CPSC 426/526
Handling Configuration Errors

Ennan Zhai
Computer Science Department
Yale University

Credits: Some slides from Tianyin Xu’s talks
Recall: Lec-15

• In lec-15, we learned:
  - Failure root causes in distributed systems
  - How to handle failures in general
  - Case study: Failure-handling code errors
Lecture Roadmap

• What is the configuration?
• Handling configuration errors
• Case Study: Spex
• Case Study: ConfigV
How many of you have made mistakes when configuring systems?
Root causes of “high-severity” issues in major storage companies. In SOSP’11.
You are not alone!

We talk about software configuration today!

- Misconfiguration: 31%
- Bugs: 20%
- Hardware: 15%
- Customer Environment: 25%
- 9%

Root causes of “high-severity” issues in major storage companies. In SOSP’11.
MySQL Configuration File

Configuration File

MySQL

Specify parameters

```
datadir = /var/lib/mysql
tmpdir  = /tmp
lc-messages-dir = /usr/share/mysql
skip-external-locking
#
# Instead of skip-networking the default is now to listen only on
# localhost which is more compatible and is not less secure.
bind-address = 127.0.0.1
#
# * Fine Tuning
#
key_buffer = 16M
max_allowed_packet = 16M
thread_stack = 192K
thread_cache_size = 8
#
# This replaces the startup script and checks MyISAM tables if needed
# the first time they are touched
myisam-recover = BACKUP
#max_connections = 100
#table_cache = 64
#thread_concurrency = 10
#
# * Query Cache Configuration
#
query_cache_limit = 1M
query_cache_size = 16M
#
# * Logging and Replication
```
<Datacenter Location="A" ProxyIPRange="10.25.252.9/29">
  <MachinePool Name="A0" FillFactor="0.8"/>
    <Vlan Name="301"/>
  </MachinePool>
  <MachinePool Name="A2" FillFactor="0.8"/>
    <Vlan Name="401"/>
  </MachinePool>
</Datacenter>

<Cluster Name="C101">
  <Blade ID="02914-0..." MachinePool="A0"/>
  <Blade ID="02916-0..." MachinePool="A0"/>
  <Blade ID="02918-0..." MachinePool="A0"/>
</Cluster>

<Cluster Name="C102">
  <Blade ID="02913-0..." MachinePool="A2"/>
  <Blade ID="02915-0..." MachinePool="A2"/>
  <Blade ID="02917-0..." MachinePool="A2"/>
</Cluster>
What is the configuration?

- Configuration is defined as operations that determine the values of system settings.
What is the configuration?

- Configuration is defined as operations that determine the values of system settings.

```
struct config_int
ConfigureNamesInt[] = {
  {
    "deadlock_timeout",
    ...
    &DeadlockTimeout,...,
  },
  ...
  {
    "max_connections",
    ...
    &MaxConnections,...,
  },
  ...
};
```

- PostgreSQL-9.2.1

- `deadlock_timeout = 10`

- 80 more mappings
What is the configuration?

- Configuration is defined as operations that determine the values of system settings.
- Configuration belongs to system management.
What is the configuration?

- Configuration is defined as operations that determine the values of system settings.
- Configuration belongs to system management.
What is the configuration?

• Configuration is defined as operations that determine the values of system settings.
• Configuration belongs to system management.
• Most configurations can be parameterized, and configuration can be looked as the process of setting configuration parameters.
Configuring systems is hard!
Configuration Errors

- Configuration errors are errors in the system’s configuration settings.
Example 1

Example from NetApp storage system

InitiatorName: iqn:DEV_DOMAIN
Example 1

Example from NetApp storage system

InitiatorName: iqn:DEV_DOMAIN

**Problem Type:** Format error

**Description:** In NetApp commercial storage system, the iscsi device's initiator name (i.e., InitiatorName) only allows lowercase letters, while the user sets the name with some capital letters “DEV_DOMAIN”.
Example 1

Example from NetApp storage system

InitiatorName: iqn:DEV_DOMAIN

Problem Type: Format error

Description: In NetApp commercial storage system, the iscsi device's initiator name (i.e., InitiatorName) only allows lowercase letters, while the user sets the name with some capital letters “DEV_DOMAIN”.

