Abstract—Package-based dependency networks model which software packages depend on which other packages. Researchers and practitioners have used them to achieve a great number of analyses, including automatically warning for security vulnerability, ecosystem health and license compliance issues. However, traditional package-based dependency networks are in-precise, severely limiting their use in practice. In this paper, we present a novel and general approach named PRÄZI to construct call-based dependency networks beyond a single program, its initial prototypical implementation RUSTPRÄZI for the Rust library system CRATES.IO, an evaluation of its soundness and precision, and two sample applications with it. Our case study on security vulnerabilities showed that RUSTPRÄZI is three times more accurate than the current state of the art, package-based analyses. PRÄZI also opens the door to new applications, e.g., an analysis on the prolonged use of deprecated methods. It showed that 48% of the studied dependent packages break when a deprecated function gets removed. Several perils endanger a practical implementation of PRÄZI, affecting both its soundness and precision. We discuss and quantify them along the RUSTPRÄZI example, equipping researchers and practitioners with guidelines on how to implement PRÄZI. Finally, we also show that there is no principal objection to make PRÄZI fully sound and precise.

I. INTRODUCTION

In today’s software development, most programs comprise a growing list of other software they depend on [1]. Online Package Repositories (OPRs), see Table 1 of modern programming languages such as Java’s MAVEN CENTRAL allow the efficient (re-)use and combination of already existing functionality in one’s own program. While reuse is a core Software Engineering principle, promising higher development speed and quality [2], uninhibited reuse from OPRs bears risks [4]. These became painstakingly obvious to the JavaScript community in the “leftpad incident”, when the removal of one package lead to the breaking of hundreds of thousands of other programs depending on it [5], most transitorily through a chain of other dependencies. As a result, researchers have analyzed the dependency graphs found in OPRs from a variety of different viewpoints [4], [6]–[9], including security vulnerabilities and ecosystem health. On the practical side, companies are using dependency graphs for an array of applications: GitHub and TIDELIFT warn project owners when they (implicitly) depend on a known vulnerable library [10], [11]; Google’s operation Rosehub supplied pull requests to over 2,600 GitHub projects which imported a vulnerable version of the Apache Collections library, including the popular SPRING framework, thus fixing the many more transitive projects depending on it [12]; BLACKDUCK performs license compliance checking to avoid importing two dependencies with conflicting licenses [13].

However, these state-of-the-art analyses operating on package-based dependency networks (PDN) share one shortcoming: they use the readily accessibly dependency specification in a project’s metadata (such as MAVEN’s pom.xml) and might be imprecise, because the actual dependency use happens at the source code level. Hence, existing PDNs are sound, but an over-approximation of the real dependency use. For example, a project might have redundant dependencies to packages whose functionality is not used anymore, creating false positive warnings. False positives, are a dominating factor for the slow adoption of static analysis tools in practice [14], with developers having to sift through several false positives just to find one relevant warning [15], [16]. The faulty security warning raised by GitHub’s package-based dependency checker in Figure 1 exemplifies the confusion false positives can create.

In this paper, we present an approach to build dependency analyses not at the package but at the function call level, called PRÄZI. Call graphs represent the inter-procedural control flow of source code and thus naturally lend themselves to this objective. With PRÄZI, we build the call graph of each package and its transitive dependencies and merge them together, resulting in a call-based dependency network (CDN). With PRÄZI, we can improve the state-of-the-art PDN analyses with more
precise CDN analyses. Our technique allows the application of the above analyses on a finer level. The main advantage of PräZi is that it i) is more precise, avoiding spurious warnings such as the one in Figure 1 and ii) also opens the door to new applications on an entire OPR from change impact analyses (“Which clients break if I as a library maintainer remove this deprecated method?”) to network health (“What are the most important methods and are they tested well?”).

In the remainder of this paper, we first describe the generic PräZi approach to produce a CDN comprising the merger of several packages’ call graphs. PräZi is simple to explain but difficult to implement at scale. We thus present a prototypical implementation and an evaluation of PräZi on Rust and CRATES.IO. We describe a list of perils of how practical implementation choices can threaten the soundness and precision of PräZi, and quantify the effects of the perils in our Rust implementation RUSTPREAZI during its construction process. While our prototypical implementation is only tailored to statically dispatched functions, our evaluation of the CDN in two case studies on the propagation of security vulnerabilities and deprecated functions shows that it is still useful in practice. In comparison to traditional coarse-grained package network based analyses, PräZi eliminates false positives. In a case study with 482 warnings, RUSTPREAZI reports an accuracy which is three times higher as compared to a traditional PDN, despite reporting a lower recall score than a PDN due to non-static function dispatch.

II. BACKGROUND

In this section, we give background information over call graphs and the Rust programming language.

A. Call Graphs

A call graph is a reduced control flow graph [17], which only represents function calls and their relationship within a single program, be it an executable or a library. We can produce call graphs from static program artifacts, but there also exist techniques to infer call graphs from dynamic program execution traces [18].

Static call graphs are ideally suited to represent static function calls, but we need more advanced techniques such as aliasing to precisely deal with dynamic invocations, frequently occurring in languages such as Java [19]. Dynamic program analysis can complement static call graph generation, but it requires running the program with all potential combinations of inputs to be fully sound.

