New Applications of Software Synthesis: Verification of Configuration Files and Firewalls Repair

IFIP WG 2.3 MEETING, MOOLOOLABA, AUSTRALIA

17. 07. 2017.

RUZICA PISKAC

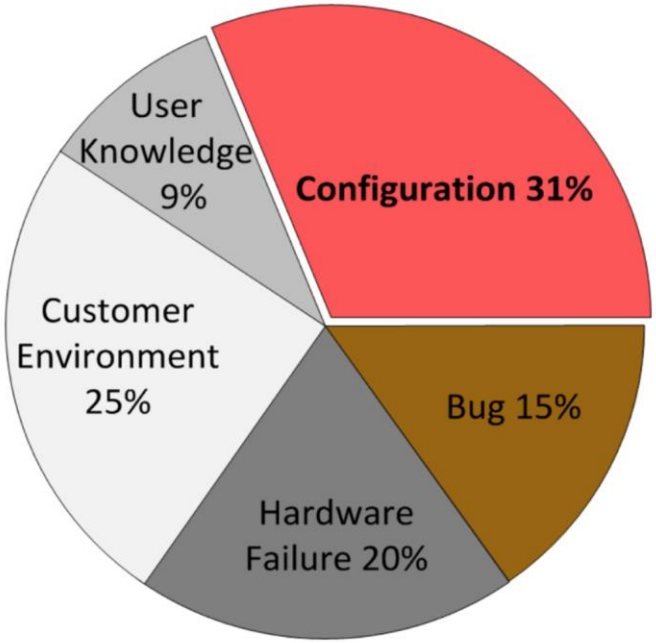
YALE UNIVERSITY

Language Learning for Verification of Configuration Files

JOINT WORK WITH MARK SANTOLUCITO AND ENNAN ZHAI

(AND AARON SHIM AND RAHUL DHODAPKAR)

Configuration errors mean downtime



Root causes of "high-severity" issues in major storage companies [Yin et al., SOSP'11]

php.ini with Apache and MySQL

```
disable_functions = ; Functions that will be disabled for security reasons
allow_url_fopen = Yes ; We allowed that they open to archives from PHP
open basedir = ;
; Colors for the way of colored syntax. highlight.string = #DD0000
highlight.comment = #FF8000
highlight.keyword = #007700
highlight.bg = #FFFFFF
highlight.default = #0000BB
highlight.html = #000000
; 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 = 16M ; Maximun 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.
log_errors = On ; That sends the errors to a file.
track_errors = On ; That $php_errormsg keeps the last Error / Warning
(boolean)
error log = /var/log/php/php4.log ; File that will keep the errors
warn_plus_overloading = Off ; We did not warn if operator + is used with
strings
; Data Handling ;
variables_order = "EGPCS" ; This directive describes the order in which
; will be registered the PHP variables (Being G=GET, P=POST, C=Cookie,
; E = System, S = Own of PHP, all is indicated like EGPCS)
```

```
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.
;extension_dir = /usr/lib/php4/apache ; Where the modules are?
enable dl = Off ; Allow or No the dynamic load of modules with
the dl() function.
; Upload files to the server;
file_uploads = On ; Allow upload files to the server.
upload max filesize = 2M ; Maximum size of the files we are
going to upload.
; Dynamic Extensions ;
extension=gd.so ; Graphics
extension=mysql.so ; Mysql
extension=ldap.so ; Ldap
extension=mhash.so ; Mhash
extension=imap.so ; Imap
extension=kadm5.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 'sendmail -t -i' by default)
; debug
[Debugger]
```

```
debugger.host = localhost ; Where is the debugge
debugger.port = 7869 ; The port it is listenin
debugger.enabled = False ; We suppose there is
; SQL Options
[SQL]
sql.safe_mode = Off ; SQL safe mode, we will d
; Mysql Options
[MySQL]
mysql.allow_persistent = Off ; We will disable
security reasons.
mysql.max persistent = -1 ; Number of persiste
when is disabled.
mysql.max_links = -1 ; Maximum number of conne
limits.
mysql.default port = 3306; Default port of mys
mysql.default_socket = ; Socket name that will
connections.
;If is void, will be use the default compilati
mysql.default host = ; No default host configu
mysql.default_user = ; No default user configu
mysql.default_password = ; No default password
; session control
[Session]
session.save_handler = files ; We saved the se
session.save path = /var/lib/php4 ; Directory
the session files.
session.use_cookies = 1 ; We will use cookies
session.name = PHPSESSID ; Name of the session
name of the cookie.
session.auto_start = 0 ; We did not initiate s
session.cookie_lifetime = 0 ; Time of life of
wait him to closes the navigator.
session.cookie path = / ; The path for which t
session.cookie_domain = ; The domain for which
session.serialize_handler = php ; Used manipul
session.gc probability = 1 ; Probability in pe
collector activates in each session.
session.gc_maxlifetime = 1440 ; After this tim
information
; will be look like garbage for the garbage co
session.referer_check = ; Verifies HTTP Refere
URLs containing ids
session.entropy_length = 0 ; Number of bytes t
file.
session.entropy_file = ; The file that will ge
session.cache limiter = nocache ; Without sess
session.cache expire = 180 ; document expirati
session.use_trans_sid = 0 ; To use translate s
compilation time.
```

php.ini with Apache and MySQL

```
disable_functions = ; Functions that will be disabled for security reasons
allow_url_fopen = Yes ; We allowed that they open to archives from PHP
open basedir = ;
; Colors for the way of colored syntax. highlight.string = #DD0000
highlight.comment = #FF8000
highlight.keyword = #007700
highlight.bg = #FFFFFF
hig
hig
; M
exp                                          e Web server if it is
ins
; R
max                                            of script.
mem                                          n consume the script.
; E
err                                          l the errors and
war
dis
display_startup_errors = Off ; That does not show the errors of PHP
starting.
log_errors = On ; That sends the errors to a file.
track_errors = On ; That $php_errormsg keeps the last Error / Warning
(boolean)
error_log = /var/log/php/php4.log ; File that will keep the errors
warn_plus_overloading = Off ; We did not warn if operator + is used with
strings
; Data Handling ;
variables_order = "EGPCS" ; This directive describes the order in which
; will be registered the PHP variables (Being G=GET, P=POST, C=Cookie,
; E = System, S = Own of PHP, all is indicated like EGPCS)
```

ERROR:
SEGFAULT

```
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.
;extension_dir = /usr/lib/php4/apache ; Where the modules are?
enable_dl = Off ; Allow or No the dynamic load of modules with
the dl() function.
; Upload files to the server;
file_uploads = On ; Allow upload files to the server.
upload_max_filesize = 2M ; Maximum size of the files we are
going to upload.
; Dynamic Extensions ;
extension=gd.so ; Graphics
extension=mysql.so ; Mysql
extension=ldap.so ; Ldap
extension=mhash.so ; Mhash
extension=imap.so ; Imap
extension=kadm5.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 'sendmail -t -i' by default)
; debug
[Debugger]
```

```
debugger.host = localhost ; Where is the debugge
debugger.port = 7869 ; The port it is listenin
debugger.enabled = False ; We suppose there is
; SQL Options
[SQL]
sql.safe_mode = Off ; SQL safe mode, we will d
; Mysql Options
[MySQL]
mysql.allow_persistent = Off ; We will disable
security reasons.
mysql.max_persistent = -1 ; Number of persiste
when is disabled.
mysql.max_links = -1 ; Maximum number of conne
limits.
mysql.default_port = 3306; Default port of mys
mysql.default_socket = ; Socket name that will
connections.
;If is void, will be use the default compilati
mysql.default_host = ; No default host configu
mysql.default_user = ; No default user configu
mysql.default_password = ; No default password
; session control
[Session]
session.save_handler = files ; We saved the se
session.save_path = /var/lib/php4 ; Directory
the session files.
session.use_cookies = 1 ; We will use cookies
session.name = PHPSESSID ; Name of the session
name of the cookie.
session.auto_start = 0 ; We did not initiate s
session.cookie_lifetime = 0 ; Time of life of
wait him to closes the navigator.
session.cookie_path = / ; The path for which t
session.cookie_domain = ; The domain for which
session.serialize_handler = php ; Used manipul
session.gc_probability = 1 ; Probability in pe
collector activates in each session.
session.gc_maxlifetime = 1440 ; After this tim
information
; will be look like garbage for the garbage co
session.referer_check = ; Verifies HTTP Refere
URLs containing ids
session.entropy_length = 0 ; Number of bytes t
file.
session.entropy_file = ; The file that will ge
session.cache_limiter = nocache ; Without sess
session.cache_expire = 180 ; document expirati
session.use_trans_sid = 0 ; To use translate s
compilation time.
```

php.ini with Apache and MySQL

```
disable_functions = ; Functions that will be disabled for security reasons
allow_url_fopen = Yes ; We allowed that they open to archives from PHP
open basedir = ;
; Colors for the way of colored syntax. highlight.string = #DD0000
highlight.comment = #FF8000
highlight.keyword = #007700
highlight.bg = #FFFFFF
hig
hig
; M
exp                                    e Web server if it is
ins
; R
max                                      of script.
mem                                    n consume the script.
; E
err                                    l the errors and
war
dis
display_startup_errors = Off ; That does not show the errors of PHP
starting.
log_errors = On ; That sends the errors to a file.
track_errors = On ; That $php_errormsg keeps the last Error / Warning
(boolean)
error_log = /var/log/php/php4.log ; File that will keep the errors
warn_plus_overloading = Off ; We did not warn if operator + is used with
strings
; Data Handling ;
variables_order = "EGPCS" ; This directive describes the order in which
; will be registered the PHP variables (Being G=GET, P=POST, C=Cookie,
; E = System, S = Own of PHP, all is indicated like EGPCS)
```

