ProXray: Protocol Model Learning and Guided Firmware Analysis

Farhaan Fowze, Dave Tian, Grant Hernandez, Kevin Butler, and Tuba Yavuz

Abstract—The number of Internet of Things (IoT) devices has reached 7 billion globally in early 2018 and are nearly ubiquitous in daily life. Knowing whether or not these devices are safe and secure to use is becoming critical. IoT devices usually implement communication protocols such as USB and Bluetooth within firmware to allow a wide range of functionality. Thus analyzing firmware using domain knowledge from these protocols is vital to understand device behavior, detect implementation bugs, and identify malicious components. Unfortunately, due to the complexity of these protocols, there is usually no formal specification available that can help automate the firmware analysis; as a result significant manual effort is currently required to study these protocols and to reverse engineer the device firmware. In this paper, we propose a new firmware analysis methodology using symbolic execution called ProXray, which can learn a protocol model from known firmware, and apply the model to recognize the protocol relevant fields and detect functionality within unknown firmware automatically. After the training phase, ProXray can fully automate the firmware analysis process while supporting user’s queries in the form of protocol relevant constraints. We have applied ProXray to the USB and the Bluetooth protocols by learning protocol constraint models from firmware that implement these protocols. We are then able to map protocol fields and identify USB functionality automatically within all 6 unknown USB firmware while achieving more than an order of magnitude speedup in reaching protocol relevant targets in unknown Bluetooth firmware. Our model achieved high coverage of the USB and Bluetooth specifications for several important protocol fields. ProXray provides a new method to apply domain knowledge to firmware analysis automatically.

Index Terms—Protocol Learning, Model Extraction, Firmware, Symbolic Execution.

1 INTRODUCTION AND MOTIVATION

The number of Internet of Things (IoT) devices has reached 7 billion globally in early 2018 [1]. These devices have penetrated into almost every aspect of society, such as medicine and industrial control systems. Due to their interactions with the physical world, e.g., via sensors and actuators, and low-cost microcontroller architectures, IoT devices have a wide attack surface, which can be exploited to cause significant damage as in the case of the Mirai botnet [2]. Analyzing the firmware of IoT devices for safety and security is becoming critical.

IoT devices usually implement communication protocols such as USB and Bluetooth within firmware to allow a variety of functionality. Different vendors often have their own protocol stack implementations based on their interpretation of a protocol specification. Unfortunately, there is usually no formal specification to check whether the firmware implements the protocol(s) correctly. Transforming an informal protocol specification into a formal representation is often infeasible due to the complexity of these protocols. For instance, the core Bluetooth 5.0 specification [3] has almost 3,000 pages excluding different application protocols (a.k.a., profiles).

Protocol related bugs contribute to the IoT’s wide attack surface. Linux kernel maintains a long list of “unusual” USB devices called “quirks” [4], which essentially violate some parts of the USB specification. Over 80 bugs were found within the Linux USB subsystem as well in the past year [5]. Even worse, attackers can exploit these vulnerabilities to steal sensitive information or achieve privilege escalation within target systems. BlueBorne [6] allows attackers to inject worms into Android devices by exploiting a Bluetooth protocol stack vulnerability within Android. BleedingBit [7] enables attackers to break into an enterprise environment undetected by exploiting a Bluetooth Low Energy (BLE) vulnerability within TI devices. In the worst case, attackers can reprogram the firmware to add malicious functionality, such as BadUSB attacks [8].

Accordingly, existing firmware analysis frameworks either ignore the domain knowledge from these protocols, e.g., treating firmware the same as typical binary executables [9], [10], [11], [12], or require manual effort to study these protocols and reverse engineer the firmware [13]. The complexity of protocol implementations, i.e., complex dataflow and implicit calls through function pointers, leaves pure NLP-based approaches and light-weight static analysis ineffective for reasoning about protocol-relevant behavior. Dynamic symbolic execution, on the other hand, serves as a feasible approach to solving this problem due to its precise memory model.

In this paper, we propose a new firmware analysis methodology called ProXray, which utilizes dynamic symbolic execution. As illustrated in Figure 1, ProXray can learn a protocol model from known firmware and apply the model to recognize the protocol and to identify functionality within unknown firmware automatically.

ProXray has three stages. To learn a protocol model without a formal specification of the protocol, we use various path prioritization heuristics for symbolic execution to extract protocol field constraints by running some known firmware implementing the protocol. The protocol model is essentially the collection of those constraints. Once the protocol model is available, ProXray can apply the model
to unknown firmware to recognize the protocol usage by mapping binary execution into protocol field constraints using symbolic execution. After the firmware is tagged with protocol fields automatically, ProXray accepts queries in the form of protocol relevant data constraints, leverages the protocol model again to accelerate path explorations in symbolic execution, and generates answers.

The first stage of our approach, in which we extract protocol constraints from the firmware in the training set, can be viewed as an instantiation of the generic feature extraction approach of Eisenbarth et al. [46]. The novelty of our approach stems from the fact that we represent the extracted features formally and combine these formal models extracted from different source code to generate a protocol constraint model. The benefit of having a protocol model is that when we do not have the source code available for a system under analysis and, therefore, cannot apply the approach of Eisenbarth et al. [46], we can reverse engineer the protocol relevant features of the code via Protocol Field Discovery and leverage this information in Protocol Model Guided Symbolic Execution.

We have applied ProXray to the USB protocol by learning a USB protocol model from 23 known USB firmware images. We are then able to automatically map protocol fields and identify USB functionality by answering queries like “Does the firmware contain a USB keyboard functionality?” within all 6 unknown USB firmware images. We have also applied ProXray to the Bluetooth protocol and achieved 100% precision in mapping the variables that correspond to the two fields, wValue and bRequest, respectively, which are instrumental in determining the functionality of USB devices. For the Bluetooth protocol, we achieved 100% precision for the HCI event type and the L2CAP signaling command code fields.

We design a protocol guided symbolic execution algorithm that identifies functionality and answers queries about unknown firmware.

We have applied our methodology to the USB protocol and the HCI and L2CAP layers of the Bluetooth protocol and evaluated ProXray using the firmware that runs on MSP430 and Intel 8051 microcontrollers. Our experimental results show at least an order of magnitude speedup for the Bluetooth firmware and up to 73.8 times speedup in reaching USB protocol related targets for MSP430 firmware and 1.5 times speedup for 8051 firmware. Our approach can correctly match functionality of every USB firmware in our test set from both MSP430 and Intel 8051 architectures.

The rest of the paper is organized as follows. Section 2 provides background information on the symbolic execution of firmware. Section 3 presents an overview of our methodology. Section 4 presents the individual steps of our approach. Section 5 applies our methodology to the USB and Bluetooth protocols. Section 6 presents experimental results. Section 7 discusses threats to validity and other challenges. Section 8 presents the related work, and Section 9 concludes and considers future work.

2 BACKGROUND

2.1 Protocol Model Learning

Model learning refers to inferring a model of software components [1]. This model could be a Hidden Markov-Chain Model (HMM), for example, or relations between firmware directly. Our experiments show that our constraint-based search prioritization heuristics perform better than the baseline symbolic execution (BSE) and extract up to 1.54 and 1.76 times more unique constraints on average for the USB and the Bluetooth protocol, respectively.

- We present a discovery algorithm that uses the protocol model to recognize the protocol usage within unknown firmware automatically. Our experiments show that our approach achieves on average 100% and 75% precision in mapping the variables that correspond to the two fields, wValue and bRequest, respectively, which are instrumental in determining the functionality of USB devices. For the Bluetooth protocol, we achieved 100% precision for the HCI event type and the L2CAP signaling command code fields.

- We design a protocol guided symbolic execution algorithm that identifies functionality and answers queries about unknown firmware.

- We have applied our methodology to the USB protocol and the HCI and L2CAP layers of the Bluetooth protocol and evaluated ProXray using the firmware that runs on MSP430 and Intel 8051 microcontrollers. Our experimental results show at least an order of magnitude speedup for the Bluetooth firmware and up to 73.8 times speedup in reaching USB protocol related targets for MSP430 firmware and 1.5 times speedup for 8051 firmware. Our approach can correctly match functionality of every USB firmware in our test set from both MSP430 and Intel 8051 architectures.

The rest of the paper is organized as follows. Section 2 provides background information on the symbolic execution of firmware. Section 3 presents an overview of our methodology. Section 4 presents the individual steps of our approach. Section 5 applies our methodology to the USB and Bluetooth protocols. Section 6 presents experimental results. Section 7 discusses threats to validity and other challenges. Section 8 presents the related work, and Section 9 concludes and considers future work.

2 BACKGROUND

2.1 Protocol Model Learning

Model learning refers to inferring a model of software components [1]. This model could be a Hidden Markov-Chain Model (HMM), for example, or relations between
objects, class hierarchies, or implemented protocols. For the latter, protocol state transitions are often set as the learning target. Depending on the availability of the code in some analyzable form, either a black-box or a white-box method can be applied. White-box methods learn the model by analyzing the source code or the binary executable. When it is not feasible to analyze the implementation of a system under analysis, black-box methods are used to infer the state machines by observing the inputs and outputs of the program. Both passive and active learning are possible. In passive learning, the training data is labeled upfront. In active learning, however, labeling is performed on the specific instances when explicit queries are submitted. Unlike previous protocol learning techniques (see Section 3 for a detailed discussion), ProXray targets learning of protocol constraints rather than a state machine of the protocol. ProXray is essentially a white-box and a passive model learning method.

2.2 Firmware Analysis using Symbolic Execution

Classical symbolic execution uses symbolic values for inputs and executes the instructions symbolically to propagate symbolic data flow among the program variables. When a branch instruction that involves a symbolic condition gets executed, multiple successors may potentially be created to represent the feasible paths in the program. Each path is associated with a symbolic expression, the path condition, to represent all decisions made on the symbolic inputs along that path.

