Prevenção de Ataques de Não-Terminação baseados em Estouros de Precisão

Raphael Ernani Rodrigues and Fernando Magno Quintão Pereira

Departamento de Ciência da Computação – UFMG – Brasil
{raphael,fernando}@dcc.ufmg.br

Resumo Dizemos que um programa é vulnerável a um ataque de não terminação quando um adversário pode lhe fornecer valores de entrada que façam algum de seus laços iterar para sempre. A prevenção de ataques desse tipo é difícil, pois eles não se originam de bugs que infringem a semântica da linguagem em que o programa foi feito. Ao contrário, essas vulnerabilidades têm origem na aritmética modular inteira de linguagens como C, C++ e Java, a qual possui semântica bem definida. Neste artigo nós apresentamos uma ferramenta que detecta tais problemas, e que saneia código vulnerável. A detecção da vulnerabilidade dá-se via uma análise de fluxo de informação; a sua cura decorre de guardas que nosso compilador insere no código vulnerável. Nós implementamos esse arcabouço em LLVM, um compilador de qualidade industrial, e testa-mo-no em um conjunto de programas que compraz mais de 2.5 milhões de linhas de código escrito em C. Descobrimos que, em média, caminhos em que informação perigosa trafega são pequenos, sendo compostos por não mais que 13 instruções assembly. A instrumentação que inserimos para prevenir ataques de não terminação aumenta o tamanho do programa saneado em cerca de 5% em média, e torna-os menos que 2% mais lentos.

Resumo We say that a program is vulnerable to a non-termination attack if (i) it contains a loop that is bounded by values coming from public inputs, and (ii) an adversary can manipulate these values to force this loop to iterate forever. Preventing this kind of attack is difficult because they do not originate from bugs that break the semantics of the programming language, such as buffer overflows. Instead, they usually are made possible by the wrapping integer arithmetics used by C, C++ and Java-like languages, which have well-defined semantics. In this paper we present the diagnosis and the cure for this type of attack. Firstly, we describe a tainted-flow analysis that detects non-termination vulnerabilities. Secondly, we provide a compiler transformation that inserts arithmetic guards on loop conditions that may not terminate due to integer overflows. We have implemented our framework in the LLVM compiler, and have tested it on a benchmark suite containing over 2.5 million lines of C code. We have found out that the typical path from inputs to loop conditions is, on the average, less than 13 instructions long. Our instrumentation that prevents this kind of attacks adds in average less than 5% extra code on the secured program. The final, protected code, is, on the average, less than 2% slower than the original, unprotected program.

1 Introdução

Um ataque de Negação de Serviços (*Denial-of-Service – DoS*) consiste em sobrecarregar um servidor com uma quantidade de falsas requisições grande o suficiente para lhe comprometer a capacidade de atender contatos legítimos. Existem hoje diversas maneiras diferentes de detectar e reduzir a efetividade desse tipo de ataque [13]. Neste artigo, contudo, descreveremos uma forma de negação de serviço que é de difícil detecção: *os ataques de não-terminação*. Um adversário realiza um ataque desse tipo fornecendo ao programa alvo entradas cuidadosamente produzidas para forçar iterações eternas sobre um laço vulnerável. Um ataque de não terminação demanda conhecimento do código fonte do sistema a ser abordado. Não obstante tal limitação, esse tipo de ataque pode ser muito efetivo, pois bastam algumas requisições para comprometer o sistema alvo. Uma vez que essa quantidade de incursões é pequena, os métodos tradicionais de detecção de negação de serviço não podem ser usados para reconhecer ataques de não-terminação. Além disso, dada a vasta quantidade de código aberto usado nos mais diversos ramos da indústria de *software*, usuários maliciosos têm a sua disposição um vasto campo de ação.

A detecção de código vulnerável a ataques de não-terminação é difícil. Tal dificuldade existe, sobretudo, porque esse tipo de ataque não decorre de deficiências de tipagem fraca, normalmente presentes em programas escritos em C ou C++. Programas escritos em linguagens fortemente tipadas, como Java, por exemplo, também apresentam a principal fonte de vulnerabilidades a ataques de não terminação: a *aritmética modular inteira*. Em outras palavras, uma operação como `a + 1`, em Java, C ou C++, pode resultar em um valor menor que aquele inicialmente encontrado na variável `a`. Esse fenômeno ocorrerá quando a variável `a` guardar o maior inteiro presente em cada uma dessas linguagens. Nesse caso, ao fim da operação, `a + 1` representará o menor inteiro possível em complemento de dois. Em outras palavras, um laço como `for (i = 0; i <= N; i++)` nunca terminará se `N` for `MAX_INT`, o maior inteiro da linguagem.

Este artigo traz duas contribuições relacionadas a ataques de não-terminação. Em primeiro lugar, ele descreve uma técnica que descobre vulnerabilidades relacionadas a esse tipo de ataque. Em segundo lugar, o artigo mostra como código pode ser protegido contra tais ataques. A nossa técnica de detecção de vulnerabilidades é baseada em análise de fluxo contaminado. Tal análise é parte do arcabouço teórico de rastreamento de informação inicialmente proposto por Denning e Denning nos anos setenta [6]. Um ataque de fluxo contaminado pode ser efetivo somente em programas cujas operações críticas dependam de dados de entrada. Em nosso contexto, uma operação crítica é o teste de controle de laço. Em conjunto com o algoritmo de detecção de vulnerabilidades, nós propomos também uma técnica para sanear programas contra ataques de não-terminação. Nossa estratégia consiste na inserção de verificações sobre as operações aritméticas realizadas pelo programa alvo. Essas verificações ocorrem durante a execução do programa, e invocam código de segurança sempre que estouros de precisão em variáveis inteira são percebidos. Nós instrumentamos somente código que controla o número de iterações em laços. Consequentemente, o arcabouço que

propomos incorre em uma perda de desempenho muito pequena, e, em nossa opinião, completamente justificável em decurso do benefício que assegura.

