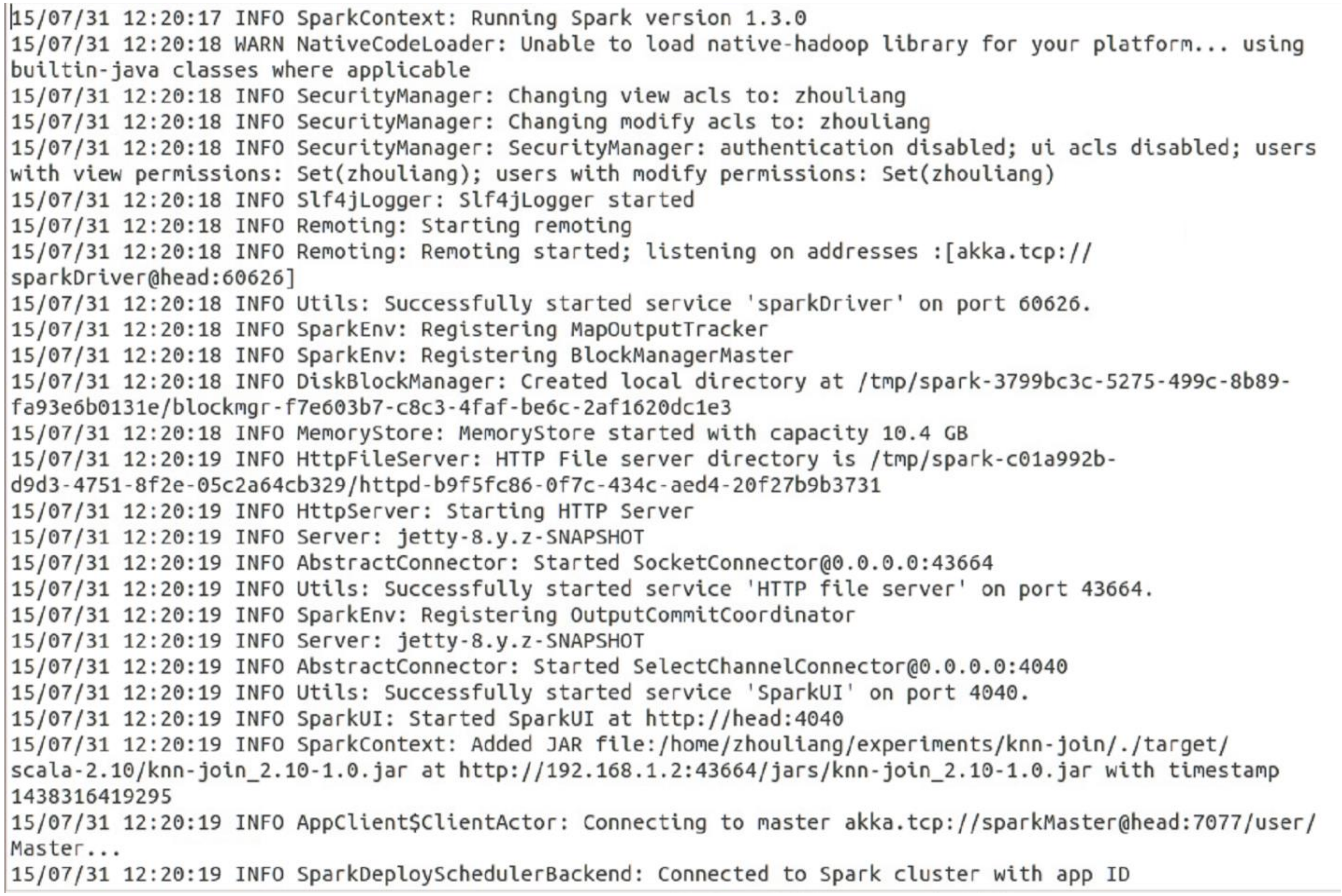
Spell: Streaming Parsing of System Event Logs

Min Du, Feifei Li

School of Computing,

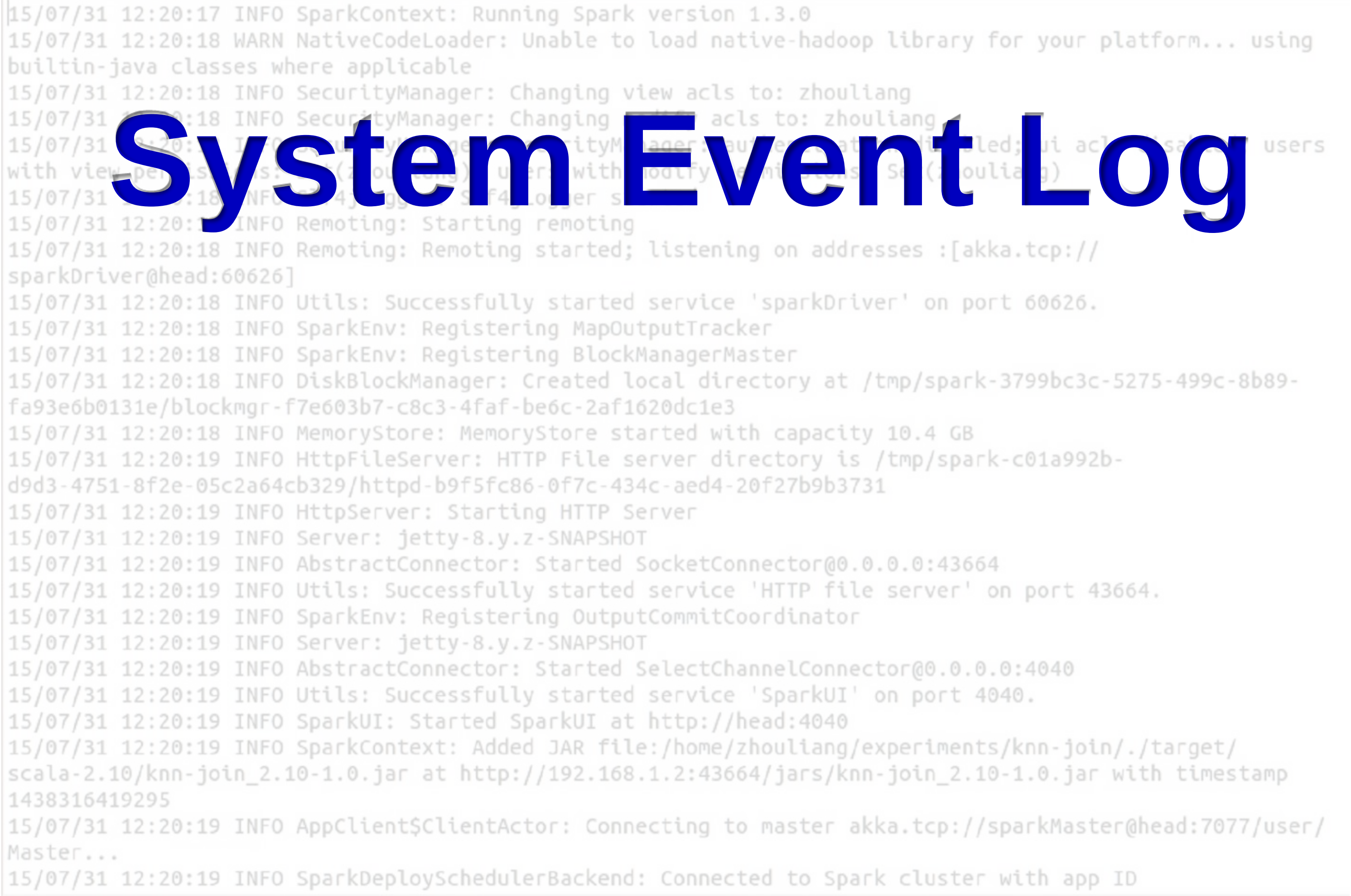
University of Utah

Background



```
15/07/31 12:20:17 INFO SparkContext: Running Spark version 1.3.0
15/07/31 12:20:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
15/07/31 12:20:18 INFO SecurityManager: Changing view acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: Changing modify acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users
with view permissions: Set(zhouliang); users with modify permissions: Set(zhouliang)
15/07/31 12:20:18 INFO Slf4jLogger: Slf4jLogger started
15/07/31 12:20:18 INFO Remoting: Starting remoting
15/07/31 12:20:18 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
sparkDriver@head:60626]
15/07/31 12:20:18 INFO Utils: Successfully started service 'sparkDriver' on port 60626.
15/07/31 12:20:18 INFO SparkEnv: Registering MapOutputTracker
15/07/31 12:20:18 INFO SparkEnv: Registering BlockManagerMaster
15/07/31 12:20:18 INFO DiskBlockManager: Created local directory at /tmp/spark-3799bc3c-5275-499c-8b89-
fa93e6b0131e/blockmgr-f7e603b7-c8c3-4faf-be6c-2af1620dc1e3
15/07/31 12:20:18 INFO MemoryStore: MemoryStore started with capacity 10.4 GB
15/07/31 12:20:19 INFO HttpFileServer: HTTP File server directory is /tmp/spark-c01a992b-
d9d3-4751-8f2e-05c2a64cb329/httpd-b9f5fc86-0f7c-434c-aed4-20f27b9b3731
15/07/31 12:20:19 INFO HttpServer: Starting HTTP Server
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SocketConnector@0.0.0.0:43664
15/07/31 12:20:19 INFO Utils: Successfully started service 'HTTP file server' on port 43664.
15/07/31 12:20:19 INFO SparkEnv: Registering OutputCommitCoordinator
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
15/07/31 12:20:19 INFO Utils: Successfully started service 'SparkUI' on port 4040.
15/07/31 12:20:19 INFO SparkUI: Started SparkUI at http://head:4040
15/07/31 12:20:19 INFO SparkContext: Added JAR file:/home/zhouliang/experiments/knn-join/./target/
scala-2.10/knn-join_2.10-1.0.jar at http://192.168.1.2:43664/jars/knn-join_2.10-1.0.jar with timestamp
1438316419295
15/07/31 12:20:19 INFO AppClient$ClientActor: Connecting to master akka.tcp://sparkMaster@head:7077/user/
Master...
15/07/31 12:20:19 INFO SparkDeploySchedulerBackend: Connected to Spark cluster with app ID
```

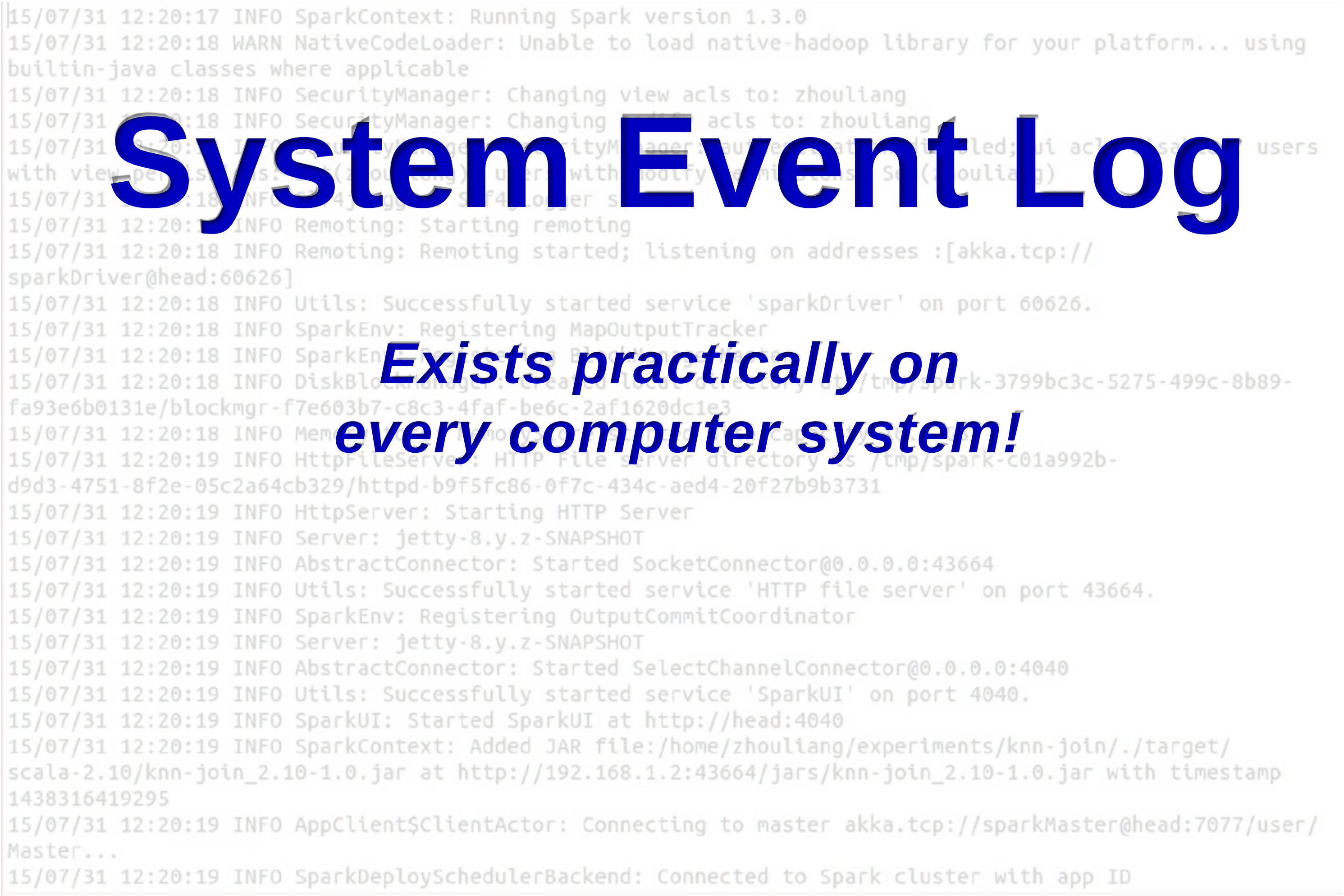
Background



System Event Log

```
15/07/31 12:20:17 INFO SparkContext: Running Spark version 1.3.0
15/07/31 12:20:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
15/07/31 12:20:18 INFO SecurityManager: Changing view acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: Changing modify acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users
with view permissions: Set(zhouliang); users with modify permissions: Set(zhouliang)
15/07/31 12:20:18 INFO Slf4jLogger: Slf4jLogger started
15/07/31 12:20:18 INFO Remoting: Starting remoting
15/07/31 12:20:18 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
sparkDriver@head:60626]
15/07/31 12:20:18 INFO Utils: Successfully started service 'sparkDriver' on port 60626.
15/07/31 12:20:18 INFO SparkEnv: Registering MapOutputTracker
15/07/31 12:20:18 INFO SparkEnv: Registering BlockManagerMaster
15/07/31 12:20:18 INFO DiskBlockManager: Created local directory at /tmp/spark-3799bc3c-5275-499c-8b89-
fa93e6b0131e/blockmgr-f7e603b7-c8c3-4faf-be6c-2af1620dc1e3
15/07/31 12:20:18 INFO MemoryStore: MemoryStore started with capacity 10.4 GB
15/07/31 12:20:19 INFO HttpFileServer: HTTP File server directory is /tmp/spark-c01a992b-
d9d3-4751-8f2e-05c2a64cb329/httpd-b9f5fc86-0f7c-434c-aed4-20f27b9b3731
15/07/31 12:20:19 INFO HttpServer: Starting HTTP Server
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SocketConnector@0.0.0.0:43664
15/07/31 12:20:19 INFO Utils: Successfully started service 'HTTP file server' on port 43664.
15/07/31 12:20:19 INFO SparkEnv: Registering OutputCommitCoordinator
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
15/07/31 12:20:19 INFO Utils: Successfully started service 'SparkUI' on port 4040.
15/07/31 12:20:19 INFO SparkUI: Started SparkUI at http://head:4040
15/07/31 12:20:19 INFO SparkContext: Added JAR file:/home/zhouliang/experiments/knn-join/./target/
scala-2.10/knn-join_2.10-1.0.jar at http://192.168.1.2:43664/jars/knn-join_2.10-1.0.jar with timestamp
1438316419295
15/07/31 12:20:19 INFO AppClient$ClientActor: Connecting to master akka.tcp://sparkMaster@head:7077/user/
Master...
15/07/31 12:20:19 INFO SparkDeploySchedulerBackend: Connected to Spark cluster with app ID
```