ERROR:
SEGFAULT

```
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.
;extension_dir = /usr/lib/php4/apache ; Where the modules are?
enable_dl = Off ; Allow or No the dynamic load of modules with
the dl() function.
; Upload files to the server;
file_uploads = On ; Allow upload files to the server.
                                       ize of the files we are
going to upload.
; Dynamic Extensions ;
extension=gd.so ; Graphics
extension=mysql.so ; Mysql
extension=ldap.so ; Ldap
extension=mhash.so ; Mhash
extension=imap.so ; Imap
extension=kadm5.so ; Kerberos
extension=cups.so ; Cupsys
extension=recode.so ; Recode
; System Log
[Syslog]
                                       isabled the definition of
syslog variables.
; mail functions
[mail function]
;sendmail_path = ;In unix system, where is located sendmail
(is 'sendmail -t -i' by default)
; debug
[Debugger]
```

```
debugger.host = localhost ; Where is the debugge
debugger.port = 7869 ; The port it is listenin
debugger.enabled = False ; We suppose there is
; SQL Options
[SQL]
sql.safe_mode = Off ; SQL safe mode, we will d
; Mysql Options
[MySQL]
mysql.allow_persistent = Off ; We will disable
security reasons.
mysql.max_persistent = -1 ; Number of persiste
when is disabled.
mysql.max_links = -1 ; Maximum number of conne
limits.
mysql.default_port = 3306; Default port of mys
mysql.default_socket = ; Socket name that will
connections.
;If is void, will be use the default compilati
mysql.default_host = ; No default host configu
mysql.default_user = ; No default user configu
mysql.default_password = ; No default password
; session control
[Session]
session.save_handler = files ; We saved the se
session.save_path = /var/lib/php4 ; Directory
the session files.
session.use_cookies = 1 ; We will use cookies
session.name = PHPSESSID ; Name of the session
name of the cookie.
session.auto_start = 0 ; We did not initiate s
session.cookie_lifetime = 0 ; Time of life of
wait him to closes the navigator.
session.cookie_path = / ; The path for which t
session.cookie_domain = ; The domain for which
session.serialize_handler = php ; Used manipul
session.gc_probability = 1 ; Probability in pe
collector activates in each session.
session.gc_maxlifetime = 1440 ; After this tim
information
; will be look like garbage for the garbage co
session.referer_check = ; Verifies HTTP Refere
URLs containing ids
session.entropy_length = 0 ; Number of bytes t
file.
session.entropy_file = ; The file that will ge
session.cache_limiter = nocache ; Without sess
session.cache_expire = 180 ; document expirati
session.use_trans_sid = 0 ; To use translate s
compilation time.
```

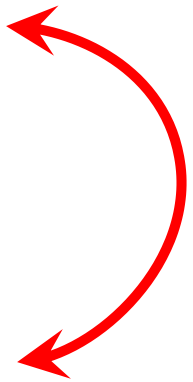
php.ini with Apache and MySQL

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

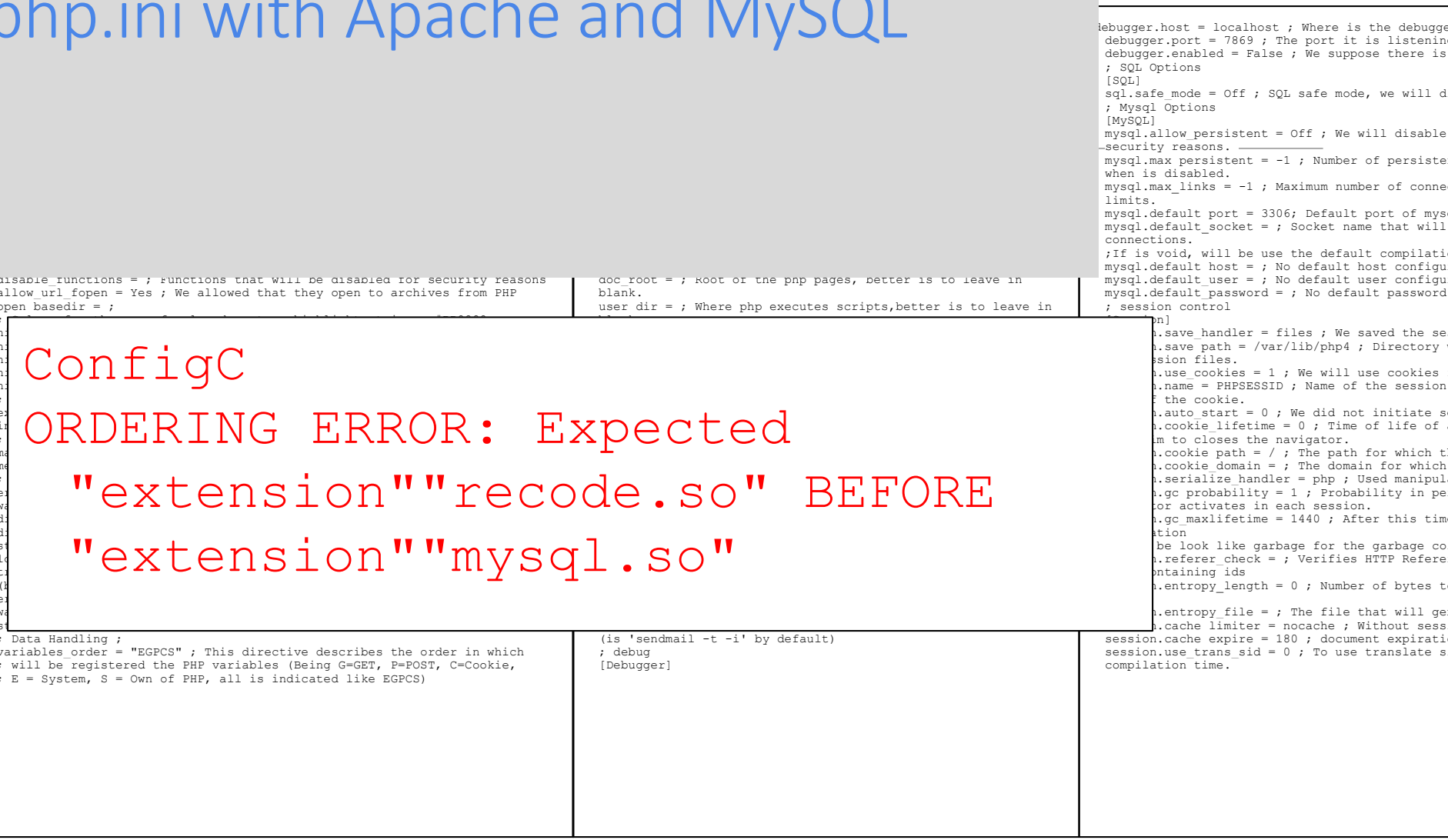
ERROR:
SEGFAULT

php.ini with Apache and MySQL

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



php.ini with Apache and MySQL



Standard MySQL install

```
#
# For explanations see
# http://dev.mysql.com/doc/mysql/en/server-system-variables.html

# This will be passed to all mysql clients
# It has been reported that passwords should be enclosed with ticks/quotes
# escpecially if they contain "#" chars...
# Remember to edit /etc/mysql/debian.cnf when changing the socket location.
[client]
port            = 3306
socket          = /var/run/mysqld/mysqld.sock
# Here is entries for some specific programs
# The following values assume you have at least 32M ram

# This was formally known as [safe_mysqld]. Both versions are currently
parsed.
[mysqld_safe]
socket          = /var/run/mysqld/mysqld.sock
nice            = 0

[mysqld]
#
# * Basic Settings
#
innodb_force_recovery = 4
user            = mysql
pid-file        = /var/run/mysqld/mysqld.pid
socket          = /var/run/mysqld/mysqld.sock
port            = 3306
basedir         = /usr/
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            =  0.0.0.0
#
# * Fine Tuning
```

```
 256M
 16M
 192K
 8
up script and checks MyISAM tables if

 touched
BACKUP
#max_connections        = 100
#table_cache            = 64
#thread_concurrency     = 10
#
# * Query Cache Configuration
#
query_cache_limit       = 16M
query_cache_size        = 48M
#
# * Logging and Replication
#
# Both location gets rotated by the cronjob.
# Be aware that this log type is a performance killer.
# As of 5.1 you can enable the log at runtime!

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

#
# Error logging goes to syslog due to
/etc/mysql/conf.d/mysqld_safe_syslog.cnf.
#
# Here you can see queries with especially long duration
#log_slow_queries       = /var/log/mysql/mysql-slow.log
#long_query_time = 2
#log-queries-not-using-indexes
#
# The following can be used as easy to replay backup logs or
for replication.
# note: if you are setting up a replication slave, see
README.Debian about
#       other settings you may need to change.
#server-id              = 1
#log_bin                        = /var/log/mysql/mysql-bin.log
expire_logs_days        = 10
max_binlog_size         = 100M
#binlog_do_db           = include_database_name
#binlog_ignore_db       = include_database_name
##
```

```
# * InnoDB
#
# InnoDB is enabled by default with a 10MB dat
# Read the manual for more InnoDB related opti
#
# * Security Features
#
# Read the manual, too, if you want chroot!
# chroot = /var/lib/mysql/
#
# For generating SSL certificates I recommend
#
# ssl-ca=/etc/mysql/cacert.pem
# ssl-cert=/etc/mysql/server-cert.pem
# ssl-key=/etc/mysql/server-key.pem


[mysqldump]
quick
quote-names
max_allowed_packet      = 16M

[mysql]
#no-auto-rehash # faster start of mysql but no

[isamchk]
key_buffer              = 16M
#
# * IMPORTANT: Additional settings that can ov
#   The files must end with '.cnf', otherwise
```

