Important Information

Latest Software
We recommend that you install the most recent software release to stay up-to-date with the latest functional improvements, stability fixes, security enhancements and protection against new and evolving attacks.

Latest Documentation
The latest version of this document is at: http://supportcontent.checkpoint.com/documentatin_download?ID=16241

For additional technical information, visit the Check Point Support Center (http://supportcenter.checkpoint.com).

For more about this release, see the R75.40VS home page (http://supportcontent.checkpoint.com/solutions?id=sk76540).

Revision History

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 July 2012</td>
<td>First release of this document</td>
</tr>
</tbody>
</table>

Feedback
Check Point is engaged in a continuous effort to improve its documentation.

Please help us by sending your comments (mailto:cp_techpub_feedback@checkpoint.com?subject=Feedback on CPcode for Check Point DLP R75.40VS Reference Guide).
Contents

Important Information ................................................................. 3
Developing Custom CPcode Functions ........................................... 8
  DLP CPcode Basics .................................................................. 8
  The func Statement .................................................................. 8
  Restriction on Executed CPcode Statements .............................. 9
CPcode Scope ............................................................................. 9
  Local Variables ....................................................................... 10
  Function Calls ....................................................................... 10
  Using Built-in CPcode Functions ............................................. 11
  Fastpaths .............................................................................. 11
Optimizing CPcode ................................................................. 12
  Optimizing Variables ............................................................. 12
  Polymorphic Variables ........................................................... 12
  Network Variables ................................................................ 12
  Deleting Variables .................................................................. 12
  Variables Or Literals ............................................................... 13
Optimizing Arrays .................................................................. 13
  Array Copies ....................................................................... 13
  Index Usage ......................................................................... 13
  Complex Indexes ................................................................... 14
Optimizing Function Calls ...................................................... 14
CPcode Lexicon ........................................................................ 15
  Character Sets ................................................................. 15
  Simple Operators .................................................................. 15
  Compound Operators ........................................................... 16
  Naming Variables .................................................................. 16
  Special Local Variables ......................................................... 17
NULL .................................................................................... 17
Reserved Words ................................................................... 18
Constants .............................................................................. 19
Configuring Constants ......................................................... 20
CPcode Data Types ............................................................. 21
  array ........................................................................... 21
  ethmac .......................................................................... 22
  int ................................................................................ 22
  ipv4host and ipv6host ...................................................... 23
  ipv4net and ipv6net .......................................................... 23
  list ................................................................................ 23
  pattern .......................................................................... 24
  str .............................................................................. 24
  tag ............................................................................. 25
CPcode Operators ............................................................. 26
  Arithmetic Precedence ......................................................... 26
  Arithmetic Operators .......................................................... 27
    + (addition) ..................................................................... 27
    - (subtraction) ............................................................... 27
    * (multiplication) ............................................................ 28
    / (division) .................................................................... 28
    % (modulus) ................................................................... 28
    - (negation) .................................................................... 29
  Relational Operators ............................................................ 29
    == (equal to) .................................................................... 30
Logical Operators .........................................................33
&& (logical AND) .........................................................34
|| (logical OR) ..........................................................34
! (logical NOT) ..........................................................35
Bitwise Operators ........................................................35
& (bitwise AND) ..........................................................35
| (bitwise OR) ..........................................................36
~ (bitwise NOT) ..........................................................36
^ (bitwise XOR) .........................................................37
<< (left shift) ............................................................37
>> (right shift) ..........................................................37
Other Operators .........................................................38
= (assignment) ..........................................................38
inside ........................................................................39
?: (conditional) ..........................................................39
CPcode Statements ......................................................41
= (assignment) ..........................................................41
{} (block) ..................................................................41
break ........................................................................42
foreach ......................................................................42
func ...........................................................................43
if .................................................................................44
requires ......................................................................44
return .........................................................................45
while ...........................................................................45
CPcode Built-in Variables ..............................................46
System Variables ..........................................................46
Feature Variables ..........................................................46
Global Variables ..........................................................47
Example with Global Variables ......................................49
Recommendations with Global Variables .........................50
CPcode Built-in Functions ..............................................51
align() ........................................................................53
arrayindices() .............................................................53
atoi() ..........................................................................54
blobbytes() .................................................................54
byte() ..........................................................................55
caseprefix() ...............................................................55
casesuffix() ...............................................................56
cat() ...........................................................................56
csv_close() .................................................................57
csv_getcols() ............................................................57
csv_getrows() ...........................................................58
csv_getval() ..............................................................58
csv_open() .................................................................59
debase64() .................................................................59
dehex() .......................................................................60
deft8() .......................................................................60
dns_expand() .............................................................60
dns_next() .................................................................60
elem() .........................................................................61
errno() ........................................................................61
ethtmac() .................................................................62
filename() .................................................................63
<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>formattedtime()</td>
<td>.63</td>
</tr>
<tr>
<td>hatoi()</td>
<td>.64</td>
</tr>
<tr>
<td>host()</td>
<td>.65</td>
</tr>
<tr>
<td>host6()</td>
<td>.65</td>
</tr>
<tr>
<td>incksum()</td>
<td>.65</td>
</tr>
<tr>
<td>index()</td>
<td>.66</td>
</tr>
<tr>
<td>io_close()</td>
<td>.66</td>
</tr>
<tr>
<td>io_delete()</td>
<td>.66</td>
</tr>
<tr>
<td>io_open()</td>
<td>.67</td>
</tr>
<tr>
<td>io_read()</td>
<td>.67</td>
</tr>
<tr>
<td>io_readlines()</td>
<td>.67</td>
</tr>
<tr>
<td>io_write()</td>
<td>.68</td>
</tr>
<tr>
<td>kiss_compile()</td>
<td>.68</td>
</tr>
<tr>
<td>kiss_pattern()</td>
<td>.68</td>
</tr>
<tr>
<td>kiss_search()</td>
<td>.68</td>
</tr>
<tr>
<td>le_long()</td>
<td>.69</td>
</tr>
<tr>
<td>le_short()</td>
<td>.69</td>
</tr>
<tr>
<td>le_ulong()</td>
<td>.70</td>
</tr>
<tr>
<td>le_ushort()</td>
<td>.70</td>
</tr>
<tr>
<td>listadd()</td>
<td>.70</td>
</tr>
<tr>
<td>listcombine()</td>
<td>.71</td>
</tr>
<tr>
<td>listlom()</td>
<td>.71</td>
</tr>
<tr>
<td>listlen()</td>
<td>.71</td>
</tr>
<tr>
<td>long()</td>
<td>.72</td>
</tr>
<tr>
<td>lsort()</td>
<td>.72</td>
</tr>
<tr>
<td>luniq()</td>
<td>.73</td>
</tr>
<tr>
<td>match()</td>
<td>.73</td>
</tr>
<tr>
<td>multi_compile()</td>
<td>.73</td>
</tr>
<tr>
<td>multi_escape()</td>
<td>.73</td>
</tr>
<tr>
<td>multi_pattern()</td>
<td>.74</td>
</tr>
<tr>
<td>multi_search()</td>
<td>.75</td>
</tr>
<tr>
<td>net()</td>
<td>.76</td>
</tr>
<tr>
<td>net6()</td>
<td>.77</td>
</tr>
<tr>
<td>pattern()</td>
<td>.77</td>
</tr>
<tr>
<td>pred()</td>
<td>.77</td>
</tr>
<tr>
<td>prefix()</td>
<td>.78</td>
</tr>
<tr>
<td>prunearray()</td>
<td>.78</td>
</tr>
<tr>
<td>regcomp()</td>
<td>.78</td>
</tr>
<tr>
<td>regexec()</td>
<td>.79</td>
</tr>
<tr>
<td>scope()</td>
<td>.80</td>
</tr>
<tr>
<td>sed()</td>
<td>.80</td>
</tr>
<tr>
<td>short()</td>
<td>.81</td>
</tr>
<tr>
<td>split()</td>
<td>.81</td>
</tr>
<tr>
<td>str()</td>
<td>.82</td>
</tr>
<tr>
<td>strcasecmp()</td>
<td>.82</td>
</tr>
<tr>
<td>strerror()</td>
<td>.82</td>
</tr>
<tr>
<td>strlen()</td>
<td>.83</td>
</tr>
<tr>
<td>strstr()</td>
<td>.83</td>
</tr>
<tr>
<td>strstr()</td>
<td>.83</td>
</tr>
<tr>
<td>sublist()</td>
<td>.84</td>
</tr>
<tr>
<td>substr()</td>
<td>.85</td>
</tr>
<tr>
<td>succ()</td>
<td>.86</td>
</tr>
<tr>
<td>suffix()</td>
<td>.87</td>
</tr>
<tr>
<td>tolower()</td>
<td>.87</td>
</tr>
<tr>
<td>toupper()</td>
<td>.88</td>
</tr>
<tr>
<td>typeof()</td>
<td>.88</td>
</tr>
<tr>
<td>typeof()</td>
<td>.88</td>
</tr>
<tr>
<td>ubyte()</td>
<td>.89</td>
</tr>
<tr>
<td>ulong()</td>
<td>.89</td>
</tr>
<tr>
<td>unpack()</td>
<td>.90</td>
</tr>
</tbody>
</table>
Chapter 1

Developing Custom CPcode Functions

In This Chapter

- DLP CPcode Basics
- The func Statement
- Restriction on Executed CPcode Statements
- CPcode Scope
- Function Calls
- Using Built-in CPcode Functions
- Fastpaths

In the DLP environment, CPcode can be used for creating new custom DLP Data Types.

Note - Throughout this guide, "data types" has two meanings: CPcode data types are the supported data objects that can be used in scripts. For example: a CPcode data type is "integer".

DLP data types are the representation of data sent by users, to identify messages that should be matched to the DLP policy. For example: a DLP Data Type is "Social Security Numbers".

DLP CPcode Basics

Each file should contain one filter function: `filter run dlp_dt` which does not receive any variables. This is the main function that will be called each time the Data Type is checked. This function should return 1 if the Data Type is matched, and 0 if not.

The func Statement

The `func` statement declares the name and statements that form the body of a function.

In most cases, the function declaration contains multiple statements. Curly braces ({} ) are used to make the function a block statement. Functions may take an arbitrary number of arguments (accessed using the $1, $2, $3, ... notation). Each argument may be any data type.

<table>
<thead>
<tr>
<th>func</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td><code>func name {[statement]}</code></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>name of the function</td>
</tr>
<tr>
<td>statement</td>
<td>one or more statements to be executed within the function</td>
</tr>
</tbody>
</table>
Developing Custom CPcode Functions

<table>
<thead>
<tr>
<th>func</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Return Value</strong></td>
<td>Use the return statement to return values from a function. Return values can be any type except function pointers.</td>
</tr>
</tbody>
</table>

**Example**

```cpp
func counter
{
    if ($count)
        $count = $count +1;
    else
        $count = 1;
}
```

CPcode functions are implicitly declared upon usage and may be called in a CPcode file before they are defined.

For example:

```cpp
initialize();
func initialize
{
    ...
}
```

Restriction on Executed CPcode Statements

The CPcode execution engine keeps a heuristic count of approximate CPcode statements executed. At the beginning of every function call, during the prolog and at the head of all loops, the count of the number of CPcode statements executed is increased and compared against the limit.

Functions that exceed the statement limit are interrupted. The system simulates an immediate return at the point of interruption. Interrupted functions are eligible to run again when their conditions are matched.

The current limits are approximately 5,000 "ticks" from the point of execution (the run() function).

CPcode Scope

Scope refers to the availability of symbols and values within an execution context. Some languages (such as C) support three scopes: function local, file and global scope. CPcode supports multiple scopes. The variables and symbols from one scope are available to another scope by referencing the variable's containing scope. This is similar to providing a fully qualified path name in a file system, and in fact, directly applies to a hierarchical CPcode type/subtype. The scope name is determined when the sensor execution engine loads the CPcode. This name is derived from the CPcode file names.

For example, assume that:
- The protocol configuration file is dlprules.cfg
- The dlprules CPcode contains a CPcode file named commands
- The CPcode for this Protection Group is commands.cpc

The directory structure would look something like this:

```
Root dir
   +---- dlprules
       +---- commands.cpc
       +---- othercommands.cpc
```

Therefore, the scope name for variables and functions in this CPcode is dlprules.commands (which is the file name of the master CPcode, followed by an underscore, and the file name of the subfunction CPcode). The scope name for protocol-wide is simply the protocol file name. For example, the scope for the protocol-wide $WORKSPACE/protocols/dlprules.cpc file is dlprules. Multiple Protection Groups can have global
variables and functions with the same name. These functions and variables will not interfere with each other because they are in different scopes. For example, both the commands and othercommands files contain a function named process_request. However, these functions exist within different scopes, dlprules_commands and dlprules_othercommands, respectively. Their fully-scoped names are dlprules_commands:process_request and dlprules_othercommands:process_request.

When you refer to global variables or functions by name, you are implicitly using variables and functions within the same scope as the reference. You can call functions and access variables in different scopes by using the fully qualified function or variable name.

You can declare values and assign them only to variables and functions in the same scope. By default, all CPcode functions and variables you define are global and exist within a specific scope. If you need local variables, you must explicitly create them.

You can use the scope() built-in function (see "scope()" on page 80) to obtain scope name.

**Local Variables**

CPcode local variables are defined by how they are named. To create a local variable, prefix the variable name with a dollar sign ($). The initial value of a local variable is NULL.

A local variable has a local scope. It exists only inside the function where it is declared. For example, the variable $count could hold the number of iterations inside a function named read_data. Another function, named get_dataStream, could also use a variable named $count. The $count variables in different functions hold different values.

Local variables are stack-allocated and are not persistent between calls.

For example:

This example function declares the display_write function, which calls the our_display and our_write functions. The values of the parameters $1 and $2 are available as local variables inside the function, but they cannot be modified.

```plaintext
func display_write
{
    if ($1 == 1)
        $a = our_display($2);
    else
        $a = our_write($2);
    return $a * 2 + 1;
}
```

**Function Calls**

To invoke a declared function, create a statement that includes the function name, followed by the parameters enclosed in parentheses.

For example:

```
show_service (80, "http");
```

To call a function and ignore the return value: `func_name();`

Example:

```
# In the following example, the expression statement calls
# the function count_hosts and ignores the returned value
count_hosts();
# The following statement passes the current line to
# the function for further processing.
another_scope:process_line("data\n");
```
Using Built-in CPcode Functions

Built-in CPcode functions use syntax conventions.

<table>
<thead>
<tr>
<th>Built-in CPcode Type</th>
<th>Description</th>
<th>Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Data Types</td>
<td>Functions that can take multiple data types for one parameter.</td>
<td>{array, int, list} where any of the data types in the curly braces is valid</td>
</tr>
<tr>
<td>Optional Parameters</td>
<td>Functions that use optional parameters.</td>
<td>[int a, int b, string str] where the parameters in the brackets are optional</td>
</tr>
<tr>
<td>Ellipses</td>
<td>Functions that can take an indeterminate number of parameters.</td>
<td>listadd(list li, any item, [...]) where the ellipsis (...) can arbitrarily match many arguments</td>
</tr>
<tr>
<td>Default Parameter Values</td>
<td>Functions that use default parameter values. If a specific argument is not passed to one of these functions, the default value is assumed.</td>
<td>align(int num, [int hash :=4]); where the default value appears after the := symbol (for example, 4)</td>
</tr>
</tbody>
</table>

Fastpaths

Several built-in CPcode functions have pre-optimized fastpaths they can follow if their most relevant parameter is the same variable as their return variable.

For example:

```cpp
string = dehex(string);
```

A performance boost is gained by working on the parameter in place without allocating a temporary variable as the return variable. This gain can be significant, especially for the complex data types (such as array, list, and string). Functions that do not return variables of a complex data type would not gain anything by using a fastpath, because their return variables are cheap to allocate. Functions that have pre-optimized fastpaths have that ability noted in their description.
Chapter 2

Optimizing CPcode

In This Chapter

- Optimizing Variables 12
- Optimizing Arrays 13
- Optimizing Function Calls 14

The CPcode Compiler optimizes the performance of the CPcode system. You can increase CPcode performance by taking advantage of certain features of the language and avoiding others. On the other hand, CPcode may have an adverse performance impact if you are not careful.

In general: avoid using too many string operations in one Data Type script, and avoid creating long loops.

Optimizing Variables

The CPcode Compiler is optimized to move as much of the run-time type checking into a compile-time type checker. All CPcode variables are cross-referenced with each statement that might assign its value to the variable in question. After all CPcode packages are compiled into an intermediate form, several passes of type inference are performed to determine the types of all variables concretely.

Polymorphic Variables

If the type of a variable cannot be resolved by the type inference process, that variable is called a polymorph. With polymorphs, the engine is forced to use runtime type checking, which is more expensive. To avoid this, assign only one type of data to a variable.