Background

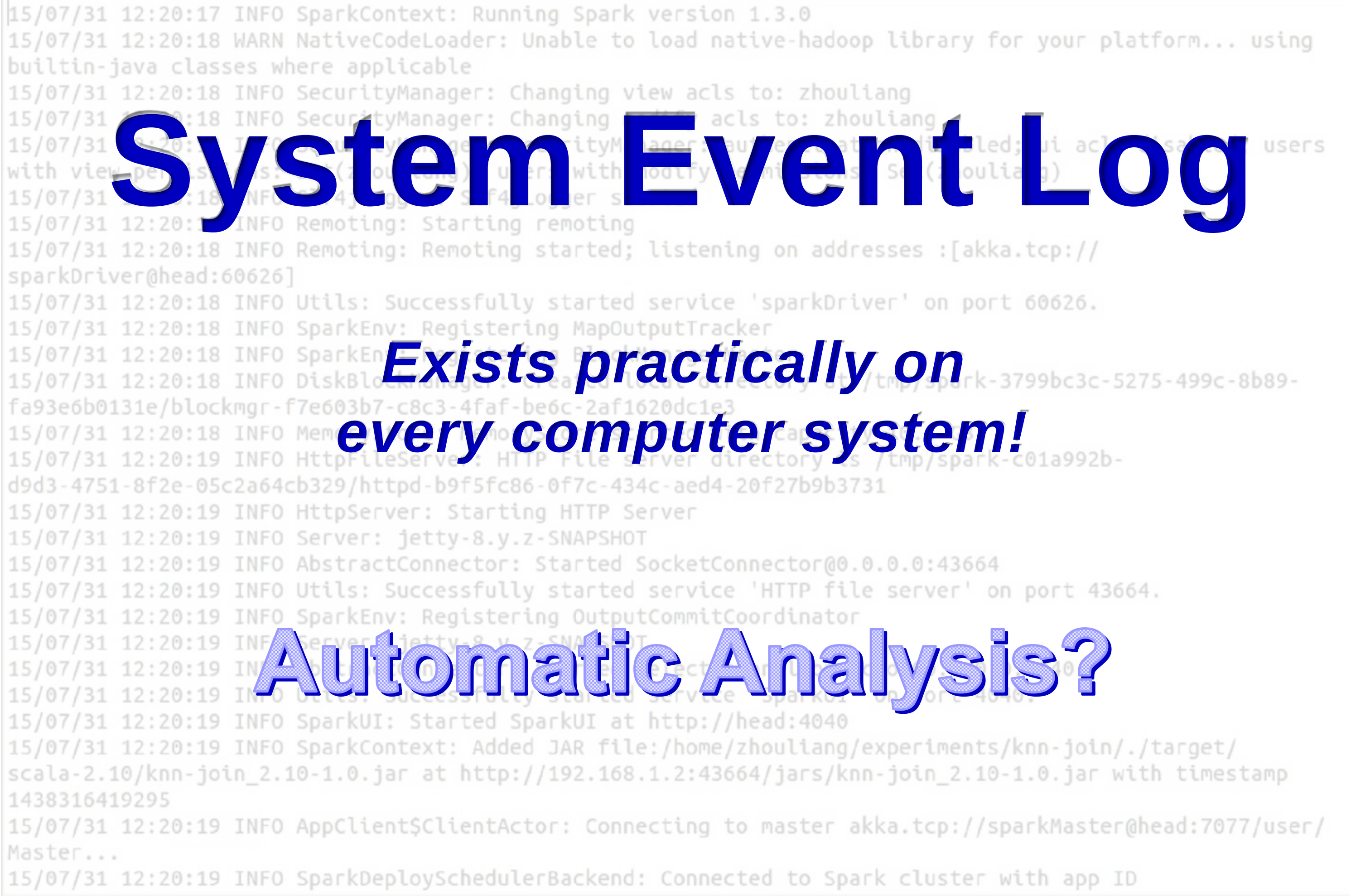


```
15/07/31 12:20:17 INFO SparkContext: Running Spark version 1.3.0
15/07/31 12:20:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
15/07/31 12:20:18 INFO SecurityManager: Changing view acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: Changing modify acls to: zhouliang
15/07/31 12:20:18 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users
with view permissions: Set(zhouliang); users with modify permissions: Set(zhouliang)
15/07/31 12:20:18 INFO Slf4jLogger: Slf4jLogger started
15/07/31 12:20:18 INFO Remoting: Starting remoting
15/07/31 12:20:18 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
sparkDriver@head:60626]
15/07/31 12:20:18 INFO Utils: Successfully started service 'sparkDriver' on port 60626.
15/07/31 12:20:18 INFO SparkEnv: Registering MapOutputTracker
15/07/31 12:20:18 INFO SparkEnv: Registering BlockManagerMaster
15/07/31 12:20:18 INFO DiskBlockManager: Created local directory at /tmp/spark-3799bc3c-5275-499c-8b89-
fa93e6b0131e/blockmgr-f7e603b7-c8c3-4faf-be6c-2af1620dc1e3
15/07/31 12:20:18 INFO MemoryStore: MemoryStore started with capacity [illegible]
15/07/31 12:20:19 INFO HttpFileServer: HTTP File server directory is /tmp/spark-c01a992b-
d9d3-4751-8f2e-05c2a64cb329/httpd-b9f5fc86-0f7c-434c-aed4-20f27b9b3731
15/07/31 12:20:19 INFO HttpServer: Starting HTTP Server
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SocketConnector@0.0.0.0:43664
15/07/31 12:20:19 INFO Utils: Successfully started service 'HTTP file server' on port 43664.
15/07/31 12:20:19 INFO SparkEnv: Registering OutputCommitCoordinator
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
15/07/31 12:20:19 INFO Utils: Successfully started service 'SparkUI' on port 4040.
15/07/31 12:20:19 INFO SparkUI: Started SparkUI at http://head:4040
15/07/31 12:20:19 INFO SparkContext: Added JAR file:/home/zhouliang/experiments/knn-join/./target/
scala-2.10/knn-join_2.10-1.0.jar at http://192.168.1.2:43664/jars/knn-join_2.10-1.0.jar with timestamp
1438316419295
15/07/31 12:20:19 INFO AppClient$ClientActor: Connecting to master akka.tcp://sparkMaster@head:7077/user/
Master...
15/07/31 12:20:19 INFO SparkDeploySchedulerBackend: Connected to Spark cluster with app ID
```

System Event Log

Exists practically on every computer system!

Background



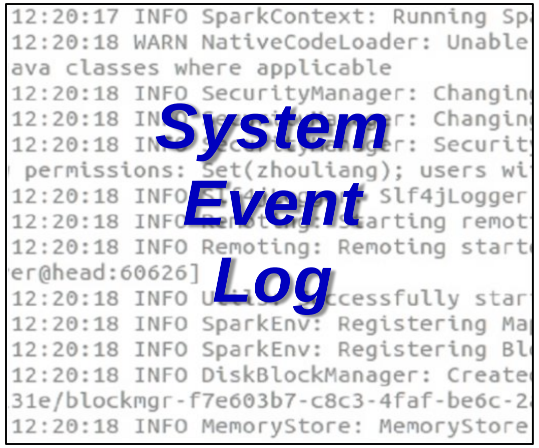
```
15/07/31 12:20:17 INFO SparkContext: Running Spark version 1.3.0
15/07/31 12:20:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
15/07/31 12:20:18 INFO SecurityManager: Changing view acls to: zhouliang
15/07/31 [illegible]:18 INFO SecurityManager: Changing [illegible] acls to: zhouliang
15/07/31 [illegible] [illegible]ityManager: [illegible] [illegible]led; ui acls [illegible] users
with view [illegible] users with modify [illegible] Set(zhouliang)
15/07/31 [illegible]:18 INFO [illegible] [illegible]er s[illegible]
15/07/31 12:20:[illegible] INFO Remoting: Starting remoting
15/07/31 12:20:18 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://
sparkDriver@head:60626]
15/07/31 12:20:18 INFO Utils: Successfully started service 'sparkDriver' on port 60626.
15/07/31 12:20:18 INFO SparkEnv: Registering MapOutputTracker
15/07/31 12:20:18 INFO SparkEn[illegible]
15/07/31 12:20:18 INFO DiskBlo[illegible] [illegible]/tmp/spark-3799bc3c-5275-499c-8b89-
fa93e6b0131e/blockmgr-f7e603b7-c8c3-4faf-be6c-2af1620dc1e3
15/07/31 12:20:18 INFO Mem[illegible]
15/07/31 12:20:19 INFO HttpFileServer: HTTP File server directory is /tmp/spark-c01a992b-
d9d3-4751-8f2e-05c2a64cb329/httpd-b9f5fc86-0f7c-434c-aed4-20f27b9b3731
15/07/31 12:20:19 INFO HttpServer: Starting HTTP Server
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 INFO AbstractConnector: Started SocketConnector@0.0.0.0:43664
15/07/31 12:20:19 INFO Utils: Successfully started service 'HTTP file server' on port 43664.
15/07/31 12:20:19 INFO SparkEnv: Registering OutputCommitCoordinator
15/07/31 12:20:19 INFO Server: jetty-8.y.z-SNAPSHOT
15/07/31 12:20:19 IN[illegible]
15/07/31 12:20:19 INFO Utils: Successfully started service 'SparkUI' on port 4040.
15/07/31 12:20:19 INFO SparkUI: Started SparkUI at http://head:4040
15/07/31 12:20:19 INFO SparkContext: Added JAR file:/home/zhouliang/experiments/knn-join/./target/
scala-2.10/knn-join_2.10-1.0.jar at http://192.168.1.2:43664/jars/knn-join_2.10-1.0.jar with timestamp
1438316419295
15/07/31 12:20:19 INFO AppClient$ClientActor: Connecting to master akka.tcp://sparkMaster@head:7077/user/
Master...
15/07/31 12:20:19 INFO SparkDeploySchedulerBackend: Connected to Spark cluster with app ID
```

System Event Log

Exists practically on
every computer system!

Automatic 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 IN                 r: Changin
12:20:18 IN                 r: 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
```

System Event Log

Strucuted Data

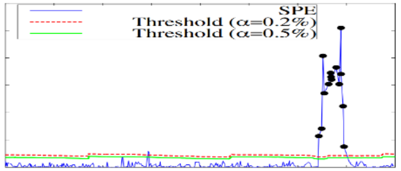
Message/Event type

Log key

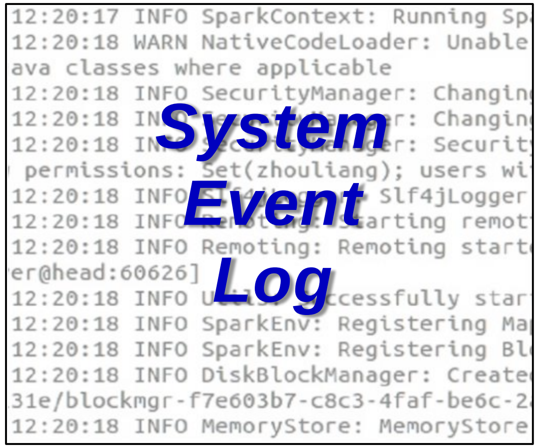
……

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

Anomaly Detection



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

Strucuted Data

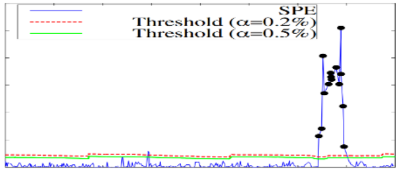
Message/Event type

Log key

……

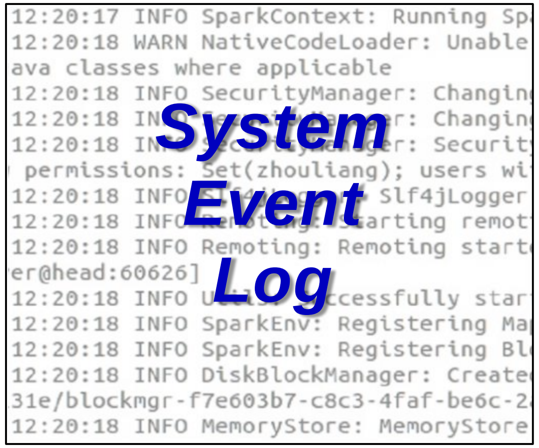
printf(“***Started service***
%s ***on port*** %d”, x, y);

Anomaly Detection



LOG ANALYSIS

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 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
```

Strucuted Data

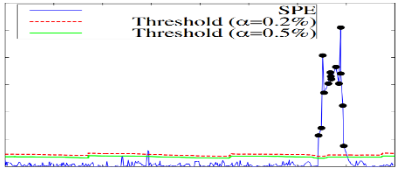
Message/Event type

Log key

……

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

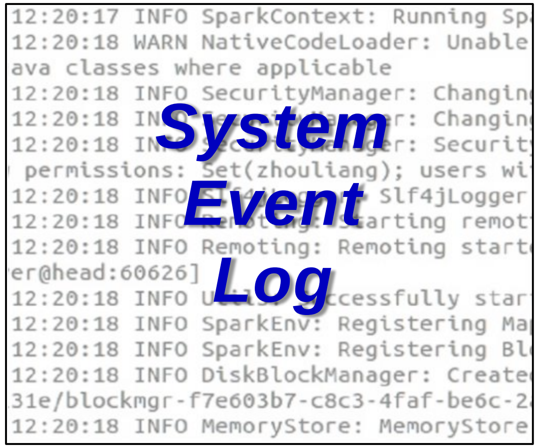
Anomaly Detection



LOG ANALYSIS

- ❑ **Message count vector:** Xu’SOSP09, Lou’ATC10, Lin’ICSE16, 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

→

Strucuted Data

Message/Event type

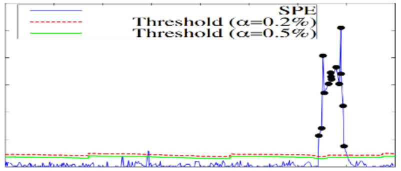
Log key

……

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

→

Anomaly Detection

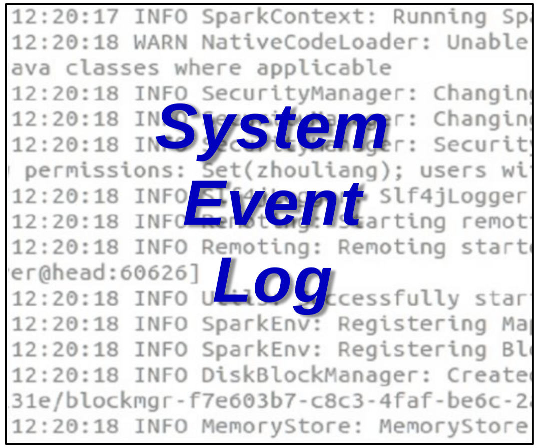


LOG ANALYSIS

- ❑ **Message count vector:**
  Xu’SOSP09, Lou’ATC10, Lin’ICSE16, etc.

- ❑ **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

Strucuted Data

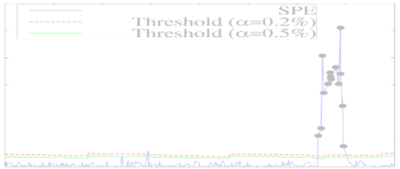
Message/Event type

Log key

……

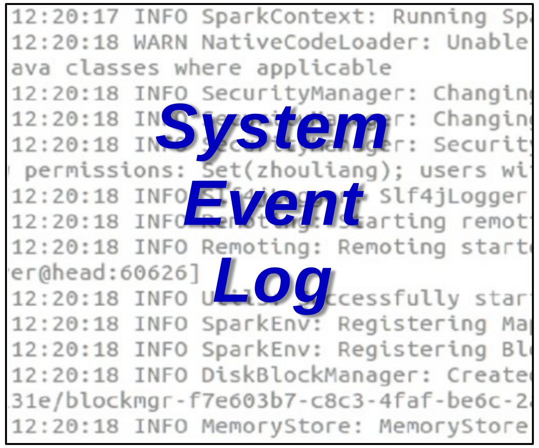
printf("***Started service*** %s ***on port*** %d", x, y);

Anomaly Detection



LOG PARSING

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

Strucuted Data

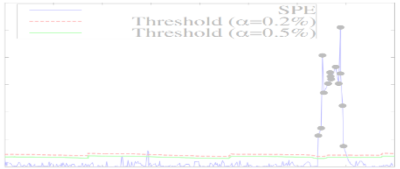
Message/Event type

Log key

……

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

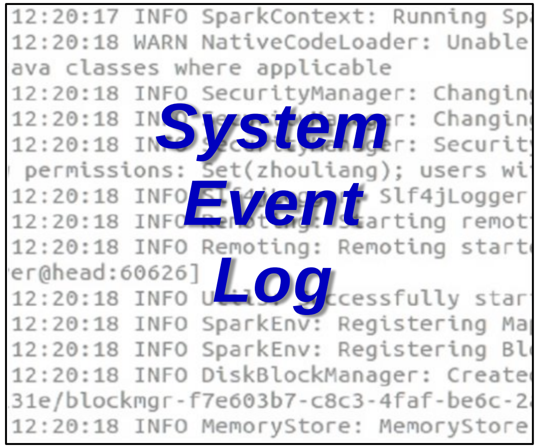
Anomaly Detection



LOG PARSING

- ❑ **Use source code as template to parse logs:**
  Xu’SOSP09

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

Strucuted Data

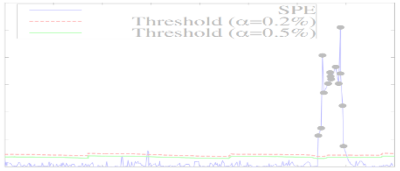
Message/Event type

Log key

……

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

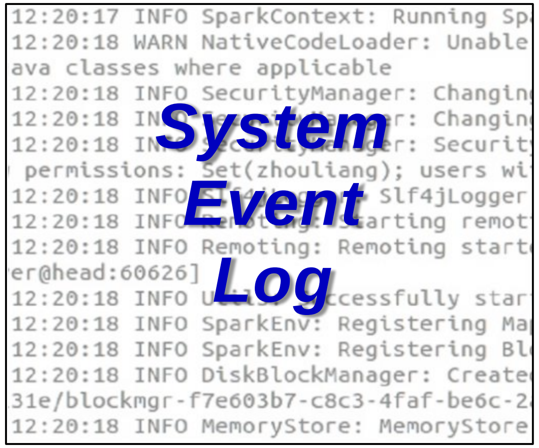
Anomaly Detection



LOG PARSING

- ❑ **Use source code as template to parse logs:**
  Xu’SOSP09
  Problem: What if we don’t have source code?

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 starte
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

Strucuted Data

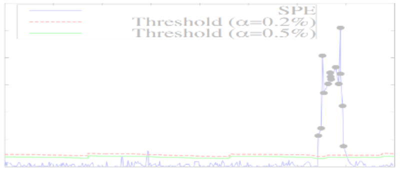
Message/Event type

Log key

……

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

Anomaly Detection

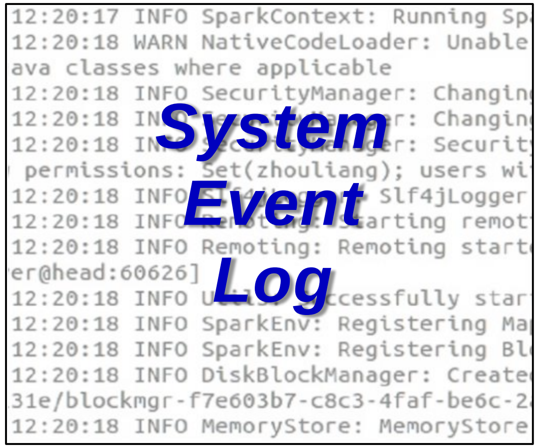


LOG PARSING

- ❑ **Use source code as template to parse logs:**
  Xu'SOSP09
  Problem: What if we don't have source code?

- ❑ **Directly parse from raw system logs:**
  Makanju'KDD09, Fu'ICDM09, Tang'ICDM10, Tang'CIKM11, 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[illegible]r: Changin
12:20:18 IN[illegible]r: Securit
 permissions: Set(zhouliang); users wi
12:20:18 INFO [illegible] Slf4jLogger
12:20:18 INFO [illegible]arting 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

Strucuted Data

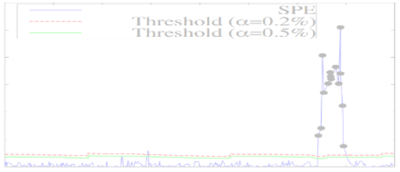
Message/Event type

Log key

……

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

Anomaly Detection



LOG PARSING

- ❑ **Use source code as template to parse logs:**
  Xu’SOSP09
  Problem: What if we don’t have source code?