Dynamic symbolic execution extends classical symbolic execution by mixing concrete and symbolic values to deal with challenging cases such as library calls and non-linear expressions. Concolic Testing [17] and Execution-Generated Testing (EGT) [16] are two specializations of dynamic symbolic execution. In Concolic Testing the program is executed with concrete input values while computing both concrete states and symbolic states for variables that have symbolic values. It uses the symbolic path expressions to generate new concrete input values that can potentially execute new parts of the code. EGT, on the other hand, uses symbolic input values, keeps a symbolic state for the relevant variables, and performs concrete computation only when all the variables are concrete. So an EGT based symbolic execution engine, such as KLEE [17], can mix concrete inputs with symbolic inputs. Please see [19] for an overview on various symbolic execution approaches. In this paper, the algorithms we present in Section 5 assume that the underlying symbolic execution engine implements the EGT approach.

Symbolic execution of firmware poses challenges that do not exist for the symbolic execution of user space applications. These include specification of the architectural elements, e.g., special function registers of a specific microcontroller architecture, and the interrupt service routines (ISRs). FIE [9] extends the KLEE symbolic execution engine to enable analysis of MSP430 firmware. It enables specification of the microcontroller specific memory layout and symbolic regions as well as ISRs. It implements an approximate interrupt scheduling policy. In previous work [13], we have extended FIE with binary execution capability and a support for Intel 8051 firmware by developing an LLVM lifter for Intel 8051 ISA and modeling the architectural elements. ProXray uses FIE with our extensions to extract USB protocol relevant constraints from a set of 23 MSP430 firmware. The combination of extracted constraints that are rewritten in terms of the protocol fields forms the learned protocol model. We have used ProXray and the learned model to a test set of firmware that consists of four MSP 430 firmware images and two Intel 8051 firmware images to show the effectiveness of our approach in reaching protocol relevant code locations, which we call targets, and identifying the specific functionality implemented by these firmware images.

3 System Overview

This section presents an overview of ProXray, which extracts a formal protocol model from known firmware and leverages it in the analysis of unknown firmware. Figure 1 shows the data flow and major processing phases.

Model Extraction Phase: The first phase involves protocol model extraction. We assume the availability of a representative set of firmware implementing the protocol of interest. This is a reasonable assumption as microcontroller vendors provide software developer packages (SDKs), e.g., the MSP430 SDK provided by Texas Instruments [19], for their boards and the associated toolchains. These developer packages demonstrate programming of both the microcontroller specific features and the communication protocols that they use. This type of program-based documentation of the protocol supplements the original protocol specification and helps developers understand how to program their firmware to use the protocol. SDKs often use identifier names that are similar to the protocol field names. Thus, in firmware that comes with an SDK, a simple text search often suffices to identify the variables that implement the protocol field names.

Our approach can leverage the domain knowledge encoded in such sample firmware to extract a constraint-based model of the protocol. As shown in Figure 1, each firmware that is used in the training set is accompanied with a mapping of the protocol fields to the memory locations (program variables) in the firmware. In this paper, we assume that the mapping information needed for the model extraction phase is generated manually by scanning the source code of the sample firmware.

The protocol model extraction stage takes the sample firmware along with the associated mapping between the protocol fields and the program variables. It performs symbolic execution on a set of training firmware in a way that prioritizes exploration of protocol relevant paths. The main information that guides the symbolic execution of this stage is the protocol field mapping and the protocol-related symbolic program constraints extracted from various branches in the firmware as paths are explored. Let us assume that the variables sdata and req of a firmware in the training set have been mapped to the USB protocol fields bRequestType and bRequest, respectively (see Section 5.1 for a brief background on the USB protocol).

In this paper, we use memory locations and program variables interchangeably. The memory location concept facilitates explanation of our algorithms that extend basic symbolic execution.
One way of guiding the exploration is to prioritize paths that have explored new symbolic regions representing some of the protocol fields. As an example, assume that a path has so far executed branch conditions involving the variable \( sdata \). If a successor of this path has recently executed a branch condition involving \( req \) while another successor has not executed such a branch condition yet then the former path should be prioritized over the latter. Another way is to prioritize paths based on the number of different branch conditions that they have executed. As an example, a path that has recently checked a new condition, e.g., \( sdata == 1 \) in addition to \( sdata == 0 \) should be prioritized over a path that has only checked \( sdata == 0 \).

Protocol relevant program constraints are captured in a canonical form, \( exp \ OP \ k \) where \( OP \) and \( k \) denote a relational operator and a constant, respectively. \( exp \) may consist of a protocol field, \( m \), or an expression that involves one or more bitwise operators manipulating some protocol fields, e.g., \( m \ BOP \ k_1 \) or \( (m_1 \ BOP_1 \ k_1) \ BOP_2 \ m_2 \), where \( m, m_1, \) and \( m_2 \) denote protocol fields, \( BOP, BOP_1 \) and \( BOP_2 \) denote bitwise operators, and \( k_1 \) denotes a constant. Basically, our atomic constraints correspond to the atomic constraints of the Bitvector Logic \([20]\). Extracted protocol-relevant constraints are rewritten by replacing each variable with the corresponding protocol field. Thus, the constraint model shown in Figure 1 consists of a set of canonical constraints, where the identifiers in the constraints correspond to some protocol field. Protocol constraints that are extracted from each firmware in the training set are eventually combined to represent the protocol constraint model. Specifically, we compute the union of all sets of canonical constraints to combine them. The protocol model is leveraged by the second and third phases of our approach to uncover functionality of unknown firmware.

**Protocol Field Discovery:** In the second phase of our approach, a given unknown firmware is analyzed to discover the mapping between the variables of the firmware and the protocol fields. It should be noted that for arbitrary firmware, which may be in binary form only, it may not be feasible to do this manually. We propose an automated technique to discover this relationship between the firmware variables/memory regions and the protocol fields. We achieve this by leveraging the protocol constraint model. Our approach employs standard symbolic execution to explore some of the paths and compares the constraints in the path condition with those in the protocol constraint model for detecting semantic matching. Those constraints that match semantically provide potential mappings. We use statistical information, i.e., the number of times a memory region is matched to a protocol field, to refine the discovered mapping.

**Protocol Model Guided Symbolic Execution:** The third phase of our approach leverages the discovered mapping for the unknown firmware to perform protocol-guided symbolic execution. The inputs consist of an unknown firmware that has not been analyzed w.r.t. to protocol use, the discovered mapping between the firmware variables and the protocol fields, and a query in the form of a protocol constraint representing the part of the protocol of interest. Note that the query helps to focus the analysis to a specific functionality of the protocol. Our guided symbolic execution rewrites the input protocol constraint to reflect all possible mappings that have been discovered. As an example, in one of the USB firmware we examined, the \( bmRequestType \) field was implemented by the variable \( tSetupPacket.bmRequestType \). So, the constraint \( bmRequestType == 161 \) as a query is rewritten as \( tSetupPacket.bmRequestType == 161 \) if \( bmRequestType \) has been mapped to \( \{tSetupPacket.bmRequestType, sdata\} \) and is rewritten as \( tSetupPacket.bmRequestType == 161 \) \( \lor \) \( sdata == 161 \) if \( bmRequestType \) has been mapped to \( \{tSetupPacket.bmRequestType, sdata\} \) to reflect all possible candidate mappings. Then we apply a customized pruning algorithm to expand the paths that satisfy the transformed protocol constraint. In this paper, we focus on two applications of protocol-guided analysis of unknown firmware: 1) checking whether the unknown firmware handles a given protocol constraint and 2) discovering the functionality class(es) an unknown firmware implements.

### 4 APPROACH

The goal of ProXray is to automatically extract a constraint-based model of a protocol using a representative set of firmware and then use this model to perform guided symbolic execution in new firmware. We present our protocol model extraction in Section 4.1, discovery of protocol elements in a new firmware in Section 4.2, and protocol guided symbolic execution in Section 4.3.

#### 4.1 Protocol Model Extraction

**Algorithm 1 Protocol Constraint Model Extraction Algorithm**

1. \( \text{ExtractProtocolConstraints}(F : \text{Firmware}, M : \text{MemLoc} \rightarrow \text{Identifier}, C_{\text{scope}} : \{\text{perPath, perGroup}\}, C_{\text{cov}} : \{\text{code, constraint, field}\}, C_{\text{window}} : N, \tau : \mathcal{R})\):
2. \( s_0 : \text{SEState} \)
3. Let \( s_0 \) denote the initial symbolic execution state/path for \( F \)
4. \( \text{global ActivePaths} \leftarrow \{s_0\} \)
5. \( \text{global SC} \leftarrow \emptyset \)
6. \( \text{global stashStack} \leftarrow \text{empty stack of SEState} \)
7. \( \text{global newPaths} \leftarrow \emptyset \)
8. \( \text{global toBeStashed} \leftarrow \text{ActivePaths} \)
9. \( \text{start the new window of size } C_{\text{window}} \text{ for executing paths in } \text{ActivePaths} \)
10. \( \text{while } \tau \text{ seconds not elapsed and } \text{ActivePaths} \neq \emptyset \text{ do} \)
11. \( \text{while exists some path in } \text{ActivePaths} \text{ for which end of window has not been reached do} \)
12. \( s \leftarrow \text{chooseNext(ActivePaths)} \)
13. \( \text{update } s's \text{ coverage based on } C_{\text{cov}} \text{ and the next instruction} \)
14. \( s'.\text{successors} \leftarrow \text{ExecuteNextInstruction}(s) \)
15. \( \text{for each } s'' \in s'.\text{successors} \text{ do} \)
16. \( \text{Let } \text{Mem} \text{ denote the memory locations that appear in } s''.\text{PC} \)
17. \( \text{ExtractAtomicConstraints(Mem, s''.\text{PC}, M)} \)
18. \( \text{end for} \)
19. \( \text{if FilterStates(C_{\text{scope}}, s) then} \)
20. \( \text{break} \)
21. \( \text{end if} \)
22. \( \text{end while} \)
23. \( \text{NextFrontierSet()} \)
24. \( \text{start the new window of size } C_{\text{window}} \)
25. \( \text{end while} \)
26. \( \text{return SC} \)

Algorithm 1 presents our approach for extracting a protocol constraint model from a single firmware \( F \). It takes...
as input a mapping $M$ from firmware memory locations to the protocol fields and configuration options that we will explain below. It uses symbolic execution to explore the paths in $F$ and returns a set of atomic constraints, $SC$, on the protocol fields.

