PTA: Finding Hard-to-Find Data Plane Bugs

Pietro Bressana, Noa Zilberman, Robert Soulé

***Abstract*—**
Bugs in network hardware can cause tremendous problems. However, programmable network devices have the potential to provide greater visibility into the internal behavior of devices, allowing us to more quickly find and identify problems. In this paper, we provide a taxonomy of data plane bugs, and use the taxonomy to derive a Portable Test Architecture (PTA) which offers essential abstractions for testing on a variety of network hardware devices. PTA is implemented with a novel data plane design that (i) separates target-specific from target-independent components, allowing for portability, and (ii) allows users to write a test program once at compile time, but dynamically alter the behavior via runtime configuration. We report 12 diverse bugs on different hardware targets, and their associated software, exposed using PTA.

***Index Terms*—debugging, testing, network programmability (SDN/NFV/In-network computing)**

I. Introduction

Bugs in network hardware can result in financial loss, security breaches, or significant downtime for essential services. Unfortunately, despite extensive testing, these bugs can be very difficult to find [1].

Example. As an example, imagine that a device drops packets when the input traffic exceeds a certain rate. How would we find and diagnose this bug? This sounds like it would be a simple bug to catch. After all, we would certainly notice packet drops. In reality, the bug—which we found in the NetFPGA [2], [3] reference projects—was not discovered for more than a decade after the platform had been introduced.

The root cause of the bug was that the input arbiter in the design was not work-conserving, i.e., packets were held in an input queue even when the output was idle. However, the bug did not reveal itself on the NetFPGA 1G board with $4\times$1Gbps interfaces, or on the SUME board with $4\times$10Gbps interfaces. It was only revealed when the design was ported to a $2\times$100Gbps Alveo board. The bug in the design was passed from one generation to the next, and the capacity of the network interface masked the defect in the internal design.

So, how could we have found and fixed this bug sooner? There are a few immediate observations that we can make.

First, the bug only appears when the traffic rate exceeded a threshold, in this case, $\sim$40Gbps aggregate throughput on SUME. So, software-based approaches like simulation or emulation, which can slow the execution of a program by a factor of 10^6 [4], would not help. Instead, we need a test framework that can generate and receive traffic at line rate.

P. Bressana is with Intel Corporation, 8048 Zürich, Switzerland
N. Zilberman is with the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK
R. Soulé is with the Department of Computer Science, Yale University, New Haven, CT 06511 USA

Second, finding this bug requires *internal* access to the data plane. Even if we could externally generate and send traffic to the device under test at the target rate (e.g., using an Ixia [5] or Spirent [6] platform), we need a way to distinguish a limitation of the network interface from the inefficient implementation of the input arbiter.

Third, we see that the same reference design was used on several hardware targets, including NetFPGA 1G, 10G, SUME, and Alveo. Writing tests is time intensive, and having to repeat test-writing efforts for each target would be onerous. Just as the reference design can be ported across targets, we want the tests to be portable across hardware targets.

Although we have focused on an FPGA in this example, similar bugs and observations hold for programmable ASICs, although at greater scale. Consider trying to replicate the same test scenario on a Tofino 2 ASIC, which has $128\times$100G ports.

Prior Work. Verifying, debugging, testing, and validating network hardware is a well studied area. A range of software-based approaches have been proposed, including simulation or emulation and formal verification [7], [8], [9]. However, none of these offer a comprehensive solution. Simulators or emulators may not faithfully model actual deployments, and, as already mentioned, cannot test scale-related bugs. And, verifiers cannot catch several types of bugs, such as bugs in the compiler, performance bugs, or bugs to due to under-specification in the language.

Therefore, network operators often augment software-based techniques with hardware-based testing. For this purpose, equipment vendors such Ixia [5] and Spirent [6] sell highly-specialized platforms, which can generate and receive traffic at line rate. Unfortunately, these devices provide limited visibility because they function as external black-box testers. Moreover, the cost of such platforms is considerable. Prior research efforts [10], [11] offer lower cost solutions with similar intents, but they are limited, in terms of features, scale, and performance, e.g., OSNT [10] does not scale beyond 4 ports, and none of them work with non-Ethernet packets.

Problem and Approach. This paper addresses the problem of how to develop a high-performance, comprehensive, portable test framework for network devices. The key idea is to leverage a portion of the resources in programmable network hardware—including SmartNICs and programmable ASICs—for testing. Programmable network hardware is an attractive option for use with testing for two reasons. First, it can send and receive traffic at high rates by design. Second, it can be adapted for use-cases beyond traditional forwarding, such as in-network computing [12], [13], [14], [15].

Challenges. Using programmable network hardware as testing devices, rather than forwarding devices, presents a significant challenge, because testing and forwarding are fundamentally

different. In particular, we identify three, high-level challenges: (i) active vs. reactive logic, (ii) dynamic processing behavior, and (iii) portability.

First, at the most basic level, forwarding devices are reactive, meaning that they execute logic only on the arrival of an incoming packet. In contrast, testing is an active process. A tester generates test stimuli in the form of test packets, and then checks a post-condition.

Second, testing devices require much more flexibility than forwarding devices. When used for forwarding, the data plane functionality of programmable NICs and switches only changes in limited ways, e.g., it might forward packets out a different port, depending on control plane configurations. But, the forwarding pipeline is not altered during operation.

In contrast, exhaustive testing often requires significant adaptation and permutation, dynamically changing the behavior depending on the needs of the test. As an example, imagine that we want to generate a variety of packets with different header sizes, similar to how Dumitru et al. [16] check for security exploits. Changing the data plane implementation of the test program for every permutation would result in significant overhead, in terms of compilation and installation, which can take hours on some platforms.

Third, the test architecture must be portable across a range of heterogeneous target devices. To provide portability, we need to identify a set of abstractions that are flexible and powerful enough to test for a variety of possible data plane bugs, but can be generally implemented on a range of devices.

Contributions. To address these challenges, we propose a new *data plane architecture* for data plane testing. We use the term data plane architecture in the same way that it is used in the P4 programming language [17]. It identifies the programmable blocks and their data plane interfaces. Essentially, it is a contract between a data plane program and a hardware target.

The P4 open-source community has begun to standardize a few data plane architectures, including the Portable Switch Architecture (PSA) [18] for network switches, and the Portable NIC Architecture (PNA) [19] which models NICs. This paper introduces the Portable Test Architecture (PTA).

Overall, this paper makes the following contributions:

- The requirements for PTA are derived from a taxonomy of bug types in programmable network devices and we detail bugs that we have found in commercial and open-source software and hardware using the tool.
- Driven by the requirements of the bug taxonomy, PTA offers a small but powerful set of abstractions to support debugging.
- PTA has a novel data plane design that: (i) separates target-specific from target-independent components, allowing for portability, and (ii) allows users to write a test program once at compile time, but dynamically alter the behavior via dynamic re-configuration.
- PTA complements prior work on automatic test packet generation [20], fuzz testing [21], [22], [23], and software validation [8], by providing a framework for running workloads generated by those tools on actual hardware. To demonstrate how PTA can be used in conjunction with existing tools, we have developed a proof-of-concept integration with P4v [8]. Users can extract assumptions and assertions from an annotated P4 program, and map them to a hardware test configuration.
- PTA uses programmable network hardware for testing, which differs from traditional forwarding in key ways. We present a set of lessons we've learned and assumptions that were challenged in the design of the framework.
- PTA is publicly available under an open-source license [24].

Key Results. We have implemented PTA for two different hardware targets: the NetFPGA SUME platform [3] and the Barefoot Tofino ASIC. We have used the framework to evaluate several P4 programs and two P4 compilers. Using PTA, we were able to identify 12 diverse bugs. These bugs are drawn from a broad spectrum of classes of bugs, demonstrating that PTA provides a comprehensive testing solution. Moreover, these bugs were in heavily-used, heavily-tested commercial and open-source systems.

This paper extends our earlier conference paper [25] with a more detailed description of the PTA reference design. It also includes an expanded discussion of the bugs found with PTA.

II. Requirements and Constraints

The design of PTA navigates the tension between developing an expressive framework that can test for a wide range of bugs, but can be implemented on a diverse set of hardware targets. Below, we discuss these requirements in more detail by first developing a taxonomy of error types and then discussing the constraints imposed by different hardware.

A. Data Plane Bug Taxonomy

A wide range of bugs can occur in network devices. These bugs can be due to incorrect program logic (i.e., functional bugs); or due to problems in compiler, target hardware architecture or others. Below, we provide a taxonomy of the types of bugs that a test framework must be able to detect. These error types provide requirements that motivate the design of PTA. Note that although PTA can be used to test both fixed-function and programmable hardware, our taxonomy highlights bugs that may be unique to programmable network hardware (e.g., compiler bugs), and may not be comprehensive.

Functional Bugs. A functional bug is one in which the functionality provided by the network device is not the same as the functionality intended by the programmer. Functional bugs can occur in both the data plane and in the control plane. An example data plane bug would be not supporting IPv6 headers where such functionality was supposed to be supported. An example control plane bug would be not filling all the required entries in a given size table.