Nós implementamos todas as ideias que discutimos neste artigo em LLVM, um compilador de qualidade industrial [11]. Na seção 4 descreveremos uma série de experimentos que validam nossa análise. Ao analisar os programas presentes na coleção SPEC CPU 2006, fomos capazes de descobrir 12.304 laços que são influenciados por dados provenientes de entradas públicas, isto é, que podem ser manipuladas por um adversário. Dentre esses laços, pelo menos 920 estão sujeitos à não terminação. Para obter esse número, utilizamos um padrão muito simples: procuramos por laços cujo teste de parada é do tipo `i <= N`, sendo `N` dependente da entrada. Uma vez que nos atemos a esse tipo de condição de parada, especulamos que a quantidade de laços vulneráveis presente nos programas de SPEC CPU 2006 seja bem maior que o valor que apuramos. A nossa instrumentação – usada para impedir os ataques de não-terminação – mostra-se extremamente eficiente. Os testes que inserimos antes de cada operação aritmética que pode levar a um ataque de não terminação custa-nos uma perda de desempenho de menos que XX%.

2 Ataques de Não-Terminação

De acordo com Appel e Palsberg [1, p.376], um laço natural é um conjunto de nodos S do grafo de fluxo de controle (CFG) de um programa, includindo um nodo cabeçalho H, com as seguintes três propriedades:

- a partir de qualquer nodo em S existe um caminho que chega a H;
- existe um caminho de H até qualquer nodo que faz parte de S;
- qualquer caminho de um nodo fora de S para um nodo em S contém H.

A *condição de parada* de um laço é uma expressão booleana $E = f(e_1, e_2, \ldots, e_n)$, sendo cada $e_j, 1 \leq j \leq n$ um valor que contribui para a computação de E. Seja P um programa que possui um laço L limitado por uma condição de parada $E = f(e_1, e_2, \ldots, e_n)$. Dizemos que P é vulnerável a um ataque de não terminação quando as duas condições abaixo são verdadeiras sobre ele:

1. existe um subconjunto $E' \subseteq \{e_1, e_2, \ldots, e_n\}$ de valores que dependem de um conjunto $I = \{i_1, i_2, \ldots, i_m\}$ de dados lidos a partir da entrada do programa.
2. existe uma atribuição de valores $i_1 \mapsto v_1, i_2 \mapsto v_2, \ldots, i_m \mapsto v_m$ que, ao influenciar E', faz com que o laço L não termine.

Note que a nossa definição de vulnerabilidade de não-terminação requer a noção de dependência de dados. Se um programa P possui uma instrução que usa a variável u e define a variável v, então v *depende* de u. Dependências de dados são transitivas, e embora possam ser cíclicas, não são necessariamente comutativas.

Ilustraremos ataques de não terminação via o programa mostrado na Figura 1(a). Esse programa calcula o fatorial de um número inteiro na linguagem C. O padrão que rege essa linguagem de programação não determina o tamanho do tipo `int`. Essa informação depende da implementação do compilador

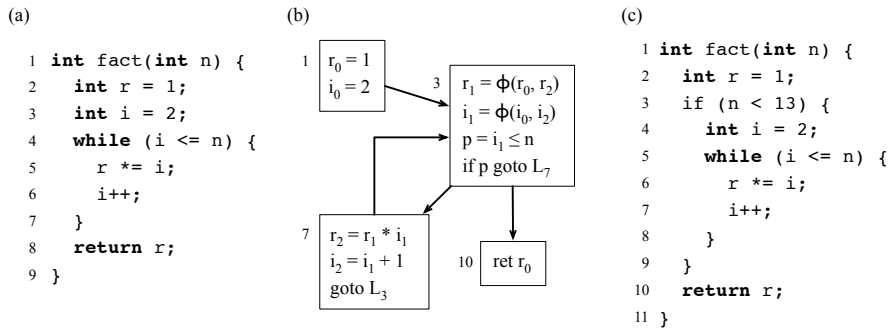

Figura 1. (a) Uma função em C, que calcula o fatorial de um número inteiro. (b) O CFG da função `fact`. (c) Exemplo de laço cuja condição de parada depende de valores de entrada mas que sempre termina.

usado. Entretanto, é usual que inteiros sejam representados como números de 32 bits em complemento de dois. Nesse caso, o maior inteiro representável é $MAX\_INT = 2^{31} - 1 = 2,147,483,647$. Se o parâmetro `n` for igual a MAX_INT, então a condição da linha 4 sempre será verdadeira, e o laço nunca terminará. A não-terminação ocorre porque quando `i` finalmente chega a MAX_INT, a soma $i + 1$ nos retorna o menor inteiro possível, isto é, -2^{31}.

O programa da figura 1(a) é vulnerável a ataques de não-terminação. Para explicitar tal fato, a figura 1(b) mostra o grafo de fluxo de controle do programa. Esse CFG está convertido para o formato de atribuição estática única (SSA) [5]. Usaremos essa representação de programas porque ela facilita a nossa análise de dependência de dados. Os blocos básicos que começam nos rótulos 3 e 7 formam um laço natural, segundo a definição de Appel e Palsberg. Esse laço é controlado pela condição de parada $i_1 \leq n$. A variável n, o limite do laço, é dependente da entrada. Existe um valor de n, a saber MAX_INT, que força o laço a não-terminar.