To circumvent the problem of aliasing, many call graph generators sacrifice precision for soundness, blowing up the

precise CDN analyses. Our technique allows the application of the above analyses on a finer level. The main advantage of PräZi is that it i) is more precise, avoiding spurious warnings such as the one in Figure 1 and ii) also opens the door to new applications on an entire OPR from change impact analyses (“Which clients break if I as a library maintainer remove this deprecated method?”) to network health (“What are the most important methods and are they tested well?”).

In the remainder of this paper, we first describe the generic PräZi approach to produce a CDN comprising the merger of several packages’ call graphs. PräZi is simple to explain but difficult to implement at scale. We thus present a prototypical implementation and an evaluation of PräZi on Rust and CRATES.IO. We describe a list of perils of how practical implementation choices can threaten the soundness and precision of PräZi, and quantify the effects of the perils in our Rust implementation RUSTPREAZI during its construction process. While our prototypical implementation is only tailored to statically dispatched functions, our evaluation of the CDN in two case studies on the propagation of security vulnerabilities and deprecated functions shows that it is still useful in practice. In comparison to traditional coarse-grained package network based analyses, PräZi eliminates false positives. In a case study with 482 warnings, RUSTPREAZI reports an accuracy which is three times higher as compared to a traditional PDN, despite reporting a lower recall score than a PDN due to non-static function dispatch.

II. BACKGROUND

In this section, we give background information over call graphs and the Rust programming language.

A. Call Graphs

A call graph is a reduced control flow graph [17], which only represents function calls and their relationship within a single program, be it an executable or a library. We can produce call graphs from static program artifacts, but there also exist techniques to infer call graphs from dynamic program execution traces [18].

Static call graphs are ideally suited to represent static function calls, but we need more advanced techniques such as aliasing to precisely deal with dynamic invocations, frequently occurring in languages such as Java [19]. Dynamic program analysis can complement static call graph generation, but it requires running the program with all potential combinations of inputs to be fully sound.

To circumvent the problem of aliasing, many call graph generators sacrifice precision for soundness, blowing up the
In this section, we devise a generic technique, PRÄZI, to systematically analyze a set of packages residing in an OPR to construct a CDN. We argue that PRÄZI is generic enough to be applied to any programming environment that features i) a way of expressing dependency information between packages, and ii) tooling to generate call graphs for a package. As the overview in Figure 4 shows, PRÄZI first has to resolve and retrieve all packages (and their dependencies) to be analyzed. It then has to generate call graphs for them, unify these to avoid name clashes and link them together in one giant call graph, the CDN. In the following sections we present each step of PRÄZI in detail, along with a set of perils that may affect the soundness and completeness of the produced CDN.

A. Resolving Dependencies and Retrieving Packages

PRÄZI starts with a pre-defined set of packages for which a CDN should be built, the seed set. Packages in the seed set should have dependencies that can be resolved within the context of an OPR. The first step for PRÄZI is to resolve all dependencies in the seed set in a recursive fashion, until it has calculated the full transitive closure of the packages and their associated versions it needs to retrieve.

Dependency resolution is complicated by the need to build a full tree of package versions not only for the present, but also for the past, and possibly future. Almost all current package managers allow developers to specify ranges of dependency versions, often in the semantic versioning format. For example, any dependency version with a leading 1. fulfills the version range 1.* (e.g. 1.0, 1.8, or 1.20.2). The package manager will generally choose the latest available version at the time of its invocation. This complicates the retroactive resolution of dependency versions.

Suppose that package A depends on version 1.* of package B, as shown in Figure 3a. Package B releases versions 1.1 at t₁ and 1.2 at t₂ (t₁ < t₂) (Figure 3b). If the dependency resolution happens at t, where t₁ < t < t₂, then the dependency manager will select version 1.1. However, if dependency resolution happens at t > t₂, it will select version 1.2, even though A’s dependency specification (1.*) remained unchanged. Removal or black-listing of packages, a practice supported by many OPRs, further complicate retroactive resolution.

To deal with this issue, PRÄZI expects the nodes in a PDN to be timestamped. It resolves dependency constraints by linking dependents to all versions of their dependencies that would satisfy the constraints (in the example above, A would thus depend both on B 1.1 and B 1.2). This creates an over-approximation of the actual PDN; to resolve the exact dependencies for package C version v₁ released at timestamp t₁, PRÄZI removes all nodes, and the corresponding edges, whose release timestamp tᵦ is tᵦ > t₁; it then performs a breadth-first search of the dependency graph from Cᵥ₁. If it finds a node (e.g., A) that links to multiple versions of another node (e.g., B), it must apply an equivalent selection strategy to that of the actual package manager. Usually, this strategy is to select the latest of those versions.
After resolving dependencies, PräZi must download the releases of the identified package versions. What this step will retrieve depends on the call graph generator requirements for the target programming language. For example, in languages where the call graph generator can work on intermediate formats (e.g., Java), binary packages may be sufficient; in most other languages though, PräZi needs to retrieve the source code of the package.

**Peril 1.** For the PräZi method to be complete, all packages specified in the seed set must be retrievable at analysis time along with their metadata. Not all OPRs can guarantee this: for example, until the leftpad incident, NPM allowed developers to remove packages.