Performance Bugs. Performance bugs are related to aspects such as the maximum throughput or packet rate of a certain design, how certain packet sizes affect the throughput, whether congestion control is handled properly, and more. For performance testing, for example, the user must be able to continuously fill the pipeline with packets of a certain size and check that no packets are dropped or lost at the output. Another performance aspect is the ability to mix packet sizes

in explicit ways, which exercise different parts of a design (e.g., programmable data plane, schedulers, memory access).

Compiler Bugs. Although compilers are tested with scrutiny, there may be bugs. There are at least two classes of compiler bugs. The first class of errors regards functionality bugs, e.g., where a language feature is supported but the implementation is missing, or the functionality is implemented incorrectly. A second class of errors covers the compliance with the programming language specification.

Under-Specification Bugs. The extent of a programming language definition, and the diversity between target platforms, leads to cases where the language specification is not detailed, either intentionally [26] or not. This can lead to unexpected or unintended behaviors, for example, if the specification does not detail whether the initialization of a header should be to zero, or if can remain unpopulated and random.

Architecture Bugs. Similar programs may target different data plane architectures, and even perfect programs may be susceptible to bugs in the underlying device architecture. One immediate class of device architecture limitations is access hazards to tables, such as read-after-write. A second class of bugs uncovers limitations of the data plane architecture, such as a proprietary module (e.g., an extern) that is not responding within the expected time. Another class of bugs relates to the integration of different modules in the architecture, such as a mismatch in the connection of interfaces.

Security Vulnerabilities. Network devices can suffer from security vulnerabilities just like any other device, and programmable network devices introduce new threat vectors. Security vulnerabilities are commonly the result of a different class of bugs, and are highlighted due to their importance and the need for targeted tests. A test framework should allow users to quickly and efficiently test a large number of such security threats. One such example would be looking for the "Meltdown" [27] equivalent of a programmable data plane: can you craft a packet that would allow you to read the contents of previous packets, various tables, or memories? Another example is backdoors in the program, whether in the original users code or introduced as a by-product of the compilation process. The hardware test can reduce the security risk by testing the deployed program as it runs on the platform.

B. Heterogeneous Targets

The diversity of data plane bugs implies that a testing architecture should be flexible and expressive. However, the design of the architecture is necessarily constrained by the capabilities of the target hardware. We briefly summarize these below. We focus our discussion on two devices that are on opposite, extreme ends of the spectrum: FPGAs and ASICs. Other types of network devices, such as those based on System on Chip (SoC), fall between these two extremes [28].

FPGAs. Field Programmable Gate Arrays (FPGA) have a given set of resources, but provide users with extreme flexibility and full programmability. As long as a design does not exhaust resources, and users can compile the design while maintaining the constraints they set (e.g., on timing), FPGAs can implement almost any logical operation, with different levels of complexity.

ASICs. Like FPGA, ASICs also have a finite amount of resources. But, in contrast to FPGAs, they have a set device architecture. While ASICs have become programmable—significantly more so than in the past—their programmability is constrained to the architecture. Note that CPU architectures impose similar constraints, e.g., programming on an x86 CPU is different from programming an ARM core or RISC-V. The main advantages of switch ASICs over FPGA-based switches is that they achieve much higher clock rate (and therefore higher throughput), offer increased scale (e.g., number of ports), and use resources more efficiently.

Constraints. These differences between hardware targets hinder portability. It is often said that P4 allows users to write target independent programs. But, this is not true. A program written in P4, like programs written in other languages, is tied to the target architecture. Examples of architecture specific properties include externs, initialized values (of registers, memories and other stateful elements), and timestamp taking, among others. Portability issues are not always a property of complex hardware design. They can result from mundane aspects, such as the number of bits assigned on the metadata bus to indicate the egress port number (which may differ between an 8-port switch and a 256-port one, in order to minimize resource usage).

III. Debug Abstractions

One of the main challenges in designing PTA is identifying the core set of abstractions to support debugging. We adopt a requirement driven design process. Based on the taxonomy in the previous section, we systematically explored the necessary abstractions for each of those classes of bugs. The set of abstractions is intended to be minimal, so that it can be readily supported by diverse hardware. At the same time, it is intended to encompass the set of functions needed for testing.

To illustrate the process, we first walk through the running example—i.e., the input arbiter bug from the NetFPGA reference project from Section I—before summarizing the complete set of PTA debug abstractions.

A. Requirement Driven Design

So, how might a developer find and isolate the performance bug in the input arbiter? Because the module is (incorrectly) not work conserving, we clearly need to be able to generate packets at data-path rate, creating controlled back-to-back arrival events to the arbiter.

Many bugs (e.g., functional, compiler) would depend on a particular data plane program, suggesting that the debug framework needs a method to load a data plane image. However, in this case, the bug is in the architecture of our target device, and therefore independent from the data plane program that we would load. To test the architecture, we need access to the low-level abstractions offered by the hardware, including the metadata bus, stateful ALUs, and any externs provided by the architecture of additional hardware modules.

Abstraction	Description
Load_Image	Load the image file to the target
Init_Counters	Initialise the counters in the target
Init_Registers	Initialise the registers in the target
Generate_Packets	Generate test packets
Collect_Results	Collect raw results from target's registers

Table I: PTA's user-facing abstractions.

Abstraction	Description
Metadata_Bus	Layout of the metadata bus
Stateful_ALU	Architecture of the Stateful ALUs
Extern	Architecture of the Extern modules
Register_Read	Interface for reading hardware registers
Register_Write	Interface for writing hardware registers

Table II: PTA's back-end abstractions.

We need modules to initialize and check the values of stateful elements, e.g., counters and registers. This allows us to confirm the number of packets sent and processed by the pipeline, and more generally, application specific logic.

Finally, to detect the presence of dropped packets (again at line rate), we need a way to collect and inspect output packets.

B. Core Abstractions

By following this requirements driven design process, we identify two classes of abstractions: user-facing abstractions (Table I) and back-end abstractions (Table II). User facing abstractions are used to specify the functionality of a test. Back-end abstractions represent the target device's architecture .

User-Facing Abstractions. Users writing tests will be using user facing abstractions, similar to functions. As these abstractions are not target-specific, a test will be written only once. The abstractions are used to load the program image (`Load_Image`), initialize registers (`Init_Registers`) and counters (`Init_Counters`) and to generate and collect packets (`Generate_Packets` and `Collect_Results`). Note that although packet generation is exposed via user-facing abstractions, it is adapted to a target (e.g. to vary transmission rate) using user transparent back-end abstractions.

Back-End Abstractions. Back-end abstractions are used to specify a network-device target, and are called by any tests using this target device. The back-end abstractions library includes both the layout of the metadata bus (`Metadata_Bus`), that is used by PTA as a configuration channel, and the architecture specification of both stateful ALUs (`Stateful_ALU`) and extern modules (`Extern`). Since the hardware components of the framework are usually accessed through a register interface, PTA provides two additional abstractions for reading and writing registers (`Register_Read` and `Register_Write`).

IV. Portable Test Architecture

Building on the core abstractions, PTA provides a comprehensive hardware data plane testing solution. PTA is *programmable*, meaning that the tool can be customized to the particular testing needs of the user for a diverse set of bugs. It is also *re-configurable*, meaning that new tests can be run via dynamic re-configuration (e.g., using register access), rather than re-programming (e.g., requiring a new image file). PTA allows for *integration with existing tools*, providing prior work on automatic test packet generation [20], fuzz testing [21], and software validation [8] with a path to run on hardware. When used to test devices with programmable data planes, PTA allows *access to internal state*, providing detailed fault localization. PTA allows users to test network devices in *real time* at full line rate, and test results are *reproducible*. We expect PTA to be deployed out of band, i.e., in parallel to live traffic. It does not incur additional latency or alter the traffic.

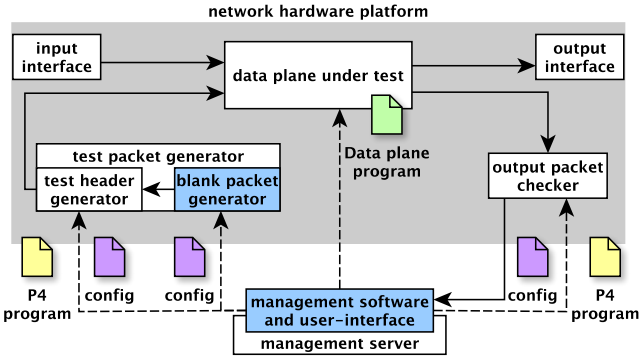

Figure 1: The proposed architecture: target specific infrastructure (light-blue), portable suite of P4 test programs (yellow) and test-specific configurations (purple).