Standard MySQL install

ERROR:
Fails to start

```
#
# For explanations see
# http://dev.mysql.com/doc/mysql/en/server-system-variables.html

# This will be passed to all mysql clients
# I
# e
# R
[cl
por
soc
# H
# T

# T
par
[my
soc
nic

[my
#
# * Basic Settings
#
innodb_force_recovery = 4
user            = mysql
pid-file        = /var/run/mysqld/mysqld.pid
socket          = /var/run/mysqld/mysqld.sock
port            = 3306
basedir         = /usr/
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            =  0.0.0.0
#
# * Fine Tuning
```

```
 256M
 16M
 192K
 8
up script and checks MyISAM tables if

 touched
BACKUP
#max_connections        = 100
#table_cache            = 64
#thread_concurrency     = 10
#
# * Query Cache Configuration

limit       = 16M
size        = 48M

and Replication

ion gets rotated by the cronjob.
hat this log type is a performance killer.
you can enable the log at runtime!

            = /var/log/mysql/mysql.log

ing goes to syslog due to
conf.d/mysqld_safe_syslog.cnf.

# Here you can see queries with especially long duration
#log_slow_queries       = /var/log/mysql/mysql-slow.log
#long_query_time = 2
#log-queries-not-using-indexes
#
# The following can be used as easy to replay backup logs or
for replication.
# note: if you are setting up a replication slave, see
README.Debian about
#       other settings you may need to change.
#server-id              = 1
#log_bin                        = /var/log/mysql/mysql-bin.log
expire_logs_days        = 10
max_binlog_size         = 100M
#binlog_do_db           = include_database_name
#binlog_ignore_db       = include_database_name
##
```

```
# * InnoDB
#
# InnoDB is enabled by default with a 10MB dat
# Read the manual for more InnoDB related opti
#
# * Security Features
#
# Read the manual, too, if you want chroot!
# chroot = /var/lib/mysql/
#
# For generating SSL certificates I recommend
#
# ssl-ca=/etc/mysql/cacert.pem
# ssl-cert=/etc/mysql/server-cert.pem
# ssl-key=/etc/mysql/server-key.pem


[mysqldump]
quick
quote-names
max_allowed_packet      = 16M

[mysql]
#no-auto-rehash # faster start of mysql but no

[isamchk]
key_buffer              = 16M
#
# * IMPORTANT: Additional settings that can ov
#   The files must end with '.cnf', otherwise
```

Standard MySQL install

ERROR:
Fails to start

```
#
# For explanations see
# http://dev.mysql.com/doc/mysql/en/server-system-variables.html

# This will be passed to all mysql clients
# I
# e
# R
[cl
por
soc
# H
# T

# T
par
[my
soc
nic

[my
#
# * Basic Settings
#
innodb_force_recovery = 4
user            = mysql
pid-file        = /var/run/mysqld/mysqld.pid
socket          = /var/run/mysqld/mysqld.sock
port            = 3306
basedir         = /usr/
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            =  0.0.0.0
#
# * Fine Tuning
```

```
 256M
 16M
 192K
 8
up script and checks MyISAM tables if

 touched
BACKUP
#max_connections        = 100
#table_cache            = 64
#thread_concurrency     = 10
#
# * Query Cache Configuration

limit       = 16M
size        = 48M

and Replication

ion gets rotated by the cronjob.
hat this log type is a performance killer.
you can enable the log at runtime!

            = /var/log/mysql/mysql.log

ing goes to syslog due to

# Here you can see queries with especially long duration
#log_slow_queries       = /var/log/mysql/mysql-slow.log
#long_query_time = 2
#
# The following can be used as easy to replay backup logs or
for replication.
# note: if you are setting up a replication slave, see
README.Debian about
#       other settings you may need to change.
#server-id              = 1
#log_bin                        = /var/log/mysql/mysql-bin.log
expire_logs_days        = 10
max_binlog_size         = 100M
#binlog_do_db           = include_database_name
#binlog_ignore_db       = include_database_name
##
```

```
# * InnoDB
#
# InnoDB is enabled by default with a 10MB dat
# Read the manual for more InnoDB related opti
#
# * Security Features
#
# Read the manual, too, if you want chroot!
# chroot = /var/lib/mysql/
#
# For generating SSL certificates I recommend
#
# ssl-ca=/etc/mysql/cacert.pem
# ssl-cert=/etc/mysql/server-cert.pem
# ssl-key=/etc/mysql/server-key.pem


[mysqldump]
quick
quote-names
max_allowed_packet      = 16M

[mysql]
#no-auto-rehash # faster start of mysql but no

[isamchk]
key_buffer              = 16M
#
# * IMPORTANT: Additional settings that can ov
#   The files must end with '.cnf', otherwise
```

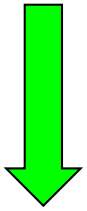
Standard MySQL install

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

ERROR:
Fails to start

Standard MySQL install

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



```
general_log = 1
general_log_file=/var/log/mysql/mysql.log
```

```
ConfigC
TYPE ERROR: Expected a Int with P=1.0 for
 "general_log[mysqld]"
```

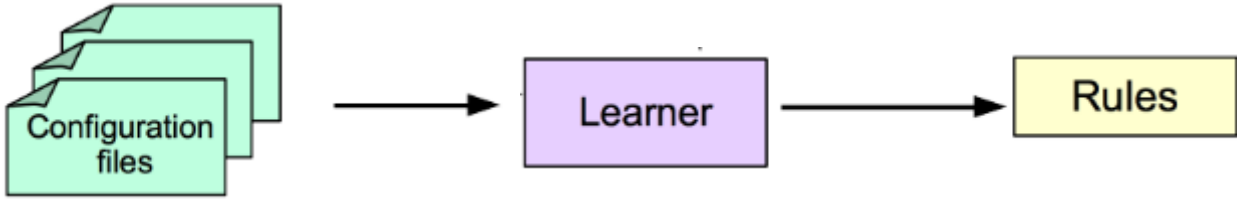
Technical Challenges

- Configuration files are large and hard to debug
- System error messages are cryptic
- Limited specification or documentation
- Keywords do not have types or hard-coded constraints

Goal: **Automatically Generated Rules**

- Automatic specifications
- Automatic verification

Learning configuration file languages from examples



A sample *Training Set* of configuration files

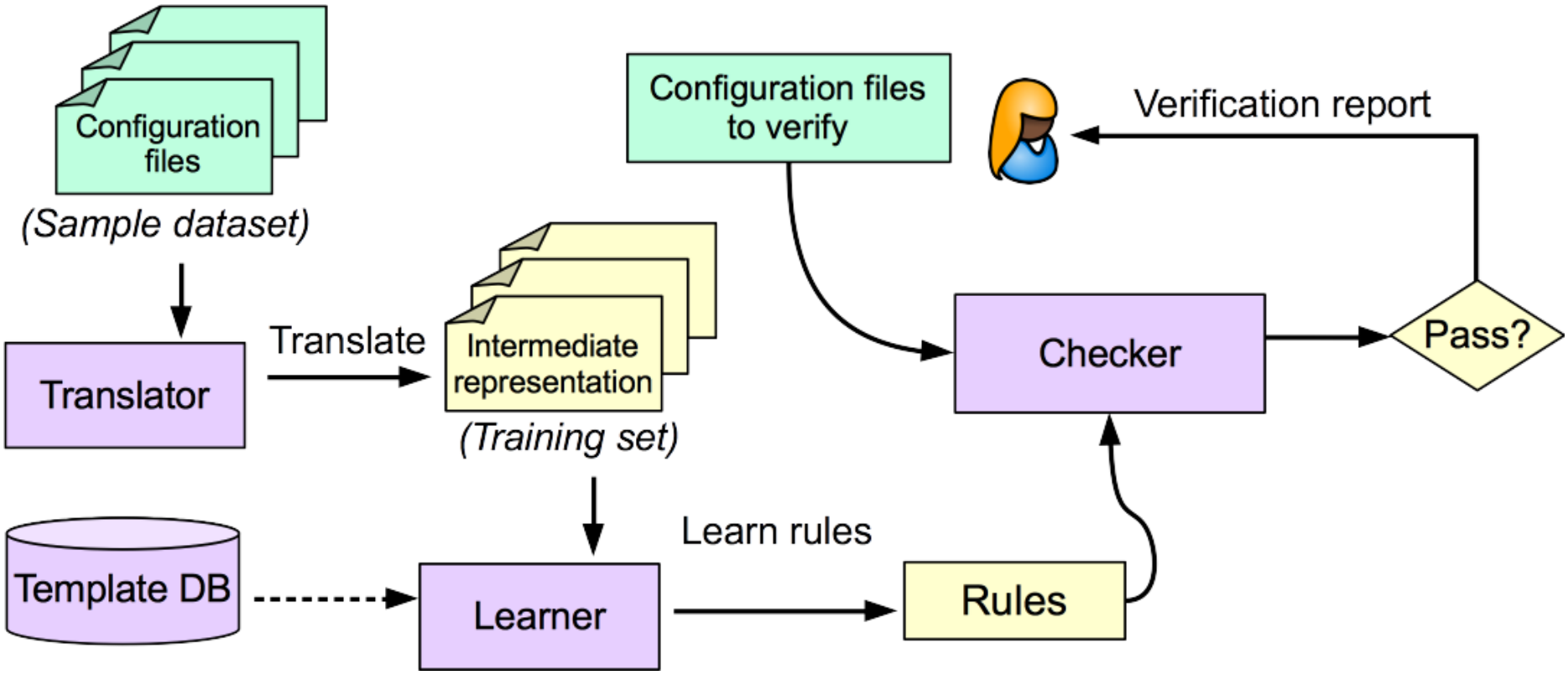
Rules to be used in verification of user's configuration files