**Peril 2.** To generate the PDN, PräZi needs the dependency metadata descriptor for each package. Several package managers, including NPM and Cargo, do not check whether the dependency metadata are well-formed or do not reference resources local to the developer’s workstation when new package versions are uploaded. This may lead to missing packages in the PDN, and subsequently, our CDN.

**Peril 3.** Precise runtime dependency resolution complicates accurate construction of dependency sets, and may affect the replicability of the PräZi process. The version of dependent packages may have significantly changed externally without any changes to the importing package. Thus, two consecutive builds of the CDN may be different.

### B. Generating Call Graphs

After all package dependencies have been resolved and fetched locally, the PräZi technique generates a call graph for each package version. Depending on the call graph generator implementation and the programming environment, the package may need to be built. PräZi treats the call graph generator as a pluggable component. Any implementation that adheres to the following requirements is suitable for PräZi: i) Function types and their arguments must be fully resolved. For example, a Rust call graph generator will resolve a call to function `parse(input: &str)` in struct `Url` that resides in package `url` as `Url::parse(input: &str)`, where `str` is the Rust standard string type. ii) The graph output format is an edge list of function name pairs. The last step of the PräZi technique involves merging together all individual package-level CDNs into a single CDN. This process consists of aggregating all individual package call graphs and filtering out duplicate nodes. The end result is the CDN corresponding to the input seed set.

### C. Generating Unique Function Identifiers (UFIs)

To merge the call graphs for each individual package version into a single CDN, PräZi needs to ensure that the contained function signatures are globally unique. Complications can arise in case one program imports multiple versions of the same package, packages include similar function names (e.g., `log(x: &str)`) in languages with flat namespaces, or duplicate function names when processing function calls across OPRs. To solve these issues, PräZi prepends all calls, including function names and the types in their arguments, with three attributes: i) OPR name, ii) package name and iii) package version. For example, the UFI for the above function `parse(input: &str)`, which can be found in the crate `io::crates::url::1.6.1::Url::parse(input: &str)`. This approach ensures that the call graph generator can accurately identify function calls across different packages and versions.

### D. Unifying Call Graphs

The last step of the PräZi technique involves merging together all individual package-level CDNs into a single CDN. This process consists of aggregating all individual package call graphs and filtering out duplicate nodes. The end result is the CDN corresponding to the input seed set.

### V. Implementing PräZi for Rust

We chose Rust to showcase the practical feasibility and scalability of PräZi for a number of reasons: As a new language, Rust’s OPR, crates.io, is relatively contained compared to NPM’s 650,000 packages (see Section II-B). LLVM, Rust’s standard compiler, can output call graphs as a side-artifact of compilation. Thus, we can have high confidence in their correctness.

These facts set Rust apart from almost all legacy languages, such as C and C++, where building and dependency management is usually done at the operating system level. Moreover, it is hard to define what exactly the OPR of, e.g., Java is,
as Maven Central also contains build artifacts from Scala, Groovy, Closure, JRuby and Kotlin. Dynamic dispatch, which is hard to capture via static call graphs, is common practice in Java and C. Moreover, with Cargo, Rust has a unified way to build projects, unlike Java, where projects can use Ant, Maven, or Gradle. Finally, in contrast to highly dynamic languages such as JavaScript, Rust as a strongly typed language lends itself to static analysis, and thus, to building accurate call graphs. In fact, the Rust language documentation itself advocates static over dynamic dispatch [25].

A. Resolving Dependencies and Retrieving Packages

CRATES.IO hosts an official up-to-date index of its packages in a GitHub repository [26]. We clone a snapshot of the index at revision b76c5ac (16th February 2018) and populate our seed set with all its entries. This set constitutes of 79,724 releases (i.e., package versions) from 13,991 unique packages. The complete seed set makes the step of resolving dependencies to fetch missing packages redundant. To download a specific package from CRATES.IO, we use its dedicated API [27]. In total, we could download and uncompress 79,701 package versions from our seed set.

Peril 1 We are missing 23 package versions due to unauthorized access and 8 package versions due to malformed tar archive headers.

Peril 2 Because CRATES.IO does not validate build manifests (i.e., the Cargo.toml file) in released packages in CRATES.IO, our downloaded set may contain invalid build manifests. A package with an invalid build manifest is not compilable, and hence cannot be used as a dependency in other projects. Using Cargo’s manifest validator tool, we could identify 477 package versions with invalid manifests. We further exclude “packages” that are client applications, excluding 6,277 non-library package releases. In total, 72,947 package versions (i.e., 92% of our pre-defined seed set) can be used to build the Rust CDN.

Peril 3 In RUSTPRAZI, we have simplified the retroactive dependency problem and created a network that is only valid for one point in time, at revision b76c5ac. We now want to quantify how different the PDN looks when we build it a day, a week, a month, 3 months and 6 months later. We construct the new PDNs by keeping the source code from b76c5ac, and only updating the CRATES.IO index to the new timestamps. This ensures that the code will not change, but will fetch new dependencies, if available. From Table II, we can observe that CRATES.IO is very volatile: even within one month, a quarter of the dependencies have resolved to a different version. Thus, if having an accurate complete history is a concern for a practical implementation of PRAZI, it seems vital to follow the general approach outlined in Section IV-A.

