A Case for Remote Attestation in Programmable Dataplanes

Nik Sultana  
Illinois Institute of Technology  
Deborah Shands  
SRI International  
Vinod Yegneswaran  
SRI International

ABSTRACT

Programmability is a double-edged sword. It can better tailor solutions to problems, optimize resource use, and inexpensively patch deployed equipment. But programmability can also be abused to undermine the security of hardware and that of its unwitting users. Remote Attestation (RA) is a class of techniques to provide integrity assurance to remote users of resources such as hardware, OSs and applications. It is used to establish well-defined trust relationships among mutually distrustful principals who provide, use or delegate remote resources. RA could benefit, for example, tenants of a datacenter or users of IoT equipment such as health monitors.

This position paper considers how RA can be used to enable dynamic assessments of network security characteristics through automated generation, collection, and evaluation of rigorous evidence of trustworthiness. We introduce a set of use cases, sketch how the Copland and NetKAT languages can be combined and extended to make network-aware attestation policies, and propose an extension of P4-programmable hardware to enforce this mechanism in the network.

CCS CONCEPTS

• Networks → Network security;

KEYWORDS

Remote Attestation, Programmable Networking

1 INTRODUCTION

Programmability is a double-edged sword. It can better tailor solutions to problems, optimize resource use, and inexpensively patch deployed equipment. But programmability can also be abused to undermine the security of hardware and that of its unwitting users.

The abuse of programmable network equipment was central to the “Athens Affair” [20, 24], a cyberattack on a cellular network operator that “targeted the conversations of specific, highly placed government and military officials.” This enabled eavesdropping on the private communications of the prime minister of Greece and least 100 other high-ranking officials. The attacker patched software running on programmable network equipment to duplicate digitized voice data streams associated with a specific list of phone numbers and direct the duplicate streams to other cellular phones, enabling eavesdropping. The rogue software patch activated existing, but unused, lawful-intercept functions of the equipment. The operators of the network were unaware that their equipment had been subverted. The attack came to light only by accident, when an upgrade resulted in a noticeable malfunction. The pervasive programmability of modern wired and wireless networks is risky without rigorous assurance of exactly which programs are running on the network equipment.

Remote Attestation (RA) [9] could have helped the network operator targeted in the Athens Affair to detect that rogue software was running on its equipment. RA is a class of techniques for validating remote resources, such as the remote execution of a program and the hardware on which it is executing. RA can verify a claim that a specific program is running without verifying claims about the program’s correctness.

In this paper, we sketch a path to tackle both the specification and the mechanism for RA on programmable dataplanes. We describe a specification approach that combines the state-of-the-art approaches of Copland [14] and NetKAT [2] which are used to reason about RA and SDN respectively. For the mechanism, we describe an extension of the Protocol Independent Switch Architecture (PISA) [7] that can participate in RA protocols by attesting its code as sketched in Fig. 2.
The security of programmable networking equipment has not yet caught up with its flexibility, and research is needed to address the gap. Unfortunately, the security risks introduced by programmability can undermine the benefits of using programmable network hardware. For example, there has been excellent progress on leveraging programmable networking for monitoring [15, 16, 25], but without RA an adversary can replace a monitoring program with one that produces false readings, perhaps as part of a Denial-of-Service or Confused Deputy attack on the network being monitored.

The networking community needs fundamental techniques for using RA on programmable network hardware. Theoretical and practical RA techniques developed for other targets offer helpful stepping stones. For example, practical RA techniques for hosts [5, 10] can be reused for programmable network hardware (e.g., for securing the boot sequence and providing secure signing and verification of data.) Similarly, abstractions [9] and reasoning techniques about confinement [22] can be repurposed for programmable dataplanes.

Because existing RA techniques abstract away the network and focus on peers, new techniques are needed to enable dynamic, automated assessment of security-critical characteristics of networks implemented on programmable hardware. The central hypothesis in this position paper is that RA can be used to enable dynamic assessments of network security characteristics through automated generation, collection, and evaluation of rigorous evidence of trustworthiness.

A realization of the hypothesis is sketched in the remainder of the paper which introduces use cases, describes how NetKAT and Copland can be combined to express the use cases, and discusses the extension of a P4 [6] switch to support the RA. Related work is mentioned throughout the paper.

2 MOTIVATING USE CASES

We describe motivating practical situations in which having RA-capable programmable dataplanes improves the security of network users and operators.

UC1: Configuration Assurance. Using the wrong dataplane program can have serious consequences for the operator and its users. For example, using the wrong firewall, forwarding table, or load-balancer could degrade the network’s security and performance. In this use case, RA protects against untrusted or unwanted dataplane programs that might have been mistakenly or deliberately swapped for the intended version. In the “Athens Affair” from §1, the equipment’s software was deliberately modified to exfiltrate traffic, but it could also have been modified to malfunction and cause a denial-of-service.

