

Wycieki pamięci - analiza popularnych języków programowania

Mariusz Ziętek, Szymon Pitula

29 kwietnia 2019

Spis treści

1	Wprowadzenie	3
2	Metody i narzędzia do wykrywania wycieków	3
3	Wycieki pamięci w języku C/C++	5
4	Wycieki pamięci w języku Rust	9

1 Wprowadzenie

Wycieki pamięci to problem związany z alokacją i nie zwalnianiem pamięci programu, która nie jest już potrzebna. Zjawisko to powoduje, że obszar pamięci który nie jest już używany przez program pozostaje oznaczony jako zajęty i nie może zostać ponownie wykorzystany. Problem ten jest najbardziej widoczny w aplikacjach, które działają ciągle lub przez długi okres czasu. Program taki będzie rezerwował coraz więcej pamięci, powodując znaczny spadek wydajności i w końcu zakończenie procesu. Wyciek pamięci jest stosunkowo trudny w wykryciu, ponieważ jego objawy nie są widoczne od razu po uruchomieniu programu. Problem ten jest spotykany w różnych językach programowania, w szeroko rozumianym programowaniu aplikacji. W swojej pracy chcielibyśmy przybliżyć ten problem czytelnikom, pokazać przyczyny i skutki wycieków pamięci oraz metody diagnozowania i unikania tego problemu w najpopularniejszych językach programowania: C/C++, Rust, Java, Python. Każdy z tych języków posiada inny system zarządzania pamięcią oraz wymaga innego udziału użytkownika w tym procesie.

2 Metody i narzędzia do wykrywania wycieków

Podstawowym narzędziem używanym przez nas do detekcji wycieków jest **Valgrind**. Jest to program dostarczający narzędzia do debugowania i profilowania działania programów. Najpopularniejszym narzędziem jest **Memcheck**, który będzie używany w tym projekcie najczęściej. Pozwala on na wykrycie błędów związanych z pamięcią (w tym wycieków), które mogą prowadzić do nieprzewidywalnego zachowania programu lub jego zakończenia. Istotną wadą programu jest jego wpływ na wydajność monitorowanej aplikacji. Valgrind potrafi spowolnić działanie programu nawet trzydziestokrotnie. W przypadku tego projektu nie jest to jednak duży problem, ponieważ przytoczone programy będą stosunkowo proste, aby przekaz był jak najbardziej przejrzysty.

Aby uruchomić aplikację z kontrolą Valgrind wystarczy dopisać odpowiednią flagę na przykład:

```
valgrind --leak-check=yes myprog arg1 arg2
```

Powyższe polecenie pozwoli na uruchomienie narzędzia *Memcheck* (domyślnie), wraz ze szczegółową kontrolą wycieków. Jeżeli Valgrind wykryje problem z pamięcią lub wyciek, zwróci odpowiednią wiadomość. Przykładowy raport wygląda następująco:

```
==19182== Invalid write of size 4
==19182==      at 0x804838F: f (example.c:6)
==19182==      by 0x80483AB: main (example.c:11)
==19182== Address 0x1BA45050 is 0 bytes after a block of size
==19182==      40 alloc'd
==19182==      at 0x1B8FF5CD: malloc (vg_replace_malloc.c:130)
==19182==      by 0x8048385: f (example.c:5)
==19182==      by 0x80483AB: main (example.c:11)
```

Informacje które tu widzimy to:

- 19182 to ID procesu, zwykle nie jest ważne.
- Pierwsza linia ("Invalid write ...") wskazuje na rodzaj problemu. W tym przypadku nieudana próba zapisu do pamięci.
- Poniżej znajduje się ślad stosu i nazwy funkcji mówiące o tym, gdzie wystąpił problem. Czytanie tej sekcji może być trudne w przypadku dużych programów, ułatwieniem może być czytanie od dołu.