Settings for Rule Learning

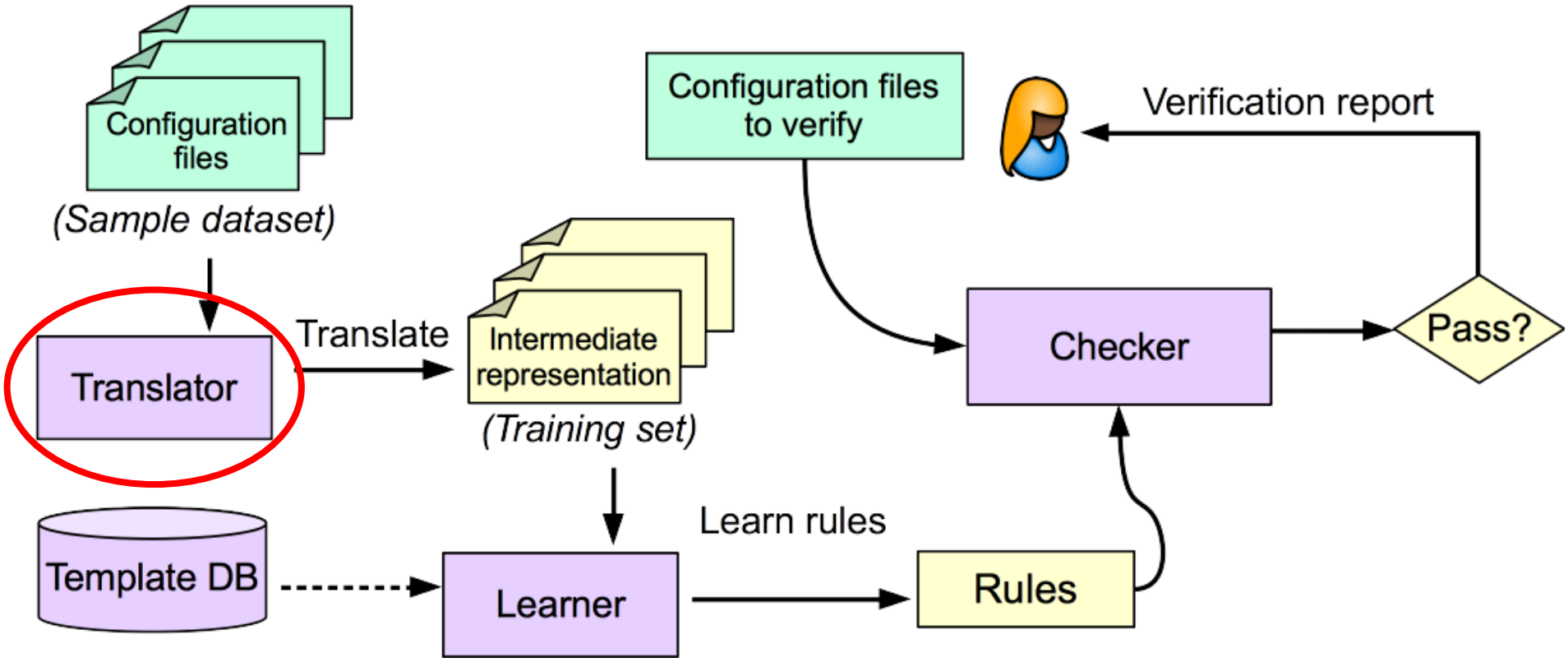
	Training Set contains Errors?	Training Set labeled?*
ConfigC [CAV16]	No	Labeled
ConfigV	Mostly No	Unlabeled

*These are all unsupervised learning settings - labels are on files, not rules

System overview



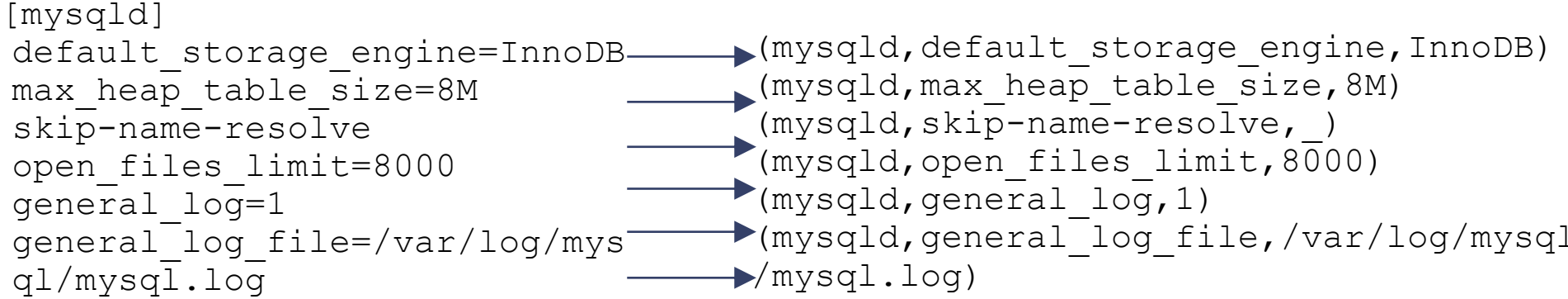
System overview



Translator

```
[mysqld]
default_storage_engine=InnoDB
max_heap_table_size=8M
skip-name-resolve
open_files_limit=8000
general_log=1
general_log_file=/var/log/mysql/
mysql.log
```

Translator

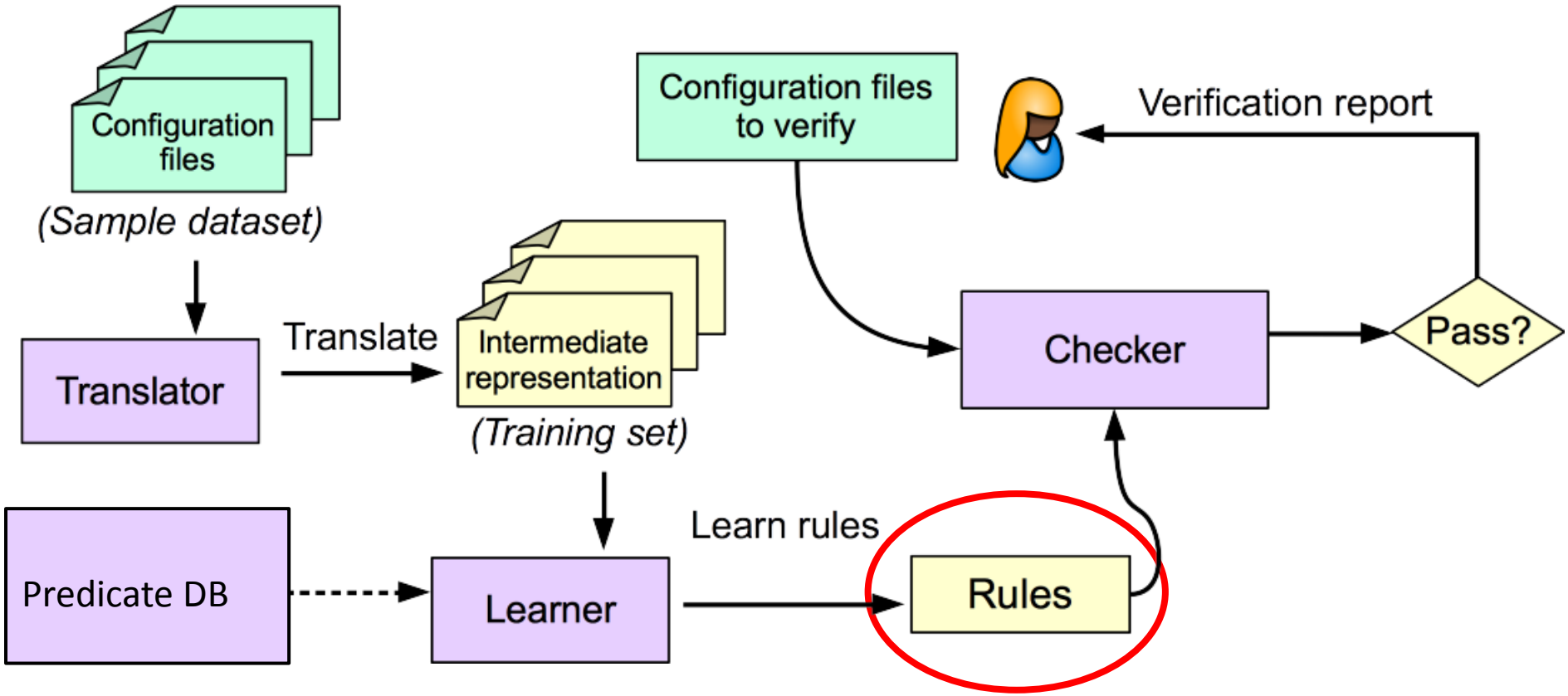

[mysqld]
default_storage_engine=InnoDB
max_heap_table_size=8M
skip-name-resolve
open_files_limit=8000
general_log=1
general_log_file=/var/log/mysql/mysql.log
(mysqld,default_storage_engine,InnoDB)
(mysqld,max_heap_table_size,8M)
(mysqld,skip-name-resolve,_)
(mysqld,open_files_limit,8000)
(mysqld,general_log,1)
(mysqld,general_log_file,/var/log/mysql/mysql.log)

Translator

```
context1
 keyword=value          ⟶  (context1_keyword,value)
 keyword=value          ⟶  (context1_keyword,value)

 context2
 keyword=value          ⟶  (context2_keyword,value)
 keyword=value          ⟶  (context2_keyword,value)
 keyword=value          ⟶  (context2_keyword,value)
```

System overview



Association Rule Learning [Agrawal83]

Learning rule of the form

$r = \{S_0, ..., S_{|S|}\} \in \text{valid}$

$\Rightarrow \{T_0, ..., T_{|T|}\} \in \text{valid}$

S and T are source and target sets of words.