B. Generating Call Graphs

To generate call graphs, we use the LLVM call graph generator, version 4.0.0, which works by analyzing a program’s LLVM Intermediate Bytecode Representation (IR).

Peril 4 Not being Rust-specific, LLVM may miss Rust-specific calling conventions, leading to a potentially incomplete Rust CDN. To evaluate its shortcomings, we consider all possible ways [25] to define or call a function in Rust (Table III). We then construct examples that exercise a specific function call or definition, and generate the call graph representing these cases. After inspecting the generated call graph, we document the support of each feature in Table III.

Overall, we can identify that the LLVM call graph generator is not able to infer non-static dispatch calls or macro invocations.

![Diagram of CRATES.IO CDN pipeline]

Fig. 4: Our approach to generate a CDN for the CRATES.IO OPR.

<table>
<thead>
<tr>
<th>Revision</th>
<th>Time point</th>
<th># (Changed) Packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>b76c5ac</td>
<td>Feb 16’18</td>
<td>297,757</td>
</tr>
<tr>
<td>6e9b751</td>
<td>1 day (Feb 17’18)</td>
<td>429 (0.1%)</td>
</tr>
<tr>
<td>76a24f9</td>
<td>1 week (Feb 23’18)</td>
<td>8,484 (2.9%)</td>
</tr>
<tr>
<td>eb7b311</td>
<td>1 month (Mar 16’18)</td>
<td>55,651 (23.0%)</td>
</tr>
<tr>
<td>a4bc79d</td>
<td>3 months (May 11’18)</td>
<td>82,960 (38.0%)</td>
</tr>
<tr>
<td>75ff77</td>
<td>6 months (Aug 3’18)</td>
<td>97,509 (48.7%)</td>
</tr>
</tbody>
</table>

TABLE II: Retroactive dependency changes.
use C
ARGO

RATES
are incorrectly pointing the download source of a dependency
due to missing path-based dependencies. These dependencies
compilation errors reveals that a large number of builds fail
wards compatible with the stable compiler. By swapping the
features from a nightly compiler release, which are not back-
the compatible compiler versions; a package may use unstable

fell after installing native dependencies.

Furthermore, it is only able to infer generic function definitions
if instantiations of it exist.

To generate the LLVM IR of the Rust packages, we need to
compile our set of 72,947 package versions. Compiling such
a large set of packages is a challenging task because many
environmental factors influence it. It is also an important step
because compile failures affect the completeness of the Rust
CDN.

We perform the build step in several compilation rounds to
achieve maximum completeness. We first use a stable version
of the compiler, and then iteratively analyze compilation logs
to tackle the common failure reasons. The compilation itself
ran for almost three days in parallel using a build server
with an Intel Xeon E5-2690 v4 CPU with 14 hyper-threaded
cores clocked at 2.6GHz, 128GB RAM and seven 2TB SSDs
formatted with ZFS in RAIDZ-1 (RAID-5 equivalent) mode,
on Ubuntu 16.04.3 LTS.

Table IV shows the number of successful compilations along
with the total time for each compilation round. In the first
round, we successfully compile 51% of our set using the
rustc stable 1.22.1 (2017-11-22) compiler. Unfortunately,
a Rust package’s build manifest does not specify the
compatible compiler versions; a package may use unstable
features from a nightly compiler release, which are not back-
wards compatible with the stable compiler. By swapping
the stable compiler version for the nightly version rustc 1.24.
0-nightly (2017-12-06), we compile an additional
4,972 package version releases. Analysis of the remaining
compilation errors reveals that a large number of builds fail
due to missing path-based dependencies. These dependencies
are incorrectly pointing the download source of a dependency
to a local directory instead of CRATES.IO. To resolve this, we
use CARGO’s internal dependency source rewrite feature in
and compile 2,644 additional package versions. Finally, we
observe that several package releases are using native library
dependencies which are not installed on Ubuntu 16.04; with
them installed, we compile an extra 1,862 package releases.

Peril [5] Despite our best efforts, we could not compile
23,063 (31%) package releases. To understand why they fail
to compile, we analyze the compiler errors and classify them
into five categories in Table V. The majority seem to relate
to actual programming faults in the packages, in particular the
Rust type checker (e.g., E0277, E0599, E0425), syntactical
errors and invalid specifications for conditional compilation.
A common reason for these error messages is the improper
use of Traits. Overall, we can compile 69% of total releases
and at least one release for 88% of packages, roughly double
the ratio of previous attempts [33].

C. Generating Unique Function Identifiers

For each constructed call graph in our set of packages,
we parse the function names to append it with OPR and
package-specific information. The Rust compiler mangles
function identifiers in source code to flat C++ namespace-
like representations [34]. To construct a UFI from a Rust-
mangled identifier, we prepend namespaces in the identifier
with appropriate package names and versions, and also with an
OPR-qualifier (i.e., CRATES.IO). Due to nested type structures
in functions, portions of an identifier can have nested names-
paces which complicates the UFI construction. To untangle
the nested namespaces in an identifier, we develop a parser
that produces syntax tree representations from Rust-mangled
function names. With a syntax tree, in a simplified way, we
can access and prepend individual namespaces in an identifier.