Valgrind dzieli wycieki pamięci na kilka kategorii:

```
==6051== LEAK SUMMARY:
==6051==      definitely lost: 16 bytes in 1 blocks
==6051==      indirectly lost: 0 bytes in 0 blocks
==6051==      possibly lost: 0 bytes in 0 blocks
==6051==      still reachable: 0 bytes in 0 blocks
==6051==      suppressed: 0 bytes in 0 blocks
```

- *definitely lost* oznacza, że wyciek jest pewny i należy go naprawić.
- *indirectly lost* informuje o wycieku w strukturach bazujących na wskaźnikach. Przykład: jeżeli korzeń drzewa binarnego spowoduje *definitely lost*, wszystkie dzieci spowodują *indirectly lost*
- *possibly lost* zwykle oznacza to samo co *definitely lost*, za wyjątkiem sytuacji, gdy jest on spowodowany umyślnie. Występuje gdy w programie znajdują się wskaźniki nie wskazujące na początek bloku.
- *still reachable* występuje w momencie, gdy program nie zwolnił pamięci, ale mógłby to zrobić. Zwykle nie jest to znaczący problem.
- *suppressed* wskazuje na błędy, które zostały wyciszone. Powinno się je ignorować.

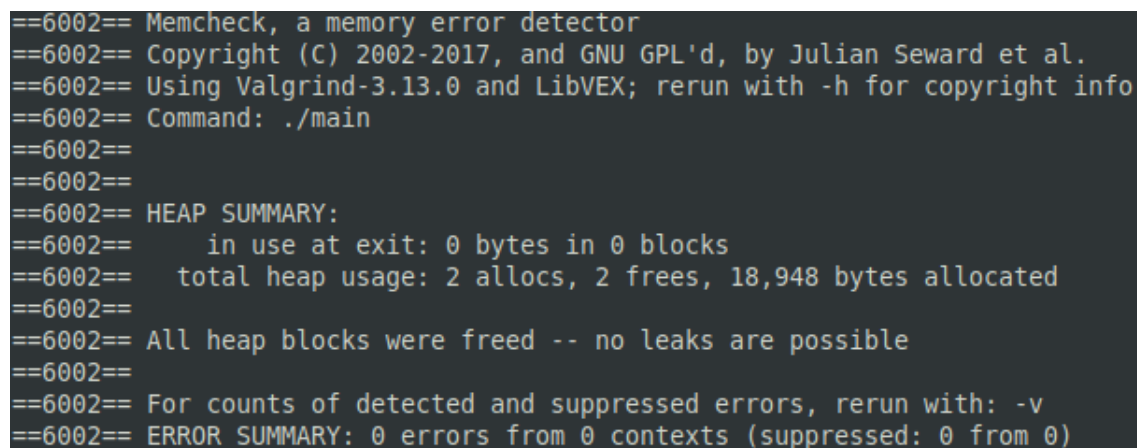
Korzystając z programy Valgrind można zatem wysledzić lokalizację błędu, określić jego przyczynę i rodzaj. Jest więc bardzo użytecznym narzędziem, które może pomóc w wyeliminowaniu błędów oprogramowania, a także w zwiększeniu wydajności. Dodatkowo Valgrind może działać na większości dostępnych platform i wspiera każdy język programowania. Narzędzie to powinno zatem być używane możliwie jak najczęściej (ograniczenia spowodowane są czasem działania programu), aby tworzone oprogramowania było jak najlepszej jakości.