For example, if you initially assign a string to the variable and later assign an integer to the same variable, the variable will be tagged as polymorphic. If one variable is tagged as polymorphic, variables that depend on it will likely also be tagged polymorphic.

Network Variables

Do not mix IP addresses and networks inside a single variable. This will cause the variable to become a polymorph.

The CPcode Compiler tracks IPv4 addresses, IPv4 networks, IPv6 addresses, and IPv6 networks as separate types. However, it condenses them during code generation, because the execution engine considers them to be a unified type.

Deleting Variables

Each variable is either global or local. You may delete a variable by assigning a NULL to it (such as foo = NULL). This reclaims the memory that the variable occupies. You must nullify global variables when they become irrelevant to reclaim their memory.

You do not need to explicitly nullify local variables, unless you intend to reuse or reassign them. The CPcode Compiler will automatically nullify local variables when they fall out of scope, and will handle their destruction more efficiently than could be done with CPcode statements.
**Variables Or Literals**

In CPcode, there is no performance penalty in using a variable over a constant literal.

In this example, both statements take the same code path, and with the exception of cache and paging, they take the same amount of time:

```cpcode
$foo = 1; $bar = $foo;
$foo = ($bar = 1);
```

Two exceptions to this rule are **constant folding** and **conditionals**.

- **Constant folding** - The CPcode Compiler will attempt to evaluate operations on integer literals at compile time. Therefore, '1 + 1' translates to '2', and in the example:

  ```cpcode
  FOO = 1;
  BAR = 1;
  BLETCH = FOO + BAR;
  ```

- **Conditionals** - The literal will be evaluated during the compilation process, so the condition will be removed and only one code block will be compiled. For example:

  ```cpcode
  func absindex {
    if (feature.offsetindex) {
      return offsetindex($1,$2,$3);
    }
    else {
      $i = index(substr($1,$3),$2);
      if ($i == -1 || $i == NULL) {
        return -1;
      }
      else {
        return $i + $3;
      }
    }
  }
  ```

**Optimizing Arrays**

These guidelines will help you write CPcode scripts that use arrays and are optimized for performance.

**Array Copies**

Avoid writing CPcode that could result in a duplicate array. Copying an array can consume unnecessary CPU time and memory.

To avoid duplicate arrays:

- Do not use arrays or lists as a key to an array.
- Do not use arrays or lists as a value in an array.
- Do not use a polymorphic variable to contain an array.

**Index Usage**

Choose a simple data type to use as an array index. The index governs the speed of array operations.

- Use a simple data type, such as an integer, for best performance.
- Using blobs of the same size gives the worst performance. Blob comparison first checks the lengths of the blobs. If they are of different lengths, the comparison will be faster than if the blobs are of identical length, which means that their buffers must be walked.
**Complex Indexes**

It is common practice to track the number of times a TCP or UDP port-tuple has caused an alert, while sending the alert only once. If IP addresses are not of interest, it is better to combine the two 16-bit ports into a single integer than to use the blobbytes() function to create a blob. For example:

```plaintext
$key = ($src_port << 16) | $dst_port;
```

is better than:

```plaintext
$key = blobbytes($src_port, $dst_port);
```

**Optimizing Function Calls**

If used carefully, CPcode functions will simplify the writing of quality CPcode. If used carelessly, performance will suffer or the CPcode may fail to compile.

To ensure that functions do not adversely affect performance:

- Do not write recursive CPcode functions unless necessary.
- Do not use polymorphic variables in a function call.
Chapter 3

CPcode Lexicon

In This Chapter

<table>
<thead>
<tr>
<th>Character Set</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCII Alphanumeric</td>
<td>CPcode supports all ASCII alphanumeric characters.</td>
</tr>
<tr>
<td>Line Continuation</td>
<td>CPcode is not a line-oriented language. It does not have line continuity characters. Where line breaks occur, they must use UNIX-style newline characters, not the Windows-style carriage return and linefeed characters.</td>
</tr>
<tr>
<td>White Space</td>
<td>You can use spaces and horizontal tabs around elements to increase readability and to follow formatting conventions. Comments are also processed as white space. Note: You must use spaces on either side of the &quot;inside&quot; operator.</td>
</tr>
<tr>
<td>Comment</td>
<td>A hash mark (#) indicates the beginning of a comment. The comment includes everything between the hash mark and the end of the line.</td>
</tr>
</tbody>
</table>

Simple Operators

Simple operators are one character.

<table>
<thead>
<tr>
<th>Character</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>addition</td>
</tr>
<tr>
<td>-</td>
<td>subtraction</td>
</tr>
<tr>
<td>*</td>
<td>multiplication</td>
</tr>
<tr>
<td>Character</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
</tr>
<tr>
<td>/</td>
<td>division</td>
</tr>
<tr>
<td>%</td>
<td>modulus</td>
</tr>
<tr>
<td>-</td>
<td>negation</td>
</tr>
<tr>
<td>=</td>
<td>assignment</td>
</tr>
<tr>
<td>&amp;</td>
<td>bitwise AND</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>~</td>
<td>bitwise NOT</td>
</tr>
<tr>
<td>^</td>
<td>bitwise XOR</td>
</tr>
<tr>
<td>!</td>
<td>logical NOT</td>
</tr>
<tr>
<td>&gt;</td>
<td>greater than</td>
</tr>
<tr>
<td>&lt;</td>
<td>less than</td>
</tr>
</tbody>
</table>

**Compound Operators**

Compound operators are two or more characters.

The "inside" operator searches a list for specified values and returns a Boolean.

<table>
<thead>
<tr>
<th>Characters</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;&lt;</td>
<td>left shift</td>
</tr>
<tr>
<td>&gt;&gt;</td>
<td>right shift</td>
</tr>
<tr>
<td>&amp;&amp;</td>
<td>logical AND</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>==</td>
<td>equal to</td>
</tr>
<tr>
<td>!=</td>
<td>not equal to</td>
</tr>
<tr>
<td>&gt;=</td>
<td>greater than or equal to</td>
</tr>
<tr>
<td>&lt;=</td>
<td>less than or equal to</td>
</tr>
<tr>
<td>?:</td>
<td>conditional</td>
</tr>
</tbody>
</table>

**Naming Variables**

- Variable names can contain only ASCII alphanumeric characters and underscores.
- Variable names must begin with an alphabetic character.
- Variable names are not case sensitive.
Variable names are limited to 128 characters in length.

Local variables must be defined using the dollar sign character ($) as the first character of their name.

You cannot use reserved words as variable names. However, variable names can contain reserved words. For example, if_available is a valid variable name even though if is a reserved word.

### Special Local Variables

- **$#** - the number of parameters passed to a function or filter
- **$n** - nth parameter ($1 is the first parameter; $2 is the second parameter, and so on)

You cannot assign values to these special local variables, and they do not contain the same value across invocations.

**For example:**

```cpcode
# This example makes use of the $# variable
func foo
{
    sum = bar(1, 2, 3);
    # "There were three parameters to bar 1 2 3"
    # sum == 6
    sum = bar(4, 5);
    # "There were two parameters to bar 4 5"
    # sum == 9
    sum = bar();
    # "Incorrect call to bar()"
    # sum == NULL
}

func bar
{
    if ($# == 3) {
        print("There were three parameters to bar", $1, $2, $3);
        return $1 + $2 + $3;
    } else if ($# == 2) {
        print("There were two parameters to bar", $1, $2);
        return $1 + $2;
    } else {
        print("Incorrect call to bar()");
        return NULL;
    }
}
```

### NULL

NULL is a special variable that is a member of every data type.

You can compare expressions against NULL.

**For example:**

```cpcode
# Is $foo an error? (Is it invalid?)
if ($foo == NULL)

# Is $foo not an error? (Is it valid?)
if ($foo != NULL)
```
# Is this division in error
if ( ($foo = $a / $b) == NULL )

NULL is also used to clear variables and free their memory.

**For example:**

```c
func clean_up {
    COUNT = NULL;
}
func count_connections {
    if (COUNT[$1] == NULL) {
        COUNT[$1] = 1;
    } else {
        if (COUNT[$1] > CONNECTION_THRESHOLD) {
            complain($1)
        }
    }
}
```

Most built-in functions will return a NULL when there is a problem with an input parameter. In some instances, the CPcode Compiler will detect the error and reject the statement.

**For example:**

```c
# Most Built-in functions can return a NULL if they
# don't like the parameters. The compiler will actually
# reject this statement because it knows it is an error.
if (($length = listlen(42)) == NULL)
```

If you set a variable or any array element to NULL, that variable or array element will be discarded.

**For example:**

```c
# Make the variable $foo invalid
$foo = NULL;
# Make the variable in array at index '1' invalid
$array[1] = NULL;
```

NULLS are also commutative.

**For example:**

```c
$a = NULL;
# $b becomes invalid (and the equivalent of NULL)
$b = $a + 1;
```

### Reserved Words

Do not use CPcode statement names or built-in function names as variable names.

In addition, you cannot use these words as variable names:

- call
- continue
- filter
- func
- global
Constants

The CPcode Compiler evaluates operations on integer constants at compile time. Therefore, '1 + 1' translates to '2', and this example:

```cpp
if (1) {
    foo();
} else {
    bar();
}
```

translates to `foo();`

<table>
<thead>
<tr>
<th>Constant Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decimal</td>
<td>+ - 0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Hexadecimal</td>
<td>Prefix 0x and: 0 1 2 3 4 5 6 7 8 9 A B C D E F a b c d e f</td>
</tr>
<tr>
<td>IPv4host</td>
<td>Dotted quad format, 0 to 255. Example: 10.0.0.1</td>
</tr>
<tr>
<td>IPv4net</td>
<td>IPv4net constants in dotted quad (may use netmask notation) or Classless Inter-Domain Routing (CIDR) format. The CIDR element numbers are 0 to 32. Example: 192.0.2.0:255.255.255.0 (dotted quad with netmask) 192.0.2.0/24 (CIDR)</td>
</tr>
<tr>
<td>IPv6net</td>
<td>IPv6/NUM notation, where NUM is 1 to 128. Example: ::/64 ::1/128</td>
</tr>
<tr>
<td>MAC</td>
<td>MAC address, hexadecimal, 00 to ff, each octet must contain two digits. Example: 00:80:00:EA:8C:3A</td>
</tr>
<tr>
<td>String</td>
<td>Sequence of characters enclosed in double or single quotes. Can contain any single-byte character. If the string crosses lines, it will contain the newline character.</td>
</tr>
</tbody>
</table>
In a string constant, you can use special characters with the backslash to create a new meaning.

<table>
<thead>
<tr>
<th>Escape Character</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>\n</td>
<td>newline</td>
</tr>
<tr>
<td>\r</td>
<td>carriage return</td>
</tr>
<tr>
<td>\t</td>
<td>tab</td>
</tr>
<tr>
<td>\xnn</td>
<td>arbitrary byte value, where nn is any hex digit</td>
</tr>
<tr>
<td>\</td>
<td>backslash</td>
</tr>
<tr>
<td>\0</td>
<td>NULL byte</td>
</tr>
<tr>
<td>&quot;</td>
<td>double quote required when using a double quote inside a string enclosed with double quotes (for example, &quot;The &quot;Daily Report&quot;)</td>
</tr>
<tr>
<td>'</td>
<td>single quote required when using a single quote inside a string enclosed with single quotes (for example 'yesterday's data')</td>
</tr>
</tbody>
</table>

**Configuring Constants**

You can create a values file to configure specified values for a constant. The values file name must match the CPcode source file name, except for the file extension.

For example:

- CPcode file - time_travel.cpc
- Values file - time_travel.values

Content of time_travel.cpc:

```cpc
filter run dlp_dt() {
    echo("value of DELOREAN_SPEED is ", DELOREAN_SPEED);
}
```

Content of time_travel.values:

```plaintext
name DELOREAN_SPEED
text Speed
desc The speed the car needs to get to,
     desc in order to time travel.
mode scalar
88
```

**Note** - When you upload a CPcode file to SmartDashboard you must also upload the values file.
Chapter 4

CPcode Data Types

In This Chapter

<table>
<thead>
<tr>
<th>Type</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>array</td>
<td>21</td>
</tr>
<tr>
<td>ethmac</td>
<td>22</td>
</tr>
<tr>
<td>int</td>
<td>22</td>
</tr>
<tr>
<td>ipv4host and ipv6host</td>
<td>23</td>
</tr>
<tr>
<td>ipv4net and ipv6net</td>
<td>23</td>
</tr>
<tr>
<td>list</td>
<td>23</td>
</tr>
<tr>
<td>pattern</td>
<td>24</td>
</tr>
<tr>
<td>str</td>
<td>24</td>
</tr>
<tr>
<td>tag</td>
<td>25</td>
</tr>
</tbody>
</table>

This chapter details the types of data that CPcode supports - not to be confused with the Data Types that you are creating for DLP policies.

Parameters passed to, or returned from, CPcode functions and filters must be of a valid CPcode data type.

Most built-in CPcode functions return NULL when there is a problem with an input parameter. NULL is a special variable that is a member of every data type.

**array**

**Description**
A one-dimensional array of values

**Syntax**
array_name[element]

**Range**
The range for each element depends on the data type of that element.
For example: `b = a["hello"];`

**Comments**
- The index into the array can be any valid CPcode data type.
- Multiple data types can be used as an index into the same array.
- Each value in an array can be any valid CPcode data type.
- Data types within an array can be mixed.

To delete an element from an array, set the element to NULL.
To delete an array, set the array to NULL.
For example: `$array = NULL;`
```c
# define constants to use in emulating a structure
# with an array:
USERNAME = 1;
AUTHENTICATED = 2;
STATE = 3;
LINE = 4;

# build up the structure:
...
$fake_structure[USERNAME] = $username;
$fake_structure[AUTHENTICATED] = 0;
$fake_structure[STATE] = STATE_PENDING_AUTH;
...
if ($fake_structure[STATE] < STATE_MSG_DONE) {
    # Buffer up the current message for something
    # XXX Should be careful about the length
    $fake_structure[MSG] = cat($fake_structure[MSG], "new data")
} else {
    # Try each virus signature
    foreach $virus_sig inside (VIRUS_REGEXES) {
        if (regexec($virus_sig, $fake_structure[MSG])) {
            do_alert($fake_structure);
            break;
        }
    }
    # Free the current message.
    $fake_structure[MSG] = NULL;
}
```

**ethmac**

**Description**  An Ethernet MAC address.

**Syntax**  Six octets of hexadecimal values separated by colons.

**Range**  00:00:00:00:00:00 to ff:ff:ff:ff:ff:ff

**Example**  

```plaintext
# Correct usage
our_mac_address = 00:60:08:31:14:b7;

# Incorrect usage
our_mac_address = 0:60:8:31:14:b7;
```

**int**

**Description**  An integer.

**Syntax**  A decimal or hexadecimal representation written with a prefix of 0x.

**Range**  (-2^31) to (2^31 - 1) or -2,147,483,648 to +2,147,483,647

**Comments**  
- Octets with a value less than 0x10 must start with a leading zero.
- **int** is a signed data type, even when taken as a result of the **ulong** built-in function.

**Example**  

```
2001
0xAA
```
ipv4host and ipv6host

<table>
<thead>
<tr>
<th>ipv4host</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
</tr>
<tr>
<td><strong>Range</strong></td>
</tr>
<tr>
<td><strong>Example</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ipv6host</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Range</strong></td>
</tr>
</tbody>
</table>

ipv4net and ipv6net

<table>
<thead>
<tr>
<th>ipv4net</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Syntax</strong></td>
</tr>
<tr>
<td><strong>Range</strong></td>
</tr>
<tr>
<td><strong>Example</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ipv6net</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Range</strong></td>
</tr>
</tbody>
</table>

list

<table>
<thead>
<tr>
<th>list</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
</tbody>
</table>
**Syntax**

[element,element]

**Range**

The range for each element depends on the data type of that element. Data types may be mixed within a list. While a list can contain any number of elements, it is best to keep the list short.

**Comments**

Spaces are not allowed.

**Example**

```
interesting_ports = [21, 23, 80, 443, 2001];
a = [1, 2, 3];
$b=1;
a = [$b, $b+$b, 4-$b];
```

---

**pattern**

**Description**

A regular expression returned by the `regcomp()` or `pattern()` functions, that can be either passed to the `regexec()` function or directly compared against a string.

---

**str**

**Description**

An arbitrary sequence of bytes.

**Syntax**

- Text, enclosed within single or double quotes.
- Hex numbers for each byte, separated by a backslash, enclosed within single or double quotes

**Range**

"" (empty string) to no explicit limit

The str data type is not a C-style NULL-terminated string. It may contain binary data including any number of NULL characters.