Impact: A storage share cannot be recognized
Example 2

Example from OpenLDAP

```
include schema/ppolicy.schema
overlay ppolicy
```

This entry is missing from the configuration file

**Problem Type:** Entry missing error

**Description:** In order to use the password policy (i.e., ppolicy) overlay, user needs to first include the related schema in the configuration file. But the user did not do this.
Example from OpenLDAP

```
include schema/ppolicy.schema
overlay ppolicy
```

This entry is missing from the configuration file

**Problem Type:** Entry missing error

**Description:** In order to use the password policy (i.e., `ppolicy`) overlay, user needs to first include the related schema in the configuration file. But the user did not do this.

**Impact:** The LDAP server fails to work
Example 3

Example from MySQL

general_log = /var/log/mysql/mysq1.log
Example from MySQL

```
general_log = /var/log/mysql/mysql.log
```

**Problem Type:** Value type error

**Description:** The parameter “general_log” should be an integer, rather than path (string). In MySQL, there is another parameter “general_log_file” used to point the log path.
Example 3

Example from MySQL

general_log = /var/log/mysql/mysql.log

**Problem Type:** Value type error

**Description:** The parameter “general_log” should be an integer, rather than path (string). In MySQL, there is another parameter “general_log_file” used to point the log path.

**Impact:** Log cannot be correctly written.
Example 4

Example from MySQL DB

```bash
mysql's config
max_connections = 300
...
php's config
mysql.max_persistent = 400
```
Example from MySQL DB

```
mysql's config
max_connections = 300
...
php's config
mysql.max_persistent = 400
```

**Problem Type:** Value correlation error

**Description:** When using persistent connections, the `mysql.max_persistent` in PHP should be no larger than the `max_connections` in MySQL. User did not conform to this constraint.

---

The max allowed persistent connections specified in PHP is larger than the max connection specified in MySQL.
Example from MySQL DB

```
mysql's config
max_connections = 300
... ...
php's config
mysql.max_persistent = 400
```

The max allowed persistent connections specified in PHP is larger than the max connection specified in MySQL.

Problem Type: Value correlation error

Description: When using persistent connections, the mysql.max_persistent in PHP should be no larger than the max_connections in MySQL. User did not conform to this constraint.