3 Wycieki pamięci w języku C/C++

Konstrukcja tego języka pozwala na wykonanie wycieków pamięci na wiele sposobów. Wybraliśmy te z nich, które w możliwie czytelny sposób pokażą istotę problemu i takie, które są często popełniane. Poniżej przedstawione zostały wycieki związane z alokacją pamięci i zwalnianiem nieużywanego obszaru w przypadku korzystania ze wskaźników. Do alokacji i zwalniania pamięci w języku C++ służą operatory *new* oraz *delete* (można także korzystać z operatorów języka C, np. - *alloc*, *malloc*, *free*. W języku tym odpowiedzialność za zarządzanie pamięcią leży po stronie programisty. Za pomocą *new* rezerwuje się pamięć potrzebną do przechowania obiektu, a gdy ta nie jest już potrzebna należy ją zwolnić. Jeżeli programista nie użyje operatora *delete* do zwolnienia pamięci, dojdzie do wycieku. Można to zaobserwować już w bardzo prostych programach. Poniższy przykład pokazuje różnicę w wykorzystaniu przez program pamięci w przypadku, gdy zostaje ona poprawnie zwolniona i w przypadku, gdy do zwolnienia nie dojdzie. Pierwszy fragment kodu obrazuje poprawne zarządzanie pamięcią:

```
void problematic_function();
int main() {
    problematic_function();
    return 0;
}
void problematic_function(){
    int *data = new int;
    *data = 15;
    delete data;
}
```

Pamięć została zwolniona poprawnie, zatem nie powinien wystąpić wyciek. Potwierdzeniem tej tezy są wyniki programu Valgrind:



```
==6002== Memcheck, a memory error detector
==6002== Copyright (C) 2002-2017, and GNU GPL'd, by Julian Seward et al.
==6002== Using Valgrind-3.13.0 and LibVEX; rerun with -h for copyright info
==6002== Command: ./main
==6002==
==6002==
==6002== HEAP SUMMARY:
==6002==     in use at exit: 0 bytes in 0 blocks
==6002==   total heap usage: 2 allocs, 2 frees, 18,948 bytes allocated
==6002==
==6002== All heap blocks were freed -- no leaks are possible
==6002==
==6002== For counts of detected and suppressed errors, rerun with: -v
==6002== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0)
```

Raport Valgrind mówi o tym, że zaalokowano i zwolniono dwie struktury, zatem wyciek nie jest możliwy. Całkowita użyta przez program pamięć to 18,948 bajtów.

W kolejnym kroku z kodu programu zostanie usunięta linia odpowiedzialna za zwolnienie pamięci:

```
void problematic_function();
int main() {
    problematic_function();
    return 0;
}
void problematic_function(){
    int *data = new int;
    *data = 15;
}
```

Zmieniony program zostaje przetestowany przez Valgrind i wyniki są następujące:

```
==6055== Memcheck, a memory error detector
==6055== Copyright (C) 2002-2017, and GNU GPL'd, by Julian Seward et al.
==6055== Using Valgrind-3.13.0 and LibVEX; rerun with -h for copyright info
==6055== Command: ./main
==6055==
==6055==
==6055== HEAP SUMMARY:
==6055==   in use at exit: 4 bytes in 1 blocks
==6055==   total heap usage: 2 allocs, 1 frees, 18,948 bytes allocated
==6055==
==6055== 4 bytes in 1 blocks are definitely lost in loss record 1 of 1
==6055==    at 0x483087B: operator new(unsigned int) (in /usr/lib/valgrind/vgpreload_memcheck-x86-linux.so)
==6055==    by 0x108657: problematic_function() (in /home/dsk/Desktop/OiAK-Proj/CppTablica/main)
==6055==    by 0x10862C: main (in /home/dsk/Desktop/OiAK-Proj/CppTablica/main)
==6055==
==6055== LEAK SUMMARY:
==6055==   definitely lost: 4 bytes in 1 blocks
==6055==   indirectly lost: 0 bytes in 0 blocks
==6055==   possibly lost: 0 bytes in 0 blocks
==6055==   still reachable: 0 bytes in 0 blocks
==6055==   suppressed: 0 bytes in 0 blocks
==6055==
==6055== For counts of detected and suppressed errors, rerun with: -v
==6055== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```

Tym razem Valgrind informuje o wycieku. Według raportu po zakończeniu programy w użyciu zostały 4 bajty, co zgadza się z rozmiarem niezwolnionej liczby całkowitej. Dodatkowo mamy też informację, że wyciek jest definitywny, czyli bezwzględnie wystąpił. Valgrind informuje także o miejscu wystąpienia błędu. Korzystając z tej informacji można szybko zlokalizować błąd. Z informacji wynika, że przyczyną jest operator *new* w funkcji *problematicfunction()* wywołanej w *main()*. Informacja ta jest oczywiście prawdziwa, co łatwo zauważyć analizując kod programu.

Kolejny przykład prezentuje często popełniany błąd, który łatwo popełnić.

