Liljestrand, Hans; Gauhar, Zaheer; Nyman, Thomas; Ekberg, Jan-Erik; Asokan, N.

Protecting the stack with PACed canaries

Published in: SysTEX ’19, October 27, 2019, Huntsville, ON, Canada

DOI: 10.1145/3342559.3365336

Published: 27/10/2019

Document Version
Peer reviewed version

Published under the following license: Unspecified

Please cite the original version:
Protecting the stack with PACed canaries

Hans Liljestrand
Aalto University, Finland
Huawei Technologies Oy, Finland
hans@liljestrand.dev

Zaheer Gauhar
Aalto University, Finland
zaheer.gauhar@aalto.fi

Thomas Nyman
Aalto University, Finland
thomas.nyman@aalto.fi

Jan-Erik Ekberg
Huawei Technologies Oy, Finland
Aalto University, Finland
jan.erik.ekberg@huawei.com

N. Asokan
University of Waterloo, Canada
asokan@acm.org

Abstract
Stack canaries remain a widely deployed defense against memory corruption attacks. Despite their practical usefulness, canaries are vulnerable to memory disclosure and brute-forcing attacks. We propose PCan, a new approach based on ARMv8.3-A pointer authentication (PA), that uses dynamically-generated canaries to mitigate these weaknesses and show that it provides more fine-grained protection with minimal performance overhead.

1 Introduction
Run-time attacks that exploit memory errors to corrupt program memory are a prevalent threat. Overflows of buffers allocated on the stack are one of the oldest known attack vectors [12, 17]. Such exploits corrupt local variables or function return addresses. Modern attacks techniques—such as return-oriented programming (ROP) [16] and data-oriented programming (DOP)[6]—can use this well-known attack vector to enable attacks which are both expressive and increasingly hard to detect. The fundamental problem is insufficient bounds checking in memory-unsafe languages such as C / C++. Approaches for hardening memory-unsafe programs have been proposed, but tend to incur high performance overheads, and are therefore impractical to deploy [19]. An exception is a technique called stack canaries [2], which is both efficient and can detect large classes of attacks. Consequently stack canaries are widely supported by compilers and used in all major operating systems today [2, 9].

Widely deployed stack canary implementations suffer from one or more of the following weaknesses: they 1) rely on canary values that are fixed for a given run of a program [2]; 2) store the reference canary in insecure memory, where an attacker can read or overwrite it [9]; or 3) use only a single canary per stack frame and therefore cannot detect overflows that corrupt only local variables.

The recently introduced ARMv8.3-A pointer authentication (PA) [1] hardware can be used to verify return addresses [14], effectively turning the return address itself into a stack canary. However, PA on its own is susceptible to reuse attacks, where an attacker substitutes one authenticated pointer with another [8]. State-of-the-art schemes harden PA return-address protection to ensure that protected return address as statistically unique to a particular control-flow path, and therefore cannot be substituted by an attacker [7].

We propose fine-grained PA-based canaries that: 1) protect individual variables from buffer overflow, 2) do not require secure storage for reference canaries, 3) can use existing return-address protection [7] as an anchor to produce canaries which are statistically unique to a particular function call, and 4) are efficient, since they can leverage hardware PA instructions both for canary generation and verification.

Our contributions are:
• PCan: A fine-grained, and efficient PA-based canary scheme (Section 5).
• A realization of PCan for LLVM (Section 6).
• Evaluation showing that PCan is more secure than existing stack canaries and has only a small performance impact (Section 7).

2 Background
A stack canary[2] is a value placed on the stack such that a buffer overflow will overwrite it before corrupting the return address (Figure 1). The buffer overflow can be detected by verifying the integrity of the canary before return.