A. Overview

Imagine that a user wants to verify a certain data plane. We assume that the user has some information about the data plane functionality, e.g., the P4 program, but not all information of the dataplane, e.g. hardware-target's micro-architecture. The user will need to devise a test plan that covers the range of potential bugs, covered in Section II-A. Most of these tests are generic (e.g., performance tests) while some are use-case specific (e.g., P4-program specific functional tests). Next, following this plan, imagine that the user wants to verify that this data plane runs at line rate for different packet sizes. To run this test, we need three components: a packet generator, to inject packets into the data plane; an output checker, to assert that post-conditions hold at the end of the test; and a management component, to run the test. All these components are illustrated in Figure 1, which shows the high-level design of PTA. Both the generation and checker modules are implemented in hardware, while the **management component** is a set of software programs.

Even for this simple example, there are a range of parameters and scenarios to be tested: How many packets should be sent? What sizes should the packets be? At what rate should be packets sent? And, at what rate do we expect the output packets to arrive? What protocols are being used in the packet headers? Hand-writing tests for each of these scenarios would be tedious, and possibly error-prone.

To help reduce the burden, PTA separates tests into two parts: the programmable part and the re-configurable part. The programmable part can be thought-of as data plane specific. Users can, for example, write a program to generate packets with different protocol headers, and write a checker to validate the emitted headers. The **re-configurable part** is test-specific. It is a control plane configuration of the programmable part, that allows users to change parameters such as packets sizes, sending rates, etc.

The programmable part of a test can be divided into infrastructure that is target-independent and target-dependent. The **target-independent** infrastructure is written in P4, and controls, for example, the definition of protocol headers. The **target-dependent** infrastructure is the device-specific functionality (e.g., generation of blank packets).

Note that although PTA's programmable parts are implemented in P4, the data plane under test does not need to be written in P4. PTA can be used to test data planes designed using a variety of different workflows and languages, including high level synthesis, C/C#, and HDLs (e.g., Verilog).

It is important to stress that all of the components of PTA are implemented *inside* the target network platform. This provides PTA with several important advantages. First, it allows PTA to test the data plane while avoiding the surrounding hardware, including the network interfaces. A failure of a test can guarantee that the cause is not in the interfaces but in the tested data plane. Second, it enables testing the device at line rate and at real time. Testing a device at line rate is challenging due to the cost of external traffic generators (e.g., Ixia [5]), making it outside the reach of many users. Thus, the internal data path may have a certain speed-up over the external interfaces, making it very hard to create and detect hazard scenarios such as read-after-write in two consecutive clock cycles, or certain cross-traffic scenarios that lead to consistency issues. Finally, PTA allows users to test and debug their data plane in the field, without additional equipment, and without changing the physical settings.

B. Target-Independent Test Infrastructure

The target-independent infrastructure of PTA is a set of P4 programs used for packet header generation and output checking. Both components are implemented as a sequence of match-action tables and a set of registers that change control flow. The entries in these tables are re-configurable, and provide the flexibility to support different tests.

PTA includes a set of default programs that developers may use to generate packets with standard protocols (e.g. Ethernet, IP, TCP, UDP, etc.) and to check for common conditions. Thus, for many test scenarios, users need not write any P4 code themselves. To support custom protocols and to check for data plane specific test scenarios (e.g., to generate a packet with a Paxos protocol header [15] and test for a specific post-condition), users can expand on these default programs for custom-protocols using P4.

Packet Header Generator. The packet header generator takes blank input packets, and turns them into stimulus packets injected to the data plane under test. The P4 program defines the protocols that need to populate the header and properties of the contents. Because tests are written in P4, developers can use any protocol that is implementable in P4, and can easily add custom headers, different fields, change the ordering of headers, and more. For example, an empty packet entering the test header generator will be emitted as a standard TCP/IP packet, with a certain sequence number and valid checksum. Combined with a blank packet generator, the test header generator will control packet size and contents, traffic pattern (e.g., inter-packet gap), and may even intentionally craft illegal stimulus packets. The output of the test header generator connects to the input of the data plane under test.

Output Packet Checker. The output of the data plane under test is connected to the output packet checker. The output packet checker, implemented in P4 and shown in Figure 1, can be programmed to expect specific values or sequences of values within the returned packets. It compares these values against the input traffic stream (e.g., to detect packet drop, reordering, or other points of failure). The stages within the checker's stages support different types of functionality, such as matching specific header fields, or comparing metadata bus values. The outcome of each check is stored in a memory. Typical types of stages include an ALU, that performs both logic and arithmetic operations over headers and metadata, and CAM and TCAM blocks that compute simple matches against header and metadata fields. The number of received packets is an example of a common functionality implemented using counters.

C. Target-Dependent Test Infrastructure

P4-based data planes are packet driven, and do not generate packets without a stimulus. For this reason, PTA uses a blank packet generator that creates empty packets, feeding the test packet generator's P4 pipeline. By a blank packet, we mean a packet with no header fields and no payload. The blank packet generator is target specific. For example, some ASIC switches (e.g., Tofino) already have a built in-packet generator, while other devices (e.g., FPGA) require a dedicated implementation. Even if a packet generator already exists within the device, it varies in features and properties between devices, and is therefore target specific.

Additional target specific infrastructure is focused on the connectivity of PTA: connecting the output of the test packet generator to the data plane under test, and connecting the output of the data plane under test to the output packet checker. This connectivity depends on the hardware architecture of the device, where PTA is internal to the device. For example, NetFPGA has a 256-bit wide AXI-4 streaming bus, and the output of the test packet generator is connected to NetFPGA's input arbiter. On other devices, the bus type and width will vary, as well as the connection points.

Finally, PTA includes 4 additional functions implemented in hardware that are useful for testing: random number generation, counters, time-stamping, and a method to swap fields. These are used as externs in the P4 programs.

D. Re-Configuration

To fully explore the parameter space for a test, PTA allows users to re-configure tests. To re-configure a test, a user writes a simple test script. The script allows users to change features such as the packet rate burst length, the gap length between packet transmissions, the packet sizes, the payload sizes, etc. It also allows users to set initial meta data flags and fields.

All configurations updates are control plane changes that happen dynamically at runtime. This allows users of PTA to explore a wide variety of test scenarios without having to recompile the test programs and install a new image—which can take a long time for some targets.

E. Interactions with the Control Plane

PTA monitors the data plane, and can identify and verify packets going to the control plane. It does not have direct visibility into the control plane. However, the functionality of a networked-program is a combination of the data plane program and the control plane configuration. During tests, PTA treats both as a single unit. A correctly programmed data plane with a misconfigured control plane may lead to a test failure, for example if a missing entry in a table leads to packet drop. In this context, PTA can indicate why a test has failed (e.g., a packet was dropped), but not what was the source of the failure (i.e., data or control plane). As we describe in Section IX-B (test#05), PTA successfully detects control plane related bugs. Another advantage of PTA's approach is that it further enables testing different control plane configurations.

F. Management and User-Interface

While P4 programs define the type of packets that can be generated by PTA, the management software defines the properties of the test. It is responsible for configuring the control and data planes, triggering tests, collecting and processing results, and declaring Pass/Fail. For this purpose, PTA includes a set of Python libraries to help manage the tests.

Test packet generation is determined by configuring the blank packet generator. This includes both information about the generated packet stream (e.g., number of packets, packet size, burst properties) as well as the input metadata accompanying the packet (e.g., path through the generation data plane, which headers to add).

Test results are specified using assertions. The management software reads from the output packet generator the results of the test (e.g. number of packets received), and compares them to the expected values set as Pass criteria.

G. Test Generation

The definition of validation tests is always a challenging task, which we address in Section X. In this section, we focus on the generation of pre-defined tests.

First, PTA reuses verification tests as validation tests through an automated integration with P4v, as discussed in Section VI. This means that tests previous written and executed for P4v are compiled and run on the hardware target using PTA.

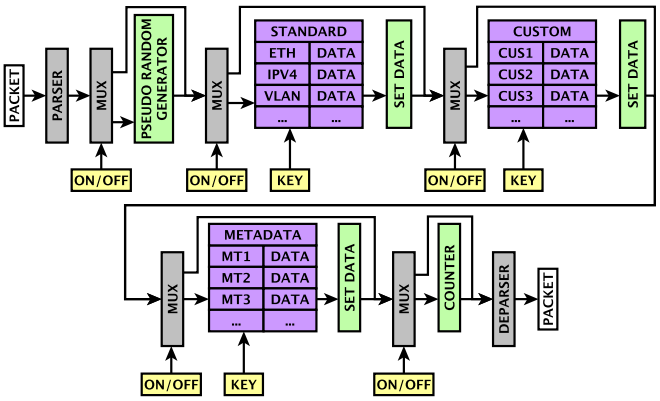

Figure 2: PTA Reference Design: test packet generator

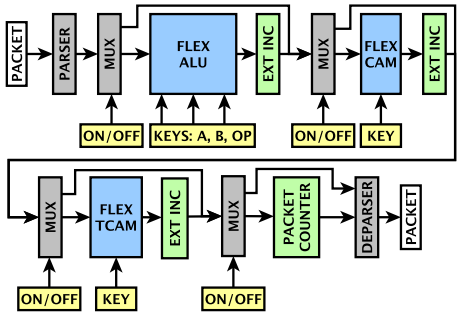

Figure 3: PTA Reference Design: output packet checker

Aspects that can not be tested using verification tools, such as P4v, must be written by the user. The user has a high degree of freedom in test generation. PTA includes a test library that allows users to specify test pre-conditions and post-conditions in a Python scripts. These are automatically translated to a configuration of PTA. Example test scripts are provided in [24]. A lot of these tests, though not all, can be defined once and then be reused as data plane programs change or for regression tests (as is the case with the NetFPGA).