E.g.

r = {bread, peanut butter} ∈ shopping list

⇒ {jelly} ∈ shopping list

Generalizing Association Rule Learning

Learning rules of the form

$r = [S_0, ..., S_{|S|}] \in \text{valid}$

$\Rightarrow \text{valid} \vdash \mathbf{p}\ ([S_0, ..., S_{|S|}], [T_0, ..., T_{|T|}])$

S and T are source and target lists of words.

E.g.

r = [bread, peanut butter] ∈ shopping list

⇒ shopping list ⊢ **purchased together** ([bread, peanut butter] , [jelly])

Association Rule Learning for configuration files

Learning rules of the form

$$r = [S_0, \ldots, S_{|S|}] \in C$$

$$\Rightarrow C \vdash \mathbf{p}\ ([S_0, \ldots, S_{|S|}], [T_0, \ldots, T_{|T|}])$$

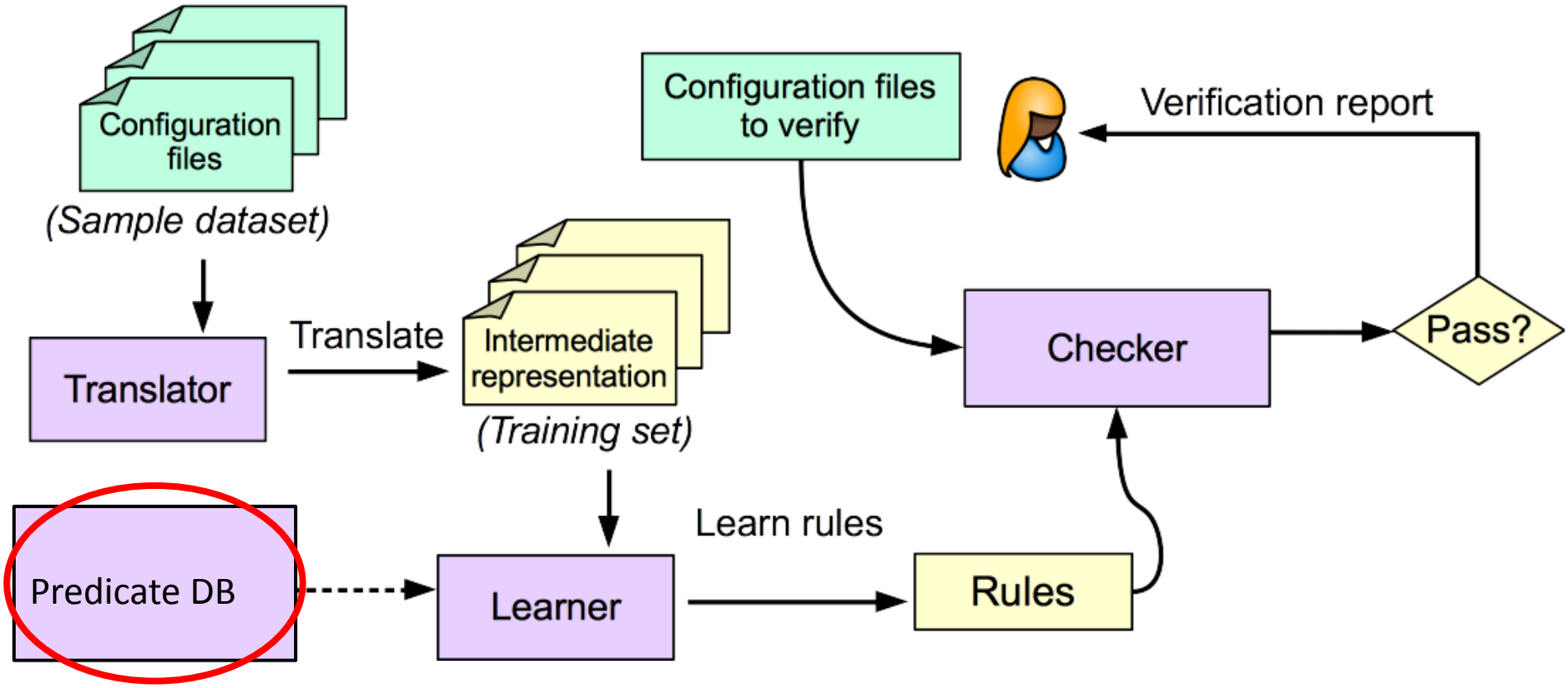
S and T are source and target lists of keywords in the configuration file.

E.g.

$$r = [\text{extension recode.so}] \in C$$

$$\Rightarrow C \vdash \mathbf{order}\ ([\text{extension recode.so}], [\text{extension mysql.so}])$$

System Overview



Predicate Database

Predicate Family	***Type***	***General Forms***
Ordering	(*, *) -> Bool	X before Y
Keyword Correlation	(*, *) -> Bool	X in same file as Y
Type	(*) -> Bool	X has type Integer
Equality	(a,a) -> Bool	X=Y
Coarse Grain	(Int, Int) -> Bool (Size, Size) -> Bool	X=Y, X > Y, X < Y
Fine Grain	(Int, Int, Int) -> Bool (Int, Size, Size) -> Bool (Size, Int, Size) -> Bool	X*Y=Z, X*Y > Z, X*Y < Z

Probabilistic Typing

File #1
[server]
Foo = ON
[client]
Bar = 1

File #2
[server]
Foo = ON
[client]
Bar = ON

File #3
[server]
Foo = OFF
[client]
Bar = OFF

Learning the rule...

$[Foo] \in C$

$\Rightarrow C \vdash$ **equal** ([Foo] , [Bar])

Where

equal :: (a,a) -> Bool

Probabilistic Typing

File #1
[server]
X = 15M
Y = 4096
Z = 4MB

File #2
[server]
X = 5GB
Y = 3MB
Z = 10MB

File #3
[server]
X = 10M
Y = 1M
Z = 3M

Need to *avoid* the rule...

$[X,Y] \in C$

$\Rightarrow C \vdash \_*\_> ([X,Y], [Z])$

Where

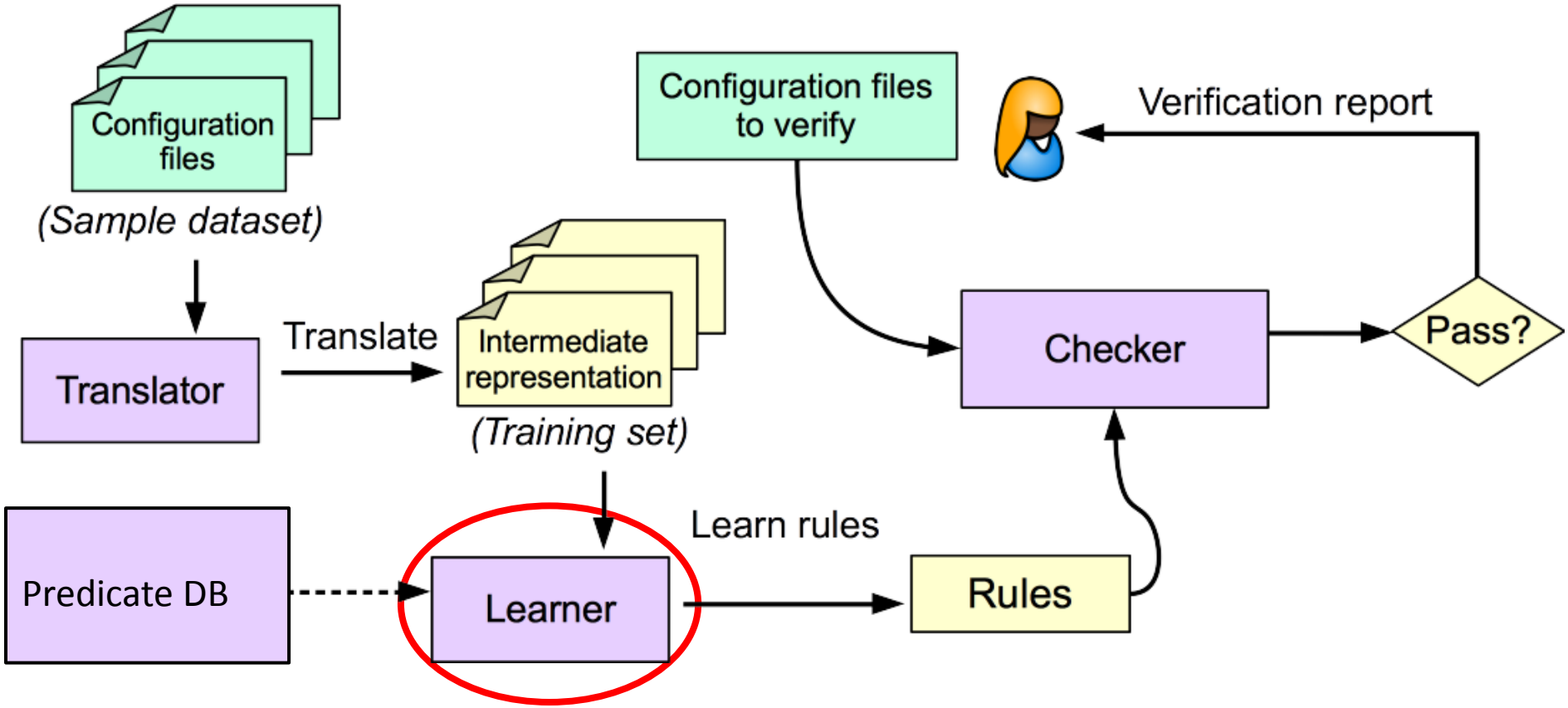
`_*_>` :: (Size,Int,Size) -> Bool

$$\frac{c_{int} = |\{\forall C \in \mathcal{TR}.\ \forall (k,v) \in C.\ v \in \mathbb{Z}\}| \qquad c_{bool} = |\{\forall C \in \mathcal{TR}.\ \forall (k,v) \in C.\ v \in \{0,1,ON,OFF\}\}|}{k : \tilde{\tau}[int = c_{int}, bool = c_{bool}]}\ \text{PTYPE}$$