The Rust mangled function identifiers share some similari-
ties with Rust syntax. To build the parser, we use the parser
combinator framework, syn [35] for parsing Rust source code.
To validate our parser implementation, we create a corpus
containing all function identifiers from our set of constructed
call graphs, and then attempt to parse each identifier into a
syntax tree. The constructed corpus consists of 111,876,036
function identifiers. In the first attempt, we use the default Rust
parser to process our corpus. In total, 6,030,730 identifiers
generate a parse error. After adding support for brackets (i.e.,
<...>), we reduce the parse errors to 4 million errors. After
inspecting the error logs, we identify that the parser fails

<table>
<thead>
<tr>
<th>Build Round</th>
<th>#Releases</th>
<th>#Packages</th>
<th>Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRATES.IO</td>
<td>72,947</td>
<td>12,307</td>
<td>—</td>
</tr>
<tr>
<td>1. Rustc stable</td>
<td>40,366 (55%)</td>
<td>9,376 (76%)</td>
<td>33.8</td>
</tr>
<tr>
<td>2. Rustc nightly</td>
<td>+4,972 (+7%)</td>
<td>+976 (+8%)</td>
<td>+13.5</td>
</tr>
<tr>
<td>3. Cargo.toml fixes</td>
<td>+2,644 (+4%)</td>
<td>+244 (+2%)</td>
<td>+5.8</td>
</tr>
<tr>
<td>4. Native dependencies</td>
<td>+1,862 (+3%)</td>
<td>+235 (+2%)</td>
<td>+16.5</td>
</tr>
<tr>
<td>Σ</td>
<td>49,844 (69%)</td>
<td>10,831 (88%)</td>
<td>69.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Failure reason</th>
<th>#Builds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compile error w. error code</td>
<td>13,509 (58%)</td>
</tr>
<tr>
<td>Compile error wo. error code, of which 7,272 (31%)</td>
<td></td>
</tr>
<tr>
<td>… code parsing errors</td>
<td>1,486</td>
</tr>
<tr>
<td>… conditional compilation errors</td>
<td>1,058</td>
</tr>
<tr>
<td>… dependency resolution errors</td>
<td>719</td>
</tr>
<tr>
<td>… type checking errors</td>
<td>278</td>
</tr>
<tr>
<td>… other errors</td>
<td>3,711</td>
</tr>
<tr>
<td>Custom build script failure</td>
<td>2,127 (9%)</td>
</tr>
<tr>
<td>Missing system dependencies</td>
<td>137 (&lt; 1%)</td>
</tr>
<tr>
<td>Miscellaneous errors</td>
<td>18 (&lt; 1%)</td>
</tr>
</tbody>
</table>
for namespaces that contain Impl structures and anonymous closures. After resolving these cases, we reduce the parse error rate to 0.24%. Therefore, we are able to annotate multiple packages with package information in a single identifier with high confidence.

D. Unifying Call Graphs

To generate a single CDN, we merge nodes with the same UFI (from different call graphs) into a single node. Without merging nodes, a naïve concatenation of all individual call graphs resulted in a graph with 60,410,714 nodes and 178,308,144 edges. After merging nodes on the function name, we reduced the graph to 7,034,536 nodes and 19,511,485 edges, merging on average 8.6 nodes onto a single node. In total, 6,983,046 function nodes have an origin in CRATES.IO. The 51,490 remaining nodes are standard library functions from core, std, or packages which are not hosted on CRATES.IO. Overall, we have a definition for 6,882,760 nodes, 97.8% of all nodes, i.e., we could expand their internal call flow similar to io::crates::Lib2::used in Figure 4. This high percentage demonstrates the internal completeness of the buildable part of CRATES.IO and that merging nodes, core to PRÄZI, is useful and occurs often.

E. Evaluation of the Rust CDN

One purpose of the Rust CDN (and PRÄZI at large) is to present a more precise view of the dependency relationships between software components in comparison to a traditional PDN. Since both networks operate on a different abstraction level, we cannot directly compare them. We can, however, reverse-engineer a new, possibly more precise PDN called PCDN ⊆ PDN by “uplifting” the CDN. Only then can we compare it to the original PDN. We construct PCDN by including a package dependency ⟨A, B⟩ in PCDN’s edge set if there is at least one call from a function in package A to one in B in the CDN. This, of course, sacrifices precision in the CDN and thus is an absolute lower bound for possible improvements.

As a caveat, the PRÄZI perils may result in missing calls across packages in the CDN, which in turn makes the PCDN miss dependency relationships. Therefore, our evaluation focuses on quantifying the differences between PDN and PCDN and qualitatively exploring their causes. To make the comparison fair, we remove from the PDN all packages that could not be compiled because those could never appear in the PCDN. We then extract the subset of dependency links present in the CDN, but absent in the PCDN. For each dependency, we determine through manual code inspection whether it is correctly or incorrectly absent in the PCDN.

The PCDN contains 42,827 nodes and 110,762 edges. The PDN contains the same number of nodes but has 129,535 edges. A set difference on the edges of the two networks shows that 18,042 edges (i.e., 14%) are not in the PCDN. Qualitative evaluation of all 18,042 different edges is practically infeasible, as it relies on manual work. Instead, we select a statistically representative subset of its edges using Cochran’s sample size formula [36]. Selecting from a homogeneous set of edges, at a 95% confidence level with a confidence level interval of 5%, we need a sample of \( n = 381 \) edges that the first author investigated. In Table VI, we break down the results into dependencies that i) should be and are absent in the PCDN and ii) should be present in it, but are not.