3 Detecção e Prevenção de Não-Terminações

Nesta seção descreveremos nossa técnica para detectar vulnerabilidades de não-terminação. Esse algoritmo de detecção fornece os subsídios necessários a uma segunda técnica que introduzimos neste artigo: o *saneamento de laços*.

3.1 Detecção Automática de Não-Terminação

Dizemos que um laço é *alcançável* quando as condições que o controlam usam valores que dependem de dados de entrada do programa. Note que um laço alcançável não é necessariamente vulnerável. A título de exemplo, o programa da

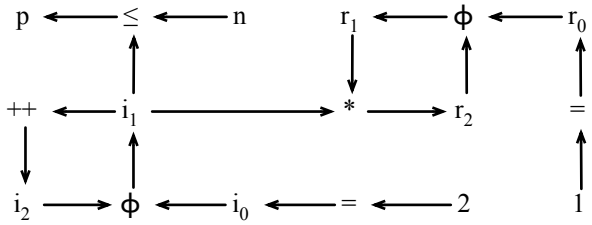

Figura 2. (a) Grafo de dependências da função `fact`, construído a partir do CFG visto na figura 1(b).

Figura 1(c), uma sutil alteração da função `fact` inicialmente vista na figura 1(a), termina para qualquer entrada, embora ele contenha um laço alcançável. Utilizamos o *grafo de dependências de dados* para determinar laços alcançáveis. Esse grafo é definido da seguinte forma: para cada variável v no programa, nós criamos um nodo n_v, e para cada instrução i no programa, nós criamos um nodo n_i. Para cada instrução $i : v = f(\ldots, u, \ldots)$ que define uma variável v e usa uma variável u nós criamos duas areastas: $n_u \leftarrow n_i$ e $n_i \leftarrow n_v$. O grafo de dependências que extraímos a partir d CFG visto na figura 1(b) é mostrado na figura 2.

Um caminho entre uma entrada do programa, e o predicado que controla o laço é uma condição necessária para um ataque de não-terminação. O grafo de dependências de nosso exemplo apresenta tal condição: existe um caminho que une o nodo correspondente ao parâmetro `n`, uma entrada, ao nodo que corresponde a `p`, o predicado de controle do laço. Esse tipo de caminho, uma vez contruído o grafo, pode ser encontrado em tempo linear no número de arestas do grafo - normalmente proporcional ao número de nodos - via a simples busca em profundidade ou largura.

Diremos que um laço é *vulnerável* quando ele é alcançável, e, além disso, sua condição de parada é dependente de alguma operação *cíclica* passível de estouro de precisão. Seguindo a definição de laços de Appel e Palsberg, uma operação cíclica é qualquer instrução que ocorre no corpo S do laço. Por exemplo, no CFG da figura 1(b), as instruções $i_2 = i_1 + 1$ e $r_2 = r_1 \times i_1$ são cíclicas. O laço daquele exemplo encaixa-se em nossa definição de vulnerabilidade, pois sua condição de parada é alcançável a partir da entrada, e depende de uma instrução cíclica passível de estouro de precisão: $i_2 = i_1 + 1$.

A nossa definição de vulnerabilidade inclui muitos laços que não são concretamente vulneráveis, tais como aquele visto na figura 1(c). Seria possível utilizar técnicas computacionalmente intensivas, tais como algoritmos de satisfabilidade, para refinar a nossa definição, eliminando alguns desses falsos positivos. Tal abordagem já foi utilizada em trabalhos anteriores ao nosso [2,4,9,15,17]. Por outro lado, os próprios autores desses trabalhos reportam que dificilmente suas técnicas poderiam lidar com programas muito grandes. Nós optamos por usar

uma definição mais conservadora de laço vulnerável para termos uma ferramenta prática. Nós sanearemos todo laço considerado perigoso, inclusive aqueles que, devido à nossa definição liberal de vulnerabilidade, de fato não o são. Ainda assim, conforme mostraremos na seção 4, o impacto dessa instrumentação é negligível.

3.2 Saneamento de Laços

Uma vez encontrado um caminho vulnerável, passamos à fase de saneamento de laços. Um laço pode ser saneado via a inserção de testes que detectam e tratam a ocorrência de estouros de precisão inteira. Nós inserimos tais testes sobre as operações aritméticas cíclicas que controlam a condição de parada do laço. Continuando com o nosso exemplo, o laço alvo possui dois blocos básicos: o primeiro começa no rótulo três, e o segundo começa no rótulo sete. O laço possui duas operações aritméticas cíclicas, todas ocorrendo no segundo bloco básico. Dentre essas operações, aquela no rótulo sete é inofensiva: ela define a variável r_2, que não participa da condição de parada do laço. Por outro lado, a operação no rótulo oito, que define a variável i_2, é usada no cálculo daquela condição, e precisa ser instrumentada.

Novamente, o grafo de dependências ajuda-nos a encontrar quais operações precisam ser instrumentadas para sanear um laço controlado por um predicado p. Nesse caso, usamos o seguinte critério para determinar se uma operação $i : v = f(v_1, \ldots, v_n)$ precisa ser instrumentada:

- existe um caminho no nodo n_i até o nodo n_p.
- O nodo n_i encontra-se em um ciclo.

A título de exemplo, a operação de incremento ++ no grafo de dependências visto na figura 2 precisa ser instrumentada. Em primeiro lugar, porque essa operação encontra-se em um ciclo. Em segundo lugar, porque existe um caminho do nodo n_{++} até o nodo n_p.

