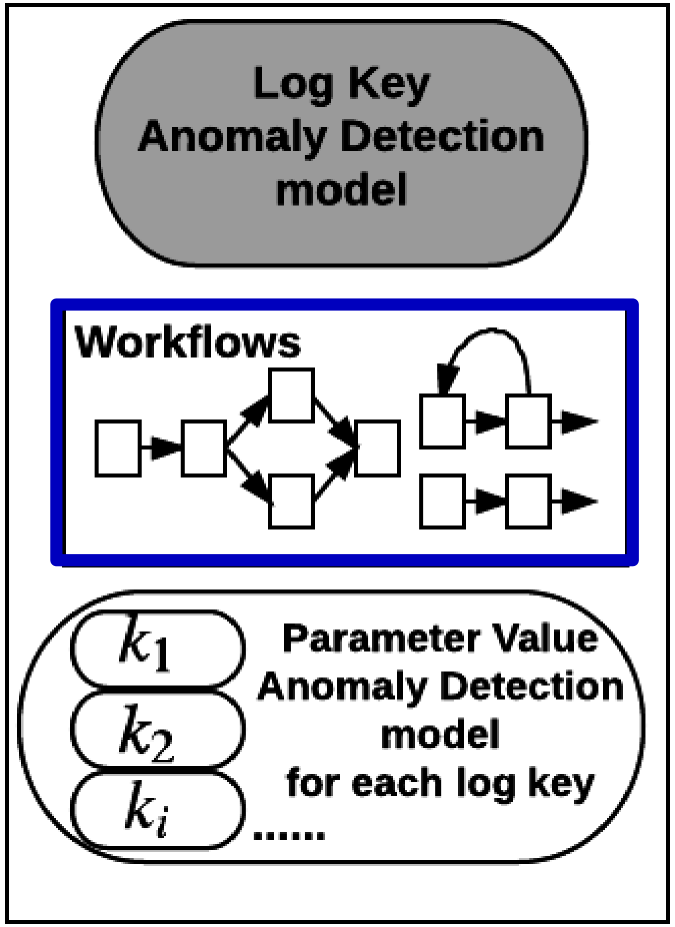


Workflow Construction



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

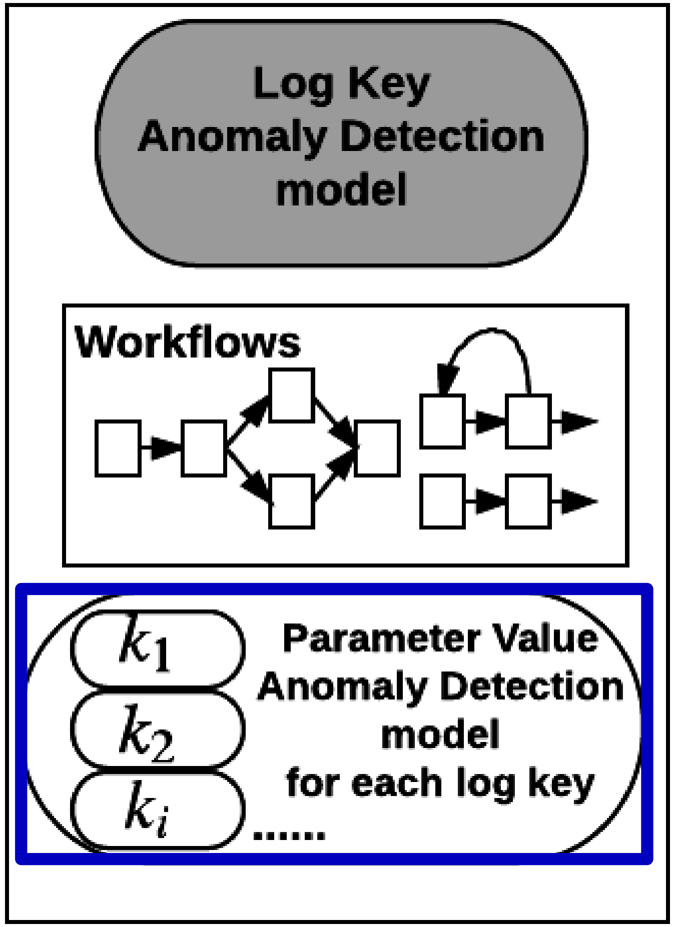
	k_1	…	k_j	…	k_n
k_1	$p_d(1, 1)$		$p_d(1, j)$		
…					
k_i	$p_d(i, 1)$		$p_d(i, j) = \frac{f_d(k_i, k_j)}{d \cdot f(k_i)}$		
…					
k_n	$p_d(n, 1)$		$p_d(n, j)$		

$f_d(k_i, k_j)$: 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:

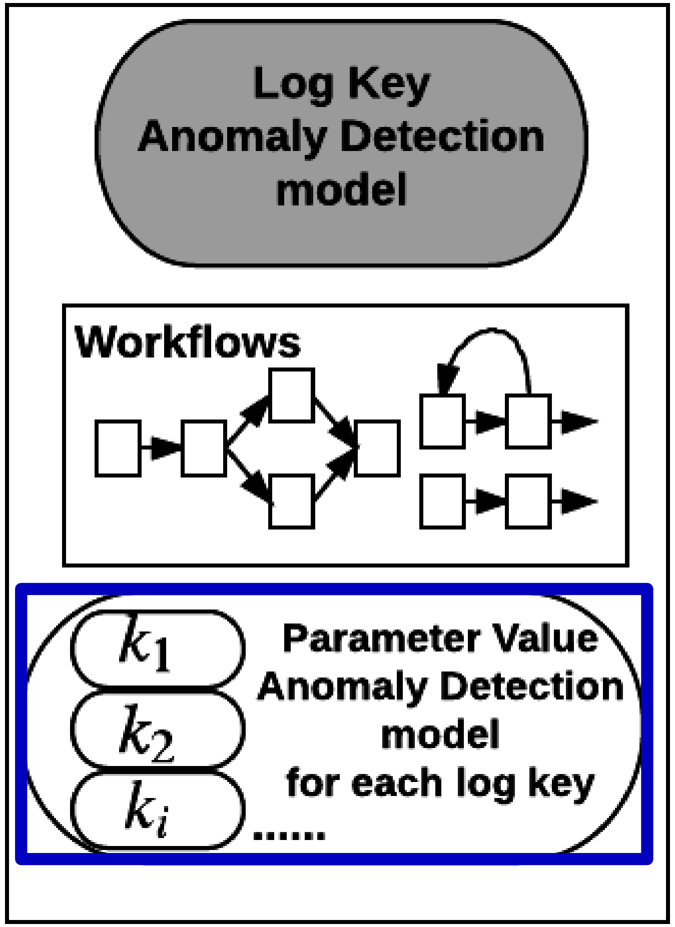
Log messages of a particular log key:

$t_2\!: Took\ 0.61\ seconds\ to\ deallocate\ network$ …

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

….

Parameter Value Anomaly Detection model



Example:

Log messages of a particular log key:

$\boldsymbol{t_2}$: $\boldsymbol{Took\ 0.61\ seconds\ to\ deallocate\ network}$ …

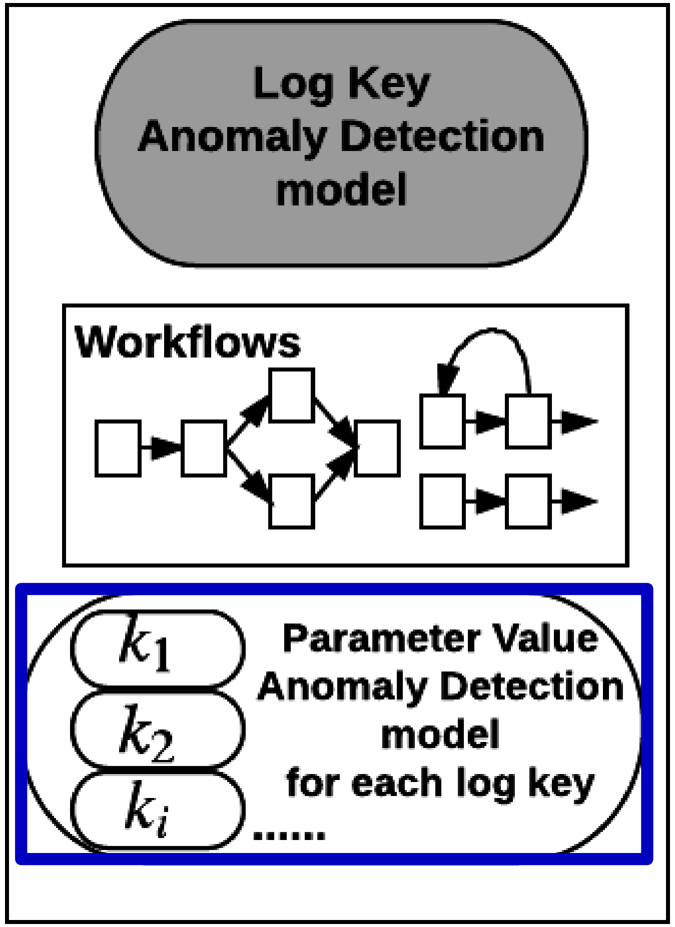
$\boldsymbol{t'_2}$: $\boldsymbol{Took\ 1.1\ seconds\ to\ deallocate\ network}$ …

….

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: $Took\ 1.1\ seconds\ to\ deallocate\ network$ …

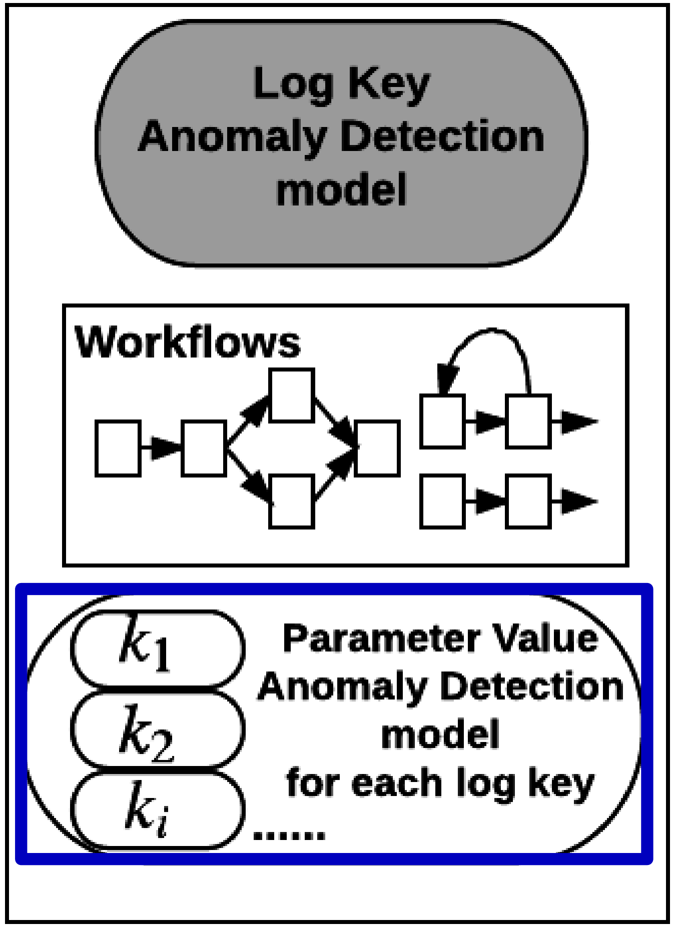
….

Parameter value vectors overtime:

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

Multi-variate time series data anomaly detection problem!