- ❑ **Directly parse from raw system logs:**
  Makanju’KDD09, Fu’ICDM09, Tang’ICDM10, Tang’CIKM11, etc.
  Problem: Offline batched processing, some very slow.

Our approach

Spell, **a structured *S*treaming *P*arser for *E*vent *L*ogs using an *L*CS (longest common subsequence) based approach.**

Our approach

Spell, **a structured *S*treaming *P*arser for *E*vent *L*ogs using an *L*CS (longest common subsequence) based approach.**

Example:

Two log entries:

Temperature (41C) exceeds warning threshold
Temperature (42C, 43C) exceeds warning threshold

Our approach

Spell, **a structured *S*treaming *P*arser for *E*vent *L*ogs using an *L*CS (longest common subsequence) based approach.**

Example:

Two log entries:

Temperature (41C) exceeds warning threshold

Temperature (42C, 43C) exceeds warning threshold

LCS:

*Temperature * exceeds warning threshold*

Our approach

Spell**, a structured *S*treaming *P*arser for *E*vent *L*ogs using an *L*CS (longest common subsequence) based approach.**

Example:

Two log entries:

Temperature (41C) exceeds warning threshold

Temperature (42C, 43C) exceeds warning threshold

LCS:

*Temperature * exceeds warning threshold*

Naturally a message type!

printf(“Temperature %s exceeds warning threshold”)

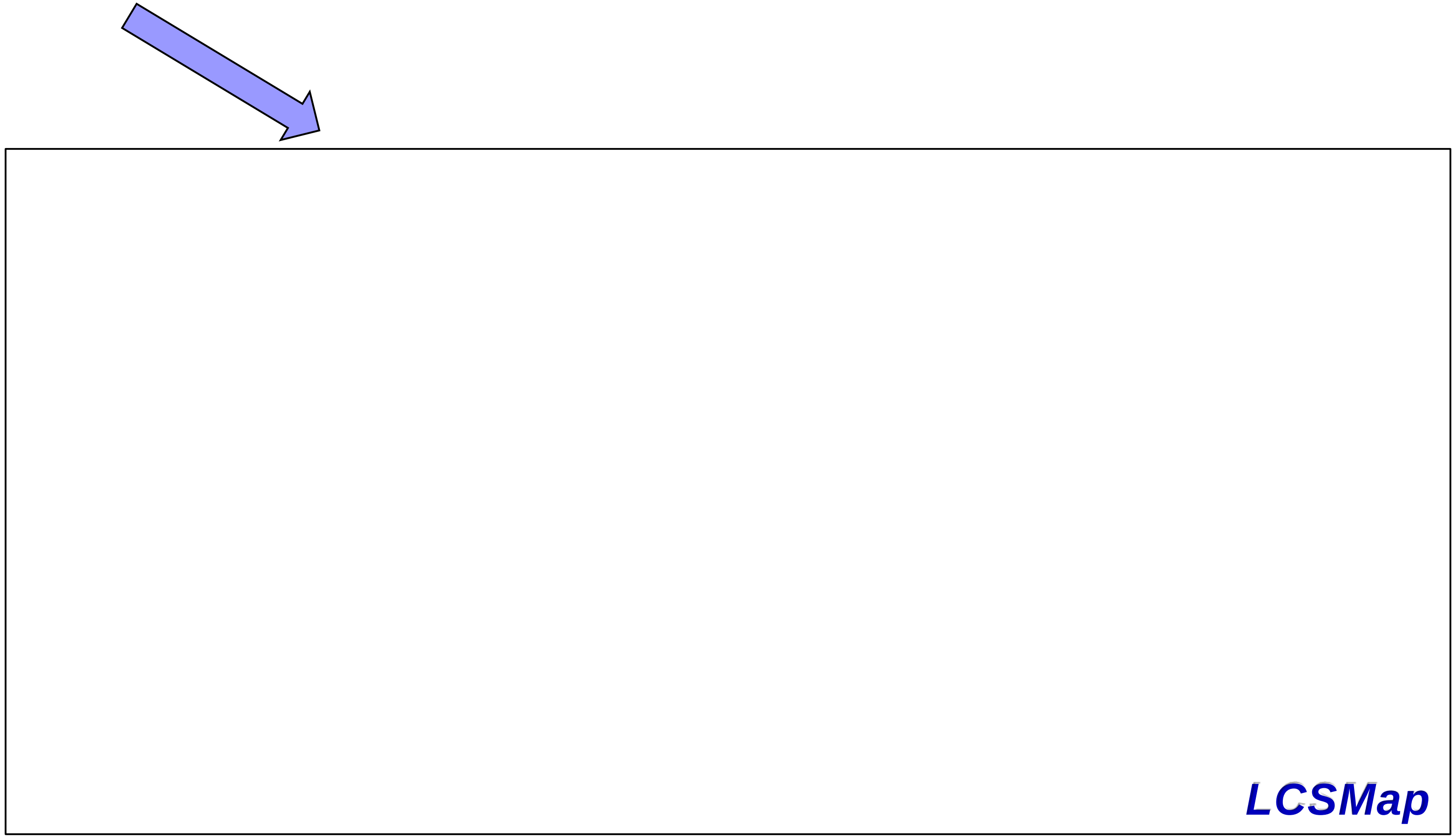
SPELL – Basic workflow

Add new log entry into LCSMap in a streaming fashion, update existing message type if *length(LCS) > 0.5 * length(new log entry)*

LCSMap

SPELL – Basic workflow

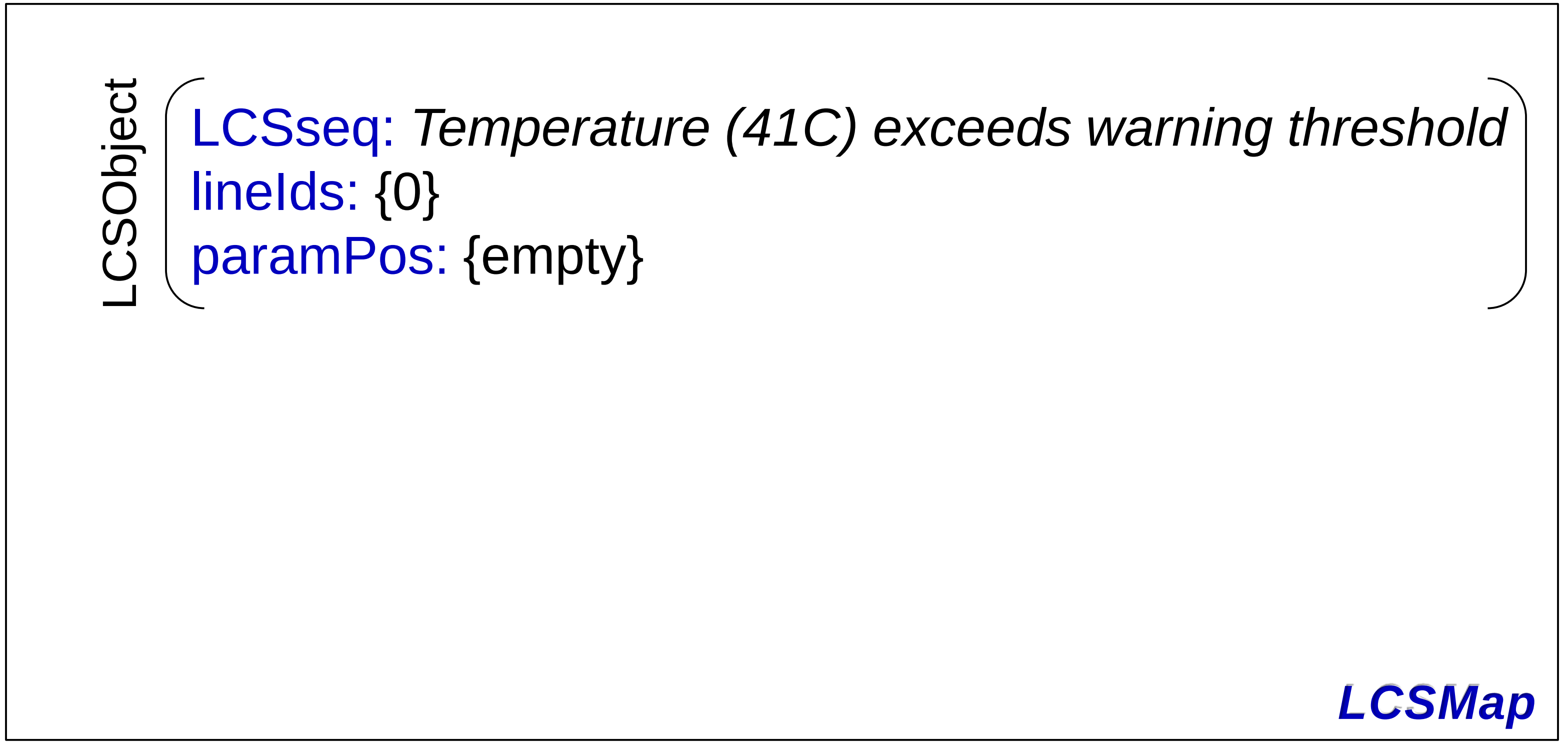
new log entry: *Temperature (41C) exceeds warning threshold*



LCSMap

SPELL – Basic workflow

new log entry:



LCSObject

LCSseq: *Temperature (41C) exceeds warning threshold*

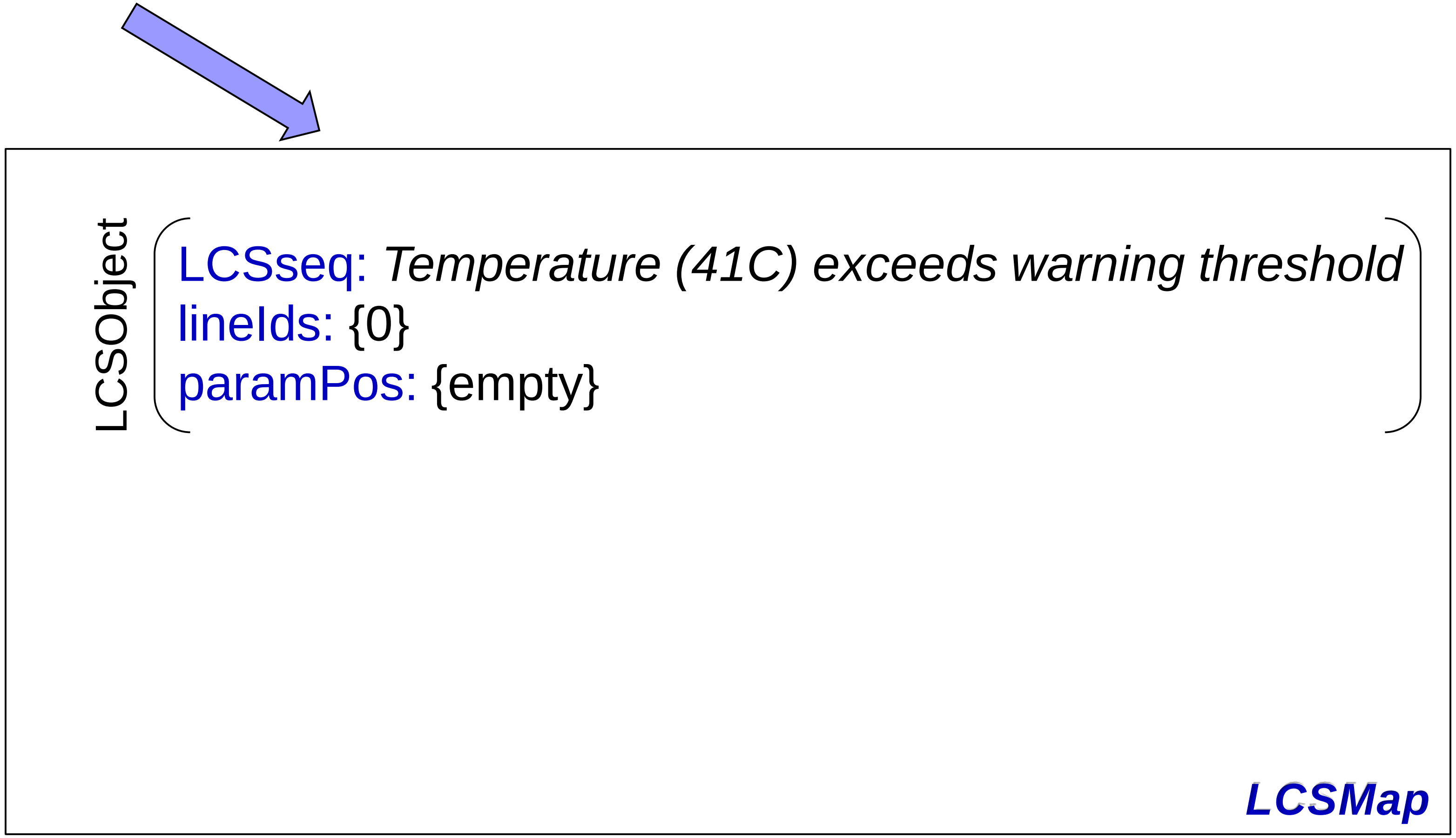
lineIds: {0}

paramPos: {empty}

LCSMap

SPELL – Basic workflow

new log entry: *Temperature (43C) exceeds warning threshold*



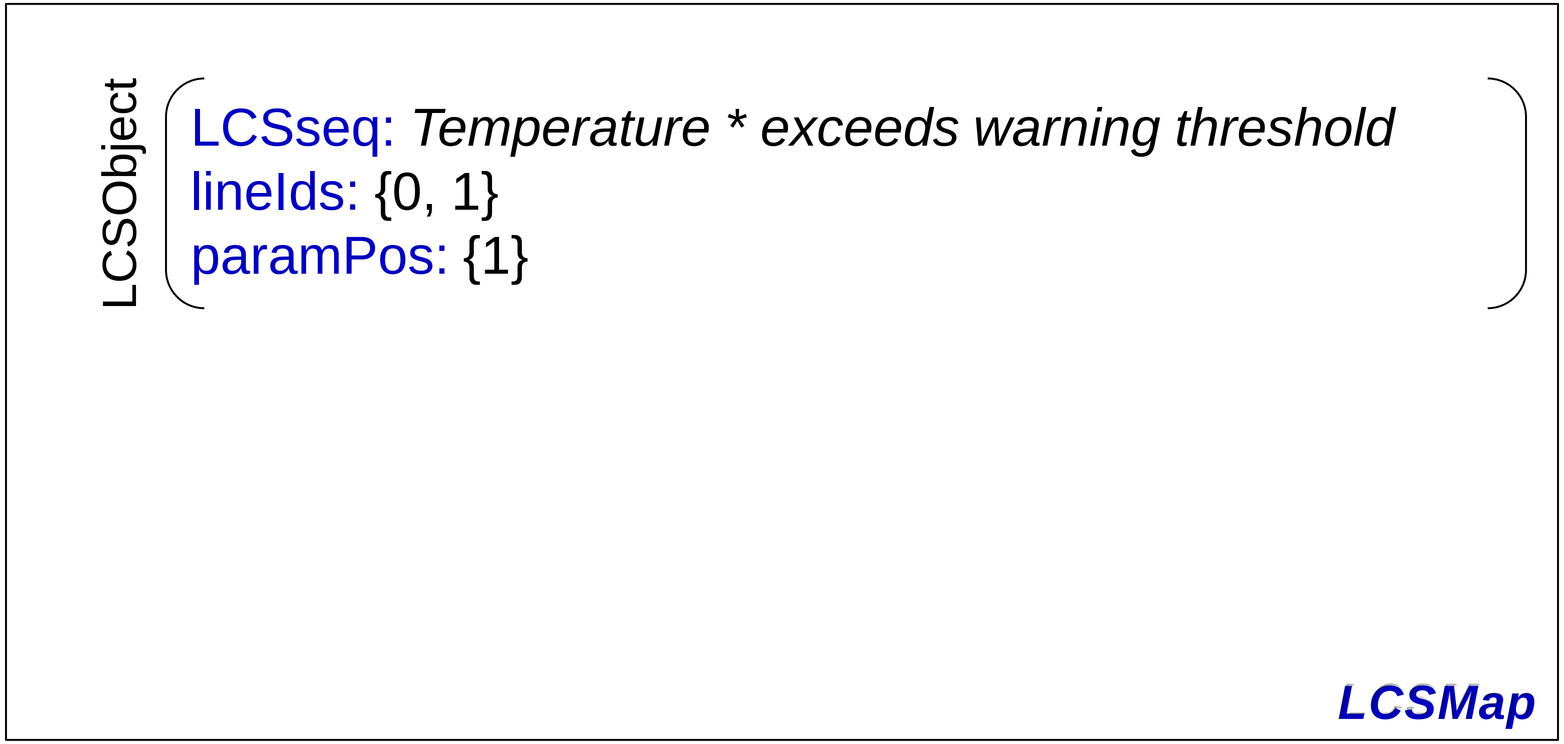
LCSObject

LCSseq: *Temperature (41C) exceeds warning threshold*
lineIds: {0}
paramPos: {empty}

LCSMap

SPELL – Basic workflow

new log entry:



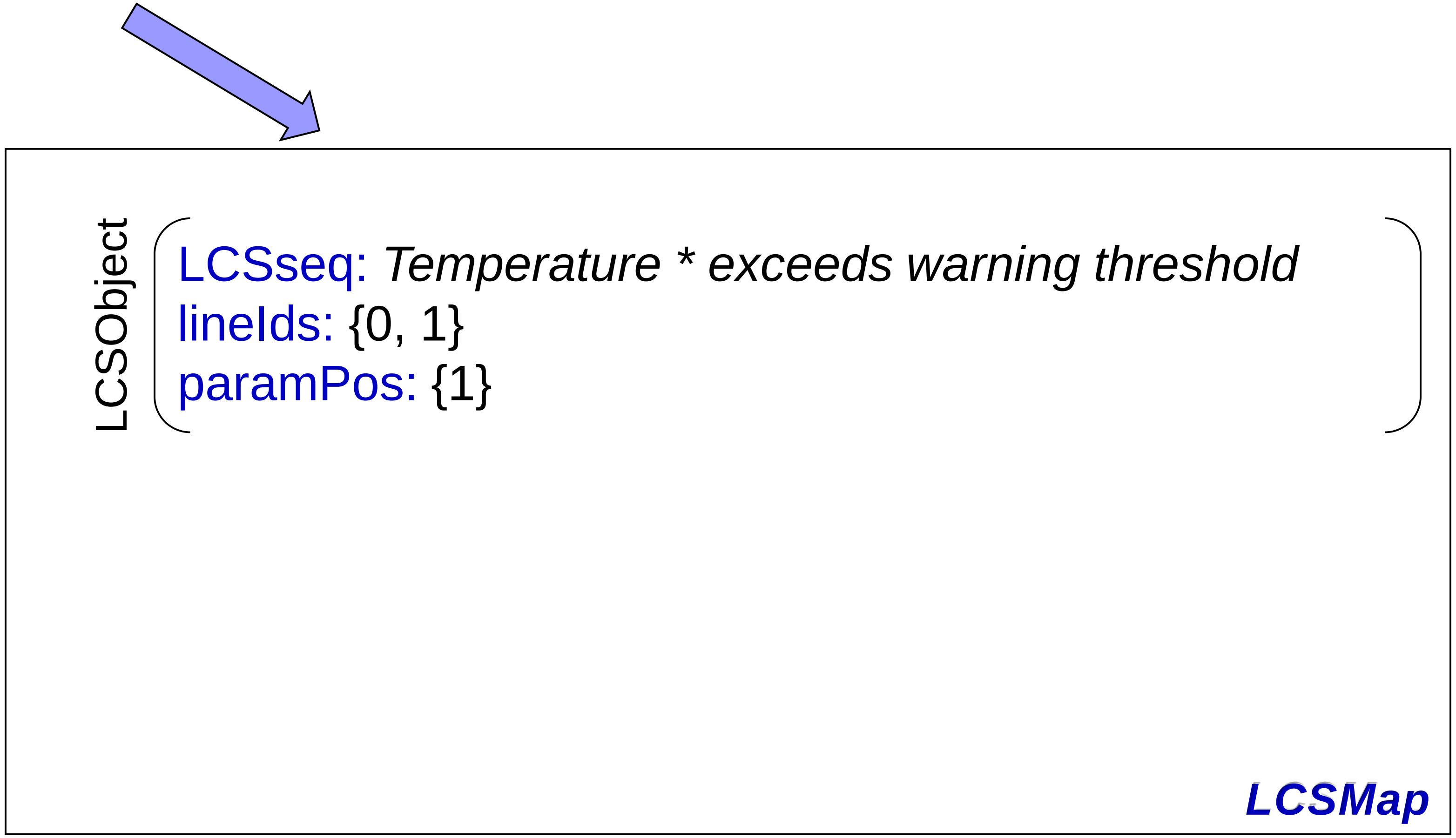
LCSseq: *Temperature * exceeds warning threshold*

lineIds: {0, 1}

paramPos: {1}

SPELL – Basic workflow

new log entry: *Command has completed successfully*



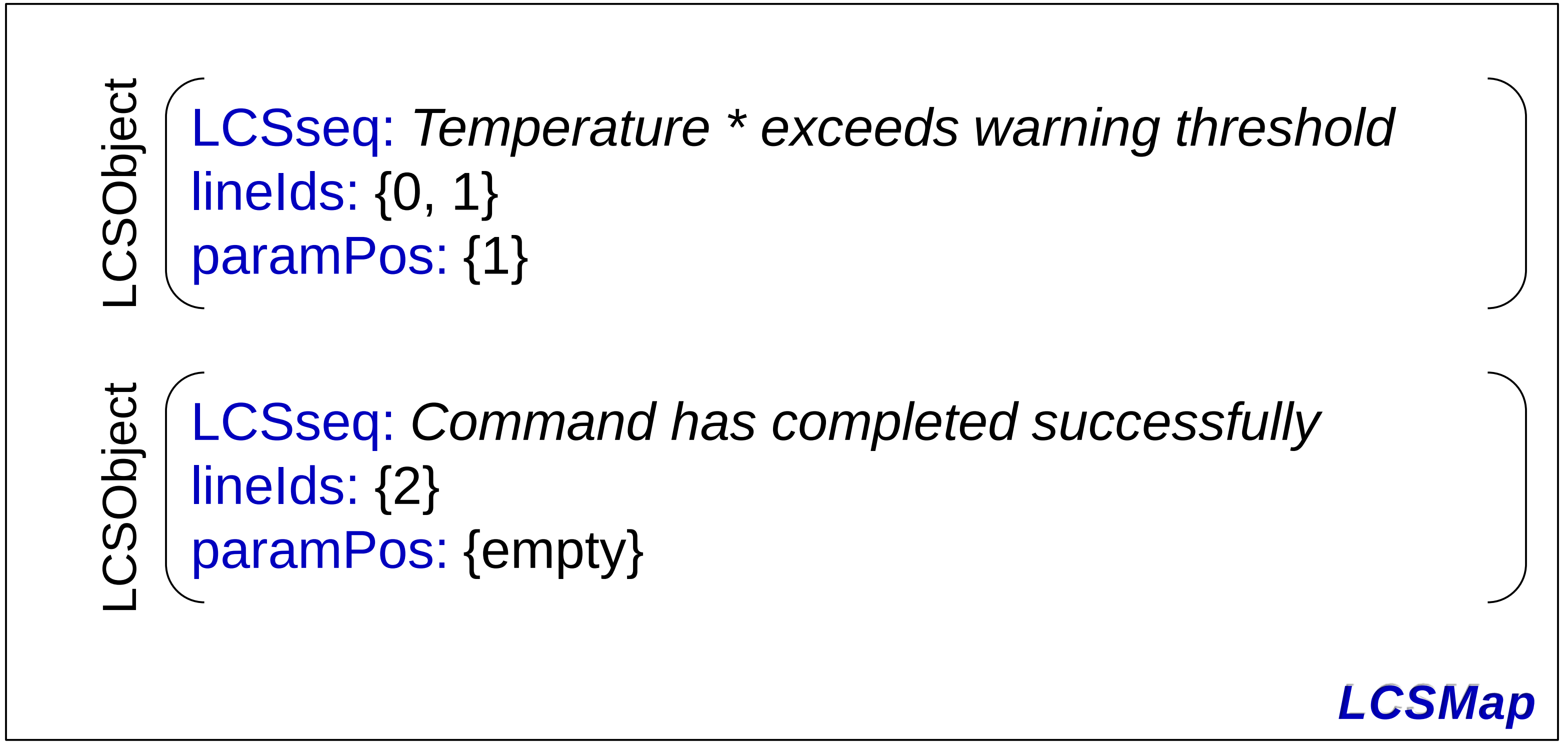
LCSObject

LCSseq: *Temperature * exceeds warning threshold*
lineIds: {0, 1}
paramPos: {1}

LCSMap

SPELL – Basic workflow

new log entry:



LCSObject

LCSseq: *Temperature * exceeds warning threshold*
lineIds: {0, 1}
paramPos: {1}

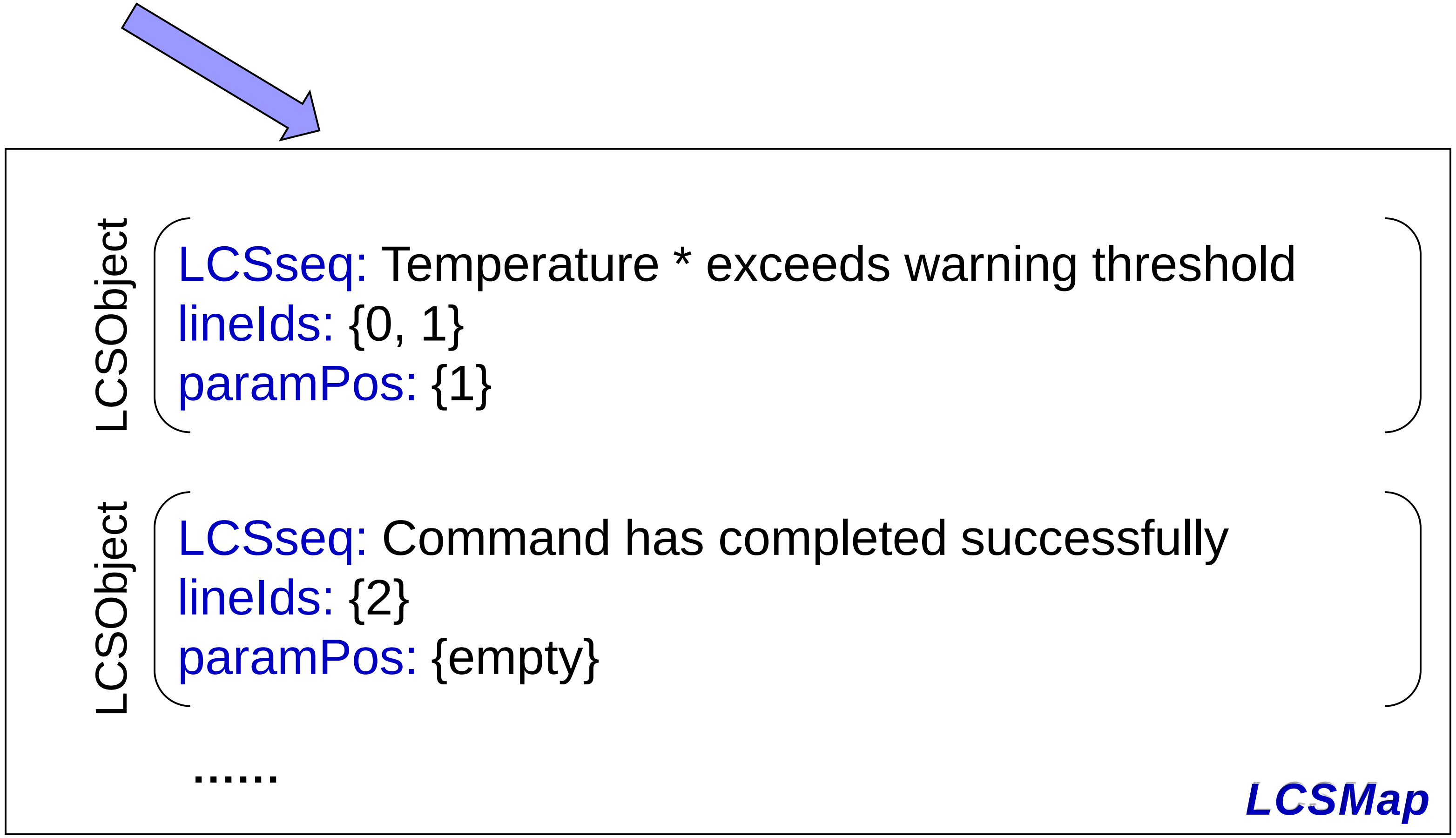
LCSObject

LCSseq: *Command has completed successfully*
lineIds: {2}
paramPos: {empty}

LCSMap

SPELL – Basic workflow

new log entry: *……*



LCSObject

LCSseq: Temperature * exceeds warning threshold
lineIds: {0, 1}
paramPos: {1}

LCSObject

LCSseq: Command has completed successfully
lineIds: {2}
paramPos: {empty}

……

LCSMap

SPELL – Improvement on efficiency

To compute LCS of two log entries, each one has $O(n)$ length:

SPELL – Improvement on efficiency

To compute LCS of two log entries, each one has $O(n)$ length:

Naïve way: Dynamic Programing

SPELL – Improvement on efficiency

To compute LCS of two log entries, each one has $\boldsymbol{O(n)}$ length:

Naïve way: Dynamic Programing

Time complexity:

To compare a log entry with an existing message type: $O(n^2)$

To compare a new log entry with $O(m)$ existing message types: $O(mn^2)$

SPELL – Improvement on efficiency

To compute LCS of two log entries, each one has $\boldsymbol{O(n)}$ length:

Naïve way: Dynamic Programing

Time complexity:

To compare a log entry with an existing message type: $O(n^2)$

To compare a new log entry with $O(m)$ existing message types: $O(mn^2)$

Can we do better?

SPELL – Improvement on efficiency

Observation.

For a complex system,

number of log entries: millions

number of message types: hundreds

SPELL – Improvement on efficiency

Observation.

For a complex system,

number of log entries: millions

number of message types: hundreds

For example:

Blue Gene/L log:

4,457,719 log entries, 394 message types

Hadoop log used in Xu'SOSP09:

11,197,705 log entries, only 29 message types

SPELL – Improvement on efficiency

Observation.

For a complex system,

number of log entries: millions

number of message types: hundreds

For example:

Blue Gene/L log:

4,457,719 log entries, 394 message types

Hadoop log used in Xu'SOSP09:

11,197,705 log entries, only 29 message types

For a majority of new log entries, their message types already exist in LCSMap!

SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

Existing message types:

A B C

A C D

A D

E F

SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

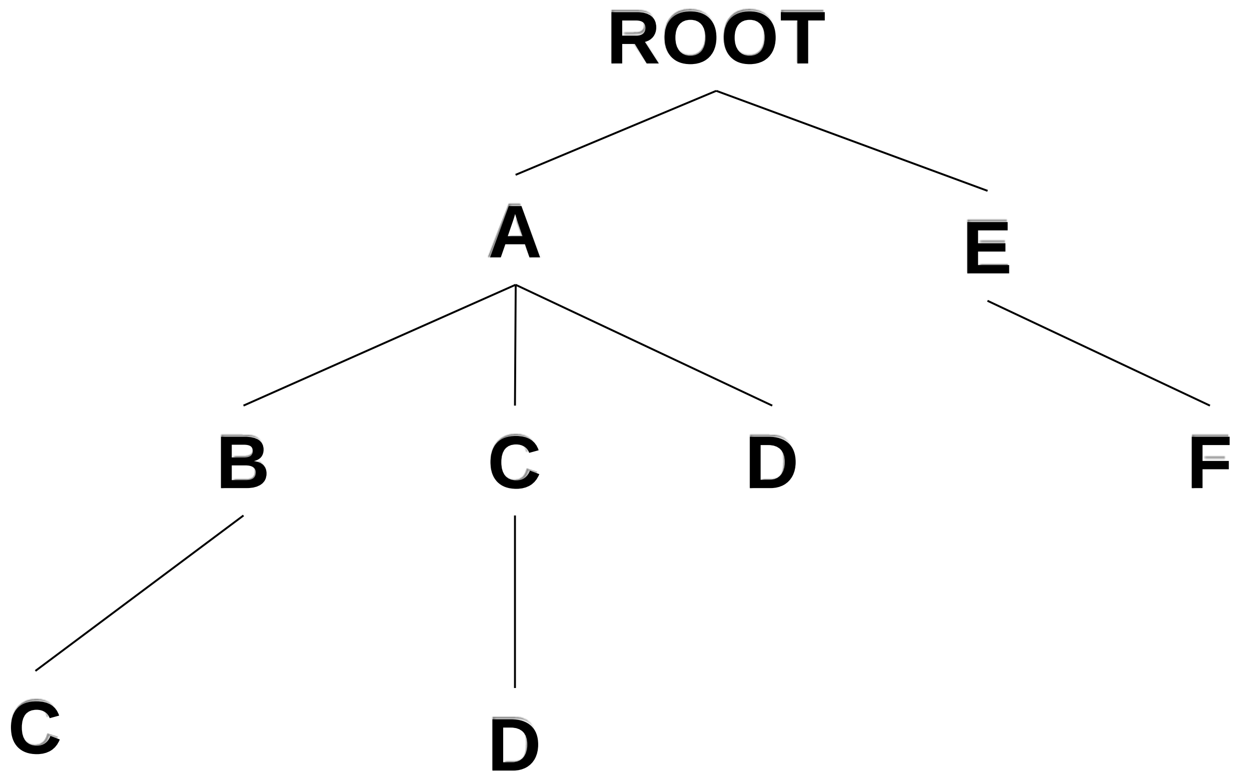
Existing message types:

A B C

A C D

A D

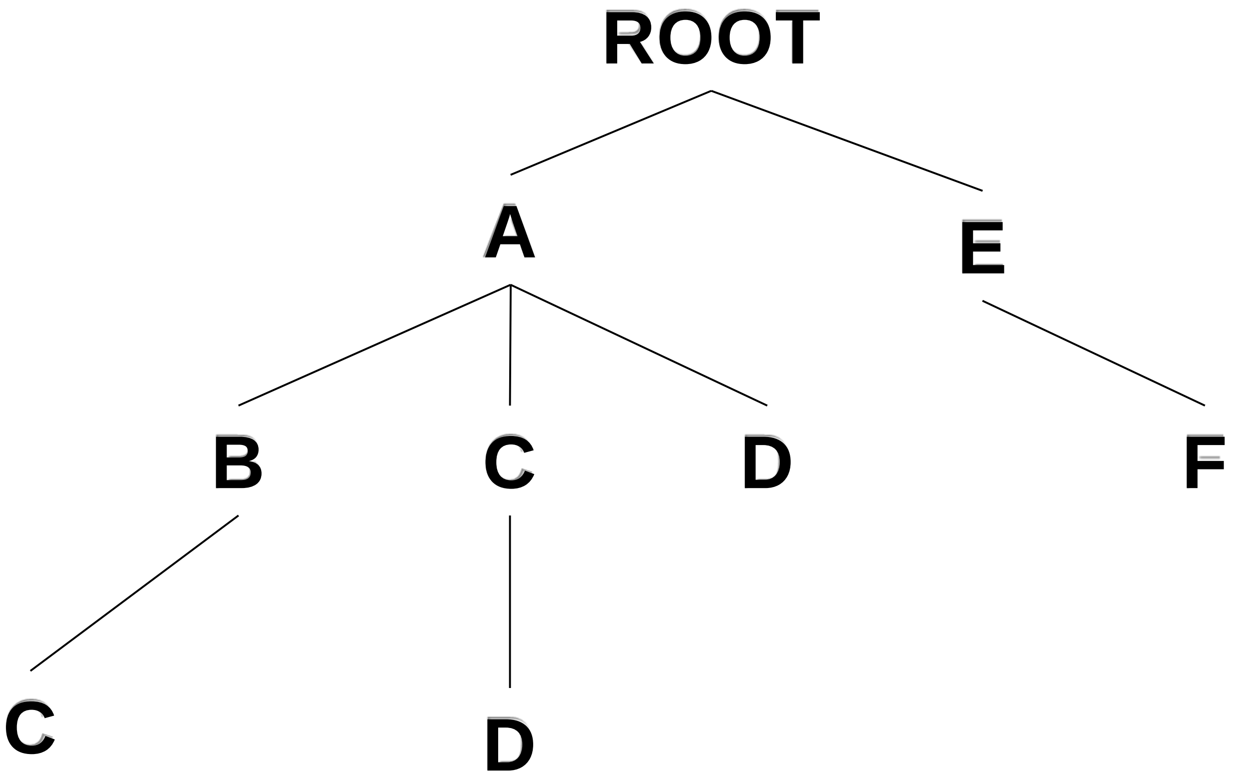
E F



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

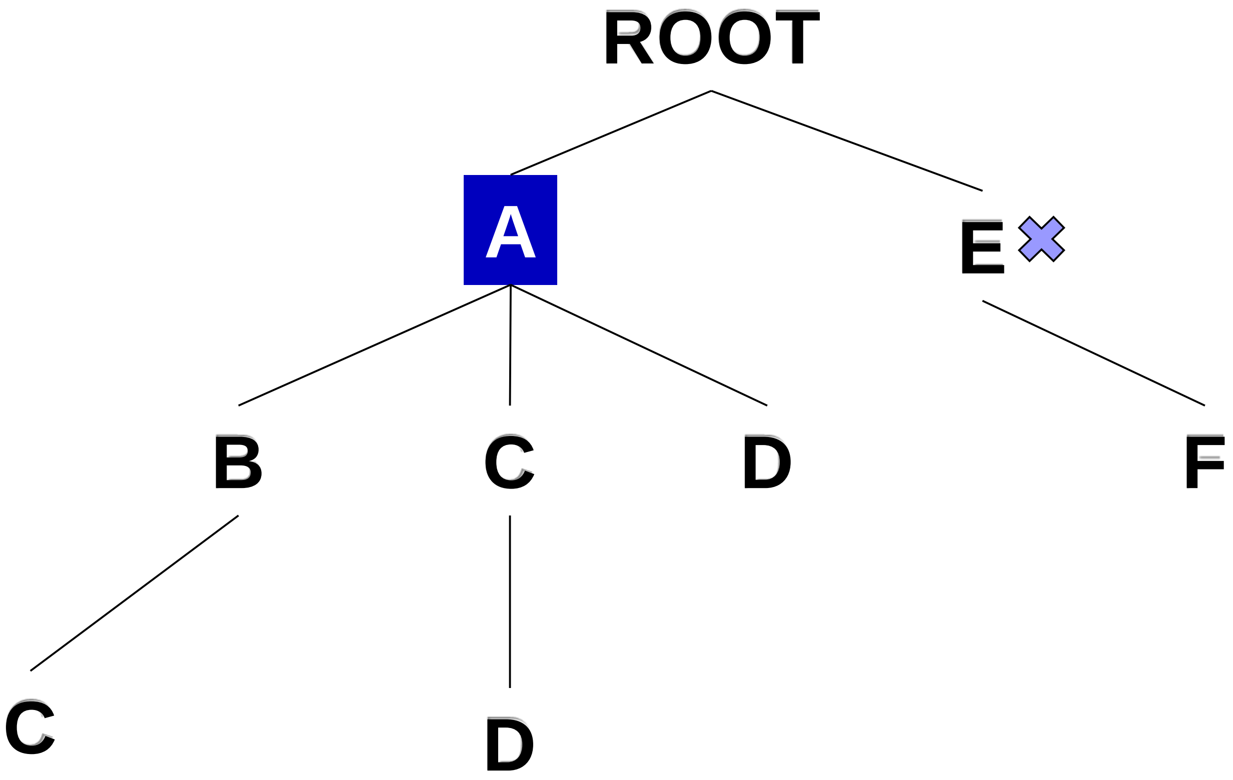
New log entry: *A B P C*



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

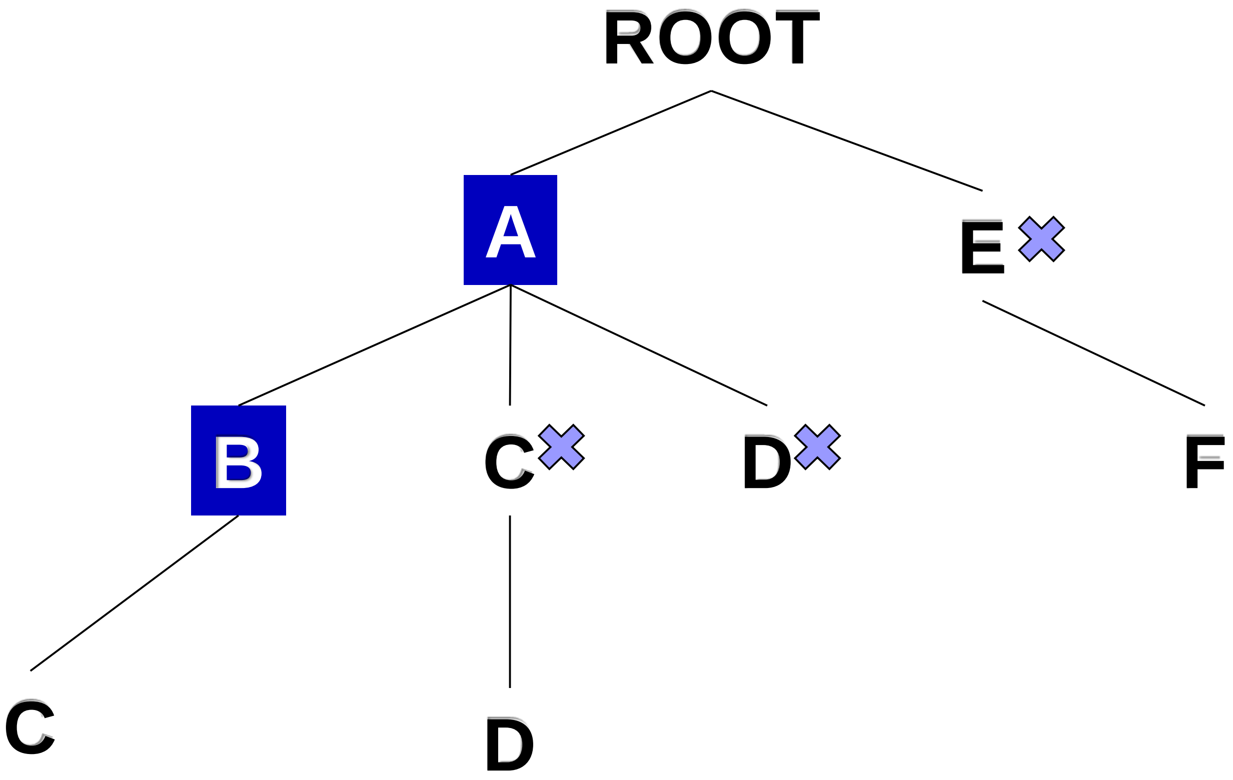
New log entry: **A** *B P C*



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

New log entry: **A** **B** *P* *C*

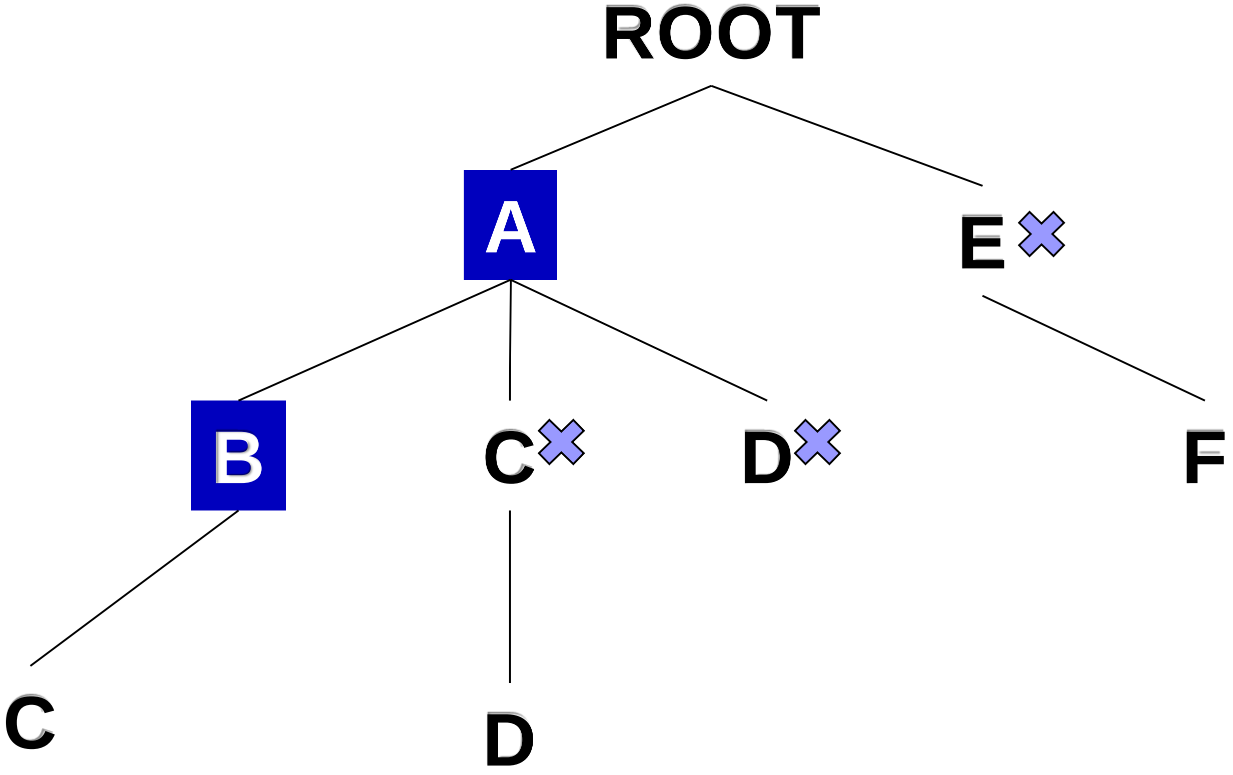


SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

New log entry: **A** **B** *P* *C*

Parameter:

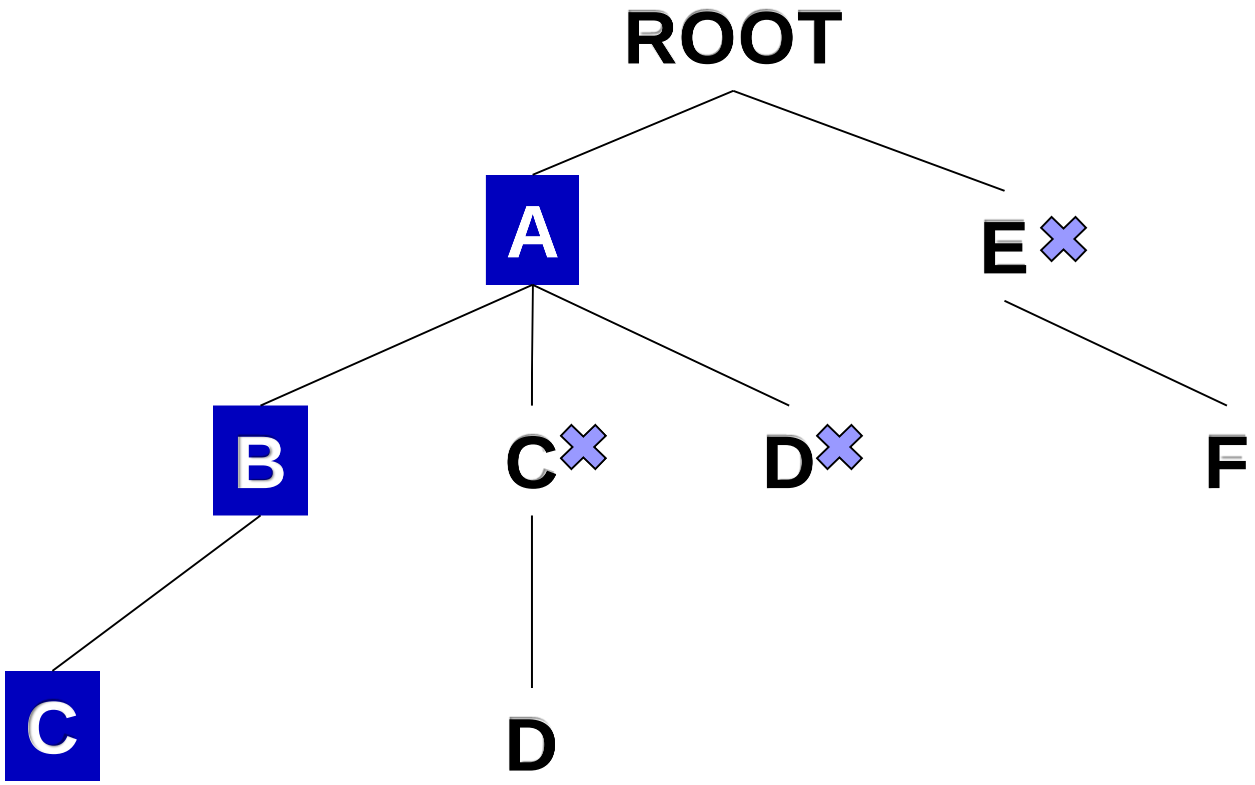


SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

New log entry: **A B** *P* **C**

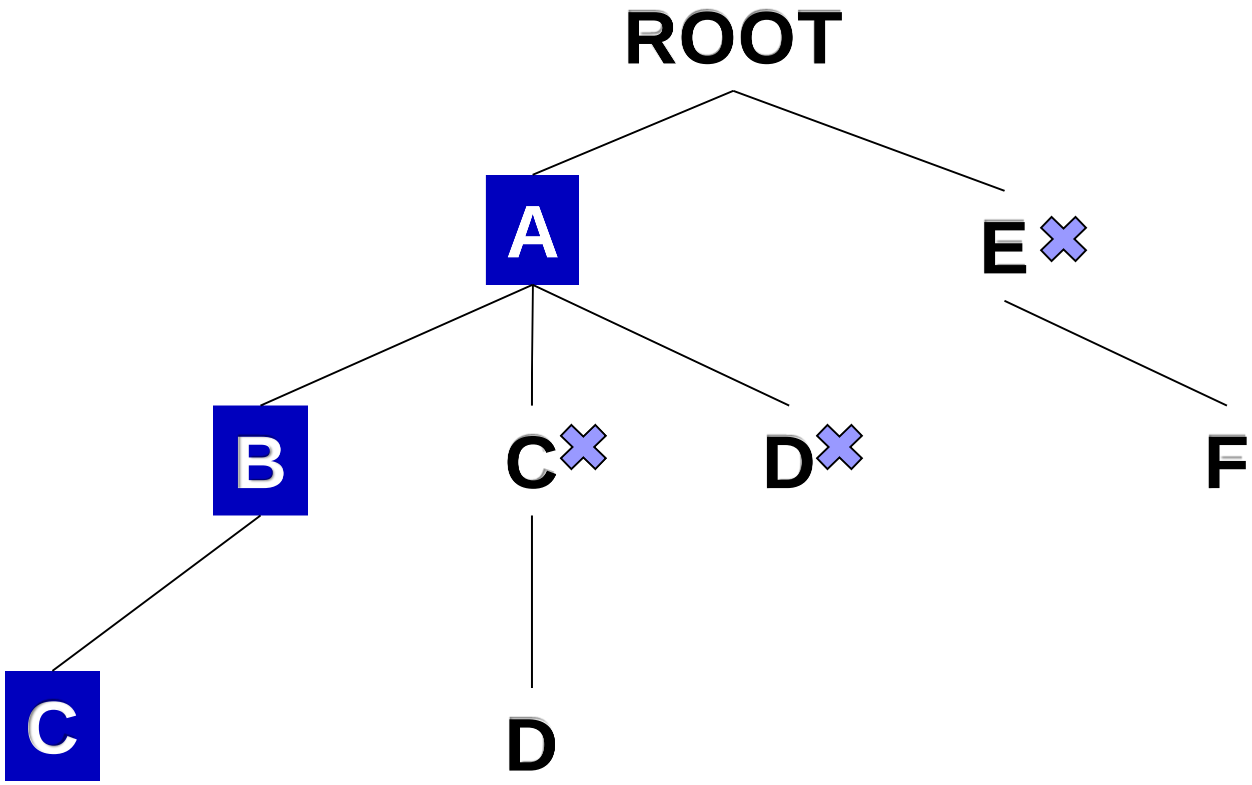
Parameter:



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

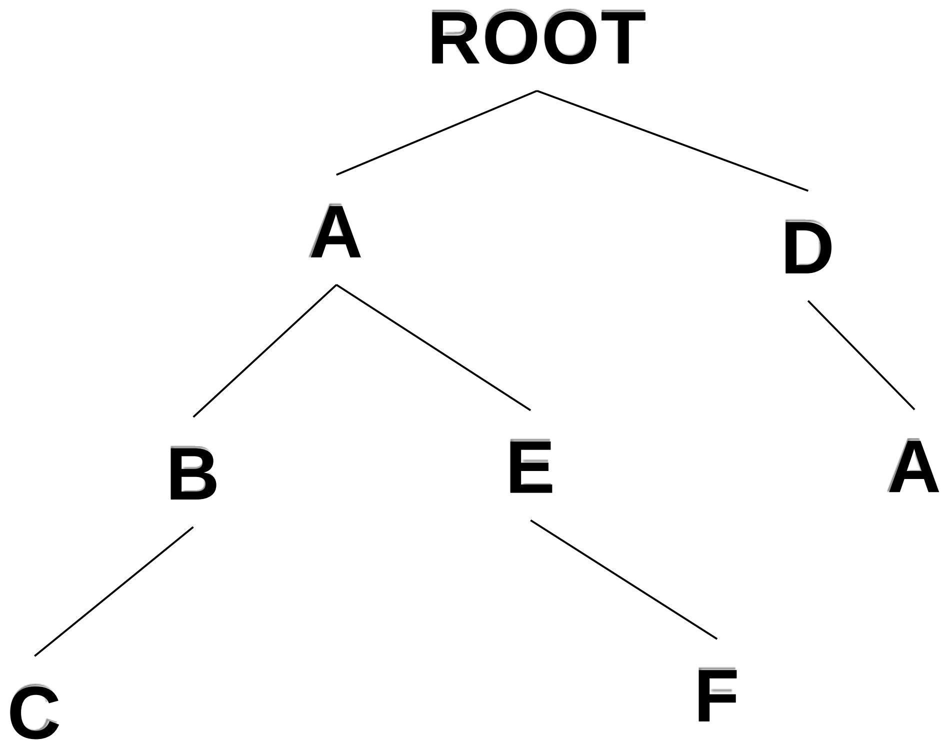
Time Complexity:
$\boldsymbol{O}(\boldsymbol{n})$ for each log entry