Our qualitative evaluation shows that our Rust CDN can identify several cases (35%) where a regular PDN would definitely lead to false positives reported to developers. At the same time, there is a substantial amount of dependencies that the PCDN seems to fail to capture. While 65% sounds discouraging for RUSTPRÄZI at first, in the following, we will characterize and explain them in detail, and show that none are a theoretical limitation of PRÄZI.

Of the missing dependency links, almost half are due to shortcomings in the LLVM call graph generator. They come as no surprise, as we found out that we can only claim soundness for static dispatch (see Section V-B). Plugging a better call graph generator in, switching to a language with better generators (such as Java), or not using these language features thus resolves the problem, which is not inherent to PRÄZI. The remaining cases represent possible future improvements to PRÄZI. A fifth of the missing cases are data-type only dependencies that a call graph cannot capture, e.g., importing a struct of another package. This suggests going beyond call graphs for PRÄZI. Another third of the missing dependencies are due to conditional compilation: a function is only compiled-in if the appropriate feature toggles are on. We mitigate this issue when, we rerun the call graph generation process for every possible feature toggle combination. We have thus shown that none of the missing cases is a principal shortcoming of PRÄZI.

To verify the generalizability of this evaluation, the first two authors conducted an inter-rater reliability study. We cross-validated 20 randomly selected pairs of dependencies. After an independent assessment and comparison of the results,
both raters agreed that 19 ratings of the main rater were correct \( p_0 = \frac{19}{20} = 0.95 \). The naïve likelihood of a random agreement is \( p_c = 0.5 \). This gives us a Cohen’s \( \kappa \) of \( \frac{p_0 - p_c}{1 - p_c} = \frac{0.95 - 0.5}{1 - 0.5} = 0.9 \) \[37\], which signifies (almost) perfect agreement \[38\], increasing trust in the correctness and generalizability of the manual inspection.

VI. CASE STUDIES

To demonstrate the effectiveness of our approach, we present two cases studies, namely security vulnerability propagation and function deprecation.

A. Security Vulnerability Propagation

Perhaps the most common application of dependency networks is the study of the spread of security vulnerabilities \[4\], \[7\], \[39\]. Companies, such as BLACKDuck \[13\], TIDELIFT \[11\], and GITHUB help projects identify whether they are affected by publicly disclosed vulnerabilities in their dependencies.

With our first case study, we aim to assess whether the call-based representation of an OPR could yield higher precision in the security vulnerability propagation case and how severe its soundness issues would be in such a real-world test. We compute how nine security advisories from Rust’s security advisory database, RUSTSEC \[40\], affect other packages, using both our Rust CDN and PDN. Table \text{VII} presents a detailed overview of the advisories we examined along with the number of affected package versions per network. From the initial set of nine advisories, we could not analyze the security-framework advisory because it is macOS-specific, sodiumoxide because its vulnerable versions generate build failures, and the openssl advisory because it is related to configuration rather than vulnerable code. From the six remaining ones, we skip cookie because it does not have any callers and smallvec, whose vulnerability is in a generic function, which the current version of the Rust CDN cannot cover due to limitations in the call graph generator (see Section \text{V}-E).

By construction, our Rust CDN is precise, but could miss function calls to dependent packages, while the Rust PDN is sound, but overapproximates the number of used packages (see Section \text{V}-E). To compute the accuracy of the two networks, we need to establish a ground truth: for each security advisory, we collect the direct dependents of the vulnerable package; then, we manually investigate whether there exists a function call from a dependent to any function of the vulnerable package. We compare the Rust PDN and the Rust CDN (converted to a PCDN, as in Section \text{V}-E) against the ground truth and create a confusion matrix:

i) **True Positive** (TP) means correctly flagging a package as vulnerable when a vulnerable function call exists.

ii) **False Positive** (FP) means falsely flagging a package as vulnerable when there is no invocation of a vulnerable function.

iii) **False Negative** (FN) means flagging a package as not vulnerable when there is at least one evident call of a vulnerable function.

iv) All remaining cases are **True Negatives** (TN).

Finally, we use the standard binary classification metrics precision, recall, and accuracy to compare their performance.

From the results in Table \text{VII} we observe that the PCDN reports a much lower number of affected package versions, on average, 83% fewer affected packages than the PDN. By analyzing the direct dependencies of the vulnerable packages, we establish that a high percentage of these affected packages are in fact false positives in the PDN. This score would be even more in the advantage of our PCDN were we to look at the full transitive closure (which we have not done because of the high manual workload associated with it). Although our PCDN has a lower recall score than a PDN, our PCDN has an accuracy which is three times higher than the state-of-the-art PDN, signifying that CDNs yields very high precision benefits over traditional PDNs, even in real case scenarios with suboptimal tools (see Section \text{V}).

B. Deprecation Impact Analysis

As packages evolve, their public API changes to accommodate improved functionality. As a consequence, functions can become obsolete. Several programming languages, have a special mechanism to annotate obsolete functions, either in the API documentation (e.g., in Python) or as a language feature (e.g., in Java). While annotating functions as deprecated is a common practice among developers, cleaning up deprecated code is a far more challenging task. In a qualitative study, Sawant et al. report that API producers are “wary about removing deprecated features from their API” and “mostly have no preset protocol for removal” \[41\]. The reason is that developers cannot know the impact of such cleanups.