Instrumentação de Saneamento. Para evitar que estouros de precisão venham a causar a não-terminação de laços, nós inserimos testes no código binário do programa alvo. O código que constitui cada um desses testes é formado por uma guarda, mais um tratador de eventos. Nossas guardas usam as condições mostradas na figura 3 para verificar a ocorrência de estouros de precisão. Atualmente instrumentamos quatro tipos diferentes de instrução: adição, subtração, multiplicação e arredamentos para a esquerda, conhecidos como *shift left*. As operações de adição, subtração e multiplicação podem ser com ou sem sinal aritmético.

Os testes são implementados como sequências de operações binárias, executados logo após a instrução guardada. Para ilustrar esse ponto, mostramos, na figura 4, o código necessário para instrumentar uma soma com sinal de duas variáveis. Essa figura mostra código no formato intermediário usado por LLVM, o compilador que utilizamos para implementar as idéias descritas neste artigo. Omitimos, nesse exemplo, o código do tratador de evento de estouro, pois ele simplesmente invoca uma rotina implementada em uma biblioteca dinamicamente

Instrução	Verificação
$x = o_1 \ +_s \ o_2$	$(o_1 > 0 \wedge o_2 > 0 \wedge x < 0) \ \vee$ $(o_1 < 0 \wedge o_2 < 0 \wedge x > 0)$
$x = o_1 \ +_u \ o_2$	$x < o_1 \vee x < o_2$
$x = o_1 \ -_s \ o_2$	$(o_1 < 0 \vee o_2 > 0 \vee x > 0) \ \vee$ $(o_1 > 0 \vee o_2 < 0 \vee x < 0)$
$x = o_1 \ -_u \ o_2$	$o_1 < o_2$
$x = o_1 \ \times_{u/s} \ o_2$	$x \neq 0 \Rightarrow x \div o_1 \neq o_2$
$x = o_1 \ll n$	$(o_1 > 0 \wedge x < o_1) \vee (o_1 < 0 \wedge n \neq 0)$
$x = \downarrow_n \ o_1$	$\text{cast}(x, \text{type}(o_1)) \neq o_1$

Figura 3. Overflow checks. Usamos $\downarrow_n$ para descrever a operação que trunca em n bits. O subscrito s indica uma operação aritmética com sinal, e o subscrito u indica uma operação sem sinal.

compartilhada. Conforme podemos observar pela figura, uma guarda aumenta o código instrumentado substancialmente. Nesse exemplo em particular a verificação requer a inserção de 14 novas instruções no programa guardado. Embora tal crescimento a princípio possa parecer proibitivamente grande, os experimentos que mostraremos na seção 4 indicam que somente uma parcela muito pequena das instruções do programa alvo precisam ser guardadas. Consequentemente, o custo, em termos de crescimento de código e perda de desempenho, é negligível.

4 Resultados Experimentais

Nós implementamos as técnicas descritas neste artigo em LLVM versão 3.3. Nossa implementação foi testada em uma máquina Intel® Core™ i7-3770, com 16 Gigabtyes de RAM, e 3.40 GHz de Clock. Executamos nossa análise com sucesso sobre o arcabouço de testes do LLVM, um conjunto de programas com mais de 4.3 milhões de linhas de código C. No restante desta seção mostraremos somente resultados obtidos sobre os programas disponíveis em SPEC CPU 2006.
Definição de Entrada de Dados. As entradas de dados são as funções que um adversário pode usar para forçar a não-terminação de um programa. Nos experimentos apresentados nesta seção, consideraremos como entrada de dados os seguintes valores:

- os argumentos do método `main`, isto é, as variáveis `argc` e `argv`;
- o resultado retornado por funções *externas*;
- ponteiros passados como argumento de funções *externas*.

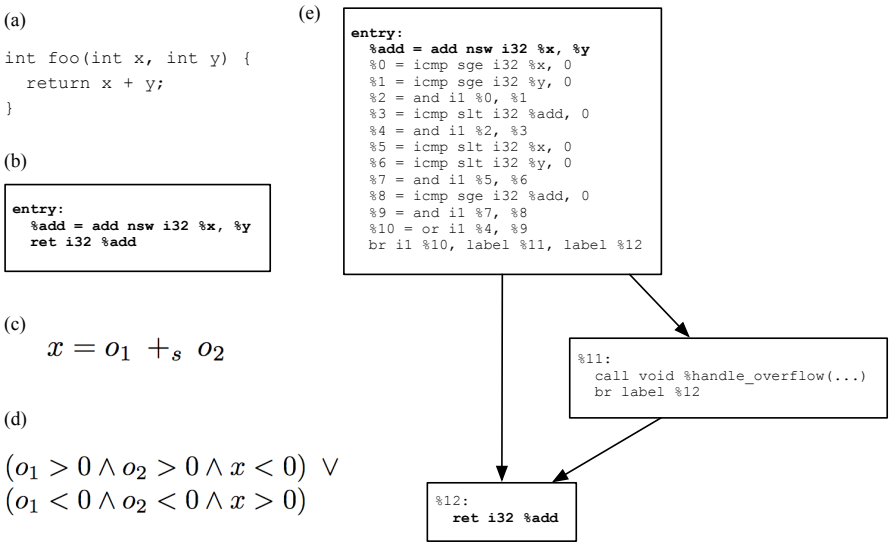

Figura 4. (a) Programa que será instrumentado. (b) Representação do programa em bytecodes LLVM. (c) Operação que está sendo instrumentada. (d) Teste de detecção de estouro de precisão. (e) Programa instrumentado, em bytecodes LLVM.