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

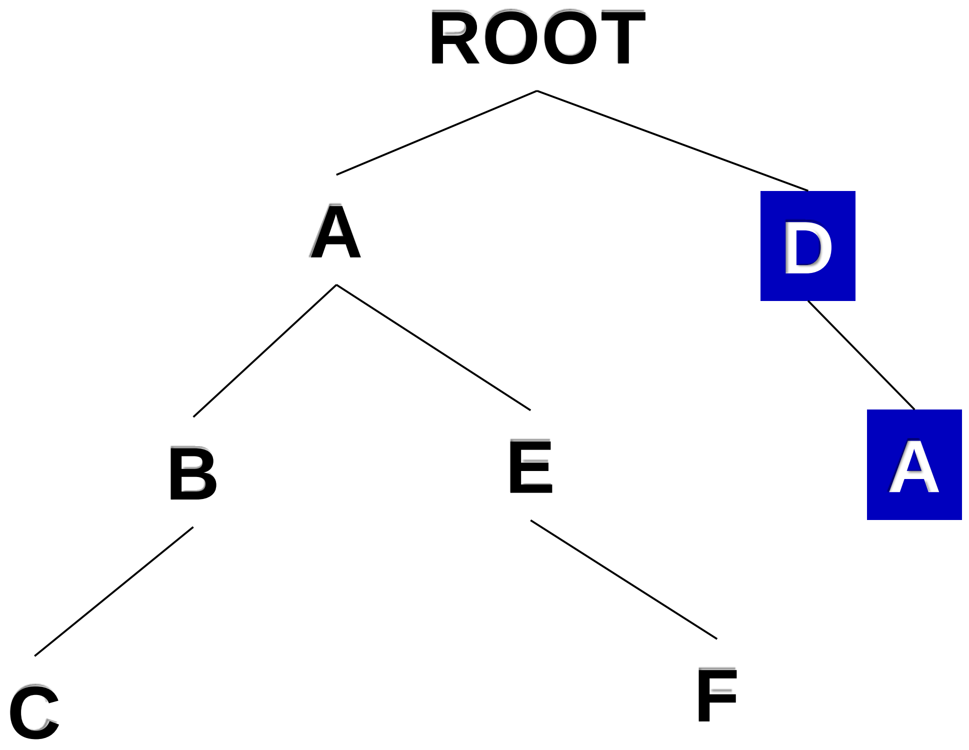
Problem:
New log entry: *D A P B C*



SPELL – Improvement on efficiency

Improvement 1: Prefix Tree

Problem:
New log entry: D A *P B C*
Matches D A



SPELL – Improvement on efficiency

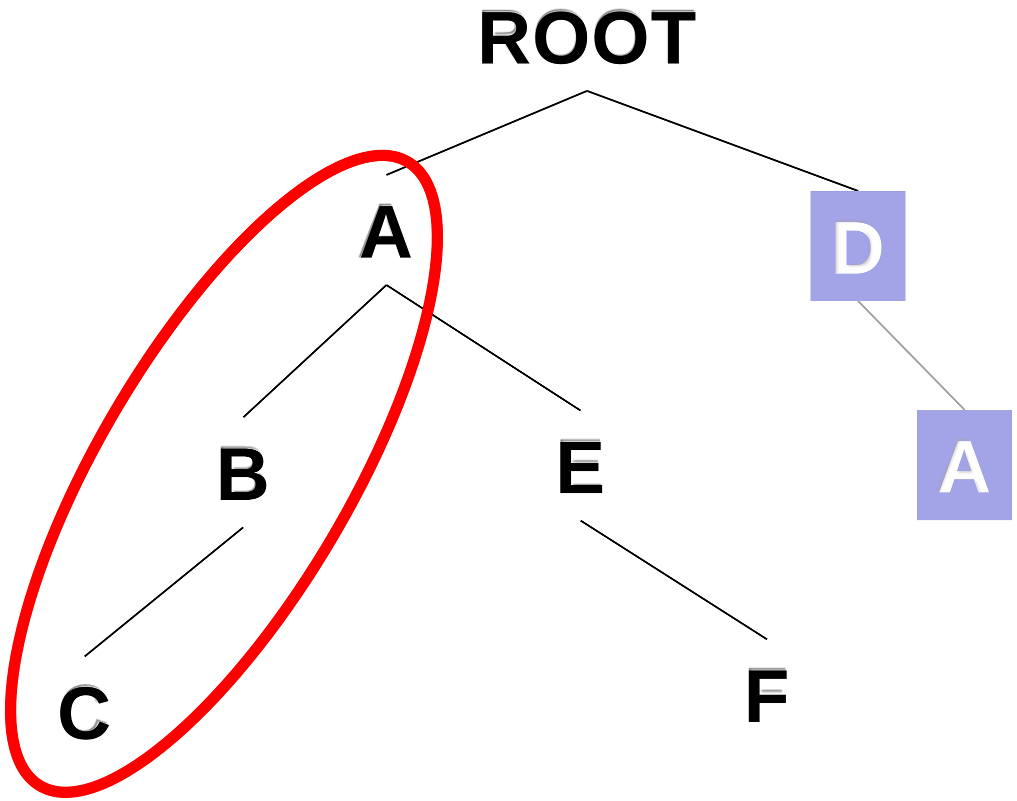
Improvement 1: Prefix Tree

Problem:

New log entry: D *A* *P* *B* *C*

Matches D A

Should be: A B C



SPELL – Improvement on efficiency

Improvement 2: Inverted Index

Existing message types:
A B C
A E F
D A

A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:

D
A
B
P
C

A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:

D
A
B
P
C

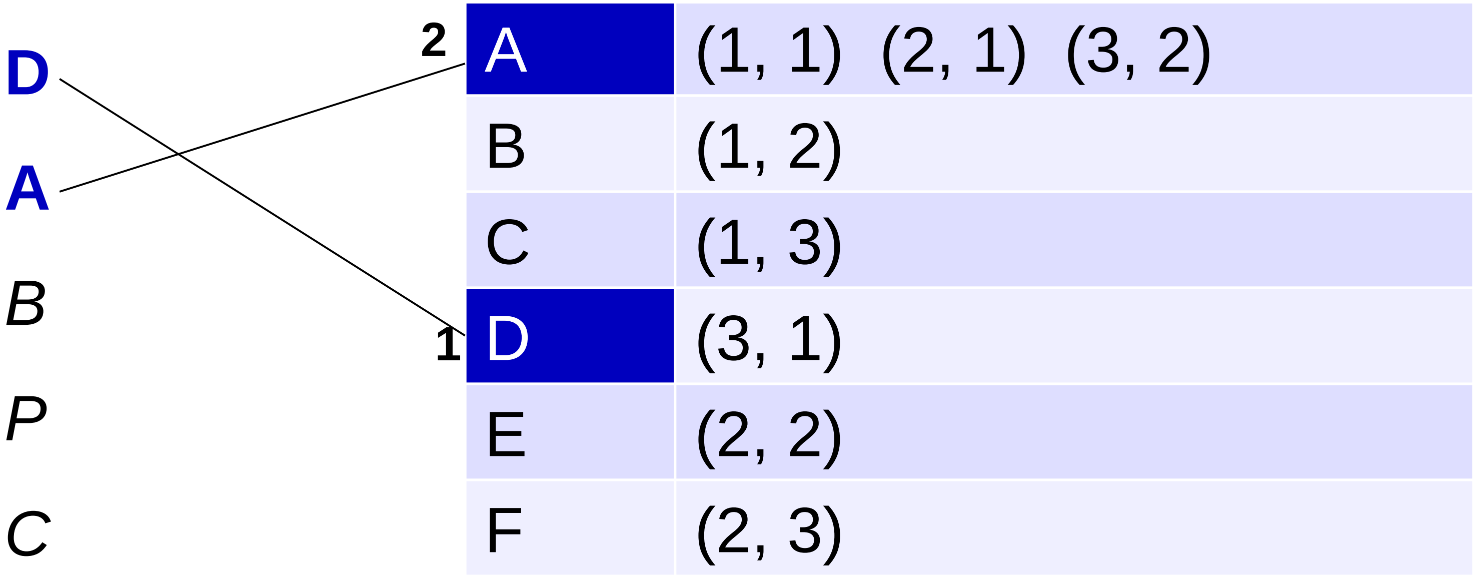
1

A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:

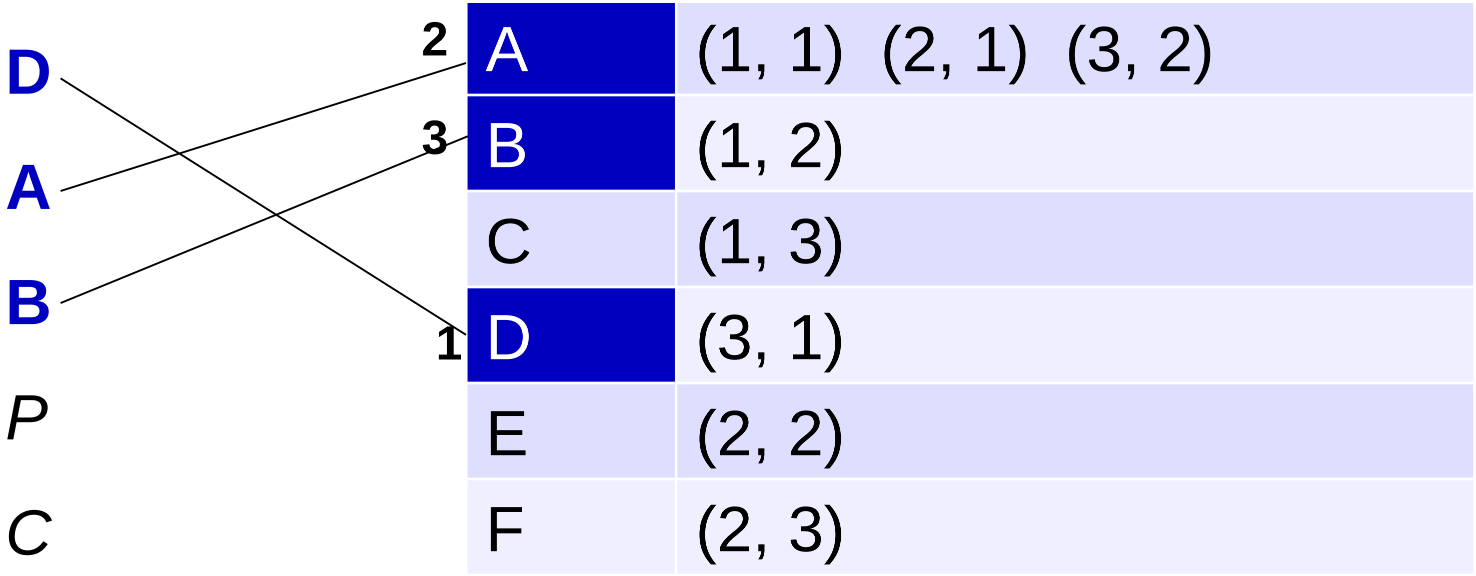


A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:

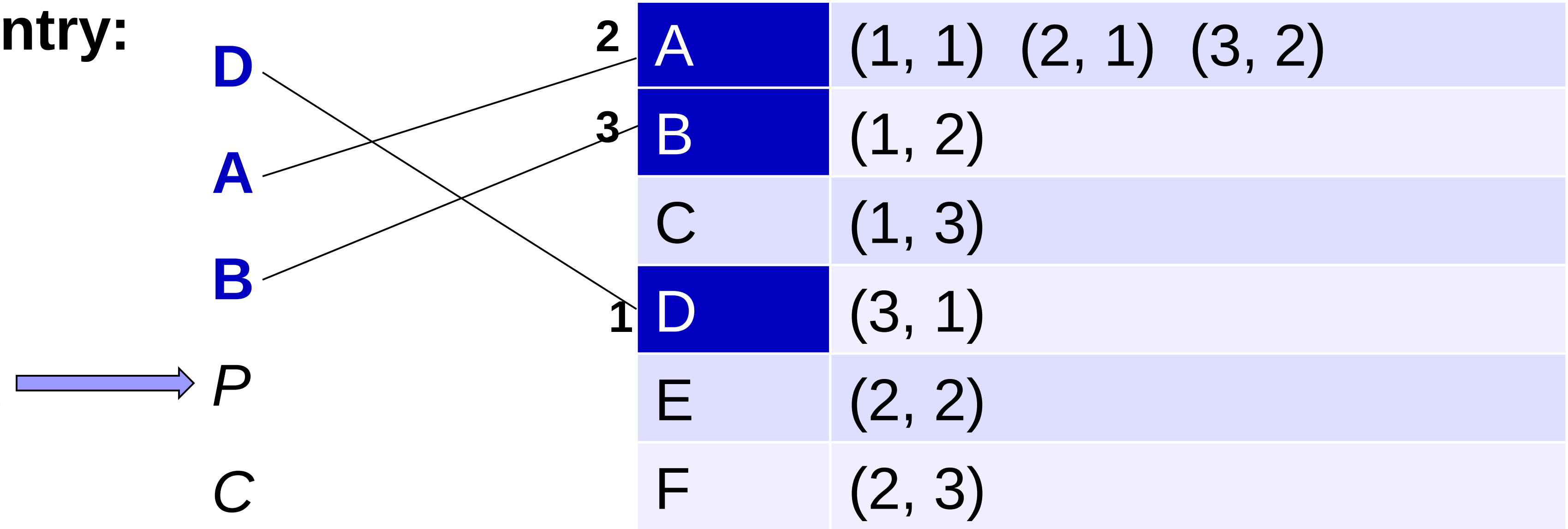


A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:



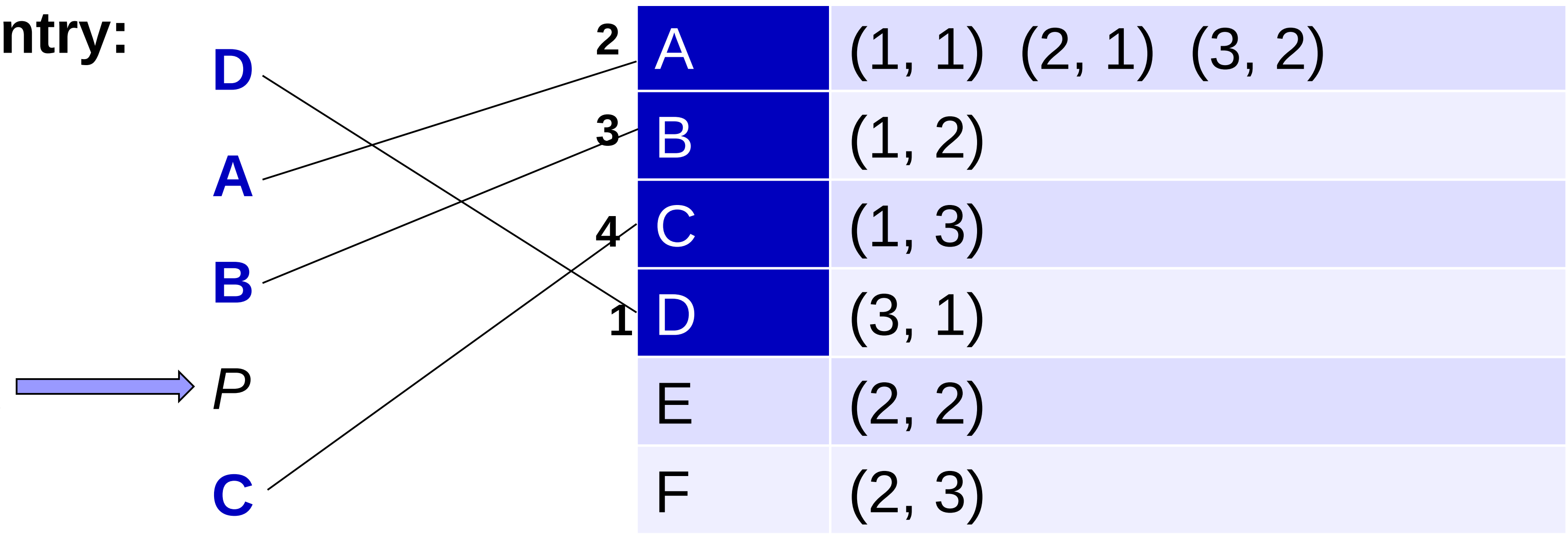
*No match,
parameter position*

A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:



No match,
parameter position

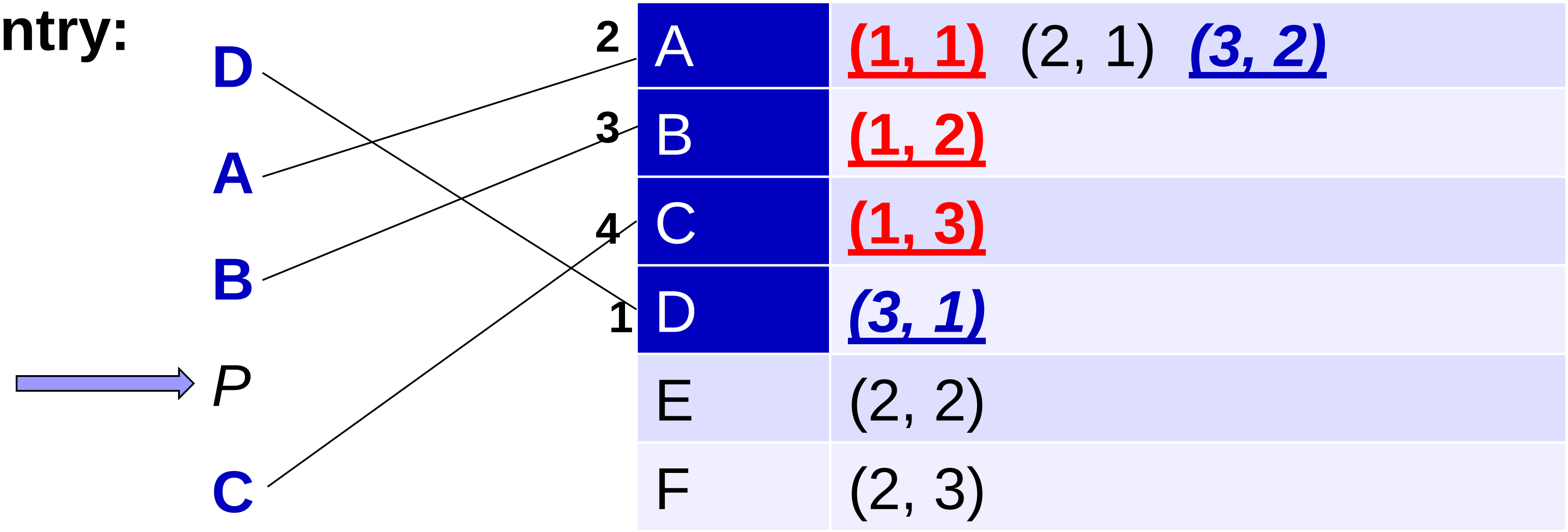
A	(1, 1) (2, 1) (3, 2)
B	(1, 2)
C	(1, 3)
D	(3, 1)
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

New log entry:

No match,
parameter position



A	**(1, 1)** (2, 1) ***(3, 2)***
B	**(1, 2)**
C	**(1, 3)**
D	***(3, 1)***
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

Time complexity: $O(cn)$
for each log entry, $c < m$

A	**(1, 1)** (2, 1) ***(3, 2)***
B	**(1, 2)**
C	**(1, 3)**
D	***(3, 1)***
E	(2, 2)
F	(2, 3)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

Time complexity: $O(cn)$
for each log entry, $c < m$

A	**(1, 1)** (2, 1) ***(3, 2)***
B	**(1, 2)**
C	**(1, 3)**
D	***(3, 1)***
E	(2, 2)
F	(2, 3)

For remaining log entries, compare it with each message type using simple DP.