Using RA in this use case would involve producing evidence at some granularity and frequency (at most, per hop and per packet) along the paths of a network flow, and giving peers a signed and suitably redacted form of that evidence. For example, the evidence for a packet $p$ could indicate that $p$ reached switch $S_1$ on a specific network port, was processed by firewall_v5.p4 and forwarded to $S_2$ which was running ACL_v3.p4, that then forwarded $p$ to a DPI appliance that forwarded it back to $S_2$, that forwarded the packet to the current node.

UC2: (Authentication) Path Evidence as a Security Factor. The evidence gathered along a path can be used as a factor for authentication [1]. For example, a user that forgets their password or connects from a new device could be permitted limited access to a resource if they can prove that they are connecting from their home via an acceptable network path. This evidence can be used to weakly authenticate different peers and complement other authentication methods.

In addition to chaining together evidence of the forwarding decisions made on programmable dataplanes, this evidence can attest to other packet processing or filtering. Our confidence in a path can be bolstered if the path has specific features (e.g., if it crosses a specific series of firewalls or appliances as discussed in UC1.)

UC3: (Authorization) Path Evidence as a Tag. In UC2, evidence was used for authentication. Evidence can also be used in authorization decisions that affect intra- or inter-domain

---

1Instead of revealing their actual serial number, switches could be assigned a per-user pseudonym by the operator.

2Programs can also be assigned pseudonyms that can be lifted by an auditor’s request or court order.
Remote Attestation in Programmable Dataplanes

The decision to forward packets can depend on whether those packets have been processed by a set of appliances, such as an IDS, firewall, and scrubber. Packets can be treated differently based on evidence that they have never left a particular segment of the network. RA can provide an evidential basis to use cases like those described for FlowTags [11].

Path evidence could be used for DDoS mitigation: while under attack, a network could drop traffic for which it lacks path-based evidence. The next section describes caching of evidence and varying the level of detail and sampling frequency of evidence to lower the overhead of providing and consuming evidence.

UC4: (Auditing) Evidence as Documentation. This use case has two complementary sub-cases: (A): Programmable dataplanes can run filters that match packet types and fields [8]. These features can be used to characterize malware communications with command-and-control nodes [17], which in turn can fingerprint the presence of malware. This can be captured as evidence that is used to justify other actions, such as applying for a court order to deactivate that malware [12]. (B): The subsequent action to deactivate the malware can also be documented in a similar way (i.e., as an appraisable set of network interactions) and stored for later use as evidence to show the limited and focused action that was taken to deactivate the malware, to prove compliance with the authorizing court order. Thus RA can be used to provide evidence on integrity, transparency, and compliance of network-related activities to third-parties at a later time [26].

A similar audit trail can be compiled as evidence when gathering data from sensors [13], such as those for atmospheric radiation sampling, as evidence of the data’s provenance and to help ensure that the sensor traffic has not been spoofed.

UC5: Cross-Referenced Attestation. Evidence from host-based and network-based attestation could be composed together to offer a more complete picture of how host-based processes interacted with the network, and how the network processed their traffic. This can help detect and stop exfiltration attacks by checking whether outward traffic patterns have been authorized by an unmodified application.

This composed evidence could also be used to show that the host that produced a specific flow was running a specific version of a network stack or protocol implementation. This enables the enforcement of policies that are sensitive to the network behavior of software. For example, TLS packets that were produced by a verified implementation [3] could be allowed to leave the network, while packets produced by un-verified implementations are blocked.

Another application of this use case involves trusted redaction of evidence for compliance certification of information processing within a cloud environment: path evidence could be processed to redact details sensitive to the enterprise customer before giving the redacted evidence to a compliance officer. By using host-based RA, the customer can meet regulatory compliance obligations without disclosing unnecessary, sensitive information to the regulator.

3 THREAT MODEL

We assume that evidence-producing hardware components (e.g., those that initialize a chip or generate a digital signature) are trustworthy: they are correctly designed and manufactured to generate tamper-evident evidence. Larger components or products into which trustworthy components are integrated (e.g., switches, NICs) are not assumed to be trustworthy. Adversaries may attempt to exploit insecure intermediate nodes as well as limitations on hardware resources to potentially compromise path attestation. An adversary may also perform supply-chain or organizational insider attacks to actively interfere with hardware, software, and staff. RA leverages outputs from trusted components to derive security guarantees across distributed systems that are deployed on third-party nodes assembled by untrusted manufacturers and run by untrusted operators (and their employees).

