

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Tomasz Kępa

Student no. 359746

Detecting Anti-Adblockers using Differential Execution Analysis

Master's thesis
in COMPUTER SCIENCE

Supervisor:
dr Konrad Durnoga
Institute of Informatics

August 2019

Supervisor's statement

Hereby I confirm that the presented thesis was prepared under my supervision and that it fulfils the requirements for the degree of Master of Computer Science.

Date

Supervisor's signature

Author's statement

Hereby I declare that the presented thesis was prepared by me and none of its contents was obtained by means that are against the law.

The thesis has never before been a subject of any procedure of obtaining an academic degree.

Moreover, I declare that the present version of the thesis is identical to the attached electronic version.

Date

Author's signature

Abstract

Ads are the main source of income of numerous websites. However, some of them are fairly annoying which causes many users to use adblocking browser extensions. Some services, in turn, use specialized scripts to detect such plug-ins and silently report them or block some content as a punishment. The goal of this thesis is to build a pipeline for detecting such scripts based on a differential execution analysis, a method provided by other authors in 2018. Such a mechanism can be used later to analyze the prevalence of anti-adblockers on Polish websites or to build an extension capable of circumventing such scripts.

Keywords

dynamic analysis, differential execution analysis, javascript, anti-adblockers, ads

Thesis domain (Socrates-Erasmus subject area codes)

11.3 Informatics, Computer Science

Subject classification

Software and its engineering. Dynamic analysis

Tytuł pracy w języku polskim

Wykrywanie skryptów blokujących rozszerzenia typu AdBlock w przeglądarkach

Contents

Introduction	5
1. Basic concepts	7
1.1. Definitions	7
1.2. Adblockers	7
1.3. Anti-adblockers	8
1.4. Differential Execution Analysis	9
1.5. Detecting anti-adblockers	9
1.6. JavaScript execution model	10
2. Trace collection	11
2.1. Methods overview	11
2.2. Dynamic in-JavaScript code injection	11
2.3. Static code injection	12
2.3.1. Web Tracing Framework	12
2.3.2. Iroh	12
2.4. Engine instrumentation	13
3. Trace collection by V8 instrumentation	15
3.1. V8 architecture	15
3.1.1. JS bytecode	16
3.1.2. JS built-in functions	17
3.2. V8 usage in Chromium	17
3.3. Chrome's extensions architecture	17
3.4. V8's <i>--trace</i> flag	18
3.5. Bytecode injection	19
3.6. Controlling Chrome programatically	23
4. Trace analysis	25
4.1. Parsing	25
4.2. Trace untangling	27
4.3. Trace alignment	30
4.4. Trace matching using Stable Marriage Problem	31
4.5. Artificial examples	32
4.6. Noise filtering	35

5. Evaluation	37
5.1. The pipeline	37
5.2. Methodology	37
5.3. Negative examples	38
5.4. Positive examples	40
5.4.1. Websites with unsatisfactory results	42
5.4.2. Detected anti-adblockers	46
5.5. Small scale study	49
5.6. Conclusion	49
Listings	51
List of Figures	53
List of Tables	55
Bibliography	57
Appendix A	61

Introduction

In modern-era Internet most webpages operate for profit. Most of them, however, choose to provide free content in exchange for displaying paid ads. Unfortunately, not all websites play fair. Some of them concentrate on displaying as many ads as possible and generate traffic by using click-baits and other shady practices. Even websites with valuable content can have overwhelming amount of ads. This leads to grave dissatisfaction of some portion of users. To make their browsing experience better, they turn to use of ad-blocking extensions.

The most popular adblockers operate based on user-curated lists. They have been around for such a long time that they are really effective. In order to circumvent them, one has to put an ongoing effort to be able to catch up. It is such a common occurrence that even some browsers have built-in ad-blocking functionality (e.g. Opera).

The amount of users utilizing ad-blockers cannot be ignored. In 2019 there was over 615 million devices worldwide with adblocker installed [33]. This, in turn, leads to loss of revenue for many businesses. To combat this, some of them choose to deploy anti-adblockers, i.e. scripts that detect adblocking extensions and react in some way. They can generate a visible warnings or even even block the website's content entirely.

The anti-adblockers come in many different fashions. Some of them are simple, custom-made scripts that set up an obvious bait and check if there was a reaction typical to adblocking extension (e.g. a file names *ads.js* has not been loaded). There are also more effective solutions that use a number of tricks to make the analysis and counteraction harder.

Some ad providers and publishers recognize users' discontent and start initiatives like Acceptable Ads [1]. The idea is to whitelist ads that meet certain criteria (mostly non-intrusiveness). Others care less about users and just improve the anti-adblocking solutions to make sure that ads continue to bring profit.

The people behind adblockers have mixed approaches to anti-adblocking warnings. Some of them choose to let all of them be, others only if they are dismissible (and thus less effective from the point of view of the site owners). Naturally, there are also other players who develop extensions which block all such warnings, regardless of their intrusiveness. This leads to an arms race.

A regular study of anti-adblocking scripts can help both sides of the barricade. Knowing how such scripts behave can help create better methods of detecting them. That, in turn, can lead to creation of better blocking tools. On the other hand, being able to automatically detect some scripts usually means that it is also possible to block them. As a consequence, studying different solution and detection methods can lead to better anti-adblockers as well.

In 2018 Zhu et al. [46] proposed a sophisticated method of automatically detecting such scripts. This method uses what is called a Differentiation Execution Analysis. The whole approach consists of several steps. First, at least two JavaScript execution traces have to be collected. One being the website's code executed in an environment without any adblocking extensions and second with such plugin active. Those two traces are later compared and checked if there are any differences between them. If there are, they can be attributed to

anti-adblocking scripts.

There are lots of difficulties in getting the whole mechanism to work. First of all, trace collection has to be hand-made. There is no off-the-shelf solution that would meet the requirements of this method. Second, JavaScript event-based execution model makes it problematic to compare execution traces as the order and number of events can greatly differ. This issue is solved by processes called trace slicing (or untangling) and trace matching. The former is a method of gathering execution events into subtraces that correspond to different events. The latter is a process of pairing subtraces from two different execution traces. Third, a website can incorporate tons of noise. The webpage can utilize some Random Number Generators, there can be network errors, the content can be generated or selected dynamically.

The authors of the method give a glimpse of how to battle each of the difficulties, but leave out a lot of details in most cases. The contribution of this thesis are the following:

- The entire method has been implemented end-to-end and each step explained in detail.
- A novel approach based on Stable Marriage Problem to solve trace matching problem has been introduced.
- A new approach to filtering execution noises has been introduced and tested.
- The entire pipeline has been evaluated and used to conduct a small-scale study on the most popular websites in Poland.
- An overview of anti-adblocking methods has been presented, based on the pipeline evaluation.

The implementation has been manually tested on 100 websites and proved to be a useful tool in detecting and analysing anti-adblocking scripts. The analysis of those websites resulted in identification of the most popular mechanisms and solutions.

The study conducted on 300 most popular webpages in Poland revealed that anti-adblocking scripts are far more widespread than what can be conjectured by just studying visible reactions.

Chapter 1

Basic concepts

1.1. Definitions

- Execution event – each occurrence of control evaluating some expression, entering or leaving a control statement etc.
- Execution trace – a series of execution events collected during program execution. It is dependent both on program structure and its input (also implicit such as randomly generated numbers).
- Positive trace – an execution trace collected with an active adblocking extension.
- Negative trace – an execution trace collected without an adblocking extension.
- Execution index – a concept formally introduced by Xin et al. [45]. For our purposes we can define it as any function that uniquely identifies execution points. In our case it will be a statement source map information (file name and precise location of the statement) with the current function stack.

1.2. Adblockers

Online advertising market is growing each year. The Interactive Advertising Bureau's report for 2018 [24] states that the Internet advertising revenues surpassed \$100 billion annually in the US in 2018. Moreover, according to the same report, online advertising is the biggest ad media, approximately 30% bigger than TV.

What it means for the Internet users? That there will be more and more ads. Some sources estimate that an average Internet user is served 11250 ads per month [13].

Not surprisingly, ads are the main source of income for numerous websites. By displaying ads, authors are able to provide valuable content free of charge.

However, there are multiple reasons why users may not want to see online advertisements [33]:

- There are websites whose sole purpose is to earn money by displaying as many ads as possible without providing any interesting or original content. They often generate traffic using click-baits and similar shady practices.
- Ads often lengthen pages loading times. The slower the connection, the more annoying it becomes.

- Ads increase webpage payload size, which generates higher cost on mobile connections.
- Some ads track users, which raises privacy concerns.
- Ads can be used to spread malware (in this case it is called Adware) [3].
- Some of them are annoying.
- They occupy the screen space, leaving less area for the content, which can be especially frustrating on devices with small screens.

One of the solutions is to use special browser extensions, called ad blockers (or adblockers), that prevent ads from being displayed. Their work is based on community-curated list of filters. Those filters are used to first identify some portions of website as ads and later remove them.

The removal depends on the type of an ad. Whole page overlays can simply be not shown without disrupting the website UI. On the other hand, after banners removal, there remains some blank space, which is usually fixed by repositioning adjacent elements. In some cases, particularly when ads are served from domain of some ad provider, the requests that download ads can be blocked, thus saving the bandwidth and reducing data usage.

There is some controversy concerning morality of use of such extensions. One of the most popular ad-blocking extension, uBlock Origin states in its manifesto, that it is the users' right to have control over what content should be accepted in their browser [37]. As if other arguments for using adblockers were not enough of a justification.

Some ad-serving recognize users' frustration and create initiatives like Acceptable Ads [1]. In this program, advertisements meeting certain criteria are whitelisted by the committee. The whitelist is active by default in adblocking extensions that join the initiative. Some bigger ad providers are also charged for being whitelisted.

1.3. Anti-adblockers

In 2017 PageFair prepared a report on adblockers usage [33]. Worldwide, 11% users use ad blocking solutions. This may seem small, but it is not uncommon for some markets to around 20% or higher penetration (e.g. USA – 18%, Germany – 29%, Indonesia – 58%).

Such high percentage of users not seeing ads means lost income for many businesses. To recover lost revenue, many websites started deploying anti-adblocking scripts. Their goal is simple – when they detect that content blocking extension is present, they take some action, potentially mitigating the problem. The action can vary from simply just reporting the use of extension to the backend to blocking the content entirely.

The aforementioned PageFair report studied so-called "adblock walls", i.e. mechanism preventing users from seeing the website content when they have an adblocker enabled. The study shows that 90% of adblockers users have come upon an adblock wall. More interestingly, 74% of them leave the website when confronted with such a wall.

Anti-adblockers come in many variants. There are simple, custom scripts written specifically for one service, but there are also some sophisticated scripts provided by third parties, designed to be easily integrated on any site.

One rather simple example is presented by company offering "Adblock Analytics" service [2]. In their example they add a file named *ads.js* with a short JavaScript code that adds a hidden div block with an unique id. Ad-blocking extensions usually block files named like

that. All that remains to detect the extension is to check whether the `div` element was indeed added to the DOM tree. If not, it is a sign that an adblocker is active.

Another example is the "BlockAdBlock" module [35]. Similarly to the first example, it first sets up a bait, i.e. a component that should look like an ad to the adblocker, but then the detection phase is a little bit more thorough. It runs the check several times and it inspects several properties of the bait to make sure it has not been tampered with. If something changed – an adblocker is detected. It also allows to register callbacks for both outcomes of the analysis.

Even people that base their websites on content managements systems have an easy access to such scripts. Most notably, around 1 in 3 of the top 10 million most popular website runs on WordPress [44] and in this case there are multiple extensions available, varying in price and capabilities [5].

Naturally, there are multiple users that do not want to disable their adblocker to see the website's content. As a reaction, there are several solutions developed to combat anti-adblockers. Usually, they focus on providing new filtering lists, that can be used by the existing adblocking extensions [30, 34]

Unfortunately, those solutions have rather poor effectiveness, as reported by Zhu et al. [46]. This inability to effectively defuse anti-blockers was their most important reason to develop a method of detecting such scripts.

1.4. Differential Execution Analysis

Differential Execution Analysis is a dynamic program analysis method. Its goal is to pinpoint exact differences in two executions of the same program with different inputs or other conditions (e.g. network or memory errors). Good overview of the method is presented by Johnson et al. [25]

The analysis is usually performed by collecting an execution trace for each run and then comparing them using trace alignment. Trace alignment is a process of identifying which fragments of execution traces are common, where they diverge and when they converge again. The exact algorithm is discussed in section 4.3.

The results can be useful in various scenarios, i.e. debugging a program or during security analysis.

1.5. Detecting anti-adblockers

Zhu et al. in their paper "Measuring and Disrupting Anti-Adblockers Using Differential Execution Analysis" [46] introduced a novel method of detecting anti-adblockers using Differential Execution Analysis. The work presented in this thesis is based on that method. The differences and new ideas are explained in the later sections.

The premise is quite simple. They collect execution traces of JavaScript code on given website first without any content blocking and then with an adblocker turned on. Afterwards, they analyze such traces using Differential Execution Analysis. Any differences in execution in both cases are attributed to anti-adblocking activities of the website.

While the idea is pretty straightforward, there are multiple challenges here. First, trace collection is not a trivial task, especially when the interest exceeds just function entry and exit events. The authors instrument V8 engine to achieve the task, but do not provide much details, apart from briefly stating that their instrumentation is embedded into native code generation process.

Second, due to JavaScript execution model, described in detail in section 1.6, execution traces of different events are interleaved. To battle this issue, traces have to be sliced into subtraces and analyzed pairwise. In a language with a simpler execution model this step would be unnecessary. Authors also do not explain their method of how to pair the subtraces. They just mention that all $m \times n$ pairs have to be analyzed.

Lastly, the biggest challenge is to combat execution noises, e.g. page randomness or variable content. Authors solve the issue by loading the same page three times with an adblocker and three times without and use redundant traces to generate a black list of execution differences.

1.6. JavaScript execution model

JavaScript concurrency model is based on an "event loop" [31]. The engine is essentially single-threaded and concurrency is implemented by utilizing a message queue. This queue processes events one by one, to completion, i.e. a function corresponding to the message starts with a new, empty stack and its processing is done when the stack becomes empty.

The easiest way to add new events to the queue is by calling *setTimeout* or *setInterval*. Furthermore, all callbacks attached to DOM events (e.g. *onClick*) are executed by adding an event to the queue.

It is worth noting that execution of functions can be intertwined, e.g. when generators are used. This is the reason why trace slicing is needed.

Each iframe and browser tab has its own execution environment that include a separate message loop, more on that in section 3.2.

Chapter 2

Trace collection

2.1. Methods overview

There are a few distinct ways to obtain execution trace of JavaScript code.

The simplest (and most limited) methods use only mechanisms present in the language. More elaborate inject special tracing code to the analyzed script. The last kind modify the engine to produce the desired data.

2.2. Dynamic in-JavaScript code injection

JavaScript is a very dynamic language. For this reason, it is relatively easy to write code that will modify each function present in the environment to log each entry and exit, possibly along with all the arguments and return value [36].

Listing 2.1 is an example of a code that instruments all functions in a selected object to log their name and arguments when they are called. Function *instrument* simply goes through all properties of an object and replaces each function with a new function that first logs function name and all provided arguments and then calls the original function. This function can be easily extended to also log return value and instrument all subobjects recursively.

However, this is not enough for the needs of Differential Execution Analysis. The most obvious limitation is that it is not possible to instrument control statements. Another major shortcoming is that function can be instrumented only after they are defined. It is quite an obstacle, because JavaScript allows to define function practically anywhere and it is very common to use anonymous functions as callbacks. It is not possible to instrument such callbacks without modifying the instrumented code, which brings us the static injection methods.

```
1 function instrument(obj, withFn) {  
2   for (const name of Object.keys(obj)) {  
3     const fn = obj[name];  
4     if (typeof fn === 'function') {  
5       obj[name] = (function() {  
6         return function(...args) {  
7           withFn(name, ...args);  
8           return fn(...args);  
9         }  
10      })();  
11    }  
12  }  
13 }  
14
```

```

15 const o = {
16     f(w) {
17         /* function body */
18     }
19 };
20
21 instrument(o, console.log);
22
23 o.f("hello");

```

Listing 2.1: Dynamic instrumentation in JavaScript

2.3. Static code injection

Less limited approach is to statically rewrite the instrumented code and inject tracing code wherever it is needed. The upside of such approach is that there are several ready to use frameworks. The downsides will be pointed out when discussing each solution.

2.3.1. Web Tracing Framework

One notable solution is Web Tracing Framework developed by Google [22]. The main use of this framework is to profile web applications to find performance bottlenecks. The functionality is similar to that of *Performance* tab in Chromium developer tools.

Notwithstanding, one of its advanced features is closer to our needs. It allows the user to first instrument JavaScript sources and then collect execution traces and inspect them in a special app.

Having to instrument all source code is cumbersome, especially when we try to analyze code on some arbitrary websites. For this reason Web Tracing Framework also offers an extension and proxy server that cooperate to instrument all JavaScript code online, when it is loaded into the browser.

Unfortunately, this solution has a few deal-breaking downsides:

- It logs only function entry and exit events.
- Logging format is not public
- It is a bit dated, new JavaScript features may not be traced properly
- Function defined using *eval* or *Function* will not be traced

2.3.2. Iroh

Iroh [20] is the most complete solution based on static code injection. Just like Web Tracing Framework, Iroh also needs to patch the code first, but its capabilities go well beyond what the previous solution offers. It allows the user to register arbitrary callbacks to practically any element of JavaScript's Abstract Syntax Tree (AST). It means that this tool is able to instrument *if* statements. This use case is even included in the official examples.