By linking dependent functions together, PRÄZI enables us to perform change impact analysis at the OPR level. Using an PRÄZI CDN, developers can estimate the impact of removing deprecated functions, both to direct API clients and transitively. To demonstrate this, we calculate the impact of function deprecation within CRATES.IO.

In Rust, deprecated functions can be annotated with a \# [deprecated] attribute: the Rust compiler will fail compilation if a program links to a deprecated function (unless a \# [allow(deprecated)] is specified). To find deprecated functions, we extract function signatures prepended with a \# [deprecated] attribute. In total, we find 721 deprecated function signatures from 190 package versions in 43 unique packages. We only consider deprecated functions and their callers, i.e., functions who call the deprecated functions directly. Among the 190 package versions, only 42 package versions have callers, reducing our search space to 43 deprecated functions. Then, we manually match the deprecated function to its UFI in the Rust CDN. We are able to find a UFI for 24 deprecated functions; the remaining ones are missing due to the reasons identified in Table \text{VII}. We perform a propagation analysis (similar to Section \text{VI}-A) for the 24 cases. In total, 13 of 24 deprecated functions have calling functions in other packages and together affect 163 package versions,
TABLE VII: Results for the security advisory propagation analysis.

<table>
<thead>
<tr>
<th>Package</th>
<th>Function</th>
<th>PDN Affect</th>
<th>PCDN Affect</th>
<th>#Packages</th>
<th>PDN Affect</th>
<th>PCDN Affect</th>
<th>#Cases</th>
<th>PDN Affect</th>
<th>PCDN Affect</th>
<th>#Cases</th>
<th>PDN Affect</th>
<th>PCDN Affect</th>
<th>#Cases</th>
<th>PDN Affect</th>
<th>PCDN Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>base64</td>
<td>encode_config_buf</td>
<td>145</td>
<td>51</td>
<td>227</td>
<td>129</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
<td>0.25</td>
<td>0.68</td>
<td>0.25</td>
<td>0.25</td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cookie</td>
<td>parse_inner, max_age</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hyper</td>
<td>Headers::set</td>
<td>21</td>
<td>3</td>
<td>2</td>
<td>1.00</td>
<td>0</td>
<td>1</td>
<td>1.05</td>
<td>1.05</td>
<td>0.5</td>
<td>1.05</td>
<td>1.05</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>smallvec</td>
<td>insert_many</td>
<td>1,581</td>
<td>0</td>
<td>325</td>
<td>61</td>
<td>61</td>
<td>1</td>
<td>0.62</td>
<td>0.62</td>
<td>1.00</td>
<td>0.71</td>
<td>0.71</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tar</td>
<td>unpack_in</td>
<td>502</td>
<td>31</td>
<td>61</td>
<td>291</td>
<td>0.25</td>
<td>1</td>
<td>0.43</td>
<td>0.43</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>untrusted</td>
<td>skip_and_get_input</td>
<td>5,655</td>
<td>564</td>
<td>61</td>
<td>291</td>
<td>0.25</td>
<td>1</td>
<td>0.43</td>
<td>0.43</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Σ (or weighted average)</td>
<td>8,016</td>
<td>649</td>
<td>482</td>
<td>1</td>
<td>1</td>
<td>0.30</td>
<td>0.53</td>
<td>0.53</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VII. IMPLICATIONS

While researchers and practitioners widely use PDNs meta-information for various decision making tasks, e.g. related to updates [7] or security [39], it seems that their precision is a largely overlooked aspect. Indicatively, in the case of CRATES.IO, 35% of the dependency links included in a dependency network extracted from metadata should not be there. The work we present in this paper uncovers several implications in the way that both researchers and practitioners use PDNs, which we briefly present below.

A. Implications for Researchers

Our study showed that the generation of sound, yet precise call graphs remains an important research problem for the feature-richness of modern languages. The quality of a PRázI CDN is directly dependent on the quality of the call graph generator. In the Rust case, a Rust-specific call graph generator needs to be developed as a plug-in to the compiler in order to accommodate for features such as macros that are only visible during compilation, and ii) be always up to date with the latest language features.

We also demonstrated the volatility and fast-paced nature of dependency resolution in OPRs by comparing timed snapshots of CRATES.IO. It seems crucial to include this aspect in future studies on OPRs.

Finally, we showed the granularity and precision that PRázI offers by opening the door to many new analyses not possible before, for example change impact analysis across an OPR. There are several strands of research opportunities to improve PRázI internally (which we list in Section V-E) or externally, e.g. by applying it to a different programming language.

B. Implications for Practitioners

With PRázI, we have described and implemented a technique that can reduce the number of false positive in a wide array of current applications that suffer from bad precision. Practitioners could improve our prototypical RUSTPRázI implementation to make it sound in all cases relevant to them.

Due to the large number of broken packages on CRATES.IO, builders of OPRs should consider validating packages before publishing them, similar to CPAN [42] or CRAN [43]. Moreover, to make the entire build chain setup of packages more reproducible, builders of OPRs should perhaps have similar goals to what Debian wants to achieve [44].