SPELL – Improvement on effectiveness

We could be wrong.
More heuristics to adjust the result.

Example:

boot (command 2359) Error: Console-Busy Port already in use
wait (command 3964) Error: Console-Busy Port already in use

Initial message type:

***** *(command *) Error: Console-Busy Port already in use*

SPELL – Improvement on effectiveness

We could be wrong.
More heuristics to adjust the result.

Example:

boot (command 2359) Error: Console-Busy Port already in use
wait (command 3964) Error: Console-Busy Port already in use

Initial message type:

***** *(command *) Error: Console-Busy Port already in use*

Solution: Split heuristic

If a parameter position has very few unique tokens:

*boot (command *) Error: Console-Busy Port already in use*
*wait (command *) Error: Console-Busy Port already in use*

SPELL – Improvement on effectiveness

We could be wrong.
More heuristics to adjust the result.

Example:

Fan speeds (3552 3534 3375 4354 3515 3479)
Fan speeds (3552 3534 3375 4299 3515 3479)
Fan speeds (3552 3552 3391 4245 3515 3497)
Fan speeds (3534 3534 3375 4245 3497 3479)
Fan speeds (3534 3534 3375 4066 3497 3479)

Initial message type:

*Fan speeds (* 3552 * 3515 *)*
*Fan speeds (3534 3534 3375 * 3497 3479)*

SPELL – Improvement on effectiveness

We could be wrong.
More heuristics to adjust the result.

Example:
Fan speeds (3552 3534 3375 4354 3515 3479)
Fan speeds (3552 3534 3375 4299 3515 3479)
Fan speeds (3552 3552 3391 4245 3515 3497)
Fan speeds (3534 3534 3375 4245 3497 3479)
Fan speeds (3534 3534 3375 4066 3497 3479)

Initial message type:
*Fan speeds (* 3552 * 3515 *)*
*Fan speeds (3534 3534 3375 * 3497 3479)*

Solution: Merge heuristic
Merge similar message types together:
*Fan speeds: (*)*

Evaluation

Methods to compare:

IPLoM (Makanju’KDD09):

Partition log file using 3-step heuristics (log entry length, etc.)

CLP (Fu’ICDM09)

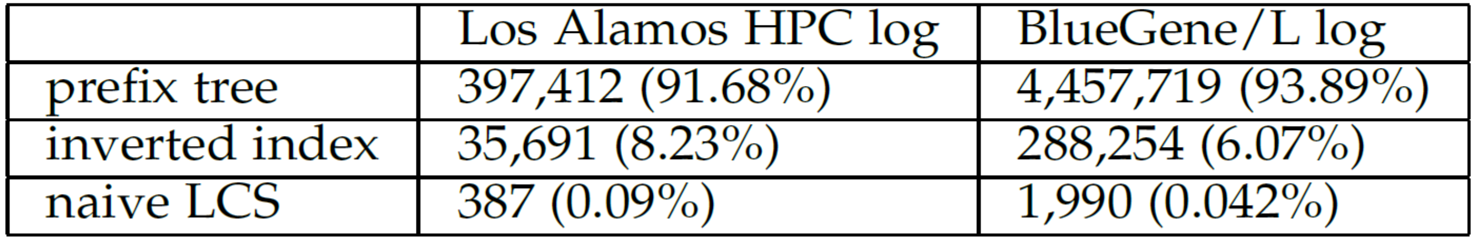
Cluster similar logs together based on weighted edit distance

Log dataset:

Log type	Count	Message type ground truth
Los Alamos HPC log	433,490	Available online
BlueGene/L log	4,747,963	Available online
Openstack Cloud log	87,519	Manually parsed from source code

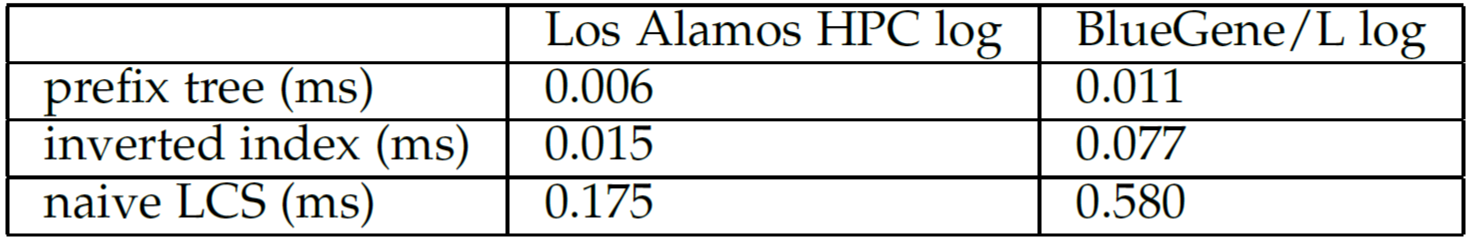
Evaluation - Efficiency

Number (Percentage) of log entries returned by each step



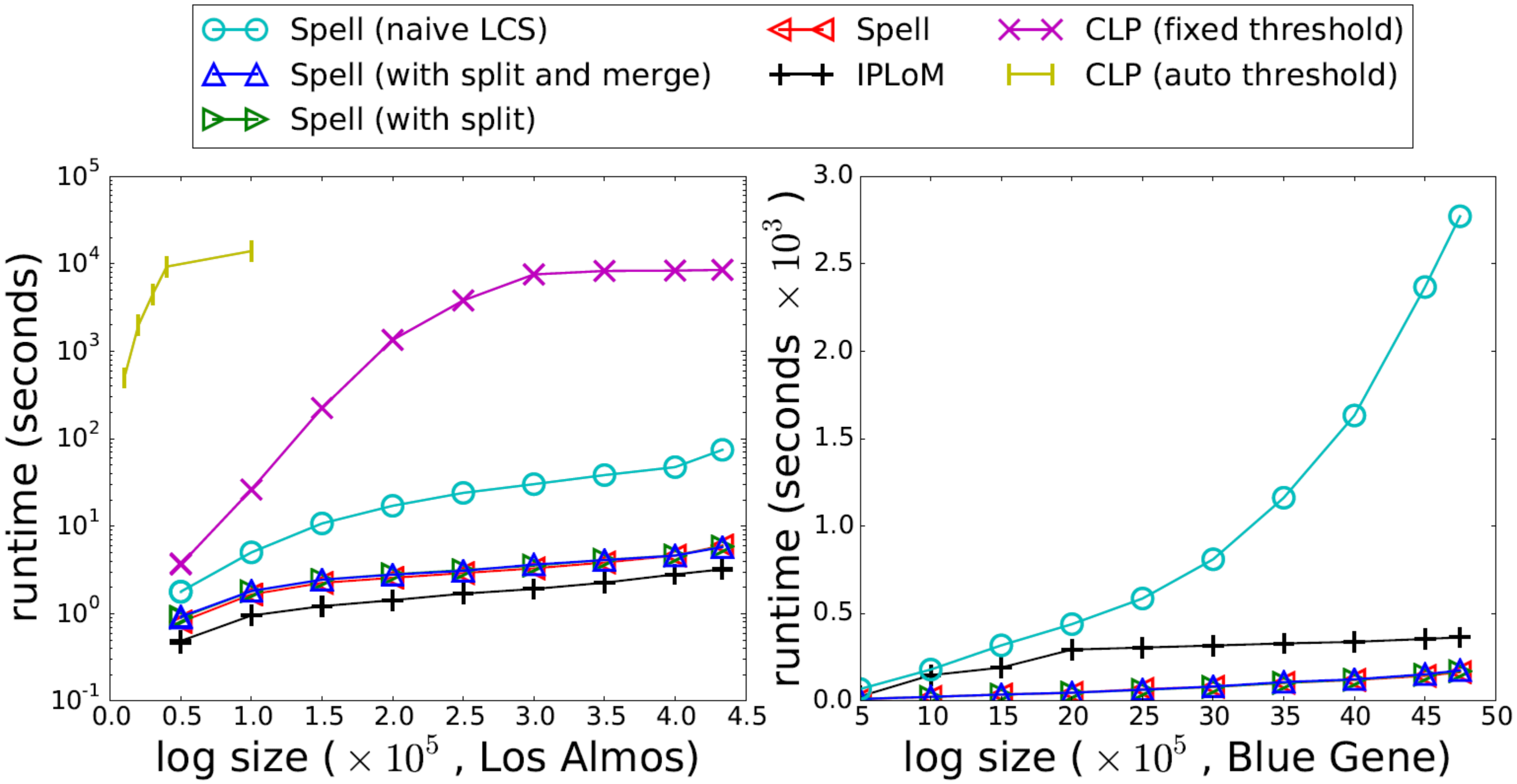
	Los Alamos HPC log	BlueGene/L log
prefix tree	397,412 (91.68%)	4,457,719 (93.89%)
inverted index	35,691 (8.23%)	288,254 (6.07%)
naive LCS	387 (0.09%)	1,990 (0.042%)

Amortized cost of each message type lookup step in Spell

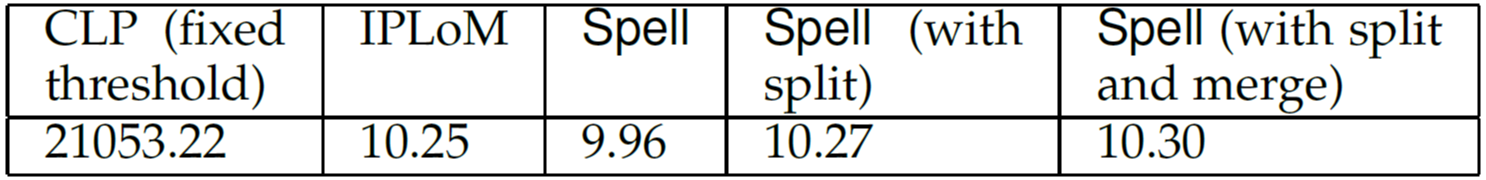


	Los Alamos HPC log	BlueGene/L log
prefix tree (ms)	0.006	0.011
inverted index (ms)	0.015	0.077
naive LCS (ms)	0.175	0.580

Evaluation - Efficiency

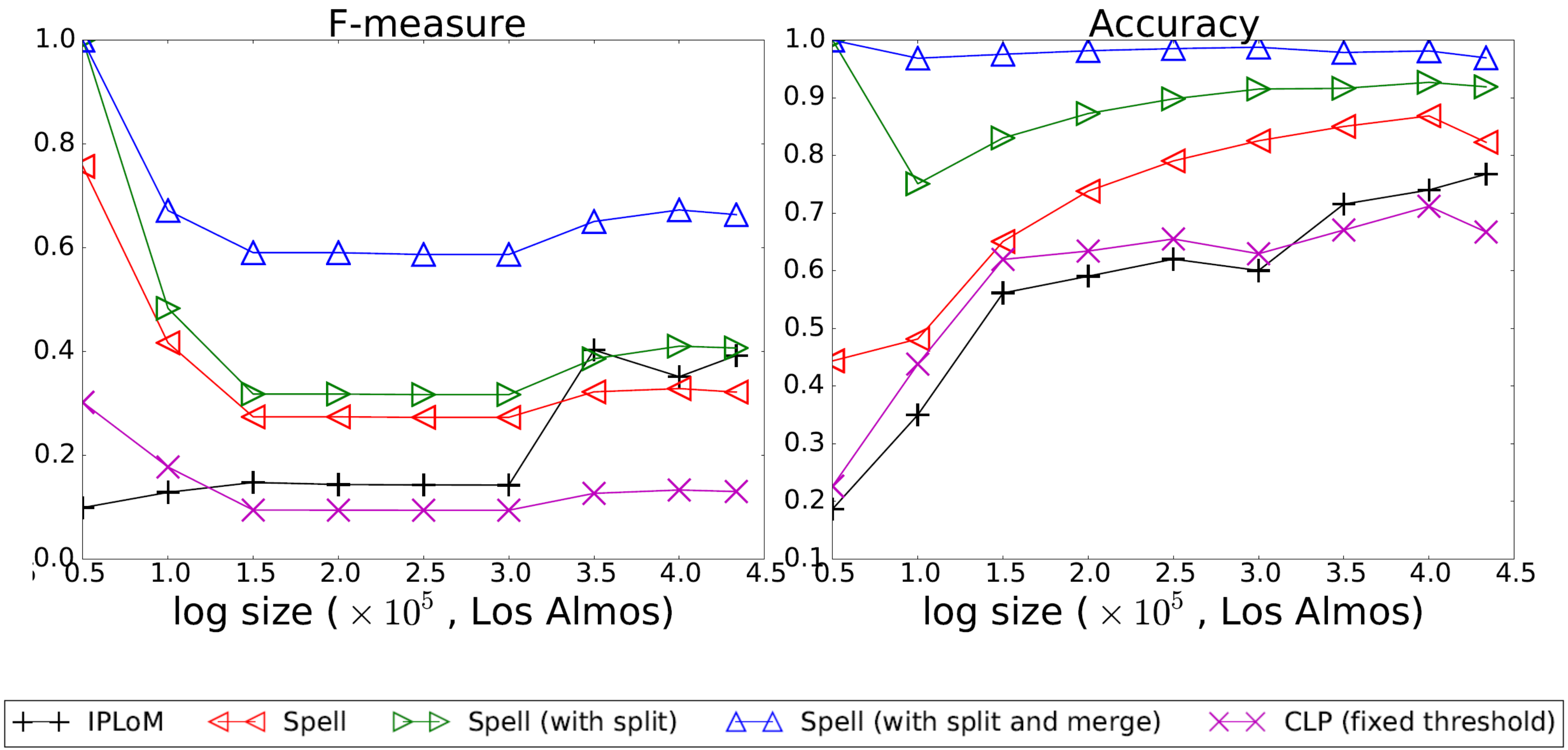


Openstack:

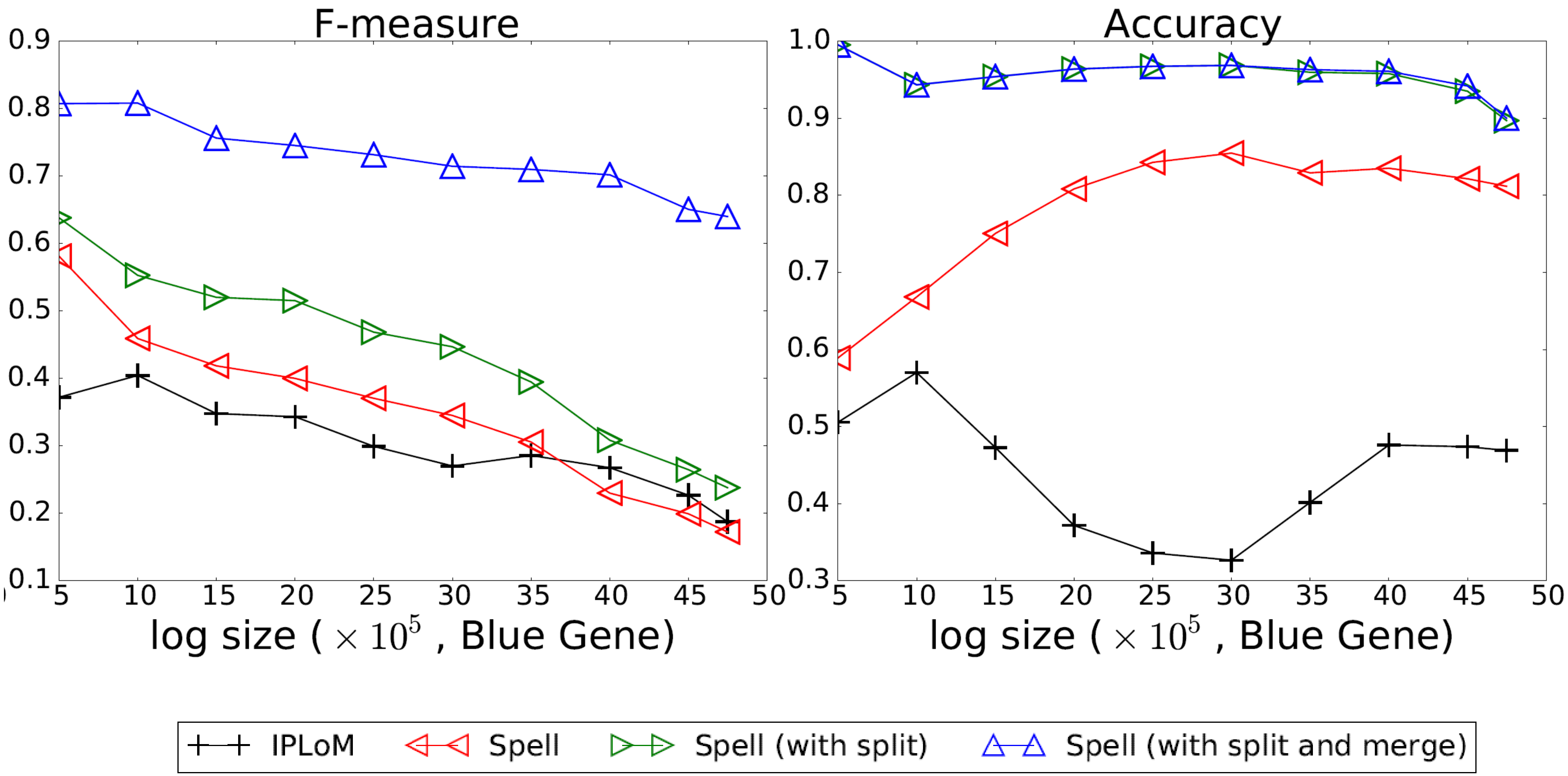


CLP (fixed threshold)	IPLoM	Spell	Spell (with split)	Spell (with split and merge)
21053.22	10.25	9.96	10.27	10.30