Unfortunately, the framework does not offer proxy that could instrument code loaded into browser on the fly. There is also another, more general concern – performance of such solution may not be acceptable.

2.4. Engine instrumentation

The last option is to modify the JavaScript engine itself to produce execution traces. The most striking benefit is that the engine already has all required info and the solution does not require the analyzed code to be modified. Another advantage is the performance. Implementing tracing code directly in the engine means that there is less indirection. The code does not need to be interpreted by the engine, it is a native code that is called from JavaScript.

Unfortunately, such instrumentation has to be written almost from scratch. Nevertheless, due to the most flexibility and performance advantages, this solution has been chosen for this implementation.

The same choice has been made by Zhu et al. [46] for their implementation, but they did not share their code.

More details on how to instrument the engine in chapter 3.

Chapter 3

Trace collection by V8 instrumentation

3.1. V8 architecture

Most modern browsers do not implement JavaScript interpreter directly. Rather, they utilize a more specialized program called JavaScript engine, which they usually embed.

V8 is an engine used by the most popular browser, Chrome [41], which in June 2019 had over 80% market share [42].

Let us start with introducing some V8-specific glossary [39]:

- Isolate – an instance of V8. There can be more than one Isolate used by one process of embedding application.
- Context – a concept of global variable scope in V8. Each Context has its own global variables and prototype chain. Each iframe has its own separate Context. There can be multiple Context in one Isolate, but due to site isolation (see section 3.2), each iframe is run in another process and has its own Isolate.

V8 processes JavaScript code in several steps. In this thesis we will focus on steps directly related to implementing trace collection.

In short, JS code is first parsed into AST, which contains source map information. In the next step V8 traverses the entire AST and emits bytecode for each node. The bytecode is V8-specific and reflects the architecture of V8's abstract machine. More on that in section 3.1.1.

It is worth noting that while user-defined functions are translated to bytecode, most built-in functions are implemented in a different way. We will take a closer look at them in section 3.1.2.

Only after the code is translated into bytecode, it is finally executed. At this stage there are two kinds of functions. First – those defined in JavaScript, represented in bytecode, second – builtins defined in other ways and already compiled into native code. This distinction is not important to the user, as those functions do not differ in JavaScript, and can easily call each other. It will become important once we try adding instrumentation code.

At some point during execution, functions that are called very often with the same argument types, can be compiled into native code by its TurboFan Just In Time compiler [11]. If later the same function is called with some argument of some other type, it simply gets deoptimized into bytecode.

The engine’s architecture is focused on achieving superior performance, while conforming to all standards and not jeopardizing security.

3.1.1. JS bytecode

V8’s interpreter, Ignition, is a register machine with an accumulator register [21]. While all other registers have to be specified when used as arguments, accumulating register is implicit, it is not specified by bytecodes that use it.

This section is supposed to be only a shallow dive into V8’s bytecode. We will have a look at one simple example to be able to understand what is going on in section 3.5.

Let us have a look at a simple JavaScript function and see the bytecode produced by Ignition.¹ Listing 3.1 shows a naive implementation of a function calculating factorial. It includes function call (line 5) because the interpreter is lazy and otherwise the function would not be compiled. List 3.2 presents bytecode generated by V8’s interpreter for that function.

```
1 function factorial(n) {
2     return n <= 1 ? 1 : n * factorial(n - 1);
3 }
4
5 console.log(factorial(3));
```

Listing 3.1: Calculating factorial in JavaScript

```
1 [generated bytecode for function: factorial]
2 Parameter count 2
3 Register count 3
4 Frame size 24
5   18 E> 0x167d37c1eb2a @    0 : a5          StackCheck
6   28 S> 0x167d37c1eb2b @    1 : 0c 01        LdaSmi [1]
7   37 E> 0x167d37c1eb2d @    3 : 6b 02 00      TestLessThanOrEqual a0,
   [0]
8       0x167d37c1eb30 @    6 : 99 06          JumpIfFalse [6] (0
   x167d37c1eb36 @ 12)
9       0x167d37c1eb32 @    8 : 0c 01          LdaSmi [1]
10      0x167d37c1eb34 @   10 : 8b 15          Jump [21] (0x167d37c1eb49
   @ 31)
11  48 E> 0x167d37c1eb36 @   12 : 13 00 02      LdaGlobal [0], [2]
12      0x167d37c1eb39 @   15 : 26 fa          Star r1
13      0x167d37c1eb3b @   17 : 25 02          Ldar a0
14  64 E> 0x167d37c1eb3d @   19 : 41 01 04      SubSmi [1], [4]
15      0x167d37c1eb40 @   22 : 26 f9          Star r2
16  52 E> 0x167d37c1eb42 @   24 : 5d fa f9 05      CallUndefinedReceiver1 r1,
   r2, [5]
17  50 E> 0x167d37c1eb46 @   28 : 36 02 01      Mul a0, [1]
18  69 S> 0x167d37c1eb49 @   31 : a9          Return
19 Constant pool (size = 1)
20 0x167d37c1ead9: [FixedArray] in OldSpace
21   - map: 0x0a0b39f807b1 <Map>
22   - length: 1
23       0: 0x167d37c1e741 <String[#9]: factorial>
24 Handler Table (size = 0)
```

Listing 3.2: Ignition bytecode for function *factorial*

The listing starts with the parameter count, register count and frame size. The first one may be baffling at first since *factorial* takes only one number as an argument. We have to remember that all JavaScript functions also take implicit arguments *this*.

¹V8 prints out bytecode when flag `--print-bytecode` is provided

After the header, the actual bytecode is listed. The first number and letter, e.g. "18 E" in line 5 identifies expression (E) or statement from the source file. The number is an offset in characters. It is followed by the code's address in memory, offset (in bytes) from function start and the code itself in a hexadecimal form. All of them are rather useless for us. The most useful parts are the codes in human-readable form at the end of each line.

Once we recall that most codes use accumulating register, reading the code becomes relatively self-explanatory. To make things easier, the use of accumulating register is reflected in the code's name, e.g. *LdaConstant [0]* loads constant numbered 0 to the accumulating register.

We should now be ready to interpret each line of *factorial*'s bytecode. Upon function entry (line 5) the validity of the stack is checked. Later, integer 1 is loaded into accumulating register. Next, argument 0 (symbol *a0*), which happens to be n , is tested to be less than or equal to the value stored in the accumulating register (1). If not, the jump (to line 11 in the listing) is performed. If the conditional was true, integer 1 is loaded into accumulating register, then the control jumps to the last instruction which returns from the function. The return value is always stored in the accumulating register, so 1 is returned. If the conditional was true, and we are in line 11, constant 0 is loaded, which happens to be the name of our function. Later, that name is stored in register *r1*, argument 0 is loaded into the accumulating register, 1 is subtracted and the result is stored in register *r2*. Next, function of name stored in *r1* (*factorial*) is called with value stored in register *r2* ($n - 1$). The result of the call, stored in the accumulating register, is then multiplied by the first argument (n). The result of multiplication is already in the accumulating register and the function can now return.

3.1.2. JS built-in functions

According to V8's documentation post [38], JavaScript built-in functions can be implemented in three different ways. They can be written in JavaScript directly, implemented in C++ (runtime functions) or defined using an abstraction called Code Stub Assembly (CSA). The post, however, is slightly dated. Since then, new abstraction, Torque [40], has been added.

It is not important to know how to write CSA or Torque code. It is only important to remember that most built-in functions are implemented in a different way than a user-defined ones.

3.2. V8 usage in Chromium

Today's usage of V8 in Chromium is determined in large part by security concerns [10]. For us, the most important decision was made after discovery of Spectre [27] and Meltdown [28] vulnerabilities. To increase protection against attacks based on those two vulnerabilities, Chrome's team decided to make Site Isolation enabled by default. [16]

Each website can have multiple iframes – the default one and some embedded ones. All of them are run in a separate process and, as a consequence, have their own instances (Isolates) of V8.

3.3. Chrome's extensions architecture

In the current model, Chrome's extensions may consist of the following components [15]:

- Manifest – a file describing an extension, listing all its files and capabilities.

- UI Elements – code adding extension’s user interface.
- Options Page – a page allowing the user to customize the extension.
- Background script – a file with callbacks for browser events. Run only when an event with a registered callback occurs. It runs in its own Context, in a separate process
- Content script – extension’s code that is run in the page’s Context. This code can read and modify DOM of the website and communicate with parent extension via messages or storage. It can also access a limited subset of Chrome’s APIs directly, mostly those needed to communicate with parent extension [14].

3.4. V8’s *--trace* flag

Usually the easiest way to implement some new functionality is to find a code that provides a similar functionality and extend it. In case of tracing, such base is provided by the V8’s *--trace* flag.

First, we will inspect what this flag can do. The example from the section 3.1.1 (Listing 3.1) will be reused. Listing shows console output of V8 with *--trace* flag enabled. The output is well-formated and self-explanatory. Unfortunately, it has some shortcomings. First, there is no source map info. Second, due to the nature of JavaScript, function traces can get intertwined (more on this in section 1.6. And since stack information is limited to just the stack depth, it may be impossible to untangle events in some cases.

```

1 1: ~(this=0x1bd8e6501521 <JSGlobal Object>) {
2 2:   ~factorial(this=0x1bd8e6501521 <JSGlobal Object>, 3) {
3 3:     ~factorial(this=0x1bd8e6501521 <JSGlobal Object>, 2) {
4 4:       ~factorial(this=0x1bd8e6501521 <JSGlobal Object>, 1) {
5 4:         } -> 1
6 3:       } -> 2
7 2:     } -> 6
8 6
9 1: } -> 0x1bc7923804d1 <undefined>

```

Listing 3.3: V8’s output for *factorial* with *--trace* flag

Nevertheless, this flag is a good starting point. Let’s have a deeper dive into how it works.

We have already seen the bytecode for *factorial* in listing 3.2. Listing 3.4 shows the bytecode for the same function when *--trace* flag is enabled. There are two differences compared to the bytecode produced without tracing flag. First – there is a call to runtime function *TraceEnter* before *StackCheck* upon function entry. Second, just before returning, the result is saved to register *r0* and runtime function *TraceExit* is called with return value as its argument. *TraceExit* stores its argument back in accumulating register so there is no change in semantics of the inspected code.

```

1 [generated bytecode for function: factorial]
2 Parameter count 2
3 Register count 3
4 Frame size 24
5      0x315c6429eb32 @      0 : 61 a7 01 fb 00      CallRuntime [TraceEnter],
      r0-r0
6 18 E> 0x315c6429eb37 @      5 : a5              StackCheck
7 28 S> 0x315c6429eb38 @      6 : 0c 01              LdaSmi [1]
8 37 E> 0x315c6429eb3a @      8 : 6b 02 00              TestLessThanOrEqual a0,
      [0]

```

```

9      0x315c6429eb3d @ 11 : 99 06      JumpIfFalse [6] (0
    x315c6429eb43 @ 17)
10      0x315c6429eb3f @ 13 : 0c 01      LdaSmi [1]
11      0x315c6429eb41 @ 15 : 8b 15      Jump [21] (0x315c6429eb56
    @ 36)
12      48 E> 0x315c6429eb43 @ 17 : 13 00 02      LdaGlobal [0], [2]
13      0x315c6429eb46 @ 20 : 26 fa      Star r1
14      0x315c6429eb48 @ 22 : 25 02      Ldar a0
15      64 E> 0x315c6429eb4a @ 24 : 41 01 04      SubSmi [1], [4]
16      0x315c6429eb4d @ 27 : 26 f9      Star r2
17      52 E> 0x315c6429eb4f @ 29 : 5d fa f9 05      CallUndefinedReceiver1 r1,
    r2, [5]
18      50 E> 0x315c6429eb53 @ 33 : 36 02 01      Mul a0, [1]
19      0x315c6429eb56 @ 36 : 26 fb      Star r0
20      0x315c6429eb58 @ 38 : 61 a8 01 fb 01      CallRuntime [TraceExit],
    r0-r0
21      69 S> 0x315c6429eb5d @ 43 : a9      Return
22      Constant pool (size = 1)
23      0x315c6429eae1: [FixedArray] in OldSpace
24      - map: 0x3806c0d807b1 <Map>
25      - length: 1
26      0: 0x315c6429e741 <String[#9]: factorial>
27      Handler Table (size = 0)

```

Listing 3.4: Ignition bytecode for function *factorial* with `--trace` enabled

3.5. Bytecode injection

Finally, we have come to the gist of the current chapter – tracing implementation. Once we have seen how `--trace` flag works, we can improve it to our needs.

The entire implementation requires a few steps:

1. Adding two new flags – one for turning on our tracing (`--trace-dea`), and one for specifying the file with tracing info
2. Preparing a function that prints out the entire stack with source map information
3. Preparing a set of new runtime functions, one for each type of event we want to trace
4. Injecting calls to runtime functions in the appropriate places

We will not delve too deeply on how to implement each part. Points 1 and 3 are pretty straightforward. Both of them just require adding declaration to special header files.

Point 2 seems the hardest, but luckily there is a similar function in the Chromium codebase. It is worth noting that it is possible to recover source map information during runtime and have the entries contain precise line and column info. However, it turned out to be too slow for use in Chromium, as there was a noticeable slowdown. Almost no webpage could finish loading in a reasonable time. The reason for such poor performance is that those locations are not stored directly in AST. Rather, only offset in characters is stored. To recover line and column info it is necessary to first go through a few levels of abstraction to get the source code and then calculate the coordinates by traversing it. As a consequence, only character offset is displayed. It is always possible to recover more friendly line, column location later, so it is not that big of a deal.

The challenge of part 4 is to have the right offsets of statements/expressions. It is the easiest with functions, as location of beginning of their definition is stored in an object representing the function during runtime. In case of usual statements it is harder as their locations are available only during parsing and bytecode generation stages. To have those offsets accessible during runtime, they are stored as code constants and passed as arguments to runtime functions that print out the code events.

Listing 3.5 shows bytecode generated by Ingition with our tracing flag enabled. Similarly to the original tracing, there are calls to runtime functions upon entry (line 5) and just before returning (lines 23-24). The new part is that also the information which branch was taken is logged. In lines 7-8 the offset information is stored in register *r0* which is later used by runtime functions that perform the actual logging (lines 12, 15)

```

1 [generated bytecode for function: factorial]
2 Parameter count 2
3 Register count 4
4 Frame size 32
5      0x3dd338d9eb3a @      0 : 61 a9 01 fb 00      CallRuntime [TraceDeaEnter
6      ], r0-r0
7      18 E> 0x3dd338d9eb3f @      5 : a5          StackCheck
8      28 S> 0x3dd338d9eb40 @      6 : 12 00          LdaConstant [0]
9      0x3dd338d9eb42 @      8 : 26 fb          Star r0
10     0x3dd338d9eb44 @     10 : 0c 01          LdaSmi [1]
11     37 E> 0x3dd338d9eb46 @     12 : 6b 02 00      TestLessThanOrEqual a0,
12     [0]
13     0x3dd338d9eb49 @     15 : 99 0b          JumpIfFalse [11] (0
14     x3dd338d9eb54 @ 26)
15     0x3dd338d9eb4b @     17 : 61 af 01 fb 01      CallRuntime [
16     TraceDeaIfStmtThen], r0-r0
17     0x3dd338d9eb50 @     22 : 0c 01          LdaSmi [1]
18     0x3dd338d9eb52 @     24 : 8b 1a          Jump [26] (0x3dd338d9eb6c
19     @ 50)
20     0x3dd338d9eb54 @     26 : 61 b0 01 fb 01      CallRuntime [
21     TraceDeaIfStmtElse], r0-r0
22     48 E> 0x3dd338d9eb59 @     31 : 13 01 02      LdaGlobal [1], [2]
23     0x3dd338d9eb5c @     34 : 26 f9          Star r2
24     0x3dd338d9eb5e @     36 : 25 02          Ldar a0
25     64 E> 0x3dd338d9eb60 @     38 : 41 01 04      SubSmi [1], [4]
26     0x3dd338d9eb63 @     41 : 26 f8          Star r3
27     52 E> 0x3dd338d9eb65 @     43 : 5d f9 f8 05      CallUndefinedReceiver1 r2,
28     r3, [5]
29     50 E> 0x3dd338d9eb69 @     47 : 36 02 01      Mul a0, [1]
30     0x3dd338d9eb6c @     50 : 26 fb          Star r0
31     0x3dd338d9eb6e @     52 : 61 aa 01 fb 01      CallRuntime [TraceDeaExit
32     ], r0-r0
33     69 S> 0x3dd338d9eb73 @     57 : a9          Return
34 Constant pool (size = 2)
35 0x3dd338d9eae1: [FixedArray] in OldSpace
36 - map: 0x0f45366007b1 <Map>
37 - length: 2
38     0: 35
39     1: 0x3dd338d9e741 <String[#9]: factorial>
36 Handler Table (size = 0)

```

Listing 3.5: Ignition bytecode for function *factorial* with *--trace-dea* enabled

Listing 3.6 shows the output produced by the new flag. Each execution event entry consists of event type, location (optional, depends on event type) and full call stack. Events returning value also log that value, but it is not used later in the pipeline. Stack is represented as a

list of locations. Each location consists of function name, file of origin and offset in the file (in characters). The `-1` that appears after the offset is a remnant of the implementation that recovers full position info (line, column). The next part of the pipeline expects two numbers here. This way it is easy to turn on full position recovery and have the pipeline still working.