As funções externas são a união dos seguintes três conjuntos:

- funções que não foram declaradas em nenhum dos arquivos que compõem o programa compilado;
- funções sem corpo;
- funções que podem ser chamadas via um ponteiro de funções.

Grafo de dependências de dados. A figura 5 mostra informações estáticas a respeito dos grafos de dependências dos programas analisados. Como ponto de referência, mostramos o tamanho absoluto de cada programa, em número de instruções. Em média 14% dos valores que cada programa manipula podem conter informações vindas da entrada de dados. Esses valores são produzidos por funções externas, para as quais não é possível saber se há ou não contaminação por dados externos. Observa-se que existem mais nós de operação no grafo do que instruções no programa. Esse fato verifica-se porque nossa análise é interprocedural. Assim, inserimos nós de operação para criar dependências entre parâmetros formais a parâmetros reais e também para ligar valores de retorno aos valores que recebem o resultado de funções.

A figura 5 também apresenta a quantidade de nós de memória. Esses nós são vértices do grafo de dependência que representam os ponteiros e os blocos de memória alocados no código do programa. A fim de determinar se dois ponteiros são sinônimos ou não, implementamos a análise de apontadores de Andersen [**?**].

Benchmark	Insts.	Entr.	Ops.	Vars.	Mems.	Arestas
433.milc	24,971	3,995	24,799	20,435	6,901	74,599
444.namd	77,922	15,964	78,043	72,866	10,468	232,792
447.dealII	483,614	98,180	512,374	441,576	131,986	1,516,202
450.soplex	67,808	11,182	69,870	56,172	21,764	204,159
470.lbm	3,788	239	3,787	3,490	626	10,859
400.perlbench	288,429	25,224	287,050	230,982	98,886	819,872
401.bzip2	17,999	1,862	18,007	15,019	4,423	53,414
403.gcc	830,861	65,118	830,054	660,772	321,279	2,440,279
429.mcf	2,851	373	2,897	2,354	905	8,608
445.gobmk	146,298	21,420	152,342	167,197	21,146	458,405
456.hmmer	62,704	8,487	63,004	51,549	20,727	182,736
458.sjeng	25,473	2,610	25,169	26,313	2,522	73,355
462.libquantum	6,562	921	6,552	5,845	940	19,142
464.h264ref	141,772	11,995	141,606	106,292	45,219	409,813
471.omnetpp	96,929	15,989	101,197	88,819	31,686	305,170
473.astar	9,386	1,506	9,476	8,199	1,982	28,067
483.xalancbmk	648,941	132,976	689,176	569,780	244,874	1,971,945
Total	2,936,308	418,041	3,015,403	2,527,660	966,334	8,809,417

Figura 5. Dados do grafo de dependência dos programas analisados. **Insts.**: número de instruções do programa. **Entr.**: número de instruções que podem representar entradas de dados. **Ops.**: número de nós de operação no grafo de dependência. **Vars.**: número de nós que representam variáveis. **Mems.**: número de nós que representam blocos de memória. **Arestas**: número de arestas do grafo.

Sinônimos, isto é, apontadores que indicam regiões de memória sobrepostas, são agrupados em um mesmo nó. Quanto mais precisa a análise de ponteiros, menores os conjuntos contidos em cada uma dessas unidades. A análise de Andersen é um compromisso entre eficiência e precisão.

Analisando a figura 5, constatamos que os grafos de dependências são esparsos. Nos programas de SPEC CPU 2006, a razão entre o número de vértices e o número de arestas é 1.38. Essa relação fica ainda mais evidente quando extraímos o coeficiente de determinação entre essas duas grandezas – número de vértices e arestas. Obtemos o valor de 0.99, o que indica uma forte correlação linear entre os dois valores. Essa baixa densidade ocorre porque programas reais variáveis tendem a ser usadas um número baixo de vezes. Boissinot *et al.* [**?**] demonstraram, empiricamente, que a maior parte das variáveis é usada somente uma vez no programa, e mais de 99% das variáveis são usadas menos que cinco vezes. Assim, a maior parte dos vértices que representam variáveis em nossos grafos de dependências possuem grau de saída inferior a cinco.

Benchmark	Laços (L)	Alcançáveis	Caminhos	Vulneráveis		% (II/L)
				I	II	
433.milc	329	146	25	138	0	0.00%
444.namd	484	438	9	408	2	0.41%
447.dealII	4,759	3,493	12	2,657	73	1.53%
450.soplex	542	513	14	453	175	32.29%
470.lbm	18	0	0	0	0	0.00%
400.perlbench	1,160	1,034	14	315	72	6.21%
401.bzip2	211	151	19	95	24	11.37%
403.gcc	3,824	2,966	24	1,297	310	8.11%
429.mcf	39	10	7	2	0	0.00%
445.gobmk	1,082	588	17	499	25	2.31%
456.hmmer	681	376	8	255	84	12.33%
458.sjeng	235	139	15	109	2	0.85%
462.libquantum	79	44	10	41	3	3.80%
464.h264ref	1,614	193	17	161	9	0.56%
471.omnetpp	363	280	8	201	41	11.29%
473.astar	88	53	8	50	18	20.45%
483.xalancbmk	2,212	1,880	10	1,061	82	3.71%
Total	17,720	12,304	13	7,742	920	5.19%

Figura 6. Informações estáticas inferidas pela análise de não-terminação. **Laços**: número de laços no programa. **Alcançáveis**: quantidade de laços que são dependentes de dados produzidos a partir de canais de entrada. **Caminhos**: tamanho médio do menor caminho de dependência de dados da entrada até a operação de controle do laço. **Vulneráveis**: número de laços que preenchem nossos requisitos de vulnerabilidade. **I**: laços vulneráveis segundo a definição da seção 3.1. **II**: laços vulneráveis controlados por comparações do tipo `i <= N`.