One challenge of symbolic execution is the path explosion problem. The exponential growth in the number of paths slows down the progress made for each path. Depending on the goal of the underlying analysis, the achieved coverage may be far from the ideal. Our goal is to explore the paths of the firmware that implement the protocol functionality while traversing as diverse set of protocol relevant paths as possible. Although symbolic execution engines come with path exploration heuristics such as those based on random selection and coverage, we need customized heuristics to maximize the unique number of protocol constraints extracted.

We thus designed three types of “knobs” to fine tune our model extraction process. The first knob, $C_{conv}$, configures the type of coverage we would like to measure as the exploration progresses. In addition to traditional code coverage, we also consider constraint coverage and protocol field coverage. The second knob, $C_{scope}$, configures the scope of coverage computation. One possible scope is per-path and the other possibility is per-group (considering all active paths). The third knob, $C_{window}$, configures the duration of coverage computation, denoting the number of blocks executed.

Algorithm 1 keeps track of a frontier set of symbolic execution states or paths, $ActivePaths$, that is initialized with the initial state (line 4). It keeps a stack of stashed paths, $stashStack$, representing all paths other than the current frontier set and to be considered later. The algorithm runs until a time bound $\tau$ is reached or the frontier set becomes empty. Each time $ActivePaths$ is initialized, a window of size $C_{window}$ starts (lines 9 and 24). Until the window ends, paths in $ActivePaths$ execute as in standard symbolic execution by choosing a path from the frontier set (line 12), updating coverage based on the type of coverage configuration (line 13), and executing the next instruction to compute possible successors (line 14).

After executing an instruction for a path $s$, the path conditions of the successors are analyzed (lines 15-18) using Algorithm 2 to check the appearance of a new protocol relevant constraint. Algorithm 2 uses the protocol field mapping $M$ to identify such constraints (lines 3-4), which are stored in the set $SC$ (line 5).

Algorithm 2 An algorithm for extracting atomic constraints from a given expression using a mapping from the memory locations to the protocol fields.

```
1: ExtractAtomicConstraints(Mem : \mathcal{P}(MemLoc), E : Expression, M : MemLoc \rightarrow Identifier)
2: M_{Mem} \leftarrow \{ m \mid m \in MemLoc \land (M(m) \neq undef) \}
3: for each \chi \in ATOMIC(E) and m \in M_{Mem} do
4: if m \in Var(\chi) then
5: SC \leftarrow SC \cup \chi[M(m)/m]
6: end if
7: end for
```

Figure 2 shows a sample code snippet from the MSP430 SDK for the USB protocol [19]. In this SDK, class specific descriptors are defined by class specific code (lines 5-18). The common functionality (lines 21-58) uses data structures defined by the class specific functionality, e.g., the tUsbRequestList array, to check for the matching request types and the requests. Figure 3 shows the extracted protocol relevant constraint on the protocol field $bRequest$ that corresponds to the condition at line 32 when the request matches the fields of the array entry on lines 7-16.

Figure 4 shows a sample code snippet from an implementation of the Bluetooth protocol [21]. The transport layer function $hci_transport_h4_block_read$ (line 22) reads the raw HCI packet and calls a callback function through the function pointer $packet_handler$ (line 23), which gets resolved to the $packet_handler$ function defined on line 28. The $packet_handler$ function checks the HCI packet type and calls the $event_handler$ function for HCI event type packets. Line 41 shows one of the cases handled by the $event_handler$ function that involves checking the HCI Command Complete event for the sub-event $hci_read_local_name$. To check this particular case, it uses the $HCI_EVENT_IS_COMMAND_COMPLETE$ macro and according to the definition of this macro (lines 16-17) it checks two conditions: 1) the event code corresponding to $HCI_EVENT_COMMAND_COMPLETE$ (line 16) and 2) the sub-event code that consists of two bytes and when read in little endian mode corresponds to the opcode of the $hci_read_local_name$ command (line 17). Figure 5 shows the extracted atomic constraint that corresponds to the condition at line 17. In this case the bitwise operators are the bitwise Or and the shift left (Shl) operators.

In addition to $ActivePaths$, Algorithm 1 keeps two sets of paths, $newPaths$ and $toBeStashed$, denoting the next set of frontier paths and the paths from the current frontier set that will be stashed away. Filtering of paths into the next frontier set and stashing of others are performed by Algorithm 3 and depends on the configured scope of the analysis. At the beginning of each window, $newPaths$ and $toBeStashed$ are initialized to an empty set and to $ActivePaths$, respectively (lines 7-8 of Algorithm 1 and 11-12 of Algorithm 4). If the scope is per group (lines 2-14) then filtering does not happen until all the paths in the frontier set reach the end of the current window. While inside the window, the algorithm updates the frontier set (line 4) and the paths to be stashed away (line 5).

When a path reaches the end of the current window, it is removed from the frontier set (line 7). If such a path achieves new coverage based on the type of coverage, its successors are added to the next frontier set (line 9) and it is not considered for stashing (line 10). Otherwise, its successors are stashed (line 12).

If the scope is per path (lines 15-19) then new coverage is checked after each instruction execution (line 15) and handled immediately even if the end of the window is not reached yet. As in the per group case, when a path is filtered the next set of frontier paths and the set of stashed paths get updated (line 16 and 17). Unlike in the per group case, when a path is filtered it signals to stop executing the paths in the current frontier set (line 18).

After executing the paths in the current frontier set, Algorithm 4 calls Algorithm 2 to compute the next frontier set (line 23). Algorithm 4 pushes the stashed paths, if any, onto the stack (line 3) and updates the frontier set using the
Fig. 2: Code snippets from the USB SDK for MSP430 MCUs [19].

Fig. 3: The extracted USB protocol constraint (the `USB_MSC_GET_MAX_LUN` request) that corresponds to the condition on line 32 in Figure 2 when the request matches the fields of the array entry bw lines 7-16 in Figure 2.

paths that have been filtered, if any (line 6). On the other hand, if no paths could be filtered from the frontier set, the next set of frontier paths is received from the top of the stack of stashed paths (lines 8 and 9).

4.2 Protocol Field Discovery

The goal of protocol field discovery is to analyze an arbitrary firmware known or suspected to implement a protocol functionality and identify the set of potential memory locations corresponding to each data field of the protocol. The idea is to use symbolic execution to explore paths of the firmware under analysis and utilize the protocol constraint model that has been extracted as presented in Section 4.1 for this discovery process.
Algorithm 3 An algorithm for filtering the symbolic execution states based on relevance to extracting protocol relevant constraints.

1: FillerStates($C_{scope} : \{ perPath, perGroup \}, s : SEState): boolean
2: if $C_{scope} = \text{perGroup}$ then
3: if not end of current window for $s.successors$ then
4: $ActivePaths \leftarrow ActivePaths \cup s.successors \\{s\}$
5: $toBeStashed \leftarrow toBeStashed \cup s.successors \\{s\}$
6: else
7: if $s$ covers new based on $C_{cov}$ in the current window then
8: $newPaths \leftarrow newPaths \cup s.successors$
9: $toBeStashed \leftarrow toBeStashed \\{s\}$
10: else $\triangleright$ no new coverage by the end of the window
11: $toBeStashed \leftarrow toBeStashed \cup s.successors \\{s\}$
12: end if
13: end if
14: else if $s$ covers new based on $C_{cov}$ then $\triangleright C_{scope} = \text{perPath}$
15: $newPaths \leftarrow newPaths \cup s.successors$
16: $toBeStashed \leftarrow toBeStashed \\{s\}$
17: return true
18: end if
19: end if
20: return false

Fig. 4: Code snippets from a Bluetooth stack implementation [21].

Algorithm 5 presents the details of our protocol field discovery algorithm. The inputs include a set of atomic constraints over protocol fields $PM$, the set of protocol fields $PF$, and a time threshold $\tau$. The algorithm returns mappings between the protocol fields and the set of memory locations. It collects candidate mappings as it explores program paths by running the standard symbolic execution algorithm. When a path condition is updated, it checks to see if each constraint in the path condition matches some
Algorithm 4 An algorithm for computing the next frontier set.
1: Next Frontier Set()
2: if toBeStashed ≠ ∅ then
3:   stashStack.push(toBeStashed)
4: end if
5: if newPaths ≠ ∅ then
6:   ActivePaths ← newPaths
7: else
8:   ActivePaths ← stashStack.top()
9: end if
10: if newPaths ≠ ∅ then
11:   ActivePaths ← newPaths
12: end if
13: end for
14: end for
15: end if
16: return Mnewest

Algorithm 5 Protocol Field Discovery Algorithm
1: DiscoverProtocolFields(F : Firmware, PM : T(Constraint), PF : T(Identifier), τ : R : Identifier → T(MemLoc)
2: Let M, Mnewest : Identifier → T(MemLoc)
3: M, Mnewest ← λx.∅
4: Let Freq : MemLoc → Identifier → N
5: Freq ← λx.0
6: Let s0 denote the initial symbolic execution state/path for F
7: ActivePaths ← {s0}
8: while τ seconds not elapsed and ActivePaths ≠ ∅ do
9:   s ← chooseNext(ActivePaths)
10:   s.successors ← ExecuteNextInstruction(s)
11:   ActivePaths ← ActivePaths ∪ s.successors \ {s}
12:   for each s′ ∈ s.successors do
13:     for each cpc ∈ ATOMIC(s′,PC) do
14:       for each pf ∈ Identifier do
15:         if ∃ x ∈ PM.(isValid(c → c[pf]/m)) then
16:           M ← M[pf] = M[pf] \ {m}
17:           Freq ← Freq[m,pf] = Freq[m,pf] + 1
18:       end if
19:     end for
20:   end for
21: end while
22: Mnewest ← Apf.(m | Freq[m,pf] = MAXpf,cPF(Freq[m,pf]))
23: return Mnewest
24: end if
25: end for
26: while τ seconds not elapsed and ActivePaths ≠ ∅ do
27:   s ← chooseNext(ActivePaths)
28:   s.successors ← ExecuteNextInstruction(s)
29:   filtered, pruned ← false
30:   filteredPaths ← ∅
31:   for each s′ in s.successors do
32:     if s′.PC ∧ Q′ ≠ false then
33:       filtered ← true
34:     end if
35:     if s′.PC ∧ Q′ = false then
36:       pruned ← true
37:     end if
38:   end for
39:   if filtered ∧ pruned then
40:     ActivePaths ← filteredPaths
41:     else
42:     ActivePaths ← ActivePaths ∪ filteredPaths \ {s′}
43:   end if
44: end while
45: return ActivePaths