The following C-style special characters can be used when creating a str data type:

- \n - newline
- \r - carriage return
- \t - tab
- \\ - backslash
- \xnn - an arbitrary byte value for this character, where nn is any hex digit

**Comments**

Only the CPcode application engine can generate a pattern data type. You cannot define it as a constant literal when developing CPcode or build it directly out of literals.

Operations on large strings can be memory intensive.

**Example**

```
foo = "string";
foo = "\x73\x74\x72\x69\x6e\x67";
```
**tag**

**Description**  Tags are descriptors for triggered filters, which you can use to turn off the filter during later execution.

**Comments**  There is an implicit parameter to filters - $1$ - which is the filter tag for the filter ("Special Local Variables" on page 17).
Chapter 5

CPcode Operators

In This Chapter

Arithmetic Precedence 26
Arithmetic Operators 27
Relational Operators 29
Logical Operators 33
Bitwise Operators 35
Other Operators 38

CPcode operators determine how CPcode operates on one or more expressions within a statement.

Arithmetic Precedence

<table>
<thead>
<tr>
<th>Order</th>
<th>Oper.</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>()</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[]</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>!</td>
<td>logical NOT</td>
</tr>
<tr>
<td></td>
<td>~</td>
<td>bitwise NOT</td>
</tr>
<tr>
<td>3rd</td>
<td>inside</td>
<td>Boolean, dependent upon whether an item is inside a list</td>
</tr>
<tr>
<td>4th</td>
<td>*</td>
<td>multiplication</td>
</tr>
<tr>
<td></td>
<td>/</td>
<td>division</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>modulus</td>
</tr>
<tr>
<td>5th</td>
<td>+</td>
<td>addition</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>subtraction</td>
</tr>
<tr>
<td>6th</td>
<td>&lt;&lt;</td>
<td>left bit shift</td>
</tr>
<tr>
<td></td>
<td>&gt;&gt;</td>
<td>right bit shift</td>
</tr>
<tr>
<td>7th</td>
<td>&lt;</td>
<td>relational less than</td>
</tr>
<tr>
<td></td>
<td>&lt;=</td>
<td>relational less than or equal to</td>
</tr>
<tr>
<td></td>
<td>&gt;</td>
<td>relational greater than</td>
</tr>
<tr>
<td></td>
<td>&gt;=</td>
<td>relational greater than or equal to</td>
</tr>
</tbody>
</table>
### Arithmetic Operators

In this section

<table>
<thead>
<tr>
<th>Operator</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>addition</td>
</tr>
<tr>
<td>-</td>
<td>subtraction</td>
</tr>
<tr>
<td>*</td>
<td>multiplication</td>
</tr>
<tr>
<td>/</td>
<td>division</td>
</tr>
<tr>
<td>%</td>
<td>modulus</td>
</tr>
<tr>
<td>-</td>
<td>negation</td>
</tr>
</tbody>
</table>

#### + (addition)

**Description**
The addition operator adds expression2 to expression1.

**Syntax**
```
int expression1 + int expression2
```

**Operands**
- int expression1 - an expression
- int expression2 - an expression

**Return Value**
```
int - the sum of the two expressions
```

**Example**
```
# This example adds 1 to the variable $count
$count = $count + 1;
```

#### - (subtraction)

**Description**
The subtraction operator subtracts expression2 from expression1.

**Syntax**
```
int expression1 - int expression2
```
Operands

- int expression1 - an expression
- int expression2 - an expression

Return Value

int - the difference between the two expressions

Example

# Subtract 1 from the $time_left variable
$time_left = $time_left - 1;

*(multiplication)*

Description

The multiplication operator multiplies expression1 by expression2.

Syntax

int expression1 * int expression2

Operands

- int expression1 - an expression
- int expression2 - an expression

Return Value

int - the product of the two expressions

Example

# Multiply number of minutes by 60
# to determine number of seconds
seconds = minutes * 60;

/(division)*

Description

The division operator divides expression1 by expression2

Syntax

int expression1 / int expression2

Operands

- int expression1 - an expression
- int expression2 - an expression

Return Value

int - the quotient

Example

# Divide number of seconds by 60
# to determine the # number of minutes
minutes = seconds / 60;

%(modulus)*

Description

The modulus operator divides expression1 by expression2 and returns the remainder.

Syntax

int expression1 % int expression2

Operands

- int expression1 - an expression
- int expression2 - an expression

Return Value

int - the remainder
Example  # Use division and modulus by 60
            # to convert seconds to minutes and seconds
            seconds = 318;
            minutes = seconds / 60;
            seconds = seconds % 60;

- (negation)

Description The negation operator displays the negative value of an expression.

Syntax
- int expression

Operands int expression - an expression

Return Value int - the negation of the expression

Example  # Assign the negative of variable $b to variable $a
            $a = -$b;

Relational Operators

In this section

== (equal to) 30
!= (not equal to) 31
> (greater than) 32
>= (greater than or equal to) 32
< (less than) 33
<= (less than or equal to) 33
== (equal to)

Description
The equal to operator compares expression2 to expression1 for equality, using different evaluation criteria for each data type.

- array - compares each index and the corresponding item
- ethmac - compares as a set of six octets. Each octet in expression1 must be equal to the matching octet in expression2
- int - compares by value
- ipv4host - compares as a set of four octets. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv4net - compares as a set of eight octets or a set of four octets and the netmask. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv6host - compares as a set of 16 octets. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv6net - compares as a set of 32 octets or a set of four octets and the netmask. Each octet in expression1 must be equal to the matching octet in expression2.
- list - compares by each element of each list. Each element in expression1 must be equal to the matching element in expression2
- pattern - compares by pattern
- str - compares by value, with regard to case

You can compare expressions that have different data types, for example:

- ipv4host and ipv4net - if the host is in the network, the expressions are considered equal
- str and pattern - if the string is the same as the pattern, the expressions are considered equal

Syntax

```
expression1 == expression2
```

Operands

- expression1 - an expression
- expression2 - an expression

The following data types are valid for these operands:

- array
- ethmac
- int
- ipv4host
- ipv4net
- ipv6host
- ipv6net
- list
- pattern
- str

Return Value

- 1 - expression1 is equal to expression2
- 0 - expression1 is not equal to expression2
Example

```plaintext
# This example compares the subject to NULL, and if it is
# not NULL, then other processing occurs
if ($subject == NULL)
    $subject = $1;
else
    $subject = cat( $subject, $1);
return;
```

!= (not equal to)

Description

The not equal operator compares expression2 to expression1 for inequality, using different evaluation criteria for each data type.

- array - compares each index and the corresponding item
- ethmac - compares as a set of six octets. Each octet in expression1 must be equal to the matching octet in expression2
- int - compares by value
- ipv4host - compares as a set of four octets. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv4net - compares as a set of eight octets or a set of four octets and the netmask. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv6host - compares as a set of 16 octets. Each octet in expression1 must be equal to the matching octet in expression2.
- ipv6net - compares as a set of 32 octets or a set of four octets and the netmask. Each octet in expression1 must be equal to the matching octet in expression2.
- list - compares by each element of each list. Each element in expression1 must be equal to the matching element in expression2
- pattern - compares by pattern
- str - compares by value, with regard to case

You can compare expressions that have different data types, for example:

- ipv4host and ipv4net - if the host is not in the network, the expressions are not equal.
- str and pattern - if the string is not the same as the pattern, the expressions are not equal.

Syntax

expression1 != expression2
Operands

- expression1 - an expression
- expression2 - an expression

The following data types are valid for these expressions:
- array
- ethmac
- int
- ipv4host
- ipv4net
- ipv6host
- ipv4net
- list
- pattern
- str

Return Value

- 1 - expression1 is not equal to expression2
- 0 - expression1 is equal to expression2

Example

# Compare value in $s_nevermind to 1
if ($s_nevermind != 1)
    webdet_server_process($v);

> (greater than)

Description

The greater than operator compares expression2 to expression1 and determines whether expression1 is greater than expression2.

Syntax

int expression1 > int expression2

Operands

- int expression1 - an integer expression
- int expression2 - an integer expression

Return Value

- 0 - expression1 is not greater than expression2
- 1 - expression1 is greater than expression2

Example

# Compare string length of variable $blob to 2048
if (strlen($blob) > 2048)
    $blob = "";

>= (greater than or equal to)

Description

The greater than or equal to operator compares expression2 to expression1 and determines whether expression1 is greater than or equal to expression2.

Syntax

int expression1 >= int expression2

Operands

- int expression1 - an integer expression
- int expression2 - an integer expression

Return Value

- 0 - expression1 is not greater than or equal to expression2
- 1 - expression1 is greater than or equal to expression2
Example  # Compare return value of the byte function to 0
if (byte ($sblob, bloblimit) >= 0)
    $state = -1; ""

< (less than)

Description  The less than operator compares expression2 to expression1 and determines whether expression1 is less than expression2.

Syntax  int expression1 < int expression2

Operands  • int expression1 - an integer expression
          • int expression2 - an integer expression

Return Value  • 0 - expression1 is not less than expression2
              • 1 - expression1 is less than expression2

Example  # Compare value in $state variable to 0
if ($state < 0) return;

<= (less than or equal to)

Description  The less than or equal to operator compares expression2 to expression1 and determines whether expression1 is less than or equal to expression2.

Syntax  int expression1 <= int expression2

Operands  • int expression1 - an integer expression
          • int expression2 - an integer expression

Return Value  • 0 - expression1 is not less than or equal to expression2
              • 1 - expression1 is less than or equal to expression2

Example  # Compare the string length of the variable $cmd to 0
if (strlen ($cmd) <= 0)

Logical Operators

In this section

&& (logical AND)  34
|| (logical OR)  34
! (logical NOT)  35
&& (logical AND)

**Description**
The logical AND operator evaluates both sides of the expression. If both expressions evaluate to true (they are non-zero and non-NULL), the result is true. If either, or both, expression evaluates to false, the result is false.

The CPcode logical AND operator, like the C language && operator, is a short-circuit operator. If the first expression evaluates to false, the second expression is not evaluated.

**Syntax**
expression1 && expression2

**Operands**
- expression1 - an expression
- expression2 - an expression

**Return Value**
- 0 - the expression is false
- 1 - the expression is true

**Example**
```plaintext
if (ALERT_ON_ROOT && index($data, "root") >= 0) {
    alert(...);
}
```

# If ALERT_ON_ROOT is set to non-zero, index() will be called.
# If ALERT_ON_ROOT is set to zero, the value of the conditional will be false and index() will not be called.
# Compiler generates code similar to:
# if (ALERT_ON_ROOT) {
#    if (index($data, "root") >= 0) {
#        alert(...);
#    }
# }

|| (logical OR)

**Description**
The logical OR operator evaluates both sides of the expression. If either or both expressions evaluates to true (non-zero and non-NULL), the result is true.

This operator is a short circuit operator, like the || operator in the C language. If the first expression evaluates as true, the second expression is not evaluated.

**Syntax**
expression1 || expression2

**Operands**
- expression1 - an expression
- expression2 - an expression

**Return Value**
- 0 - the expression is false
- 1 - the expression is true

**Example**
```plaintext
# because of short circuiting, strlen() will not be called in this
# example
$string = "foo";
if ($string || strlen($strlen) < 20)
```
! (logical NOT)

Description  The logical NOT operator computes the logical negation of the expression. True indicates that the expression is zero or NULL.

If the expression is true, the result is false.
If the expression is false, the result is true.

Syntax

! expression

Operands  expression - an expression

Return Value

- 0 - the expression is not zero
- 1 - the expression is zero

Example

$string = "foo";
if (! $string || strlen(string) < 20)

Bitwise Operators

In this section

& (bitwise AND) 35
| (bitwise OR) 36
~ (bitwise NOT) 36
^ (bitwise XOR) 37
<< (left shift) 37
>> (right shift) 37

& (bitwise AND)

Description  The bitwise AND operator evaluates identically positioned bits in both expressions. If both bits are 1, the operator sets the corresponding bit in the result to 1. If either bit is 0, the operator sets the corresponding bit in the result to 0. The following table summarizes this behavior:

<table>
<thead>
<tr>
<th>If bit in expression1 is</th>
<th>and bit in expression2 is</th>
<th>then bit in result will be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Syntax

int expression1 & int expression2

Operands

- int expression1 - an expression
- int expression2 - an expression

Return Value

- int - the bitwise evaluation of both expressions
Example

# Bitwise AND on the value in the variable foo
# and the number 1
if ((foo & 1) == 1)
    echo ("foo is odd\n");

| (bitwise OR)

Description

The bitwise OR operator evaluates identically positioned bits in both expressions. If both bits are 0, the operator sets the corresponding bit in the result to 0. If either bit is 1, the operator sets the corresponding bit in the result to 1.

Syntax

```
type expression1 | type expression2
```

Operands

- `type expression1` - an expression
- `type expression2` - an expression

Return Value

`type` - the bitwise evaluation of the two expressions

Example

# This example performs a bitwise OR on the values in the variables a and b, and displays the result of 7
a = 5;
b = 3;
c = a | b;
echo ("c is ", c, "\n");

~ (bitwise NOT)

Description

The bitwise NOT operator inverts the bit values in an expression. If the bit is 0, the operator sets the corresponding bit in the result to 1. If the bit is 1, the operator sets the corresponding bit in the result to 0. The following table summarizes this behavior:

<table>
<thead>
<tr>
<th>If bit in expression is</th>
<th>then bit in result will be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Syntax

```
~ type expression
```

Operands

`type expression` - an expression

Return Value

`type` - the bitwise evaluation of the expression

Example

# Do a bitwise NOT on the variable a where a = ~ a;
# this example clears the least significant bit in the variable b
b = b & ~ 1;
**^(bitwise XOR)**

**Description**
The bitwise XOR operator evaluates identically positioned bits in both expressions. If both bits are the same, the operator sets the corresponding bit in the result to 0. If the bits are different, the operator sets the corresponding bit in the result to 1. The following table summarizes this behavior:

<table>
<thead>
<tr>
<th>If bit in expression1 is</th>
<th>and bit in expression2 is</th>
<th>then bit in result will be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Syntax**
```
int expression1 ^ int expression2
```

**Operands**
- int expression1 - an expression
- int expression2 - an expression

**Return Value**
int - the bitwise evaluation of the two expressions

**Example**
```
# Do a bitwise XOR on the low 4 bits of the value in
# the variable $a
$a = $a ^ 0xf;
```

---

**<< (left shift)**

**Description**
The left shift operator shifts all of the bits in expression1 to the left by expression2 bits.

**Syntax**
```
int expression1 << int expression2
```

**Operands**
- int expression1 - an expression
- int expression2 - an integer representing the number of bits by which expression1 should be shifted

**Return Value**
int - the result of the bitwise shift

**Example**
```
# This example determines whether the Nth bit is set
$tmp = 1 << Nth;
if ($1 & $tmp)
    echo ("Nth bit set\n");
```

---

**>> (right shift)**

**Description**
The right shift operator shifts all of the bits in expression1 to the right by expression2 bits. This is a logical shift, (for example, 0x80000000 >> 1 == 0x40000000).

**Syntax**
```
int expression1 >> int expression2
```

**Operands**
- int expression1 - the integer to be shifted
- int expression2 - an integer representing the number of bits by which expression1 should be shifted

**Return Value**
int - the result of the bitwise shift
Example

# Find the version of the IP packet payload contained in the # variable $1
$version = ubyte($1,0) >> 4;

Other Operators

In this section

= (assignment) 38
inside 39
?: (conditional) 39

= (assignment)

Description The assignment operator sets variable equal to expression. You can use assignments within expressions (often referred to as built-in assignments).

Syntax
variable = expression

Operands
- variable - a global variable, local variable, or array element
- expression - an expression. All data types are valid.

Return Value
variable - the value assigned by the expression

Note - Built-in assignments must be enclosed inside parentheses, for example:
# this is correct:
$foo = ($bar = NULL);

Example

# Use the assignment operator to set variables to # different data types
$services[ "telnet" ] = 1;
router_mac_address = 00:c0:7b:6b:ee:2b;
$count = 318830;
router_IP_address = 208.244.85.1;
our_network = 208.244.85.0:255.255.255.0;
our_network = 208.244.85.0/24;
web_ports = [ 80, 8000, 8080, 2001 ];
company_name = "Check Point";
# The value of the Built-in assignment expression will # be the left hand side of the assignment after it is assigned.
$foo = 1;
$bar = ($foo = $foo + 1);
# The value of $bar is 2
# The value of $foo is also obviously 2
inside

Description The inside operator searches for expression1 in the list represented by expression2. The result is true if expression1 is an element of the list represented by expression2.

Syntax expression1 inside expression2

There must be a space on either side of the inside operator.