Laços Alcançáveis e Vulneráveis. A figura 6 mostra a quantidade de laços alcançáveis e vulneráveis que encontramos por programa. A noção de "laço vulnerável"é definida na seção 3.1. A quantidade de laços vulneráveis que reportamos representa um limite inferior no número de estruturas de iteração que precisamos instrumentar para prevenir ataques de não terminação baseados em estouro de precisão. A figura indica que aproximadamente 70% de todos os laços do programa são alcançáveis. Dessa quantidade, a metade é vulnerável. Cerca de 12% dos laços vulneráveis são controlados por comparações do tipo `i <= N`. Esse tipo de condição é particularmente perigosa, pois, conforme visto no exemplo da figura 1, se o limite `N` for o maior inteiro possível, então a condição será sempre verdade. A figura 6 mostra que os caminhos entre as entradas de dados e as condições de parada de laços são geralmente pequenos. Por exemplo, os dez

Benchmark	Insts.	Arits.	Instrumentação		Crescimento	
			I	II	I	II
433.milc	24,971	1,101	150	0	9.49%	0.00%
444.namd	77,922	3,136	932	2	15.75%	0.04%
447.dealII	483,614	14,910	3,762	77	6.40%	0.21%
450.soplex	67,808	1,779	771	261	17.83%	6.09%
470.lbm	3,788	1,130	0	0	0.00%	0.00%
400.perlbench	288,429	7,983	655	226	2.92%	1.09%
401.bzip2	17,999	1,684	240	56	17.20%	4.61%
403.gcc	830,861	14,338	2,072	473	3.26%	0.83%
429.mcf	2,851	158	4	0	2.00%	0.00%
445.gobmk	146,298	11,856	932	76	9.73%	0.78%
456.hmmer	62,704	3,210	419	143	10.08%	3.37%
458.sjeng	25,473	2,138	167	22	9.97%	1.31%
462.libquantum	6,562	593	68	4	14.49%	0.98%
464.h264ref	141,772	13,398	411	102	3.81%	0.88%
471.omnetpp	96,929	2,029	249	46	3.91%	0.73%
473.astar	9,386	639	99	37	16.46%	6.13%
483.xalancbmk	648,941	12,528	1,562	116	2.52%	0.24%
Total	2,936,308	92,610	12,493	1,641	5.02%	0.81%

Figura 7. Impacto da instrumentação no código saneado. **Insts.**: número de instruções no benchmark. **Arits.**: número de instruções que podem causar estouro de precisão inteira. **Instrumentação**: quantidade de testes inseridos para sanear o programa. **Crescimento**: razão entre o tamanho do programa instrumentado e o tamanho do programa original. **I**: análise considerando todos os laços vulneráveis. **II**: análise considerando somente os laços controlados por condições do tipo `i <= N`.

caminhos vulneráveis em `429.mcf` possuem em média sete instruções. Conclui-se que um auditor que procure por vulnerabilidades a ataques de não-terminação tem, em geral, de analisar sequências relativamente pequenas de código.

Instrumentação. A figura 7 mostra que nossa instrumentação tem um impacto ínfimo sobre o tamanho do programa guardado. Nossos programas possuem poucas operações de tipo inteiro passíveis de estouro de precisão – em média somente 3.15% das instruções são desse tipo. Dessas instruções, um número ainda menor é usado dentro de laços vulneráveis. No caso médio, cada laço vulnerável custou-nos a criação de 1.64 guardas, como aquela vista na figura 4(e). O aumento de tamanho do programa guardado é pequeno, conforme podemos observar nas duas últimas colunas da figura 7. O maior crescimento, observado em `450.soplex`, foi de 17.83%. Na média, entretanto, os programas cresceram apenas 5% em número de instruções. Um dos benchmarks, `470.lbm` não recebeu

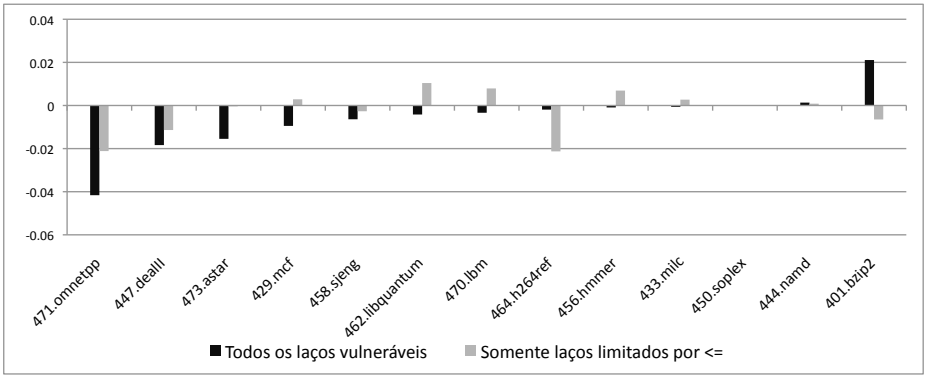

Figura 8. Variação no tempo de execução dos programas instrumentados.

qualquer instrumentação, uma vez que ele não possui laços controlados por dados de entrada. Todos os laços desse benchmark dependem de constantes criadas dentro do próprio programa.