Execution events that are logged in the current implementation:

- Function enter and exit
- Generator enter, suspend, yield (exit is indistinguishable from a normal function exit)
- If statement then/else paths
- Ternary expression truthy/falsy value (same as if statement paths)

It is easy to extend the implementation to also log execution events associated with loops, but there is a tradeoff between coverage and size of log files created, so it was not done.

```

1 FUNCTION ENTER
2   0: ~ at factorial.js @@ 0,-1 ()
3   =
4   --
5 FUNCTION ENTER
6   0: ~factorial at factorial.js @@ 18,-1 ()
7   1: ~ at factorial.js @@ 0,-1 ()
8   =
9   --
10 IF STMT - ELSE
11   2: ~factorial at factorial.js @@ 35,-1
12   0: ~factorial at factorial.js @@ 18,-1 ()
13   1: ~ at factorial.js @@ 0,-1 ()
14   =
15   --
16 FUNCTION ENTER
17   0: ~factorial at factorial.js @@ 18,-1 ()
18   1: ~factorial at factorial.js @@ 18,-1 ()
19   2: ~ at factorial.js @@ 0,-1 ()
20   =
21   --
22 IF STMT - ELSE
23   3: ~factorial at factorial.js @@ 35,-1
24   0: ~factorial at factorial.js @@ 18,-1 ()
25   1: ~factorial at factorial.js @@ 18,-1 ()
26   2: ~ at factorial.js @@ 0,-1 ()
27   =
28   --
29 FUNCTION ENTER
30   0: ~factorial at factorial.js @@ 18,-1 ()
31   1: ~factorial at factorial.js @@ 18,-1 ()
32   2: ~factorial at factorial.js @@ 18,-1 ()
33   3: ~ at factorial.js @@ 0,-1 ()
34   =
35   --
36 IF STMT - THEN
37   4: ~factorial at factorial.js @@ 35,-1
38   0: ~factorial at factorial.js @@ 18,-1 ()
39   1: ~factorial at factorial.js @@ 18,-1 ()
40   2: ~factorial at factorial.js @@ 18,-1 ()
41   3: ~ at factorial.js @@ 0,-1 ()
42   =

```

```

43 --
44 FUNCTION EXIT
45   0: ~factorial at factorial.js @@ 18,-1 ()
46   1: ~factorial at factorial.js @@ 18,-1 ()
47   2: ~factorial at factorial.js @@ 18,-1 ()
48   3: ~ at factorial.js @@ 0,-1 ()
49 -> 1
50 =
51 --
52 FUNCTION EXIT
53   0: ~factorial at factorial.js @@ 18,-1 ()
54   1: ~factorial at factorial.js @@ 18,-1 ()
55   2: ~ at factorial.js @@ 0,-1 ()
56 -> 2
57 =
58 --
59 FUNCTION EXIT
60   0: ~factorial at factorial.js @@ 18,-1 ()
61   1: ~ at factorial.js @@ 0,-1 ()
62 -> 6
63 =
64 --
65 6
66 FUNCTION EXIT
67   0: ~ at factorial.js @@ 0,-1 ()
68 -> 0x0da3956804d1 <undefined>
69 =
70 --

```

Listing 3.6: V8's output for *factorial* with *--trace-dea* flag

Careful reader might have noticed that in listing 3.1 in the last line there is a call to *factorial* and the result is passed to *console.log*, but only the former call is logged. The reason for this is that the implemented solution works only for functions defined in JavaScript (it is also true for the default *--trace* flag) Functions defined in other ways (see section 3.1.2) are not logged. It is certainly possible to add logging to each one of them by modifying their code or by modifying CSA/Torque compiler, but it has not been done here. The amount of work required to instrument also those function seems disproportionate to the potential improvements. Another justification is that they are a black box anyway, there is no JavaScript code corresponding to them and we cannot see branch divergences happening inside them.

The last obstacle worth noting is the Chrome's process separation. As explained in section 3.2 there are multiple instances of V8 running at the same time, in different processes. When some flag is passed to Chrome, all instances see the same value. Therefore, if some file is passed, all processes will write to the same file, possibly resulting in interlacing and unusable output. An easy workaround is to create a new file for each Isolate and later just select only the interesting one (the one that corresponds to the analyzed website). There will be always at most one such file (theoretically a website could not contain any JavaScript code) and it is easy to select it by greping by source location. All other files can be deleted.

V8 Isolate is generally pretty oblivious to what code it runs. Website code is the same to V8 as extension code. After all, the environment (Web API, DOM, etc.) is provided by the browser. Nevertheless, it is preferable not to log adblocker events. Such extensions have really long lists of filters and their intialization takes 1-2 seconds when the browser is not producing execution traces. When it does, the initialization takes more than half a minute on a slow computer and the trace itself occupies 4.5GB. Again, we can resort to a trick: inspect the printed event and stop tracing in a given Isolate when some extension-specific file has been

encountered. As a result, extension code is not traced and it has no noticeable performance deterioration.

3.6. Controlling Chrome programmatically

Adblocking extensions need a second or two to initialize their code. Even in a stock browser, opening a website immediately after the browser starts can result in ads not being blocked. As a consequence, simply passing the website address through console is not enough when we want to collect traces automatically. The extension will simply not be fully initialized and website will not be blocked. Some more sophisticated solution is needed.

Fortunately, there is a framework for building end-to-end tests – Selenium [9]. It has binding in all major languages, in our case we will use Python.

Selenium is capable of doing any interaction with the website that user can do. It communicates directly with a specialized WebDriver, different for each browser. In case of Chrome it is called ChromeDriver. WebDriver issues the command to the browser through the debugging interface, which is a special port with well-defined communication protocol. And that is all we know about it, as we are not using any of its sophisticated features, just connecting to the browser, opening a website and closing it after some time.

Chapter 4

Trace analysis

4.1. Parsing

Although the format presented in chapter 3 is really simple, parsing it can prove to be challenging. The sole reason is the size of the files. It is really common for websites to produce files of size in the range of a few Gigabytes. Some can even output as much as 48 GB in about 2 minutes!

The analyzing program was written in Haskell [23]. Haskell is a purely functional, lazy, statically typed language. It has been chosen because it is relatively easy to write parsers in a language like this. Further, static typing with pattern matching proves convenient when manipulating well-structured data like log entries. Last but not least, automatically derived instances ¹ help avoid writing tedious code and focus on implementing non-trivial parts.

All that being said, Haskell is a bit like C++ – inexperienced user can easily make mistakes that render the code slow and memory-greedy.

The logging format was described in section 3.5. Each entry consist of event type, and several locations. Each location comprises function name, source file and position in the file.

At first, internal format for the event was exactly the same. One object consisted of two strings and two numbers (it could be one number, but this way line and column info logging can be turned on any time).

The first mistake, specific to Haskell, was to use *String* type to represent function and file names. The use of default *String* to represent textual data makes the code inefficient because it is a linked lists of *Chars*, i.e. to store one character 9 bytes of memory are used. It is so wrong and inefficient to use this type that it is not worth trying to profile such implementation.

That error was fixed by changing the representation to *ByteString* type from *bytesting* package [19]. *ByteStrings* are represented internally as real byte arrays, not linked lists. This change was accompanied by rewrite of the parsing code to *attoparsec* [12], which can operate of *ByteStrings* as an input.

The code was now much less memory-consuming and faster, but it still seemed to consume too much memory. Profiling proved that suspicions were warranted. The profiling diagram is shown on figure 4.1. All profiling diagrams in this section were collected by running the analyzing program on the same set of 6 input files. First three occupy 48MB each, the last three 4MB each. All files combined occupy around 150MB, while peak memory usage of the program is above 900MB.

¹Without going into details, classes in Haskell are a bit like interfaces in object-oriented languages, e.g. class *Ord* defines objects that can be compared

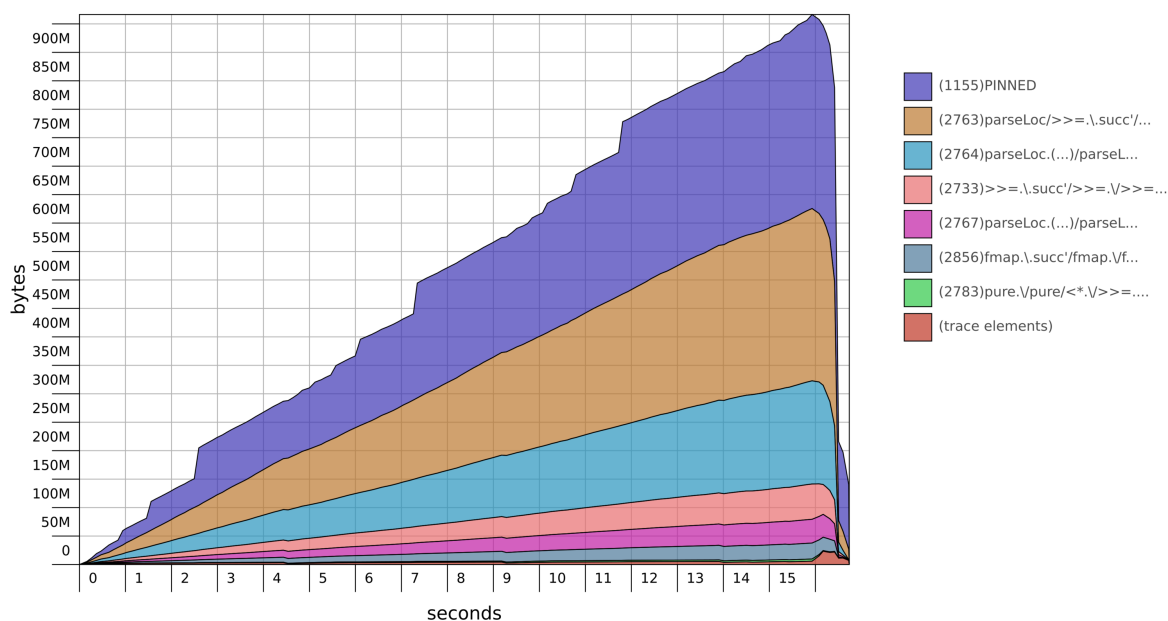


Figure 4.1: Memory usage of the first implementation using ByteStrings as internal strings representation

Looking at the diagram, pinned memory seemed to be never freed during the execution. Also, traces seemed to reside in memory for too long and occupy too much space.

The first problem was a reflection of how Haskell keeps *ByteStrings*. They are stored in pinned memory, which is kept on a heap but not managed by garbage collector [29] and cannot be moved around. As a result, it leads to fragmentation of the heap and cannot be effectively managed. The fix turned out to be simple. There is a more suitable storage format – *ShortByteString* (also part of the *bytestring* package). It is managed by GC and stored just like usual heap object, which can be moved around and easily garbage collected. ²

The second problem was harder to track down. This time the cause was the laziness. Laziness can improve the performance when some objects are never used and the language never needs to fully compute them. However, when we know that all objects will eventually be used, it is more efficient to calculate them as soon as possible. The enforcement of eager evaluation seems to be an obscure feature, but when we know what is going on, it is fine. The improvement was visible (figure 4.2) but it was not the end, as peak usage of almost 240MB is still greater than 150MB.

It seems suspicious that the entire file has to be kept in memory to be parsed (pinned memory on figure 4.2 represents input files read into memory). After all, if it was written in C++, probably one line at a time would be read and processed. So why this parser consumes so much memory?

This is what *attoparsec* states about incremental input:

Note: incremental input does not imply that *attoparsec* will release portions of its internal state for garbage collection as it proceeds. Its internal representation is equivalent to a single *ByteString*: if you feed incremental input to a parser,

²So why even use *ByteString*? Because it is stored in pinned memory, it can be passed to foreign functions. Also, it is more efficient when the strings are really long and have long lifespan

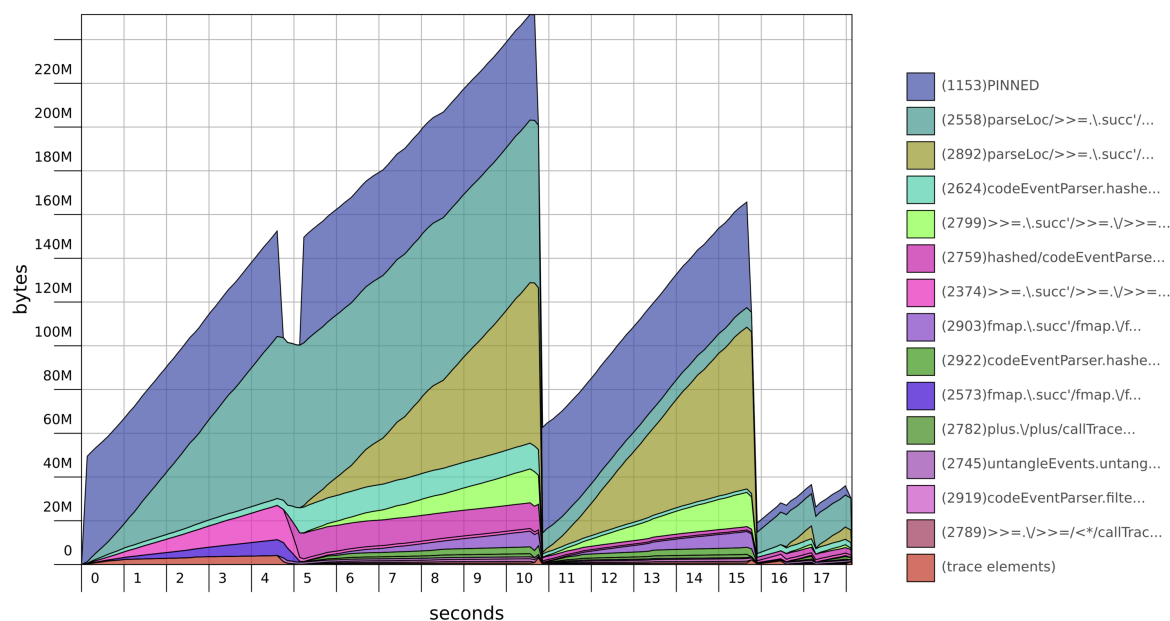


Figure 4.2: Memory usage after switching to ShortByteString and forcing eager evaluation

it will require memory proportional to the amount of input you supply. (This is necessary to support arbitrary backtracking.)

So, our parser keeps the whole file contents in memory to be able to backtrack. But it is not necessary with such a simple format. Fortunately, it is possible to improve this behaviour. Currently, the parser tries to parse the entire file into a list of events. To be sure that it can backtrack, it keeps the whole input in memory. Instead, we can ask the parser to parse only one event. It will do it and return the part of the input that was not parsed yet. We can repeat that in a loop and parser will never keep more memory than just a few kilobytes. Figure 4.3 shows the improvement.

Last, but not least, the internal representation of the event can be greatly improved. Locations consist of file names and function names. But both sets are limited! Usually there are only a few sources containing just a few hundreds of unique functions. We can create a map of all source and function names and just keep appropriate ids in location objects. Just 4 numbers instead of 2 strings and 2 numbers!

This representation also has another major benefit – comparisons of such objects are much faster.

The final memory consumption is presented in figure 4.4

4.2. Trace untangling

Due to the nature of JavaScript execution model (see section 1.6), execution events corresponding to different JavaScript events or functions can be intertwined.

For this reason, all events forming a trace have to be untangled into subtraces, each corresponding to one script or callback.

Let us recall that all events are logged by our instrumented Chromium with their call stacks. Now, we have three cases:

Naturally, some care has to be taken to ensure that the right stack is compared. Some events change the call stack, e.g. function enter and exit, some do not, e.g. "if statement – then". Here, the stack *after* the event occurred is saved and it is appropriately modified (one location can be added, removed or nothing is changed) when comparing with past events stacks.

Listing 4.1 presents the pseudocode of the algorithm.

```

1 def untangleEvents(events):
2     closedTraces := {}
3     openTraces := {}
4
5     for event in events:
6         if event.type == FunctionEntry and |event.stack| == 1:
7             openTraces.insert([event])
8         else:
9             matchingSubtrace := findMatchingSubtrace(event, openTraces)
10            openTraces.remove(matchingSubtrace)
11            matchingSubtrace.append(event)
12
13            if event.type == FunctionExit and |event.stack| == 0:
14                closedTraces.insert(matchingSubtrace)
15            else:
16                openTraces.insert(matchingSubtrace)
17
18    return closedTraces
19
20 def findMatchingSubtrace(ev, traces):
21     st := case
22         ev.type in [FunctionEnter, GeneratorEnter] → tail ev.stack
23         ev.type in [FunctionExit, GeneratorSuspend] → (ev.loc : ev.stack)
24         otherwise → ev.stack
25
26     return trace whose last event's stack equals st

```

Listing 4.1: Pseudocode of the trace untangling algorithm

The above code is rather uncomplicated, but it is slow when implemented naively. The most wasteful part is finding the trace with the right stack in *findMatchingSubtrace*. To make it faster, two things were done:

- Location objects are just 4 numbers instead of 2 strings and 2 numbers. This makes stack comparisons one or two orders of magnitude faster
- Open traces are kept in a map indexed by stack of the last event. The lookup becomes logarithmic instead of linear. This optimization is especially important when there are lots of intertwined subtraces. It also must be noted that this optimization alone would not help much if the comparisons between stacks were slow

One last caveat – it is possible to enter a function in JavaScript and never return. When an exception is thrown and there is no catch block in a function, the error will be propagated higher and control can possibly leave multiple functions without any of them returning any value. Those throw events will not be logged, therefore this situation has to be taken care of in the analyzing program. The solution is the following: when there is no open subtrace with a matching stack, the program will check if some stack will match with any number of elements removed from the top. If yes, then we assume that it is due to the try-catch block.