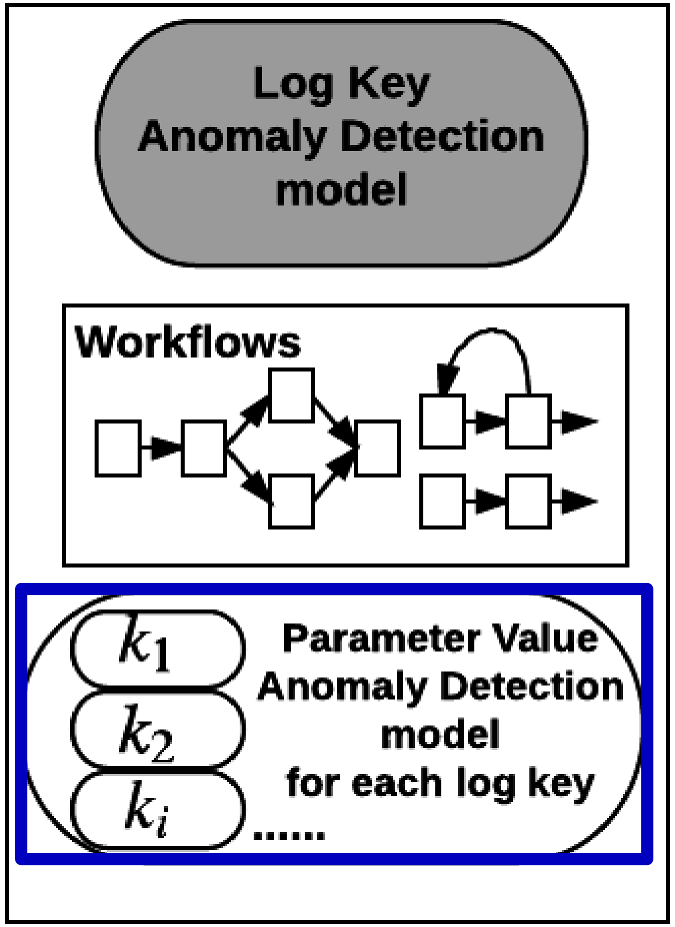
Parameter Value Anomaly Detection model



Multi-variate time series data anomaly detection problem

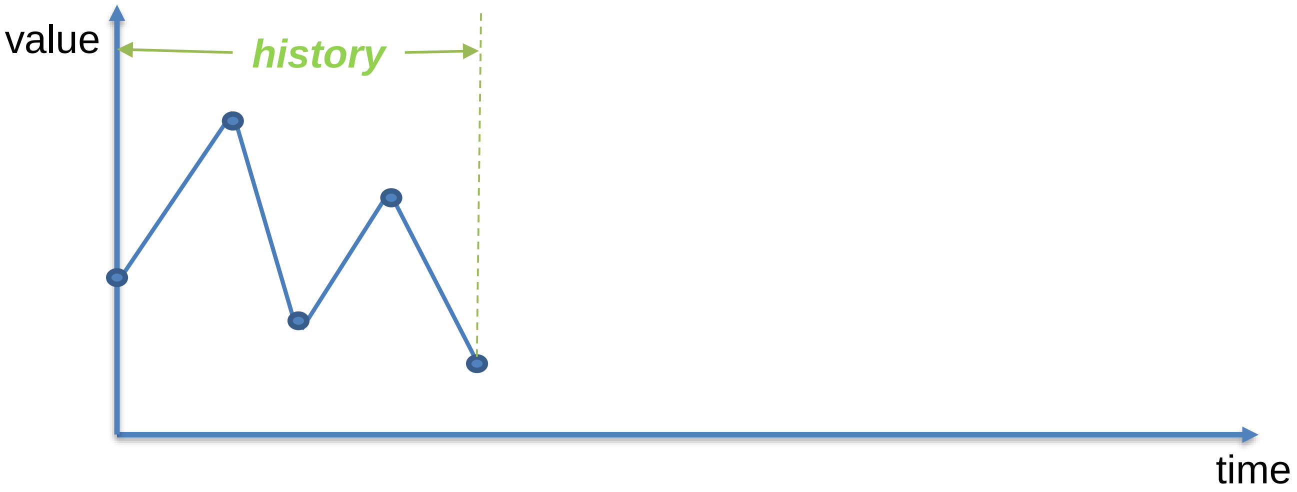
- ✓ Leverage LSTM-based approach;
- ✓ A parameter value vector is given as input at each time step;
- ✓ An anomaly is detected if the mean-square-error (MSE) between prediction and actual data is too big.

Parameter Value Anomaly Detection model

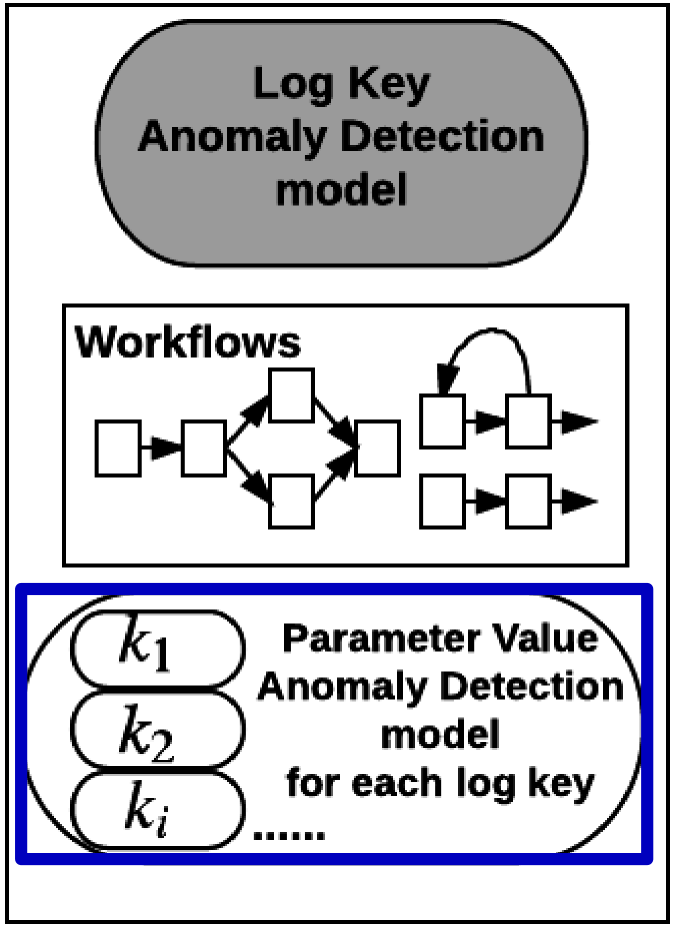


Multi-variate time series data anomaly detection problem

- ✓ 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.

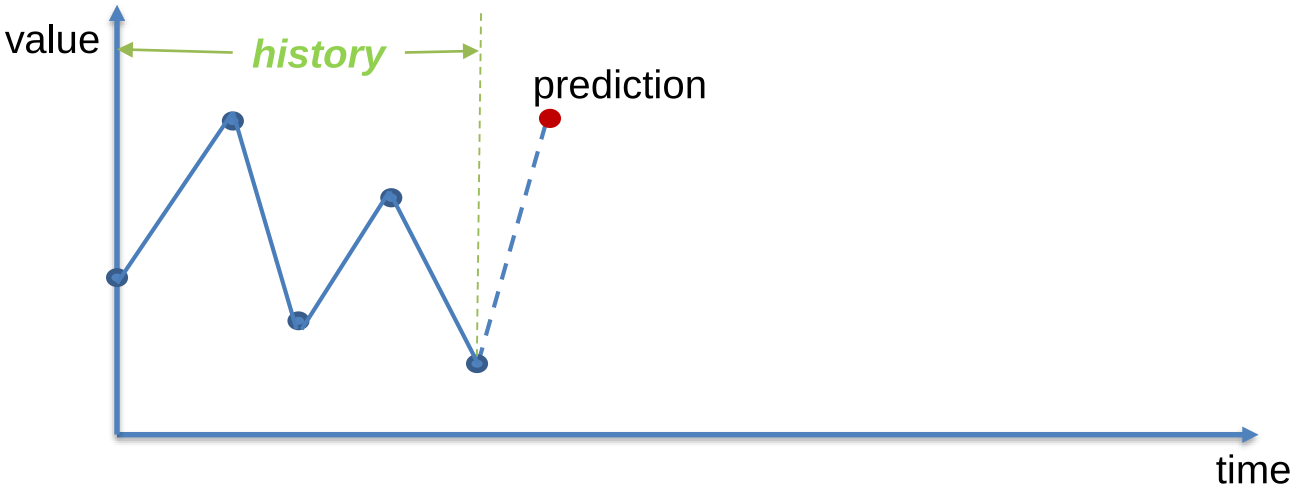


Parameter Value Anomaly Detection model

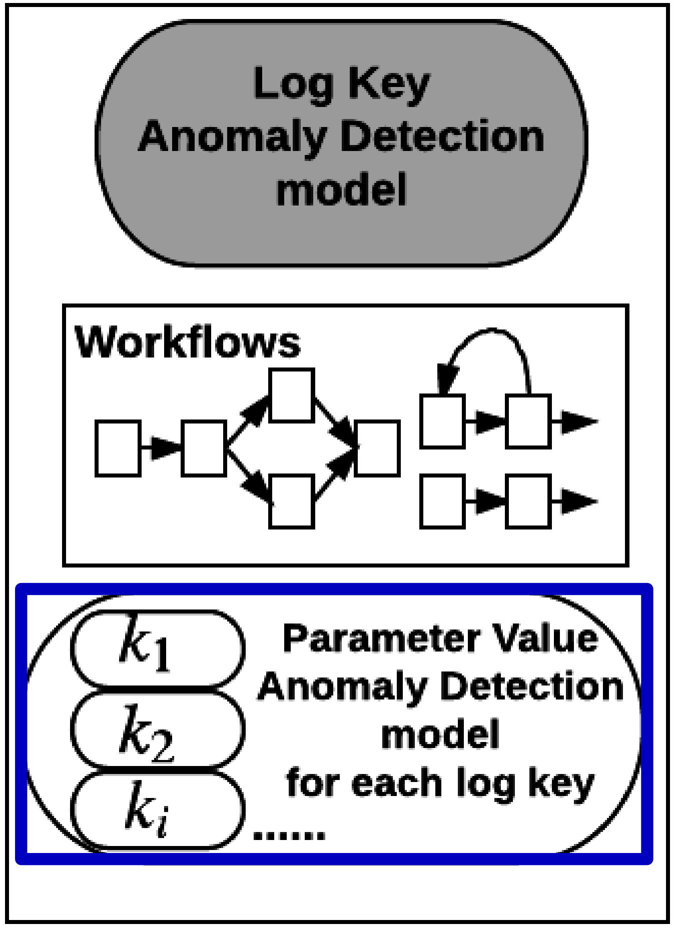

Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

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.

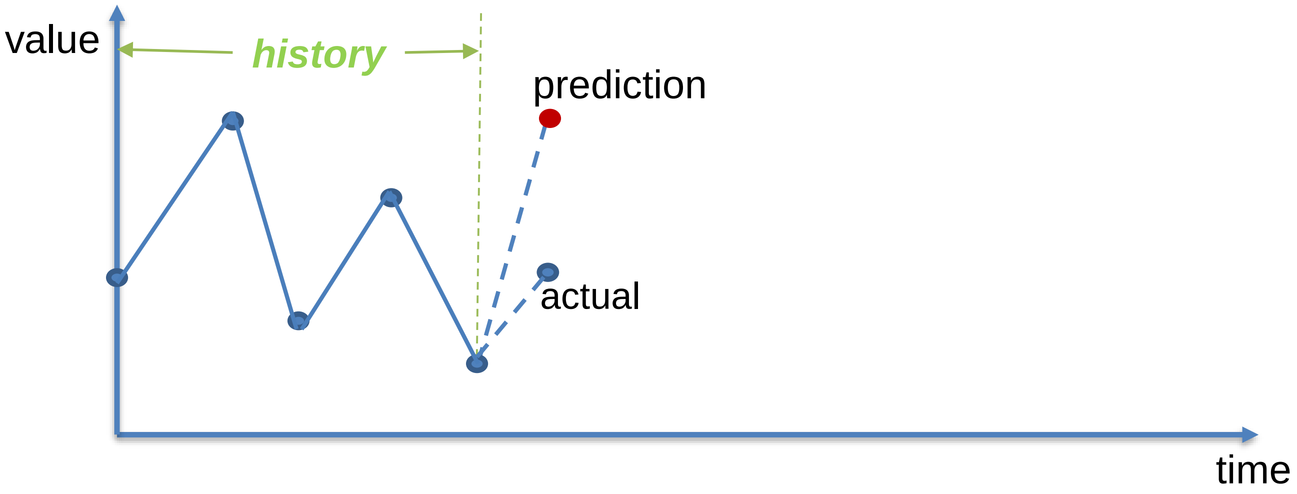


Parameter Value Anomaly Detection model

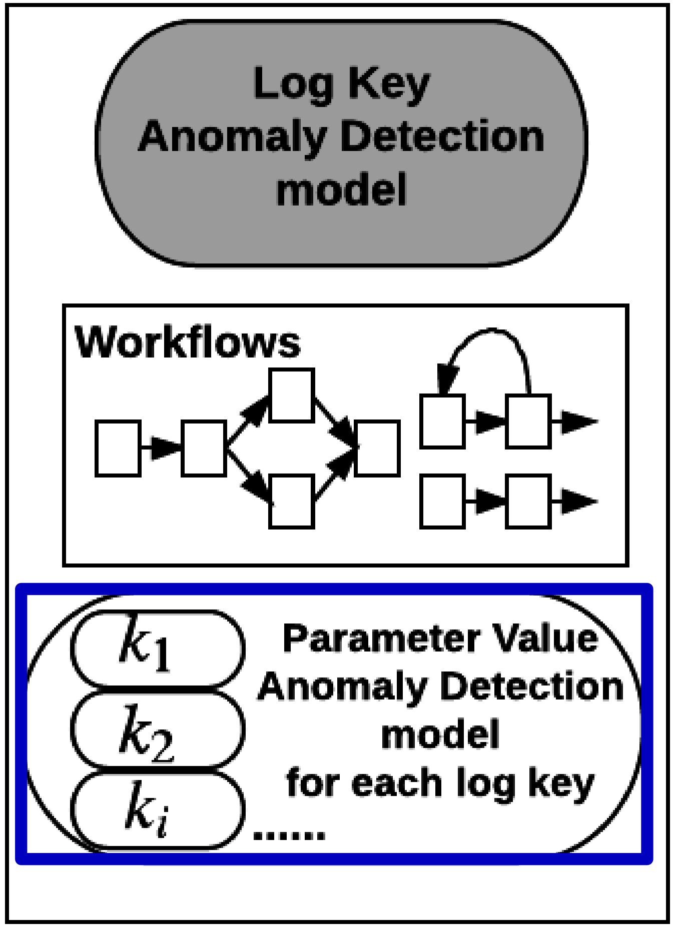


Multi-variate time series data anomaly detection problem

- ✓ 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.