PTA supports some types of fuzz testing, such as random payload or random header contents, and these were used in a few functional tests described in Section IX-B.

PTA can generate any possible test packet sequence, as long as all the headers of each generated test packet can fit in the available stages of the test packet generator pipeline.

Although PTA is an open-loop framework, "stateful" testing can be implemented using its Python test library. Each of the example test scripts provided in the project repository [24] could be extended to generate new sequences of test packes, based on the outcome of previous checks, and for repeating this procedure indefinitely.

The open-source nature of PTA means that users can also change the hardware infrastructure to support new or different tests that were not considered by the authors of this paper.

V. PTA Reference Design

As an initial design for PTA, we have developed the PTA Reference Design, which emphasizes rapid deployment. The PTA Reference Design includes both packet generator and output checker programs, and allows users to "program-once

This article has been accepted for publication in IEEE/ACM Transactions on Networking. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNET.2022.3214062

(data-plane), reconfigure-many-times (control-plane)" to run a large set of tests over different data planes.

The PTA Reference Design supports both standard headers (e.g., Ethernet, IPv4, etc.) and custom, user-defined headers, as well as user-controlled metadata. Users can run different types of tests (e.g., functional, performance) and tests generated from PTA's integration with P4V (§VI). In designing PTA Reference Design, we adopted a requirements-driven approach in which we examined prior work on programmable data planes [29], [15], [30], [31], [32] to identify both common and dedicated features, while attending to resource constraints.

Figures 2 and 3 present the concept of PTA Reference Design, for the test packet generator and the output packet checker, correspondingly. Each module is represented by a single P4 program, supporting a range of match-action tables. The entries within these tables are configurable and provide the flexibility to support different test programs.

a) Reference Test Packet Generator: The reference test packet generator contains match-action stages of three different types: "standard" stages populate widely-used standard headers; "custom" stages populate headers unique to the program under test; and "metadata" stages set fields with the pipeline's metadata bus. Although custom headers can sometimes fit within "standard" tables, they are often wide or require significant resources [31], [15]. Therefore, the test packet generator includes dedicated shallow and wide "custom" tables to optimize the use of resources. Our reference program is organized as three "standard" tables followed by "custom" tables, which is the common case in the programs we examined.

In PTA Reference Design, the metadata bus is used in a unique way, indicating to each stage if it is "on" or "off", and choosing which header is generated at each stage (based on values within the tables). The metadata bus settings are read from a table, thus enabling many different pipeline configurations using a single P4 program.

Note that there is a critical trade-off between flexibility and available device resources. The flexibility provided by the generator increases with the number of stages, meaning that with more resources, more headers and packet combinations can be generated. Although this trade-off constraints scalability, it keeps the design simple to implement and configure. The reconfigurable component of the metadata bus is relatively small, and does not consume many stages in the pipeline. Most of the metadata bus is reserved for both the blank packet generator and the data plane under test.

b) Reference Output Packet Checker: The packet checker is required to handle not only headers, but also fields and bits within headers. Moreover, the checker must handle operations (e.g, AND) computed on values from headers and fields to compare with an expected output. To provide maximum flexibility, we prefer to make each pipeline stage more complex, rather than increasing the number of pipeline stages. As in the packet generator, stages can be turned on or off, but pipeline stages look different. Figure 4 shows the design of the three types of pipeline stages: "ALU", "CAM" and "TCAM". In the figure, CAM and TCAM are depicted together, since there are minimal differences in the two designs.

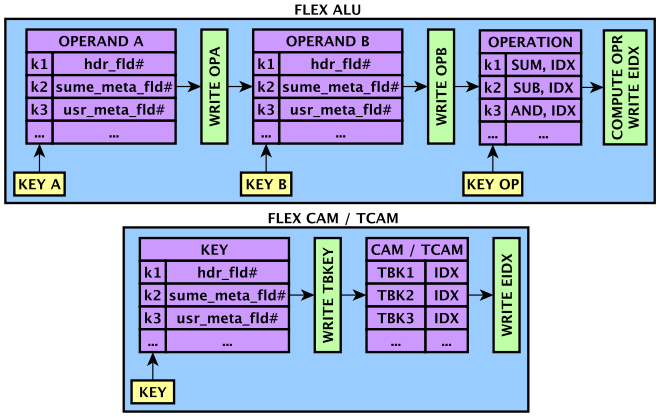

Figure 4: Components of the reference output packet checker

ALU stages, each made from three match-action stages, are used for preparing values to be checked. Each of the first two stages is used to load one operand, and the third stage executes the operation and stores the result to a register. Operands can be both header and metadata fields. Supported operations include addition, subtraction, AND, OR and XOR.

CAM and TCAM stages are used to implement fast binary and ternary matching, correspondingly. Each CAM/TCAM stage is composed of a pipeline of two match-action stages, one for selecting the header/metadata field and the second to match and write the result to a dedicated metadata variable. The second stage provides flexibility for the match operation, since P4 requires that the field to be matched is known at compile-time. Only the first stage is be configured by the metadata bus. The second stage is configured through the control plane before executing a test.

c) Example Configuration.: To illustrate the operation of PTA Reference Design, assume a user program processing packets with four headers: three standard headers (Ethernet, IPv4 and UDP), and one proprietary header that includes two fields. The program under test is supposed to XOR the first application header field with the metadata field indicating the network interface that received the packet and write the result to the second application header field. Assuming PTA Reference Design was already loaded on the platform, only the test needs to be configured.

In order to generate the test headers, we configure the generator to use three "standard" tables to generate the Ethernet, IPv4 and UDP headers, and one "custom" table to generate the proprietary header. In this scenario, two of the "custom" tables will not be used and will be turned off. A seventh table will be used to generate a metadata bus that feeds the data plane under test. The size of the blank packet generated is configured as well.

At the checker, we map the XOR operation to the first ALU (stage 1) and the related match to the first CAM (stage 2). The ALU at stage 3 is turned off. We use the CAM at stage 4 to check that packets are correctly forwarded to the output network interfaces. Using the control plane, we populate the tables in the three active stages with the addresses of the fields to match, the operation the ALU stage is supporting and the

index of the registers to which the results will be stored. In stage 1 we select the first application header field as the first operand and the metadata field indicating the network interface that received the packet as the second operand. We set the ALU to compute a XOR operation. In stage 2, we select the metadata field containing the result of the XOR operation to be matched against the table. Finally, we select the metadata field indicating the network interface to which the processed packet has been forwarded to be match against the table at stage 4. The results of the checks compute at stages 2 and 4 will be written to registers.

d) Implementation.: PTA Reference Design is currently supported on NetFPGA SUME. Because SDNet, used by P4-NetFPGA [30], limits the minimum size of table to 64 entries, and tables can not share memory, the reference program instantiates multiple identical tables, leading to sub-optimal use of resources. Consequently, the number of "custom" headers supported is limited. An upcoming release of P4-NetFPGA on Xilinx Alveo board is expected to solve these resource limitations.

VI. Integration with a Verifier

There has been significant prior work on workload and test case generation. This work covers a broad range of techniques, including automatic test packet generation [20], fuzz testing [21], and software validation [8]. PTA provides a path for these tools to run their generated tests on hardware. As a proof-of-concept about how such tools could integrate with PTA, we developed a prototype P4v-to-PTA translator.

Many software verification tools, including P4v [8] and Assert-P4 [9], are based on Hoare logic, which provides a formal system for reasoning about the correctness of computer programs. The central feature of Hoare logic is the Hoare triple. A Hoare Triple is of the form $\{P\}\ c\ \{Q\}$, where P is the precondition, Q is the postcondition, and c is the command, i.e., a piece of code that changes the state of the computation. A verifier can translate these assumptions and assertions into logical formulas, and use an automated theorem-prover to check if there is an initial state that leads to a violation.

The P4v-to-PTA translator parses an annotated P4 program and automatically extracts the assumptions (i.e., the Ps) and assertions (i.e., the Qs). For each assumption, the tool collects a list of test headers to be generated and identifies suitable values for each field. It then generates the P4 code and configuration for the test packet generator. For each assertion, the tool generates the output packet checker's data plane and the accompanying configuration, as illustrated in Figure 5.

This process is fully automated. To test an annotated P4 program, a user simply needs to place the P4 code in a specified folder. The PTA tool not only generates the test generator and checker P4 programs for PTA along with the associated configuration, but also compiles and runs the user's application on Tofino. Furthermore, the management scripts collect test results and present them to the user. The entire process is similar to the using the P4v command line tool, but runs real, extensive hardware validation. We expand this discussion in [24], by describing an example test program.