The initially proposed canaries were randomly generated 32-bit values assigned at process startup and stored within the process memory [2]. The canaries must remain confidential to prevent an attacker A from avoiding detection by writing the correct canary when triggering the buffer overflow. Terminator canaries [3], consisting of string terminator values (e.g., 0x0, EOF, and 0xFF) can prevent A from using string operations to read or write whole canaries, thereby thwarting run-time canary harvesting. Another approach is to re-generate canaries at run-time, for instance by masking them with the return address [4]. However, such techniques rely on the secrecy of the masking value.
Multi-threaded and forked processes can be vulnerable to guessing attacks if the canaries are shared. If the child process or thread is restarted after a crash, \( A \) can execute a large number of guesses without resetting the canary. Moreover, such attacks can utilize incremental—e.g., byte-for-byte—guessing to efficiently find the canary value \([10]\).

Strong adversaries with arbitrary memory read or write access can trivially circumvent any canary based solution; using reads alone allows \( A \) to first read the correct canaries from memory and then perform a sequential overwrite that writes back the correct canaries while corrupting other data.

2.1 Stack canaries in modern compilers

Modern compilers such as LLVM/Clang and GCC provide the \(-fstack-protector\) feature that can detect stack-buffer overflows\(^1\). It is primarily designed to detect stack overflows that occur in string manipulation. The default \(-fstack-protector\) option includes a canary only when a function defines a character array that is larger than a particular threshold. The default threshold value in GCC and LLVM is 8 characters, but in practice the threshold is often lowered to 4 to provide better coverage. However, a stack overflow can occur on other types of variables. The \(-fstack-protector-all\) option adds a canary to all functions. However, it can incur a substantial use of stack space and run-time overhead in complex programs.

The \(-fstack-protector-strong\) option provides a better trade-off between function coverage, run-time performance, and memory cost of stack canaries. It adds a canary to any function that 1) uses a local variable’s address as part of the right-hand side of an assignment or function argument, 2) includes a local variable that is an array, regardless of the array type or length, and 3) uses register-local variables.

Today \(-fstack-protector-strong\) is enabled by default for user-space applications in major Linux distributions, such as Debian and its derivatives\(^2\). The \(-fstack-protector\) protects non-overflowing variables by rearranging the stack

---

\(^1\)https://lists llvm.org/pipermail/cfe-dev/2017-April/053662.html

\(^2\)https://wiki.debian.org/Hardening
A stack-buffer overflow must always corrupt a canary. Reference canaries must not be modifiable by $\mathcal{A}$. Each canary value should be statistically unique.

3 Adversary Model

In this work we consider an adversary $\mathcal{A}$ that attempts to compromise the memory safety of a user-space process by exploiting a stack-buffer overflow. We do not consider adversaries at kernel or higher privilege levels. However, $\mathcal{A}$ can:

1) trigger any existing stack-buffer overflow, 2) use stack-buffer over-reads to read memory and 3) repeatedly restart the process and any child processes or threads in an attempt to brute force canaries. Adversaries with arbitrary memory read or write access cannot be thwarted with canary-based approaches and are beyond the scope of this work.

We assume that $\mathcal{A}$ can analyze the target binary and therefore knows the exact stack layout of functions (barring variable-length buffers). This enables $\mathcal{A}$ to target individual local variables reachable from a particular buffer overflow without overflowing canaries past the local variables (Figure 4). If $\mathcal{A}$ further manages to exploit a buffer over-read or other memory disclosure vulnerability, they could also overflow past the canary by simply replacing the correct canary value during the overflow.

4 Requirements

To detect linear buffer overflows we require a design that fulfills the following requirements:

- **R1** Each canary value should be statistically unique.
- **R2** Reference canaries must not be modifiable by $\mathcal{A}$.
- **R3** A stack-buffer overflow must always corrupt a canary.

5 Design

We propose PCan, a PA-based canary design that employs multiple function-specific canaries. By placing canaries after any array that could overflow (Figure 5), PCan can detect overflows that only corrupt local variables. This prevents $\mathcal{A}$ from performing precise overflows that corrupt only local variables without overwriting the canary (R3). To exploit an overflow without detection $\mathcal{A}$ is instead forced to learn the correct canary values and write them back into place.