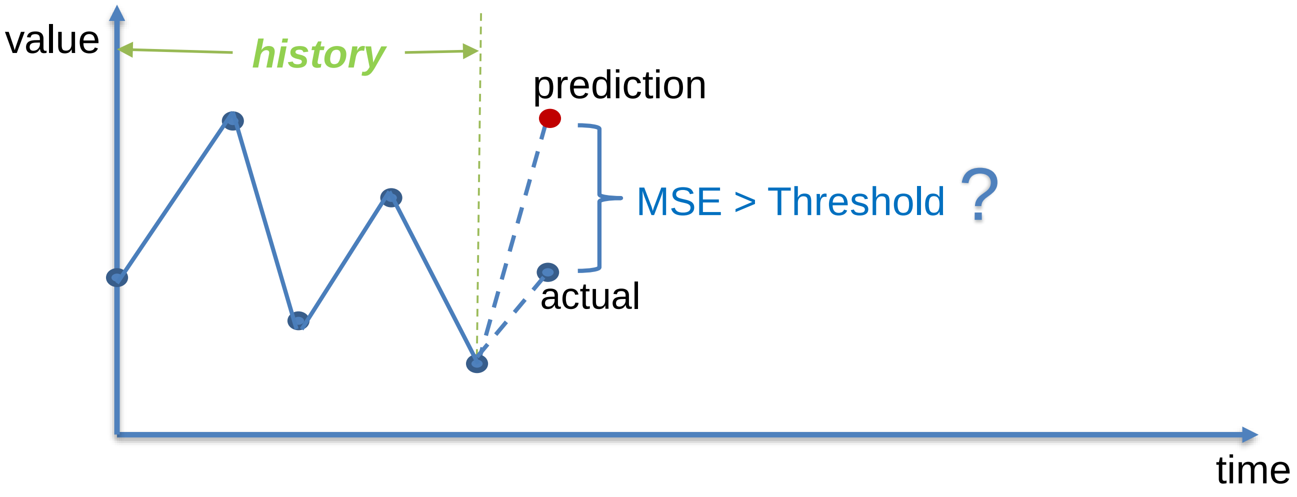


Parameter Value Anomaly Detection model

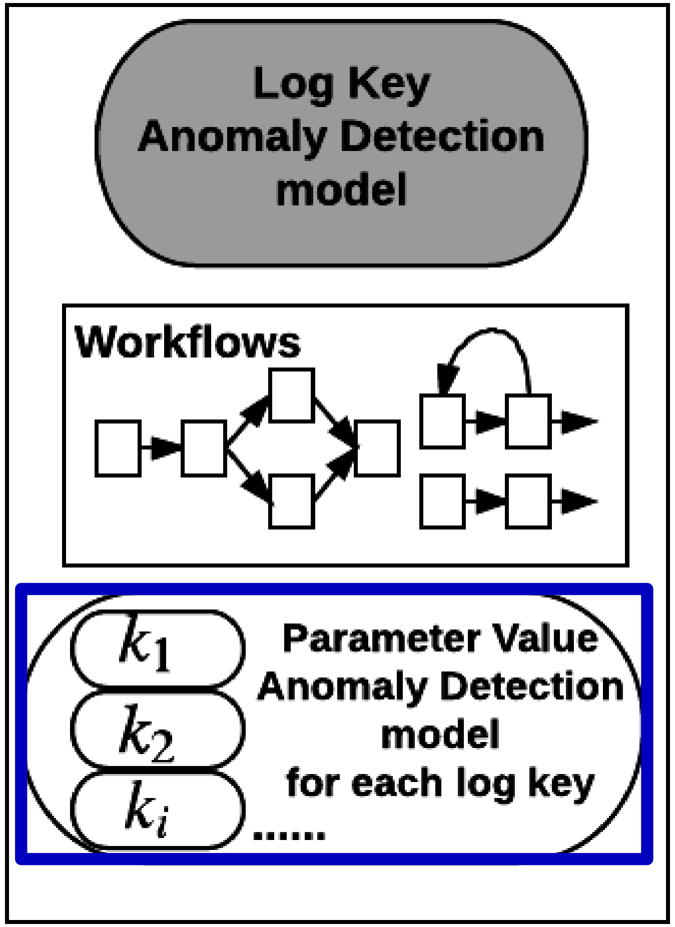


Multi-variate time series data anomaly detection problem

- ✓ 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.

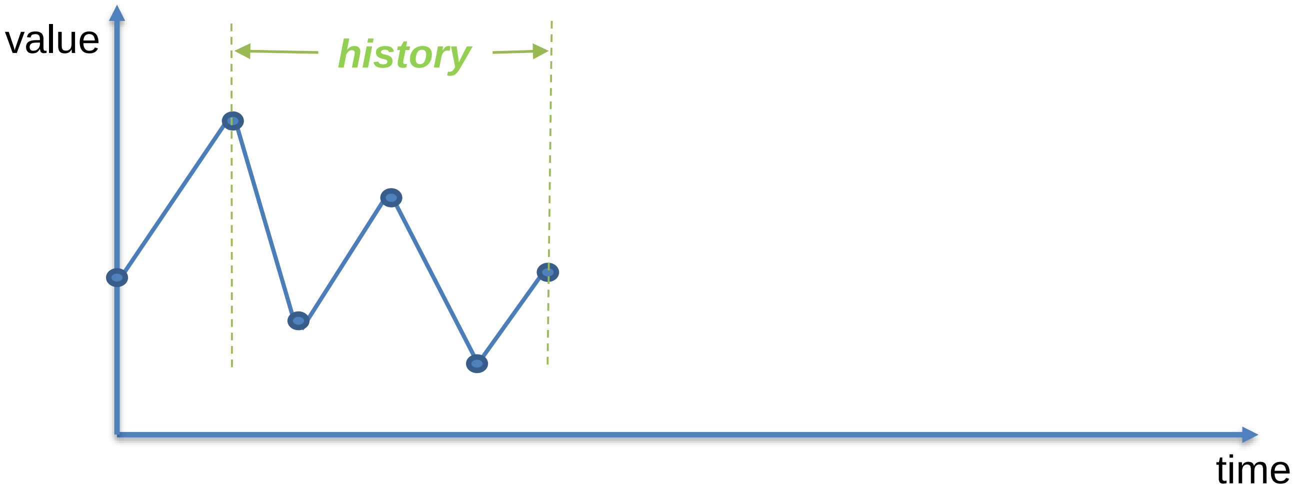


Parameter Value Anomaly Detection model

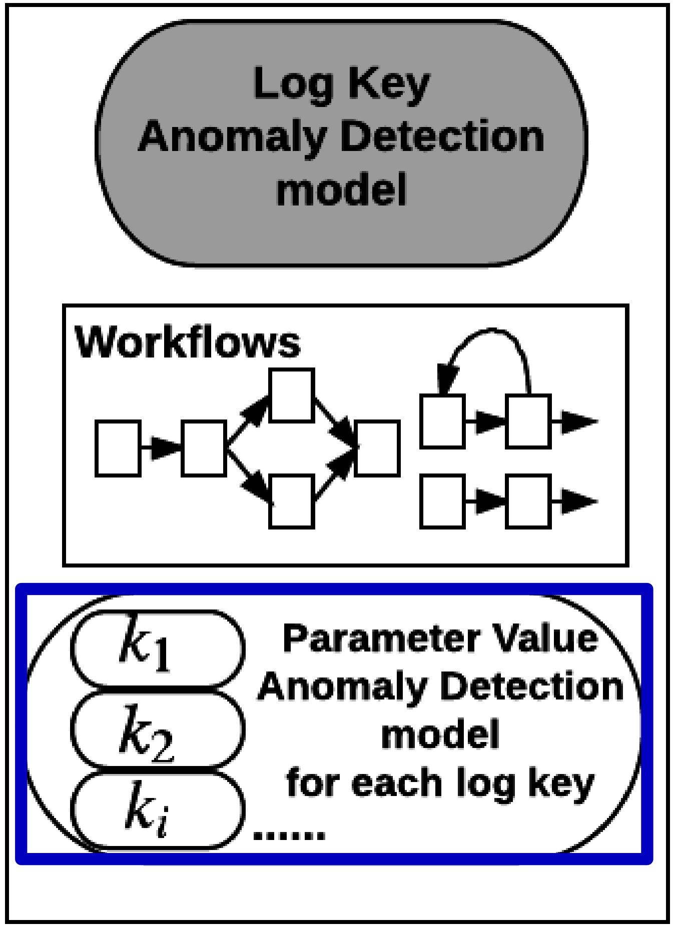


Multi-variate time series data anomaly detection problem

- ✓ 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.

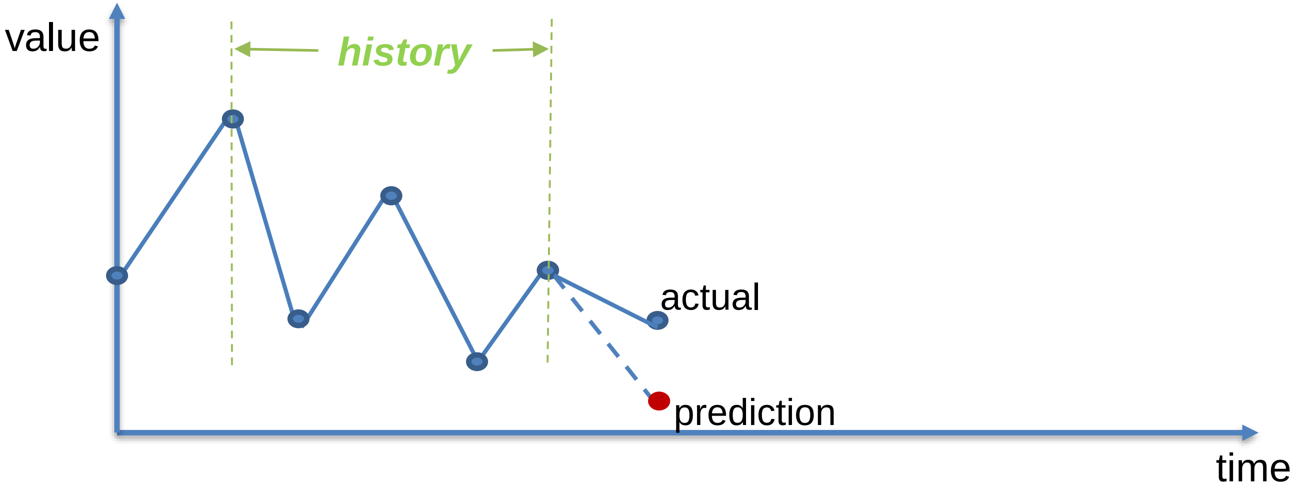


Parameter Value Anomaly Detection model

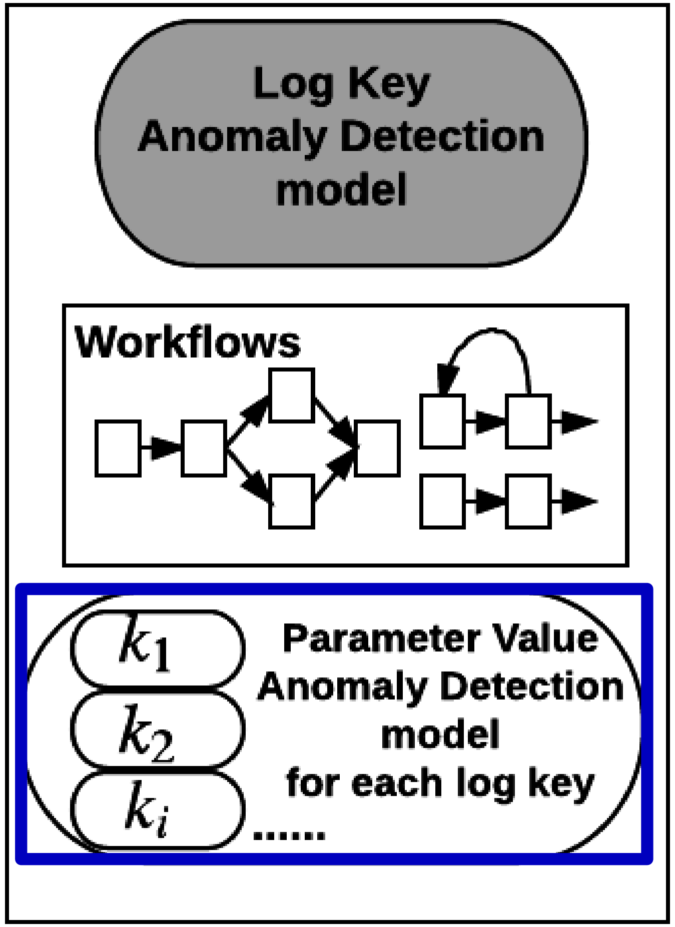


Multi-variate time series data anomaly detection problem

- ✓ 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.