4.3. Trace alignment

Trace alignment is a technique of identifying execution differences of two runs of the same program. To be able to align two traces, events in both have to be assigned execution indices. In our the execution index is the event type, its location and stack at the time of its occurrence. In this thesis execution event is often used interchangeably execution event as everything that is logged when an event occurs is the index.

The algorithm used in this implementation is based on an article by Johnson et al. [25] The result of trace alignment is an execution trace diff. It is a list of events of three kinds:

- Common – an event that occurred in both traces
- Left – an event that occurred only in the left trace
- Right – an event that occurred only in the right trace

Each diff event includes the associated execution event. Listing 4.2 present the pseudocode of the algorithm.

```
1 def alignTraces(leftTrace, rightTrace):
2     results := []
3     score := 0
4
5     if leftTrace.size() == 0 || rightTrace.size() == 0
6         || leftTrace[0] != rightTrace[0]:
7         return (-1, [])
8
9     while leftTrace.size() > 0 || rightTrace.size() > 0:
10        if leftTrace.size() == 0:
11            results.append(map(Right, rightTrace))
12            rightTrace := []
13        elif rightTrace.size() == 0:
14            results.append(map(Left, leftTrace))
15            leftTrace := []
16        else:
17            (leftEvent:leftTail) := leftTrace
18            (rightEvent:rightTail) := rightTrace
19
20            if leftEvent == rightEvent:
21                score := score + 1
22                results.append(Common(leftEvent))
23                leftTrace := leftTail
24                rightTrace := rightTail
25            elif leftEvent.stack.size() <= rightEvent.stack.size():
26                result.append(Right(rightEvent))
27                rightTrace := rightTail
28            else:
29                result.append(Left(leftEvent))
30                leftTrace := leftTail
31
32    return (score, results)
```

Listing 4.2: Pseudocode of the trace alignment algorithm

The algorithm output includes a score, which is a number of common events. This concept was not present in the original article and its purpose will be explained in section 4.4.

The algorithm is rather simple. We assume that matching traces have to start with the same event. If they do not (lines 5-6), we just terminate and return score of -1, which signifies that traces do not match.

Apart from the case when one of the traces ends, there are only two cases:

- Both unprocessed traces start with the same event (that includes stacks equality), line 20 – in that case we add an event of type Common and remove one event from both lists
- The top events do not match (lines 25 and 28) – in that case the event with larger stack is removed first. The reasoning is the following. The bottom of the larger stack is the same as the shorter stack. If we first process the shorter stack and if it shrinks, we lose some common events later. If we process the larger stack first, the shorter one is "frozen". The larger stack may grow, but it doesn't matter since it is already divergent. When it shrinks to the size of the shorter stack, it will be a reconvergence point.

4.4. Trace matching using Stable Marriage Problem

We already know how to align two traces (section 4.3), but we have not answered a different question – which pair of substraces should we align?

This thesis presents a novel approach, which uses Stable Marriage Problem to answer this question. First, we should define what SMP is and later explain how it is adapted to our needs.

The original statement of the problem is the following [43]: given equally sized sets of men and women and their matrimonial preferences, find a stable matching, i.e. matching in which there exists no pair of man and woman in which they both would have better partner than the currently assigned one.

The problem can be solved using Gale-Shapley algorithm. Listing 4.3 presents its pseudocode.

```

1 def stableMatching(men, women):
2     for m in men:
3         m.matching = free
4     for w in women:
5         w.matching = free
6
7     while there exists a free man m who still has a women to propose to:
8         w := first women on m list that he has not yet proposed to
9         if w.matching == free:
10            m.matching := w
11            w.matching := m
12        else:
13            m' := w.matching // current provisional partner of w
14            if w.preferences[m] > w.preferences[m']:
15                m'.matching = free
16                m.matching = w
17                w.matching = m
18            else:
19                pass // w and m' remain engaged

```

Listing 4.3: Pseudocode of the Gale-Shapley (deferred acceptance) algorithm

This is how SMP is used to solve subtrace matching problem:

- In the first phase all pair of substraces are aligned and similarity score (number of common events) is calculated – the score is the marital preference

- In the second phase, adapted Gale-Shapley algorithm is run and only matched subtraces are used. Their diff is simply retrived from a special map as it was calculated already in the first phase.

All men have a list of women sorted by their preferences (from highest to lowest). At the beginning everyone is free. Then, in each round each free man m proposes to first woman w he has not yet proposed to. Woman can reply "maybe" if she is free or is engaged to m' and prefers m over m' . Otherwise she rejects. So women only "trade up" and men propose whenever they are free to the best possible partner. In the end everybody is engaged and the marriages are stable.

Time complexity of this algorithm is $O(n^2)$, where n is the number of men or women. There is at most n rounds (some man proposes to every woman) and in each round there are at most n proposals made.

In the original problem statement the number of woman is equal to the number of men and everybody gets engaged in the end. In our case the number of men and women does not have to be the same. We assume that the more numerous group has the role of women. We allow some pairs of men and women to not match to each other (preference score equal to -1), therefore in the end some people may not be engaged.

Let us see what happens with these two changes and the same procedure as in the original version:

- The men should still propose in the order from highest preference to the lowest. All man will still get the best possible match this way, because all women that rejected must have had some better options.
- The women should accept anybody when they do not have any tentative match and only trade up later. If there is a man that a woman prefers over a current match and he did not propose, he must have had someone who he preferred more and who accepted.

Consequently, the algorithm is almost the same. The only exception is that men do not propose to women with score -1. Also, the result is different – in the end there usually will be many people without a pair. However, those who have a pair will form a stable marriage.

Such usage of Stable Matching Algorithm allows us to form the best pairs of subtraces to align. There is no point in aligning the traces that do not have the same initial event (the aligning algorithm has to start from an anchor point), so we skip those pairs with assigning score -1. Further, the best pairs are selected by the number of common event. If two traces A and B are exactly the same, the only possibility that they will not be matched is when there is one longer trace C that contain entire A and B . But, if there is also a longer trace D that contains entire A and B , it is very likely that the extra part will be matched with extra part of C and, as consequence, C and D will be matched. If there is no such trace D and C is matched is either A or B , it is still fine, as we have more calls to some function in one group and we cannot be sure which subtraces truly correspond to each other.

4.5. Artificial examples

This section presents the algorithm work on some simple, hand-crafted examples run in standalone V8.

First example is based on our already-familiar example of *factorial* (listing 3.1).

First, we will call the function with $n = 3$, then with $n = 4$ and compare the traces. This demonstrates how "if-then-else" divergences are detected. Listing 4.4 shows the output of the program.

We see that after 3 *factorial* calls the traces diverge and one of them takes the "else" path and makes one more recursive call (lines 19-30). The other taken the "then" path (lines 31-33). After than one extra recursive call both traces immediately reconverge (line 34).

```

1          COMMON:
2 Event: FunctionEnter
3 "~" at "factorial.js" @@ 0,-1
4          COMMON:
5 Event: FunctionEnter
6 "~factorial" at "factorial.js" @@ 18,-1
7          COMMON:
8 Event: IfStmtElse
9 "~factorial" at "factorial.js" @@ 35,-1
10         COMMON:
11 Event: FunctionEnter
12 "~factorial" at "factorial.js" @@ 18,-1
13         COMMON:
14 Event: IfStmtElse
15 "~factorial" at "factorial.js" @@ 35,-1
16         COMMON:
17 Event: FunctionEnter
18 "~factorial" at "factorial.js" @@ 18,-1
19                                     RIGHT:
20 Event: IfStmtElse
21 "~factorial" at "factorial.js" @@ 35,-1
22                                     RIGHT:
23 Event: FunctionEnter
24 "~factorial" at "factorial.js" @@ 18,-1
25                                     RIGHT:
26 Event: IfStmtThen
27 "~factorial" at "factorial.js" @@ 35,-1
28                                     RIGHT:
29 Event: FunctionExit
30 "~factorial" at "factorial.js" @@ 18,-1
31         LEFT:
32 Event: IfStmtThen
33 "~factorial" at "factorial.js" @@ 35,-1
34         COMMON:
35 Event: FunctionExit
36 "~factorial" at "factorial.js" @@ 18,-1
37         COMMON:
38 Event: FunctionExit
39 "~factorial" at "factorial.js" @@ 18,-1
40         COMMON:
41 Event: FunctionExit
42 "~factorial" at "factorial.js" @@ 18,-1
43         COMMON:
44 Event: FunctionExit
45 "~" at "factorial.js" @@ 0,-1

```

Listing 4.4: Trace diff of *factorial* run with n=3 and n=4

The next example demonstrates that sometimes it is possible to find a divergence even when built-in functions are involved. Listing 4.5 is a short example involving *Array.prototype.map*. In the first run it was called with an array of length 3, in the second the array had one less item. Listing 4.6 shows the diff. We can see that in one run *multiplyByTwo* is entered 2 times and in the other 3 times. Naturally, if the arrays were of the same length, but with different content, this difference would not be caught this way.

```

1 const multiplyByTwo = (x) => 2 * x;

```

```

2
3 const nat = [...Array(2).keys()];
4
5 const even = nats.map(multiplyByTwo);

```

Listing 4.5: Simple JavaScript example of *Array.prototype.map* usage

```

1          COMMON:
2 Event: FunctionEnter
3 "~" at "map.js" @@ 0,-1
4          COMMON:
5 Event: FunctionEnter
6 "~multiplyByTwo" at "map.js" @@ 22,-1
7          COMMON:
8 Event: FunctionExit
9 "~multiplyByTwo" at "map.js" @@ 22,-1
10         COMMON:
11 Event: FunctionEnter
12 "~multiplyByTwo" at "map.js" @@ 22,-1
13         COMMON:
14 Event: FunctionExit
15 "~multiplyByTwo" at "map.js" @@ 22,-1
16         RIGHT:
17 Event: FunctionEnter
18 "~multiplyByTwo" at "map.js" @@ 22,-1
19         RIGHT:
20 Event: FunctionExit
21 "~multiplyByTwo" at "map.js" @@ 22,-1
22         COMMON:
23 Event: FunctionExit
24 "~" at "map.js" @@ 0,-1

```

Listing 4.6: Trace diff of *Array.prototype.map* run with array of length 3 and 2

The last example demonstrates generator-related events in action. Listing 4.7 shows an example of factorial rewritten to use generators. It was run two times, with $n = 2$ and $n = 3$. Listing 4.8 shows the resulting diff. We can see that when the generator is created, "Function Enter" event is emitted, followed by "Generator Suspend" (lines 7-12). Events "Generator Enter" and "Generator Yield" are used only when *Generator.prototype.next* is called. Last, but not least, "Generator Yield" is always followed by "Generator Suspend".

```

1 function *factorial(n) {
2   let acc = 1;
3   while (n > 1) {
4     acc = acc * n;
5     n = n - 1;
6     yield [acc, n];
7   }
8 }
9
10 function print_factorial(num) {
11   const gen = factorial(num);
12
13   while (true) {
14     const ret = gen.next();
15     if (ret.done) break;
16     const [acc, n] = ret.value;
17     console.log(`Acc: ${acc}, n: ${n}`);
18   }
19 }

```

```

20
21 print_factorial(2);

```

Listing 4.7: Example of JavaScript code using generators

```

1          COMMON:
2 Event: FunctionEnter
3 "~" at "generators.js" @@ 0,-1
4          COMMON:
5 Event: FunctionEnter
6 "~print_factorial" at "generators.js" @@ 143,-1
7          COMMON:
8 Event: FunctionEnter
9 "~factorial" at "generators.js" @@ 19,-1
10         COMMON:
11 Event: GeneratorSuspend
12 "~factorial" at "generators.js" @@ 19,-1
13         COMMON:
14 Event: GeneratorEnter
15 "~factorial" at "generators.js" @@ 19,-1
16         COMMON:
17 Event: GeneratorYield
18 "~factorial" at "generators.js" @@ 96,-1
19         COMMON:
20 Event: GeneratorSuspend
21 "~factorial" at "generators.js" @@ 19,-1
22         COMMON:
23 Event: GeneratorEnter
24 "~factorial" at "generators.js" @@ 19,-1
25         RIGHT:
26 Event: GeneratorYield
27 "~factorial" at "generators.js" @@ 96,-1
28         RIGHT:
29 Event: GeneratorSuspend
30 "~factorial" at "generators.js" @@ 19,-1
31         RIGHT:
32 Event: GeneratorEnter
33 "~factorial" at "generators.js" @@ 19,-1
34         COMMON:
35 Event: FunctionExit
36 "~factorial" at "generators.js" @@ 19,-1
37         COMMON:
38 Event: IfStmtThen
39 "~print_factorial" at "generators.js" @@ 231,-1
40         COMMON:
41 Event: FunctionExit
42 "~print_factorial" at "generators.js" @@ 143,-1
43         COMMON:
44 Event: FunctionExit
45 "~" at "generators.js" @@ 0,-1

```

Listing 4.8: Trace diff of generator-using *factorial* run with n=2 and n=3

4.6. Noise filtering

One of the biggest challenges in the entire method is filtering the noise. The noise on the website comes in many different flavours:

- Some content can be based on Random Number Generators.

- Content may be dynamic and different with each refresh of the page, e.g. Facebook feed.
- Code that is not related to the website, e.g. trackers, analytics

The difficult part is to filter out as much data as possible without removing valuable information.

In the original article authors collected redundant traces, diffed together all positive traces (and then also negative traces) and generated a blacklist of execution differences that are guaranteed to be noise.

Here the approach is different. The positive traces are all put together and the common subtraces are extracted. The same thing is done with the negative traces. At the end the whole SMP and alignment analysis (see sections [4.3](#) and [4.4](#)) is run only on those common subtraces.

The effectiveness and limitations of this approach is discussed in the chapter [5](#).

Chapter 5

Evaluation

5.1. The pipeline

The goal of this work was to automate as many steps of the anti-adblocking analysis as possible. Ideally, the system should be able to read a list of websites, collect traces, analyze each of them and produce a meaningful result.

The entire system is brought together by a program written in Python. It is controlled by an extensive set of flags and logs each step to make troubleshooting easier. The pipeline comprises following steps:

- A website (or list of websites) to analyze is read as an argument or from a file when special flag is provided.
- If selected, 3 positive and 3 negative traces are collected for each website (before each browser run, all cookies are deleted). After each run, all unwanted traces are deleted (they may come from extensions or unwanted iframes).
- If selected, the analysis is run and results are saved in a selected location.
- Optionally, the traces are deleted.

The system counts the number of traces with execution differences. If there is at least one such pair of traces found, we assume that it is due to anti-adblocking activities. The output includes the diff of those traces as well as all unmatched ones.

5.2. Methodology

The methodology is the same as in the original article [46]. First, the effectiveness of the method is tested on sets of negative and positive examples. Then, the system is used to conduct a small study.

One important difference from the original approach is that, unless the page has a visible anti-adblocking reaction on the main page, the websites are always tested on some subpage. The reason is that some news pages show anti-adblocking warnings only in the articles.

The original authors composed negative examples set from 100 websites that do not display ads. The reasoning behind this is that they are guaranteed not to contain anti-adblocking scripts. However, such choice of examples as a benchmark seems flawed. One of the key challenges in this method is noise filtration. If there are no ads, then there are practically no sources of execution differences between runs with and without an adblocker. A better choice

of websites to evaluate on would be a mix of websites without ads and sites that display ads, but are known not to contain anti-adblockers. Unfortunately, finding such websites is not easy as it entails a detailed analysis of the entire code served by each site. Certainly not a task for one person to be done in a reasonable timespan. For this reason, here the negative examples are also only sites that do not serve ads, exactly 50 of them.

The original authors had an extensive list of websites with anti-adblockers collected in their previous studies (almost 700 positions). Unfortunately, they do not share that list. For this reason, a similar list of positive examples had to be created from scratch. It is a mix of websites encountered during daily browsing and sites appearing in the Polish anti-adblock filter list [32]. To avoid time-consuming code analysis at the stage of composing the list, all selected websites contain visible anti-adblock warnings. Other important note – the list contain at most handful of websites from Alexa 500 list [7] to keep the list usable for the small scale study.

The adblocking extensions have different policies when it comes to their official, default filters. Adblock Plus behaves politely and respects publishers’ right to ask users to show some support and disable the extension, whatever form it takes [4]. On the other hand, uBlock Origin employs a different rule. If the warning is not dismissible or is annoying, it will be blocked [26].

Since usage of Adblock Plus may expose a wider variety of adblock walls and warnings, it was chosen for the experiments.

5.3. Negative examples

The list of websites without ads, on which the system was tested is presented in the table 5.1. The number of execution differences found is listed in the last column.