Algorithm 6 Protocol Guided Symbolic Execution Algorithm
1: ProtocolGuidedSymEx(F : Firmware, Q : Constraint, M : Identifier → T(MemLoc), τ : R : T(SEState)
2: Let Q = V[1..N] ∧ ∨ T(yj)
3: Let T(q) = \{ V[m,pf] M[pf] = ∅ \} otherwise \ case 2  \}
4: Let Q′ = V[1..N] ∧ ∨ T(yj)
5: Let s0 denote the initial symbolic execution state/path for F
6: ActivePaths ← {s0}
7: while τ seconds not elapsed and ActivePaths ≠ ∅ do
8:   s ← chooseNext(ActivePaths)
9:   s.successors ← ExecuteNextInstruction(s)
10: filtered, pruned ← false
11: filteredPaths ← ∅
12: for each s′ in s.successors do
13:   if s′.PC ∧ Q′ ≠ false then
14:     filtered ← true
15:   end if
16:   if s′.PC ∧ Q′ = false then
17:     pruned ← true
18:   end if
19: end for
20: if filtered ∧ pruned then
21:   ActivePaths ← filteredPaths
22: else
23:   ActivePaths ← ActivePaths ∪ filteredPaths \ {s′}
24: end if
25: end while
26: return ActivePaths

Algorithm 4 presents our protocol model guided symbolic execution algorithm. The input includes a reachability query, Q, in the form of a protocol relevant constraint (the query in Figure 3), discovered mappings from protocol fields to a set of memory locations, M (generated by Algorithm 3), and a time threshold τ. The algorithm first transforms the protocol relevant input query into a program specific constraint by utilizing the mapping M (line 4). If a protocol field is mapped to a nonempty set of memory locations (line 3, case 1) then each memory location is considered to be legitimate separately and the query is rewritten by replacing the protocol field with that memory location. Constraints that are obtained through rewriting are combined using the disjunction operator. However, if a protocol field could not be mapped to a memory location (line 3, case 2) then the constraint is replaced with true.

The algorithm has some mappings between protocol fields and observed between a memory location and a protocol field that appears in the matching constraint from the constraint in the protocol model. If so, it maps the memory location that matches the constraint in Figure 3 (line 15 of Algorithm 5) and the memory region SETUPDAT (line 16 of Algorithm 5) to the protocol field bRequest (line 16 of Algorithm 5).
The algorithm keeps a set of frontier paths, ActivePaths, and starts the symbolic execution from the initial symbolic execution state for $F$. In standard symbolic execution, it chooses the next path to execute (line 8) and executes the instruction to produce the successors (line 9). Some of the successors may be filtered (lines 13-16) and some of them may be pruned (line 17-19). It keeps a set to record which successors of the current path gets filtered in filteredPaths, which has been initialized to an empty set at line 11. Path condition of each successor will be checked to decide if it satisfies part of the transformed protocol constraint. If so, the successor will be added to the set of filtered paths (line 15). If there are filtered paths as well as pruned paths, the frontier set is updated with the filtered successors of the current path. Otherwise, the frontier set is expanded with the filtered successors (line 24). The intuition behind this is to detect branches where decisions related to the query are made and to aggressively retrace the active paths with the successors of the current path that make decisions consistent with the query. At branches without query related decisions, the algorithm preserves the current active set by expanding it with all the successors.

In one extreme case, each protocol field might be mapped to an empty set and the transformed query would evaluate to true. In the other extreme case, each protocol field might be mapped to a large set of memory locations yielding a large transformed query. In both cases, steering the execution to the relevant part of the program will not be effective as effective pruning will not be achieved. Moreover, the latter will have an additional overhead in terms of constraint solving due to the size of the transformed query. So the effectiveness of Algorithm 5 in steering the execution to the desirable part of the program depends on the precision of the mapping between the protocol fields and the memory locations provided as an input.

As an example, consider the BadUSB code given in Figure 6 and a query that describes the USB standard request for reporting of the HID functionality: $bRequest = 0x06 \land wValue \land bRequest \land \text{GET_DESCRIPTOR}$. Assume that Algorithm 5 has mapped the $bRequest$ field and the $wValue$ field to $\{\text{SETUPDAT[1]}, \text{SETUPDAT[3]}\}$ and $\{\text{SETUPDAT[3]}\}$, respectively, in the context of the BadUSB firmware. This mapping gets passed to Algorithm 5 along with the query. The query would be rewritten as $(\text{SETUPDAT[1]} = 0x06 \lor \text{SETUPDAT[3]} = 0x06) \land \text{SETUPDAT[3]} = 0x22$ to generate a constraint, as shown on line 4 of Algorithm 5 that refers to the memory regions of the BadUSB firmware. Using this constraint, the guided symbolic execution stage, as shown on lines 7-26 of Algorithm 6 prunes the paths that satisfy the conditions of the lines 45, 53, 54, 62, and 63 and reaches the target location of reporting the HID functionality by filtering paths that satisfy the conditions of the lines 44, 52, and 65 of the BadUSB firmware.

5 Applying ProXray to the USB and Bluetooth Protocols

5.1 Universal Serial Bus

We choose the USB protocol as one of our case studies for ProXray because of its ubiquity in embedded systems and IoT devices. On one hand, the core USB specification, e.g., USB 2.0 [23] or USB 3.0 [24] is still approachable with...
few hundreds of pages. On the other, its ability to support versatile functionality via different USB classes also reflects the challenges in firmware analysis.

Each USB class defines one kind of functionality. The most common classes are Communication Device Class (CDC), Human Interface Device (HID), and Mass Storage Class (MSC). Different USB classes introduce their own request/response messages, which follow the standard request/response structure defined by the USB spec. As shown in Figure 7, all USB devices follow the same procedure called enumeration once plugged into the host machine. Initiated by the host, this procedure is used to provide the device configuration information, including GetDeviceDescriptors, GetConfigDescriptors, and GetInterfaceDescriptors. Once the enumeration phase is complete, the corresponding device driver loaded by the OS starts to serve the device using USB class-specific requests.

All USB requests start with a Setup packet, as shown in Figure 8. This is an 8-byte structure, containing 1-byte bmRequestType and bRequest fields, and 2-byte wValue, wIndex, and wLength fields. bmRequestType is a bitmap determining data transfer direction, type, and recipient. bRequest is the request code defined by the USB and class-specific specs. Both wValue and wIndex are separated into low and high bytes, which act as parameters passed by a given request type. wLength shows the number of bytes to be transferred during the data stage if one exists.

5.2 Bluetooth

While USB dominates wireline connections for embedded systems and IoT devices, Bluetooth provides a most common way to connect with these devices wirelessly, especially after the introduction of Bluetooth Low Energy (BLE) and Bluetooth Mesh. Comparing to the USB protocol, Bluetooth is a more complex protocol suite, including different layers within the protocol stack. As shown in Figure 9 from the bottom up we have Bluetooth Module, Host Controller Interface (HCI), Logical Link Control and Access Protocol (L2CAP), a variety of sub protocols used to support different Bluetooth Profiles, and finally the Applications.

The HCI layer is the bottom part of the Bluetooth software stack passing HCI packets to the hardware Bluetooth modules. Packets at this layer follow typical TLV format. There are four types of HCI packets in total, including Command, Event, ACL, and SCO. As its name implies, the L2CAP layer maintains logical connections between different Bluetooth devices, and can be treated as the transport layer within the stack (e.g., TCP and IP). Similar to TCP, L2CAP provides the foundation of different application protocols defined by the Bluetooth specification. Given their importance, we focus on the protocols fields within the HCI and the L2CAP layers.

5.3 Model Generation and Analysis

FIE is a firmware analysis tool that leverages KLEE to perform symbolic execution on MSP430 firmware. We extend FIE to record all conditions evaluated on any protocol field while executing the training firmware set. The core execution engine of FIE considers every read from a symbolic memory region independently and as a new version of that location on the executing path. As a result, every read from a specific memory creates a new node in the underlying abstract syntax tree (AST) used for representing symbolic expressions. This procedure helps in capturing the dynamic interactions between the firmware and its environment. However, this complicates our model extraction as we have to deal with multiple versions of the same memory location. To get the unique constraints imposed on each protocol field, we first find the memory regions used in each branch condition. We developed a custom expression AST traversal to emit the atomic constraints from the branch conditions. If the memory region in the atomic constraint turns out to be a protocol field, we rewrite the conditional expression in terms of that protocol field. For every such expression we evaluate uniqueness of that constraint using validity checking interface of the Simple Theorem Prover (STP), which is the SMT solver that FIE uses. If the new conditional expression turns out to be different from all the constraints in our model, we accept it as a newly found
unique constraint. We store all the unique constraints in the syntax of the KQuery [29] language.

We also use the KQuery language in specifying protocol relevant queries for guided execution. Based on the protocol field mapping found for the firmware under test, we rewrite these queries in terms of the mapped memory regions of the firmware. We utilize the caching solver provided by KLEE in all phases of our approach to minimize the runtime overhead.

6 Evaluation
To evaluate the effectiveness of ProXray in the context of the four research questions we mentioned in Section 1, we applied it to the USB and Bluetooth protocols. For the USB protocol we used the USB development package available from MSP’s USB developer site [19] and two Intel 8051 firmware, Phison BadUSB firmware [22] and EzHID firmware [30]. The MSP430 package is provided as an example for USB firmware developers working on the MSP430 architecture. It contains a rich set of example firmware images with each demonstrating a different use-case of the USB protocol. In general, a single firmware image focuses on a specific device class, except for the composite firmwares, which combine functionalities from multiple classes. For the Bluetooth protocol, we used the BTStack framework [21] that implements the Bluetooth Core Specification version 4.0 and provides a variety of embedded firmware examples, eight of which were ported to an MSP 430 architecture. We chose BTStack due to being a fully open source implementation and having MSP 430 examples that we could analyze with FIE.