$$\frac{k : \tilde{\tau} \qquad p_{int}(\tilde{\tau})}{k :: int}\ \text{INT} \qquad \frac{k : \tilde{\tau} \qquad p_{bool}(\tilde{\tau})}{k :: bool}\ \text{BOOL}$$

$$\frac{k_1 :: \tau \qquad k_2 :: \tau}{eq(k_1, k_2) :: Rule}\ \text{EQ} \qquad \frac{k1, k2 \in C \qquad k1 \neq k2}{ord(k1, k2) :: Rule}\ \text{ORDER}$$

System overview



Learning a rule

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

$$support(r) = \frac{\left| \{C \in \textit{Training Set} \mid S_r \cup T_r \subseteq C\} \right|}{\left| \textit{Training Set} \right|}$$

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

$$confidence(r) = \frac{\left| \{C \in \textit{Training Set} \mid C \vdash p_r(S_r, T_r)\} \right|}{support(r) * \left| \textit{Training Set} \right|}$$

Learning a rule : example

Foo ∈ C

⇒ C ⊢ **Eq** ([Foo], [Bar])

File #1
[server]
Foo = ON
[client]
Bar = 1

File #2
[server]
Foo = ON
[client]
Bar = OFF

File #3
[server]
Foo = OFF

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

$$confidence(r) = \frac{\left|\{C \in Training\ Set \mid C \vdash p_r(S_r, T_r)\}\right|}{support(r) * \left|Training\ Set\right|} = \frac{|\{File\ \#1\}|}{2/3 * 3} = \frac{1}{2}$$

ConfigC

$threshold_{support}$ = 0%

$threshold_{confidence}$ = 100%

Only need one example to learn

Every example is correct

ConfigV

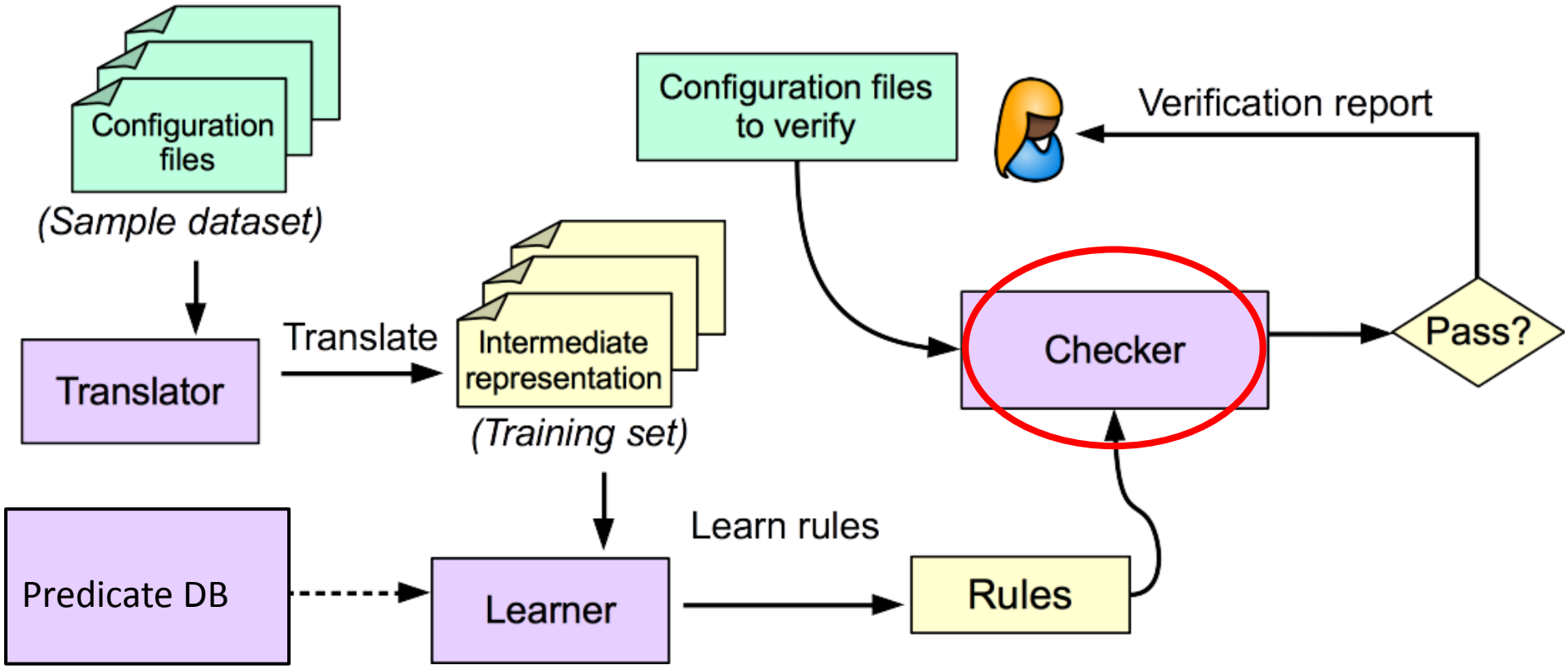
$threshold_{support}$ = ~10%

$threshold_{confidence}$ = ~90%

Need a few examples to learn

Most examples are correct

System overview



ConfigC Evaluation Suite

- Selected files that our tool can handle
  - Relations are in Template DB
  - Configuration files are representative of relations
- Learned from 20 real-world configurations
- A subset of the dataset of industrial configuration file from [Xu et al, FSE ‘15]
  - https://github.com/tianyin/configuration_datasets

ConfigC Evaluation Suite

Error Type	***Passing Tests***	***False Positives***
Missing Entry	5/5	1, 0, 0, 0, 4
Type Error	5/5	0, 0, 0, 0, 0
Keyword Orderings	5/5	0, 2, 1, 0, 6
Value Relations	4/5	0, 0, 0, 1, 0

ConfigC is guaranteed to detect all true positives*, but overestimates.
A false positive is a reported error that is not actually an error.

ConfigV Evaluation Suite

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

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

Class of Error	*Rules Learned*	*Errors Detected*	*Support*	*Confidence*
Order	13	62	6%	94%
Missing	53	55	2%	71%
Type	92	389	12%	70%
Fine-Grain	213	324	24%	91%
Coarse-Grain	97	237	10%	96%

Number of false positives in ConfigV?

	ConfigV MySQL	*Encore MySQL*	*Encore Apache*	*Encore PHP*
False Positive Rate	11%-18%	13%	21%	32%

Select 25 files with known errors and asked industry experts to rate reports

MongoDB expert rated 13/70 errors as definitely false positives.
Microsoft expert rated 8/70 errors as definitely false positives.

False positive rate on par with EnCore [Zhang et al. 2014]

Ranking reported errors

Order of reported errors is also important.
More critical and higher confidence error should go to the top.

Construct a graph with
- keywords as vertices
- rules as edges
- rule confidence as edge weight

The higher the degree of a keyword,
- the more important any rule with that keyword

Reports with extra processing steps

Errors	*None*	*RG*	*PT*	*RG ∧ PT*
Order	12/12 11/11 9/9	3/12 2/11 3/9	5/5 3/3 4/4	3/5 3/3 3/4
Missing	6/10 2/3 2/3	2/10 3/3 3/3	2/4 2/3 2/2	2/4 3/3 3/3
Fine-Grain	30/34 23/25 20/23	18/34 9/25 14/23	6/7 8/9 6/7	3/7 3/9 5/7
Coarse-Grain	29/32 22/23 10/12	29/32 2/23 4/12	11/14 9/10 4/4	4/14 2/10 2/4

X/Y means known true positive in position X among Y errors

Automated Firewall Repair With Example Based Synthesis

JOINT WORK WITH WILLIAM HALLAHAN AND ENNAN ZHAI

Firewall Example

iptables script

```
# All TCP sessions should begin with SYN
iptables -A INPUT -p tcp ! --syn -m state --state NEW -j DROP


# Accept inbound TCP packets
iptables -A INPUT -m state –-state ESTABLISHED,RELATED -j ACCEPT
iptables -A INPUT -p tcp --dport 22 -m state --state NEW -j ACCEPT
iptables -A INPUT -p tcp --dport 80 -m state --state NEW -j ACCEPT
iptables -A INPUT -p tcp --dport 81 -m state --state NEW -j ACCEPT
```