Do not confuse the inside operator, whose role is to return a Boolean value, with "inside" used in the syntax of a foreach or declare statement.

Operands
- expression1 - an expression. Valid data types are any that can be added to a list.
- expression2 - a list

Return Value
- 0 - expression1 is not found inside expression2
- 1 - expression1 is found inside expression2

Example
# Assume $path contains extracted path element
# from a data source to test against.
# This example looks at $path
# and tests it against elements in a list variable,
# and if a match is found,
# sets a local variable to 1.

SPOOLER_WINDOWS_ROOT_PATHS="c:\winnt","c:\windows"; $bad = 0;

foreach $path_var inside (SPOOLER_WINDOWS_ROOT_PATHS) {
    if (cat($path_var,"\system32") == $path) {
        $bad = 1;
        break;
    }
}

?: (conditional)

Description The conditional operator evaluates an expression for a condition. It functions as an IF-THEN-ELSE shortcut.

Syntax variable = expression ? value1 : value2

Operands
- expression - the expression to be evaluated
- value1 - the expression that will be assigned to ret_value if expression is true
- value2 - the expression that will be assigned to ret_value if expression is false

Return Value variable - the value returned from the evaluation, which is value1 if expression is true or value2 if expression is false

Note - CPcode does not support the use of the conditional operator on the left-hand side of an expression. For example: ((bar > 50) ? bar : $foo) = $counter +1; is not allowed.
Example

# A counter that will forgive you for starting at NULL
# 'counter = counter + 1;'
# will not work unless counter initialized to be
# something other than NULL at some point
counter = counter ? counter + 1 : 1;
# To limit $foo to numbers less than or equal to 100
foo = bar < 100 ? bar : 100;
Chapter 6

CPcode Statements

In This Chapter

= (assignment) 41
{} (block) 41
break 42
foreach 42
func 43
if 44
requires 44
return 45
while 45

This section lists valid CPcode statements.

If you are familiar with the C language, you should exercise caution when writing CPcode. While CPcode statements appear to be C-like, there are differences. CPcode may not support some C language statements.

= (assignment)

Description  An assignment statement assigns the value on the right-hand side of the equal sign (=) to the variable on the left-hand side. You can use the assignment statement to set the value of a variable.

Syntax  
variable = expression;

Parameters  
- variable - the variable to which the expression will be assigned
- expression - the expression to be assigned to the variable

Example  
# Use the assignment statement to set the value of our_router to 192.168.10.1.
our_router = 192.168.10.1;

{} (block)

Description  Block notation groups together multiple CPcode statements. The block starts with the left curly bracket and ends with the right curly bracket. All statements inside these brackets are part of the block.

Syntax  
{statement_list}

Parameters  statement_list - a list of CPcode statements
Example
# The block statement groups together several
# statements into a user-defined function
func counter
{
    return ($count = $count + 1);
}

break
Description
The break statement ends execution of the innermost foreach or while statement in
which it appears. You can use the break statement to exit from a loop.
Syntax
break;
Parameters
N/A
Example
# Go through the buffer line-by-line.
# If check_line finds what it is looking for,
# it will return 1, and we should exit
while ($eol = index($buf, "$n") > -1) {
    $line = substr($buf, 0, $eol);
    $buf = substr($buf, $eol + 1);
    if (check_line($line)) {
        break;
    }
}

foreach
Description
The foreach statement executes the same statements for each element of a list or
indices of an array.
When used with a list, the foreach statement sets the variable to the elements of a list
and executes the statements once for each element. The sensor execution engine
processes the elements in the order in which they appear in the list.
When used with an array, the foreach statement sets a variable to the index value of the
array and executes the statements once for each index value. The order in which the
index values appear is not defined.
Syntax
foreach $variable inside [list|array] { statements }
Parameters
- $variable - a local variable
- list - a variable, constant, or expression that evaluates to a list
- array - a variable, constant, or expression that evaluates to an array
- statements - statements to be executed for each element in the list
Notes
Do not modify a list or array inside the loop. While the CPcode Compiler normally prevents this, it could cause stability issues with the gateway or appliance. For example, do not code:

```cpcode
foreach $foo inside array {
    if (array[$foo] > 1)
        array[$foo] = NULL;
}
foreach $foo inside list {
    list = listadd(list, $foo * 2);
}
```

Example

```cpcode
# Do a set of actions for each recipient listed in a mail message
foreach $tmp inside ($to_list) {
    if (mail_list:present)
        mail_list:rec($start_time, $src_addr, $fromaddr, $dst_addr, $tmp, $subject);
    if (mail_od:present)
        mail_od:rec($start_time, $src_addr, $fromaddr, $dst_addr, $tmp, $subject);
    if (mail_tricks:present)
        mail_tricks:rec($start_time, $src_addr, $fromaddr, $dst_addr, $tmp, $subject);
}

# Do a set of actions for each element in an array.
$services["telnet"] = 1;
$services["http"] = 2;
$services["ssl"] = 3;
foreach $tmp inside ($services) {
    echo( $tmp, " is element ", $services[$tmp], "\n" );
}
```

This function prints:

telnet is element 1
http is element 2
ssl is element 3

func

Description
The func statement declares the name and statements that form the body of a function. In most cases, the function declaration contains multiple statements.

Syntax
```
func name {{statements}}
```

Parameters
- name - the name of the function
- statement - one or more statements to be executed within the function

Example
```
func counter
{
    return $1 + 1;
}
```
if

Description  The if statement evaluates an expression, and if the expression evaluates to true, executes a statement. If the expression evaluates to false, the statement is not executed. If the value of the expression evaluates to false and there is an else clause, the else clause statements are executed.

Syntax
- if (expression) statement
- if (expression) statement else statement

Parameters
- expression - the expression to be evaluated
- statement - the statement to be executed if the expression is true

Parentheses are required around the expression parameter.

If the expression being evaluated is complex (for example, contains && or ||), the evaluation will terminate at the earliest opportunity.

Notes
In the following example, the if statement is used to determine if there is any data in the $blob variable. If the $blob variable does not contain any data, the statement assigns data to it. If the $blob variable does contain data, the statement after the else clause adds to the data already in the variable.

```cpp
if ($blob == NULL)
    $blob = $my_data;
else
    $blob = cat ($blob, $my_data);
```

Example

```cpp
# some_function will never be called because its results will not affect the value of the expression.

$false = 0;
$true = 1;
if ($false && some_function() != NULL { }
if ($true || some_function() != NULL { }
```

requires

Description  The "requires" statement specifies variables that must be set. If any variables are NULL, the rest of the CPcode is not loaded.

You can use the "requires" statement to control whether a protocol or Protection Group is run. If a required variable is not set, the Protection Group or protocol is not loaded.

Syntax

```
requires variable1 variable2 variable3;
```

Parameters
- variable_ - one or more variables that must contain non-NULL values before the remaining CPcode is loaded.

Example

```cpp
# The requires statement specifies that the value for the ourdomainname variable must be set for the protocol to load.

requires ourdomainname;
```
**return**

**Description**  
The return statement exits a function and, optionally, returns a value.

**Syntax**  
return;
return [expression];

**Parameters**  
expression - any expression

**Example**  
# The return statement exits the if statement and
# returns the value in the $line variable.
$xxx = index($line, ">");
if ($xxx < 0)
  return $line;

**while**

**Description**  
The while statement evaluates an expression and repeatedly executes the statement as long as the expression remains true, as long as it is non-zero and non-NULL.

**Syntax**  
while (expression) statement;

**Parameters**  
- expression - expression to be evaluated
- statement - the statement to be executed

**Example**  
The while statement walks along the contents of $cblob. It starts at whatever value is in $n when the while statement is entered. As long as the $nth byte of $cblob is greater than zero, it keeps going. When it sees a byte that is less than or equal to zero, it stops.
while (byte($cblob,$n) > 0) {
  $n = $n + 1;
}
Chapter 7

CPcode Built-in Variables

In This Chapter

System Variables 46
Feature Variables 46
Global Variables 47

This section lists the packet variables of interest to CPcode developers. Built-in variables gather data from the packet just received and provide easy access to information about the connection stream.

System Variables

System variables extract data from the system running a Check Point appliance or gateway.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>system.maxticks</td>
<td>Returns the maximum number of ticks that can be used.</td>
<td>int</td>
</tr>
<tr>
<td>system.ticks</td>
<td>Returns the current number of ticks that the CPcode has used out of its allotted amount.</td>
<td>int</td>
</tr>
<tr>
<td>system.time</td>
<td>Returns the system time, in seconds, since the beginning of the epoch (Jan 1 1970 UTC).</td>
<td>int</td>
</tr>
</tbody>
</table>

Feature Variables

Feature variables provide access to supported execution engine features and revision information (such as when a built-in CPcode function was changed).

Feature variables are evaluated at compile time. For example:

```plaintext
if (feature.filename >= 20030412) {
    #Supports new Microsoft OS's filename scheme
    $file = filename($line);
} else {
    #Do it in CPcode
    file = filename_ms($line);
}
```

translates at compile to either:

```plaintext
$file = filename($line);
```
or

```plaintext
$feature = feature;
```
$file = filename_ms($line);

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>feature.&lt;function_name&gt;</td>
<td>Information about when the built-in CPcode function specified by <code>&lt;function_name&gt;</code> was last changed. <code>&lt;function_name&gt;</code> must be the name of a valid, built-in function.</td>
<td>int</td>
</tr>
</tbody>
</table>

### Global Variables

DLP functions in CPcode can use any of the CPcode functions and variables that are not traffic related (notice that UDP and TCP variables are not supported).

In addition, there are DLP specific variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>policy.internal_domains</td>
<td>List of domains (strings), that come from &quot;My Organization&quot; object in the policy.</td>
<td>[&quot;domain1.com&quot;,&quot;domain2.com&quot;]</td>
</tr>
<tr>
<td>message.protocol</td>
<td>IP protocol.</td>
<td>&quot;SMTP&quot; &quot;FTP&quot; &quot;HTTP&quot;</td>
</tr>
<tr>
<td>message.source</td>
<td>The source of the message.</td>
<td>&quot;<a href="mailto:dan@domain.com">dan@domain.com</a>&quot; &quot;190.1.1.1&quot;</td>
</tr>
<tr>
<td>message.source_username</td>
<td>The username of the message sender or an empty string if the sender’s email is not mapped to a username.</td>
<td>&quot;dan&quot;</td>
</tr>
<tr>
<td>message.dests</td>
<td>List of destinations.</td>
<td>In SMTP: [&quot;<a href="mailto:sam@domain.com">sam@domain.com</a>&quot;,&quot;<a href="mailto:joe@domain.com">joe@domain.com</a>&quot;] In FTP: [&quot;154.1.0.0&quot;,&quot;154.1.1.0&quot;]</td>
</tr>
<tr>
<td>message.dests_username</td>
<td>The username of the message’s destination or an empty string if the recipient’s email is not mapped to a username.</td>
<td>[&quot;sam&quot;, &quot;joe&quot;]</td>
</tr>
<tr>
<td>message.mail_subject</td>
<td>For SMTP - contains the mail subject (in the recipient's part). For HTTP/FTP - empty.</td>
<td>[&quot;RE: New List&quot;]</td>
</tr>
<tr>
<td>message.header_names</td>
<td>Contains protocol header fields.</td>
<td>MIME headers</td>
</tr>
<tr>
<td>message.header_values</td>
<td>Contains protocol header values.</td>
<td>MIME headers</td>
</tr>
<tr>
<td>message.num_of_files</td>
<td>For SMTP - Contains the number of attachments</td>
<td>3 - 3 attachments</td>
</tr>
<tr>
<td></td>
<td>For HTTP/FTP - Contains the number of files</td>
<td>3 - 3 files</td>
</tr>
<tr>
<td>message.uid</td>
<td>ID of the message</td>
<td>&quot;[5483F78E-D5AF-8E60-CB28-283968ED426A]&quot;</td>
</tr>
<tr>
<td>message.mail_body</td>
<td>Contains the text of the mail body. Relevant only for SMTP messages.</td>
<td>&quot;Hello my name is Alice … Regards, Alice&quot;</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>part.content</td>
<td>A buffer containing the text of a part. It is either the text of a mail's body, or text of a file (attached to mail, uploaded in http or ftp).</td>
<td>&quot;This is the body.... ..... Regards &quot;</td>
</tr>
<tr>
<td>part.id</td>
<td>Running index stating the part number in a file (starting from 0)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>part.type</td>
<td>Describes the type of the part.</td>
<td>&quot;Recipients&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Content&quot;</td>
</tr>
<tr>
<td>part.file_type_id</td>
<td>An identifier for file type. See “Defining New File Types” in the DLP Administration Guide (<a href="http://supportcontent.checkpoint.com/solutions?id=sk76540">http://supportcontent.checkpoint.com/solutions?id=sk76540</a>) for the full list of different types.</td>
<td>11 for “SmartWare”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>311 for “Quattro Pro Win 12.0”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>169 for “Java Script”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“OfficeWrite”</td>
</tr>
<tr>
<td>part.file_family_type</td>
<td>The main family/category type of the file.</td>
<td>“Presentation”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Word Processor”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The complete list is in Note 3 below.</td>
</tr>
<tr>
<td>part.name</td>
<td>The name of the attachment or the name of the file uploaded.</td>
<td>&quot;body&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;file.doc&quot;</td>
</tr>
<tr>
<td>part.is_binary</td>
<td>This variable equals 1 when the part.content is the binary representation of the file. If the variable is 0, part.content contains the textual representation of the file.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>part.is_file</td>
<td>This variable equals 1 when the scanned part is an attachment, otherwise it equals 0.</td>
<td>Relevant only for SMTP.</td>
</tr>
<tr>
<td>part.is_corrupted</td>
<td>This variable equals 1 when the scanned part is considered corrupted. Files are considered corrupted files when DLP cannot extract their content. This variable is relevant to CPcode that runs when the Data Type is matched (Cpcode defined in Advanced Matching).</td>
<td>1 - File is corrupted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 - File is not corrupted</td>
</tr>
<tr>
<td>part.sub_filenames</td>
<td>This field is a list of the names of the sub-files.</td>
<td>Images in Word Document: [&quot;object001&quot;,&quot;object002&quot;,&quot;object3&quot;]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Files in an archive: [&quot;notepad.exe&quot;,&quot;mydoc.doc&quot;,&quot;demo.tar&quot;]</td>
</tr>
<tr>
<td>part.sub_family_types</td>
<td>This field is a list of the main family names of the sub-files.</td>
<td>[&quot;Images&quot;,&quot;Images&quot;,&quot;Images&quot;]</td>
</tr>
</tbody>
</table>
### CPcode Built-in Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>part.sub_file_size</td>
<td>This field is a list of the file sizes in bytes of the sub-files.</td>
<td>[156, 41234, 36552 ]</td>
</tr>
<tr>
<td>verify.str</td>
<td>A string that was matched to a pattern (The pattern was defined in a data type).</td>
<td>&quot;0404787875&quot;</td>
</tr>
<tr>
<td>verify.index</td>
<td>An index in CONTENT variable to the end of the matched string.</td>
<td>50</td>
</tr>
</tbody>
</table>

**Notes:**

1. Variables marked with a 1 require the existence of one of these:
   - A Cooperate Template Data Type in the DLP policy Rule Base with the `consider template images` feature enabled.
   - A File Marking rule in the DLP policy Rule Base.
   - `always_extract_embedded_images` field in `dlp.conf` is set to 1.
2. For all part.sub variables:
   - A file can contain several sub-files. All sub variables refer to these sub-files:
     - Embedded images in office document.
     - Files listed in archives.
     - There might be cases where names are automatically generated.
     - Elements in the lists `part.sub_filenames`, `part.sub_family_types` and `part.sub_file_size` correspond to each other.
   - The fields are relevant only when `part.type` equals “Content”.
   - Sub-files of Archives are relevant only for CPcode that runs when the Data Type is matched (Cpcode defined in Advanced Matching).
3. For `part.file_family_type`, the family can be one of these:
   - "None"
   - "Word Processor"
   - "Spreadsheet"
   - "Presentation"
   - "Viewer"
   - "Text"
   - "Markup"
   - "Multimedia"
   - "Email"
   - "Images"
   - "Executable"
   - "Archive"
   - "Database"
   - "V-Calendar"
   - "Drawings"
   - "Graphic Design"
   - "JPEG"
   - "GIF"
   - "TIF"
   - "ZIP"
   - "RAR"
   - "TAR"
   - "PDF"
   - "Visio"
   - "Professional Images"
   - "Microsoft Office files"
   - "Unrecognized"
   - "Unsupported Archive"
   - "Other"