We divide all the firmware into the training set and the testing set for each protocol type. The USB training set contains 23 firmware images from the MSP430 package, implementing 3 different USB classes, including Communication Device Class (CDC), Human Interface Device (HID), and Mass Storage Class (MSC). The USB testing set contains 6 firmware, including another 4 firmware from the MSP430 package and 2 Intel 8051 firmware. Table 1 shows the individual USB firmware images and their size while Table 2 shows the number of firmware images by the USB class type along with the quantitative details on the unique constraints extracted for each. Table 1 also shows the individual firmware images used for Bluetooth experiments. The sizes include the size of the BTStack core. We evaluate our Protocol Model Extraction (6.1) using the training set, Field Discovery (6.2), and finally Guided Execution (6.3) using the testing set.

6.1 Protocol Model Extraction
In this subsection, we answer the question: Does constraint-based search prioritization perform better than the baseline symbolic execution (BSE)? We have found out that all of our heuristics provided in Algorithm 1 perform better than the baseline symbolic execution (BSE), i.e., they extract more protocol information from the training firmware set in a given amount of time compared to BSE as shown in Figure 10, which shows the average number of unique constraints extracted over a period of 15 minutes (900s). Figure 11 and 12 shows model extraction data for USB, Bluetooth, HCl, and Bluetooth L2CAP protocols, respectively. The graphs compare the performance of the heuristics on firmware where BSE showed its best performance, i.e, BSE’s number of extracted constraints were highest in the given time among all the firmware using that protocol. Our heuristics achieved much better performance than BSE in all cases.

The USB host request in the protocol is dependent on these five fields: bmRequestType, bRequest, wValue, wIndex, and wLength. For Bluetooth, the requests are based on

<table>
<thead>
<tr>
<th>Firmware Name</th>
<th>Lines of Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td></td>
</tr>
<tr>
<td>USB CDC</td>
<td></td>
</tr>
<tr>
<td>C0_SimpleSend</td>
<td>18,764</td>
</tr>
<tr>
<td>C1_LedOnOff</td>
<td>18,855</td>
</tr>
<tr>
<td>C2_ReceiveData</td>
<td>18,764</td>
</tr>
<tr>
<td>C3_EchoToHost</td>
<td>18,493</td>
</tr>
<tr>
<td>C4_PacketProtocol</td>
<td>18,429</td>
</tr>
<tr>
<td>C5_SendDataWaitUntilDone</td>
<td>18,323</td>
</tr>
<tr>
<td>C6_SendDataBackground</td>
<td>18,269</td>
</tr>
<tr>
<td>USB MSC</td>
<td></td>
</tr>
<tr>
<td>M2_SDCardReader</td>
<td>19,128</td>
</tr>
<tr>
<td>M3_MultipleLUN</td>
<td>20,403</td>
</tr>
<tr>
<td>M4_DoubleBuffering</td>
<td>19,053</td>
</tr>
<tr>
<td>M5_CDROM</td>
<td>23,055</td>
</tr>
<tr>
<td>USB HID</td>
<td></td>
</tr>
<tr>
<td>H0_SimpleSend</td>
<td>18,777</td>
</tr>
<tr>
<td>H1_LedOnOff</td>
<td>18,850</td>
</tr>
<tr>
<td>H2_ReceiveData</td>
<td>18,540</td>
</tr>
<tr>
<td>H3_EchoToHost</td>
<td>18,404</td>
</tr>
<tr>
<td>H4_PacketProtocol</td>
<td>18,476</td>
</tr>
<tr>
<td>H5_SendDataWaitUntilDone</td>
<td>18,491</td>
</tr>
<tr>
<td>H6_SendDataBackground</td>
<td>18,432</td>
</tr>
<tr>
<td>H8_Keyboard</td>
<td>18,876</td>
</tr>
<tr>
<td>H7_Mouse</td>
<td>18,649</td>
</tr>
<tr>
<td>H9_Remote_Wakeup</td>
<td>18,865</td>
</tr>
<tr>
<td>H10_ReceiveData_EncryptDecrypt</td>
<td>18,911</td>
</tr>
<tr>
<td>H11_LedOnOff_EncryptDecrypt</td>
<td>18,724</td>
</tr>
<tr>
<td>Test Set</td>
<td></td>
</tr>
<tr>
<td>USB</td>
<td></td>
</tr>
<tr>
<td>CC1_term2term</td>
<td>18,946</td>
</tr>
<tr>
<td>CH1_term2hidDemo</td>
<td>18,640</td>
</tr>
<tr>
<td>CHM1_term2hidDemo2LUN</td>
<td>24,287</td>
</tr>
<tr>
<td>HH1_hidDemo2hidDemo</td>
<td>18,647</td>
</tr>
<tr>
<td>BadUSB_Firmware</td>
<td>1,696</td>
</tr>
<tr>
<td>EzHID_Firmware</td>
<td>8,683</td>
</tr>
<tr>
<td>Bluetooth</td>
<td></td>
</tr>
<tr>
<td>spp_counter</td>
<td>47,397</td>
</tr>
<tr>
<td>spp_flowcontrol</td>
<td>47,389</td>
</tr>
<tr>
<td>sdp_general_query</td>
<td>47,378</td>
</tr>
<tr>
<td>sdp_rfcomm_query</td>
<td>47,368</td>
</tr>
</tbody>
</table>

TABLE 1: The list of firmware examples analyzed and their lines of code generated from David A. Wheeler’s SLOCCount tool.
Fig. 10: Performance of heuristics compared to BSE. The figure shows average number of constraints recovered at different time intervals for MSP430 USB firmware.

Fig. 11: Performance of different heuristics compared to baseline symbolic execution (BSE) for USB. $C_{\text{window}}$ is 5 for the heuristics.

Fig. 12: Performance of different heuristics compared to baseline symbolic execution (BSE) for Bluetooth HCI. $C_{\text{window}}$ is 5 for the heuristics.

<table>
<thead>
<tr>
<th>Type</th>
<th>FW#</th>
<th>CBE Min.</th>
<th>CBE Max.</th>
<th>CBE Avg.</th>
<th>BSE Min.</th>
<th>BSE Max.</th>
<th>BSE Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>7</td>
<td>37</td>
<td>42</td>
<td>40</td>
<td>18</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>HID</td>
<td>12</td>
<td>34</td>
<td>41</td>
<td>38</td>
<td>19</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>MSC</td>
<td>4</td>
<td>34</td>
<td>39</td>
<td>37</td>
<td>27</td>
<td>30</td>
<td>29</td>
</tr>
</tbody>
</table>

TABLE 2: USB protocol constraints extracted from the training firmware set by Constraint Based Extraction (CBE) and Baseline Symbolic Execution (BSE). CBE corresponds to $C_{\text{Type}} = \text{constraint}, C_{\text{Scope}} = \text{perPath}$. Since CBE achieved the best extraction results over time amongst our heuristics, we compare the results against BSE here. The diversity of the firmware set in terms of functionality can be seen from this table.

packet_type, event_type, HCI command, and the L2CAP signaling command code (sig_cmd_code). To extract the protocol model from a given firmware, first we manually find the addresses of these protocol relevant fields within that firmware and pass this information as an input to Algorithm 1 so that the relevant memory locations can be tracked throughout the symbolic execution. For USB, the protocol fields had the same name in the firmware as in the specification (lines 24, 31, and 32 in Figure 2). In Bluetooth firmware, the protocol fields corresponded to the various elements of an array named hci_packet_with_pre_buffer (line 20, Figure 4). So, the protocol fields could be easily mapped since the hci_packet_with_pre_buffer array was always used to pass data from the HCI layer to the L2CAP layer and the conditional statements in the firmware mostly used different elements of this array.
We have used three configuration parameters, \( C_{\text{cov}} \), \( C_{\text{scope}} \), and \( C_{\text{window}} \), to control the exploration during constraint extraction. \( C_{\text{cov}} \) denotes the coverage criteria, which can be code, constraint or field coverage. Code coverage is used to prioritize paths that cover new instructions, constraint coverage prioritizes paths that lead to more unique condition extraction, and field coverage prioritizes paths based on the visibility of multiple fields in a certain path.

Depending on the firmware’s implementation, each of these criteria has unique advantages. For a small firmware with few branches just doing code based coverage may be enough to extract information. For a large firmware with many branches, on the other hand, constraint coverage is more suitable, since it produced better yield in a short amount of time by prioritizing paths that provide new constraints. Field coverage showed faster extraction of constraints deeper in the code which involved multiple protocol fields.

\( C_{\text{scope}} \) denotes the scope for which the coverage is evaluated and can be perPath or perGroup. \( C_{\text{window}} \) specifies the granularity of the window for executing the paths in the frontier set. It is given in terms of the number of basic blocks. \( C_{\text{scope}} \) combined with \( C_{\text{window}} \) determines the frequency of action taken based on \( C_{\text{cov}} \). In a perPath scoping, the frontier set gets updated every time new coverage is achieved. However, for perGroup scoping, updating the frontier set is delayed until all paths have executed \( C_{\text{window}} \) number of blocks. There are six possible configurations based on the values of \( C_{\text{cov}} \) and \( C_{\text{scope}} \). We have combined these with window sizes from one through ten. We observed that too small a window size does not let paths cover much and too large a window size lets all paths cover something new. We found window sizes of 4-7 performs better in all cases than others. The graph in Figure 11 shows results for \( C_{\text{window}} = 5 \). Due to space restrictions we will discuss the best performing combinations for each \( C_{\text{cov}} \) here. We should note that all configurations discussed below could extract on average 33 unique constraints for USB by the end of the analysis window, which was set to 900 seconds. For Bluetooth HCI and L2CAP layers the average number of unique constraints extracted are 59 and 29, respectively, within an analysis window of 900 seconds.

1) \( C_{\text{cov}} = \text{code}, C_{\text{scope}} = \text{perGroup} \): This configuration is the closest to BSE. Focusing on paths with better code coverage reduced the number of paths to execute and led to more constraint extraction. As Figure 10 shows, this configuration extracts the least number of unique constraints compared to the other two configurations in the early phases of the analysis. Using perPath scope with code coverage achieves worse performance as only one path is chosen until it stops achieving new coverage. So, it may cause divergence from the protocol relevant parts of the code.