Firewall Example

iptables script

```
# All TCP sessions should begin with SYN
iptables -A INPUT -p tcp ! --syn -m state --state NEW -j DROP

# Accept inbound TCP packets
iptables -A INPUT -m state –-state ESTABLISHED,RELATED -j ACCEPT
iptables -A INPUT -p tcp --dport 22 -m state --state NEW -j ACCEPT
iptables -A INPUT -p tcp --dport 80 -m state --state NEW -j ACCEPT
iptables -A INPUT -p tcp --dport 81 -m state --state NEW -j ACCEPT

# Block IP address
iptables -A INPUT -s 172.168.14.6 -j DROP
```

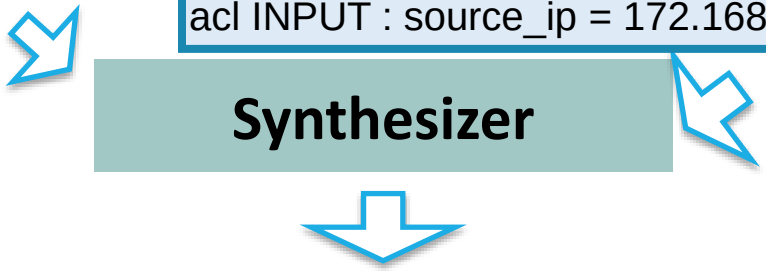
Synthesis Example

iptables script

```
iptables -A INPUT -p tcp ! --syn -m state --state NEW -j DROP
iptables -A INPUT -m state –-state ESTABLISHED,RELATED -j ACCEPT
iptables -A INPUT -p tcp --dport 22 -m state --state NEW -j ACCEPT
...
```

example

acl INPUT : source_ip = 172.168.14.6 => DROP



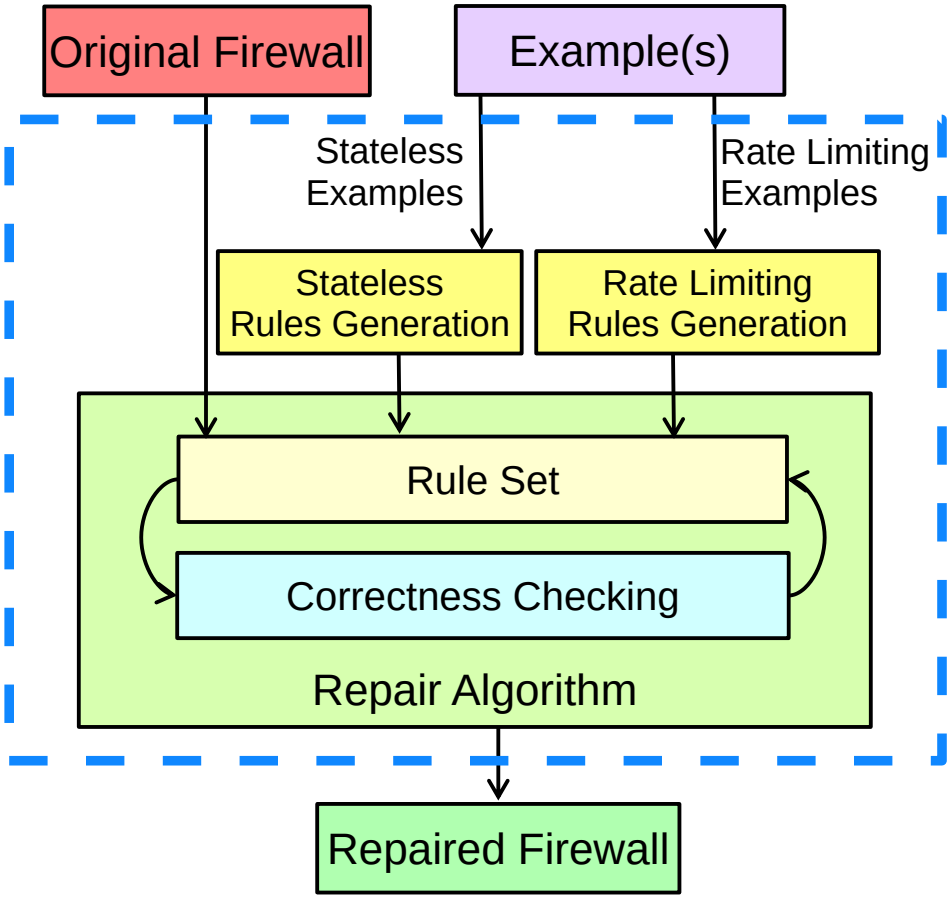
Repaired iptables script

```
iptables -A INPUT -p tcp ! --syn -m state --state NEW -j DROP
iptables -A INPUT -s 172.168.14.6 -j DROP
iptables -A INPUT -m state –-state ESTABLISHED,RELATED -j ACCEPT
iptables -A INPUT -p tcp --dport 22 -m state --state NEW -j ACCEPT
...
```

Challenges

- How do we determine what rules to add (particularly in the rate limiting case)?
- How do we determine where to insert these new rules into the existing firewall?
- How do we verify our changes do not have a larger than intended affect?

System Overview



Stateless Firewall SMT Model

Functions:

- acl_num()
- acl_length(a)
- rule_target(a, r)
- protocol(p)
- source_port(p)
- destination_port(p)
- …

- Predicates:
  - valid_acl(a) ≅ 0 ≤ a < acl_num()
  - valid_rule(a, r) ≅ valid_acl(a) ∧ 0 ≤ r < acl_length(a)
  - matches_criteria(p, a, r)
  - matches_rule(p, a, r)
  - reaches(p, a, r)
  - reaches_end(p, a)

Stateless Firewall SMT Model (Continued)

$$\forall p, a, r.$$
$$\text{reaches}(p, a, r) \wedge \text{matches\_criteria}(p, a, r)$$
$$\leftrightarrow$$
$$\text{matches\_rule}(p, a, r)$$

$$\forall p, a, r.$$
$$\text{valid\_rule}(p, a, r) \wedge \text{valid\_rule}(p, a, r + 1) \wedge$$
$$\text{reaches}(p, a, r) \wedge \neg\text{matches\_rule}(p, a, r)$$
$$\rightarrow$$
$$\text{reaches}(p, a, r + 1)$$

Stateless Firewall SMT Model (Continued)

We want rules that apply to all packet and rules…

…but want to avoid universal quantification

Given an upper limit on the number of packets, possible since we have a finite number of ACLs and rules

Repairing Stateless Firewalls

Rules can be translated directly, or almost directly, from examples:

```
Acl INPUT : source_ip = 199.83.127.227 AND destination_port = 22 => ACCEPT
```

```
iptables –A INPUT -p 6 -s 199.83.127.227 -–dport 22 -j ACCEPT
iptables –A INPUT -p 17 -s 199.83.127.227 -–dport 22 -j ACCEPT
```

Synthesizing Stateless Rules

We want to offer guarantee of correct routing on packets that resemble the example, but not change the behavior of the firewall on other packets

Convert the original and correct firewall to an SMT and call them M and M', resp. Then check:

$$
\begin{aligned}
&\forall p. \\
&\quad (\text{terminates\_with}(M, p) = \text{terminates\_with}(M', p) \\
&\qquad \vee (\text{matches\_example}(p, e) \wedge \text{reaches}(M, p, a, 0)) \\
&\quad) \\
&\wedge \\
&\quad (\text{matches\_example}(p, e) \wedge \text{reaches}(M', p, a, 0) \rightarrow \text{terminates\_with}(M', p) = ak)
\end{aligned}
$$

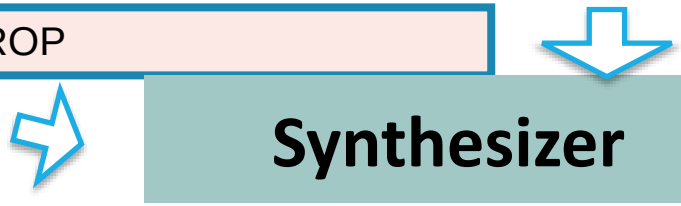
Synthesis Example (Rate Limiting)

example

acl INPUT: protocol = 17 AND time = 10 => ACCEPT,
acl INPUT : protocol = 17 AND time = 15 => DROP,
acl INPUT : protocol = 17 AND time = 20 => ACCEPT

iptables script

iptables -A INPUT -p tcp –-tcp-flags SYN,ACK SYN -j DROP



Repaired iptables script

iptables -A INPUT -m limit --limit 6/minute --limit-burst 1 -p 17 -j ACCEPT
iptables -A INPUT -p 17 -j DROP
iptables-A INPUT -p tcp --tcp-flags SYN,ACK SYN -j DROP

Rate Limiting

```
iptables -A INPUT -m limit --limit 6/minute --limit-burst 3 -j ACCEPT
iptables –A INPUT -j DROP
```