Table 5.1: Results of system evaluation on negative examples – pages without ads

	URL	Diff. count
1	http://administracja.sgh.waw.pl/en/dsl/Pages/default.aspx	0
2	http://architektura.um.warszawa.pl	0
3	http://es6-features.org/	0
4	http://kulikowisko.pl/kursy-prawa-jazdy/	0
5	http://sejm.gov.pl/Sejm8.nsf/poslowie.xsp	0
6	http://www.klubfocus.pl/o-nas	0
7	http://www.mdkopoczno.pl/index.php?cat=historia-mdk	0
8	http://www.mit.edu/research/	0
9	http://www.niebokopernika.pl/repertuar/	0
10	http://www.nsa.gov.pl/sprawozdania-roczne.php	0
11	https://amu.edu.pl/wiadomosci/aktualnosci/doktoranci/rekrutacja-do-szkoly-doktorskiej-uam2	0
12	https://arxiv.org/archive/astro-ph	0
13	https://awa.zh.ch/internet/volkswirtschaftsdirektion/awa/en/standortfoerderung/lebensraum/steuern.html	0
14	https://developer.mozilla.org/en-US/docs/Web/JavaScript/EventLoop	0
15	https://docs.djangoproject.com/pl/2.2/intro/	0
16	https://docs.python.org/3/whatsnew/3.7.html	0
17	https://mansfeld.pl/webdesign/strony-one-page-zalety-wady-mity/	0
18	https://maps.qed.ai/map/browse?page=1&selected_facets=region%3AAfrica	0
19	https://phillipreeve.net/blog/comparison-sigma-art-35mm-1-2-35mm-1-4-and-40mm-1-4/	0
20	https://pl.wikipedia.org/wiki/Bokeh	0

21	https://theoatmeal.com/comics/how_walk_human	0
22	https://uni.wroc.pl/przewodnik-dla-pierwszorocznych/	0
23	https://v8.dev/docs/webassembly-opcode	0
24	https://wiki.archlinux.org/index.php/Zsh	0
25	https://www.chromium.org/for-testers/bug-reporting-guidelines	0
26	https://www.fbi.gov/news/stories/rcfls-follow-the-modern-evidence-trail-081219	0
27	https://www.gatesfoundation.org/What-We-Do/Global-Health/Malaria	0
28	https://www.haskell.org/community/	0
29	https://www.icrc.org/en/war-and-law/treaties-customary-law/geneva-conventions	0
30	https://www.il.pw.edu.pl/index.php/pl/studia/i-stopnia-inzynierskie/kierunkowe-efekty-ksztalcenia	0
31	https://www.kgof.edu.pl/rejestracja/	0
32	https://www.mimuw.edu.pl/rekrutacja	0
33	https://www.napaluchu.waw.pl/czekam_na_ciebie/zasady_adopcji/	0
34	https://www.nsa.gov/what-we-do/cybersecurity/	0
35	https://www.oi.edu.pl/l/35/	0
36	https://www.oie.int/en/for-the-media/onehealth/	0
37	https://www.planetarium.edu.pl/noc-spadajacych-gwiazd-w-parku-slaskim/	0
38	https://www.pwpw.pl/Strefa_wiedzy/Zabezpieczenia/zabezpieczenia-dokumentow.html	0
39	https://www.stanford.edu/research/	0
40	https://www.w3.org/TR/css-grid-1/	0
41	https://www.who.int/immunization/newsroom/new-measles-data-august-2019/en/	0
42	https://www.word.waw.pl/egzaminy/wolne-terminy	0
43	https://www.zoo.lodz.pl/zwiedzanie-zoo/godziny-otwarcia/	0
44	http://kinoelektronik.pl/mazowieckie-centrum-kultury-medialnej-kultura-edukacja-sztuka-filmowa/	1
45	https://ethz.ch/en/research.html	1
46	https://szkolaimpro.pl	1
47	https://www.paperswithcode.com/about	1
48	https://www.whitehouse.gov/briefings-statements/remarks-president-trump-mass-shootings-texas-ohio/	1
49	https://www.worldanimalprotection.org/our-work/protecting-animals-communities	1
50	https://zoo.waw.pl/edukacja/oferta-edukacyjna	1

Ideally, there should be no execution differences on any website listed here. Nonetheless, a few differences were identified (positions 44-50). Let us have a look why.

kinoelektronik.pl

The origin of the difference is unclear. There does not seem to be any element blocked by AdBlock. Also, there are no blocked network requests. But for some reason there are differences in jQuery traversals.

ethz.ch

The cause of the difference is the tracking code.

szkolaimpro.pl

The difference is due to the progress bar.

www.paperswithcode.com

Execution divergence stems from Sentry – an error reporting solution.

www.whitehouse.gov

The origin of the difference is unclear.

www.worldanimalprotection.org

Similar case to kinoelektronik.pl. The difference is reflected in jQuery traversals, but it is unknown why the DOM trees are different.

zoo.waw.pl

The same case as kinoelektronik.pl and www.worldanimalprotection.org.

5.4. Positive examples

The list of websites with visible anti-adblock reactions is presented in the table 5.2. Each row consists of the URL, number of traces with execution differences, relevance of the differences and the anti-adblocker type.

Table 5.2: Results of system evaluation on positive examples – pages with adblock walls

	URL	Diff. count	Diff rel.?	Type
1	http://jakimasklad.pl/	0	N	BlockAdBlock + eval
2	http://www.srebrnestawy.pl/	0	Y	Custom, var set in file
3	https://www.cyberciti.biz/faq/ram-size-linux/	1	N	Custom, var set in file
4	http://zmiksowani.pl/	1	N	Custom, element-based bait
5	https://torrentcity.pl/torrenty/opis/id/326075/age_of_wonders_planetfall_-_premium_edition_2019_-_v1_004_36544_dlcs__bonus_content_multi8-pl_gog_exe	1	N	Custom, element-based bait + error handler
6	http://www.nadajemy.pl/	1	Y	Custom, var set in file
7	https://darkw.pl/	1	Y	Custom, check ads visibility
8	https://gragieldowa.pl/notowania/akcje/GR_WIG20	1	Y	Custom, check ads visibility
9	https://waw4free.pl/	1	Y	Adblock Detect
10	https://www.gry-planszowe.pl/forum/	1	Y	Custom, element-based bait
11	https://zw.pl	1	Y	Custom, element-based bait
12	http://www.audiklub.org/dane/a6/c6/color/100.html	2	Y	Custom, var set in file
13	https://e-dokument.org/sadowe-zazalania/zazalenie-powoda-na-postanowienie-sadu-odmawiajace-podjecia-zawieszonego-postepowania/	2	Y	BlockAdBlock + eval
14	https://kursbootstrap.pl/tutorial/	2	Y	Custom, element added in file
15	https://napisy24.pl/	2	Y	BlockAdBlock
16	https://who-called.eu/pl/178533310	3	Y	Custom, check ads visibility
17	https://www.wgrane.pl/	3	Y	Custom, element added in file
18	https://wzory-cv.com/przyklady.html	3	Y	BlockAdBlock + eval
19	https://megane.com.pl/	4	Y	Custom, check ads visibility
20	https://polscygracze.pl/	4	Y	BlockAdBlock + fallback
21	https://ogloszenia.plock.pl/	5	Y	Custom, var set in file
22	https://take-kino.pl/	5	Y	Custom, check ads visibility
23	https://www.pajacyk.pl	7	Y	Custom, element added in file
24	https://www.comperia.pl/	8	N	Custom, check ads visibility

25	https://komorkomania.pl/37730,dji-osmo-mobile-3-oficjalnie-jest-skladany-i-tanszy-od-poprzednika?src01=c801b	9	Y	Custom, check ads visibility
26	https://manjaro.pl/kilka-slow-o-nieoficjalnych-repozytoriach-arch-linux/	9	Y	Abd_Detector, element-based bait
27	https://maciejgalas.fotoblogia.pl/11821,leica-summicron-m-35-mm-f-2-asph	10	Y	Custom, check ads visibility
28	https://www.magazynbieganie.pl/fotostory/bieganie-dla-poczatkujacych-3-proste-sposoby-na-poprawe-formy/	10	Y	Abd_Detector, element-based bait
29	https://kulturalnemedi.pl/muzyka/lindsey-stirling-na-dwoch-koncertach-w-polsce/	11	Y	an_message_display, check ads visibility
30	https://www.filmweb.pl/film/John+Wick-2014-698144	11	Y	Custom, var set in file
31	https://www.telepolis.pl/artykuly/testy-sprzetu/honor-20-pro-smartfon-telefon-test-recenzja	12	Y	Custom, var set in file
32	http://lubliniec.info/artykul/12377/ostrezenie-meteorologiczne-drugiego-stopnia-dla-powiatu-lublinieckiego	13	Y	Custom, var set in file
33	http://www.krs-online.com.pl/	13	Y	Custom, var set in file
34	http://www.serialosy.pl/zagraniczne-seriale9/agenci-tarczy/24685-agenci-tarczy-s06e11-napisy-pl.html	13	Y	BlockAdBlock + eval
35	https://kimcartoon.to/Caroon/Squidbillies-Season-12	15	Y	Custom, loads scripts
36	https://pomorska.pl/100-najpopularniejszych-stron-internetowych-w-polsce-lista/ar/7198080	16	Y	Custom, element added in file
37	https://www.businessinsider.com/cheapest-michelin-star-lunch-new-york-city-2018-10	19	Y	BlockAdBlock
38	https://polskiedzieje.pl/polecane-ksiazki/podlary-i-wolnosc-dariusz-terlecki.html	23	Y	an_message_display, check ads visibility
39	https://jbzdy.net/	24	Y	Custom, var set in file
40	http://psychologia-ogolna.wyklady.org/	25	Y	BlockAdBlock + eval
41	https://mmorpg.org.pl/	35	N	BlockAdBlock + fallback
42	https://applemobile.pl/recenzja-bose-sleepbuds-1200-zlotych-za-sluchawki-przez-ktore-nie-posluchasz-wlasnej-muzyki/	36	Y	an_message_display, check ads visibility
43	https://burzowo.info/	41	Y	Custom, check ads visibility
44	https://www.cinemablend.com/	46	Y	Admiral
45	http://chelmki.eu/bedzie-wiecej-polaczen-kolejowych/	49	Y	KillAdBlock
46	https://player.pl/seriale-online/na-wspolnej-1001-1500-odcinki,5134/odcinek-1163,S03E1163,149527	60	N	Custom, loads scripts
47	https://www.polska-ie.com/	94	N	Custom, element-based bait
48	https://www.motocykl-online.pl/artykuly/Prawo-jazdy-na-motocykl-2019-kat-A-A1-A2-Jak-je-zrobic,12709,1	170	Y	BlockAdBlock
49	https://www.sonyalpharumors.com/the-first-sony-a7riv-versus-a7riii-size-comparison/	181	Y	an_message_display, check ads visibility
50	https://www.pasazer.com/	215	Y	Custom, var set in file

In most cases the system is able to find execution differences, which is a very good sign. In the following sections we will analyze when the system fails and discuss how the detected solutions work.

The analysis of 50 websites incorporating adblock walls allows us to create a simple classification of anti-adblocking mechanisms:

- Bait-based

- File as a bait
 - * Variable value is changed
 - * An element injected into DOM
 - * Error handler on network request
- DOM element as a bait
- Real ads checks
 - Visibility inspection
 - Load verification

We can also enumerate all encountered solutions:

- Custom scripts
- Off the shelf modules:
 - BlockAdBlock [35]
 - Adblock Detect [18]
 - Kill AdBlock [8]
 - Admiral’s Recover [6]
 - *an_message_display*
 - *Adb_Detector*

The most popular solutions are custom scripts and open-source BlockAdBlock (already mentioned in 1.3). Other open-source scripts include Adblock Detect and KillAdBlock. One website also used a paid solution – Admiral platform. The origin of the last two solutions is unclear, but the code was clearly the same (the names listed are distinctive functions names, not the real names of those modules)

Some custom scripts can be really widespread – some publishers own a number of domains and it makes sense to create one robust solution and use it in every website. One example is Wirtualna Polska group (positions 25 and 27), which uses their own custom anti-adblocking module.

Another important note – in most cases a significant part of differences found originated in third-party ads or tracking code (Google Analytics, AdWords, Gemius trackers, Twitter buttons, Facebook buttons etc.), but that was not inspected further.

5.4.1. Websites with unsatisfactory results

[jakimasklad.pl](#)

Brief analysis of the code shows that the site incorporates BlockAdBlock [35]. Turns out that, in this case, the module’s obfuscation is effective enough to circumvent our system. It is interesting that a number of different websites use the same solution and the system was able to pinpoint the differences in those cases.

www.srebrnestawy.pl

In this case the code that activates the warning seems very simple and there is no reason why it should not be detected by the system. Listing 5.1 shows the code.

```
1 $(function(){
2     ...
3     if (advertPolicy === false && typeof window.advertChecker === "undefined"
4     ) {
5         browserAlert.append(
6             '<div class="alert alert-warning alert-dismissible">' +
7             '<div class="container">' +
8             '<a href="#" class="close" data-dismiss="alert" aria-
9             label="close" title="close">x</a>' +
10            '<div class="alert-body">
11                <div class="text">
12                    <p>Do prawidłowego działania i wyświetlania się strony
13                    należy wyłączyć wtyczkę adblocka z Twojej przeglądarki!</p>
14                </div>
15            </div>
16        </div>');
17     }
18     ...
19 }
```

Listing 5.1: Anti-adblocking script of www.srebrnestawy.pl

The reason why it was not detected is that the bait was based on file *advert.js*, which is not loaded even when Adblock is disabled, due to a HTTP 404 error. Consequently, the warning is displayed even when Adblock is disabled and the system is right – there are no execution differences.

www.cyberciti.biz

In this case the difference found corresponds to PageAd reaction (it is good that it was found) but the check that activates the adblock warnings was not found.

Anti-adblocker used on this page is simple, works in the same way as the one found on site www.audiklub.pl. This site, however, uses Cloudflare's Rocket Loader [17].

Rocket Loader is a powerful mechanism developed by Cloudflare that is able to cache JavaScript code and defer its execution until the entire webpage is rendered. The way it works is the following:

- Before serving the page, the server inspects the website and changes all JS scripts to some type not identified by the browser as JavaScript code. The type also has to be unique so that those scripts are easy to find later.
- The server also encloses a special JavaScript module that activates when the entire webpage is loaded.
- JavaScript module finds all scripts with changed type and executes them.

Since the module that executes the scripts is a JavaScript code, the system should still be able to detect the anti-adblocking code. Unfortunately, it cannot and the most probable reason seems to be a bit aggressive noise filtering. If some function does a DOM traversal and the content changes a little on each pages load, it is probable that it will be filtered out. It is suspected that it is the case here. i.e. Rocket Loader has to find all scripts (DOM traversal) and execute them as part of one long function.

[torrentcity.pl](#)

This is an example of a rather sophisticated custom anti-adblocking script. Listing 5.2 shows the code that performs the check.

```
1 function e() {
2     var t = "adsSupported";
3     var r = window.document.createElement("script");
4     r.setAttribute("async", "async");
5     r.setAttribute("src", c + "ads.js");
6     r.addEventListener("error", function() {
7         f(false);
8     });
9     r.addEventListener("load", function() {
10        var e = 0;
11        function r() {
12            if (!window[t]) {
13                if (e++ < 10)
14                    window.setTimeout(r, 100);
15                else
16                    f(false);
17            } else {
18                setTimeout(function() {
19                    if (!window[t]) {
20                        f(false);
21                    } else {
22                        f(window[t].offsetHeight !== 0);
23                        if (window[t].parentElement)
24                            window[t].parentElement.removeChild(window[t]);
25                        return;
26                    }
27                }, 50)
28            }
29        }
30        r();
31    });
32    window.document.body.appendChild(r);
33 }
```

Listing 5.2: Anti-adblocking script of [torrentcity.pl](#)

The script first sets up a bait by injecting a new script element into DOM. It is set up to download a file named *ads.js*, which should be blocked by any adblocker.

Next, the error handler is set up to activate the reaction. If the load succeeds, it still makes sure that are are actually displayed.

Functions called as event handlers or via *setTimeout* are a weak spot of Differential Execution Analysis method, not just this exact implementation and there is no known way to automatically utilize such unpaired events.