Impact: "Too many connections" error generated
Configuration File in Reality

### `php.ini` with Apache and MySQL

```ini
; Engine
engine = On ; Activates PHP
short_open_tag = On ; allows to use < to simplify <php
asp_tags = Off ; we do not allow to include Apache ASP (PAS) or IIS
proc
comic
output
cont
after each block.

; Safe Mode
safe_mode = Off ; We do not want the safe way
safe_mode_exec_dir = ; Directory where PHP is executed
safe_mode_include_dir = Directory where PHP will search PHP libraries
safe_mode_allowed_env_vars = PHP ; Only is allowed to the users
; to create system variables that begin with PHP.
safe_mode_protected_env_vars = LD_LIBRARY_PATH ; List of system variables
; that can not be changed by security reasons.
disable_functions = ; Functions that will be disabled for security reasons
allow_url_fopen = Yes ; We allowed that they open to archive from PHP
open_basedir = ;
highlight_string = #D00000
highlight.comment = #FF8800
highlight.style = #0099FF
highlight.regex = #000000
highlight.snip = #0000FF

; Misc
expose_php = On ; It indicates in the message of the Web server if it is installed or no.

; Resource Limits
max_execution_time = 30 ; Maximum time of execution of script.
memory_limit = 100 ; Maximum memory allowed that can consume the script.

; Error handling and logging
error_reporting = E_ALL ; We indicated that shows all the errors and warnings.
display_errors = Off ; Does not print in screen.
display_startup_errors = Off ; That does not show the errors of PHP starting.

; Track errors
error_log = /var/log/php/php4.log ; Log file that will keep the errors
warn_plus_overloading = Off ; We do not warn if operator + is used with strings

; Data Handling

; Variables order is "EGFCS" ; This directive describes the order in which
; will be registered the PHP variables (Being G=GET, F=POST, C=Cookie, 
; E = System, S = Own of PHP, all is indicated like EGFCS)

register_globals = Off ; We do not want that the EGFCS are registered like globals.
register_argc_argv = Off ; We did not declare ARGV and ARGPC for
; the use to square size of sending POST that will
; added fro GFC(GET/POST/Cookie

magic_quotes_runtime = Off ; Quotes added for system generated
; data.
; for example from SQL, exec(), etc.
magic_quotes_sybase = Off ; Use Sybase style added quotes.
(escape ' with ' instead of ")

; PHP default type of file and default codification.
default_mime_type = "text/html"
default_charset = "iso-8859-1"

; Routes and directories;
include_path = 

doc_root = ; Root of the php pages, better is to leave in blank.
user_dir = ; Where php executes scripts,better is to leave in blank.

; Upload files to the server;
file_uploads = On ; Allow upload files to the server.
upload_max_filesize = 2M ; Maximum size of the files are
; going to upload.

; Dynamic Extensions
extension=gd.so ; Graphics
extension=mysql.so ; Mysql
extension=idap.so ; ldap
extension=sha.so ; Mhash
extension=imap.so ; Imap
extension=kads.so ; Kerberos
extension=cups.so ; Cupsys
extension=recode.so ; Recode

; System Log
[Syslog]
define_syslog_variables = Off ; We disabled the definition of syslog variables.

; mail functions
[mail function]
sendmail_path = ; In unix system, where is located sendmail (is
; "variables_order -t -l" by default)

debug [Debugger]

debugee.host = localhost ; Where is the debug
debugee.port = 7869 ; The port it is listening
debugee.enabled = false ; We suppose there is

; SQL Options
[SQL]
sql.safe_mode = Off ; SQL safe mode, we will di

; MySQl Options
[MySQl]
mySQl.allow_persistent = Off ; We will disable
; security reasons.
mySQl.max_persistent = -1 ; Number of persistent
; when is disabled.
mySQl.max_links = -1 ; Maximum number of connect
mySQl.default_port = 3306 ; Default port of mySQl
mySQl.default_socket = ; Socket name that will have
; connections.

; If is void, will be use the default compilation
mySQl.default_host = ; No default host configure
mySQl.default_user = ; No default user configure
mySQl.default_password = ; No default password configure

; session control
[session]
session.save_handler = files ; We saved the ses
session.save_path = /var/lib/php4 ; Directory w
session.files,

session.use_cookies = 1 ; We use cookies in
session.name = PHPSESSID ; Name of the session
; of the cookie.
session.auto_start = 0 ; We did not initiate ses
session.cookie_lifetime = 0 ; Time of life of a
; wait him to close the navigator.

; The path for which the session
; cookie_domain = ; The domain for which ses
; session.serialize_handler = php ; Used manipulat
; session.gc_probability = 1 ; Probability in per
; -colletor activates in each session.

session.gc_maxlifetime = 1440 ; After this time infor
; will be look like garbage for the garbage col
session.use_reffer_check = ; Verifies HTTP Referer
session.cookies_contain_ids = 0 ; Session entropy length
session.entropy_file = ; The file that will gen ses
session.cache_limit = nocache ; Without ses
session.cache_expire = 180 ; Document expiration
session.use_trans_sid = 0 ; To use translate si

```
Where is the error?

ERROR: SEGFAULT
Where is the error?