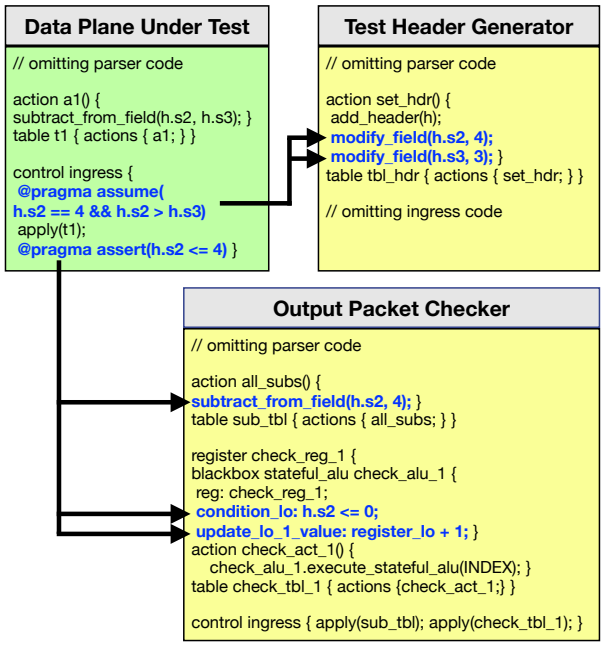

Figure 5: Integration with P4v.

Using PTA as a runtime tester, we were able to identify two bugs in the Barefoot Networks SDE compiler, related to saturating integers (Section IX-A). These bugs would not be caught by formal verification. The programs that suffered from these bugs were logically correct, but still exhibited unexpected behavior.

VII. Implementation

We have implemented PTA on two target devices: NetFPGA SUME platform and Barefoot Networks' Tofino ASIC. The code base, which includes both the implementations, is open source, and available on GitHub [24].

FPGA Implementation. An FPGA-based prototype is implemented on the NetFPGA SUME platform [3] using the P4→NetFPGA workflow [30]. It uses Xilinx's SDNet[33] 2018.2 and supports P4_16.

PTA builds upon the NetFPGA SUME reference architecture, which is composed of a data plane that processes traffic arriving from four independent network interfaces and a host (over PCIe). PTA taps to the architecture through a sixth input interface and a sixth output interface. Our implementation of PTA on NetFPGA follows the design outlined in Section IV.

ASIC Implementation. We also implemented PTA on Barefoot Networks' Tofino, a 6.5Tbps programmable Ethernet switch ASIC [34]. The Tofino ASIC provides either two or four hardware packet processing pipelines, depending on the ASIC model. The four pipelines of Tofino allow us to implement PTA using an architecture similar to the FPGA architecture: the packet generator and checker are implemented within separate pipelines, and the data plane under test is loaded in a dedicated pipeline. We note that the current Barefoot Networks SDE control plane software does not allow users to manage different programs loaded on to different data planes. This limitation is not inherent, and we expect the functionality

will be supported in future releases. However, as a temporary work-around, we performed our debug experiments using three switches instead of one. As P4v supports only P4_14, our implementation of P4v-to-PTA supports only Tofino using P4_14, extensively by Barefoot at time of development, and not the P4_16-based P4-NetFPGA.

VIII. Evaluation

Validation. We validate PTA independently from any data plane under test, both on Tofino and on NetFPGA. The validation uses external traffic generation (e.g., OSNT [10]) and capture tools (e.g., Endace DAG) to confirm assumptions such as traffic rate and contents. Barefoot further confirmed to us that the packet generator built inside the Tofino chip runs at line-rate. We conduct a functional validation of PTA, testing using both external and internal tools (counters, logic analyzer) to examine each feature. Testing of programmable data planes began only once the PTA infrastructure was tested.

Performance. We have evaluated and confirmed that both NetFPGA SUME and Tofino-based programs run at line rate, using the setup previously described and ranging packet sizes from 64B to 1514B on NetFPGA. PTA implementation on NetFPGA does not allow for congestion propagation into PTA's pipelines, meaning that any flow control indication leads to packet drop outside the modules. For both NetFPGA and Tofino, support for congestion control within the pipeline is the same as for any other programmable dataplane [33]. We showed these properties of PTA in Section IX-D, by testing networked programs for line-rate and identifying bugs.

Resource Consumption. PTA introduces two new modules to a device. On the generator side, NetFPGA programs required between 2 and 4 pipeline stages, using one table and 1-2 externs, and Tofino implementations required 2 tables. On the checker size, NetFPGA programs required between 5 and 7 pipeline stages, using two tables and 3-5 externs. On Tofino, 7 tables and 5 stateful ALUs were required in the checker. A breakdown of these results is provided in [24].

We report the resources overhead introduced by PTA, but caution that it is difficult to quantify resources in a meaningful way, since the amount used depends on the program under test and the compiler. For example, on NetFPGA the compiler requires that all tables have least 64 entries, even if 16 entries would be sufficient. A newer version of SDNet (2019.1), not currently supported by NetFPGA, is more resource efficient.

On NetFPGA, representing an FPGA-based use case, the resource overhead of PTA (i.e, average of logic and memory use) never exceeded 15%, which was for the experiments with NDP [35], compiled with SDNet 2018.2. In many cases (e.g., INT, Learning Switch) this number drops to 9%. The blank packet generator required just 0.13% logic overhead, and no memory. Detailed resource consumption is provided in [24].

On Tofino, PTA tested a data-plane program on one pipeline using other pipelines. Since resources are not shared between pipelines, PTA does not "take away" resources from the data plane under test. ASIC resources are given, and PTA easily fits, using the resources noted above.

Metric	Device Property	Example
Max # of headers in a single test	PHV size	4Kb
Max # of checks in a single test	# of Stateful ALUs	40
Max # of packets in a single test	Counter width	4 billion
Max test speed	Pipeline Bandwidth	1.6 Tbps
Max packet size	Max Transmission Unit	1514 B

Table III: Additional Evaluation Metrics. Example values are indicative of the proof-of-concept implementations.

Test Completion Time. The PTA run time includes four components: platform setup (i.e., downloading an FPGA bit file), configuration, test execution time, and results collection and report. The test execution time is test-dependent, i.e., it depends on how long a user wants to send traffic, the number of parameters to explore, etc. For the tests that we ran on NetFPGA, the average overall time was $\sim$110s, including all four components, though for some tests this number was reduced to $\sim$70s. Out of that, the platform setup time, which is a one time process, is $\sim$20s, and test re-configuration, including populating tables, is in the order of seconds. An exhaustive performance test on NetFPGA SUME, which tests throughput under each and every supported packet size, with a billion packets per packet size, was $\sim$3000s.

Additional Metrics. Many of PTA's performance metrics, summarized in Table III, are a property of the hardware target, not PTA. For example, the number of headers depends on the size of the packet header vector (PHV), and the number of verified aspects in a test (e.g., dropped packets, correct headers checks) depends on the number of stateful ALUs in the device.

IX. Bugs Found

Our implementations of PTA enabled us to uncover bugs within different programs and architectures, while covering use cases discussed in Section II. Table IV provides a partial list of tests run and bugs found using PTA. The table indicates the name of the program we used for the data plane under test, a brief description of the category of bug, the hardware platform, and whether or not the test passed. Note that when the program name is *Any*, it indicates that the bug was not tied to a particular program. We discuss these particular bugs as they highlight the diversity of test cases that PTA enables.

A. Compiler Checks

PTA was able to find or confirm three bugs in version 8.9.1 of the Barefoot SDE compiler. These bugs were found when integrating with P4v, discussed in Section VI. In the first bug, (test #01), the compiler generated incorrect byte swapping code (e.g., between big and little endian). This bug has been fixed in the 9.0.0 release of the SDE. The second bug is related to "saturating" an attribute in header fields (test #02); header fields marked as "saturating" always collapse to their minimum value after a "subtract" operation is computed on them. The third bug prevents the implemented data plane from correctly processing "signed" header fields (test #03). Although represented in two's complement form, "signed" fields are treated as unsigned numbers in the hardware, thus

```
// Parse packet headers by specifying state
// machine transitions.
parser Parser(packet_in b,
           out Parsed_packet p,
           inout sume_metadata_t sume_metadata) {

   state start {
      b.extract(p.ethernet);
      transition select(p.ethernet.etherType) {
         IPV4_TYPE: parse_ipv4;
         default: reject;
      }
   } // Eliding IPv4 parser
}
```

Figure 6: Subset of a P4 program that reject non-IPv4 packets. The behavior of the bold line is unspecified.

generating incorrect results. We reported these bugs to the developer, and they have since been fixed.

B. Functional Tests

We discovered several functional bugs in multiple designs implemented on NetFPGA SUME, including the Verilog and P4 Learning Switch designs, and NDP [35]. First, packets with invalid source MAC addresses pass through the data plane and reach the output network interfaces, even though they should have been dropped before traversing the pipeline (test #04). This bug differs from the Parse Reject bug (test #08), as the value within the header should be banned, not the header itself. Furthermore, the issue is Ethernet compatibility, not compatibility with the P4 specification. Second, we find that, when the number of entries written to the MAC lookup table exceeds the size of the table, the write pointer will wrap around, and the first entry will be over-written (test #05).