In contrast to traditional approaches, PCan avoids exposing reference canaries in memory (R2). Instead, canaries are re-generated or verified directly using PA. $\mathcal{A}$ thus cannot manipulate the reference canaries, and must instead leak the specific on-stack canary or attempt a brute-force attack.

The canaries are generated with PA, using a modifier $m$ consisting of a 16-bit function identifier and the least-significant 48 bits from SP: $m = SP + 2^{16} + \text{function-id}$. This modifier makes canaries function-dependent and, when SP differs, different for each call to the same function (R1). Because the canaries are generated at run-time and the PA keys are randomly set on each execution, the generated canaries are also statistically unique for each execution. To avoid detection $\mathcal{A}$ must acquire the exact stack-canary belonging to the specific function call and cannot rely on pre-calculated canaries or those belonging to other function calls.

5.1 PA-based canaries

PA-based return-address protection [7, 8, 14] already effectively serves as a canary by detecting return-address corruption. We therefore propose a design that can be efficiently and easily integrated with existing return-address protection schemes, but also provide a stand-alone setup. The first canary in a function’s stack frame, protecting the return-address, is either a pacga-generated stand-alone canary or the signed return address:

$$C_0 = \begin{cases} \text{pacga}(SP, m) & \text{if stand-alone} \\ \text{signed_return_address} & \text{if combined} \end{cases}$$
We denote a canary loaded from the stack with $C'$ to indicate that it might have been corrupted by $F$. Verification of $C'_0$ is done either by re-generating the stand-alone pacga canary or by relying on the return-address protection to verify it. To verify using pacga we re-generate $C_0$ and check that $C_0 = C'_0$.

Subsequent canaries, $C_i$, $i > 0$, consist of signed pointers to the previous canary:

$$C_i = \text{pacda}(Cptr_{i-1}, m), i > 0$$

where $Cptr_i$ is a pointer to $C_i$. Verification of $C_i$, $i > 0$ is done by authenticating and loading the canary to retrieve $C'_{i-1}$. If any $C'_i$ is corrupted, authentication fails, causing the subsequent load to fault (Section 2.2). A successful chain of loads will yield $C_0$, which is then verified as detailed above.

Our stand-alone scheme is more powerful than --msign-return-address in that it does not rely solely on the SP value. However, other schemes might provide better protection for the return address. For instance, PACStack [7] proposes a scheme that uses statistically unique modifiers to protect return addresses by maintaining the head of a chain of PACs in a single reserved register. We propose that PCan could be combined with such a mechanism by defining $m$ as the PACStack authentication token $auth_i$ and $C_0$ as the PACStack protected return address. Because the $m$ in this case would be statistically unique to a specific call-flow this would also harden the canaries $C_i$ for $i > 0$.

6 Implementation

We implement PCan as an extension to LLVM 8.0 and using the stand-alone pacga approach (Section 5.1). To instrument the LLVM Intermediate Representation (IR) we added new LLVM intrinsics for generating and verifying PCan canaries. These intrinsics, along with instructions for storing and loading the canaries, are added through IR transformations before entering the target-specific compiler backend. We define corresponding target-specific intrinsics to leverage built-in register allocation before converting the intrinsics to hardware instructions in the pre-emit stage.

6.1 Canary creation

To instrument the function prologue PCan locates LLVM alloca instructions that allocate buffers in the entry basic block of each function. A new 64-bit allocation for the canaries is added after each existing alloca. Intrinsics for generating the canaries and storing them are then added. The instrumented code will generate a larger stack-frame to accommodate the canaries and include code to generate and store the canary values (Listing 1).

6.2 Canary verification

To verify canaries in the function epilogue, PCan loads them in reverse order, starting from the last $C_n$ (Listing 2). Each canary $C'_i$ is authenticated using autda and then dereferenced to acquire the next canary $C'_{i-1}$. To verify the final canary, $C'_0$, PCan first re-generates $C_0$ and then performs a value comparison. Upon failure, an error handler is invoked, otherwise the function is allowed to return normally. As suggested in Section 5.1, the final canary can be replaced with a return-address protection scheme. The return address then serves as a canary that is verified using the corresponding protection scheme (e.g., --msign-return-address).