Tempo de execução. Uma vez que o número de instruções inseridas nos programas é tão pequeno, o crescimento de seu tempo de execução é irrisório. Para justificar tal fato, alegamos que nenhuma das instruções presentes em uma guarda realiza operações demoradas, como acesso a memórias lentas. Executamos todos os programas instrumentados, passando-lhes suas entradas de referência, conforme especificado no manual de uso de SPEC CPU 2006, e as diferenças de tempo de execução são mostradas na figura 8. Cada programa foi amostrado 30 vezes, e o resultado que apresentamos é a média aritmética dessas trinta amostras. A margem de erro é negligível. Testamos dois modos de instrumentação. No primeiro deles guardamos todos os laços considerados vulneráveis contra estouros de precisão. Nesse caso, observamos que os programas instrumentados ficaram 0.61% mais lentos. No segundo modo de instrumentação guardamos somente os laços cujas condições de controle usam comparadores do tipo `<=`. Registramos nesse segundo experimento que os programas modificados ficaram 0.24% mais lentos. Caso houvéssemos instrumentado todas as operações aritméticas nos programas disponíveis em SPEC CPU 2006, a taxa de lentidão seria de 3.24%. Não mostramos esse resultado na figura, para não comprometermos a sua leitura. Em alguns programas, como `401.bzip2`, por exemplo, pudemos registrar diminuição do tempo de execução. Não encontramos razão aparente para tal comportamento, pois não efetuamos qualquer otimização sobre o código instrumentado.

5 Trabalhos Relacionados

Este trabalho aborda temas relacionados a diferentes áreas da análise estática e dinâmica de programas, a saber: teoria de fluxo de informação, detecção de

estouros de precisão inteira e análise de não-terminação. Além disso, este trabalho utiliza o conceito de *grafos de dependências*, inicialmente proposto por Ferrante *et al.* [8]. Em nosso caso, o grafo de dependência dá-nos a estrutura de dados básica sobre a qual caminhos que levam à não-terminação podem ser encontrados. Esses grafos, contudo, historicamente vêm se prestando a muitos outros propósitos, como escalonamento de instruções, detecção de condições de corrida e propagação de constantes, por exemplo.

Neste artigo usamos o grafo de dependências para rastrear o fluxo de informação contaminada. O rastreamento de fluxo de informação é uma grande sub-área dentro do campo de análise estática de programas [6]. Existem duas formas principais de rastrear a informação. Pode-se traçar o fluxo de dados a partir de operações sigilosas até entradas que um adversário pode ler. Esse modo de rastreamento é popularmente conhecido como detecção de vazamento de segredos [10]. E, no sentido inverso, pode-se traçar o fluxo de informação de entradas que um adversário pode manipular até operações críticas dentro do programa [16]. Essa categoria inclui nosso trabalho, além de diversos outros tipos de vulnerabilidades, tais como *Injeção de Código SQL* [18], *Injeção de Scripts* [14] e *Ataques de Estouro de Buffer* [12].

Nós instrumentamos código considerado vulnerável para detectar estouros de precisão que podem levar à não-terminação. Esses mesmos testes já foram usados com vários outros objetivos em trabalhos anteriores. O mais importante trabalho nessa área deve-se, provavelmente, a Brumley *et al.* [3]. O grupo de David Brumley desenvolveu uma ferramenta, RICH, que instrumenta cada operação aritmética passível de estouro de precisão inteira em um programa C. A principal conclusão daquele trabalho foi que esse tipo de instrumentação não compromete sobremaneira o desempenho do programa modificado. RICH, por exemplo, aumenta o tempo de execução dos programas instrumentados em menos que 6% em média. Outro trabalho importante nesse campo foi publicado por Dietz *et al.* [7]. Esse grupo implementou IOC, uma ferramenta que, assim como RICH, detecta a ocorrência de estouros de precisão em operações aritméticas inteiras. Porém, ao contrário de Brumley *et al.*, Dietz *et al.* usaram sua ferramenta para desenvolver um amplo estudo sobre a ocorrência de estouros em programas reais. Nosso trabalho difere desses outros em propósito: estamos interessados em prevenir ataques de não terminação; e em método: nós instrumentamos somente uma pequena parte dos programas alvo.

Finalmente, nosso trabalho relaciona-se com outros que também procuram detectar, estaticamente, a não-terminação de programas. A maior parte desses trabalhos utilizam análise simbólica de código para criar expressões que levem um laço à não terminação. Exemplos desse tipo de pesquisa incluem os trabalhos de Burnim *et al.* [4], Brockschmidt *et al* [2] e Veroyen *et al.* [17]. Esses trabalhos não levam em consideração possibilidade de não-terminação devido à estouros de precisão, tampouco procuram detectar possíveis vulnerabilidades baseadas em negação de serviço. Existem, contudo, trabalhos na linha de detecção de não-terminação que são bastante próximos do nosso.

Um trabalho que prova não-terminação, mesmo em face de estouros de precisão deve-se à Gupta *et al.* [9]. Gupta, assim como os trabalhos anteriormente relacionados, utiliza análise simbólica para provar a não-terminação de programas. A ferramenta implementada por Gupta *et al.*, denominada TNT, é capaz de encontrar uma expressão algébrica que leva um laço de programa a iterar para sempre. Porém, TNT não aponta quais laços podem ser controlados a partir da entrada do programa. Por outro lado, a ferramenta SAFERPHP, proposta por Son *et al.* [15] possui exatamente esse objetivo. SAFERPHP analisa o código de programas escritos em PHP, procurando por laços que um adversário pode controlar, com o propósito, justamente, de evitar ataques de não-terminação. A principal diferença entre nosso trabalho, e aquele de Son *et al.*, é que, enquanto nossa ferramenta busca detectar a não-terminação devido à estouros de precisão inteira, SAFERPHP considera a aritmética de precisão infinita. Além disso, tanto SAFERPHP quanto TNT utilizam execução simbólica sobre caminhos possíveis no programa alvo. Essa abordagem, em nossa opinião, não é prática. Testemunho disso é o fato de tais ferramentas terem sido usadas, até a presente data, somente para analisar programas muito pequenos.

