TinyVM

工程源码: https://github.com/jakogut/tinyvm

一个用纯C写的小小型虚拟机，模拟了部分x86汇编。可以执行基本的算术、逻辑运算，支持标签和函数调用。一个执行流，其他功能全无。

支持的指令如下:

const char \*tvm\_opcode\_map[] = {"nop", "int", "mov", "push", "pop", "pushf", "popf", "inc", "dec", "add", "sub", "mul", "div", "mod", "rem","not", "xor", "or", "and", "shl", "shr", "cmp", "jmp", "call", "ret", "je", "jne", "jg", "jge", "jl", "jle", "prn", 0};

寄存器如下:

const char \*tvm\_register\_map[] = {"eax", "ebx", "ecx", "edx", "esi", "edi", "esp", "ebp", "eip", "r08", "r09", "r10", "r11", "r12", "r13", "r14", "r15", 0};

其中除了INT，其他都有实现。

VM实现的逻辑是，先从用上述指令编写的程序文件开始解析，依次进行预处理，词法分析，解析出指令流和分配变量地址空间。最后再一条一条地执行指令。而PRN指令是专用来做print的。

在做预处理时主要用的数据结构是哈希表。将``宏’’字符串生成哈希值，相同的哈希值用链表链接起来以解决冲突.

类型声明如下:

typedef struct tvm\_htab\_node\_s

{

char \*key;

int value;

void \*valptr;

struct tvm\_htab\_node\_s \*next; //链接具有相同哈希值的其他node

} tvm\_htab\_node\_t;

typedef struct tvm\_htab\_s

{

unsigned int num\_nodes;

unsigned int size;

tvm\_htab\_node\_t \*\*nodes; //一个nodes\*数组, 大小4096

} tvm\_htab\_t;

比如虚拟机执行的程序文件内容

Start:

Mov eax, 0

Mov ebx, 1

Loop:

Mov ecx, ebx

Add ebx, eax

Mov eax, ecx

Prn ebx

Cmp ebx, 1000000|h

Jl loop

End:

上述保存文件名为fib.vm

执行命令: $ ./tvmi.exe fib.vm > catofpei.log

经过词法分析后得到的tokens如下:

[0:0] start:

[1:0] mov [1:1] eax [1:2] 0

[2:0] mov [2:1] ebx [2:2] 1

[3:0] loop:

[4:0] mov [4:1] ecx [4:2] ebx

[5:0] add [5:1] ebx [5:2] eax

[6:0] mov [6:1] eax [6:2] ecx

[7:0] prn [7:1] ebx

[8:0] cmp [8:1] ebx [8:2] 1000000|h

[9:0] jl [9:1] loop

[10:0] end:

而最终得到的指令流如下：

[0] mov

[1] mov

[2] mov

[3] add

[4] mov

[5] prn

[6] cmp

[7] jl

[8] nop

[9] nop

[10] nop

VM最后一步一步执行指令:

void tvm\_step(tvm\_t \*vm, int \*instr\_idx)

打印出斐波那契数列.

核心东西如下：



这个VM支持的功能太少了，既不能分段更不能分页，没有中断更没有多线程，不过也算小小小..虚拟机..了

作者列了个TO-DO列表:

|  |
| --- |
| TODO: |
|  |  |
|  | (Basic) |
|  | - NASM style local labels |
|  | - Implement defining bytes/words/dwords |
|  | - Add C style defines to preprocessor |
|  | - Move parsed TVM program code into the VM's virtual address space |
|  |  |
|  | (Advanced) |
|  | - Fix/refactor the debugger (it doesn't work) |
|  | - Interrupts |
|  | - SDL or GLFW based screen for outputting the contents of a framebuffer |
|  | - JIT compilation |
|  | - C interface |
|  | - C Library written in TVM code |
|  |  |
|  | Please send patches to Joseph Kogut <joseph.kogut(at)gmail.com> |

貌似… @\_@#!