7 Evaluation

Due to lack of publicly available PA-capable hardware we have used an evaluation approach similar to prior work [7, 8]. We used the ARMv8-A Base Platform Fixed Virtual Platform (FVP), based on Fast Models 11.5, which supports ARMv8.3-A for functional evaluation. For performance evaluation we used the PA-analogue from prior work [7] and performed measurements on a 96board Kirin 620 HiKey (LeMaker version) with an ARMv8-A Cortex A53 Octa-core CPU (1.2GHz) / 2GB LPDDR3 SDRAM (800MHz) / 8GB eMMC, running the Linux kernel v4.18.0 and BusyBox v1.29.2.

7.1 Performance

We evaluated the performance of PCan using the SPEC CPU 2017\textsuperscript{3} benchmark, and running it on the HiKey board. We cross-compiled the benchmarks on an x86 system using whole program LLVM\textsuperscript{4}, and timed the execution of the individual benchmark programs using the time utility. Results are reported normalized

\begin{lstlisting}[language=C, label=listing1] 1 canary-creation: 2   mov   x8, sp 3   movk  x8, #3, lsl #48 ; x8 ← mod 4   pacga x10, sp, x8 ; x10 ← C_0 5   sub   x9, x29, #0x8 ; x9 ← Cptr_0 6   pacda x9, x10 ; x9 ← C_1 7   str   x9, [sp, #40] ; store C_i 8   stur x10, [x29, #-8] ; store C_0  
\end{lstlisting}

\begin{lstlisting}[language=C, label=listing2] 1 canary-verification: 2   ldr   x8, [sp, #40] ; x8 ← C'_i 3   mov   x8, sp 4   movk  x8, #3, lsl #48 ; x8 ← mod 5   autda x8, x29 ; authenticate C'_{i-1} 6   ldr   x8, [x8] ; x8 ← C'_0 7   pacga x9, sp, x9 ; x9 ← C_0 8   cmp   x8, x9 ; check C_0 = C'_0  
\end{lstlisting}

\textsuperscript{3}https://www.spec.org/cpu2017/

\textsuperscript{4}https://github.com/travitch/whole-program-llvm
to a baseline measured without PCan instrumentation and compiled with `-fno-stack-protector` (Table 1). We compare this baseline to two different setups; one using only `-fstack-protector-strong` and another using `-fno-stack-protector` and PCan instrumentation. Our results indicate that PCan incurs a very low overhead with a geometric mean of 0.30%. In some cases `-fstack-protector-strong` caused the benchmarks, we suspect this is caused by it rearranging the stack. Measurements were repeated 20 times and all binaries were compiled with `-O2` optimizations enabled.

### 7.2 Security

The initial pacga canaries used by PCan provide similar security to traditional canaries. To perform an overflow while avoiding detection \( \mathcal{A} \) must achieve the following goals: 1) find the location of canaries in relation to the overflowed buffer, 2) leak the specific canary values on the stack, and 3) write back the correct canaries when performing the buffer overflow. In our adversary model step 1) is trivial; \( \mathcal{A} \) can inspect the binary to analyze the stack layout. Step 2) could be achieved by leaking or modifying the in-memory reference values, but because PCan generates canaries on-demand, \( \mathcal{A} \) is forced to leak the values from the stack (R2). Moreover, because the canaries are statistically unique to a function and SP value \( \mathcal{A} \) cannot rely on finding just any canary and substitute it with one in the overflowed stack frame (R1). This limits the scope of attacks, as both the memory leak and overflow must happen within the lifetime of the attacked stack frame. By using multiple canaries—one after each buffer—PCan can detect overflows that only touch local variables (R3). Based on our evaluation PCan thus provides comprehensive protection with an overhead similar to currently deployed defenses.