[player.pl](#)

An example of a really robust solution, probably used on every site belonging to TVN media company. Listing 5.3 shows the checks. The website utilizes more than one check to increase detection effectiveness. The checks itself are similar to what we have already seen. The first one creates an element as a bait and check if it is visible. The second one uses different bait, a file, and reacts within callbacks. The system fails here because the script uses those techniques that are the weak spot of the presented method.

```

1 var isAdBlockEnabled = function() {
2     if (void 0 === window.isAdBlockEnabledCheckCounter && (window.
3         isAdBlockEnabledCheckCounter = 0),
4         !document.body)
5         return window.isAdBlockEnabledCheckCounter < 15 && setTimeout(
6             function() {
7                 window.isAdBlockEnabledCheckCounter++,
8                 isAdBlockEnabled()
9             }, 10),
10         window.adBlockEnabled = !1;
11     var e = document.createElement("div");
12     e.setAttribute("class", "sponsored-ad adoSlave advertisement pub_300x250
13     pub_300x250m pub_728x90 text-ad textAd text_ad text_ads text-ads text-ad-
14     links"),
15     e.setAttribute("style", "width: 1px !important; height: 1px !important;
16     position: absolute !important; left: -1000px !important; top: -1000px !
17     important;"),
18     document.body.appendChild(e);
19     var t = $(e)
20     , i = $(e).is(":visible");
21     return t.remove(),
22     window.adBlockEnabled = !i,
23     !i
24 };
25 // ...
26 v.isAdBlockEnabledAgain = function(e, t) {
27     var i = nuvi.settings.accessProtocol + "//s1-player5.cdntvn.pl/advert.js"
28     ;
29     -1 !== i.indexOf("tvwisla.com.pl") && -1 === i.indexOf("cdntvn.stage.
30     online") && (i = (i = i.replace("player5", "s1-player5")).replace(".stage.
31     online", "-cdntvn.stage.online")),
32     $.ajax({
33         url: i,
34         type: "GET",
35         cache: !1,
36         async: !1,
37         success: function() {
38             return v.adblockEnabledAgain = !1,
39             "function" == typeof e && e(),
40             !1
41         },
42         error: function() {
43             return v.adblockEnabledAgain = !0,
44             v.options().adBlockCheck && !tvn24ScrollTvnCheck() ? "function"
45             == typeof t && t(v) : "function" == typeof e && e(),
46             !0
47         }
48     })
49 }

```

Listing 5.3: Anti-adblocking script of [player.pl](#)

Other websites

In case of other websites where relevant execution divergence was not found the reasons were the same – either the check was intertwined between other activities in a way that filtering would discard it or callbacks were utilized.

5.4.2. Detected anti-adblockers

This section presents the interesting methods that have not been covered yet, but have been detected. Simple examples were preferred to avoid showing a cluttered code.

www.audiklub.pl

This is the simplest anti-adblocking script one can think of. The website includes just one advertisement, but when an adblocker is detected, the entire content is blocked (listing 5.4).

```
1 window.onload = function() {
2     if( window.canRunAds === undefined ){
3         // adblocker detected, show fallback
4         $('div#contentDiv').html("<b>Wylacz adblock na tej stronie. Tu jest
        tylko jedna mala niewinwazyjna reklama.</b>");
5         //showFallbackImage();
6     }
7 }
```

Listing 5.4: Anti-adblocking script of www.audiklub.pl

First, the trap is set up. It is a file named *ads.js*. Its content is just one line, which sets *canRunAds* variable to true.

The main check is performed in the *window.onload* callback. The script simply checks if *canRunAds* variable has been set to true. If not, it replaces the content div with an anti-adblocker warning.

kursbootstrap.pl

The script on this website serves as an example of an anti-adblocker that uses file as a bait but the check is performed by injecting an element into DOM instead of setting a global variable (listing 5.5).

```
1 // In "advertisement.js"
2 document.write('<div id="tester">an advertisement</div>');
3
4 // In the main page, inline
5 if (document.getElementById("tester") != undefined) {
6     document.write('<div class="alert alert-success alert-dismissible" role="
        alert"><button type="button" class="close" data-dismiss="alert"><span aria
        -hidden="true">&times;</span><span class="sr-only">Zamknij</span></button
        ><strong>Dziekuje!</strong> Widze, ze nie uzywasz AdBlocka. To dobrze:)
        Reklamy sa nieodlacznna czescia darmowych stron i pozwalaja autorom na ich
        rozwoj. Jesli doceniasz prace innych, nie uzywaj AdBlocka na tych
        witrynach.</div>');
7 } else {
8     document.write('<div class="alert alert-danger alert-dismissible" role="
        alert"><button type="button" class="close" data-dismiss="alert"><span aria
        -hidden="true">&times;</span><span class="sr-only">Zamknij</span></button
        ><strong>Oj! niedobrze:</strong> Wyglada na to, ze uzywasz AdBlocka.
        Reklamy na moim blogu nie sa inwazyjne i w zaden sposob nie przeszkadzaja
        w odbiorze tresci. Reklamy sa nieodlacznna czescia darmowych stron i
        pozwalaja autorom na ich rozwoj. Jesli to mozliwe wylacz AdBlocka dla tej
        domeny. Dziekuje!</div>');
9 }
```

Listing 5.5: Anti-adblocking script of kursbootstrap.pl

[who-called.eu](#)

The code from this website demonstrates how anti-adblocking script can detect if real ads are displayed, without using a bait (listing 5.6). The detecting function is run 5 seconds after the page was loaded. It then checks if the content of the element displaying ads is empty (contains 0 non-whitespace characters).

```
1 window.onload = function () {
2     setTimeout(function () {
3         var ad = document.querySelector("ins.adsbygoogle");
4         if (ad && ad.innerHTML.replace(/\s/g, "").length == 0) {
5             $('#no-ads-info').show();
6         }
7         console.log('jest ok');
8     }, 5000);
9 };
```

Listing 5.6: Anti-adblocking script of [who-called.eu](#)

[kimcartoon.to](#)

This website utilizes a check based on making sure that advertising scripts are loaded (listing 5.7). It seems, though, that the example has a bug. Perhaps it would be desirable to invert the conditional and change the variable only when one of the scripts fails to load (currently it is enough that one of them is loaded correctly)

```
1 if (adbWarn == null) {
2     $.ajaxSetup({
3         cache: false
4     });
5     var alb = true;
6     $.getScript('https://pubmatic.com/wp-content/themes/pubmatic/js/jquery.
7 alignHeight.js?ver=1.0').done(function(script, textStatus) {
8         alb = false;
9     });
10    $.getScript('https://propellerads.com/wp-content/plugins/radiantthemes-
11 addons/tabs/js/radiantthemes-tab-element-four.js').done(function(script,
12 textStatus) {
13        alb = false;
14    });
15    $.getScript('https://www.bebi.com/js/plugins.js').done(function(script,
16 textStatus) {
17        alb = false;
18    });
19    setTimeout(function() {
20        if (alb) {
21            $('#adbWarnContainer').css({
22                display: "block"
23            });
24        }
25    }, 8000);
26 }
```

Listing 5.7: Anti-adblocking script of [kimcartoon.to](#)

[e-dokument.org](#) and others using BlockAdBlock with *eval* obfuscation

BlockAdBlock is one of the most popular off-the-shelf anti-adblocking solutions. The obfuscated code is quite unreadable to a human (listing 5.8). Fortunately, the string to be evaluated

contains "BlockAdBlock" and we can be sure that it was the solution used. Since BlockAdBlock is an open-source module, interested reader can check out its code on github [35].

```

1 // Place this code snippet near the footer of your page before the close of
  the /body tag
2 // LEGAL NOTICE: The content of this website and all associated program code
  are protected under the Digital Millennium Copyright Act. Intentionally
  circumventing this code may constitute a violation of the DMCA.
3 eval(function(p, a, c, k, e, d) {
4   e = function(c) {
5     return (c < a ? '' : e(parseInt(c / a))) + ((c = c % a) > 35 ? String
      .fromCharCode(c + 29) : c.toString(36))
6   };
7   if (!''.replace(/~/, String)) {
8     while (c--) {
9       d[e(c)] = k[c] || e(c);
10    }
11    k = [ function(e) { return d[e] } ];
12    e = function() { return '\\w+' };
13    c = 1;
14  }
15  while (c--) {
16    if (k[c]) {
17      p = p.replace(new RegExp('\\b' + e(c) + '\\b','g'), k[c]);
18    }
19  }
20  return p;
21 }(,q,Q='\\',27='21\\';1N(q i=0;i<12...' ) // the string is truncated

```

Listing 5.8: Anti-adblocking script of e-dokument.org

polscygracze.pl

Another website that uses BlockAdBlock, but adds a fallback in case the module is blocked by an adblocker (listing 5.9). It is worth mentioning that the idea of making the id of adblock warning element random is good, but the string should be dynamic. If it is static it can still be added to adblockers' filter lists.

```

1 function adBlockDetected() {
2   $('#07th202231').show();
3 }
4 function adBlockNotDetected() {
5   $('#07th202231').hide();
6 }
7 if (typeof blockAdBlock === 'undefined') {
8   adBlockDetected();
9 } else {
10  blockAdBlock.onDetected(adBlockDetected).onNotDetected(adBlockNotDetected);
11 }

```

Listing 5.9: Anti-adblocking script of polscygracze.pl

Other off-the-shelf solutions

Other solutions utilize similar mechanisms to those already described. They can be more polished and use more than one check, but they do introduce new ideas. For this reason they are not described further here. Interested reader can check their documentation.

5.5. Small scale study

The experiment has been conducted on almost 300 top websites from Alexa 500 list (the version for Poland) [7]. The list of visited websites and results is attached in Appendix A. Some websites were skipped because they contain adult content or their main domain do not contain any website, e.g. content delivery networks. Those websites, instead of URLs, have comments (lines starting with –).

Some results are marked with "TO" (timeout) instead of a number of differences found. This is because there is a limited time for each trace collection (30 seconds). After that the browser is closed. In some cases the browser times out instead of closing. The system then waits another 90 seconds and closes the browser forcefully. Usually this is a sign of a website producing enormously large execution trace (some can reach almost 50GB in 2 minutes). Such websites are also skipped.

Although the number of ads, social media links and similar elements was not counted, it seems that the more elements of this kind a website incorporates, the more likely it is that there will be an anti-adblocking reaction of some kind. It is especially true in case of third-party ads, which more often than not report an adblocker usage.

Out of 281 visited websites, 33 contained visible anti-adblocking warnings. The system detected anti-adblocking reactions in 163 of them. 103 were found to be free of anti-adblockers. The rest (15 pages) timed out. Looking at the results, it seems that anti-adblocking scripts are widespread. Silent anti-adblocking reactions are over 5 times more frequent than adblock walls (163 reactions detected vs 33 adblock walls detected). Over 50% of visited websites exhibits some anti-adblocking reactions. It is not that surprising if we consider that most websites contain advertisements and the biggest ads providers report adblock usage.

5.6. Conclusion

The performed tests show that the system is capable of detecting anti-adblocking scripts in many scenarios. The exact numbers are not presented purposefully – the tests should be conducted on larger set of websites and the methodology should be different when it comes to negative examples.

The presented solution certainly has some weak spots. The first one is event coverage. It is easy to hide anti-adblockers by utilizing callbacks. The second one is the data filtering, which in some cases may filter out too much data, and in others it lets through too many traces, which limits its usefulness (the analysis becomes much harder, the optimum would be to always have less than 10 trace diffs for manual verification).

Here are some ideas that might alleviate both problems:

- Make an extension that would put a thin wrapper around some built-in functions. Such wrapper would only call the original built-in function, but the entry and exit would be logged by the system. It might be useful for functions such as *getElementById* or direct DOM manipulations.
- Manually analyze and tag the results of more traces (at least hundreds of them) and try some machine learning method to filter resulting diffs. A first try could be to use random decision forests.

The system is also a bit too slow. The trace collection should have less overhead on the browser. For large scale study spending 10 minutes on each website to collect traces is too long. One idea is to, instead of constructing a strings map during analysis, have Chrome

construct it. This, in conjunction with a use of some binary format, will result in much more compact files and less time wasted on writing to files.

Last but not least, the conducted experiments give lots of insights on how to build an effective anti-adblocker. Here are the tips:

- Use more than one detection method, preferably utilizing callbacks.
- Obfuscate the code by using minification combined with eval to make manual analysis extremely hard.
- Randomize ids of injected adblock warnings to make it harder for adblockers to filter them out.

Listings

2.1. Dynamic instrumentation in JavaScript	11
3.1. Calculating factorial in JavaScript	16
3.2. Ignition bytecode for function <i>factorial</i>	16
3.3. V8's output for <i>factorial</i> with <i>--trace</i> flag	18
3.4. Ignition bytecode for function <i>factorial</i> with <i>--trace</i> enabled	18
3.5. Ignition bytecode for function <i>factorial</i> with <i>--trace-dea</i> enabled	20
3.6. V8's output for <i>factorial</i> with <i>--trace-dea</i> flag	21
4.1. Pseudocode of the trace untangling algorithm	29
4.2. Pseudocode of the trace alignment algorithm	30
4.3. Pseudocode of the Gale-Shapley (deferred acceptance) algorithm	31
4.4. Trace diff of <i>factorial</i> run with n=3 and n=4	33
4.5. Simple JavaScript example of <i>Array.prototype.map</i> usage	33
4.6. Trace diff of <i>Array.prototype.map</i> run with array of length 3 and 2	34
4.7. Example of JavaScript code using generators	34
4.8. Trace diff of generator-using <i>factorial</i> run with n=2 and n=3	35
5.1. Anti-adblocking script of www.srebrnestawy.pl	43
5.2. Anti-adblocking script of torrentcity.pl	44
5.3. Anti-adblocking script of player.pl	45
5.4. Anti-adblocking script of www.audiklub.pl	46
5.5. Anti-adblocking script of kursbootstrap.pl	46
5.6. Anti-adblocking script of who-called.eu	47
5.7. Anti-adblocking script of kimcartoon.to	47
5.8. Anti-adblocking script of e-dokument.org	48
5.9. Anti-adblocking script of polscygracze.pl	48

List of Figures

4.1. Memory usage of the first implementation using ByteStrings as internal strings representation	26
4.2. Memory usage after switching to ShortByteString and forcing eager evaluation	27
4.3. Memory usage after preventing parser from keeping the entire input file in the memory	28
4.4. Final memory usage	28

List of Tables

5.1. Results of system evaluation on negative examples – pages without ads	38
5.2. Results of system evaluation on positive examples – pages with adblock walls	40

Bibliography

- [1] Acceptable Ads initiative. Acceptable Ads. <https://acceptableads.com>, 2019. [Online; accessed 6-August-2019].
- [2] Adblock Analytics. Detect Adblock. <https://www.detectadblock.com>, 2019. [Online; accessed 29-July-2019].
- [3] Adblock authors. What is Adware? How to Identify and Prevent Adware. <https://blog.getadblock.com/what-is-adware-ce135d866898>, 2019. [Online; accessed 6-August-2019].
- [4] Adblock authors. Why doesn't Adblock bypass anti-adblock walls? <https://help.getadblock.com/support/solutions/articles/6000199143-why-doesn-t-adblock-bypass-anti-adblock-walls->, 2019. [Online; accessed 15-August-2019].
- [5] Adelina Tuca. 5 best anti adblock wordpress plugins. <https://themeisle.com/blog/best-anti-adblock-wordpress-plugins/>, 2018. [Online; accessed 6-August-2019].
- [6] Admiral. Admiral: A Proven Platform Helping Over 12,000 Publishers Worldwide. <https://getadmiral.com/products>, 2019. [Online; accessed 19-August-2019].
- [7] Alexa. Top Sites in Poland. <https://www.alexa.com/topsites/countries/PL>, 2019. [Online; accessed 4-August-2019].
- [8] badmetevils. Kill Adblock. <https://www.drupal.org/project/killadblock>, 2018. [Online; accessed 19-August-2019].
- [9] Baiju Muthukadan. Selenium with Python. <https://selenium-python.readthedocs.io>, 2019. [Online; accessed 6-August-2019].
- [10] Ben L. Titzer, Jaroslav Sevcik. A year with Spectre: a V8 perspective. <https://v8.dev/blog/spectre>, 2019. [Online; accessed 2-August-2019].
- [11] Ben Titzer et al. TurboFan JIT Design. <https://docs.google.com/presentation/d/1s0EF4MlF7Le07uq-uThJSulJlTh--wgLeaVibsbb3tc/edit#slide=id.p>, 2019. [Online; accessed 29-July-2019].
- [12] Bryan O'Sullivan. attoparsec: Fast combinator parsing for bytestrings and text. <http://hackage.haskell.org/package/attoparsec>, 2019. [Online; accessed 5-August-2019].
- [13] Christopher Elliott. Yes, There Are Too Many Ads Online. Yes, You Can Stop Them. Here's How. https://www.huffpost.com/entry/yes-there-are-too-many-ads-online-yes-you-can-stop_b_589b888de4b02bbb1816c297, 2017. [Online; accessed 6-August-2019].