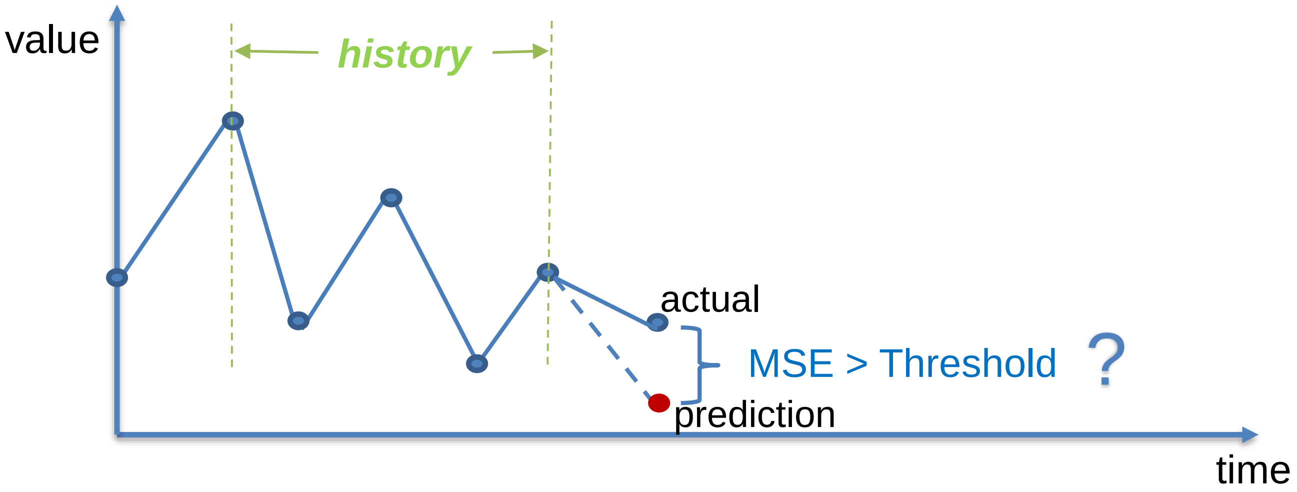
Parameter Value Anomaly Detection model



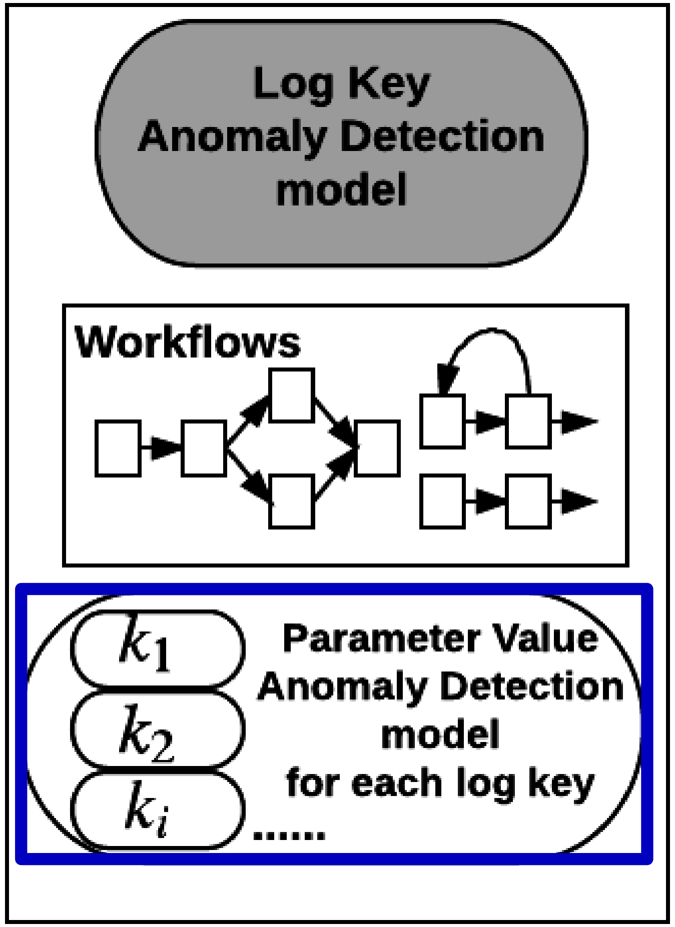
Log Key Anomaly Detection model
Workflows
k_1
k_2
k_i
Parameter Value Anomaly Detection model for each log key
......

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.

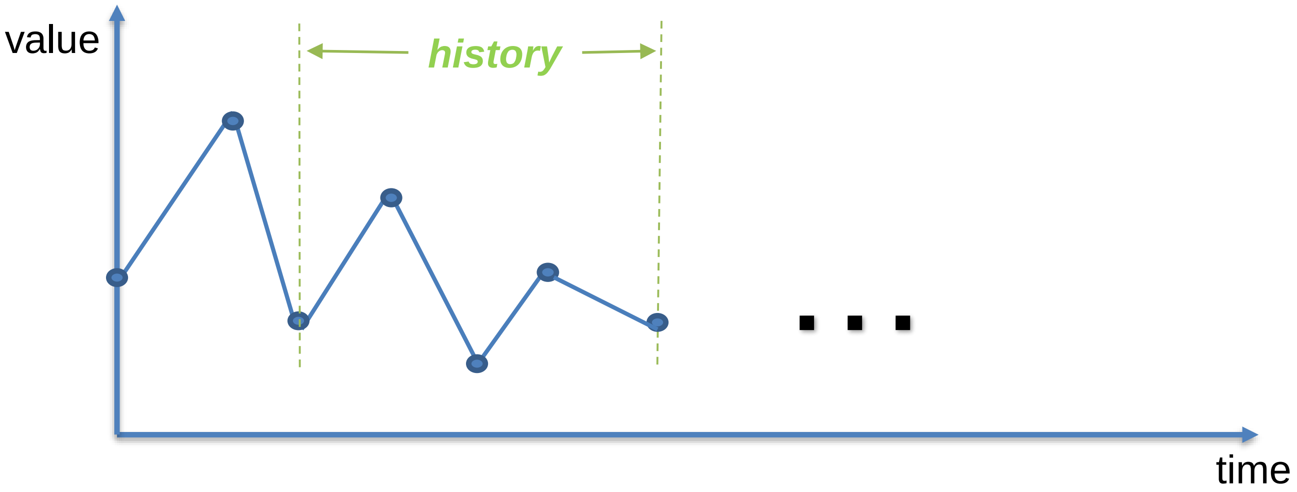


Parameter Value Anomaly Detection model



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?

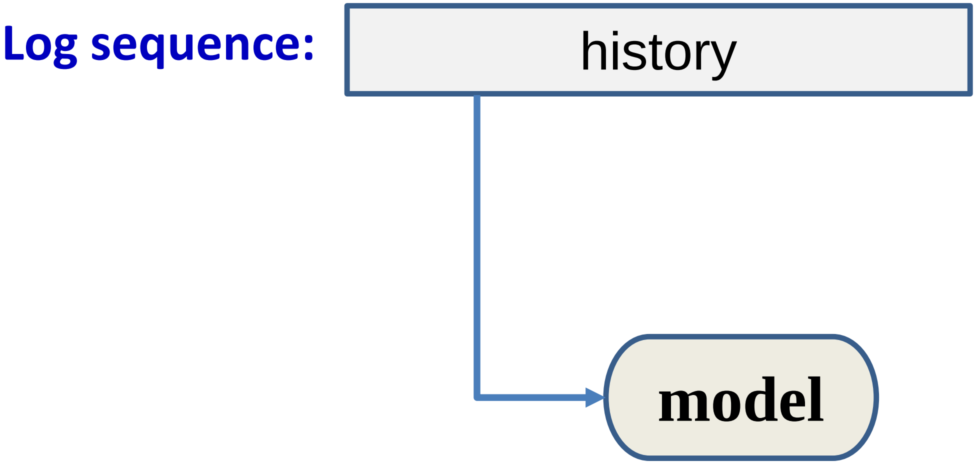
LSTM model online update

Q: How to handle false positive?

Log sequence: history

LSTM model online update

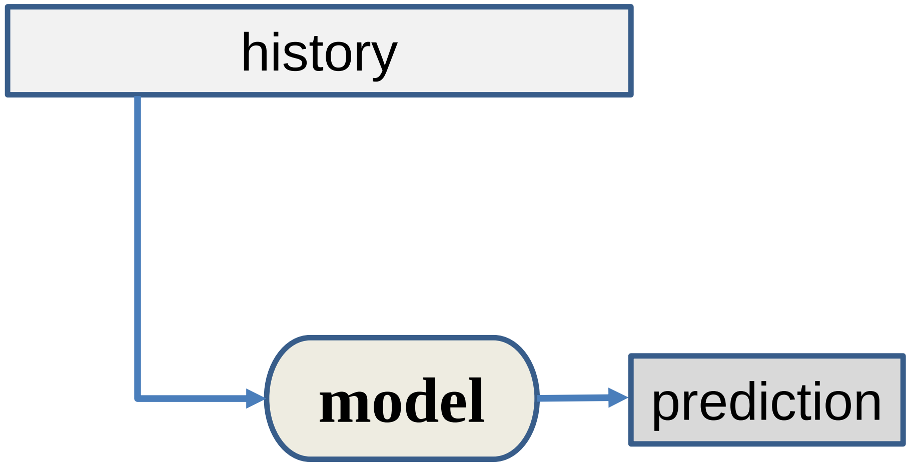
Q: How to handle false positive?



LSTM model online update

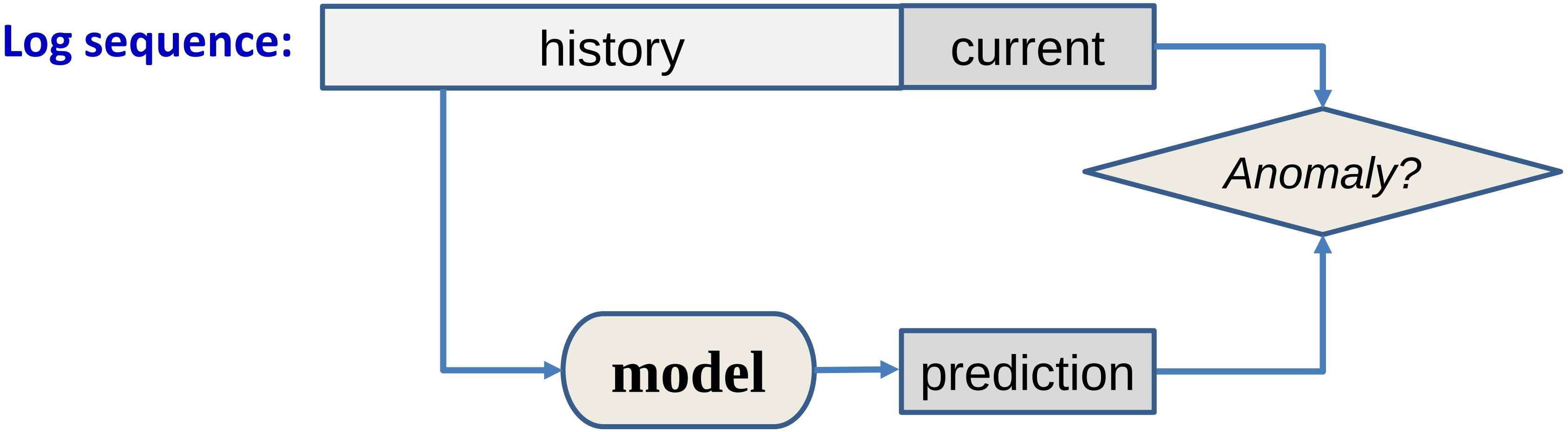
Q: How to handle false positive?