**Example with Global Variables**

This script shows a CPcode Data Type that is turned on when there are more than 2 external domains in an Email message.
filter run dlp_dt() {
    # This data type is active for SMTP only, for the recipient part
    if( message.protocol != "SMTP" || part.type != "Recipients" )
        return 0;
    $num_of_external_domains = 0;
    $external_domain_list = [];

    foreach $recipient inside message.dests {
        # get the domain of the mail address
        $domain = global:get_domain_from_address( $recipient );

        # check if the domain is external
        if( !global:is_internal_domain($domain) ) {
            if( !global:is_member( $external_domain_list , $domain ) ) {
                # Add it to the external domains list
                $external_domain_list = listadd( $external_domain_list ,
                $domain );
                # Check if we reached the max of external domains
                $num_of_external_domains = $num_of_external_domains + 1;
                if( $num_of_external_domains >= MAX_EXTERNAL_DOMAINS )
                    return 1;
            }
        }
    }
    return 0;
}
Chapter 8

CPcode Built-in Functions

In This Chapter

align() 53
arrayindices() 53
atoi() 54
blobbytes() 54
byte() 55
caseprefix() 55
casesuffix() 56
cat() 56
csv_close() 57
csv_getcols() 57
csv_getrows() 58
csv_getval() 58
csv_open() 59
debase64() 59
dehex() 60
deutf8() 60
dns_expand() 60
dns_next() 61
elem() 61
erro() 62
ethmac() 62
filename() 63
formattedtime() 63
hatoi() 64
host() 65
host6() 65
incksum() 65
index() 66
io_close() 66
io_delete() 66
io_open() 67
io_read() 67
io_readlines() 67
io_write() 68
kiss_compile() 68
kiss_pattern() 68
kiss_search() 69
le_long() 69
le_short() 69
le_ulong() 70
le_ushort() 70
listadd() 70
listcombine() 71
listglom() 71
listlen() 71
long() 72
lsort() 72
luniq() 73
match() 73
multi_compile() 73
multi_escape() 73
multi_pattern() 74
multi_search() 75
net() 76
net6() 77
pattern() 77
pred() 77
prefix() 78
prunearray() 78
regcomp() 78
regexec() 79
scope() 80
sed() 80
short() 81
split() 82
str() 82
strcasecmp() 82
strerror() 83
strlen() 83
strstr() 83
strtype() 84
sublist() 85
substr() 86
succ() 86
suffix() 87
tolower() 87
toupper() 88
typeof() 88
ubyte() 89
ulong() 89
unpack() 90
url_flush() 90
url_needs_refresh() 91
url_open() 91
url_readall() 92
url_readlines() 92
ushort() 92
**align()**

The `align()` function bit aligns data items that are larger than one byte (for example, a record where a four-byte integer only occurs at an address that is a multiple of four bytes).

### Syntax

```
align(int num, [int hash :=4]);
```

### Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>int num</td>
<td>The integer to align.</td>
</tr>
<tr>
<td>int hash</td>
<td>The single bit bitmask to align on.</td>
</tr>
</tbody>
</table>

### Return Value

`int` - The aligned integer.

### Example

```
align(0) == 0
align(1) == 4
align(2) == 4
align(3) == 4
align(4) == 4
align(5) == 8
align(0,2) == 0
align(1,2) == 2
align(2,2) == 2
align(3,2) == 4
align(4,2) == 4
```

---

**arrayindices()**

The `arrayindices()` function creates a sorted list of an array's associative indices.

### Syntax

```
arrayindices(array arr);
```

### Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>array arr</td>
<td>The array from which to create the list.</td>
</tr>
</tbody>
</table>

### Return Value

- `list` - a list of all the array's indices

It is illegal to modify an array being iterated through using a `foreach` statement. Therefore, it is necessary to first create a list of all the array's indices, and then use `foreach` to iterate through that list.
Example

```php
array[0] = "zero";
array[1] = "one";
foreach $iterator inside (arrayindices(array)) {
    echo("array[", $iterator, "] = ", array[$iterator], 
        
    );
    # Output:
    # array[0] = zero
    # array[1] = one
    # The next statement would be illegal if iterating
    # over the array instead of a copy of its indices.
    array[$iterator] = NULL;
}
```

**atoi()**

The atoi() function converts a string to an integer representation. If you specify an index, the conversion begins at that location inside the string.

The atoi() function ignores any white space leading up to the number.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>atoi(string str, [int index := 0]);</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>string str</td>
<td>The string that is to be converted to an integer.</td>
</tr>
<tr>
<td>int index</td>
<td>The point in the string at which to begin the conversion.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Return Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>• int - the integer value of the string, str</td>
</tr>
<tr>
<td>• 0 - the string is empty, does not contain digits, or contains non-digits before the first digit in the string</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>atoi(&quot;192&quot;) == 192</td>
</tr>
<tr>
<td>atoi(&quot;   192   &quot;) == 192</td>
</tr>
<tr>
<td>atoi(&quot;  1 9 2 &quot;, 4) == 2</td>
</tr>
<tr>
<td>atoi(&quot;192ABC&quot;) == 192</td>
</tr>
<tr>
<td>atoi(&quot;ABC&quot;) == 0</td>
</tr>
<tr>
<td>atoi(&quot;ABC192&quot;, 3) == 192</td>
</tr>
<tr>
<td>atoi(&quot;&quot;&quot;) == 0</td>
</tr>
<tr>
<td>atoi(42) == error</td>
</tr>
</tbody>
</table>

**blobbytes()**

The blobbytes() function creates a blob from the byte values of an expression. Integers are an exception - their lowest significant bytes are used. For example, if the integer is 0x23, only one byte is used; if the integer is 0x1234, two bytes are used; etc.

The cat() function should, preferably, be used when concatenating strings.

The blobbytes() function has been optimized to generate a blob from non-strings, and is usually used to generate a hash index for an array.
### CPcode Built-in Functions

#### blobbytes(...);

**Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>The values to be added to the blob/string.</td>
</tr>
</tbody>
</table>

**Return Value**

- string - a blob/string
  
  The overall return string for the blobbytes() function cannot be larger than 8KB.

**Example**

```plaintext
# Get first 4 bytes of a data packet
# but swap them around
$a = udp.blob;
$b = blobbytes(substr(a, 3, 1), substr($a, 2, 1),
substr($a, 1, 1), substr($a, 0, 1);
```

#### byte()

The byte() function returns a byte from a specified offset in a string. The returned value is signed.

You can use the ubyte() function to return an unsigned value.

The byte() function is useful for determining if all of the desired fields have been gathered (for example, to determine if the byte at offset 40 in a particular string contains the expected value).

**Syntax**

```
byte(string str, int offset);
```

**Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>string str</td>
<td>The string to be examined. The string begins at byte 0.</td>
</tr>
<tr>
<td>int offset</td>
<td>The offset location in the string for the byte to be returned.</td>
</tr>
</tbody>
</table>

**Return Value**

- int - the offset in the string as a signed value
- error - the offset was out of bounds

**Example**

```plaintext
# the eighth byte is the signed type code
$code = byte(udp.blob, 7);
```

#### caseprefix()

The caseprefix() function compares two strings and returns true if the second string is a prefix of the first string, without regard to case.

Unlike the index() and prefix() functions, the caseprefix() function performs a case-insensitive comparison of the second argument against the beginning of the first argument.

You can use the prefix() function for a case-sensitive comparison.

**Syntax**

```
caseprefix(string big, string little);
```
### casesuffix()

The `casesuffix()` function compares two strings and returns true if the second string is a suffix of the first string, without regard to case.

You can use the `suffix()` function for a case-sensitive comparison.

#### Syntax

```
casesuffix(string big, string little);
```

#### Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>string big</td>
<td>The larger string, which is the one compared against.</td>
</tr>
<tr>
<td>string little</td>
<td>The smaller string, which is compared to the end of big.</td>
</tr>
</tbody>
</table>

#### Return Value

- **1** - little is the same as the end of big, regardless of case
- **0** - little is not the same as the end of big

#### Example

```
casesuffix("ABCDEFG", "EFG") == 1
casesuffix("ABCDEFG", "efg") == 1
casesuffix("ABCDEFG", "ABC") == 0
```

### cat()

The `cat()` function concatenates arguments by appending them into a single string. Non-string arguments are converted to their textual representation before concatenation. For list arguments, each element in the list is appended to the new string.

#### Syntax

```
cat(...);
```

#### Parameters

```
... - the arguments to concatenate
```
Return Value
string - a blob/string

The cat() function will ignore NULL parameters. Therefore, a statement such as the following would be safe when $blob has not yet been initialized:

```php
$blob = cat($blob, $data);
```

When using cat() to concatenate an array or a list with an array member, the textual representation of the array must not exceed 8KB.

Example
```php
# Keep data around in a single blob
if ($blob == null)
    $blob = $newdata;
else
    $blob = cat($blob, $newdata);
```

### csv_close()

The csv_close() function closes a CSV file previously opened with csv_open().

Syntax
```php
csv_close(int csvp);
```

Parameters
Int csvp â€“ the file descriptor returned from csv_open.

Return Value
None

Example
```php
$fd = io_open("./myfile.csv");
$csvp = csv_open($fd);

// ... other code ...

csv_close($csvp);
```

### csv_getcols()

The csv_getcols() function returns the number of columns in a particular row in a CSV file previously opened with csv_open().

Syntax
```php
int csv_getcols(int csvp, int row);
```

Parameters
- Int csvp – the file descriptor returned from csv_open.
- Int row – which row to get columnar count

Return Value
Int – number of columns in the specified row

Example
```php
$fd = io_open("./myfile.csv");
$csvp = csv_open($fd);

// ... other code ...

$numcols = csv_getcols($csvp, 1);
echo("numcols : ", $numcols);

csv_close($fd);
```
### csv_getrows()

The `csv_getrows()` function returns the number of rows in a particular row in a CSV file previously opened with `csv_open()`.

<table>
<thead>
<tr>
<th>Syntax</th>
<th><code>int csv_getrows(int cvsp);</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td><code>Int cvsp</code> – the file descriptor returned from <code>csv_open</code>.</td>
</tr>
<tr>
<td>Return Value</td>
<td><code>Int</code> – number of rows in the specified CSV file</td>
</tr>
</tbody>
</table>
| Example         | `$fd = io_open("./myfile.csv");
$cvsp = csv_open($fd);

...$numrows = csv_getrows($cvsp);
$current_row = 0;
while( $current_row < $numrows) {
  echo(" numcols ", csv_getcols($cvsp, $current_row);
  $current_row = $current_row + 1;
}
csv_close($cvsp);
io_close($fd);` |

### csv_getval()

The `csv_getval()` function returns the value of a cell in a specified row and column in a CSV file previously opened with `csv_open()`.

<table>
<thead>
<tr>
<th>Syntax</th>
<th><code>string csv_getval(int cvsp, int row, int col);</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td><code>Int cvsp</code> – the file descriptor returned from <code>csv_open</code>.</td>
</tr>
<tr>
<td></td>
<td><code>Int row</code> – row from which to retrieve value</td>
</tr>
<tr>
<td></td>
<td><code>Int col</code> – column from which to retrieve value</td>
</tr>
<tr>
<td>Return Value</td>
<td><code>String</code> – string-encoded return value</td>
</tr>
</tbody>
</table>
| Example         | `$fd = url_fetch("http://www.mysite.com/myfile.csv");
$cvsp = csv_open($fd);

...$numrows = csv_getrows($cvsp,1);
$current_row = 0;
while( $current_row < $numrows) {
  echo("first value in row ",
  csv_getval($cvsp, $current_row,1));
  $current_row = $current_row + 1;
}
csv_close($cvsp);
io_close($fd);` |
### csv_open()

The `csv_open()` function opens a CSV file.

<table>
<thead>
<tr>
<th>Syntax</th>
<th><code>int csv_open(int fd);</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td><code>int fd</code> – the file descriptor returned from <code>io_open/url_fetch</code></td>
</tr>
<tr>
<td>Return Value</td>
<td><code>int</code> – file handle to parsed CSV file</td>
</tr>
</tbody>
</table>

**Example**

```php
$fd = url_fetch("http://www.mysite.com/myfile.csv");
$csvp = csv_open($fd);
...
$numrows = csv_getrows($csvp, 1);
$current_row = 0;
while ($current_row < $numrows) {
    echo("first value in row ", $current_row, ":", csv_getval($csvp, $current_row, 1));
    $current_row = $current_row + 1;
}
csv_close($csvp);
io_close($fd)
```

### debase64()

The `debase64()` function decodes a base64-encoded string.

The `debase64()` function has a pre-optimized fastpath, where `str` is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The `debase64()` function may be impacted by the change in the behavior of string variables. See `str` for additional information about this behavior.

<table>
<thead>
<tr>
<th>Syntax</th>
<th><code>debase64(string str);</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td><code>string str</code> - the string to decode</td>
</tr>
<tr>
<td>Return Value</td>
<td><code>string</code> - the decoded input string</td>
</tr>
</tbody>
</table>

The overall return string for the `debase64()` function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

**Example**

```php
# Decode email attachment
$email_attachment = debase64($email_attachment);
# IMAP4 authentication challenges and responses
# are also base64 encoded
$challenge = debase64($challenge);
```
dehex()

The dehex() function undoes the ASCII translation performed for HTTP requests. When sending special characters in a URL, the HTTP specification says to encode them as their hexadecimal values preceded by a percent sign (%). The dehex() function looks for such metacharacters and returns them to their original values.

The dehex() function has a pre-optimized fastpath, where str is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The dehex() function may be impacted by the change in the behavior of string variables ("str" on page 24).

Syntax

dehex(string str, [string delimiter]);

Parameters

- string str - a string that contains hex characters
- string delimiter - a single-character string used as the delimiter, with a default value of %.

Return Value

- string - a string that has had its characters changed back to the original values

The overall return string for the dehex() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

Example

dehex("abcd%41%42%43%44") == "abcdABCD"
dehex("abcd%4a%4b%4c%4d") == "abcdJKLM"
dehex("abcd%zz") = "abcd%zz"

deutf8()

The deutf8() function decodes UTF-8 encoding, first passing the input string through the dehex() function.

The deutf8() function has a pre-optimized fastpath, where str is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The deutf8() function may be impacted by the change in the behavior of string variables ("str" on page 24).

Syntax

deutf8(string str);

Parameters

- string str - the string to decode

Return Value

- string - the decoded input string

The overall return string for the debase64() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

Example

# Decode email attachment
$email_attachment = deutf8($email_attachment);
# IMAP4 authentication challenges and responses
# are also base64 encoded
$challenge = deutf8($challenge);

dns_expand()

The dns_expand() function decodes a DNS-encoded string. It operates safely on recursive names and operates correctly on heavily nested names.

DNS names may be compressed and some needed information may reside at the beginning or end of the string. Therefore, it is important to pass the entire string into dns_expand(), not just the relevant substring.
### Syntax

```c
string dns_expand(string str, int offset);
```

### Parameters

- **string str**: the string containing the encoded name
- **int offset**: the offset at which to start decoding the encoded name

### Return Value

- **string**: the lowercase decoded name, or an error string (for example, "Recursive pointer" and "Bad pointer").

### Example

```c
# Decode a SOA record with 2 contiguous strings
# (name and admin)
$name = dns_expand($dns_blob, $offset);
if (!($next = dns_next($dns_blob, $offset)))
    return -1;
$admin = dns_expand($dns_blob, $next);
if (!($next = dns_next($dns_blob, $next))) return -1;
$offset = $next;
# $offset will point to the next resource record
```

### dns_next()

The `dns_next()` function finds the next DNS resource record in a string, after a decoded name. It returns the offset into the string of the next DNS RR or a NULL error indication if it could not be fully or correctly decoded.

### Syntax

```c
int dns_next(string str, int offset);
```

### Parameters

- **string str**: the string containing multiple DNS resource records
- **int offset**: the offset at which to start decoding the current encoded name

### Return Value

- **int**: the offset into the string of the next resource record
- **NULL**: the encoded name could not be fully or correctly decoded

### Example

```c
# Decode SOA record which contains
# two contiguous strings (name and admin)
$name = dns_expand($dns_blob, $offset);
if (!($next = dns_next($dns_blob, $offset)))
    return -1;
$admin = dns_expand($dns_blob, $next);
if (!($next = dns_next($dns_blob, $next))) return -1;
$offset = $next;
# And $offset points to next resource record
```

### elem()

The `elem()` function extracts a single element from a list. If you do not specify an index, the function returns the first element in the list.
## Syntax

\[
\text{elem}(\text{list } li, [\text{int index := 0}]);
\]

### Parameters

- list \( li \) - the list from which to extract the element
- int index - the index to the element you want to extract. The index value starts at zero (that is, an index value of 0 extracts the first element).

### Return Value

- any - an element from the list.

Full run-time type checking is performed on the return value and any variables that depend on it.