ERROR: SEGFAULT

php.ini with Apache and MySQL

register_globals = Off; We do not want that the EGPCS are registered like globals.
register_argc_argv = Off; We did not declare ARGV and ARGCC for use to separate size of sending POST that will added fro GFC(GET/POST/Cookie).

magic_quotes_runtime= Off; Quotes added for system generated data.
;for example from SQL, exec(), etc.
magic_quotes_sybase = Off; Use Sybase style added quotes.
;(escape ' with ' instead of \')
; PHP default type of file and default codification.
default_charset = "utf8"
default_extension = "php"
default_charset = "iso-8859-1"
; Routes and directories;
include_path = . ,
doc_root = / Root of the php pages, better is to leave in blank.
userdir = Where php executes scripts,better is to leave in blank.
extension_dir = /var/lib/php4/apache Where the modules are enable.dll = Off; Allow or No the dynamic loading of modules with the dl() function.
; Upload files to the server;
fileuploads = On; Allow upload files to the server.
upload_max_filesize = 2M; Maximum size of the files we are to upload.

; Dynamic Extensions;
extension=gd.so; Graphics
extension=mysq1.so; Mysql
extension=ldap.so; Ldap
extension=ssh2.so; Mhash
extension=imap.so; Imap
extension=snmp.so; Kerberos
extension=cups.so; Cupsys
extension=recode.so; Recode

Systol
define syslog_variables = Off; We disabled the definition of syslog variables.
; mail functions
[mail function]
sendmail_path = In unix system, where is located sendmail (is 'variables_order = "S=POST, U=COOKIE, E=SYSTEM, S = Own of PHP, all is indicated like EGPCS')
Let's look at details

; Dynamic Extensions;
extension = gd.so ; Graphics
extension = mysql.so ; Mysql
extension = ldap.so ; Ldap
extension = mhash.so ; Mhash
extension = imap.sp ; Imap
extension = kadm5.so ; Kerberos
extension = recode.so ; Recode

ERROR:
SEGFAULT
Let's look at details

; Dynamic Extensions;
extension = gd.so ; Graphics
extension = mysql.so ; Mysql
extension = ldap.so ; Ldap
extension = mhash.so ; Mhash
extension = imap.sp ; Imap
extension = kadm5.so ; Kerberos
extension = recode.so ; Recode

ERROR:
SEGFAULT
Example from Apache with PHP

```php
extension = mysql.so
  ...
extension = recode.so
```

“`recode.so`” must be put before “`mysql.so`”

**Problem Type:** Ordering error

**Description:** When using PHP in Apache, the extension “`mysql.so`” depends on “`recode.so`”. Thus, the order between them matters. The user configured the order in a wrong way.

**Impact:** Apache cannot start due to segment fault
Lecture Roadmap

- What is the configuration?
- Handling configuration errors
- Case Study: Spex
- Case Study: ConfigV
Why misconfiguration is hard to handle

- Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
Why misconfiguration is hard to handle

- Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);

  **Example from MySQL**

  ```
  general_log = /var/log/mysql/mysql.log
  ```

  **Problem Type:** Value type error
  **Description:** The parameter “general_log” should be an integer, rather than path (string). In MySQL, there is another parameter “general_log_file” used to point the log path.
  **Impact:** Log cannot be correctly written.
Why misconfiguration is hard to handle

• Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
• Some misconfiguration problems are caused by legal values (46% - 61%);
Why misconfiguration is hard to handle

- Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
- Some misconfiguration problems are caused by legal values (46% - 61%);

| Related Config: | `datadir=/var/lib/mysql`  
|                 | `user = mysql` |
| Root Cause:     | user is not owner of `datadir`, causing permission denial error |

MySQL
Why misconfiguration is hard to handle

• Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
• Some misconfiguration problems are caused by legal values (46% - 61%);
• Configuration files are not programs. They lack grammar, semantics and structures.
Why misconfiguration is hard to handle

- Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
- Some misconfiguration problems are caused by legal values (46% - 61%);
- Configuration files are not programs. They lack grammar, semantics and structures.
- It is hard to reproduce the configuration errors.
Why misconfiguration is hard to handle

- Due to lack of knowledge, it is hard to check illegal values or types (38% - 53%);
- Some misconfiguration problems are caused by legal values (46% - 61%);
- Configuration files are not programs. They lack grammar, semantics and structures.
- It is hard to reproduce the configuration errors.
- Configuration error is not a bug, so developers do not care.
Whose fault?