- [14] Chrome team. Content Scripts. https://developer.chrome.com/extensions/content_scripts, 2019. [Online; accessed 2-August-2019].
- [15] Chrome team. Extensions Overview. <https://developer.chrome.com/extensions/overview>, 2019. [Online; accessed 2-August-2019].
- [16] Chrome team. Site Isolation. <https://www.chromium.org/Home/chromium-security/site-isolation>, 2019. [Online; accessed 2-August-2019].
- [17] Cloudflare Support. What does Rocket Loader do? <https://support.cloudflare.com/hc/en-us/articles/200168056-What-does-Rocket-Loader-do->, 2019. [Online; accessed 19-August-2019].
- [18] Damir Brekalo. Adblock detect. <https://dbrekalo.github.io/adblockdetect/>, 2019. [Online; accessed 19-August-2019].
- [19] Don Stewart, Duncan Coutts. bytestring: Fast, compact, strict and lazy byte strings with a list interface. <http://hackage.haskell.org/package/bytestring>, 2019. [Online; accessed 5-August-2019].
- [20] Felix Meier. Iroh - Dynamic code analysis for JavaScript. <https://maierfelix.github.io/Iroh/>, 2018. [Online; accessed 29-July-2019].
- [21] Franziska Hinkelmann. Understanding V8’s Bytecode. <https://medium.com/dailyjs/understanding-v8s-bytecode-317d46c94775>, 2017. [Online; accessed 1-August-2019].
- [22] Google LLC. Web Tracing Framework. <https://google.github.io/tracing-framework/index.html>, 2019. [Online; accessed 29-July-2019].
- [23] Haskell contributors. Haskell. <https://www.haskell.org>, 2019. [Online; accessed 5-August-2019].
- [24] IAB. IAB Internet Advertising Revenue Report 2018 Full Year Results. <https://www.iab.com/insights/iab-internet-advertising-revenue-report-2018-full-year-results/>, 2019. [Online; accessed 6-August-2019].
- [25] Noah Johnson, Juan Caballero, Kevin Zhijie Chen, Stephen McCamant, Pongsin Poosankam, Daniel Reynaud, and Dawn Song. Differential slicing: Identifying causal execution differences for security applications. In *Proceedings of the 32nd IEEE Symposium on Security and Privacy*, pages pp. 347–362. IEEE, May 2011.
- [26] Kaleigh Rogers. Why Your Ad Blocker Doesn’t Block Those ‘Please Turn Off Your Ad Blocker’ Popups. https://www.vice.com/en_us/article/j5zk8y/why-your-ad-blocker-doesnt-block-those-please-turn-off-your-ad-blocker-popups, 2018. [Online; accessed 15-August-2019].
- [27] Paul Kocher, Jann Horn, Anders Fogh, , Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz, and Yuval Yarom. Spectre attacks: Exploiting speculative execution. In *40th IEEE Symposium on Security and Privacy (S&P’19)*, 2019.
- [28] Moritz Lipp, Michael Schwarz, Daniel Gruss, Thomas Prescher, Werner Haas, Anders Fogh, Jann Horn, Stefan Mangard, Paul Kocher, Daniel Genkin, Yuval Yarom, and Mike Hamburg. Meltdown: Reading kernel memory from user space. In *27th USENIX Security Symposium (USENIX Security 18)*, 2018.

- [29] Mark Karpov. Short ByteString and Text. <https://markkarpov.com/post/short-bs-and-text.html>, 2017. [Online; accessed 5-August-2019].
- [30] Martin Brinkmann. Integrate Nano Defender with uBlock Origin to block Anti-Adblocker. <https://www.ghacks.net/2019/02/15/integrate-nano-defender-with-ublock-origin-to-block-anti-adblocker/>, 2019. [Online; accessed 6-August-2019].
- [31] MDN. Concurrency model and event loop. <https://developer.mozilla.org/en-US/docs/Web/JavaScript/EventLoop>, 2019. [Online; accessed 26-July-2019].
- [32] olegwukr. Polish Anti Adblock Filters. <https://github.com/olegwukr/polish-privacy-filters/blob/master/anti-adblock.txt>, 2019. [Online; accessed 15-August-2019].
- [33] PageFair. 2017 Adblock Report. <https://pagefair.com/blog/2017/adblockreport/>, 2017. [Online; accessed 6-August-2019].
- [34] Reek. Anti-Adblock Killer. <https://reek.github.io/anti-adblock-killer/>, 2019. [Online; accessed 6-August-2019].
- [35] sitexw. BlockAdBlock. <https://github.com/sitexw/BlockAdBlock>, 2018. [Online; accessed 6-August-2019].
- [36] StackOverflow users. Adding console.log to every function automatically. <https://stackoverflow.com/questions/5033836/adding-console-log-to-every-function-automatically>, 2017. [Online; accessed 25-July-2019].
- [37] uBlock Origin Contributors. This is uBlock’s manifesto. <https://github.com/gorhill/uBlock/blob/master/MANIFESTO.md>, 2015. [Online; accessed 27-July-2019].
- [38] V8 Team. Built-in functions. <https://v8.dev/docs/builtin-functions>, 2019. [Online; accessed 2-August-2019].
- [39] V8 team. Design of V8 bindings. https://chromium.googlesource.com/chromium/src/+master/third_party/blink/renderer/bindings/core/v8/V8BindingDesign.md, 2019. [Online; accessed 6-August-2019].
- [40] V8 Team. V8 Torque user manual. <https://v8.dev/docs/torque>, 2019. [Online; accessed 2-August-2019].
- [41] V8 Team. What is V8? <https://v8.dev>, 2019. [Online; accessed 29-July-2019].
- [42] W3Schools. Browser Statistics. <https://www.w3schools.com/browsers/>, 2019. [Online; accessed 29-July-2019].
- [43] Wikipedia contributors. Stable marriage problem — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Stable_marriage_problem&oldid=907225019, 2019. [Online; accessed 25-July-2019].
- [44] Wikipedia contributors. Wordpress — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=WordPress&oldid=909584536>, 2019. [Online; accessed 6-August-2019].

- [45] Bin Xin, William N. Sumner, and Xiangyu Zhang. Efficient program execution indexing. In *Proceedings of the 29th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '08, pages 238–248, New York, NY, USA, 2008. ACM.
- [46] Shitong Zhu, Xunchao Hu, Zhiyun Qian, Zubair Shafiq, and Heng Yin. Measuring and disrupting anti-adblockers using differential execution analysis. In *NDSS*. The Internet Society, 2018.

Appendix A

	URL	Vis. warn.	Diff. count
1	https://www.google.com/search?q=tom+cruise	N	35
2	https://www.youtube.com/watch?v=tp-wQYyrd0	N	34
3	https://www.google.pl/search?q=tom+cruise	N	35
4	https://allegro.pl/oferta/telewizor-led-55-4k-mpeg4-usb-3-0-smart-wi-fi-a-7568489180	N	13
5	https://www.facebook.com/BradPittPlanB/	N	5
6	https://www.olx.pl/oferta/hyundai-i30-1-6-crdi-90-km-CID5-IDATnhX.html	N	9
7	https://wiadomosci.wp.pl/mini-czarnobyle-seria-niefortunnych-wpadek-rosyjskiego-wojska-6412764769474689a	Y	TO
8	https://wiadomosci.onet.pl/nauka/zmarl-wybitny-rosyjski-astrofizyk-nikolaj-kardaszow-tworca-skali-rozwoju-cywilizacji/f96d48q	Y	8
9	https://pl.wikipedia.org/wiki/Bokeh	N	0
10	https://fakty.interia.pl/swiat/news-iranska-telewizja-iran-przejal-zagraniczny-tankowiec,nId,3130879	N	4
11	--Adultcontent	—	—
12	https://vk.com/club75379670	N	0
13	https://www.wiocha.pl/1589883,Zona-skarb	N	10
14	https://www.otomoto.pl/oferta/bmw-seria-8-840d-xdrive-coupe-upust-126000-pln-ID6CdinC.html	N	16
15	https://www.cda.pl/video/265923314/vfilm	N	2
16	https://www.aliexpress.com/item/32989785993.html	N	43
17	https://www.o2.pl/artykul/mlodzi-mniej-przejmujasie-alertami-rzadowego-centrum-bezpieczenstwa-6409157712651905a	Y	23
18	https://kobieta.gazeta.pl/kobieta/1,107881,25045075,pani-zacisnie-piesc-nie-dalam-rady-dlugie-paznokcie-mi-przeszkodzily.html	N	14
19	https://www.ceneo.pl/Torby_i_walizki	N	7
20	https://www.ipko.pl	N	0
21	https://www.reddit.com/r/europe/comments/cltm32/black_forest_germany/	N	TO
22	https://www.mbank.pl/indywidualny/	N	50
23	https://www.netflix.com/pl-en/	N	1
24	--Adultcontent	—	—
25	https://www.filmweb.pl/film/Top+Gun%3A+Maverick-2020-602101	Y	7
26	https://www.gumtree.pl/a-mieszkania-i-domy-do-wynajecia/krakow/kawalerka-do-wynajecia/1005563446690911255418809	N	2
27	https://www.centrum24.pl/	N	0
28	https://www.booking.com/hotel/pl/willa-nad-potokiem-ladek-zdroj.pl.html	N	TO
29	https://uwaga.tvn.pl/reportaze,2671,n/jak-polacy-pija-malpki,296314.html	N	6
30	https://www.twitch.tv/summoningsalt/videos	N	TO
31	--Adultcontent	—	—
32	https://www.instagram.com/paulnicklen/	N	6
33	https://www.wykop.pl/link/5074087/genialny-pomysl-pkp-dworzec-centralny-ma-wygladac-jak-w-latach-70/	N	3
34	https://www.yahoo.com/news/texas-woman-killed-by-police-officer-attempting-to-shoot-her-dog-083943364.html	N	TO

35	https://fili.cc/serial/czarnobyl/s01e01/12345/63706	N	0
36	https://zalukaj.com/zalukaj-film/48411/krol-lew-the-lion-king-2019.html	N	0
37	https://www.bankmillennium.pl	N	0
38	http://wyborcza.pl/7,75399,25056799,latajacy-czlowiek-tym-razem-zwyciezyl-pokonaj-kanal-la-manche.html	Y	6
39	https://zen.yandex.ru/media/clickoncar/skolko-tak-ne-sdelaet-nikogda-smotrim-na-hiphi-1-5d458a1c7cccba00ad56ec0a	N	0
40	https://chomikuj.pl/faworit2/Systemy+LINUX/Linux+Ubuntu-PL-ISO	N	1
41	http://paulinaphotographer.blogspot.com	N	1
42	https://www.pracuj.pl/praca/front-end-software-developer-wroclaw,oferta,6999667	N	2
43	https://www.money.pl/gospodarka/faktura-za-wizyte-prezydenta-mieszkaniec-szczecinka-oddal-pieniadze-ministrowi-6409902580459649a.html	N	20
44	--Adultcontent	—	—
45	https://kinokrad.co/336686-yesterday.html	N	0
46	https://www.google.ru/search?q=tom+cruise	N	9
47	--Adultcontent	—	—
48	https://www.fandom.com/articles/the-mandarin-shang-chi-father-mcu	N	1
49	https://www.roblox.com	N	0
50	https://news.mail.ru/politics/38207623/	N	138
51	https://www.otodom.pl/oferta/lux-2-pokoje-powisle-loggia-metro-park-ID41D1R.html	N	0
52	https://www.pekao24.pl	N	1
53	https://www.microsoft.com/pl-pl/p/surface-laptop-2/8XQJKK3DD91B	N	37
54	https://www.google.com.ua/search?q=tom+cruise	N	1
55	https://tygodnik.tvp.pl/43744932/albo-dzieci-albo-ekologia-jeden-nastolatek-mniej-to-ulga-dla-ziemi	N	5
56	https://www.pudelek.pl/artukul/149734/rihanna_atakuje_donalda_trumpa_po_strzelaninach_w_usa_latwiej_jest_tu_zdobyc_karabin_niz_wize/	N	6
57	https://naszemiaszto.pl/nawelizacja-prawa-budowlanego-coraz-blizej-znikna-absurdy-i/ar/c9-7282497	Y	8
58	https://corporate.payu.com	N	1
59	https://www.komputerswiat.pl/aktualnosci/sprzet/nawy-laptop-gamingowy-od-xiaomi-juz-oficjalnie-poznalismy-cene-urzadzenia/yzcmnje	Y	6
60	https://www.gemius.com/e-commerce-news/ukraine-users-attitude-illegal-content.html	N	0
61	https://rezka.ag/films/documentary/31527-vo-slavu-tmy-2019.html	N	2
62	https://sprzedajemy.pl/vw-ti-bulik-ogorek-klasykiem-do-slubu-reklama-eventy-bielsko-biala-2-85b084-nr58175417	N	1
63	https://www.fakt.pl/wydarzenia/polityka/posel-janusz-sanocki-wylamal-bramke-w-opolskim-urzedzie-wojewodzkiem-interweniowala/dp5sdx	Y	7
64	https://demotywatory.pl/4940852/Nie-wiem-czy-kierowca-autobusu-po-tym-klaskal-ale-warto	N	32
65	http://www.meteo.pl	N	0
66	https://joemonster.org/art/47440/Slynni-artysci_kiedys_i_dzis	N	8
67	http://www.gemius.pl/agencje-aktualnosci/wyniki-badania-gemiuspbi-za-kwiecien-2019.html	N	6
68	https://login.live.com	N	0
69	http://www.sport.pl/kolarstwo/7,64993,25061164,koszmarny-wypadek-na-tour-de-pologne-kolarz-jest-reanimowany.html	N	0
70	https://www.ing.pl	N	2
71	https://kwejk.pl/obrazek/3421509/strefa-wolna-od-lpg.html	Y	4
72	https://www.aliorbank.pl	N	0
73	--Adultcontent	—	—
74	https://www.elektroda.pl/rtvforum/topic3602578.html	Y	0
75	https://businessinsider.com.pl	Y	6
76	https://www.euro.com.pl/cms/tg-samsung-galaxy-a50-z-opaska.bhtml	N	9
77	https://www.kinopoisk.ru/media/podcast/3398593/	N	0
78	https://bnpparibas.pl/klienci-indywidualni	N	13

79	https://www.x-kom.pl/p/506331-smartfon-telefon-xiaomi-mi-a3-4-64gb-blue.html	N	4
80	https://www.zalando.pl/weekday-dakota-felted-jacket-kurtka-jeansowa-rocky-blue-web22t02l-k11.html	N	51
81	https://www.gry-online.pl/S018.asp?ID=2063	N	2
82	https://www.ing.pl	N	2
83	https://www.mediaexpert.pl/telewizory/telewizor-lg-led-55um7610,id-1367344	N	142
84	https://wpolityce.pl/sport/457898-banka-zapowiedzial-125-mln-zl-na-kilka-obiektow-sportowych	N	8
85	https://eblik.pl	N	0
86	--Adultcontent	-	-
87	https://jbzdy.net	Y	24
88	https://www.olx.ua/obyavlenie/razborka-shrot-rozborka-skoda-superb-i-azm-IDAPIUA.html	N	7
89	https://www.dobreprogramy.pl/ASUS-ROG-Strix-XG438Q.-Najwiekszy-monitor-4K-UHD-FreeSync-2-HDR,News,103080.html	N	1
90	https://www.amazon.com/dp/B07MMZ2LTB/	N	184
91	https://twitter.com/pittofficial	N	0
92	https://plejada.pl/newsy/piotr-schramm-w-stacji-hejt-ofiary-mowy-nienawisci-targaja-sie-na-zycie-dosc-tego/lrhsqnn	Y	19
93	https://www.pomponik.pl/plotki/news-elzbieta-ii-miala-dosc-gaf-meghan-wyznaczyla-jej-specjalna-m,nId,3132626	N	TO
94	https://www.przelewy24.pl	N	15
95	https://home.pl/office365/	N	0
96	https://www.play.pl/youtube/	N	1
97	https://www.morele.net/karta-graficzna-msi-geforce-rtx-2070-gaming-8g-8gb-gddr6-geforce-rtx-2070-gaming-8g-5411029/	N	2
98	https://imgur.com/t/dogs_are_the_best_people/iCfQe00	N	13
99	https://www.intercity.pl/pl/	N	1
100	https://www.empik.com/za-duzo-mysle-poradnik-dla-analizujacych-bez-konca-petitcollin-christel,p1222736270,ksiazka-p	N	5
101	--Adultcontent	-	-
102	https://niezalezna.pl/282703-usa-rasizm-nie-był-motywnym-strzelaniny	N	0
103	https://ok.ru/realmadrid.c.f	N	8
104	https://www.google.de/search?q=tom+cruise	N	9
105	https://www.flashscore.pl	N	0
106	https://kinogo.by/16305-avengers-endgame_2019_28-07.html	N	0
107	https://hbogo.pl/serie/czarnobyl/sezon-1	N	3
108	https://www.teleman.pl/tv/Resident-Evil-Apokalipsa-637242	N	0
109	https://sport.se.pl/kolarstwo/tragedia-na-trasie-tour-de-pologne-bjorg-lambrecht-zmarl-na-stole-operacyjnym-aa-3S6v-cL87-kLXz.html	Y	7
110	https://stackoverflow.com/questions/57365580/is-main-return-a-program	N	0
111	https://pz.gov.pl/pz/index	N	0
112	https://vod.pl/filmy/alita-battle-angel-caly-film-online/35p9esb	N	10
113	https://www.imdb.com/title/tt6095472/	N	56
114	https://prm.ua/estoniya-ne-viznavatime-rosiyski-pasporti-vidani-zhitelyam-donbasu/	N	0
115	https://wordpress.com	N	0
116	https://www.adobe.com/products/photoshop-lightroom.html	N	2
117	http://rozklad-pkp.pl	N	1
118	https://www.pkobp.pl	N	0
119	--Doesnotwork	-	-
120	http://www.poczta-polska.pl/bohateron-oddajmy-hold-powstancom/	N	0
121	https://discordapp.com	N	0
122	https://shinden.pl/series/52932-dungeon-ni-deai-wo-motomeru-no-wa-machigatteiru-darou-ka-2nd-season	N	0
123	https://pl.bab.la/zwroty/rekrutacja/list-motywyjny/polski-angielski	N	6
124	https://exsite24.pl/darmowe_ebook_download/czasopisma_zagraniczne/218673-practical-photography-august-2019.html	N	0