2) \( C_{\text{cov}} = \text{field}, C_{\text{scope}} = \text{perPath} \): This configuration performs slightly better than configuration (1) as it extracts more constraints in the early phases as shown in Figure 10. Protocol field based extraction performs better than instruction coverage because it better relates to our goal. We choose paths based on finding new protocol fields. If a path has seen at least one new protocol field we expand on that path and stop executing it when it does not reach any new protocol field in the given window. Since the scope is per path, the window size only comes into play when there is no field coverage. This configuration performed the best for the Bluetooth L2CAP layer, which has a higher variation of constraints on different fields. So focusing on paths with different fields gets the constraints faster. For the Bluetooth HCI layer and the USB protocol it is beneficial to focus on the unique constraints found on the same field than variety of fields, that is why this configuration does not have the best results there.

3) \( C_{\text{cov}} = \text{constraint}, C_{\text{scope}} = \text{perPath} \): In general, this configuration produces the best result as it extracts more constraints than the other configurations in a given amount of time as shown in Figures 10 - 12 for USB and Bluetooth HCI. This approach is most related to the goal of extraction. Since we want to extract as many unique constraints as possible from a firmware, we choose paths that provide us with constraints that have not been seen yet. Only the constraints on protocol fields are considered while evaluating paths. The perGroup scope combined with constraint coverage performs slightly worse than (3) as it executes more paths in the same time which delays extraction compared to perPath scope and performs much better than all other combinations.

6.1.1 Ground Truth Evaluation

We extracted 58 unique constraints in total across three different USB classes; see Table 2 for the number of constraints for each class of firmware. It is important to note that the extracted constraints come in different formats. For example, two constraints on bmRequestType are ((bmRequestType &
TABLE 3: Unique values contained in the equality constraints for each protocol field in the extracted model from the training set for both USB and Bluetooth (BT). Equality constraints directly compare the data read from a memory location to a specific value.

<table>
<thead>
<tr>
<th>Type</th>
<th>USB 2.0</th>
<th>CDC 1.2</th>
<th>HID 1.11</th>
<th>MSC 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>HID</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>MSC</td>
<td>5</td>
<td>11</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

BT packet_Type event_Type sig_cmd_code cmd

<table>
<thead>
<tr>
<th>Field</th>
<th>BT 4.0</th>
<th>BTStack</th>
<th>Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>packet_type</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>command</td>
<td>137</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>event_type</td>
<td>66</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>sig_cmd_code</td>
<td>23</td>
<td>16</td>
<td>5</td>
</tr>
</tbody>
</table>

TABLE 4: bmRequestType and bRequest value coverage comparing with different USB specifications based on the 23 firmware in the training set.

<table>
<thead>
<tr>
<th>Type</th>
<th>USB 2.0</th>
<th>CDC 1.2</th>
<th>HID 1.11</th>
<th>MSC 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>6/6</td>
<td>2/2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HID</td>
<td>5/6</td>
<td>–</td>
<td>3/4</td>
<td>–</td>
</tr>
<tr>
<td>MSC</td>
<td>5/6</td>
<td>–</td>
<td>–</td>
<td>2/2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>USB 2.0</th>
<th>CDC 1.2</th>
<th>HID 1.11</th>
<th>MSC 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>9/11</td>
<td>8/39</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HID</td>
<td>9/11</td>
<td>–</td>
<td>7/8</td>
<td>–</td>
</tr>
<tr>
<td>MSC</td>
<td>9/11</td>
<td>–</td>
<td>–</td>
<td>3/5</td>
</tr>
</tbody>
</table>

TABLE 5: Value coverage for Bluetooth fields in the training set. BT 4.0 and BTStack lists the number of possible values for the field in Bluetooth 4.0 specification and the number of values listed in BTStack core. Extracted shows the number of values extracted by ProXray.

<table>
<thead>
<tr>
<th>Field</th>
<th>BT 4.0</th>
<th>BTStack</th>
<th>Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>packet_type</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>command</td>
<td>137</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>event_type</td>
<td>66</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>sig_cmd_code</td>
<td>23</td>
<td>16</td>
<td>5</td>
</tr>
</tbody>
</table>

Fig. 14: bmRequestType value coverage comparing with different USB specifications for each firmware within the training set.

128 ≠ 0) and (bmRequestType == 161). Both constraints appear on the same path and the latter satisfies the former. The first is used to identify the direction of the request i.e., from host to device or device to host. The second is used to identify the specific request. We examine all extracted constraints to find all possible values for each protocol field contained, and list the number of unique values for each field in Table 3. Both bmRequestType and wIndex share the same possible values among all these classes. For wIndex, the variation is limited. In most cases, the default value is zero. For bmRequestTypes, which tells the data transfer direction, recipient, etc, each value can also be reused by different USB requests. This means that given enough USB requests, it is possible to enumerate every bmRequestType permitted for this USB class. bRequest shows minor differences among different classes due to the class-specific USB requests. Other fields vary since they heavily rely on the semantics of the USB request.

We then look into all the values found in the constraints for bmRequestType and bRequest. We first manually extract all possible values explicitly listed in different USB specifications for these fields. We compare the values found in the constraints with the ones included in the specifications. The coverage for these fields is shown in Table 4. For bmRequestType, all three classes of firmware show a minimum 83% and even 100% coverage on the standard USB 2.0. This is expected since different USB classes still follow the enumeration procedure using the standard USB requests. We also find 75% and 100% coverages on CDC 1.2, HID 1.11, and MSC 1.3 class-specific protocols accordingly based on different classes of the firmware.

For bRequest, all these three classes of firmware show an 82% coverage on the standard USB requests defined by USB 2.0. The only missing ones are “SET_DESCRIPTOR” and “SYNCH_FRAME”, which are optional or only used by audio streaming devices. CDC firmware shows the lowest coverage comparing to the CDC 1.2 specification due to a large number of requests defined by its four different sub specifications, including Public Switched Telephone Network (PSTN), Integrated Services Digital Network (ISDN), Ethernet Control Model (ECM), and Abstract Control Model (ACM). HID firmware demonstrates 87.5% coverage comparing to the HID 1.11 specification. Both the 2 requests defined by the MSC 1.3 specification but not covered by our MSC firmware, are for Lockable Mass Storage device, a different kind of MSC devices.

We further look into each firmware sample used during the model extraction, and investigate its bmRequestType and bRequest coverage. As shown in Figure 14, every firmware in our training set has over 60% coverage of the USB 2.0 specification and over 70% coverage of other USB class specifications respectively on bmRequestType. For bRequest shown in Figure 15, most firmware except the CDC class have over 50% coverage of different USB specifications. Again, the coverage limitation of the CDC class is due to its sub protocol variations defined by the spec. Note that we did not include wValue and wIndex in our ground truth study; because although important, these fields depend on bRequest rather than being self-contained.

Due to the complexity of the Bluetooth HCI and L2CAP layers, we focus on protocols fields packet_type, command,
and \textit{event\_type} within HCI and \textit{sig\_cmd\_code} within L2CAP. As shown in Table 5 for \textit{packet\_type} within HCI, we have successfully recovered 3 types out of 4. The only missing one is the “Command” type. We realize all the firmware examples we used do not build HCI Command packets directly. Instead, the BTStack lower level takes care of sending these packets automatically during the initialization. This is further proved by the recovery of \textit{command} protocol fields within HCI Command packet. While the Bluetooth spec defines over 100 different commands, the BTStack only includes 15 of them and we are able to recover 12 with 80% coverage. For \textit{event\_type} within HCI Event packets, the BTStack again only covers 40 of 60 defined by the Bluetooth spec. The coverage of our extraction is 67.6% against the BTStack implementation. For the \textit{sig\_cmd\_code} within L2CAP signaling packets, we have recovered 5 different values out of 16 defined by the BTStack implementation. A further investigation reveals a classic path explosion issue that prevents us from reaching another 5 different values.

In summary, although not 100% coverage for all possible values of each field defined by the specs, our constraints did a good job to cover the most common values of the most important fields (e.g., \textit{bRequest\_Type} and \textit{command}), which help pinpoint the usage of the USB and Bluetooth protocols and potential functionality with a high confidence. Our study of the training set proves that each firmware does provide a lot of information about different specifications, which bases our model extraction methodology using firmware. Since we extracted the constraints from the most commonly available firmware images, other firmware images that contain less common functionality, e.g., PSTN, can help us to improve our constraint set. But those are relatively rare and less used. We also notice the limitations of firmware built upon certain protocol stack implementations, e.g., the BTStack implementation. In general, with the help of more diverse firmware collection, the coverage of our constraints is expected to get better.

### 6.2 Protocol Field Discovery

In this subsection, we answer the following research question: \textit{Can we use the extracted constraint model to discover the program variables that correspond to the protocol fields in unknown firmware?}. One of our main contributions in this paper is automatic discovery of the protocol fields within a
firmware address space. For an arbitrary firmware it may be
difficult to identify the memory locations or variables that
implement the protocol fields. The source code may not be
available at all. Using Algorithm 5 explained in Section 4.2,
we could discover the protocol fields in our benchmarks
with high precision. We compute the precision of matching
a protocol field pf in a given firmware as

$$
100 \times \frac{1}{\mathcal{M}_{best}(pf)}
$$

where rc denotes the actual variable that implements pf
and \( \mathcal{M}_{best} \) is the output of Algorithm [5].

Tables 6 shows the field discovery data for the four
composite firmware from the MSP430 SDK for USB. For
USB, we could discover all fields except the bRequest field as
sometimes another field, wLength, yielded a higher matching
frequency. However, we achieve 100% precision in the
discovery of the remaining protocol fields.

Table 7 shows the field discovery data for the Intel 8051
firmware. Compared to MSP430 benchmarks, we achieve
less precision for protocol field discovery. However, we are
able to discover one protocol field, wValue, for each Intel
8051 firmware with 100% precision and an additional pro-
tocol field, bRequest, for BadUSB firmware with a precision
of 50%. In case of EzHID firmware the first 3 fields were
mapped with 100% accuracy. We found that in all of the 0%
precision cases, such as wLength for both, the actual field
was matched as a candidate. However, the correct mapping
was not selected as another variable that corresponded to
a different field (wValue for BadUSB, wValue, bRequest
for EzHID) had a higher score and prevented the correct
candidate from being included in the final set.