Log sequence:



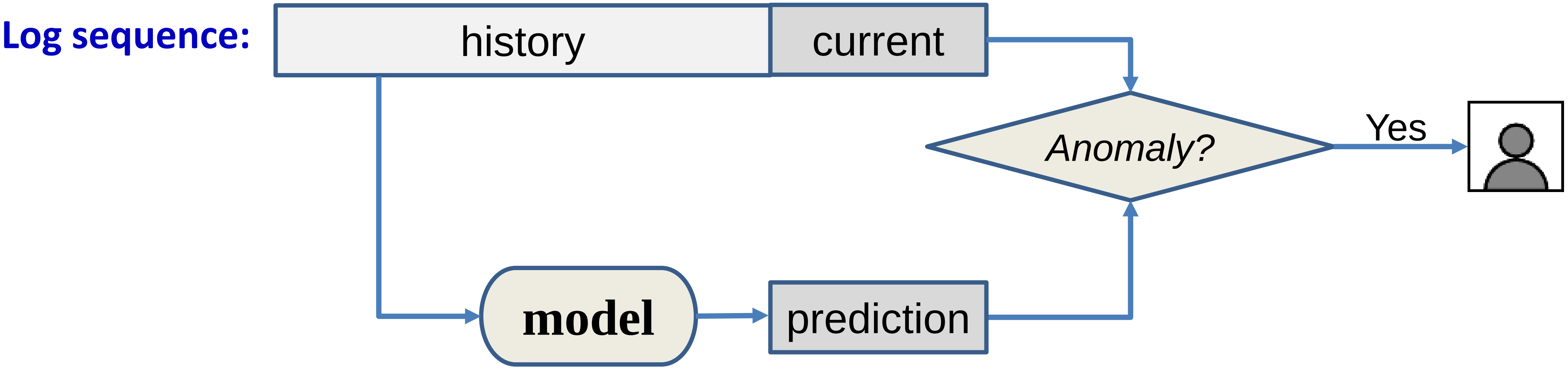
LSTM model online update

Q: How to handle false positive?



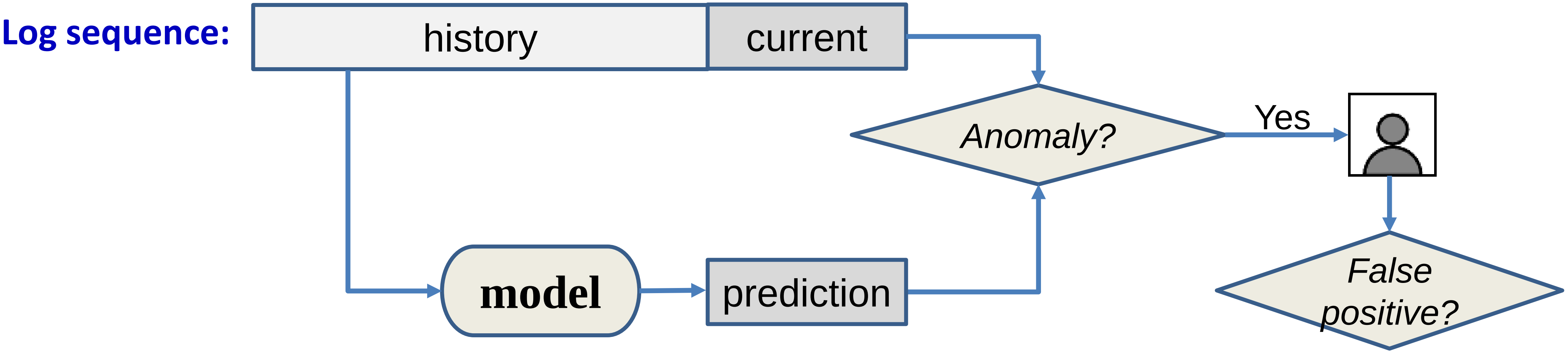
LSTM model online update

Q: How to handle false positive?



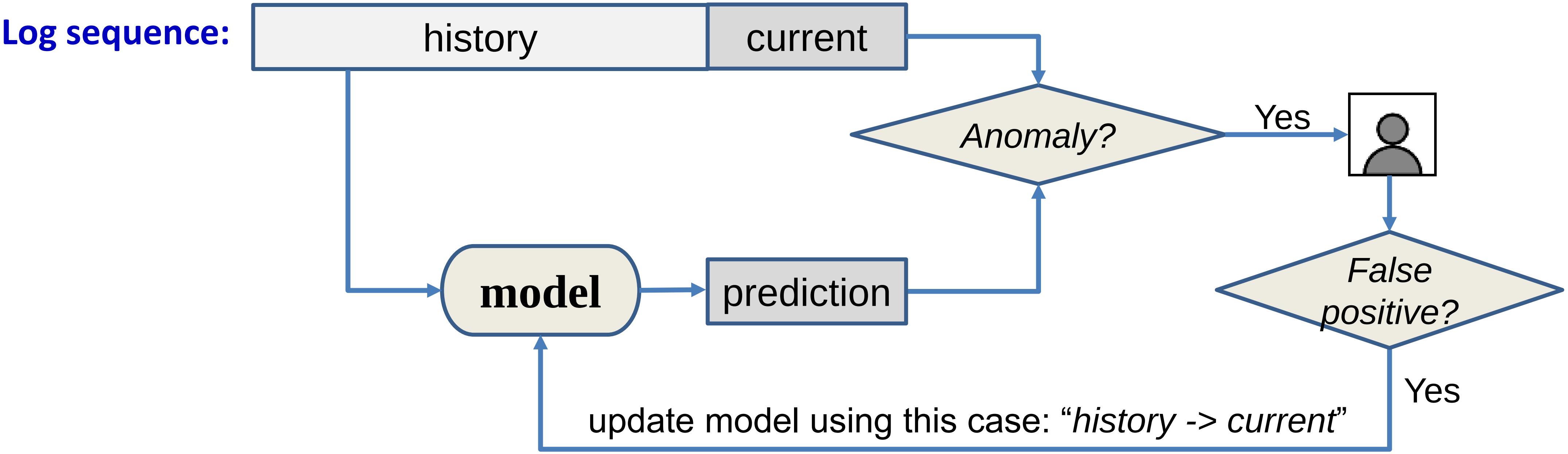
LSTM model online update

Q: How to handle false positive?

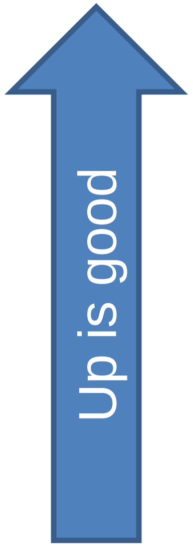


LSTM model online update

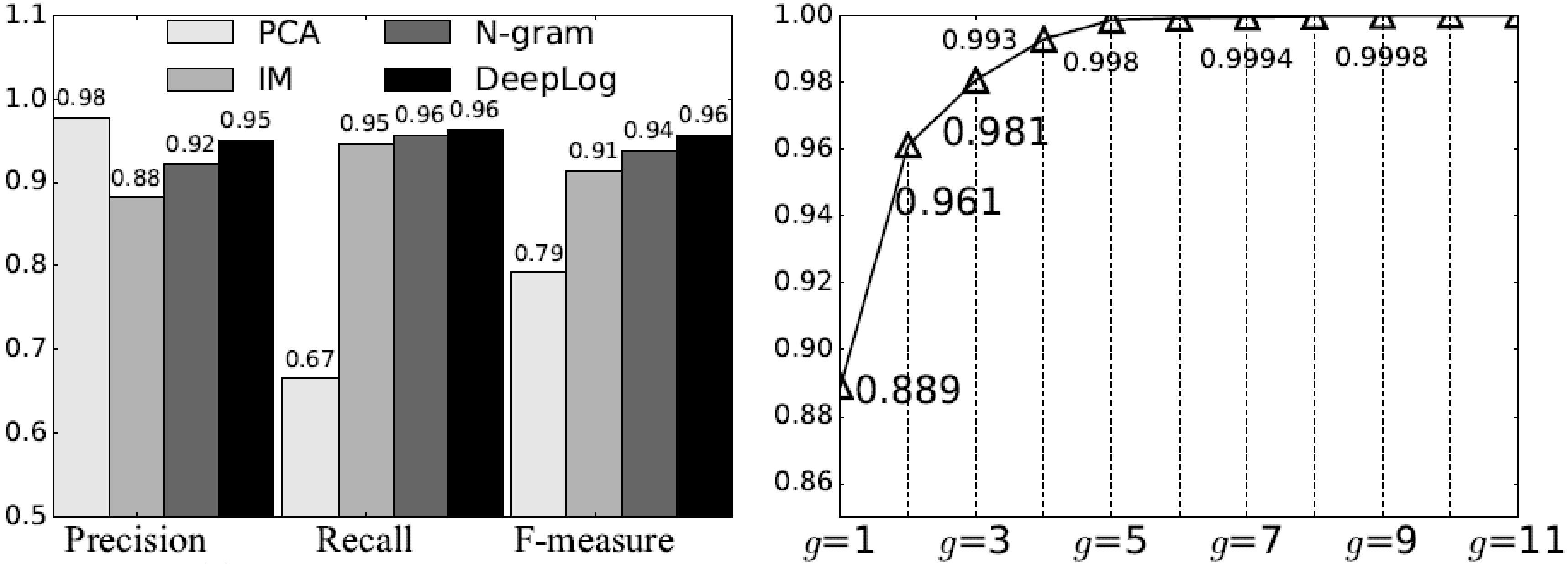
Q: How to handle false positive?



Evaluation – log key anomaly detection



Up is good



(a) Accuracy on HDFS.

(b) Cumulative probability of top g predictions.

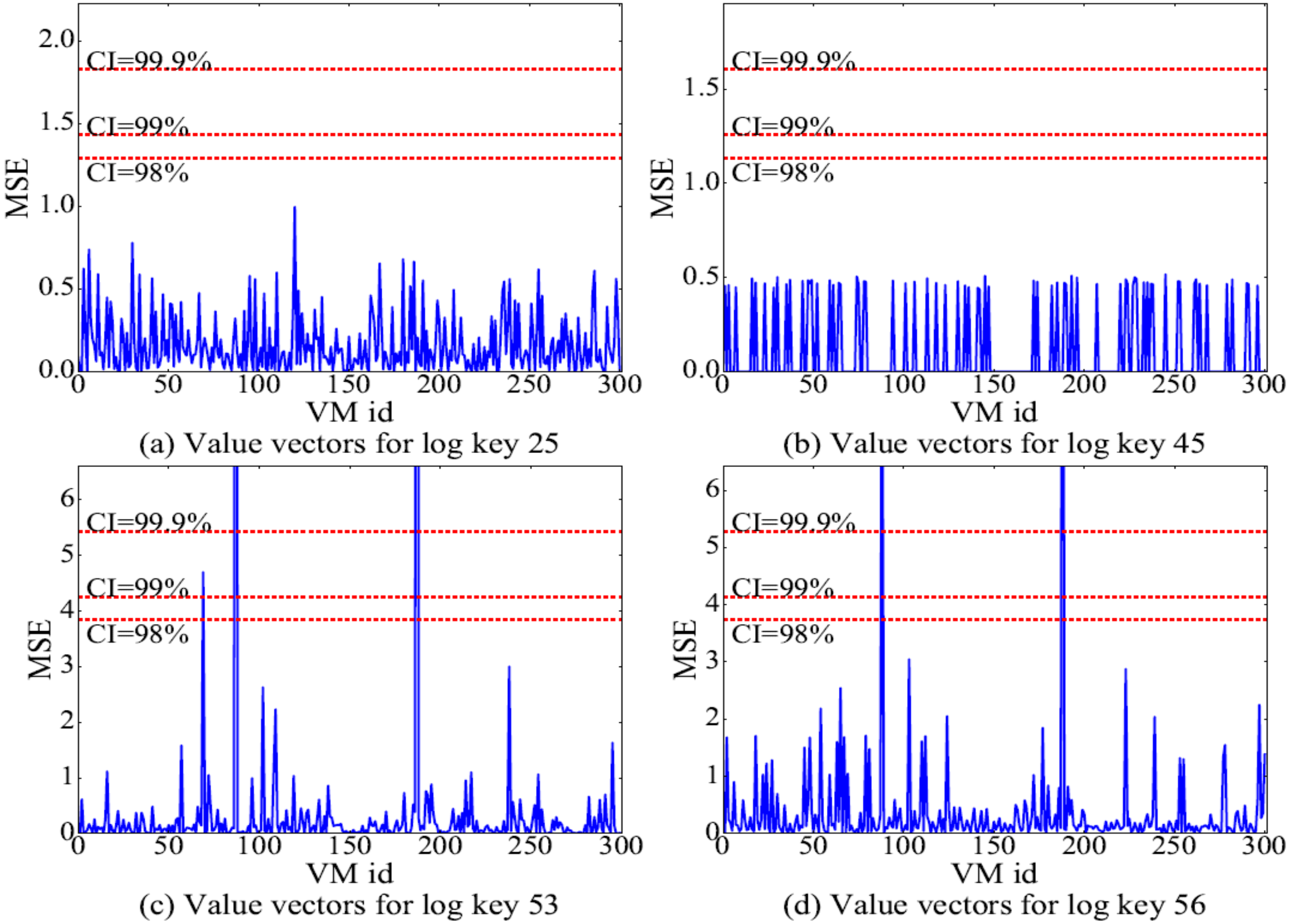
Evaluation results on HDFS log data [1].