Evaluation - Effectiveness



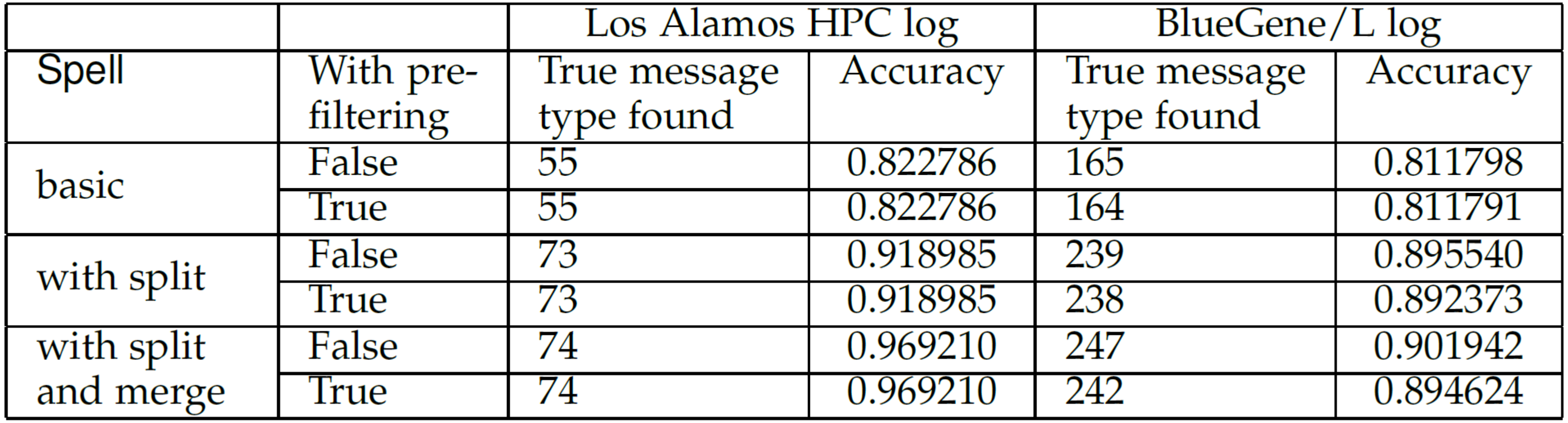
Evaluation - Effectiveness



Thank you!

Evaluation - Effectiveness

Comparison of Spell with and without pre-filter



		Los Alamos HPC log		BlueGene/L log	
Spell	With pre-filtering	True message type found	Accuracy	True message type found	Accuracy
basic	False	55	0.822786	165	0.811798
	True	55	0.822786	164	0.811791
with split	False	73	0.918985	239	0.895540
	True	73	0.918985	238	0.892373
with split and merge	False	74	0.969210	247	0.901942
	True	74	0.969210	242	0.894624

Evaluation

Methods to compare:

IPLoM (Makanju'KDD09):

Partition log file using 3-step heuristics (log entry length, etc.)

CLP (Fu'ICDM09)

Cluster similar logs together based on weighted edit distance

Log dataset:

Log type	Source	Count	Message type ground truth
Los Alamos HPC log	Available online	433,490	Available online
BlueGene/L log	Available online	4,747,963	Available online
Openstack Cloud log	Generated using CloudLab	87,519	Manually parsed from source code

Our approach

Spell, **a structured *S*treaming *P*arser for *E*vent *L*ogs using an *L*CS (longest common subsequence) based approach.**

LCS of two sequences:

The longest subsequence common to both sequences.

Our approach

Spell, a structured ***S***treaming ***P***arser for ***E***vent ***L***ogs using an ***L***CS (longest common subsequence) based approach.

LCS of two sequences:

The longest subsequence common to both sequences.

E.g. LCS of:

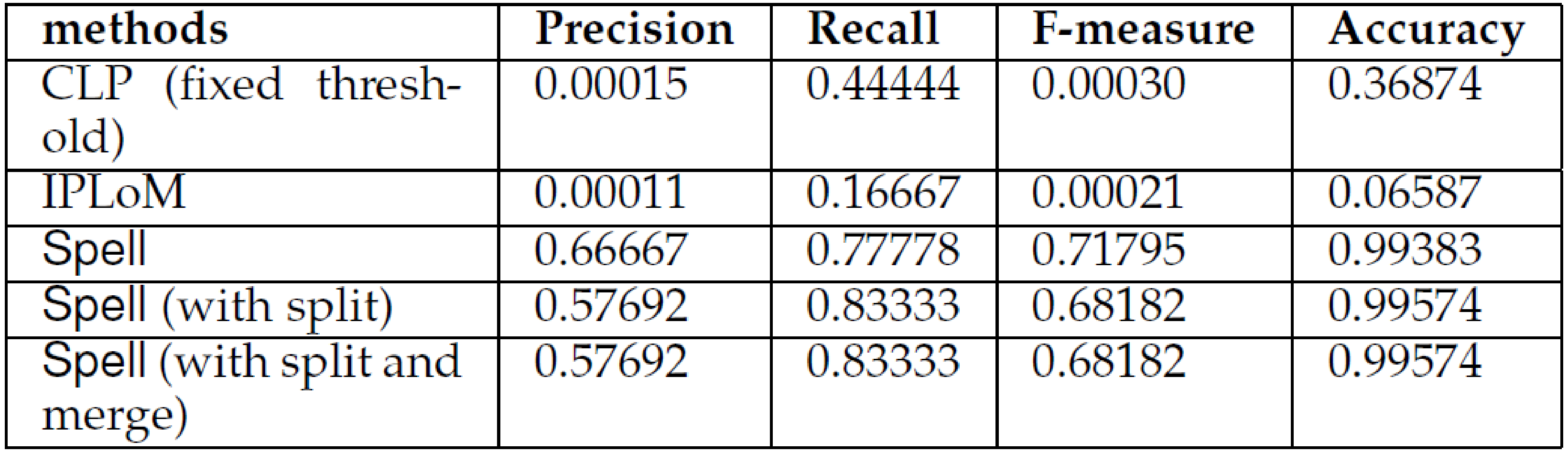
1, 3, 5, 7, 9
1, 5, 7, 10

equals:

1, 5, 7

Evaluation - Effectiveness

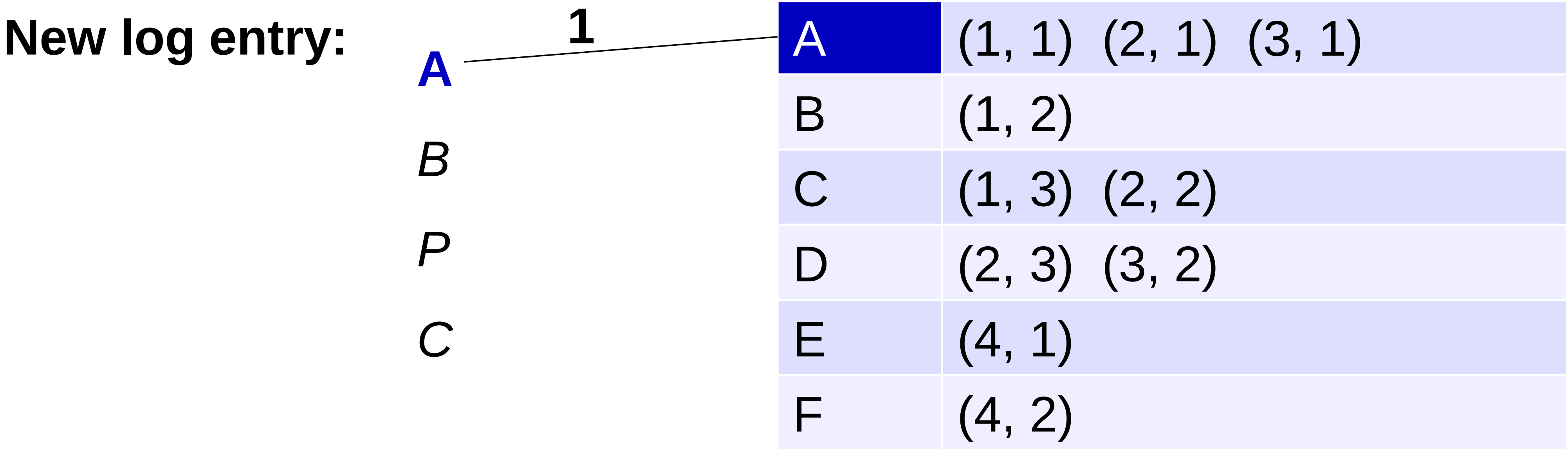
Effectiveness measures on Openstack Log



methods	Precision	Recall	F-measure	Accuracy
CLP (fixed threshold)	0.00015	0.44444	0.00030	0.36874
IPLoM	0.00011	0.16667	0.00021	0.06587
Spell	0.66667	0.77778	0.71795	0.99383
Spell (with split)	0.57692	0.83333	0.68182	0.99574
Spell (with split and merge)	0.57692	0.83333	0.68182	0.99574

SPELL – Improvement on efficiency

Improvement 2: Inverted Index



New log entry:

A

B

P

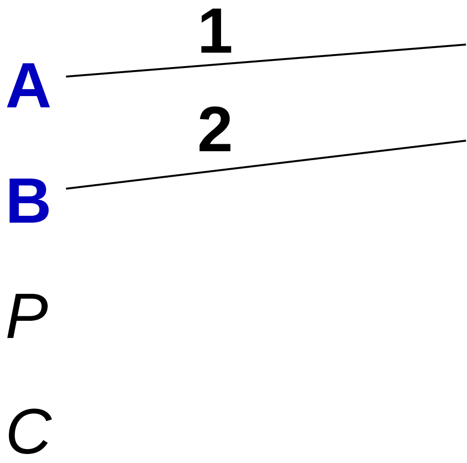
C

A	(1, 1) (2, 1) (3, 1)
B	(1, 2)
C	(1, 3) (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

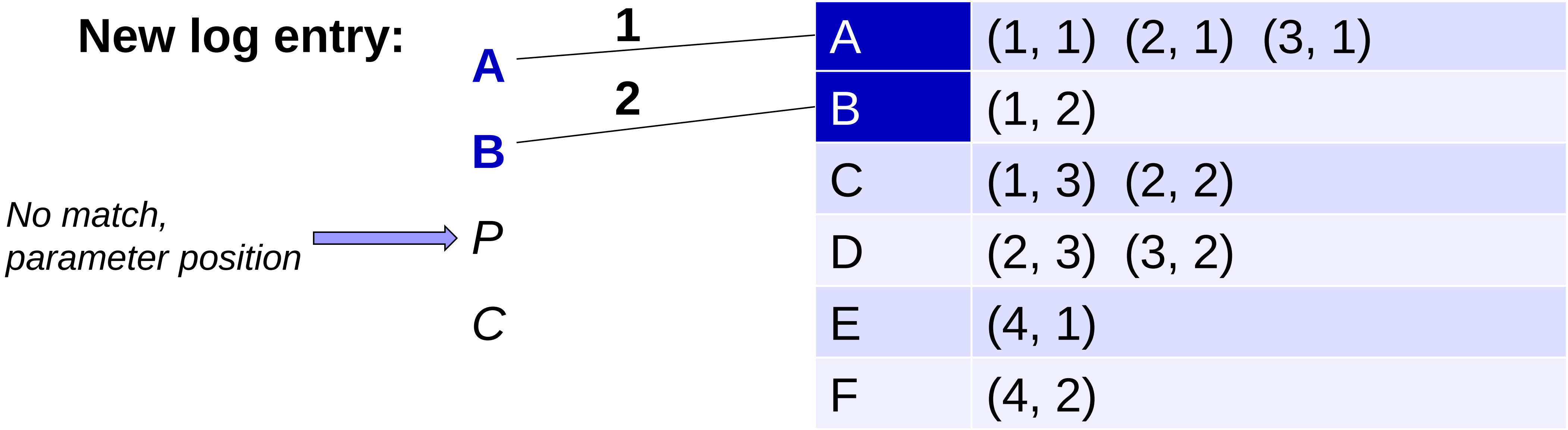
New log entry:



A	(1, 1) (2, 1) (3, 1)
B	(1, 2)
C	(1, 3) (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index



New log entry:

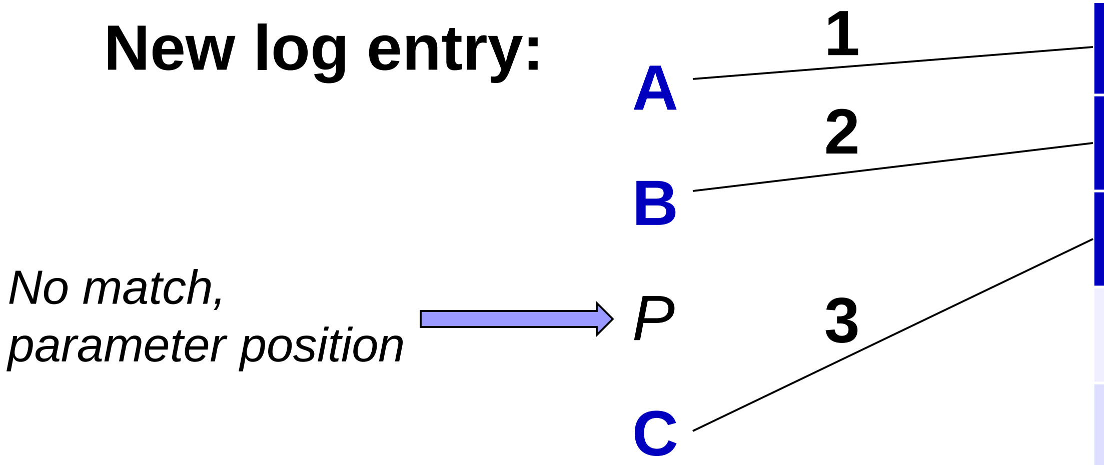
A
B
P
C

No match,
parameter position

A	(1, 1) (2, 1) (3, 1)
B	(1, 2)
C	(1, 3) (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

SPELL – Improvement on efficiency

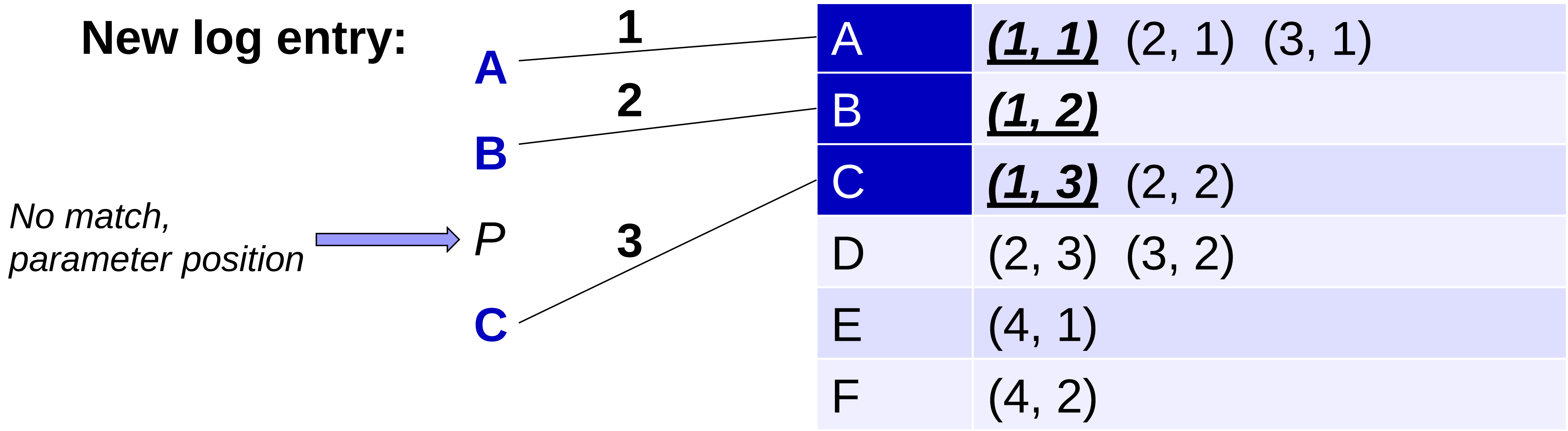
Improvement 2: Inverted Index



A	(1, 1) (2, 1) (3, 1)
B	(1, 2)
C	(1, 3) (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index



New log entry:

No match,
parameter position

A	***(1, 1)*** (2, 1) (3, 1)
B	***(1, 2)***
C	***(1, 3)*** (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

SPELL – Improvement on efficiency

Improvement 2: Inverted Index

Time complexity: $O(cn)$
for each log entry, $c < m$

A	(1, 1) (2, 1) (3, 1)
B	(1, 2)
C	(1, 3) (2, 2)
D	(2, 3) (3, 2)
E	(4, 1)
F	(4, 2)

For remaining log entries, compare it with each message type using simple DP.