Many IDEs support a “remove unused imports” analysis. PRázI allows the development of a “remove unused library” feature, cleaning and optimizing the dependency set of projects.

C. Threats to Validity

In the case of security, it is important for PRázI to guarantee that no false negatives exist, i.e., that the analysis is fully sound. The PRázI technique is only susceptible to this threat insofar as the used call graph generator is sound, which lies outside the scope of this paper. Our initial Rust implementation guarantees soundness for statically dispatched method calls, but is only “soundy” otherwise [45]. If Rust programmers avoid dynamic dispatch, our prototype is sound by construction (we have specified and measured the exceptions in Table VI). Moreover, security warnings only serve as an example application for meta-warnings attached to a program. Similarly, bug, performance, deprecation, or other advisories exist, e.g., in addition to RUSTSec’s security advisories, similar such advisories exist for performance and semantical

TABLE VIII: All called deprecated functions.

<table>
<thead>
<tr>
<th>Function</th>
<th>Package</th>
<th>PCDN</th>
<th>#Affected by Dep. Funcs</th>
</tr>
</thead>
<tbody>
<tr>
<td>OwnedKVList::&lt;new/ids/root&gt;</td>
<td>slogs::1.7.1</td>
<td>93</td>
<td>63</td>
</tr>
<tr>
<td>platform::window/display</td>
<td>wininit::0.7.6</td>
<td>91</td>
<td>50</td>
</tr>
<tr>
<td>platform::window/display</td>
<td>wininit::0.9.0</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>platform::window/display</td>
<td>wininit::0.8.3</td>
<td>44</td>
<td>14</td>
</tr>
<tr>
<td>platform::window/display</td>
<td>wininit::0.6.4</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>get_formats_list, get_name</td>
<td>cpal::0.4.6</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Σ</td>
<td>13</td>
<td>6</td>
<td>311</td>
</tr>
</tbody>
</table>
V. CASE STUDIES

The case studies are based on an implementation of the PRAZI technique. We demonstrate PRAZI's usefulness to mitigate the issues arising from conditional compilation by two case studies:

- **Case Study 1:** A Rust library with 20 versions, which includes changes in the API and the codebase. We show how PRAZI can help researchers understand the impact of these changes.
- **Case Study 2:** A Java project with 15 versions, involving changes in the project's dependencies. We demonstrate how PRAZI can help analyze the propagation of changes in the dependency network.

In both cases, PRAZI helps in understanding and mitigating the impact of changes, providing insights that are not available with other tools.

VI. DISCUSSION

The PRAZI technique is a novel approach for creating dependency networks at the function level. While it is considered a state-of-the-art technique, there are some limitations:

- **Memory footprint:** Creating PRAZI requires significant memory, especially for large projects. However, with cloud storage solutions, this can be managed.

- **Computational cost:** Generating PRAZI can be computationally expensive, especially for projects with a large number of versions. However, the tool is optimized to handle this issue.

- **Practical use:** While PRAZI can be useful in a range of scenarios, its practical use might be limited by the availability of the implementation and the need for expertise to use the tool effectively.

VII. RELATED WORK

In this section, we briefly present previous PDN-based analyses and compare them to our work.

The aftermath of the leftpad incident has led to a surge of studies around OPRs. Researchers have constructed dependency networks of OPRs to trace the impact of security problems. To study the evolution of language ecosystems, or to recommend update paths for projects.

In the area of security, notably, Kikas et al. have shown there exist packages that can break up to 30% of packages in both npm and rubygems. Moreover, Kula et al. and Decan et al. also indicate a large percentage of affected packages. However, our security case study has shown that these studies may grossly overapproximate the risk.

To the best of our knowledge, ours is the first attempt to construct a dependency network on a function call level. The closest to our work is an initial vulnerability study by Zapata et al. The authors manually established that 73.3% of 60 JavaScript projects marked as vulnerable by a dependency checker are not. The results are on similar lines as ours, saying vulnerable functions in many cases are not called.

IX. FUTURE WORK

PRAZI is a novel technique that is first to apply lightweight static analysis on whole OPRs. While it is comprehensive and can already provide valuable insights, our evaluation showed that it is not complete. In this section, we present how PRAZI can be improved and practically exploited.

First, the buildability of an OPR plays a crucial role in the completeness of the PRAZI CDN; researchers could setup several build environments to generate a maximally complete call graph. PRAZI can be combined with product line research to help mitigate the issues arising from conditional compilation.

The soundness and completeness of the PRAZI CDN can be significantly improved through language-specific call graph generators. In several mainstream languages, e.g., Java, precise call graph generators with few constraints exist, such as Soot and Wala. For the purposes of PRAZI, researchers could work on enriching the static with dynamic call graphs obtained from running a project’s tests.

To make a full PRAZI implementation practically useful, the size of the generated CDN must be tamed to cope with OPRs. The case study data is available as a replication package.

This paper makes the following contributions:

- A novel technique, PRAZI, to create precise, call-based dependency networks.
- An open source Rust implementation, PRAZI, to show its practical feasibility.
- An evaluation of PRAZI and quantification of its shortcomings.
- Two case studies on security and deprecation warnings, that demonstrate PRAZI’s usefulness.
- Derivative datasets including the CDN, its evaluation, and case study data is available as a replication package.

https://DrNXs1ALFzzQdh4r.github.io