(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



(a) Value vectors for log key 25

(b) Value vectors for log key 45

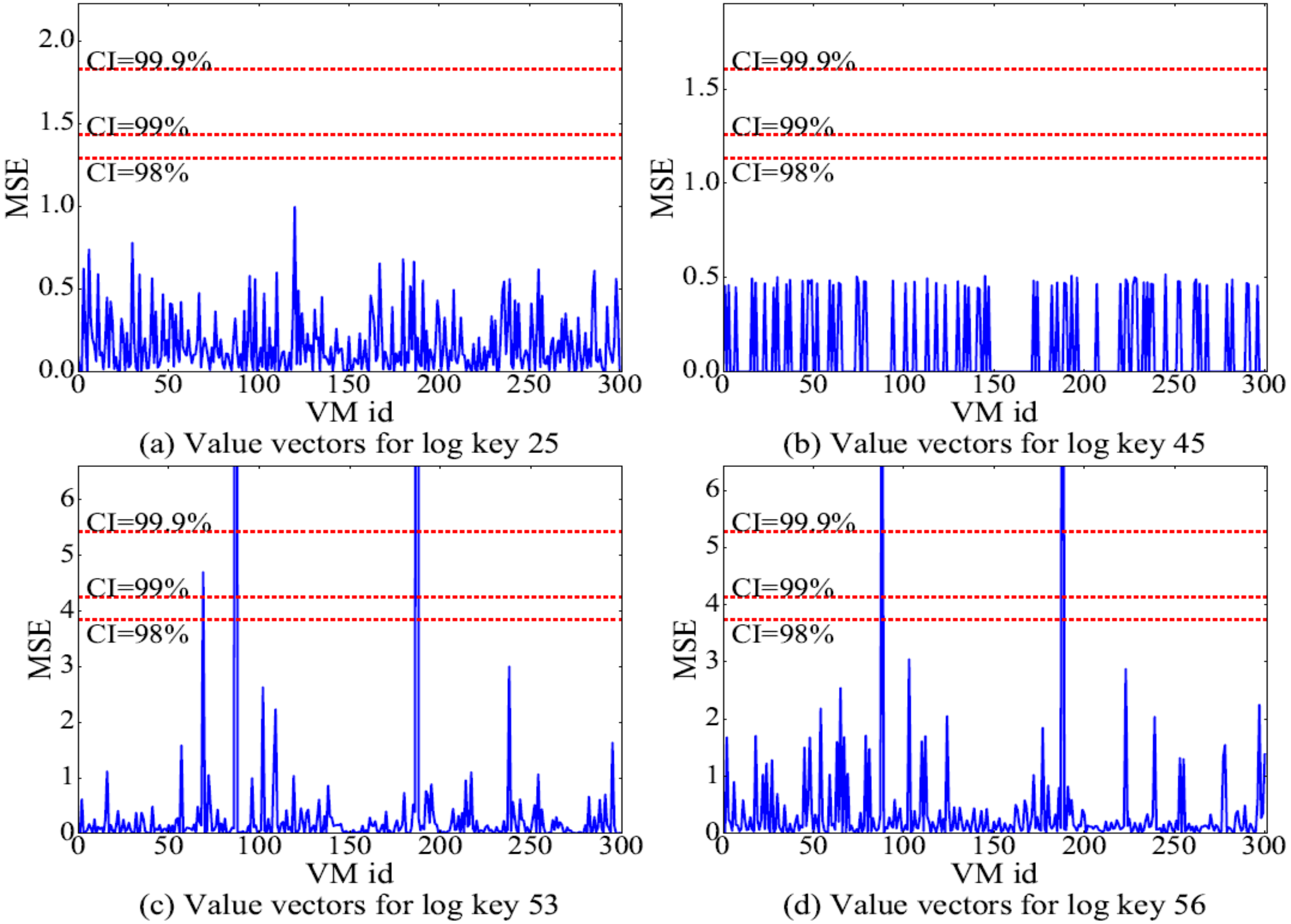
(c) Value vectors for log key 53

(d) Value vectors for log key 56

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

Evaluation – parameter value anomaly detection

MSE:
mean square error



(a) Value vectors for log key 25

(b) Value vectors for log key 45

(c) Value vectors for log key 53

(d) Value vectors for log key 56

Evaluation results on <u>OpenStack cloud log</u> with different confidence intervals (CIs)

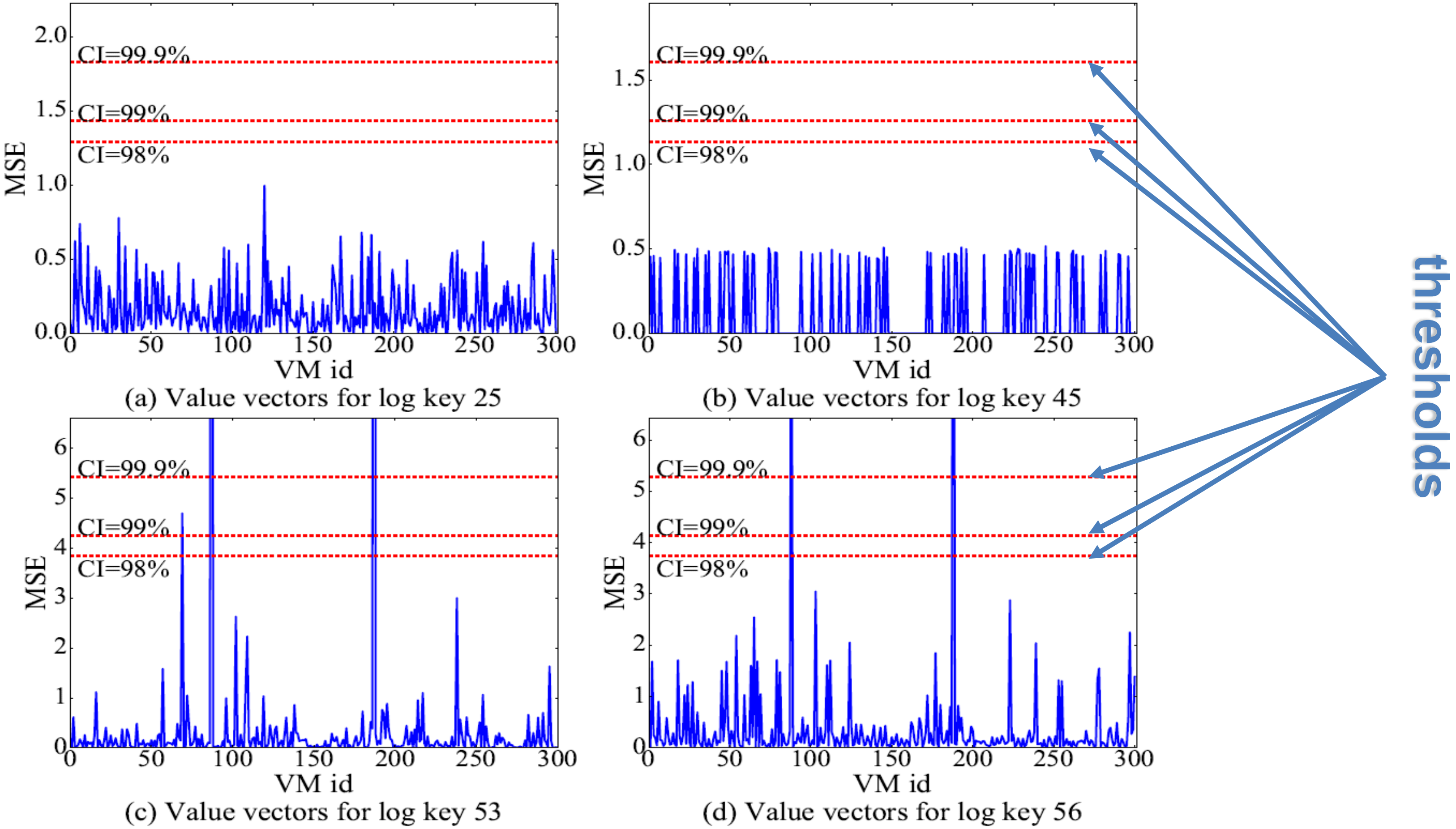
generated on CloudLab;

VM creation/deletion operations;

injected performance anomalies.