We also evaluated field discovery for the firmware in
the training set by excluding the firmware under analysis
from the model extraction phase and by using the constraint
model extracted from the remaining 22 firmware. Table 7
shows that we achieve similar precision values compared to
those for the composite MSP430 firmware.

Table 8 shows the field discovery data for the four testing
firmware for Bluetooth. In the case of Bluetooth, we could
discover both the HCI and the L2CAP fields properly. The
reason for lower precision for packet_type is due to the lack
of variety in the values of those fields. For example, the
set of values for packet_type has a lot common with the set of
values for event_type. So, when a firmware uses the common
values the mapping precision is lower. However, due to the
high variety of values for event_type the mapping has 100%
precision every time.

In terms of recall, we achieve 100% recall for the Blue-
tooth protocol. For the USB protocol, we achieved an aver-
TABLE 10: Comparison of PGSE and BSE for Bluetooth firmware w.r.t. time to target and the number of paths at the time of reaching the target. - denotes not reaching the target within a timeout of 2 hours.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Query (Q)</th>
<th>Time (secs)</th>
<th>#Paths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PGSE</td>
<td>BSE</td>
</tr>
<tr>
<td>T1</td>
<td>(packet_type0 == 0x04) &amp; (event_type0 == 0x04)</td>
<td>173.315</td>
<td>174.571</td>
</tr>
<tr>
<td>T2</td>
<td>C(T1) &amp; (packet_type1 == 0x04) &amp; (event_type1 == 0x03)</td>
<td>254.991</td>
<td>-</td>
</tr>
<tr>
<td>T3</td>
<td>C(T2) &amp; (packet_type2 == 0x04) &amp; (event_type2 == 0x0E)</td>
<td>361.009</td>
<td>-</td>
</tr>
<tr>
<td>T4</td>
<td>C(T3) &amp; (packet_type3 == 0x03)</td>
<td>399.706</td>
<td>-</td>
</tr>
<tr>
<td>T5</td>
<td>C(T4) &amp; (code == 0x02)</td>
<td>412.259</td>
<td>-</td>
</tr>
<tr>
<td>T6</td>
<td>C(T4) &amp; (code == 0x04)</td>
<td>414.205</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE 11: Constraints used for Protocol Guided Symbolic Execution of MSP430 firmware.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Query (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>(bmRequestType &amp; 0x80) \neq 0</td>
</tr>
<tr>
<td>T2</td>
<td>C(T1 &amp; bRequest == 0x06)</td>
</tr>
<tr>
<td>T3</td>
<td>C(T2 &amp; (wValue == 0x01))</td>
</tr>
<tr>
<td>T4</td>
<td>C(T3 &amp; (wIndex == 0))</td>
</tr>
<tr>
<td>T5</td>
<td>C(T4 &amp; (wLength == 0))</td>
</tr>
</tbody>
</table>

age recall of 95% for the MSP430 USB firmware, a recall of 40% for BadUSB, and a recall of 60% for EzHID. The missed fields were not selected as the best matches although they appeared in the candidate set of mappings.

6.3 Protocol Guided Execution

Based on our extracted model, we can increase the efficiency of a symbolic execution engine in terms of exploring the required paths/code locations in a goal based execution, where the goal, or the query as shown in Figure 1, is described in terms of a generic protocol constraint. This is different from preconditioned symbolic execution, where the constraint is expressed in terms of the variables/memory locations of the system under test. Our field mapping phase automatically discovers potential mappings and rewrites the constraints of the query based on these potential mappings. So this phase involves the field mapping phase of Section 5.2 and uses the same experimental setup for USB that uses the 23 MSP430 USB firmware as the training set and six firmware from both MSP430 and Intel 8051 architecture as the testing set. Similarly for Bluetooth it has the same training and testing set as earlier.

6.3.1 Target Finding

In this subsection, we answer the following research question: How effective is protocol-guided symbolic execution in driving symbolic execution to the protocol relevant targets in unknown firmware?. To achieve this, we identified several protocol relevant targets in our benchmarks. For MSP430 USB firmware, we have identified protocol relevant targets T1-T5 shown in Table 11. As seen from Figure 17, we have achieved great reduction in the number of paths that are relevant to the protocol functionality of interest. Our approach utilizes an upper bound of 73.8x speedup (T5 in C1) as shown in Figure 16. Our guided execution could reach T5 in a minimum of 1.106s (9.883s minimum for BSE) to a maximum of 2.624s (121.866s maximum for BSE). The number of paths in this case ranged from a minimum of 36 (336 for BSE) to maximum 1134 for BSE). In case of timing for T1 there are a few cases where PGSE times are marginally higher than BSE. It is because of the depth of the code point corresponding to T1. Since the code location is not that deep within execution BSE and PGSE have similar results. The deeper the code location gets with respect to execution the better PGSE performs by eliminating irrelevant code. Table 12 and Table 13 shows the constraint based targets for Intel 8051 test firmware BadUSB and EzHID respectively. It took 126 paths in 26.63s to reach T8 from Table 12 compared to 236 paths in 32.852s for BSE. From Table 11 it is evident that BadUSB_Firmware has much lower number of lines of code compared to the MSP430 firmware. Smaller code is easier for BSE to explore compared to exploring code with a high line count and more conditional paths. In case of EzHID firmware we do not see much difference between PGSE and BSE. It took 67 paths in 4.78s to reach T1 from Table 13 compared to 86 paths in 6.98s for BSE. Figure 18 and Figure 19 shows the performance of PGSE compared to BSE for the Intel 8051 test firmware. In almost all cases, PGSE has shown improvement over BSE both in terms of number of paths and time to reach targets.

For the Bluetooth HCI and L2CAP layers the performance of PGSE is far superior from BSE. Table 10 shows the targets for Bluetooth. T1 - T4 are in the HCI layer and T5 - T6 are in the L2CAP layer. event_type0 and event_type1 represent the value of event type on 1st and 2nd iteration of the main event loop. To reach the targets in the L2CAP layer, a specific sequence of events needs to occur in the HCI layer. The Bluetooth firmware code contains too many branches for BSE to achieve this sequence. For example, there are 27 cases for only event_type. Also, two types of interrupts get fired on every iteration of the main event loop. As a result, BSE could not reach any targets beyond T1. Since T1 is not that deep in the code and does not require a sequence of events, the performance of both PGSE and BSE is similar. As the execution went deeper in the code BSE could not keep up with PGSE. At the end of the two hour timeout BSE had spawned 52,289 paths but not the required one. PGSE was able to reach the all the targets with only 404 paths in 414.205 seconds.

6.3.2 Functionality Matching

In this subsection, we answer the following research question: How precise is constraint model based functional classification on unknown firmware?. Thus, as an additional capability of our protocol model extraction approach, we tried to
of attribute and value pairs. The protocol model extraction phase of our approach requires some domain knowledge. Specifically, the analyst should understand whether the protocol involves multiple layers and the format of the packet types for each layer. In our study, we consulted the core specifications of both protocols to achieve this knowledge.

This level of understanding of the protocol will guide the analyst in getting a better understanding of the corresponding SDKs. The complexity of the implementation is proportional to the complexity of the protocol. In our case, understanding the BTStack framework that implements the Bluetooth protocol took more time than understanding the MSP430 USB SDK. We should acknowledge that the documentation plays a significant role and BTStack’s documentation was very helpful in this regard.

The most important stage in understanding a protocol implementation is to identify the data structures used for the packet types of the protocol. In our experience with these two implementations, we found out that the developers try to use the terminology from the protocol as identifiers in their code. For instance, MSP430 SDK uses the identifier tSetupPacket to represent the Setup packet for the USB protocol as shown in Figure 2 (line 24). Similarly, BTStack framework uses the identifier hci_packet to represent the HCI packet for the Bluetooth protocol as shown in Figure 4 (line 20). So, we needed to identify these two data structures for all the training firmware we used in our experiments as they all used these specific data structures defined by the relevant libraries.

Despite the ease of finding the data structures that correspond to the protocol specific packet types, the data-flow in these protocol implementations is typically complicated as can be seen in the sample snippets of code provided in Figures 2 and 4. This is due to pointer arithmetic and using function pointers to implement callbacks for various event types in the protocol. These programming constructs require a precise analysis to avoid false positives. Dynamic symbolic execution provides a precise memory model and proved to be effective for protocol constraint model extraction.

We think that as more microcontroller (MCU) architectures get fully supported by the state-of-the-art symbolic execution engines such as FIE, more code bases can be explored for model extraction and model guided analysis. In this study, we have been restricted to MSP430 and 8051 architectures as these were the only MCUs that were supported by FIE with our extensions from [13]. Another challenge is the proprietary code that comes with the SDKs. As an example, we were not able to use TI’s MSP430 Bluetooth SDK due to having some 3rd party libraries in binary only.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Query (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>(bRequest &amp; 0x40) == 0</td>
</tr>
<tr>
<td>T2</td>
<td>(bRequest &amp; 0x20) == 0</td>
</tr>
<tr>
<td>T3</td>
<td>((bRequest &amp; 0x60) == 0) &amp; (bRequest == 0x05)</td>
</tr>
<tr>
<td>T4</td>
<td>((bRequest &amp; 0x60) == 0) &amp; (bRequest == 0x09)</td>
</tr>
<tr>
<td>T5</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x01)</td>
</tr>
<tr>
<td>T6</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x02)</td>
</tr>
<tr>
<td>T7</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x06)</td>
</tr>
<tr>
<td>T8</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x22)</td>
</tr>
</tbody>
</table>

TABLE 12: Constraints used for Protocol Guided Symbolic Execution of BadUSB firmware.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Query (Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x01)</td>
</tr>
<tr>
<td>T2</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x02)</td>
</tr>
<tr>
<td>T3</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x03)</td>
</tr>
<tr>
<td>T4</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x21)</td>
</tr>
<tr>
<td>T5</td>
<td>(bRequest == 0x06) &amp; (wValue == 0x22)</td>
</tr>
</tbody>
</table>

TABLE 13: Constraints used for Protocol Guided Symbolic Execution of EzHID firmware.