### Example

```cpp
# Extract the third element in the list
portlist = [21, 23, 53, 80];
third = \text{elem}(\text{portlist}, 2);
third == 53;
```

## errno()

The errno() function returns the Boolean value associated with an error.

### Syntax

\[
\text{errno}(\text{any argument});
\]

### Parameters

- any argument - the expression you want evaluated for error status

### Return Value

int - a Boolean value corresponding to the error status of the argument, where:

- 1 - the argument evaluated to an error
- 0 - the argument did not evaluate to an error

The errno() function has changed significantly beginning with CPcode v3. It now returns a Boolean integer instead of a numeric value. Unlike past CPcode versions, the CPcode execution engine will stop if a catastrophic error is encountered.

### Example

```cpp
\text{Print}("\text{Error status }", \text{errno}($\text{foo} + $\text{bar}$));
```

## ethmac()

The ethmac() function constructs an Ethernet MAC address from the supplied arguments.

### Syntax

\[
\text{ethmac}\left(\{\text{string, int}\} a, [\text{int b, int c, int d, int e, int f}]\right);
\]

### Parameters

- \{string, int\} \( a \) - either the first octet or a string with the textual representation of the MAC address
- int \( b \) - the second octet
- int \( c \) - the third octet
- int \( d \) - the fourth octet
- int \( e \) - the fifth octet
- int \( f \) - the sixth octet

### Return Value

ethmac - the MAC address
filename()

The filename() function will deobfuscate ("cleanse") a filename. It contracts .../ and . directory hiding and removes superfluous //.

The filename() function has a pre-optimized fastpath, where str is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The filename() function may obfuscate the actual occurrences when symbolic links are present. For example:

```plaintext
mkdir /tmp/link
cd /tmp/link
ln -s /usr foo
filename("/tmp/link/foo/..") == "/tmp/link"

but
/tmp/link/foo/..
points to /.
```

The filename() function may be impacted by the change in the behavior of string variables. Refer to the description of the string data type on page B-5 for additional information about this behavior.

Depending on the protocol being decoded, it may be necessary to run the input filename through the deutf8() or dehex() functions before passing it through filename().

The overall return string for the filename() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

**Syntax**

```plaintext
filename(string str, [int BackSlashFlag := 0]);
```

**Parameters**

- string str - the original filename
- int BackSlashFlag - Boolean used to treat the filename as a Microsoft Windows filename

**Return Value**

string - a cleansed filename

**Example**

```plaintext
filename("/usr/..//etc//passwd") == "/etc/passwd"
filename("/etc/././foo") == "/etc/foo"
filename("/etc/.../foo") == "/etc/.../foo"
```

formattedtime()

The formattedtime() function returns the time of day as a formatted string

The formattedtime() function may be impacted by the change in the behavior of string variables. Refer to the description of the string data type on page B-5 for additional information about this behavior.

The formattedtime() function is extremely slow and should be used judiciously.

The overall return string for the formattedtime() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.
### Syntax

```plaintext```
formattedtime(int time, [string format]);
```

### Parameters

- **int time** - the time to be formatted
- **string format** - the format to use for displaying the time. Refer to the strftime in UNIX man pages for formatting options. If a format is not specified, the following default format is used:
  ```plaintext```
  ```
  [%m/%d/%y %H:%M:%S]
  ```

### Return Value

**string** - a string containing the time in the format specified

### Example

```plaintext```

time = formattedtime(system.time);
timeonly = formattedtime(system.time, "%T %Z");
dateonly = formattedtime(system.time, "%A, %d %B %Y");

```
echo("The current system time is now ", time, 
"\n");
```

echo("The time is now ", timeonly, "\n");

echo("The date is now ", dateonly, "\n");

# Current system time = [08/06/98 16:49:24]
# Time is now 16:49:24 EDT
# Date is now Thursday, 06 August 1998
```

### hatoi()

The hatoi() function converts a two's complement hex string to an integer. If you specify an index, the conversion begins at that location inside the string.

If the leading bit is set to (1<<31), then the return value will be a negative number.

The hatoi() function ignores any white space leading up to the number. However, unlike the atoi() function, hatoi() will return a NULL if there are any characters after the hex number or if the hex value overflows the size of the CPcode integer.

### Syntax

```plaintext```
hatoi(string str, [int index := 0]);
```

### Parameters

- **string str** - the hex string that is to be converted to an integer
- **int index** - the point in the string at which to begin the conversion

### Return Value

- **int** - the integer value of the hex string, str
- **NULL** - the string is empty, contains characters that are not valid hex digits, or is too large to fit inside an CPcode integer.

### Example

```plaintext```

```{javascript}
{ 
hatoi("10") == 16
hatoi("ff") == 255
hatoi("ffXX") == NULL
hatoi(" 80000000") == -2147483648
hatoi(" ffffffffff") == NULL
hatoi("monkeys: 41", 8) == 65
hatoi(""") == NULL
hatoi(42) == NULL
}
```

host()

The host() function constructs an IPv4 host address from other parameters (for example, information in a data stream).

For example, the PORT command in FTP sends the octets of an IP address as a comma-separated, printable list. The host() function could be used to create an IP address from the list. The long() function could be used to gather information from a packet, using the single-integer syntax of the host() function to create an IP address. This can be easier than extracting each octet individually.

**Syntax**

```plaintext
host({ip, string, list, int} a, [int b, int c, int d]);
```

**Parameters**

- `{ip, string, list, int}` - a 32-bit, host-endian integer containing:
  - a list of four integers representing each octet in the address
  - the full IPv4 address
  - a string representation of an IP address
  - the first of four integers representing each IP address octet
  - int b - in the case of four octets, the second octet
  - int c - in the case of four octets, the third octet
  - int d - in the case of four octets, the fourth octet

**Return Value**

`ip` - an IPv4 address in network byte order

**Example**

```plaintext
$foo = host(208, 217, 179, 1);
$foo = host(0xD0D9b301);
$foo = host([208, 217, 179, 1]);
$foo = host("208.217.179.1");
$foo = host(208.217.179.1);
```

host6()

The host6() function constructs an IPv6 host address from the supplied parameter.

**Syntax**

```plaintext
host6(string address, [int binary]);
```

**Parameters**

- `string address` - an textual or binary IPv6 address
- `int binary` - tells host6() whether to parse the address parameter as a binary 16-byte address, where:
  - 0 - indicates that the address parameter is textual
  - 1 - indicates that the address parameter is binary

**Return Value**

`ip` - an IPv6 address in network byte order

**Example**

```plaintext
$foo = host6("::1");
$foo = host6("abcd:0123::5432:1000", 0);
```

incksum()

The incksum() function computes a 16-bit checksum, summing two characters at a time, which corresponds to the IP header checksum.

incksum() can be used to either verify or compute checksums.
### incksum()

The `incksum()` function searches for the first occurrence of the little string in the big string.

The index count starts with zero.

Using a list of little strings may not always be an ideal solution. If the list is static, it may be significantly faster to use a regular expression before calling the `index()` function.

The `prefix()` function is a faster case of the `index()` function.

**Syntax**

```
incksum({string, list} in);
```

**Parameters**

- `{string, list} in` - the string or list of strings for which the checksum should be computed

**Return Value**

- `int` - an integer checksum

**Example**

```
incksum("foo bar") == 61865
incksum(["foo", "bar"]) == 53673
incksum(["fo", "o " , "ba", r"])) == 61865
```

### index()

The `index()` function searches for the first occurrence of the little string in the big string.

The index count starts with zero.

Using a list of little strings may not always be an ideal solution. If the list is static, it may be significantly faster to use a regular expression before calling the `index()` function.

The `prefix()` function is a faster case of the `index()` function.

**Syntax**

```
index(string big, {list, string} little);
```

**Parameters**

- `string big` - the string to be searched
- `{list, string} little` - the string, or list of strings, to be located

**Return Value**

- `int` - the index of the little string into the big string
- `-1` - an occurrence of little string could not be found in big string

**Example**

```
index("Network Security", "work") == 3
index("123", "1") == 0
```

### io_close()

The `io_close()` function closes a previously opened file descriptor.

`io_*` functions do not integrate with the caching implementation provided with the `url_*` functions. If caching is desired or network access is desired, use the `url_*` functions instead.

**Syntax**

```
io_close (int fd);
```

**Parameters**

- `int fd` – file descriptor to be closed

**Return Value**

- None

**Example**

```
io_close($fd);
```

### io_delete()

The `io_delete()` function deletes a file from the file system.

`io_*` functions do not integrate with the caching implementation provided with the `url_*` functions. If caching is desired or network access is desired, use the `url_*` functions instead.
### io_delete

The `io_delete` function is used to delete a file. It takes a single parameter, `filename`, which is the path of the file to be deleted.

**Syntax**

```plaintext
io_delete (string filename);
```

**Parameters**

- `filename`: The path of the file to be deleted.

**Return Value**

- 0 on success
- Error code otherwise

**Example**

```php
$filename = "./test.txt";
io_delete($filename);
```

### io_open

The `io_open` function returns an I/O descriptor for reading or writing. The file will be automatically closed when the calling CPCode filter exits.

**Io_* functions do not integrate with the caching implementation provided with the url_* functions. If caching is desired or network access is desired, use the url_* functions instead.**

**Syntax**

```plaintext
int io_open (string filename);
```

**Parameters**

- `filename`: The path of the file to be opened.

**Return Value**

- `int` returns a valid file descriptor, NULL if the file does not exist.

**Example**

```php
$fd = io_open("./test.txt");
```

### io_read

The `io_read` function reads data from an I/O descriptor. **Io_* functions do not integrate with the caching implementation provided with the url_* functions. If caching is desired or network access is desired, use the url_* functions instead.**

**Syntax**

```plaintext
list io_read (int fd);
```

**Parameters**

- `int fd`: The file descriptor to be read.

**Return Value**

- `list` list of strings representing the contents of the file.
- NULL if file descriptor is invalid.

**Example**

```php
$fd = io_open("./test.txt");
...
$contents = io_read($fd);
io_close($fd);
```

### io_readlines

The `io_readlines` function reads data from an I/O descriptor. **Io_* functions do not integrate with the caching implementation provided with the url_* functions. If caching is desired or network access is desired, use the url_* functions instead.**
### list io_readlines(int fd, <int chomp>);

**Parameters**
- `int fd` – file descriptor to be read
- `int chomp` – optional flag indicating whether or not to remove newline characters from the lines returned. Will remove "\n" as well as "\r\n."

**Return Value**
- list – list of strings representing the contents of the file.
- NULL if file descriptor is invalid.

**Example**
```
$fd = io_open("./test.txt");
...
$contents = io_readlines($fd);
io_close($fd);
```

---

### io_write()

The `io_write()` function writes data to an I/O descriptor. Writes do not append lines to a file, rather a single `io()` call will replace the file in question.

I0_* functions do not integrate with the caching implementation provided with the url_* functions. If caching is desired or network access is desired, use the url_* functions instead.

**Syntax**
```
list io_write(int fd, list data);
```

**Parameters**
- `int fd` – file descriptor to be written to.
- `list data` – A list of strings to write into a file.

**Return Value**
None

**Example**
```
$fd = io_open(cat(scope(),"-known-ports"));
...
io_write($fd, ["80","8080","9090"]);
io_close($fd);
```

---

### kiss_compile()

A different name for the `kiss_pattern()` function. See `kiss_pattern()` (on page 68).

---

### kiss_pattern()

The `kiss_pattern()` function creates a list of patterns into a single multi-pattern search. It behaves and is used in a manner similar to the multi_pattern() function.

**Syntax**
```
String kiss_pattern(list pattern_list);
```

**Parameters**
- `list pattern_list` – this is a list of lists. Each list pair contains a pattern and a return token. Function pointers are not allowed for return tokens but any other data types are.

**Return Value**
string – the compiled pattern suitable for using with `kiss_search()`;

**Example**
```
MY_PATTERN = kiss_pattern([ ["x0",1],["x0x0",2],
["y0y0",3], ["y0y0dyne",4] ]) ;
```
kiss_search()

The kiss_search() function compares a kiss_pattern expression to a given string. It behaves and is used in a manner similar to multi_search();

<table>
<thead>
<tr>
<th>Syntax</th>
<th>list kiss_search(string kiss_pattern, string str);</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>string kiss_pattern – compiled regular expression returned from kiss_pattern()</td>
</tr>
<tr>
<td>Return Value</td>
<td>list – a list of tokens that are returned</td>
</tr>
</tbody>
</table>
| Example      | MY_PATTERN = kiss_compile([ ["x0",1], ["x0x0",2], ["y0y0",3], ["y0y0dyne",4] ])

func my_search {
    if ($matches = kiss_search(MY_PATTERN,$1)) {
        foreach $elm inside ($matches) {
            if ($elm == 1) echo ("x0");
            else if ($elm == 2) echo ("x0x0");
            else if ($elm == 3) echo ("y0y0");
            else if ($elm == 4)
                echo ("Is your name John Smith?");
        }
    }
}

le_long()

The le_long() function returns a four-byte signed integer from the specified offset within a string. It extracts the long integer in host order without converting it.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>le_long(string str, int offset);</th>
</tr>
</thead>
</table>
| Parameters   | • string str - the string to be examined, beginning at byte 0
          | • int offset - the offset in the string for the first byte to be returned |
| Return Value | • int - the offset integer, as a signed value
          | • error - the offset was out of bounds |
| Example      | le_long("^A^@^@^@", 0) == 0x00000001;
          | le_long("^@^A^@^@", 0) == 0x00000100;
          | le_long("^@^@^A^@", 0) == 0x00010000;

le_short()

The le_short() function returns a two-byte signed integer from the specified offset within a string. It extracts the short integer in host order without converting it.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>le_short(string str, int offset);</th>
</tr>
</thead>
</table>
| Parameters   | • string str - the string to be examined, beginning at byte 0
          | • int offset - the offset in the string for the first byte to be returned |
### Return Value
- int - the offset integer, as a signed value
- error - the offset was out of bounds

### Example
```c
le_short("^A^@", 0) == 0x00000001;
le_short("^@^A", 0) == 0x00000100;
```

### le_ulong()

The `le_ulong()` function returns a four-byte unsigned integer from the specified. It extracts the long integer in host order without converting it.

The `le_ulong()` function returns the same result as the `le_long()` function. It is provided for backwards compatibility. The internal representation of all CPcode integers is signed. Therefore, the `le_ulong()` function yields a signed integer.

#### Syntax
```c
le_ulong(string str, int offset);
```

#### Parameters
- string str - the string to be examined, beginning at byte 0
- int offset - the offset in the string for the first byte to be returned

#### Return Value
- int - the offset integer, as an unsigned value
- error - the offset was out of bounds

#### Example
```c
le_ulong("^A^@^@^@", 0) == 0x00000001;
le_ulong("^@^A^@^@", 0) == 0x00000100;
le_ulong("^@^@^A^@", 0) == 0x00010000;
le_ulong("\x00\xff\xff\xff", 0) == -256
```

### le_ushort()

The `le_short()` function returns a two-byte unsigned integer from a specified offset within a string. It extracts the short integer in host order without converting it.

#### Syntax
```c
le_ushort(string str, int offset);
```

#### Parameters
- string str - the string to be examined, beginning at byte 0
- int offset - the offset in the string for the first byte to be returned

#### Return Value
- int - the offset integer, as an unsigned value
- error - the offset was out of bounds

#### Example
```c
le_ushort("^A^@", 0) == 0x00000001;
le_ushort("^@^A", 0) == 0x00000100;
```

### listadd()

The `listadd()` function constructs a list at runtime, or adds items to an existing list. NULL items are added to the list.

The `listadd()` function has a pre-optimized fastpath, where `li, ...` is its most relevant parameter. It operates more efficiently if its input list is the same CPcode variable as its output list.

However, duplicating the list may be an expensive operation, as in the example:
```c
$listB = listadd($listA, 1, 2, 2);
```
### Syntax

`listadd(list li, any item, [...]);`

### Parameters
- `list li` - the existing list. If this value is NULL, the function creates an empty list and adds to that list.
- `any item` - the first item to be added to the tail of the list
- `...` - a series of items to be added to the tail of the list

### Return Value
- The original list, or the newly constructed list, with the items added to it.

### Example

```plaintext
code
badIPs = [ 208.217.179.1, 208.217.179.3 ];
badIPs = listadd(badIPs, ip.src, ip.dst);
A = [1, 2, 3];
B = [4, 5, 6];
C = listadd(A, B); # C == [1, 2, 3, [4, 5, 6]]
```

### listcombine()

The `listcombine()` function combines two lists. It appends each element from the second list to the first list.

The `listcombine()` function differs from the `listadd()` function, which adds the entire second list to the first list instead of adding each individual element of the second list.