6 Conclusão

Neste artigo nós descrevemos uma forma de ataque de negação de serviço que busca levar o programa alvo à não-terminação. Ao contrário da literatura relacionada, ate-mo-nos a ataques baseados em estouro de precisão de aritmética de inteiros. Esse tipo de fenômeno é característico de linguagens de programação como Java, C e C++. Nós definimos algumas propriedades necessárias para a efetiva realização de um ataque de não-terminação, a saber, condição de controle controlada por adversário, e por operações cíclicas passíveis de estouro de precisão. Em seguida, mostramos como eliminar a última dessas condições via guardas inseridas pelo compilador. Finalmente, mostramos experimentalmente que nossas guardas, ainda que inseridas conservadoramente, não comprometem o tempo de execução do programa instrumentado. Demonstramos assim que a prevenção de ataques de não-terminação baseados em estouros de precisão é barata e efetiva.

Neste trabalho nós adotamos uma definição muito conservadora de laços vulneráveis. De fato, muitos dos laços que indicamos como vulneráveis, em nossos experimentos, de fato não o são. Nossa decisão foi fruto de um compromisso entre a precisão e a eficiência: instrumentamos todos os laços possivelmente vulneráveis, mesmo aqueles que não são perigosos, e ainda assim mantivemos estável o tempo de execução dos programas. Por outro lado, é nossa intenção, em trabalho futuro, estreitar essa definição de laço vulnerável, a fim de fornecer a desenvolvedores uma ferramenta que lhes auxilie na descoberta de exemplos de vulnerabilidades.

Software: nossas técnicas foram todas implementadas em LLVM, e estão disponíveis publicamente na URL `http://code.google.com/p/range-analysis/`.

Referências

1. Andrew W. Appel and Jens Palsberg. *Modern Compiler Implementation in Java*. Cambridge University Press, 2nd edition, 2002.
2. Marc Brockschmidt, Thomas Ströder, Carsten Otto, and Jürgen Giesl. Automated detection of non-termination and nullpointerexceptions for java bytecode. In *FoVeOOS*, pages 123–141. Springer-Verlag, 2012.
3. David Brumley, Dawn Xiaodong Song, Tzi cker Chiueh, Rob Johnson, and Huijia Lin. RICH: Automatically protecting against integer-based vulnerabilities. In *NDSS*. USENIX, 2007.
4. Jacob Burnim, Nicholas Jalbert, Christos Stergiou, and Koushik Sen. Looper: Lightweight detection of infinite loops at runtime. In *ASE*, pages 161–169. IEEE, 2009.
5. Ron Cytron, Jeanne Ferrante, Barry K. Rosen, Mark N. Wegman, and F. Kenneth Zadeck. Efficiently computing static single assignment form and the control dependence graph. *TOPLAS*, 13(4):451–490, 1991.
6. Dorothy E. Denning and Peter J. Denning. Certification of programs for secure information flow. *Commun. ACM*, 20:504–513, 1977.
7. Will Dietz, Peng Li, John Regehr, and Vikram Adve. Understanding integer overflow in c/c++. In *ICSE*, pages 760–770. IEEE, 2012.
8. Jeanne Ferrante, Karl J. Ottenstein, and Joe D. Warren. The program dependence graph and its use in optimization. *TOPLAS*, 9(3):319–349, 1987.
9. Ashutosh Gupta, Thomas A. Henzinger, Rupak Majumdar, Andrey Rybalchenko, and Ru-Gang Xu. Proving non-termination. *SIGPLAN Not.*, 43(1):147–158, 2008.
10. C. Hammer, J. Krinke, and G. Snelting. Information flow control for java based on path conditions in dependence graphs. In *ISSSE*, pages 1–10. IEEE, 2006.
11. Chris Lattner and Vikram S. Adve. LLVM: A compilation framework for lifelong program analysis & transformation. In *CGO*, pages 75–88. IEEE, 2004.
12. Elias Levy. Smashing the stack for fun and profit. *Phrack*, 7(49), 1996.
13. David Moore, Colleen Shannon, Douglas J. Brown, Geoffrey M. Voelker, and Stefan Savage. Inferring internet denial-of-service activity. *ACM Trans. Comput. Syst.*, 24(2):115–139, 2006.
14. Andrei Alves Rimsa, Marcelo D'Amorim, and Fernando M. Q. Pereira. Efficient static checker for tainted variable attacks. In *SBLP*. SBC, 2010.
15. Sooel Son and Vitaly Shmatikov. SAFERPHP: finding semantic vulnerabilities in php applications. In *PLAS*, pages 8:1–8:13. ACM, 2011.
16. Omer Tripp, Marco Pistoia, Stephen Fink, Manu Sridharan, and Omri Weisman. TAJ: Effective taint analysis of web applications. In *PLDI*, pages 87–97. ACM, 2009.
17. Helga Velroyen and Philipp Rümmer. Non-termination checking for imperative programs. In *TAP*, pages 154–170. Springer-Verlag, 2008.
18. Gary Wassermann and Zhendong Su. Sound and precise analysis of web applications for injection vulnerabilities. In *PLDI*, pages 32–41. ACM, 2007.