When testing a P4 implementation of In-band Network Telemetry (INT) [32] on NetFPGA, we detect some missing functionality (test #06). This includes missing measurement of switch hop latency, egress port utilization, or queue congestion status. The design also does not report which rules matched while traversing the data plane or provides information about other flows traversing the same network queues.

A more serious bug in the implementation of INT is the handling of packets with a large instructions count (test #07). The INT specification [32] states that "a device would cease processing an INT packet with an Instruction Count higher than the number of instructions that it is able to support". In our test, we find that if more than five instructions are requested, the program fails to set the Bottom-of Stack (BOS) flag to the last (fifth) INT header.

C. Under-Specification Tests

Because different hardware targets have different capabilities and features, they may exhibit different behavior. And, in some cases, it would be unreasonable to force all targets to have a uniform behavior, because doing so would add unnecessary complexity to a design, or add additional performance overhead. For such situations, the language specification often leaves the implementation details as up to the compiler.

One example of such behavior is illustrated in the snippet of code in Figure 6, which shows the implementation of a parser in P4 (test #08). It includes logic to extract the Ethernet header and examine the type field of the Ethernet header. If the type field indicates IPv4, the parser will transition to the parser state for extracting IPv4 headers. Otherwise, the program will drop the packet (i.e., `reject`).

The intention of this program is that any non-IPv4 packets should be dropped. However, the behavior of the program when compiled using P4→NetFPGA [30] might run counter to user expectations—the packet is forwarded through the programmable pipeline and out of the device.

The reason is because the P4 language specification leaves the choice of how to implement a parser reject state up to the architecture. The SDNet compiler [33] does not implement the reject state as drop and P4→NetFPGA [30] does not use the reject indicator provided by the SDNet-generated module.

Technically, forwarding the rejected packet is not a bug, since the implementation is not contrary to the specification. However, it does result in unintuitive behavior that might surprise a developer. And, this behavior would not necessarily be caught by verification tools like P4v [8] or Vera [7], depending on how they model `reject`.

D. Performance Tests

We evaluate the performance of several P4 and HDL based programmable designs built upon the NetFPGA infrastructure. We first discuss bugs that are specific to a given program. We then discuss bugs that are a property of the NetFPGA infrastructure.

In the NetFPGA Reference Switch and NDP, we discover a write-after-write hazard, where the lookup table is not able to sustain subsequent entry updates with packet sizes of less than 385B, due to the write access latency (test #09). When the packet size is 385B or bigger, meaning 13 clock cycles or more between two updates, the design functions as expected.

Running a similar test on the P4-based learning switch resulted in a failure to support consecutive table updates at line rate, regardless of packet size. The root cause to this limitation is the separation of control and data planes, which means that updates to the lookup table must go through the host by design, a latency in the order of milliseconds.

An evaluation of the P4-based INT design on NetFPGA yielded interesting performance results (test #10). We find that the data plane can sustain the full internal throughput (50Gbps) only with unaligned packet sizes (e.g., 65B, 97B), but not for data path aligned packet sizes (e.g., 64B, 96B). We expect that this issue is caused by the expansion of the packet within the encrypted data plane module, beyond the INT header added to the packet. This is also the explanation proposed to us by the P4→NetFPGA designers.

E. Architecture Tests

A few of the bugs uncovered by PTA had to do with the architecture of specific designs or with the underlying hardware infrastructure (test #11). For example, initial throughput testing of both HDL-based and P4-based learning switches

Test#	Program	Category	Description	HW Plat.	Pass/Fail
01	*Any*	Compiler	Byte swapping	Tofino	Fail
02	*Any*	Compiler	Saturating	Tofino	Fail
03	*Any*	Compiler	Signed fields	Tofino	Fail
04	NDP, Switch (P4, Verilog)	Functional	Invalid MAC	NetFPGA	Fail
05	NDP, Switch (P4, Verilog)	Functional	Table wrap	NetFPGA	Fail
06	INT	Functional	INT features	NetFPGA	Fail
07	INT	Functional	Instructions count	NetFPGA	Fail
08	INT	Underspecification	Parser reject	NetFPGA	Fail
09	NDP, Switch (P4, Verilog)	Performance	Write hazard	NetFPGA	Fail
10	INT	Performance	Aligned packet sizes	NetFPGA	Fail
11	*Any*	Architecture	Input arbiter	NetFPGA	Fail
12	*Any*	Security	Meltdown	NetFPGA/Tofino	Pass/Pass
13	*Any*	Security	Read headers beyond	NetFPGA/Tofino	Fail/Pass
14	Switch (P4)	Comparison	Port MAC	NetFPGA	Fail
15	Switch (Verilog)	Comparison	Table wrap	NetFPGA	Fail

Table IV: A subset of the tests we ran and bugs found using PTA.

resulted in a large number of packet drops. The cause was found in the arbiter at the input to the data plane, that turned out not to be work conserving. An additional architecture limitation was discovered at the output of the data plane, at the output queues. Full rate traffic through the data plane under test led to packet drops at the queues, which turned out to be an intentional design choice by the NetFPGA team. They designed the overall supported outputs queues throughput to be circa 40Gbps. We note that PTA found this bug after the platform had already been in use for more than 10 years. The fix has supported 2 more recent NetFPGA-based projects.

F. Security Checks

In the course of working on this paper, we have conducted several experiments, both on P4→NetFPGA and on Tofino, trying to uncover security vulnerabilities. In our exploration, we focused on one aspect of the P4 language, which is *Undefined Behaviors* (Section G.2 of P4 Specification v1.1.0 [36]). This includes aspects such as uninitialized variables, accessing header fields of invalid headers, and accessing header stacks with an out of bounds index.

First, we tried to identify the networking-equivalent of a "Meltdown" [27] bug by attempting to infer the contents of previous packets using malformed packets (test #12). In principle, we try to infer the contents of the memory by reading a value of a non-existing header in the packet, in an attempt to use previously stored header contents. This is one form of accessing header fields of invalid headers. Positively, we find that the SDNet compiler returns a zero value for such attempts, providing stateless operation between packets.

In another test, we attempted to read headers beyond the end of the packet, with a similar motivation (test #13). In this case the result was positive as well, with the Tofino switch dropping the "aggressor" packet. SDNet does not allow such operations either, invalidating all parsed bits of the offending packet. This case is interesting, as it touches on the delicate interface between compiler and architecture. Although SDNet guards against such operations, P4→NetFPGA did not handle the error indication from SDNet. Therefore, the unprocessed and partially corrupted packet may still be emitted.

G. Comparing Designs

By comparing seemingly identical designs, we do not identify bugs, as these are covered by previous scenarios. However, we do identify gaps in specification, behavior, or performance. For example, we compare two implementations of a learning switch: one written in Verilog, and one written in P4. Both designs share the same NetFPGA infrastructure, and differ only in the data plane module. Despite the similarity, we find two differences in functionality. First (test #14), in the P4-based design, two ports cannot be assigned to the same MAC; once a Port-MAC binding has been learned, it cannot be overwritten by other packets. In contrast, in the Verilog-based design, after a Port-MAC binding has been learned, it can be overwritten by other packets. Second (test #15), in the Verilog-based design, overflow happens when exceeding the table size and the first entry is overwritten without any notice (as noted before). In the P4-based design, on the other hand, no overflow happens when exceeding the table size. In principle, such updates are expected to be silently dropped by the control plane. This is a property of the closed-source compiler, which we don't have visibility to test.

H. Ethics and Corrective Actions

Ethical issues have been considered as part of this research. We have focused on the handling of vulnerabilities and weaknesses discovered in the different designs. Vulnerabilities have been disclosed and discussed with code and platform originators, which also helped us clarify what is considered a bug, a known design limitation, or an unsupported feature. We have further taken a positive approach and contributed code fixes to open-source projects (e.g., NetFPGA), as a means to improve their quality based on our findings.

The reject limitation found in P4→NetFPGA was reported to the NetFPGA project and Xilinx Labs. Xilinx Labs have proposed a work-around that enables users to support functionality similar to reject in the pipeline, even though actions as a result of reject is not implemented in the compiler.

We discussed the architecture and performance issues in the NetFPGA Reference designs with the NetFPGA team. The NetFPGA team indicated to us that they were aware of a

minimum-access latency limitation for table updates, but not to the discovered extent. The authors of this paper have also contributed a fix to the NetFPGA input arbiter module as part of this work, as well as the packet generation module of PTA.

Bugs in the Barefoot Networks SDE were reported by entering a ticket on the FASTER portal and via personal email communication. All bugs reported in this paper have since been fixed by the developers.

X. Lessons Learned

In this section, we summarize some lessons learned through our experiences with testing network data planes.