```
struct problematicStructure{
    int indexNumber;
    float *grades = new float[1000];
};

int main() {

    problematicStructure *s1 = new problematicStructure();
    delete s1;
    return 0;
}
```

Mogłoby się wydawać, że program jest wykonany poprawnie. Struktura została zaalokowana i zwolniona. Valgrind jednak pokaże, że istnieje wyciek pamięci.

```
==6518== Memcheck, a memory error detector
==6518== Copyright (C) 2002-2017, and GNU GPL'd, by Julian Seward et al.
==6518== Using Valgrind-3.13.0 and LibVEX; rerun with -h for copyright info
==6518== Command: ./main
==6518==
==6518==
==6518== HEAP SUMMARY:
==6518==   in use at exit: 4,000 bytes in 1 blocks
==6518==   total heap usage: 3 allocs, 2 frees, 22,952 bytes allocated
==6518==
==6518== 4,000 bytes in 1 blocks are definitely lost in loss record 1 of 1
==6518==    at 0x4830F6B: operator new[](unsigned int) (in /usr/lib/valgrind/vgpreload_memcheck-x86-linux.so)
==6518==    by 0x10887B: problematicStructure::problematicStructure() (in /home/dsk/Desktop/OiAK-Proj/CppTablica/main)
==6518==    by 0x108792: main (in /home/dsk/Desktop/OiAK-Proj/CppTablica/main)
==6518==
==6518== LEAK SUMMARY:
==6518==   definitely lost: 4,000 bytes in 1 blocks
==6518==   indirectly lost: 0 bytes in 0 blocks
==6518==   possibly lost: 0 bytes in 0 blocks
==6518==   still reachable: 0 bytes in 0 blocks
==6518==   suppressed: 0 bytes in 0 blocks
==6518==
==6518== For counts of detected and suppressed errors, rerun with: -v
==6518== ERROR SUMMARY: 1 errors from 1 contexts (suppressed: 0 from 0)
```

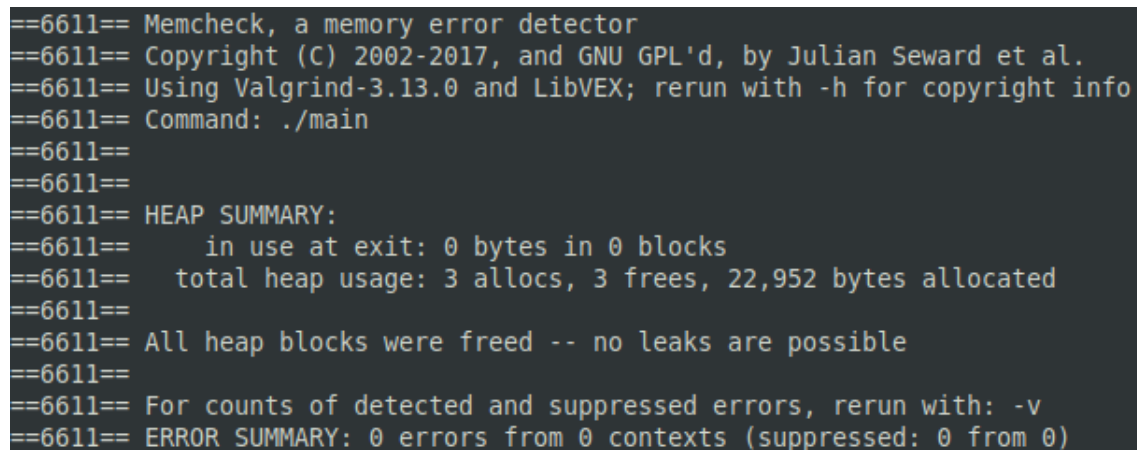
Istotą problemu jest to, że należy usunąć **całą** zarezerwowaną pamięć. Zdarza się, że programista usunie strukturę bądź klasę której już nie potrzebuje, ale zapomni o jej wewnętrznych zasobach. Valgrind informuje o definitywnym wycieku 4000 bajtów, co odpowiada 1000 liczb typu float. W przypadku usunięcia tablicy, ale nie zwolnienia pamięci zajmowanej przez strukturę, Valgrind Również znajdzie wyciek. W tym przypadku będzie to jednak tylko 8 bajtów, odpowiadających rozmiarowi struktury.

Aby całkowicie wyeliminować wyciek pamięci w programie należy zwolnić zarówno strukturę jak i tablicę w niej zawartą, pamiętając o odpowiedniej kolejności:

```
struct problematicStructure{
    int indexNumber;
    float *grades = new float[1000];
};

int main() {

    problematicStructure *s1 = new problematicStructure();
    delete[] s1->grades;
    delete s1;
    return 0;
}
```



```
==6611== Memcheck, a memory error detector
==6611== Copyright (C) 2002-2017, and GNU GPL'd, by Julian Seward et al.
==6611== Using Valgrind-3.13.0 and LibVEX; rerun with -h for copyright info
==6611== Command: ./main
==6611==
==6611==
==6611== HEAP SUMMARY:
==6611==   in use at exit: 0 bytes in 0 blocks
==6611==   total heap usage: 3 allocs, 3 frees, 22,952 bytes allocated
==6611==
==6611== All heap blocks were freed -- no leaks are possible
==6611==
==6611== For counts of detected and suppressed errors, rerun with: -v
==6611== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0)
```

Powyższe przykłady obrazują, że w języku C oraz C++ do wycieku pamięci można doprowadzić w bardzo prosty sposób. Odpowiedzialność za zarządzanie pamięcią spoczywa na programiście, zatem unikanie wycieków może być trudne. Dodatkowo pokazane przykłady obejmują wąski zakres problemu, który występuje w tym języku. Szczególnie więc warto korzystać z narzędzi takich jak Valgrind, Address Sanitizer, czy menadżer zasobów Visual Studio.

4 Wycieki pamięci w języku Rust

Rust różni się od innych języków programowania sposobem zarządzania pamięcią. Sprawdzanie poprawności kodu pod kątem błędnego użycia pamięci odbywa się już na etapie kompilacji, a nie po uruchomieniu aplikacji. Rust nie posiada Garbage Collectora, ale wprowadza mechanizm zwany **Ownership**. Własność ta tak jak w przypadku C++ zmusza programistę do ręcznego zarządzania pamięcią, jednak istotną różnicą jest to, że program napisany w Rust nie skompiluje się, jeżeli wykryty zostanie błąd. Dodatkowo Ownership ułatwia zarządzanie pamięcią, pozwalając na tylko jednego właściciela obiektu, a także poprzez automatyczne zwalnianie zasobów. Oznacza to, że deklarując zmienną, przypisywany jest do niej Ownership, czyli zmienna staje się właścicielem pewnego fragmentu pamięci. Pamięć ta zostanie automatycznie zwolniona w momencie wyjścia poza zakres użycia zmiennej (np. po wyjściu z funkcji). Rust wydaje się zatem być językiem znacznie bardziej bezpiecznym niż C++. Czy to oznacza, że wyciek pamięci jest niemożliwy? Tylko w teorii. Dokumentacja Rust mówi o tym, że język stara się kontrolować wycieki, ale nie jest to całkowicie możliwe, poprzez nieprzewidywalną naturę tego zjawiska.

W języku Rust najłatwiej zobrazować wyciek pamięci wywołując go specjalną funkcją `mem::forget`. Powoduje ona, że Rust usunie zmienną, ale nie uruchomi jej destruktoru.