Who should be responsible for configuration problem? User or developer?
## State-of-the-Arts

### Approaches for system development

<table>
<thead>
<tr>
<th>Stage</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>System design</td>
<td>Making systems easy to configure</td>
</tr>
<tr>
<td>Quality assurance</td>
<td>Hardening systems against configuration errors</td>
</tr>
</tbody>
</table>
# State-of-the-Arts

## Approaches for system development

<table>
<thead>
<tr>
<th>Stage</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>System design</td>
<td>Making systems easy to configure</td>
</tr>
<tr>
<td>Quality assurance</td>
<td>Hardening systems against configuration errors</td>
</tr>
</tbody>
</table>

## Approaches for system management

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting and validation</td>
<td>Checking the correctness of configurations</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Automating monitoring procedures</td>
</tr>
<tr>
<td>Coping with failures</td>
<td>Troubleshooting configuration errors</td>
</tr>
</tbody>
</table>
## State-of-the-Arts

### Approaches for system development

<table>
<thead>
<tr>
<th>Stage</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>System design</td>
<td>Making systems easy to configure</td>
</tr>
<tr>
<td>Quality assurance</td>
<td>Hardening systems against configuration errors</td>
</tr>
</tbody>
</table>

### Approaches for system management

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting and validation</td>
<td>Checking the correctness of configurations</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Automating monitoring procedures</td>
</tr>
<tr>
<td>Coping with failures</td>
<td>Troubleshooting configuration errors</td>
</tr>
</tbody>
</table>
Lecture Roadmap

• What is the configuration?
• Handling configuration errors
• Case Study: Spex
• Case Study: ConfigV
Spex [SOSP’13]

• Automatically infer config. constraints by analyzing source code (for developers);
• Understand error-prone config. design and vulnerabilities;
• Assist developers to make systems more robust against misconfiguration.
Spex [SOSP’13]

• Automatically infer config. constraints by analyzing source code (for developers);
• Understand error-prone config. design and vulnerabilities;
• Assist developers to make systems more robust against misconfiguration.

- Need source code access;
- Limited to basic types and dependencies.
Spex Overview

Source Code -> Mapping -> Inference -> Config Constraints

# Config file
Max_threads=50
...=...
...

# Source code
if (ThreadNum<100) {
  ...
}
Structure-Based Mapping

- PostgreSQL-9.2.1

```c
struct config_int
ConfigureNamesInt[] = {
    {
        "deadlock_timeout",
        ...
       , &DeadlockTimeout, ..., },
    ...
    {
        "max_connections",
        ...
       , &MaxConnections, ..., },
    ...
}

80 more mappings
```

Annotation:

```plaintext
@STRUCT = ConfigureNamesInt
@PAR = [config_int, 1]
@VAR = [config_int, 3]
```
What constraints can be inferred?

1. **Data type**
   - e.g., integer, float, string, boolean

2. **Data range**
   - e.g., [10, 100], {'yes', 'no'}

3. **Control dependency**
   - e.g., \(X\) dominates \(Y\)'s executions

4. **Value relationship**
   - e.g., \(X < Y\)
Data Type Inference

- Methodology
  - Check the variable’s data type and how the variable is used in syscall/libcall

```c
int ft_init_stopwords(...) {
    fd = my_open(ft_stopword_file, ...);
    ...
}
```

```c
File my_open(const char * FileName, ...) {
    ...
    fd = open((char*) FileName, Flags);
}
```

/* MySQL-5.5.29 */

# Config parameter “ft_stopword_file”

A file path
Data Range Inference

Methodology
- If the variable is compared with a constant value, inspect the branch block to decide the range.

```c
static int config_generic (...) {
    ...
    if(c->value_int < 4)
        c->value_int = 4;
    else if(c->value_int > 255)
        c->value_int = 255;
    ...
}

/* OpenLDAP-2.4.33 */
```