<table>
<thead>
<tr>
<th>Type</th>
<th>Specific Constraints</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>bRequest ∈ {32, 33, 34}</td>
<td>CDC</td>
</tr>
<tr>
<td>HID</td>
<td>bRequest == 2 ∨ wValue ∈ {33, 34}</td>
<td>HID</td>
</tr>
<tr>
<td>MSC</td>
<td>bRequest ∈ {254, 255}</td>
<td>MSC</td>
</tr>
</tbody>
</table>

TABLE 14: USB type specific constraints used in functionality matching.

identify the USB subclasses implemented by a given USB firmware. The intention here is to be able report the types of functionalities a firmware can support. For this purpose, we have first identified the class specific constraints in the model by automatically removing those that appear in more than one subclass and recorded the associated subclass type for each class specific constraint. In our benchmarks, we have come across three USB subclasses: CDC, HID, and MSC.

Table 14 shows CDC, HID, and MSC specific constraints that we extracted from the 23 MSP430 firmware.

Next the protocol field discovery of the test firmware is done in the same process as explained in Algorithm 5. Guided symbolic execution is then used to find functionality by matching the type of class specific constraints shown in Table 14. All paths in the firmware are evaluated against these constraints. If a firmware is found to conform to a specific constraint of any functionality type, we report that the firmware implements that functionality. We have correctly identified the functionality type for every MSP430 and Intel 8051 firmware in our benchmarks using this process including matching all functionalities of the 4 composite firmware. The importance of this can be seen in case of BadUSB firmware. We found that it implements HID functionalities in addition to its reported MSC functionality which indicates that the firmware has a malicious aspect.

6.4 Usability

The applicability of our approach is not limited to the USB and the Bluetooth protocols. It can be applied to any protocol that processes requests that are structured as a set
form. This prevented us from generating the LLVM bitcode for that SDK, which was needed for symbolic execution.

Finally, any mistakes related to the setup of the model extraction phase, e.g., identifying a wrong data structure for the packet type, could lead to the extraction of wrong protocol constraints. A low-quality constraint model would lead to a low-quality field discovery and, hence, to a misleading analysis result during protocol model guided analysis. Another issue could be the implementation mistakes in the SDKs. One way to deal with these issues is to use a variety of SDKs and a diverse set of firmware examples. We think that with better support for diverse set of MCU types, it will be practical to analyze a variety of SDKs and extract high quality protocol models.

7 Discussion

In this paper, we have focused on the USB protocol, and, specifically, the USB (control) requests issued by the host, and the HCI and L2CAP layers of the Bluetooth protocol. It should be sufficient to analyze different device firmware to capture these parts of the USB and the Bluetooth protocols. Our training set was restricted to MSP430 firmware as FIE has been specialized for this microcontroller architecture. Although our evaluation is based on sample firmware implementations from the vendor rather than the USB and Bluetooth firmware in the wild, these non-toy implementations provide a strong base of how the real-world firmware would be implemented.

ProXray also needs the source files of the firmware to map program variables into protocol fields during the training phase. However, access to the source file might be infeasible due to various reasons. It is not uncommon to have only the binary format of firmware available. Fortunately, it is still possible to recover the protocol fields from the binary using static analysis. One way is to look for certain binary patterns from within the binary, e.g., the binary pattern of a USB setup packet. We will include binary parsing/mapping support in our future work.

Threats to Validity: Threats to internal validity include not having a training set that is comprehensive in terms of the USB subclass functionality. However, despite this limitation, our approach could precisely discover the functionality of all USB firmware in the test set. We think that the quality of the extracted model can be improved by enriching the training set. Threats to external validity include application of ProXray to the USB and Bluetooth protocols, which do not involve relational constraints, e.g., $p_1 \text{ROP} p_2$, where $p_1$ and $p_2$ are protocol fields. Although our implementation focuses on constraints that relate some protocol field to some constant value, the implementation of our approach can be easily adapted to protocols that involve relational constraints by incorporating 1) prioritization heuristics that can compare two paths w.r.t. the potential for covering richer relational constraints in the model extraction phase and 2) matching heuristics that consider pairwise associations in the potential mappings for the protocol field discovery phase.

8 Related Work

Protocol Model Learning: Black-box model learning has been used for extracting models of the Session Initiation Protocol [31], the Europay-MasterCard-Visa protocol (EMV) [32], the Transmission Control Protocol (TCP) [33], the Transport Layer Security (TLS) protocol [34], a botnet command and control protocol [35], a smart-card reader [36], and the Secure Shell Protocol (SSH) [37]. White-box model learning has been combined with predicate abstraction and symbolic execution in [38] to systematically consider the input space for the learned protocol model. MACE [39] extracts finite state machines (FSMs) from reference implementations in an incremental way by using the extracted FSM to guide concolic execution and by using guided concolic execution to enrich the model. Our approach focuses on extraction of the protocol constraints rather than the underlying FSM of the protocol, which may not be relevant to protocols such as the USB protocol as the specification explicitly advises the firmware developers to handle any type of request at any time during the communication with the host. Also, we use the extracted models to reverse engineer unknown protocol implementations.

USB Protocol Modeling: In the hardware/software co-design, USB has been modeled as a 8-bit wide channel, where each transfer takes 8 cycles [41]. This bandwidth-and timing-based modeling was designed to ease the partitioning during the co-design, but ignored the semantics of the USB protocol. In the USBFirewall [42] work a DSL is designed to describe the syntax of the USB protocol. This DSL is used to compile into C code using Haskell, generating a “formal” USB packet parser for FreeBSD. The USB Type-C Authentication protocol [43] is formally verified in USBSoK [44] using ProVerif [45]. Unlike previous work, we build a constraint-based model based on standard USB request/response message structure. This model can be readily used by a SAT solver.

Model Extraction: There have been some interest in (semi-)automatically extracting models from software, which include extraction of relationships between features and computational units [46], object models [47], finite-state models [49], [50], [51], dependence models [52], test scenarios conforming to some temporal logic formula defined over a data-flow graph [53], state predicates to guide state exploration of a cooperating model checker [54], a statistical language model [55], framework models for symbolic execution [56], symbolic data models for web applications [57], and behavioral models for USB firmware [58]. Our approach uses symbolic execution to extract protocol relevant constraints from firmware and it uses these constraints to learn a constraint-based model for the protocol of interest. Our approach differs from related work in that the extracted model is not used to improve the analysis of firmware from which the models get extracted, which we refer to as the training set, and is used to analyze unknown firmware that is known to implement the same protocol as the firmware in the training set.

Firmware Analysis: Symbolic execution has been used for firmware analysis in FIE [9], AVATAR [10], Firmalice [11], and FirmUSB [13]. FIE models the reactive nature of firmware via scheduling interrupt functions at various gran-
Symbolic Execution with Pruning/Prioritization: Redundant paths are pruned in [59] based on the program point, the concrete write set, the path condition, and the variables that would be read after that program point. The generational search approach in [60] executes a set of child paths and uses code coverage information for prioritizing their exploration. Structural path coverage is used in [61] to prioritize the paths. In [62], structural path information is enriched with data-flow facts to include only the relevant program locations in path coverage. Dependence analysis is used in [63] to perform pruning based on symbolic path equivalence. The goal of our pruning heuristics is to achieve high coverage of the protocol of interest and, hence, our heuristics consider canonical constraints that involve the protocol fields. Our approach is similar to the generational search presented in [60] in that it also executes a set of child paths and uses coverage information for evaluating their usefulness. However, our approach also considers constraint coverage and prioritizes symbolic paths rather than feasible inputs.

Guided Symbolic Execution: Program dependence information and user provided abstraction strategies are used in [64] to compute equivalence classes of paths that can reach a specific program location. [65] uses dynamically computed relevant slice conditions to explore paths that are relevant to a slicing criteria. Control flow analysis and weakest-precondition computation are combined in [66] to generate tests that exercise code changes due to patches. Interpolation and subsumption checking are used in [67] to prune paths that do not lead to buggy program locations. Approximate weakest preconditions are used in [68] to prevent scheduling of paths that are guaranteed not to violate assertions in multithreaded programs. Abstract reachability graphs computed by a predicate abstraction based model checker are used in [69] to avoid exploration of infeasible paths. The program analysis phase of our approach discovers associations between the protocol fields and the memory locations. This information is used in guiding symbolic execution by prioritizing paths that produce constraints of the protocol model.

In [70] constraint normalization is used to efficiently support satisfiability and model counting queries for string manipulating programs. In our approach, constraint normalization enables efficient computation of constraint coverage and simplifies the formal representation of the protocol.

9 CONCLUSION

We presented a methodology for protocol model guided analysis using symbolic execution. We applied our approach to MSP430 and Intel 8051 firmware in the context of the USB and Bluetooth protocols. We extract a constrained-based model of the protocol using symbolic execution. Our path pruning and prioritization heuristics perform 1.54 and 1.76 times better than the baseline symbolic execution in terms of the number of unique constraints extracted in the context of the USB and Bluetooth protocols, respectively. Our constraint-based protocol model enables analysis of new firmware without prior knowledge of how it uses the protocol. Our protocol field discovery can find a mapping for each of the protocol fields and achieves high precision for the most significant fields of both protocols. Our protocol guided symbolic execution achieves up to 73.8 times speedup for unknown USB firmware and more than an order of magnitude speedup for unknown Bluetooth firmware in reaching the parts of the code that are relevant to a given protocol constraint. In future work we will explore other protocols as well as firmware for other microcontroller architectures.

ACKNOWLEDGMENTS

This work is supported by the US National Science Foundation under grant CNS-1815883 and by the Semiconductor Research Corporation.

REFERENCES


Farhaan Fowze received his B.Sc. from Bangladesh University of Engineering and Technology (BUET) in 2012. He is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of Florida. His research interests include model extraction, binary analysis, and program analysis. Contact at farhaan104@ufl.edu.

Tuba Yavuz Tuba Yavuz is an Assistant Professor at the Electrical and Computer Engineering Department of University of Florida. Her research interests include model extraction, model checking, and program analysis with applications to cyber-security. Yavuz received a Ph.D. in computer science from the University of California at Santa Barbara. She has been on the program committee of ICSE’19 and Usenix Security’19. She is a member of the IEEE and ACM. Contact her at tuba@ece.ufl.edu address.