The `listcombine()` function has a pre-optimized fastpath, where `list` is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

### Syntax

`list listcombine(list a, list b);`

### Parameters
- `list a` - the first list to process
- `list b` - the second list to process

### Return Value
- `list` - a list containing all the elements of both lists
- The original list, or the newly constructed list, with the items added to it.

### Example

```plaintext
code
a = [1, 2, 3];
b = [4, 5, 6];
c = listcombine(a, b);
c == [1, 2, 3, 4, 5, 6]
```

### listglom()

See `listcombine()` (on page 71)

### listlen()

The `listlen()` function returns a count of the number of elements in a list.

### Syntax

`listlen(list li);`

### Parameters
- `list li` - the list whose elements are to be counted
Return Value

- int - the number of elements in the list
- error - the parameter is not a list

The `listlen()` function does not count the elements of sublists as separate elements.

The `listlen()` function is faster than the description implies. Lists keep an internal count of their members, which this function returns.

Example

```c
portlist = [21, 23, 80];
listlen(portlist) == 3;
superlist = [1, 2, 3, [4, 5, 6]];
listlen(superlist) == 4;
```

### long()

The `long()` function returns a four-byte signed integer from a specified offset within a string. It converts the value from network byte order in the string to the host's byte order. Network byte order is big endian, therefore the most significant byte is stored at the lowest address in memory.

The `long()` function is useful for extracting values from packets.

**Syntax**

```
long(string str, int offset);
```

**Parameters**

- string str - the string to be examined, beginning at byte 0
- int offset - the offset in the string for the first byte to be returned

**Return Value**

- int - the offset integer, as a signed value
- error - the offset was out of bounds

**Example**

```c
# the structure starts with a 4-byte
# integer in network byte order.
# It will be converted to host byte
# order before placing it in $data
$data = long($struct_data, 0);
```

### lsort()

The `lsort()` function sorts the members of a list according to type, grouping elements of the same type together. Within each data type, the `lsort()` function sorts the elements in ascending order.

The order of the elements within the group is not guaranteed.

The `lsort()` function has a pre-optimized fastpath, where `li` is its most relevant parameter. It operates more efficiently if its input list is the same CPcode variable as its output list.

**Syntax**

```
lsort(list li);
```

**Parameters**

- list li - the list to be sorted

**Return Value**

- list - the sorted list
### luniq()

The luniq() function creates a list of the unique elements from another list (that is, removes duplicates). It compares adjacent elements of a list. If they are equal, the luniq() function adds only one of the elements to the new list. If they are not equal, the luniq() function adds both elements to the list.

The luniq() function determines if elements are equal using the criteria for the == operator.

To guarantee that the new list contains unique elements, ensure that the list passed to the luniq() function is sorted first. It is much faster to assign the sorted result to the list being sorted than it is to use a temporary list, either explicitly or implicitly.

<table>
<thead>
<tr>
<th>luniq()</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntax</strong></td>
</tr>
<tr>
<td>luniq(list li);</td>
</tr>
<tr>
<td><strong>Parameters</strong></td>
</tr>
<tr>
<td>list li - the list from which duplicate elements will be removed</td>
</tr>
<tr>
<td><strong>Return Value</strong></td>
</tr>
<tr>
<td>list - a new list with duplicate adjacent elements removed</td>
</tr>
<tr>
<td>error - the parameter is not a list</td>
</tr>
<tr>
<td><strong>Example</strong></td>
</tr>
<tr>
<td>foo = [ 8, 4, 2, 9 10, 6, 4 ];</td>
</tr>
<tr>
<td>echo(lsort(foo)); # Slow</td>
</tr>
<tr>
<td># [2, 4, 6, 8, 9, 10]</td>
</tr>
<tr>
<td>foo = [ 8, 4, 2, 9 10, 6, 4 ];</td>
</tr>
<tr>
<td>foo = lsort(foo); # Fast</td>
</tr>
<tr>
<td>echo(foo);</td>
</tr>
<tr>
<td># [2, 4, 6, 8, 9, 10]</td>
</tr>
</tbody>
</table>

### match()

See regexec() (on page 79).

### multi_compile()

See multi_pattern() (on page 74).

### multi_escape()

The multi_escape() function escapes any reserved special characters in a string.

If used in conjunction with the multi_pattern() function, the multi_pattern() function can then accept the string literally.
multi_escape(string str);

**Parameters**
- string str - the string to escape

**Return Value**
- string - the escaped string

**Example**
```c
{ $foo = multi_escape("foo.exe");
  # $foo == "foo\exe"
  $mp = multi_pattern([$foo, 1]);
}
```

### multi_pattern()

The `multi_pattern()` function creates a complex regular expression from a list of multiple patterns. Unlike normal regular expressions, a multi-pattern search can perform the match function in a single iteration through the search string. A typical multi-pattern search is sub 0(n).

The allowable regular expression characters are:
- +
- *
- ?
- ^
- $
- .
- [

Multi-pattern search compilation is resource intensive and should be done only once and then assigned to a global variable.

The `multi_pattern()` function should never be called from a filter.

**Syntax**
```c
multi_pattern(list pattern_list, [int case_insensitive]);
```

**Parameters**
- list pattern_list - the list of patterns and their return tokens
- int case_insensitive - tells `multi_pattern()` whether the pattern is case insensitive, where:
  - 0 - indicates that the pattern is case sensitive
  - a non-zero integer - indicates that the pattern is case insensitive

**Return Value**
- string - a complex regular expression
### Example

<table>
<thead>
<tr>
<th>Example</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>example = multi_pattern( [ [&quot;one&quot;, 1], [&quot;two&quot;, 2], [&quot;three&quot;, 3] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one three&quot;) == [1,3]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;two&quot;) == [2]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;three&quot;) == [3]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;THREE&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;[oO][nN][eE]&quot;, 1] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;ONE&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;oNe&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx oNe xxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;^[oO][nN][eE]&quot;&quot;, 1] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;ONExxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;oNexxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx oNe xxx&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;\x90\x90\x90&quot;, &quot;NOP string&quot;], [&quot;root&quot;, &quot;root login&quot;] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx \x90\x90\x90 \x90 xxx&quot;) == &quot;NOP string&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx root xxx&quot;) == &quot;root login&quot;</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ &quot;foo&quot;, &quot;bar&quot;, &quot;zah&quot; ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx foo xxx&quot;) == &quot;foo&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx zah xxx&quot;) == &quot;zah&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx dog xxx&quot;) == NULL</td>
<td></td>
</tr>
</tbody>
</table>

### multi_search()

The `multi_search()` function compares a multi-pattern expression to a string and returns the tokens associated with all matches.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>multi_search(string multi_pattern, string str);</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>example = multi_pattern( [ [&quot;one&quot;, 1], [&quot;two&quot;, 2], [&quot;three&quot;, 3] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one three&quot;) == [1,3]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;two&quot;) == [2]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;three&quot;) == [3]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;THREE&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;[oO][nN][eE]&quot;, 1] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;ONE&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;oNe&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx oNe xxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;^[oO][nN][eE]&quot;&quot;, 1] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;one&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;ONExxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;oNexxx&quot;) == [1]</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx oNe xxx&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;foo&quot;) == NULL</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ [&quot;\x90\x90\x90&quot;, &quot;NOP string&quot;], [&quot;root&quot;, &quot;root login&quot;] ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx \x90\x90\x90 \x90 xxx&quot;) == &quot;NOP string&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx root xxx&quot;) == &quot;root login&quot;</td>
<td></td>
</tr>
<tr>
<td>example = multi_pattern( [ &quot;foo&quot;, &quot;bar&quot;, &quot;zah&quot; ] );</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx foo xxx&quot;) == &quot;foo&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx zah xxx&quot;) == &quot;zah&quot;</td>
<td></td>
</tr>
<tr>
<td>multi_search(example, &quot;xxx dog xxx&quot;) == NULL</td>
<td></td>
</tr>
</tbody>
</table>
### multi_pattern()

The `multi_pattern()` function checks a string (`str`) against a multi-pattern expression (`multi_pattern`) and returns matched tokens or `NULL` if no match is found.

#### Parameters
- `string multi_pattern` - the multi-pattern expression to be compared
- `string str` - the string to be compared

#### Return Value
- `list` - the matched patterns' tokens
- `NULL` - no string match was found within the multi-pattern expression

#### Example
See the `multi_pattern()` function Examples (see "multi_pattern()" on page 74).

### net()

The `net()` function uses the supplied arguments to construct an IPv4 address from several different data types.

#### Syntax
```
net({string, list, ip, int} a, {{ip, int} b, int c, int d, int e, int f, int g, int h});
```

#### Parameters
- `{string, list, ip, int} a` - a 32-bit host-endian integer containing:
  - a list of eight integers representing each octet in the address/mask
  - a full IPv4 address
  - a string representation of an IP address and netmask
  - the first of four integers representing each IP address octet
- `{ip, int} b` - one of the following:
  - in the case of eight octets, the second octet
  - in the case of two 32-bit integers, the netmask
  - in the case of two network variables, the netmask
- `int c` - the third octet of the network address
- `int d` - the fourth octet of the network address
- `int e` - the first octet of a netmask, or the netmask in CIDR notation
- `int f` - the second octet of a netmask
- `int g` - the third octet of a netmask
- `int h` - the fourth octet of a netmask

#### Return Value
- `ip` - an IPv4 address/network in network byte order
net6()

The net6() function constructs an IPv6 network from the supplied parameter.

**Syntax**

```
net6(string network);
```

**Parameters**

- string network - a textual IPv6 network

**Return Value**

ip - an IPv6 network in network byte order

**Example**

```
$foo = net6("::1/128");
$foo = net6("abcd:0123::/32");
```

pattern()

See regcomp() (on page 78)

pred()

The pred() function finds and returns the logical predecessor of the indexed item in an array.

The pred() function allows reverse foreach-like behavior, but at a larger performance cost.

**Syntax**

```
pred(array arr, any index);
```

**Parameters**

- array arr - the array from which to extract the predecessor
- any index - the index to the item whose predecessor will be extracted

**Return Value**

any - the predecessor of the indexed item in the array (that is, the item that will be extracted)
prefix()

The `prefix()` function compares two strings and returns true if the second string is a prefix of the first string. The `prefix()` function is a faster case of the `index()` function. Use the `casprefix()` function for a case-insensitive comparison.

**Syntax**

```
prefix(string big, string little);
```

**Parameters**

- `string big` - the string against which `little` will be compared
- `string little` - the smaller string, which will be compared to the beginning of `big`

**Return Value**

`int`, where:

- `1` - `little` is the same as the beginning of `big`
- `0` - `little` is not the same as the beginning of `big`

**Example**

```
prefix("ABCDEFG", "ABC") == 1
prefix("ABCDEFG", "aBc") == 0
prefix("ABCDEFG", "DEF") == 0
```

prunearray()

The `prunearray()` function randomly removes a specified percentage of elements in an array. If CPcode uses a global state table in an array, the system stability can be put at risk if the array grows arbitrarily large and uses all of the system's memory. The `prunearray()` function should be used to randomly (and fairly) free part of those global arrays.

**Syntax**

```
prunearray(array ar, int percent);
```

**Parameters**

- `array ar` - the array to prune
- `int percent` - the percentage of the array to prune

**Return Value**

`array` - the pruned array

**Example**

```
STATE_TABLE = prunearray(STATE_TABLE, system.pressure)
```

regcomp()

The `regcomp()` function creates and compiles a regular expression from a string. The compiled regular expression can be compared against other strings using the `regex()` function.
The `regcomp()` function is not a fast operation. Therefore, it should be used only on global variables, and only once for each variable.

To mitigate the performance impact of regular expression compilation from a filter, there is a 256-entry, fully associative `regcomp` cache. However, it is still better to execute a single `regcomp()` on a global variable.

The `regcomp()` function recognizes those regular expressions that are recognized by `grep` or the `regcomp()` function on UNIX systems.

The `regcomp()` function does not work on strings that contain Nulls.

### Syntax
```
regcomp(string pattern);
```

### Parameters
- `string pattern` - the string to be converted into a regular expression

### Return Value
- `string` - a regular expression pattern

### Example
```
example = regcomp("abc+");
"abc" == example              # TRUE
"abcccc" == example           # TRUE
"defccc" == example           # FALSE
"abcdef" == example           # TRUE
"defabc" == example           # TRUE
"defabcdef" == example        # TRUE
example = regcomp("^abc");
"abc" == example              # TRUE
"abcccc" == example           # TRUE
"defccc" == example           # FALSE
"abcdef" == example           # TRUE
"defabc" == example           # FALSE
"defabcdef" == example        # FALSE
example = regcomp("^abc|^def");
"abc" == example              # TRUE
"abcccc" == example           # TRUE
"defccc" == example           # TRUE
"abcdef" == example           # TRUE
"defabc" == example           # TRUE
"defabcdef" == example        # TRUE
"Xdefabcdef" == example       # FALSE
```

### regexec()

The `regexec()` function compares a regular expression to a string, and returns an integer result indicating whether the patterns matched. This function utilizes regular expressions compiled by the `regcomp()` function.

### Syntax
```
regexec(string pattern, string str);
```

### Parameters
- `string pattern` - the regular expression to be compared
- `string str` - the string to be compared
Return Value

<table>
<thead>
<tr>
<th>Return Value</th>
<th>int, where:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 0 - patterns did not match</td>
</tr>
<tr>
<td></td>
<td>• 1 - patterns matched</td>
</tr>
</tbody>
</table>

Example

```c
example = regcomp("abc+" );
regexec( example, "defabc") == 1
example = "^abc" ;
regexec( example, "abcde") == 1
regexec( example, "defabc") == 0
```

**scope()**

The `scope()` function returns the name of the scope in which the CPcode is currently executing.

Calls to `scope()` are evaluated at compile time and are the functional equivalent of using a string literal with a scope name. For example:

```c
$my_scope = scope();
```

translates to:

```c
$my_scope = "ftp_commands" ;
```

**Syntax**

```c
scope();
```

**Parameters**

None

**Return Value**

- string - a string showing the name of the scope in which the CPcode is currently executing. The string displays as:
  - the protocol, an underscore, and the Protection Group name if this CPcode is for a Protection Group
  - the protocol if this CPcode is protocol-wide CPcode

**Example**

```c
func foo
{
  echo("debug:: foo called from", $1);
    ...
}
...
foo(scope());
```

**sed()**

The `sed()` function rewrites strings. It corresponds to the UNIX sed command (Stream EDitor), which simplifies later processing by replacing or removing items from a string.

The `sed()` function must compile the pattern into a regular expression, which is extremely slow. To alleviate the compilation bottleneck, the `sed()` function uses a 64-entry, fully associative regex cache. Using the same pattern is normally much faster than using many patterns. To this end, the cache is not LRU (least recently used); its replacement scheme is to remove the closest match to the new entry on the assumption that they were both dynamically generated from the same Protection Group. Dynamic patterns are discouraged and are penalized by low cache precedence.

**Syntax**

```c
sed(string pattern, string replace, string input,
    [int matchOnce :=0];
```
**Parameters**
- string pattern - the string pattern to be matched and replaced
- string replace - the replacement string
- string input - the input string
- int matchOnce - an optional Boolean where:
  - 0 - replace all occurrences of the pattern
  - 1 - replace the first occurrence of the pattern

**Return Value**
- string - the rewritten string

**Example**
```
sed("an efficient", "A REALLY INEFFICIENT", "sed is an efficient built-in function")
# Returns
"sed is A REALLY INEFFICIENT built-in function"
# MatchOnce = 0
sed("moo", "ribbet", "The frog says moo. The cow says moo.", 0)
# Returns
"The frog says ribbet. The cow says ribbet."
# MatchOnce = 1
sed("moo", "ribbet", "The frog says moo. The cow says moo.", 1)
# Returns
"The frog says ribbet. The cow says moo."
```

### short()

The `short()` function returns a two-byte signed integer from a specified offset within a string. It converts the value from network byte order in the string to the host’s byte order. Network byte order is big endian; therefore, the most significant byte is stored at the lowest address in memory.

The `short()` function is useful for extracting values from packets.

**Syntax**
```
short(string str, int offset);
```

**Parameters**
- string str - the string to be examined, beginning at byte 0
- int offset - the offset in the string for the first byte to be returned

**Return Value**
- int - the offset integer, as a signed value
- error - the offset was out of bounds

**Example**
```
# Structure starts with 2-byte int
# in network byte order.
# It will be converted to host byte order
# before placing it in $data
$data = short($struct_data,0);
```
split()

The split() function breaks apart and returns a list of strings derived from another string. Characters before
the specified delimiter become the first element in the list, and characters after the delimiter become the
second element in the list, and so forth.

It is permissible to split on the NULL character or other binary characters by entering hex characters (for
example, \x00xff) in the string.