# Config parameter “index_intlen”
Data range: [4, 255]
Control Dependency Inference

- Methodology
  - Check if the config variable’s usage is controlled by another config variable

```c
static TransactionId
RecordTransactionCommit() {
  ...
  if(enableFsync &&
    MinimumActiveBackends(CommitSiblings)){
    ...
  }
  /* PostgreSQL-9.2.1 */
/*All commit_siblings’s use sites are inside the func call.

# Config parameter: "fsync"

"commit_siblings" takes effect iff "fsync" is enabled

"commit_siblings"
```


Injecting Errors

ft_stopword_file dir_path

MySQL-5.5.29

Program received signal SIGSEGV, Segmentation fault.
my_mb_cctype_8bit (cs=0x1226760, ctype=0x7fffffffde00, s=0x1ad5000 <Address 0x1ad5000 out of bounds>, e=0x10185a67f <Address 0x10185a67f out of bounds>) at ./strings/cctype-simple.c:1299 1299 *ctype= cs->ctype[*s + 1];
......
Can we help real-world systems?

<table>
<thead>
<tr>
<th>Software</th>
<th>Real-world misconfig.</th>
<th>Bad reactions that can be potentially avoided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage-A</td>
<td>246</td>
<td>68 (27.6%)</td>
</tr>
<tr>
<td>Apache</td>
<td>50</td>
<td>19 (38.0%)</td>
</tr>
<tr>
<td>MySQL</td>
<td>47</td>
<td>14 (29.8%)</td>
</tr>
<tr>
<td>OpenLDAP</td>
<td>49</td>
<td>12 (24.5%)</td>
</tr>
</tbody>
</table>
Lecture Roadmap

- What is the configuration?
- Handling configuration errors
- Case Study: Spex
- Case Study: ConfigV
## State-of-the-Arts

### Approaches for system development

<table>
<thead>
<tr>
<th>Stage</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>System design</td>
<td>Making systems easy to configure</td>
</tr>
<tr>
<td>Quality assurance</td>
<td>Hardening systems against configuration errors</td>
</tr>
</tbody>
</table>

### Approaches for system management

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting and validation</td>
<td>Checking the correctness of configurations</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Automating monitoring procedures</td>
</tr>
<tr>
<td>Coping with failures</td>
<td>Troubleshooting configuration errors</td>
</tr>
</tbody>
</table>
ConfigV [OOPSLA'17]

- The first configuration verification framework;
- A proactive configuration correctness checking approach;
- We found many configuration errors in GitHub.
ConfigV [OOPSLA’17]

- The first configuration verification framework;
- A proactive configuration correctness checking approach;
- We found many configuration errors in GitHub.

What is verification?
The first configuration verification framework;
A proactive configuration correctness checking approach;
We found many configuration errors in GitHub.

Meet each other?

ConfigV [OOPSLA'17]
Why this is challenging?

Hard to express specification!
**Why this is challenging?**

Example from MySQL

```plaintext
general_log = /var/log/mysql/mysql.log
```

**Problem Type:** Value type error  
**Description:** The parameter “general_log” should be an integer, rather than path (string). In MySQL, there is another parameter “general_log_file” used to point the log path.  
**Impact:** Log cannot be correctly written.
Why this is challenging?

Example from MySQL

general_log = /var/log/mysql/mysql.log

Example from Apache with PHP

```
extension = mysql.so
... ...
extension = recode.so
```

“recode.so” must be put before “mysql.so”

**Problem Type:** Ordering error

**Description:** When using PHP in Apache, the extension “mysql.so” depends on “recode.so”. Thus, the order between them matters. The user configured the order in a wrong way.

**Impact:** Apache cannot start due to segment fault
Why this is challenging?

Example from MySQL

```
general_log = /var/log/mysql/mysql.log
```

Problem Type: Variable type
Description: The log path is defined as an integer, rather than a string, which is another parameter in the log path.
Impact: Log cannot start.