Evaluation – parameter value anomaly detection

MSE:
mean square error



(a) Value vectors for log key 25

(b) Value vectors for log key 45

(c) Value vectors for log key 53

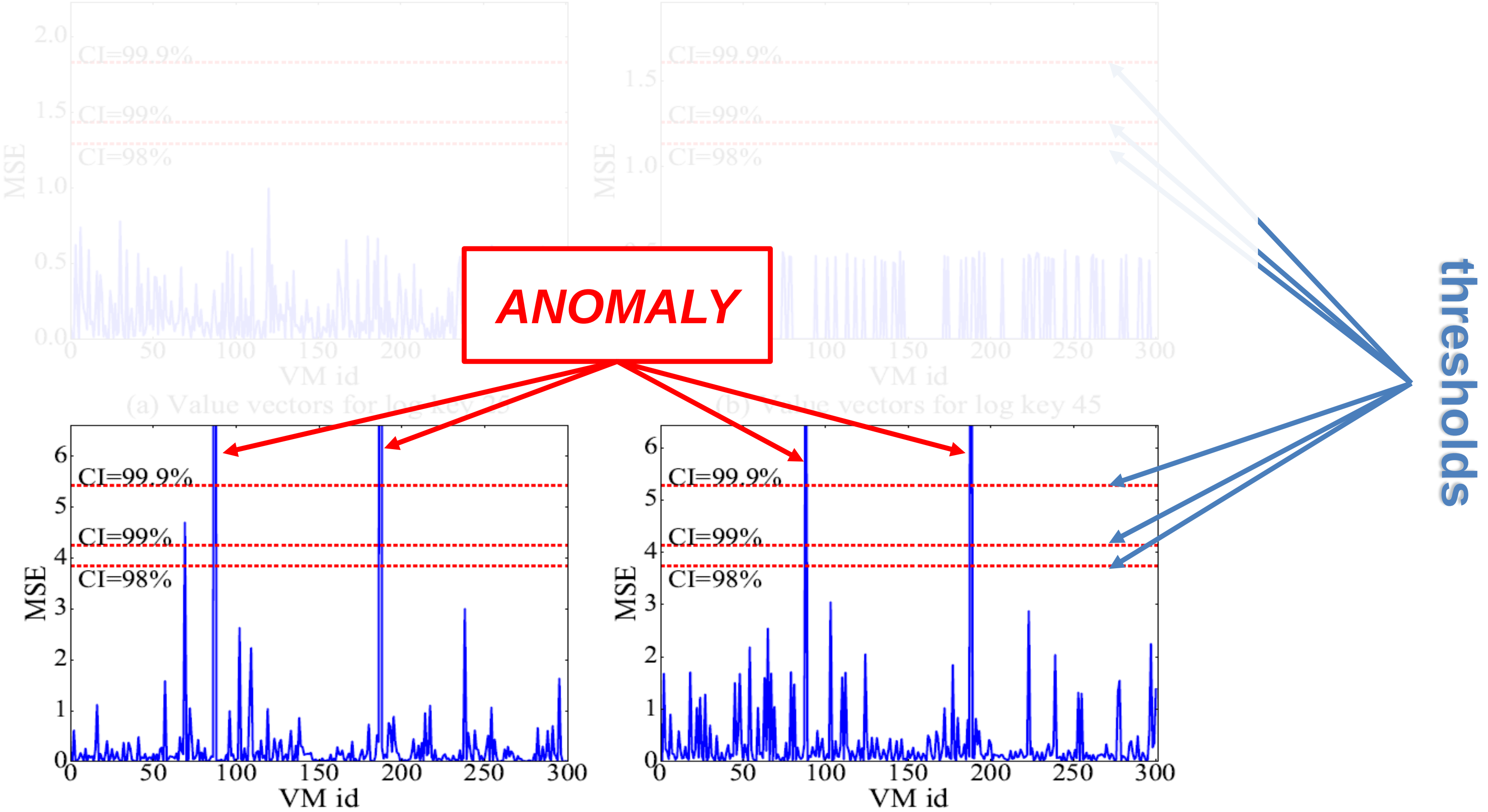
(d) Value vectors for log key 56

thresholds

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

Evaluation – parameter value anomaly detection

MSE:
mean square error



(a) Value vectors for log key

(b) Value vectors for log key 45

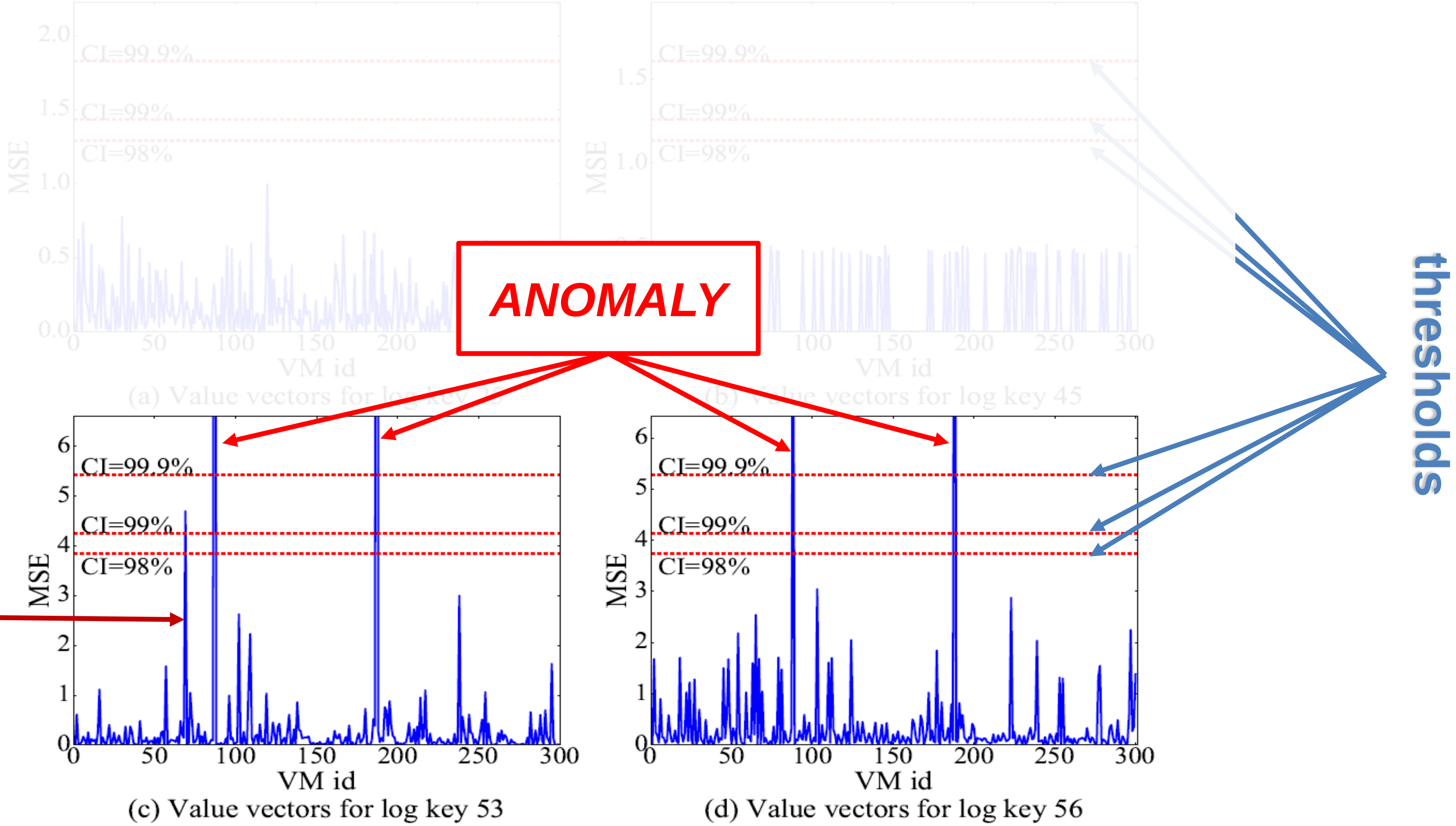
(c) Value vectors for log key 53

(d) Value vectors for log key 56

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

Evaluation – parameter value anomaly detection

MSE:
mean square error



(a) Value vectors for log key

(b) Value vectors for log key 45

(c) Value vectors for log key 53

(d) Value vectors for log key 56

ANOMALY

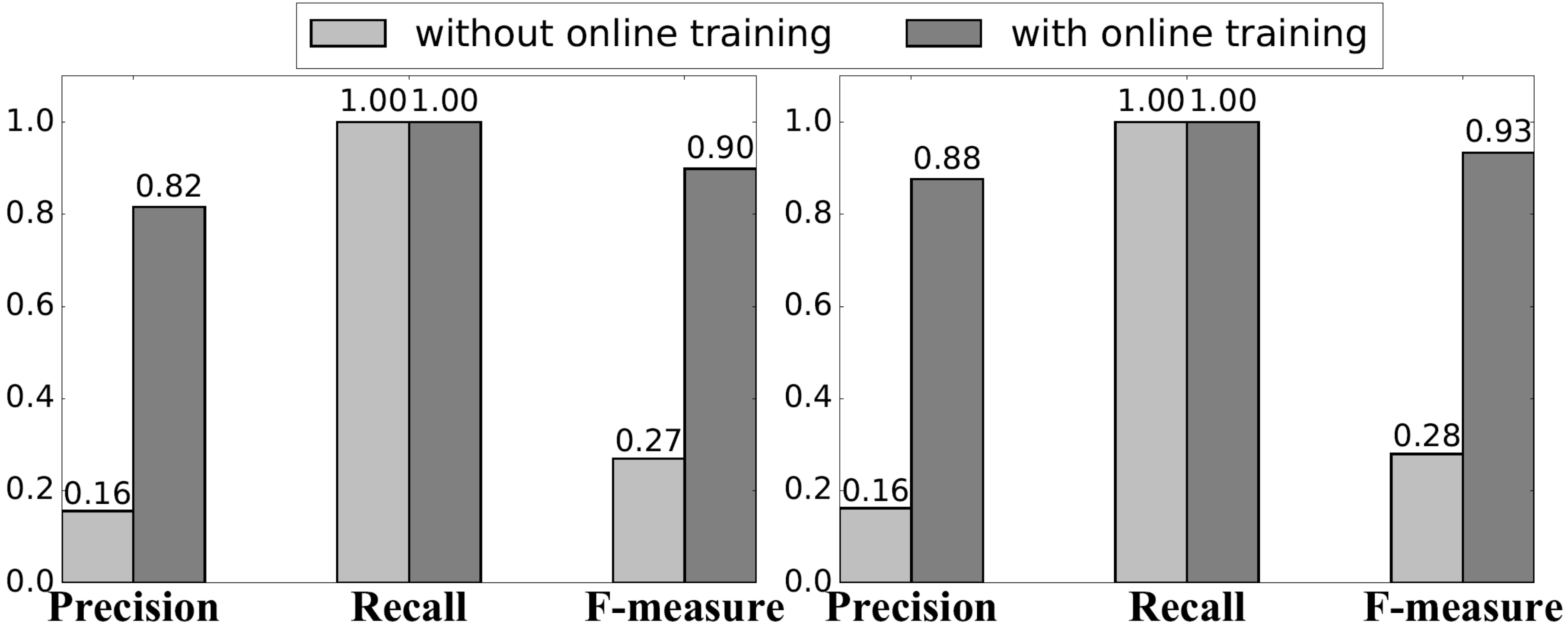
False Positive

thresholds

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

Evaluation – LSTM model online update

Up is good

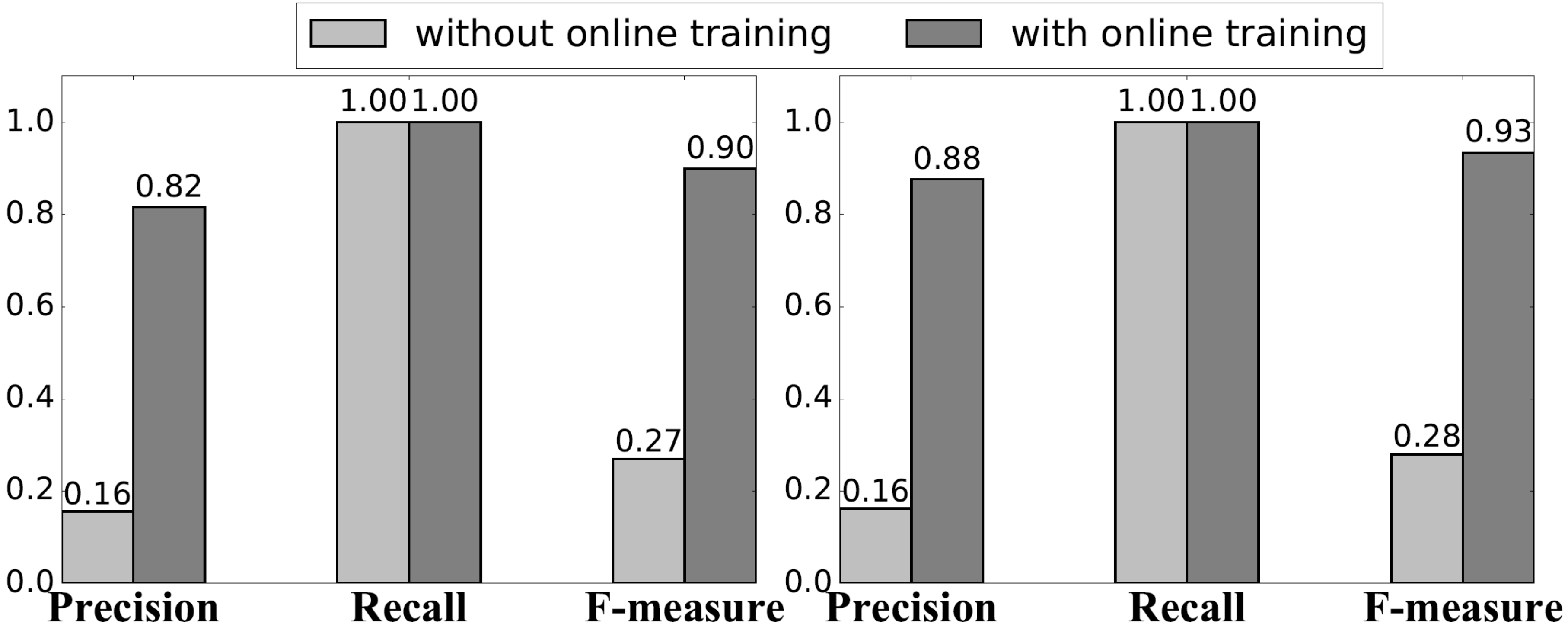


(a) First 1% dataset for training. **(b) First 10% dataset for training.**

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

Evaluation – LSTM model online update

Up is good



(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

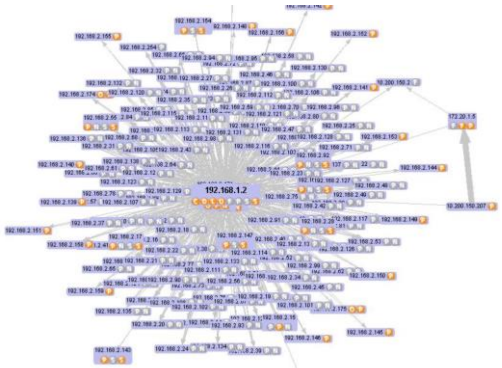
https://www.usenix.org/cfdr-data

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.