A. Extending the P4 Language

PTA is designed to provide a programmable test framework with an internal view of a network device. However, this internal view is hampered by the closed-source nature of hardware solutions, e.g., modules generated by SDNet are encrypted. Testing would be improved if users could set hooks within the code or access the state or values of certain language constructs. A hook "breaks" the data plane structure, since it allows users to inspect status at a certain point within the design. Such extensions to the P4 language would allow testing in case of a failure, or when the pipeline is stuck.

Supporting watch-points and stepping through code are required future contributions in the field of programmable network devices. Some of the bugs introduced in this work, such as the performance limitation identified in the INT design, can not be easily tracked and tested today even by compiler vendors, with full access to the code.

Such extensions to the language would benefit all, as the debuggability of network devices in production environments is one of the utmost concerns of service providers.

B. Under-Specification as a Source of Bugs

Because enforcing a uniform behavior on all hardware targets is impractical, some functionality is compiler specific (e.g., uninitialized values). Under-specification in the language may lead to bugs (Section IX) or security vulnerabilities [16].

Another form of under-specification, i.e., in the interface and division of responsibilities between the data plane and the rest of the device, can also cause errors, as shown above (test #08). Integration bugs are not uncommon in hardware design, but the problem is exacerbated where different technologies interface. This ranges from the integration of a programmable pipeline within an otherwise fixed-function switch, as well as with the integration of externs within P4 programs.

Attending to under-specification requires a closer integration of end-to-end system components. While such an integrated design is likely to reduce bugs, the disadvantage is that portability may be restricted and design time may increase.

C. Writing a Test Suite

One of the challenges in testing a network device is creating a comprehensive corpus of tests. Considering the bugs detected by PTA, we identify three classes of tests.

The first class of tests is the "validation" list of tests. This includes tests that were run during the design stage, and need to be validated on a newly produced target, e.g., tests written for P4v and later translated by PTA. It also includes traditional network tests, e.g., checking throughput under different scenarios (test #10).

A second class of tests is tests generated in a response to a bug discovered by a user, e.g., the byte swapping bug (test #01). The goal of such a test would be to (i) validate that the bug is fixed in a newer version of the program. (ii) be used as part of regression tests in future releases (iii) test deployment of bug fixes in the field. The last use case is a good example of the usefulness of PTA, as it enables testing devices in the field without physically connecting them to test equipment.

The third class of tests targets known potential points of failure that are typically hard to test. An example is testing saturation (test #02), which one would typically verify in a block-level simulation, but would be hard to trigger as part of traditional hardware validation without additional built in self text (BIST) resources. PTA enables users to craft such tests without per-test resource overhead and while specifically targeting sensitive elements in the design.

As tests are target independent in PTA, we envision building such an open-source corpus of tests for community benefit, starting with the tests included in the PTA repository [24].

D. Coverage and Test Case Generation

One of the advantages of network tests is that they can be run at line rate. On ASIC, that means exhaustive testing is feasible, since billions of packets with billions of header values can be tested every second. On the NetFPGA SUME platform, the supported packet rate is about sixty million packets per second. While these numbers are high, they are insufficient to fully cover all potential cases. For example, to test all combinations of 48bit source Ethernet MAC header, it would take about eight hours on a switch capable of processing ten billion packets per second. Testing the combination of both source and destination MAC header would take $\times 2^{48}$ longer. As switches often need to drop packets where the source and destination address are the same, or some forbidden MAC addresses, this is not an imaginary scenario. Note that with PTA, users can write such exhaustive tests, or write tests using random field values (e.g., source and destination MAC address), as well as specific scenarios (identical source and destination MAC addresses). The advantage of PTA is that there is no need to write new P4 code for these different scenarios, just re-configure the test via register access.

E. Line-rate testing

Besides reducing the test exection time and enhancing coverage through exhaustive testing, line-rate testing enabled two categories of tests in PTA: architecture tests and performance tests. Some architecture bugs, such as test #11, become visible only when the traffic exceeds a certain rate. Otherwise, it would be impossible to discover them. Performance bugs, including test #09 and test #10, require line-rate testing for pushing the performance of the data plane under test to its maximum.

XI. Limitations of PTA

Implementation. Our two implementations present different facets of PTA. The implementation on NetFPGA uses P4_16, demonstrates full integration with a device's data plane, enables in-field testing (e.g., for smartNIC applications), but lacks the performance and realism of commercial ASICs. The implementation on Tofino demonstrates the feasibility of using PTA to detect bugs on commercial ASICs, and supports full line rate, but does not support in-device integration (as the ASIC's architecture is fixed), nor in-field testing. As we show in Section VIII, it does allow detecting data plane bugs.

Test Generation. With PTA, users must still define tests themselves. PTA also helps map logical tests to physical hardware tests, as in the case of P4v-to-PTA. The problem of defining tests is a long standing research problem [21], [22], [8] which is beyond the scope of this work.

Bugs in PTA. Despite best efforts, there is no guarantee that PTA is bug free. To reduce the likelihood of bugs, we separated PTA's infrastructure from users' data plane and architecture, and validated it independently using external tools (Section VIII). We note that a user program's bug will not affect PTA's operation. Similarly, a bug in PTA will not affect a user program. In this case, PTA's result may be incorrect.

Portability. Porting PTA between targets requires implementing the abstractions in Table II. Different targets require changes to the hardware infrastructure, and validating again PTA's infrastructure (e.g., traffic generation performance and correctness). It may also require control plane changes, as interfaces differ between targets. However, this work demonstrates the portability of PTA by presenting two prototypes, targeting two different programmable network devices: an FPGA-based card and and ASIC-based switch.

XII. Related Work

PTA is related to programmable network programming languages, network testing, and data plane verification.

Network Programming Languages. Developers specify the packet-processing behavior of re-configurable ASICs using a variety of domain specific programming languages. Examples include Huawei's POF [37], Xilinx Labs' PX [38], Broadcom's NPL [39], and the P4 Consortium's P4 [17]. PTA is implemented in P4, but does not inherently depend on P4. It could be ported to any of the above languages.

Network Testing. There has been significant prior work in both industry and academia on testing fixed-function switches. The most similar to PTA is the service activation test (SAT) [40]. SAT, used by carrier Ethernet service providers, is intended to ensure that network services are configured as specified and meet the predefined Service Acceptance Criteria (SAC). SAT uses packet injection as a means to test the service, but it is closely defined and not programmable, covering only a limited set of the aspects enabled by PTA. More low-level approaches, such as ATPG [41], focus on manufacturing faults rather than bugs. FPGA debug tools, such as ILA [42], enable limited functionality testing, far less than the tests described in this paper, and are timing-affecting.

Network Testing Using P4. In-band network telemetry (INT) [32] and postcards [43] use programmable network hardware for monitoring and network-level diagnostics. In contrast, PTA focuses on detecting bugs on a device, in a compiler, or in the logic of a data plane program. Moreover, PTA is active, rather than passive, meaning that it will generate specific packets to facilitate ad-hoc and exploratory analysis.

Data Plane Verification. Several recent projects explore P4 program verification, including Vera [7], P4v [8], and Assert-P4 [9]. The details of the approaches differ, but essentially, they all translate P4 programs and some control plane state into a logical formula, and use techniques such as symbolic execution [44] to check that correctness properties are not violated (e.g., a header field in a packet is not accessed if it has not been parsed). PTA complements these efforts by providing runtime testing. PTA provides grey box testing, as often only partial information exists on the data plane under test. Grey box testing has been further motivated by prior works on router and network level, such as NetSonar [45].

XIII. Conclusion

We have presented PTA, a portable test architecture for testing data planes. PTA leverages both the P4 language and hardware design to provide flexibility and visibility into programmable network devices. We have built a prototype of PTA, and used it to detect numerous hard-to-find bugs. PTA addresses an urgent need for improved tools and techniques for data plane testing and verification.

Acknowledgments. We thank the NetFPGA core development team who helped us develop and debug PTA. We thank the anonymous shepherd and reviewers, who helped us improve this paper. We acknowledge the support from the Swiss National Science Foundation (SNF) (project 407540_167173).

References

[1] "Monitoring and Troubleshooting: One Engineer's rant," https://archive.nanog.org/meetings/nanog53/presentations/Monday/Hoose.pdf, 2011.

[2] J. W. Lockwood, N. McKeown, G. Watson *et al.*, "Netfpga–an open platform for gigabit-rate network switching and routing," in *MSE*. IEEE, 2007, pp. 160–161.

[3] N. Zilberman, Y. Audzevich, G. A. Covington, and A. W. Moore, "Netfpga sume: Toward 100 gbps as research commodity," *IEEE Micro*, September 2014.

[4] K. Camera, H. K.-H. So, and R. W. Brodersen, "An integrated debugging environment for reprogrammble hardware systems," in *Proceedings of the Sixth International Symposium on Automated Analysis-driven Debugging*, ser. AADEBUG'05, 2005, pp. 111–116. [Online]. Available: http://doi.acm.org/10.1145/1085130.1085145

[5] Ixia, "Ixnetwork," https://www.ixiacom.com/products/ixnetwork, 2019.

[6] "Spirent," https://www.spirent.com/, 2020.