Example from Apache with PHP

```
extension = mysql.so
... ...
extension = recode.so
```

“recode.so” must be put before “mysql.so”

Example from OpenLDAP

```
include schema/ppolicy.schema
overlay ppolicy
```

This entry is missing from the configuration file

Problem Type: Entry missing error
Description: In order to use the password policy (i.e., `ppolicy`) overlay, user needs to first include the related schema in the configuration file. But the user did not do this.
Impact: Apache cannot start.

Impact: The LDAP server fails to work.
Why this is challenging?

Example from MySQL

- mysql's config
  - max_connections = 300
- php's config
  - mysql.max_persistent = 400

Example from MySQL DB

The max allowed persistent connections specified in PHP is larger than the max connection specified in MySQL.

Problem Type: Value correlation error

Description: When using persistent connections, the mysql.max_persistent in PHP should be no larger than the max_connections in MySQL. User did not conform to this constraint.

Impact: “Too many connections” error generated

Impact: Apache cannot start

Impact: The LDAP server fails to work
Our Insight

Config ➔ Learn ➔ Spec/Rules
ConfigV Overview

- **Translator**
- **Predicate DB**
- **Configuration files**
  - Sample dataset
- **Intermediate representation**
  - Training set
- **Checker**
- **Rules**
- **Learn rules**
- **Verification report**
[mysqld]
default_storage_engine=InnoDB
max_heap_table_size=8M
Skip-name-resolve

(innodb)
general_log=1
general_log_file=/var/log/mysql/mysql.log

(mysqld_default_storage_engine,InnoDB)
(mysqld_max_heap_table_size,8M)
(mysqld_skip-name-resolve,)
(innodb_general_log,1)
(innodb_general_log_file,/var/log/mysql/mysql.log)
Uniform Representation

class context1
keyword1=value --------> (context1_keyword1,value)
keyword2=value --------> (context1_keyword2,value)

class context2
keyword1=value --------> (context2_keyword1,value)
keyword2=value --------> (context2_keyword2,value)
keyword3=value --------> (context2_keyword3,value)
# Predicate Database

<table>
<thead>
<tr>
<th>Predicate Family</th>
<th>Type</th>
<th>General Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordering</td>
<td>(*, *) -&gt; Bool</td>
<td>X before Y</td>
</tr>
<tr>
<td>Keyword Correlation</td>
<td>(*, *) -&gt; Bool</td>
<td>X in same file as Y</td>
</tr>
<tr>
<td>Type</td>
<td>(*) -&gt; Bool</td>
<td>X has type Integer</td>
</tr>
<tr>
<td>Equality</td>
<td>(a,a) -&gt; Bool</td>
<td>X=Y</td>
</tr>
<tr>
<td>Coarse Grain</td>
<td>(Int, Int) -&gt; Bool, (Size, Size) -&gt; Bool</td>
<td>X=Y, X &gt; Y, X &lt; Y</td>
</tr>
<tr>
<td>Fine Grain</td>
<td>(Int, Int, Int) -&gt; Bool, (Int, Size, Size) -&gt; Bool, (Size, Int, Size) -&gt; Bool</td>
<td>X<em>Y=Z, X</em>Y &gt; Z, X*Y &lt; Z</td>
</tr>
</tbody>
</table>
Association Rule Learning

Learning rule of the form

\[ r = \{S_0, \ldots, S_{|S|}\} \subseteq \text{valid} \]
\[ \Rightarrow \{T_0, \ldots, T_{|T|}\} \subseteq \text{valid} \]

S and T are source and target sets of words.

E.g.

\[ r = \{\text{bread, peanut butter}\} \subseteq \text{shopping list} \]
\[ \Rightarrow \{\text{jelly}\} \subseteq \text{shopping list} \]
\[ \{\text{beer}\} \subseteq \text{s list} \]
\[ \Rightarrow \{\text{diapers}\} \subseteq \text{s list} \]
**Association Rule Learning**