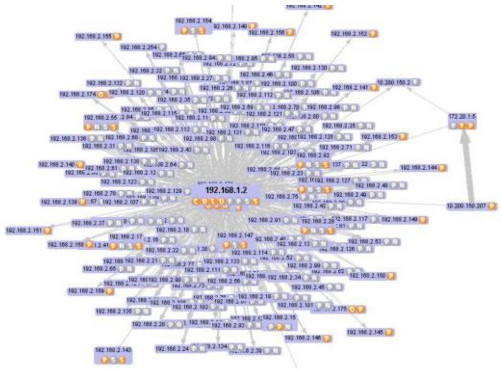


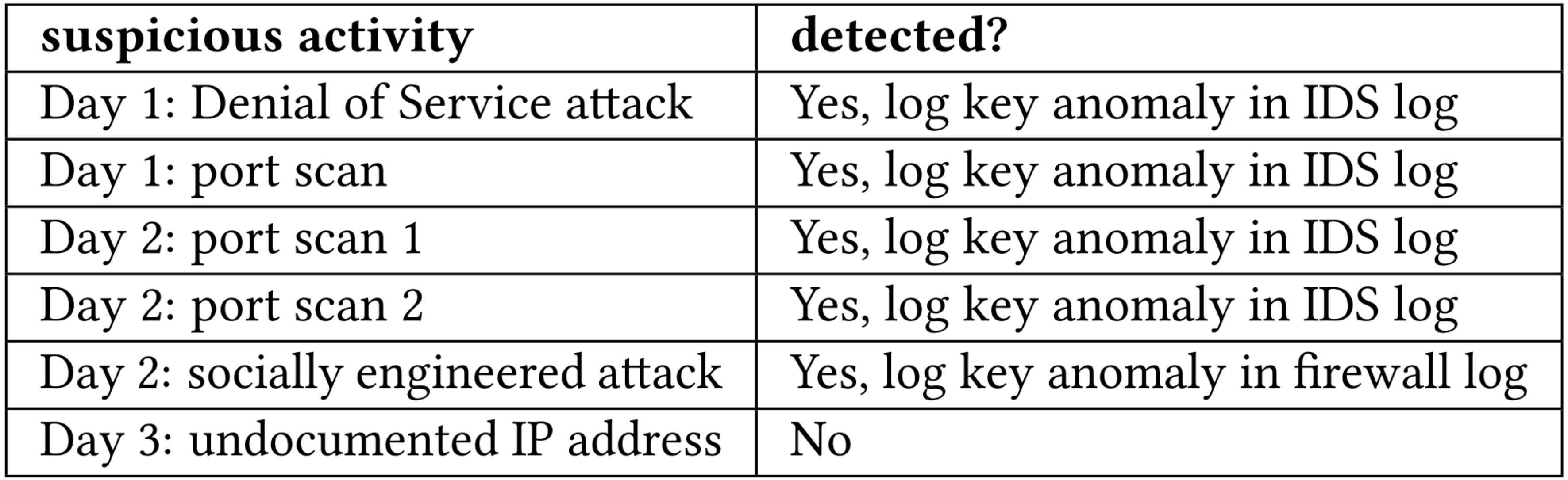
Evaluation – case study: network security log

Dataset: IEEE VAST Challenge 2011

(Mini Challenge 2 – Computer Networking Operations)

The dataset contains firewall log, IDS log, etc.

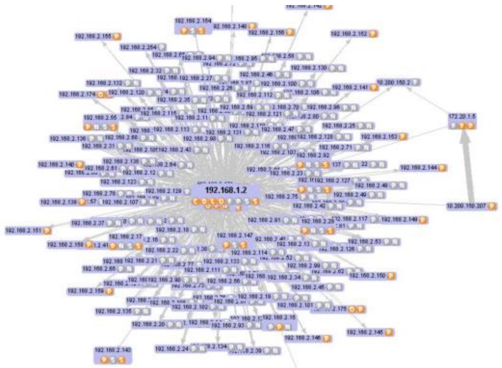




suspicious activity	detected?
Day 1: Denial of Service attack	Yes, log key anomaly in IDS log
Day 1: port scan	Yes, log key anomaly in IDS log
Day 2: port scan 1	Yes, log key anomaly in IDS log
Day 2: port scan 2	Yes, log key anomaly in IDS log
Day 2: socially engineered attack	Yes, log key anomaly in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

Detection results.

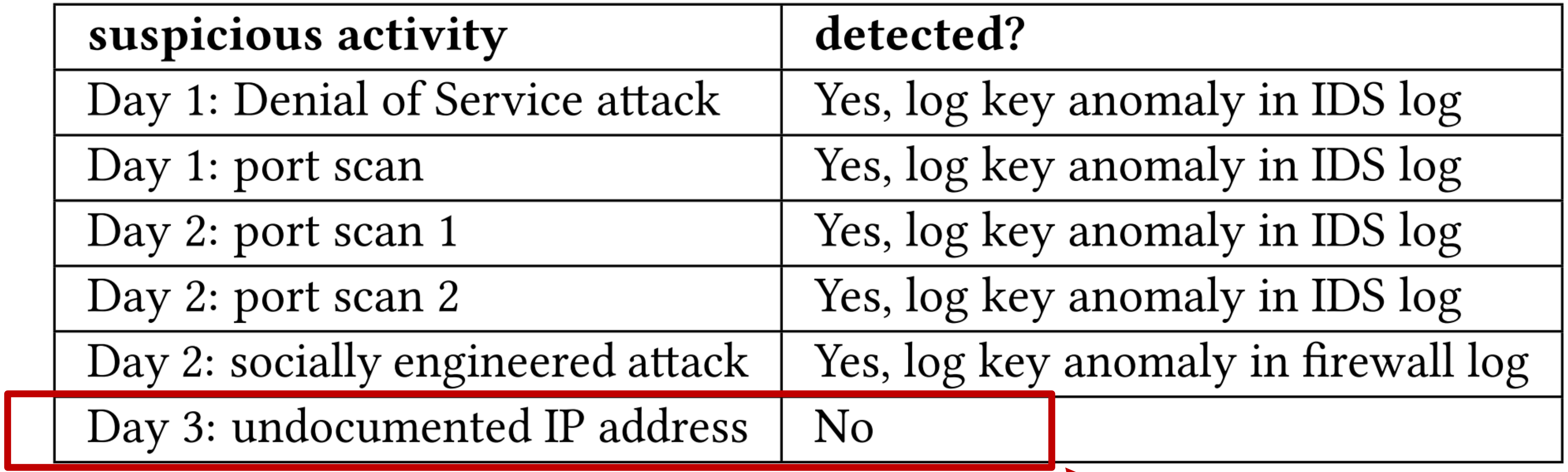
Evaluation – case study: network security log



Dataset: IEEE VAST Challenge 2011

(Mini Challenge 2 – Computer Networking Operations)

The dataset contains firewall log, IDS log, etc.

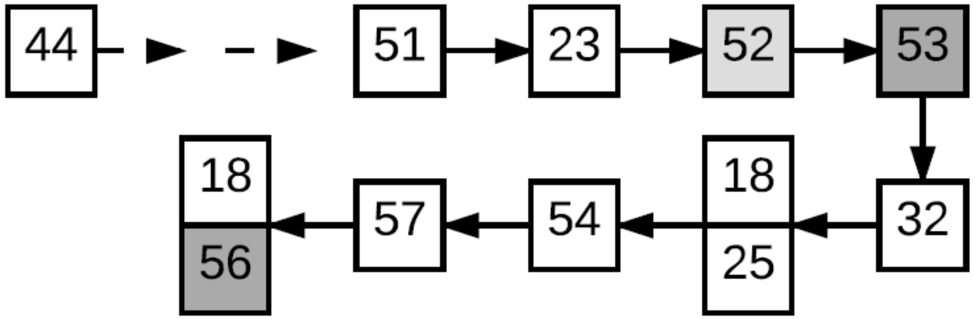


suspicious activity	detected?
Day 1: Denial of Service attack	Yes, log key anomaly in IDS log
Day 1: port scan	Yes, log key anomaly in IDS log
Day 2: port scan 1	Yes, log key anomaly in IDS log
Day 2: port scan 2	Yes, log key anomaly in IDS log
Day 2: socially engineered attack	Yes, log key anomaly in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

Detection results.

Could be fixed with prior knowledge of “documented IP”

Evaluation – workflow construction



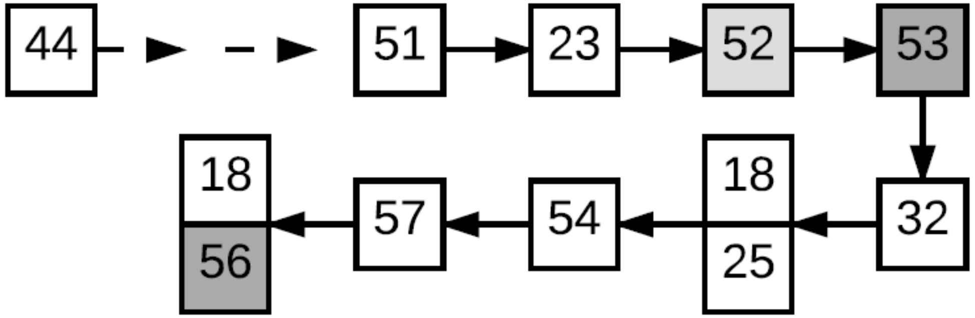
```
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTPV1.1" status: * len: * time: *
52: instance: * Creating image
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (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?



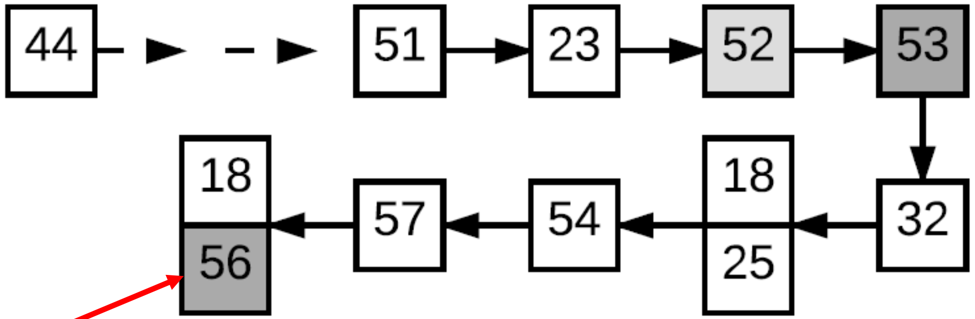
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTPV1.1" status: * len: * time: *
*52: instance: * Creating image*
53: instance: * VM Started (Lifecycle Event)
32: instance: * VM Paused (Lifecycle Event)
18: instance: * VM Resumed (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 * HTTPV1.1" 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
```

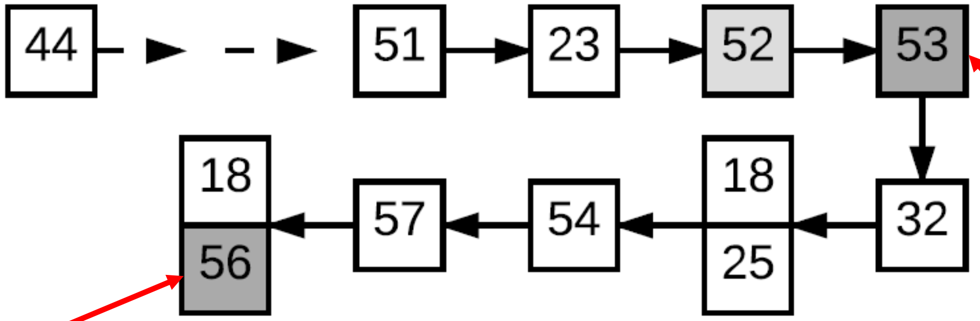
Parameter value anomaly

Constructed workflow of *VM Creation*.

(previously generated OpenStack cloud log)

Evaluation – workflow construction

How does it help to diagnose anomalies?



```
44: instance: * Attempting claim: memory * disk * vcpus * CPU
51: instance: * Claim successful
23: instance: * GET * HTTPV1.1" 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

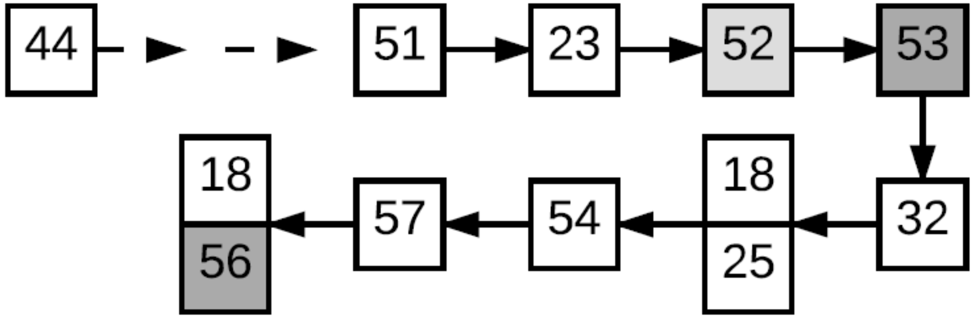
Time difference (performance) anomaly

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 * HTTPV1.1" 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

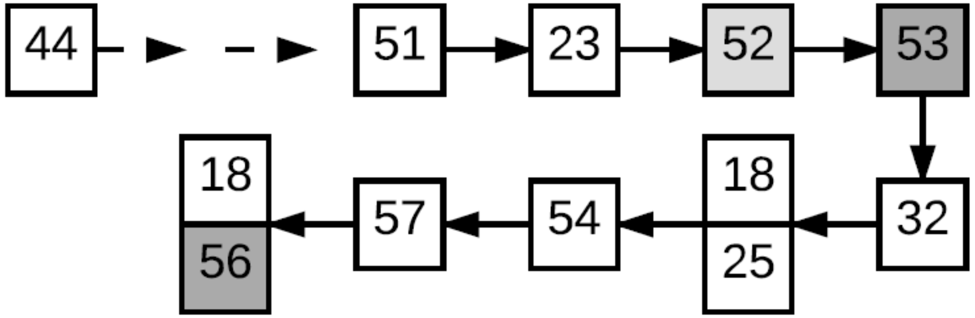
Identified anomaly:
Instance took too long to build because of the transition from 52 -> 53

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 * HTTPV1.1" 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

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

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

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