[7] R. Stoenescu, D. Dumitrescu, M. Popovici *et al.*, "Debugging p4 programs with vera," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18, 2018, pp. 518–532. [Online]. Available: http://doi.acm.org/10.1145/3230543.3230548

[8] J. Liu, W. Hallahan, C. Schlesinger *et al.*, "P4v: Practical verification for programmable data planes," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18, 2018, pp. 490–503. [Online]. Available: http://doi.acm.org/10.1145/3230543.3230582

[9] L. Freire, M. Neves, L. Leal *et al.*, "Uncovering bugs in p4 programs with assertion-based verification," in *Proceedings of the Symposium on SDN Research*, ser. SOSR '18, 2018, pp. 4:1–4:7. [Online]. Available: http://doi.acm.org/10.1145/3185467.3185499

[10] G. Antichi, M. Shahbaz, Y. Geng *et al.*, "Osnt: Open source network tester," *IEEE Network Magazine*, vol. 28, no. 5, pp. 6–12, 2014.

[11] Y. Zhou, Z. Xi, D. Zhang *et al.*, "Hypertester: high-performance network testing driven by programmable switches," in *ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, 2019, pp. 30–43.

[12] R. Miao, H. Zeng, C. Kim *et al.*, "Silkroad: Making stateful layer-4 load balancing fast and cheap using switching asics," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, ser. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM), 2017, pp. 15–28. [Online]. Available: http://doi.acm.org/10.1145/3098822.3098824

[13] X. Jin, X. Li, H. Zhang *et al.*, "NetCache: Balancing Key-Value Stores with Fast In-Network Caching," in *ACM Symposium on Operating Systems Principles (SOSP)*, Oct. 2017.

[14] ——, "NetChain: Scale-Free Sub-RTT Coordination." in *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Apr. 2018.

[15] H. T. Dang, P. Bressana, H. Wang *et al.*, "P4xos: Consensus as a network service," *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1726–1738, 2020.

[16] M. Dumitru, D. Dumitrescu, and C. Raiciu, "Can we exploit buggy p4 programs?" in *ACM SIGCOMM Symposium on SDN Research*, Jun. 2020.

[17] P. Bosshart, D. Daly, G. Gibb *et al.*, "P4: Programming Protocol-Independent Packet Processors," *SIGCOMM Computer Communication Review (CCR)*, vol. 44, no. 3, pp. 87–95, Jul. 2014.

[18] "P4$_{16}$ Portable Switch Architecture (PSA)," https://p4.org/p4-spec/docs/PSA.html, 2019.

[19] G. Brebner, "Extending the range of p4 programmability," 2018, keynote.

[20] A. Nötzli, J. Khan, A. Fingerhut *et al.*, "p4pktgen: Automated test case generation for P4 programs," in *Proceedings of the Symposium on SDN Research, SOSR 2018, Los Angeles, CA, USA, March 28-29, 2018*, 2018, pp. 5:1–5:7. [Online]. Available: https://doi.org/10.1145/3185467.3185497

[21] A.-A. Agape and M. C. Danceanu, "P4fuzz: A compiler fuzzer for securing p4 programmable dataplanes," Aalborg University, Tech. Rep., Sep. 2018. [Online]. Available: https://projekter.aau.dk/projekter/files/281195651/Master_Thesis_Project___P4_Fuzzer.pdf

[22] B. Shastry, M. Leutner, T. Fiebig *et al.*, "Static program analysis as a fuzzing aid," in *RAID'17*, 2017, pp. 26–47.

[23] K. Thimmaraju, B. Shastry, T. Fiebig *et al.*, "Taking control of sdn-based cloud systems via the data plane," in *ACM SIGCOMM Symposium on SDN Research*, 2018, pp. 1–15.

[24] "Pta, blinded reporsitory," https://github.com/pta-project-repo/pta-artifacts, 2020.

[25] P. Bressana, N. Zilberman, and R. Soulé, "Finding hard-to-find data plane bugs with a PTA," in *CoNEXT '20: The 16th International Conference on emerging Networking EXperiments and Technologies, Barcelona, Spain, December, 2020*, D. Han and A. Feldmann, Eds. ACM, 2020, pp. 218–231. [Online]. Available: https://doi.org/10.1145/3386367.3431313

[26] "Undefined behaviors - P4$_{16}$ Language Specification," https://p4.org/p4-spec/docs/P4-16-v1.0.0-spec.html#sec-undefined-behaviors, 2017.

[27] "Meltdown and Spectre," https://meltdownattack.com, 2020.

[28] Y. Tokusashi, H. T. Dang, F. Pedone *et al.*, "The case for in-network computing on demand," in *Proceedings of the Fourteenth EuroSys Conference 2019*, ser. EuroSys '19, 2019, pp. 21:1–21:16. [Online]. Available: http://doi.acm.org/10.1145/3302424.3303979

[29] V. Olteanu, A. Agache, A. Voinescu, and C. Raiciu, "Stateless datacenter load-balancing with beamer," in *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2018, pp. 125–139.

[30] S. Ibanez, G. Brebner, N. McKeown, and N. Zilberman, "The p4-netfpga workflow for line-rate packet processing," in *FPGA*. ACM, 2019, pp. 1–9.

[31] T. Jepsen, L. P. de Sousa, H. T. Dang *et al.*, "Gotthard: Network support for transaction processing," in *ACM SIGCOMM Symposium on SDN Research*. ACM, 2017, pp. 185–186.

[32] "Inband network telemetry (int)," https://github.com/p4lang/p4factory/tree/master/apps/int, 2017.

[33] Xilinx, "SDNet," http://www.xilinx.com/products/design-tools/software-zone/sdnet.html, 2014.

[34] Barefoot-Networks, "Barefoot tofino," Jun. 2018. [Online]. Available: https://barefootnetworks.com/products/brief-tofino/

[35] M. Handley, C. Raiciu, A. Agache *et al.*, "Re-architecting datacenter networks and stacks for low latency and high performance," in *SIGCOMM*. ACM, 2017, pp. 29–42.

[36] "P4$_{16}$ Language Specification Version 1.1.0," https://p4.org/p4-spec/docs/P4-16-v1.0.0-spec.html, Nov. 2018.

[37] H. Song, "Protocol-oblivious Forwarding: Unleash the Power of SDN Through a Future-proof Forwarding Plane," in *Workshop on Hot Topics in Software Defined Networking*, Aug. 2013, pp. 127–132.

[38] G. Brebner and W. Jiang, "High-speed packet processing using reconfigurable computing," *IEEE Micro*, vol. 34, no. 1, pp. 8–18, 2014.

[39] "Introduction to nplspec," 2019, https://github.com/nplang/NPL-Spec.

[40] M. E. Forum, "Carrier ethernet service activation testing (sat), technical specification mef48," https://www.mef.net/Assets/Technical_Specifications/PDF/MEF_48.pdf, 2014.

[41] P. Duhamel and J. Rault, "Automatic test generation techniques for analog circuits and systems: A review," *IEEE transactions on Circuits and Systems*, vol. 26, no. 7, pp. 411–440, 1979.

[42] C. Stroud, E. Lee, S. Konala, and M. Abramovici, "Using ila testing for bist in fpgas," in *Test Conference*. IEEE, 1996, pp. 68–75.

[43] N. Handigol, B. Heller, V. Jeyakumar *et al.*, "I know what your packet did last hop: Using packet histories to troubleshoot networks," in *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2014, pp. 71–85.

[44] C. Cadar, D. Dunbar, and D. Engler, "Klee: Unassisted and automatic generation of high-coverage tests for complex systems programs," in *Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation*, ser. OSDI'08, 2008, pp. 209–224. [Online]. Available: http://dl.acm.org/citation.cfm?id=1855741.1855756

[45] H. Zeng, R. Mahajan, N. McKeown *et al.*, "Measuring and troubleshooting large operational multipath networks with gray box testing," *Mountain Safety Res., MSR-TR-2015-55*, 2015.

Pietro Bressana is a Cloud Software Architect at Intel Corporation. He holds a bachelor's degree in Electronic Engineering and a master's degree in Computer Engineering, both from Politecnico di Milano (Italy). He received his Ph.D. in computer science from the Università della Svizzera italiana (Lugano, Switzerland). As a Ph.D. student, he visited the Networks and Operating Systems Group of the University of Cambridge (UK).

Noa Zilberman is an Associate Professor at the University of Oxford. Prior to joining Oxford, she was a Fellow and an Affiliated Lecturer at the University of Cambridge. Her research interests include computing infrastructure, programmable hardware and networking. She holds a PhD Degree in Electrical Engineering from Tel Aviv University, and is a Senior Member of IEEE.

Robert Soulé is an Assistant Professor at Yale University and a Research Scientist at Barefoot Networks, an Intel Company. Prior to joining Yale, he was an Associate Professor at the Università della Svizzera italiana in Lugano, Switzerland. He received his B.A. from Brown University, and his Ph.D. from NYU. After his Ph.D., he was a post-doc at Cornell University.