File #1

- [server]
- Foo = ON
- [client]
- Bar = 1

File #2

- [server]
- Foo = ON
- [client]
- Bar = ON

File #3

- [server]
- Foo = OFF
- [client]
- Bar = OFF

\[ p_{int} = \geq 2 \quad p_{bool} = \geq 2 \]

\[
\frac{k_1 :: \tau \quad k_2 :: \tau}{eq(k_1, k_2) :: Rule} \quad EQ
\]

- Bar : [int = 1, bool = 3]
- Bar :: bool

- Foo : [int = 0, bool = 3]
- Foo :: bool

\[ eq \ (Foo, Bar) :: Rule \]
Learning A Rule

Support - “How often does the training set contain the rule’s keyword?”

\[ \text{support}(r) = \frac{|\{C \in \text{Training Set} \mid S_r \cup T_r \subseteq C\}|}{|\text{Training Set}|} \]

Confidence - “How often is the rule true in the training set?”

\[ \text{confidence}(r) = \frac{|\{C \in \text{Training Set} \mid C \vdash p_r (S_r, T_r)\}|}{\text{support}(r) \times |\text{Training Set}|} \]
Learning A Rule

\[
\text{Foo} \in C \\
\Rightarrow C \vdash \\
\text{Eq ([Foo], [Bar])}
\]

<table>
<thead>
<tr>
<th>File #1</th>
<th>File #2</th>
<th>File #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[server]</td>
<td>[server]</td>
<td>[server]</td>
</tr>
<tr>
<td>Foo = ON</td>
<td>Foo = ON</td>
<td>Foo = OFF</td>
</tr>
<tr>
<td>[client]</td>
<td>[client]</td>
<td></td>
</tr>
<tr>
<td>Bar = 1</td>
<td>Bar = OFF</td>
<td></td>
</tr>
</tbody>
</table>

\[
support(r) = \frac{| \{ C \in \text{Training Set} \mid S_r \cup T_r \subseteq C \} |}{| \text{Training Set} |} = \frac{| \{ \text{File #1, File #2} \} |}{3} = \frac{2}{3}
\]

\[
\text{confidence}(r) = \frac{| \{ C \in \text{Training Set} \mid C \vdash p_r (S_r, T_r) \} |}{support(r) \ast | \text{Training Set} |} = \frac{| \{ \text{File #1} \} |}{2/3 \ast 3} = \frac{1}{2}
\]
ConfigV Overview

Predicate DB

Need to formalize/encode...
recode.so BEFORE mysql.so
thread_mem < total_mem
ConfigV Overview

Need to formalize/encode...
recode.so BEFORE mysql.so
thread_mem < total_mem
php.ini with Apache and MySQL

ORDERING ERROR: Expected "extension""recode.so" BEFORE "extension""mysql.so"
Standard MySQL install

general_log=/var/log/mysql/mysql.log

general_log=ON

general_log_file=/var/log/mysql/mysql.log

ConfigV

TYPE ERROR: Expected an Bool for "general_log[mysqld]"
Evaluation

Training set built from 256 industrial configuration files [Xu et al, FSE ‘15]

1000 configuration files scraped from Github (*.mycnf) for test set

<table>
<thead>
<tr>
<th>Class of Error</th>
<th>Rules Learned</th>
<th>Errors Detected</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>13</td>
<td>62</td>
<td>6%</td>
<td>94%</td>
</tr>
<tr>
<td>Missing</td>
<td>53</td>
<td>55</td>
<td>2%</td>
<td>71%</td>
</tr>
<tr>
<td>Type</td>
<td>92</td>
<td>389</td>
<td>12%</td>
<td>70%</td>
</tr>
<tr>
<td>Fine-Grain</td>
<td>213</td>
<td>324</td>
<td>24%</td>
<td>91%</td>
</tr>
<tr>
<td>Coarse-Grain</td>
<td>97</td>
<td>237</td>
<td>10%</td>
<td>96%</td>
</tr>
</tbody>
</table>
No existing effort can solve this problem
- no party is willing to share internal information
- nobody knows what parameters are correlated
Next Lecture

• In the lec-17, I will learn:
  - Correlated failure prevention
  - Independence-as-a-Service (INDaaS)
  - RepAudit