The split() function will ignore any leading delimiters in the string.

If no delimiter is specified, white space characters (a space, \t, \r, and \n) will become the default.

Syntax

\begin{verbatim}
split(string str,[string delimiter]);
\end{verbatim}

Parameters

- string str - the string to be split
- string delimiter - the parameter on which the string should be split

Return Value
list - a list of strings

Example

\begin{verbatim}
split("www.company.com", ".") == ["www", "company", "com"]
split("Foo and his pet rock") == ["Foo", "and", "his", "pet", "rock"]
split("One,Two Three;four", ",;") == ["One", "Two", "Three", "four"]
split("    ONE   TWO   ") == ["ONE", "TWO"]
\end{verbatim}

str()

The str() function converts each parameter to its textual representation and returns a list of those strings.

The str() function may be impacted by the change in the behavior of string variables. Refer to the description
of the string data type on page B-5 for additional information about this behavior.

Syntax

\begin{verbatim}
str(...);
\end{verbatim}

Parameters

- ... - parameters to convert to textual representation and put in a list

Return Value
list - a list of strings

Example

\begin{verbatim}
str(1, 2, "three", 4.5.6.7) == ["1", "2", "three", "4.5.6.7"]
\end{verbatim}

strcasecmp()

The strcasecmp() function compares two strings, character by character, ignoring case. Two characters are
considered equal if they have the same byte value, or if they are alphabetically the same character without
regard to case (for example, D is equal to d).

The strcasecmp() function is useful for comparing case-insensitive strings. Many protocols specify that
strings must be case insensitive.

The strcasecmp() function has a legacy optimization to first compare string lengths. If one string is shorter
than the other is, the character comparison is not done and <0 is returned.

The first occurrence of an unequal character causes the function to stop the comparison. If the characters
are equal, comparison continues until the end of the string is reached.
**strerror()**

The `strerror()` function returns a Boolean string indicating whether an error occurred.

**Syntax**
```
strerror(any arg);
```

**Parameters**
- any arg - the expression being evaluated

**Return Value**
- string - a string indicating whether the argument evaluated as "no error" or "unknown error"

**Example**
```
$one = 1;
$zero = 0;
strerror($one + $zero) == "no error"
strerror($one / $zero) == "unknown error"
$div = $one / $zero;
strerror($div) == "unknown error"
```

**strlen()**

The `strlen()` function returns the length, in bytes, of a string or blob.

Unlike other programming languages, CPcode does not use a trailing NULL character.

**Syntax**
```
strlen(string str);
```

**Parameters**
- string str - the string whose length is to be determined

**Return Value**
- int - the length of str

**Example**
```
strlen("123") == 3
```

**strstr()**

See `index()` (on page 66).
**strtype()**

The `strtype()` function returns an integer representing the type of data that is in a string.

The `strtype()` function is useful for performing one type of processing on data that contains only ASCII characters and another type of processing on binary data.

**Syntax**

```c
strtype(string str);
```

**Parameters**

- `string str` - the string whose data is to be evaluated

**Return Value**

`int`, where:

- 1 - NULL
- 2 - the string contains only digits
- 3 - the string contains only alphabetic characters, including ISO 8859-1 Latin 1 characters 192 through 255, except 215 and 247
- 4 - the string contains only spaces
- 5 - the string contains only white space, including ISO 8859-1 Latin-1 characters 10, 11, 12, 13, 32, and 160
- 6 - the string contains only alphanumeric characters
- 7 - the string contains only printable characters, including ISO 8859-1 Latin-1 characters 32 through 126 and 160 through 255
- 8 - the string contains only ASCII characters between 0 and 127
- 9 - the string contains only binary data
### Example

```cpp
func foo
{
    type = strtype($1);
    if (type == 9)
        echo("The data has a string type of ", type, "\n");
    else
        echo("The data ", $1, " has a string type of ", type, "\n");
}
foo("21167");
foo("Company");
foo("Company123");
foo(""");
foo("Generic Company, Inc.");
foo("www.company.com\x00");
foo("\xff \x01");
The output is:
The data 21167 has a string type of 2
The data Company has a string type of 3
The data Company123 has a string type of 6
The data has a string type of 1
The data Generic Company, Inc. has a string type of 7
The data www.company.com has a string type of 9
The data has a string type of 9
```

### sublist()

The `sublist()` function returns a list that is a subset of another list. It creates a new list, starting with the element specified as the start, and continuing for the number of elements specified for the length.

When using the `sublist()` function, remember that the first character in the string is always at zero.

If a length is not included, the new list starts at the offset and uses the remaining elements in the old list. If the starting offset is past the end of the list, the new list is empty. If the length reaches past the end of the list, the new list uses the remaining elements in the list.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>sublist(list li, int start, [int length]);</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>list li - the list to be searched</td>
</tr>
<tr>
<td></td>
<td>int start - the offset in the list for the beginning of the new list</td>
</tr>
<tr>
<td></td>
<td>int length - the number of elements of the first list to use in the second list</td>
</tr>
<tr>
<td>Return Value</td>
<td>list - the sublist</td>
</tr>
</tbody>
</table>
### Example

```plaintext
ourIP = listadd(our_IPs, 208.217.179.1, 208.217.179.2, 208.217.179.3);
ourIP = listadd(our_IPs, 208.217.179.4, 208.217.179.5, 208.217.179.6);
ourIP = listadd(our_IPs, 208.217.179.7, 208.217.179.8, 208.217.179.9);
echo(sublist(ourIP, 3, 3));
echo(sublist(ourIP, 8));
Displays:
[208.217.179.4, 208.217.179.5, 208.217.179.6]
[208.217.179.9]
```

### `substr()`

The `substr()` function creates a new string from another string, starting with the character specified at the offset start, and continuing for the number of characters specified for length.

When using the `substr()` function, remember that the first element in the list is always zero.

If a length is not included, the new string starts at the offset and uses the remaining characters in the old string.

#### Syntax

```plaintext
substr(string src, [int start :=0, int length]);
```

#### Parameters

- `string src` - the string to be examined
- `int start` - the offset in the string for the beginning of the new string
- `int length` - the number of characters of the first string to use in the second string

#### Return Value

`string` - the new string

#### Example

```plaintext
substr("Network Security Inc", 8, 8) == "Security"
substr("foo bar", 4) == "bar"
```

### `succ()`

The `succ()` function finds and returns the logical successor of the indexed item in a passed array.

The `succ()` function allows foreach-like behavior, but at a larger performance cost.

#### Syntax

```plaintext
succ(array arr, any index);
```

#### Parameters

- `array arr` - the array from which to extract the successor
- `any index` - the index to the item whose successor will be extracted

#### Return Value

`any` - the successor of the indexed item in the array (that is, the item that will be extracted)
suffix()

The suffix() function compares two strings and returns true if the second string is a suffix of the first string. Use the casesuffix() function for a case-insensitive comparison.

Syntax

```
suffix(string big, string little);
```

Parameters

- string big - the string against which little will be compared
- string little - the smaller string, which will be compared to the end of big

Return Value

int, where:
- 1 - little is the same as the end of big
- 0 - little is not the same as the end of big

Example

```
suffix("ABCDEFG", "EFG") == 1
suffix("ABCDEFG", "eFg") == 0
suffix("ABCDEFG", "ABC") == 0
```

tolower()

The tolower() function converts all English alphabetic characters from uppercase to lowercase. Other characters, including alphabetic characters in other character sets, are not converted.

The tolower() function has a pre-optimized fastpath, where str is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The tolower() function may be impacted by the change in the behavior of string variables. See str for additional information about this behavior.

The overall return string for the tolower() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

Syntax

```
tolower(string in);
```

Parameters

- string in - the string to be converted to lowercase

Return Value

string - the same string that was passed in, with all uppercase characters (A-Z) converted to lowercase characters (a-z).

Example

```
tolower("AB12fj4") == "ab12fj4"
tolower("ABCdefG") == "abcdefg"
$foo = tolower($foo); # FAST
```
toupper()

The toupper() function converts all English alphabetic characters from lowercase to uppercase. Other characters, including alphabetic characters in other character sets, are not converted.

The toupper() function has a pre-optimized fastpath, where str is its most relevant parameter. It operates more efficiently if its input string is the same CPcode variable as its output string.

The toupper() function may be impacted by the change in the behavior of string variables. Refer to the description of the string data type on page B-5 for additional information about this behavior.

The overall return string for the toupper() function cannot be larger than 8KB. You should avoid passing in an input string larger than 8KB.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>toupper(string in);</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>string in - the string to be converted to uppercase</td>
</tr>
<tr>
<td>Return Value</td>
<td>string - the same string that was passed in, with all lowercase characters (a-z) converted to uppercase characters (A-Z).</td>
</tr>
<tr>
<td>Example</td>
<td>toupper(&quot;AB12fj4&quot;) == &quot;AB12FJ4&quot;</td>
</tr>
<tr>
<td></td>
<td>toupper(&quot;ABCdefG&quot;) == &quot;ABCDEFG&quot;</td>
</tr>
<tr>
<td></td>
<td>$foo = toupper($foo); # FAST</td>
</tr>
</tbody>
</table>

typeof()

The typeof() function returns the data type of an expression.

You can use the typeof() function for error checking or for conditional operations. However, there is a performance penalty associated with this usage.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>typeof(any argument);</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>any argument - the variable whose data type will be returned</td>
</tr>
<tr>
<td>Return Value</td>
<td>string - a string indicating the data type of the argument, where:</td>
</tr>
<tr>
<td></td>
<td>alert - the argument is an alert</td>
</tr>
<tr>
<td></td>
<td>array - the argument is an array</td>
</tr>
<tr>
<td></td>
<td>ethmac - the argument is an ethmac address</td>
</tr>
<tr>
<td></td>
<td>filter tag - the argument is filter tag</td>
</tr>
<tr>
<td></td>
<td>function pointer - the argument is a function pointer</td>
</tr>
<tr>
<td></td>
<td>error - the argument is an error code</td>
</tr>
<tr>
<td></td>
<td>int - the argument is an integer</td>
</tr>
<tr>
<td></td>
<td>ipv4host - the argument is an IPv4 host address</td>
</tr>
<tr>
<td></td>
<td>ipv4net - the argument is an IPv4 network address</td>
</tr>
<tr>
<td></td>
<td>ipv6host - the argument is an IPv6 host address</td>
</tr>
<tr>
<td></td>
<td>ipv6net - the argument is an IPv6 network address</td>
</tr>
<tr>
<td></td>
<td>list - the argument is a list</td>
</tr>
<tr>
<td></td>
<td>recorder - the argument is a recorder</td>
</tr>
<tr>
<td></td>
<td>str - the argument is a string, a blob, or a pattern</td>
</tr>
<tr>
<td></td>
<td>unknown - a processing error occurred</td>
</tr>
</tbody>
</table>
Example
```cpp
if (typeof($1) != "str")
    return 0;
```

**ubyte()**

The `ubyte()` function returns a one-byte unsigned integer from the specified offset within a string.

**Syntax**

```cpp
ubyte(string str, int offset);
```

**Parameters**

- `string str` - the string to be examined, beginning at byte 0
- `int offset` - the offset in the string for the byte to be returned

**Return Value**

- `int` - the offset integer, as an unsigned value
- `error` - the offset is out of bounds

**Example**

```cpp
#the eighth byte is the unsigned type code
$code = ubyte( $struct_data, 7 );
```

**ulong()**

The `ulong()` function returns a signed four-byte integer from a specified offset within a string. It converts the value from network byte order in the string to the host's byte order. Network byte order is big endian; therefore, the most significant byte is stored at the lowest address in memory.

The `ulong()` function is useful for extracting values from packets.

The `ulong()` function returns the same result as the `long()` function. It is provided for backwards compatibility.

The internal representation of all CPcode integers is signed. Therefore, the `ulong()` function yields a signed integer.

**Syntax**

```cpp
ulong(string str, int offset);
```

**Parameters**

- `string str` - the string to be examined, beginning at byte 0
- `int offset` - the offset in the string for the first byte to be returned

**Return Value**

- `int` - the offset integer, as a signed value
- `error` - the offset is out of bounds
Example

```php
# Structure starts with a 4-byte integer
# in network byte order.
# Converted to host byte order
# before placing it in $data
$data = ulong($struct_data, 0);
# Caveat
$blob = cat(-1);
echo(ulong($blob, 0)); # -1
echo(ulong($blob, 0) >> 2); # -1
$blob = cat(-4);
echo(ulong($blob, 0) >> 1); # -2
# Be _very_ careful with
# shifting signed integers.
# Their sign bit is extended.
```

unpack()

The unpack() function extracts the individual variables from a structure that was packed using the pack() function.

**Syntax**

```php
int unpack(list schema, string packed, var1, var2, ...);
```

**Parameters**

- `list schema` - the result of a call to the packing_schema() function
- `string packed` - a packed structure
- `var1, var2, varn` - the variables in which to store data elements from the unpacked structure

**Return Value**

`int` - `!NULL` if the unpacking was successful

**Example**

```php
{
    schema = packing_schema("ip", "ip", "port");
    ...
    $struct = pack(schema, $src_ip, $dst_ip, $dst_port);
    if (ALERTS[$struct] == NULL
        ALERTS[$struct] = 1;
    else
        ALERTS[$struct] = ALERTS[$struct] + 1;
    ...
    unpack(schema, $struct, $src, $dst, $dport);
    msg = cat(ALERTS[$struct], " attacks from ", $src, " against ", $dst, ":", $dport);
    echo( $msg);
}
```

url_flush()

The url_flush() function removes a specified URL from cache.
Cache control parameters are inherited by the CPcode runtime from the calling application for use with the url_* functions.

**url_flush()**

The `url_flush()` function clears the cache for the specified URL.

**Syntax**

```
int url_flush ( URL )
```

**Parameters**


**Return Value**

None

**url_needs_refresh()**

The `url_needs_refresh()` function indicates whether or not the specified URL needs to be updated because it has been modified and/or the cached representation is stale.

**Cache control parameters are inherited by the CPcode runtime from the calling application for use with the url_* functions.**

**Syntax**

```
int url_needs_refresh ( URL )
```

**Parameters**


**Return Value**

Int – an integer with one of the following possible values:

- 0 – the cached URL is up to date.
- 1 – the cached URL is stale.
- 2 – the URL is not cached.

**url_open()**

The `url_open()` function returns a handle to the resource which can be used with the I/O builtins (io_*) to retrieve data. The user must call io_close on the handle to release it.

If the cached representation is stale and the resource has been modified, the cached representation will be refreshed before the handle is returned.

**Cache control parameters are inherited by the CPcode runtime from the calling application for use with the url_* functions.**

**Syntax**

```
int url_open ( URL, <username>, <password> )
```

**Parameters**

- **<username>** - optional parameter supporting basic HTTP authentication.
- **<password>** - optional parameter supporting basic HTTP authentication.

**Return Value**

Int – Returns a handle to the cached resource if successful, otherwise NULL if the not successful.
url_readall()

The url_readall() function returns all of the contents of the specified URL.
Cache control parameters are inherited by the CPcode runtime from the calling application for use with the url_* functions.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>string url_readall ( URL, &lt;username&gt;, &lt;password&gt; )</th>
</tr>
</thead>
</table>
• <username> - optional parameter supporting basic HTTP authentication,  
• <password> - optional parameter supporting basic HTTP authentication. |
| Return Value | string - A string containing the contents of the specified URL. |

url_readlines()

The url_readlines() function returns the contents of the specified URL as a list of strings with each entry containing a single line. Line terminators are not removed.
Cache control parameters are inherited by the CPcode runtime from the calling application for use with the url_* functions.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>list url_readlines ( URL, &lt;username&gt;, &lt;password&gt; )</th>
</tr>
</thead>
</table>
• <username> - optional parameter supporting basic HTTP authentication,  
• <password> - optional parameter supporting basic HTTP authentication. |
| Return Value | list - A list of strings. Each entry is a single line from the URL. |

ushort()

The ushort() function returns a two-byte value as an unsigned integer from a specified offset within a string. It converts the value from network byte order in the string to the host's byte order. Network byte order is big endian; therefore, the most significant byte is stored at the lowest address in memory.
The ushort() function is useful for extracting values from packets.

<table>
<thead>
<tr>
<th>Syntax</th>
<th>ushort(string str, int offset);</th>
</tr>
</thead>
</table>
| Parameters | • string str - the string to be examined, beginning byte 0  
• int offset - the offset in the string for the first byte to be returned |
| Return Value | • int - the offset integer, as an unsigned value  
• error - the offset is out of bounds |
| **Example** | # Structure starts with a 2-byte integer  
# in network byte order.  
# Converted to host byte order  
# before placing it in $data  
$data = ushort($struct_data, 0); |