125	http://seasonvar.ru/serial-22766-Doktor__Han.html	N	0
126	https://www.office.com	N	7
127	https://portal.abczdrowie.pl/gotowane-czy-surowe-jak-jesc-warzywa-aby-mialy-najwiecej-wartosci-odzywczych	N	TO
128	https://soundcloud.com/nlechoppa/shotta-flow-3	N	0
129	https://www.spotify.com/pl/	N	0
130	http://krakow.pl/aktualnosci/231889,33,komunikat,za_nami_patriotyczna_lekcja_spiewania.html	N	2
131	https://mediamarkt.pl/sonywf1000xm3	N	2
132	https://www.messenger.com	N	5
133	https://www.orange.pl/love	N	11
134	https://biznes.radiozet.pl/News/Najbogatsze-i-najbiedniejsze-gminy-w-Polsce.-Ranking-wygral-Kleszczow-a-na-ostatnim-miejscu-Zawadzkie	N	0
135	https://www.biedronka.pl/pl/kuchnia-azjatycka-01-08	N	4
136	https://www.castorama.pl/domowa-jednostka-sterujaca-24v-id-1106728.html	N	420
137	https://www.ebay.com/itm/Polarized-Sunglasses-Mens-Womens-Sport-Running-Fishing-Golf-Driving-Glasses/273348514805	N	58
138	https://www2.hm.com/pl_pl/productpage.0488561032.html	N	2
139	https://jakdojade.pl/wroclaw/trasa/	N	TO
140	https://www.trojmiasto.pl/wiadomosci/Pierwszy-odcinek-trasy-tramwajowej-na-Stogi-zostanie-jednak-otwarty-jesienia-n136680.html	Y	7
141	https://steamcommunity.com	N	0
142	https://store.steampowered.com/app/1089350/NBA_2K20/	N	563
143	--Adultcontent	-	-
144	http://www.gry.pl/grac/chrome-dino	Y	31
145	https://www.auto-swiat.pl/klasyki/oldtimer/klasyk-z-najwyzszej-polki-citroen-ds-21-presidentielle/bz1xrvq	Y	11
146	https://www.bankier.pl/wiadomosc/Donald-Trump-o-oslabieniu-juana-Chiny-manipuluja-waluta-7717876.html	N	33
147	https://inpost.pl/przesylki-paczkomaty-kurier	N	0
148	https://www.spidersweb.pl/2019/08/donald-trump-gry-wideo.html	N	TO
149	https://pl.jobble.org/jdp/-2553381723608201811/UI-%2F-Web-Developer-Warszawa	N	0
150	https://www.santander.pl	N	1
151	https://www.lotto.pl/eurojackpot/aktualnosci/ponad-1-mln-zlotych-w-eurojackpot-w-bydgoszczy-kumulacja-nadal-rosnie	N	20
152	https://www.infor.pl/prawo/powiat/rejestracja-samochodu/2758936,Podatek-przy-zakupie-auta-2019-r.html	N	8
153	--Adultcontent	-	-
154	https://blog.mozilla.org/blog/2019/06/04/when-it-comes-to-privacy-default-settings-matter/	N	0
155	https://www.pepper.pl/promocje/kupon-na-115-zl-do-hulajnog-citybee-177880	N	0
156	https://www.gamepedia.com/news/1977-defy-the-rules-of-mass-and-physics-with-morphies	N	16
157	--Adultcontent	-	-
158	http://www.reverso.net/text_translation.aspx?lang=EN	N	5
159	https://www.decathlon.pl/rower-mtb-st-100-275-id_8400336.html	N	TO
160	https://ekino-tv.pl/movie/show/ksiezniczka-labeledzi-i-krolestwo-muzyki-hd-the-swan-princess-kingdom-of-music-2019-dubbing/26187	N	37
161	https://autokult.pl/34501,kamper-nie-musi-byc-nudny-taki-transit-to-marzenie-kazdego-fana-wyscigow	Y	1
162	https://thestartmagazine.com/article/28389333-0516-450d-8f8c-cd0d8eb77087	N	253
163	https://www.tripadvisor.com/AttractionProductReview-g60763-d11451132-Statue_of_Liberty_and_Ellis_Island_Award_Winning_Small_Group_Tour-New_York_City-New.html	N	42
164	http://www.medonet.pl/zdrowie,odciski--jakie-specjalistyczne-zabiegi-pomagaja-w-leczeniu-odciskow-,artykul,1732804.html	Y	12
165	https://www.ikea.com/pl/pl/catalog/products/40443835/	N	0

166	https://www.pwn.pl/careers	N	0
167	https://www.eobuwie.com.pl/sneakersy-tommy-hilfiger-branson-8c1-fm0fm00612-twilight-005.html	N	23
168	https://natemat.pl/280829,jacy-sa-wspolczesni-mezczyzni-nie-wszyscy-sa-zli-kobiety-wymagaja-za-duzo	N	34
169	http://rutracker.org/forum/viewtopic.php?t=5370852	N	0
170	https://player.pl/programy-online/kuchenne-rewolucje-odcinki,114	Y	64
171	https://oliviawhen.tumblr.com/post/186738461661/did-you-know-its-doggust-dog-august-my-new	N	0
172	https://login.poczta.home.pl	N	1
173	https://www.mazowieckie.pl/pl/aktualnosci/aktualnosci/38711,Rzez-Woli-pamiec-cywilnych-ofiar-w-75-rocznice.html	N	1
174	--Unreachable	-	-
175	https://lowcygier.pl/promocje-cyfrowe/mass-effect-trilogy-w-wersji-cyfrowej-za-2990-zl-w-muvesklepie-gol/	Y	4
176	https://www.uw.edu.pl/droga-mleczna-w-trzech-wymiarach/	N	0
177	https://z1.fm/song/23170839	N	0
178	https://t.co	N	0
179	https://worldoftanks.eu/en/news/special-offers/Pz-Kpfw-IV-schmalturm-august/	N	0
180	https://www.nazwa.pl/cennik/cennik-wordpress-cloud/	N	0
181	--Adultcontent	-	-
182	https://www.targeo.pl	N	0
183	https://www.bing.com/search?q=tom+cruise	N	5
184	http://www.plotek.pl/plotek/7,154063,25061903,kamil-durczok-mial-kupic-wodke-na-pol-godziny-przed-kolizja.html	N	0
185	https://pclab.pl/art81578.html	N	4
186	http://www.benchmark.pl/testy_i_recenzje/gigabyte-x570-aorus-xtreme-i-ssd-m-2-na-pcie-4-0-recenzja.html	N	5
187	--Adultcontent	-	-
188	https://www.komputronik.pl/product/341252/goodram-8gb-1x8gb-1600mhz-ddr3-cl11-dimm-.html	N	67
189	https://9gag.com/gag/aQ1P42W	N	1
190	https://www.leroymerlin.pl/podlogi-drewniane-boazerie/panele-deski-parkiety/panele-podlogowe-laminowane/panele-podlogowe-dab-elbrus-ac4-8-mm-home-inspire,p469109,1188.html	N	0
191	https://wetransfer.com	N	0
192	https://smaker.pl/przepis-busz-poledwiczka,174227,kulinarnachwila.html	N	4
193	--Adultcontent	-	-
194	https://www.rp.pl/Urzednicy/308069970-Kto-moglby-zaplacic-za-loty-marszalka-Marka-Kuchcinskiego.html	N	0
195	https://www.amazon.de/Samsonite-SCure-Spinner-Koffer-Black/dp/B007WQJSGM/	N	274
196	https://badoo.com	N	0
197	https://www.deviantart.com/yeraa/art/Midsummer-808465288	N	4
198	https://inteligo.pl	N	1
199	https://parenting.pl/7-najbardziej-absurdalnych-ciazowych-przesadow	N	11
200	https://www.dotpay.pl/metody-platnosci/platnosci-odroczone/kupuj-teraz/	N	0
201	https://github.com/google	N	0
202	https://www.credit-agricole.pl/klienci-indywidualni	N	0
203	https://naekranie.pl/aktualnosci/lowcy-cieni-bedzie-serial-stacji-nowa-tv-opis-fabuly-i-data-premiery-1565118604-1565119246	N	34
204	https://www.airbnb.pl/rooms/37253565	N	0
205	https://www.skapiiec.pl/cat/308-lodowki.html	N	14
206	https://www.nocowanie.pl/noclegi/kolobrzeg/apartamenty/145748/	N	4
207	https://www.apple.com/iphone-xr/	N	0
208	https://jbdy.net	Y	24
209	https://wbijam.pl	N	0
210	https://www.taaron.pl/dla-domu/prad/zwrot-za-prad	N	0

211	https://www.itaka.pl/wczasy/bulgaria/sloneczny-brzeg/hotel-julia,BOJJULI.html	N	29
212	https://www.zapmeta.com.pl/?q=tom+cruise	N	0
213	https://nk.pl/logowanie	N	78
214	https://www.rmfm24.pl/fakty/polska/news-brunon-kwiecien-nie-zyje-zostal-skazany-za-probe-zamachu-na-,nId,3134454	N	7
215	https://www.polsatnews.pl/wiadomosc/2019-08-06/z-panstwowej-mennicy-w-meksyku-zrabowano-zlote-monety-za-ponad-2-mln-usd/	Y	4
216	https://echodnia.eu/swietokrzyskie/wicemarszalek-sejmu-malgorzata-gosiewska-skrytykowala-swietokrzyskich-parlamentarzystow-prawa-i-sprawiedliwosci-za-brak/ar/ci-14330679	Y	12
217	https://www.msn.com/pl-pl/wiadomosci/polska/bestie-z-okuninki-zgwaT1\lcili-kolegę-kijem-katowali-kilka-godzin/ar-AAFqI0k	N	TO
218	https://www.linkedin.com/	N	1
219	https://sinoptik.pl	N	0
220	https://gfycat.com/snivelingmediocredeinonychus-thanksgiving-comedy	N	8
221	https://www.autocentrum.pl/newsy/wiozla-dwie-osoby-w-bagazniku-na-trzy-miesiace-straci-prawo-jazdy/	N	3
222	https://duckduckgo.com/?q=tom+cruise&ia=web	N	1
223	http://pogodynka.pl/wiadomosci/aktualnosci/1907-22072019-ocena-meteorologiczna-i-hydrologiczna-1343	N	3
224	--Redirects	-	-
225	https://bm.pl	N	0
226	https://www.tradedoubler.com/en/publishers/	N	0
227	https://privatbank.ua	N	1
228	https://www.ryanair.com/gb/en/	N	4
229	https://www.zus.pl/o-zus/aktualnosci/-/publisher/aktualnosc/1/w-szczubrzeszynie-zus-brzmi-prosto/2816156	N	0
230	https://www.telemagazyn.pl/film/goracy-towar-502417/	N	20
231	https://login.gov.pl/login/secure	N	0
232	https://www.animezone.pl/gatunki	N	1
233	http://baskino.me/films/boeviki/19454-lyudi-v-chernom-interneshnl.html	N	18
234	https://www.znanylekarz.pl/magdalena-zychowska/dermatolog-lekarz-medycyny-estetycznej/wroclaw	N	TO
235	http://site-pl.com/?reqp=1&reqr=	N	0
236	https://en.wiktionary.org/wiki/amazing	N	0
237	--403HTTPError	-	-
238	http://kinogo.eu/9265-mstiteli-final-2019.html	N	0
239	http://lubimyczytac.pl/aktualnosci/rozmowy/12508/szczescie-solo-czyli-o-zwiazku-z-samym-soba-wywiad-z-wanda-zolcinska	Y	4
240	http://www.speedtest.pl/kim_jestesmy	N	0
241	http://ulub.pl/utACbs0ip7/bibsypl-100-obrazkow-na-minute	N	0
242	https://www.pyszne.pl/pizzeria-staromiejska	N	4
243	https://www.pinterest.com	N	TO
244	https://fotka.com	N	0
245	https://ex-torrenty.org/torrent/166668	N	0
246	https://www.gov.pl/web/gov/uslugi-dla-obywatela/#wyjazdy-i-wypoczynek	N	0
247	https://www.sadistic.pl/portal/14693	Y	2
248	https://www.praca.gov.pl/eurzad/index.eup#/inneSprawy/wyborUrzedu?dest=EURZAD	N	0
249	https://www.mp.pl/dostepnaczyniowy/ekspert/pytania-do-eksperta/211883, czy-cewnik-dializacyjny-moze-byc-wykorzystany-do-podania-przez-ratownika-lekow-podczas-resuscytacji	N	2
250	https://gratka.pl/motoryzacja/motocykl-yamaha-xvs-650-v-star/oi/12732681	N	17
251	https://vider.info/vid/+fxms588	Y	0
252	https://yandex.ua/search/?lr=143&oprnd=9981694847&text=tom%20cruise	N	0
253	https://www.tekstowo.pl/piosenka,lionel_richie,hello.html	N	4

254	https://www.twojapogoda.pl/wiadomosc/2019-07-12/wedkarze-lapcie-za-wedki-bo-ryby-beda-braly-jak-szalone-tak-dobrych-lowow-w-te-wakacje-juz-nie-bedzie/	Y	13
255	https://www.doz.pl/apteka/p131082-Estabiom_pregna_kapsulki_20_szt.	N	2
256	https://www.upc.pl/internet/kup-internet/	N	21
257	https://www.lidl-sklep.pl/PARKSIDE-Akumulatorowe-urządzenie-wielofunkcyjne-4-w-1-PKGA-20-Li-B1-bez-akumulatora-i-ladowarki/p100259868	N	0
258	http://lordsfilms.tv/26676-priklyucheniya-paddingtona-2.html	N	0
259	https://www.eset.com/pl/business/	N	1
260	https://www.lidl.pl/pl/Oferta-spozywca.htm?articleId=66407	N	2
261	https://www.przegladsportowy.pl	Y	20
262	http://search-pl.com/?reqp=1&reqr=	N	0
263	https://www.gov.pl/web/sprawiedliwosc/przelom-w-kształceniu-służby-wiezionej--slubowanie-studentow-pierwszej-w-historii-uczelnipolskiego-wieziennictwa	N	0
264	https://myanimelist.net/anime/37347/Dungeon_ni_Deai_wo_Motomeru_no_wa_Machigatteiru_Darou_ka_II	N	11
265	https://obywatel.gov.pl/kierowcy-i-pojazdy	N	0
266	https://rozetka.com.ua/xiaomi_redmi_note_7_4_64gb_black/p84686898/	N	10
267	https://www.kurnik.pl/szachy/	N	0
268	http://sklep.cyfrowypolsat.pl/telewizja-satelitarna	N	0
269	https://poradnikprzedsiębiorcy.pl/-trwa-nabor-do-xi-edycji-konkursu-innowator-mazowska	N	1
270	https://www.e-podroznik.pl	N	0
271	https://www.t-mobile.pl/telefony-i-urządzenia/telefony/samsung-galaxy-a50-samsung-galaxy-fit/SA-H-B1570.phtml?addoffer=tariprom_40335	N	1
272	https://en.sat24.com/en/forecastimages/europa/forecasttemp	N	0
273	http://home-nav.com/?reqp=1&reqr=	N	0
274	https://www.dropbox.com/?landing=dbv2	N	0
275	https://www.styl.pl/zdrowie/news-najlepsza-woda-dla-ciebie,nId,3135966	N	9
276	http://rutor.info/torrent/707276/ladja_the-rook-01h01-06-iz-08-2019-web-dlrip-lostfilm	N	0
277	https://www.popads.net	N	0
278	https://www.meczyki.pl/newsy/zwrot-akcji-w-sprawie-christiana-eriksena-manchester-united-zrezygnowal-z-walki-o-dunczyka/119664-n	Y	TO
279	https://www.tvp.info/43838474/spanikowalem-pracownik-ochrony-finalu-wosp-w-gdansk-przed-sadem	N	24
280	https://www.instalki.pl/testy/hardware/36923-xiaomi-mi-band-4-test-opaski-sportowej.html	N	1
281	https://www.wroclaw.pl/portal/dla-wygody-mieszkanow-wroclaw-stawia-na-osiedla-kompletne	N	0
282	https://genius.com/a/ana-del-rey-shares-looking-for-america-in-response-to-recent-mass-shootings	N	2
283	https://www.ipla.tv/kanaly-tv	N	0
284	https://play.eune.leagueoflegends.com/en_PL	N	0
285	http://l.ua	N	0
286	https://www.skyscanner.pl	N	TO
287	https://www.wakacje.pl/oferty/grecja/chalkidiki/metamorfosis/simeon-526145.html?od=2019-10-15,7-dni,HB,z-warszawy-modlin	N	3
288	https://vimeo.com/showcase/2380839/video/59207751	N	5
289	https://www.speedtest.net	Y	1
290	https://dorzeczy.pl/kraj/110342/szykuje-sie-najwiekszy-skandal-w-historii-polskiego-show-biznesu-znany-muzyk-jazzowy-oskarzony-o-pedofilie.html	N	0
291	https://www.efortuna.pl/pl/fortuna_z_bonusem/index.html	N	0
292	https://jelenia-gora.lento.pl/poszukujemy-opiekunki-opiekunow,8060605.html	N	0
293	https://serwisy.gazetaprawna.pl/edukacja/artykuly/1425114,500-plus-dla-nauczycieli-8-6-mln-zl-lacznie.html	N	4

294	https://www.cdaction.pl/news-57514/gamescom-2019-final-fantasy-7-remake-i-marvels-avengers-grywalne-dla-zwiedzajacych.html	N	1
295	--CDN	—	—
296	https://en.savefrom.net	N	0
297	https://mielno.webcamera.pl	Y	251
298	https://www.getinbank.pl	N	2
299	https://prom.ua/p771785763-nastolnyj-futbol-detskij.html	N	205
300	https://gazetawroclawska.pl/wypadek-na-traugutta-dzwig-wyrwal-drzwi-tramwaju-zdjecia/ar/c16-14333595	Y	21