### 8 Related Work

After the seminal article “Smashing the Stack for fun and profit” [12], the notion of canaries as a protection against buffer overflow was first introduced in StackGuard [2], and initial GCC compiler support appeared at the same time. StackGuard proposes to use a random canary, stored at the top of the stack (or in the thread local storage memory area), during program launch to thwart canary harvesting from the compiled code. The threat of canary harvesting and the added protection (especially for C) provided by terminator canaries was identified shortly thereafter [3]. The problem of canary copy and re-use was already identified by Etoh and Yoda in 2000 [4], where the stack-frame based canary protection was augmented by masking the canary value with the function return address. Later, Stackx et al. [18] argue against the futility of storing secrets in program memory, which supports using PA to generate canaries dynamically.

Another shortcoming of canary integrity are cases when the canary mechanism is subject to brute-force attacks, e.g., in the context of process forking. \( \mathcal{A} \) could use the canaries in forked child processes as oracles to perform brute-force guessing of canary values. Published solutions against this form of attack includes DynaGuard [13] and DCR [5]. Both solutions keep track of canary positions in the code, and re-initialize all canaries in a child process, at considerable performance overhead. DCR optimizes the canary location tracking by chaining canaries using embedded offsets - we inherit this notion of chaining canaries from their work, although we deploy these for canary validation whereas DCR uses the mechanism for canary rewriting. By combining the SP in the canaries PCan mitigates such attacks, but full protection would would require a similar approach of re-initializing canaries on fork. Finally, the polymorphic canaries by Wang et al. [20] optimize away the need to rewrite canaries during fork, by adding a function-specific random mask to the stack canary, which effectively removes the opportunity for systematic canary brute-forcing.

Heap protection with canaries has received much less attention than stack protection, possibly because the optimal balance between validation and performance overhead when canaries are applied to the heap remains an open problem. The first paper on the subject was Robertson et al. in 2003 [15], but a more recent mechanism — HeapSentry by Nikiforakis et al. [11] puts effort on the unpredictability (randomness) of the heap canaries. HeapSentry consists of a wrapper for the allocator and a kernel module, and exhibits overheads at around 12%. Pointer bounds checking schemes offer protections stronger than canaries alone, but in comparison incur significant performance overheads [19].

### 9 Future Work

Our current approach only protects stack-based variables with a static size. Canaries for dynamic allocations cannot
be verified in the prologue because they might be either out of scope or overwritten by later dynamic allocations, and are currently not used by PCan. To prevent attacks that corrupt dynamic allocations, we propose to add instrumentation that protects dynamic allocations based on their life-time, i.e., which verifies the associated canaries immediately when the allocation goes out of scope. The existing LLVM allocation life-time tracking could be leveraged to implement this addition without significant changes to the compiler.

We plan to refine and expand our Canary approach by using compile-time analysis—i.e., the StackSafetyAnalysis of LLVM—to omit instrumentation of buffers that can be statically shown to be safe. In some cases, however, could achieve their goal before function return, i.e., before the Canary corruption is detected. Such attacks could be detected earlier by utilizing the StackSafetyAnalysis to add checks after vulnerable steps during function execution, before the function epilogue.

We also plan to extend PCan instrumentation to cover heap allocations, similar to HeapSentry. Because the PA-keys are managed by the kernel, PCan could be used for HeapSentry-like consistency checks from within the kernel, e.g., before executing system-calls.

10 Conclusion

Canaries are a well-established protection mechanism against errors in memory-unsafe programs. We present PCan, which provides hardware-assisted integrity-protection for canaries, inhibiting the most prevalent canary-circumvention techniques. Furthermore, we propose the notion of fine-grained canaries, where canaries not only protect the return address, but also detect overflows in individual data objects. We make available our compiler prototype at https://github.com/pointer-authentication/PCan-llvm, and measure its performance impact. Finally we point out further optimizations for fine-grained canaries, as well a solution path for protecting dynamic allocations.

11 Acknowledgments

This work was supported in part by the Intel Collaborative Research Institute for Collaborative Autonomous & Resilient Systems (ICRI-CARS), the Academy of Finland under grant nr. 309994 (SELIoT), and a Google ASPIRE award.

References