Token bucket Algorithm

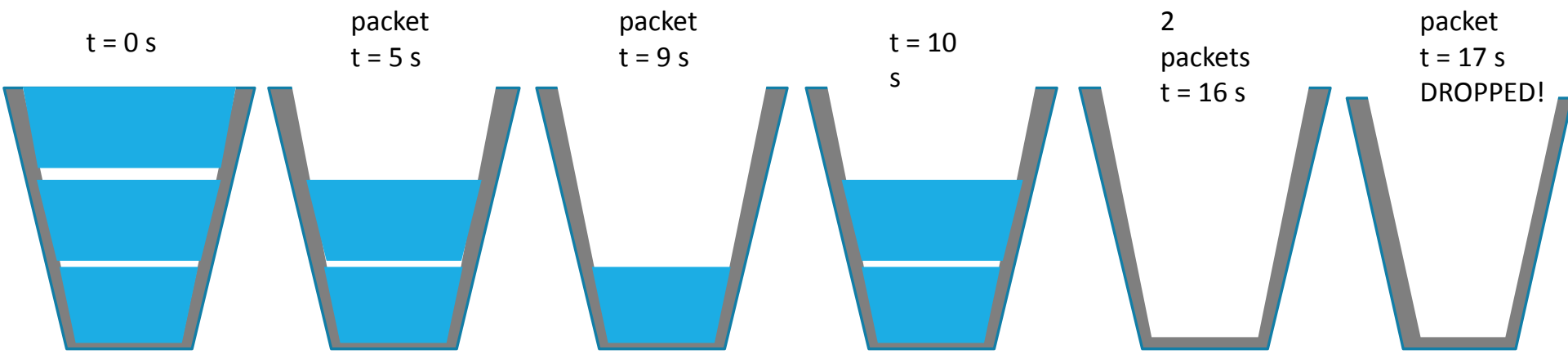
rate = 6 packets / minute

burst = 3 packets

Rate Limiting

rate = 6 packets / minute

burst = 3 packets



Rate Limiting SMT Model

Each packet p has an arrival time, arrival-time(p), in seconds

Want to only use integers

- rate = 7 packets / minute? One packet allowed every 8.571… seconds

Each limit in the model has:

- rate
- burst
- sub (a multiplyer)

Repairing Rate Limiting Firewalls

examples

```
acl INPUT : protocol = 17 AND time = 10 => ACCEPT,
acl INPUT : protocol = 17 AND time = 15 => DROP,
acl INPUT : protocol = 17 AND time = 20 => ACCEPT
```

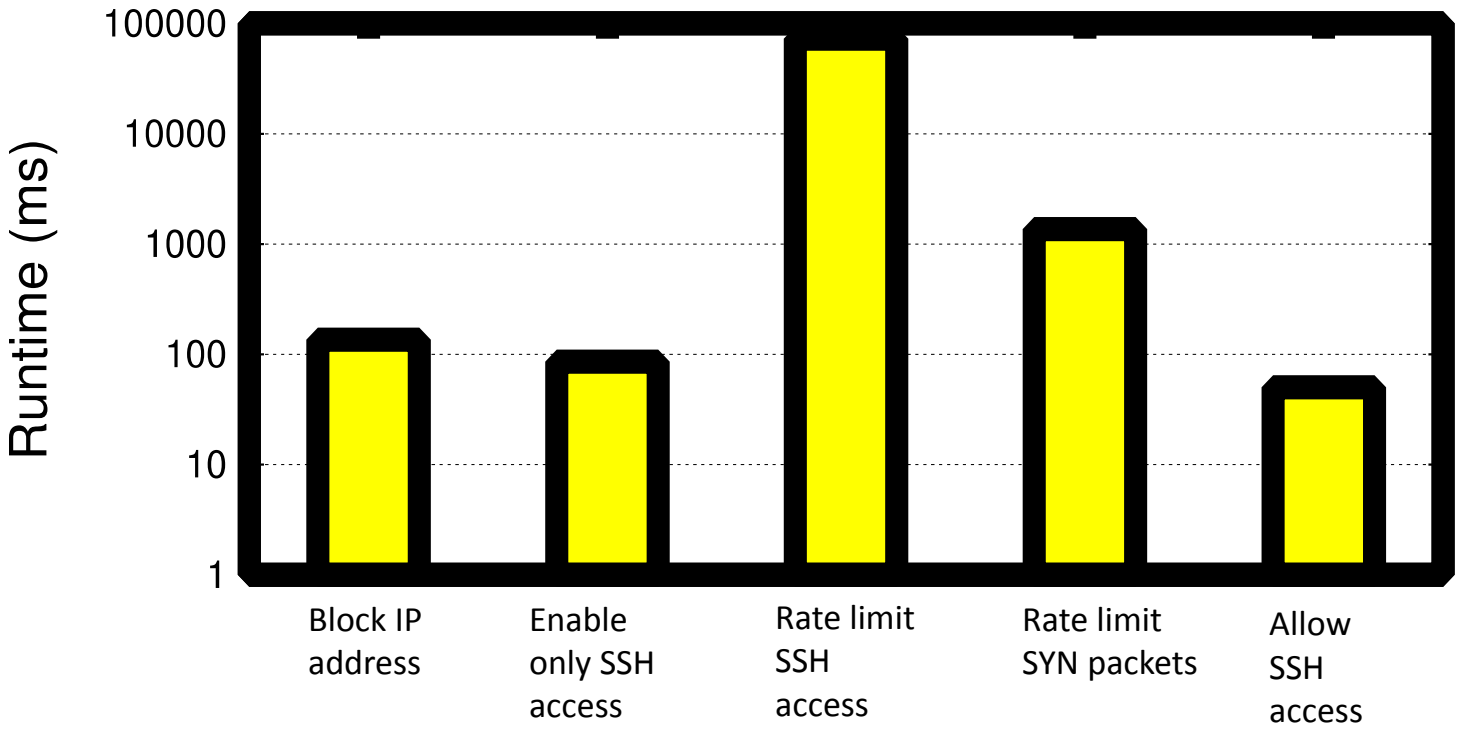
partially constructed rules

```
protocol = 17 AND limit rate1 burst1 sub1 active1 AND use1 → ACCEPT
protocol = 17 AND limit rate2 burst2 sub2 active2 AND use2 → DROP
protocol = 17 AND limit rate3 burst3 sub3 active3 AND use3 → ACCEPT
```

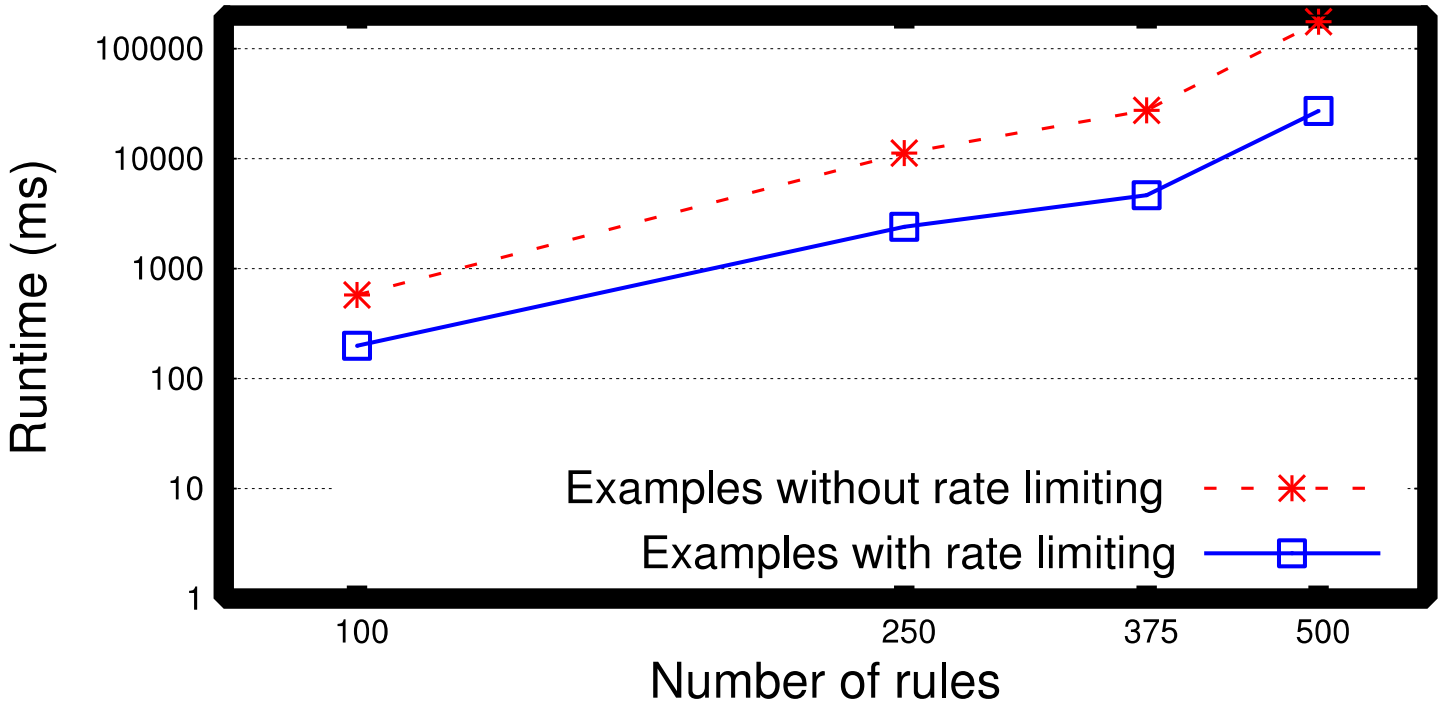
completely constructed rules

```
protocol = 17 AND limit 6 1 60 true AND true → ACCEPT
protocol = 17 AND limit 1 1 1 false AND true → DROP
protocol = 17 AND limit 1 1 1 false AND false → ACCEPT
```

Runtime of FireMason



Scalability of FireMason



Conclusions

- Presented today: a tool for verification of configuration files and a tool for firewall repair
- Software synthesis can be successfully applied to problems usually tackled by a system research community, such as repair of firewalls and verification of configuration files
- Often the specification does not exist and needs to inferred from the examples
- Important is to find a suitable model of the problem
- Both tools, ConfigC and FireMason, were tested on the real world examples