The technique described in this paper enables verification of the integrity of dataplane programs and their state. This does not, by itself, ensure confidentiality, integrity, and availability of data processed by a programmable dataplane, but it supports improved confidence in the dataplane programs that are processing data. Our technique does not overcome physical insecurity of a network device. For example, it does not protect against a malicious insider installing a passive tap to siphon network traffic, or installing an intermediary device to introduce network traffic between two network elements. RA complements other security techniques, such as physical access control and network protocols for end-to-end security, such as TLS and IPsec.

4 REMOTE ATTESTATION

Fig. 1 shows the main principals in RA. The Relying Party (RP) is the user of a remote program for which Claims (1) about its execution are met by Evidence (2) produced by the Attester on behalf of the platform executing that program. The RP presents this Evidence to an Appraiser (3) (sometimes referred to as the Verifier) which verifies the evidence to produce an Attestation Result (4). More details can be found in specifications such as RATs [4, §7].

4.1 Reasoning about Network-aware RA

To use RA to reason about network communication of multiple parties over time requires a language that enables us to describe how networking equipment is to generate and process attestation evidence.
The language needs primitives to describe principals and to chain together processing steps across the network. To work in general network settings, the language needs some special primitives. The language must enable us to (Prim1) abstract over paths, since paths might not be knowable in advance. To reason about abstract paths, the language must enable us to (Prim2) abstract over places, since the identities of intermediate hops along a path might not be known to us. The language must also enable us to (Prim3) reason about reachability, and predicate a policy on a collector of evidence being reachable by producers of evidence.

The next section describes Copland, an existing language for reasoning about RA. Later we describe our proposal to extend Copland to provide primitives (Prim1)-(Prim3).

### 4.2 The Copland Language

Copland separates the specification of RA attestation protocols from the enforcement mechanisms provided by specific hardware features. It has formal semantics [19] and a verified compiler toolchain [18]. The separation between policy and mechanism plays a central role in describing fundamental principles for flexible RA [14] that accommodates the needs of different use cases.

We introduce Copland’s syntax by adapting an example from Rowe et al. [23], in preparation for the language extension described in the next section. This example is like the host-based portion of the “verified TLS implementation” part of use case UC5 in §2: a banking website uses evidence about the client’s browser extensions to check for malware that could steal the client’s credentials. The client runs a bmon process which measures the exts process that represents the client’s browser extensions. The bank also receives measurements from av, an antivirus program running in the client device’s kernelspace.

Expressions in Copland describe measurements done by principals of specific values. Measurements are to be carried out in certain places as described by the expression. The results of measurements can be transformed and sent to other places, in composition with other measurements. Consider the example below:

\[
\begin{align*}
* \text{bank} : & \quad @\text{ks}[\text{av us bmon}] \cdot \text{++} @\text{us}[\text{bmon us exts}] \\
\text{UC5} : & \quad \sim C_1 \bowtie C_2
\end{align*}
\]

In expression (1), the overbraces are not part of Copland syntax, rather, we use them to label two Copland subexpressions \(C_1\) and \(C_2\). Here, \(*R : C\) indicates that the unique outermost principal \(R\) is requesting evidence for expression \(C\), \(R\) is the relying party, and the principals mentioned in \(C\) include attesters and appraisers.

In the banking website example shown in (1), the bank is requesting a compound measurement. Starting with the inner expression \(C_1\) (i.e., “av us bmon”) it means that the antivirus principal av is to measure principal bmon that is running in place us (userspace). The expression \(\bowtie[P]\) means that measurement \(C\) must be carried out in place \(P\)—so the measurement \(C_1\) would be carried out in ks (kernel space). \(C_2\) is used to check that bmon has not been tampered with. \(C_2\) uses bmon to measure the userspace-located exts, to scan for suspicious browser extensions.

Measurements in (1) are composed using \(\bowtie\); this means that both measurements are carried out in parallel (\(\sim\)). The evaluation of a Copland expression takes in evidence that has been accrued so far and transforms it into composite evidence. The values of \(I\) and \(r\) indicate whether evidence accrued so far is passed to each arm of the composition. Symbol “+” means that evidence is passed on, and “−” means that evidence is not passed on. Thus \(C_1 \sim C_2\) is a parallel composition that only allows information to flow back from evaluating \(C_i\).

An active adversary could replace bmon in userspace to always give a positive measurement, even when exts includes malware. Ramsdell et al. [21] describe how an active adversary who has userspace but not kernelspace control can cheat example (1) as follows: it first evaluates \(C_2\) using the corrupted bmon, then it “repairs” bmon in userspace (replacing it with the non-corrupt version), and only then would it allow \(C_1\) to take place. av would then indicate to bank that bmon is authentic. This can be mitigated by sequencing the two measurements using the \(\bowtie\) composition instead of the parallel composition, to make it more difficult for an adversary to “hide their tracks.” Rowe et al. [23] describe adversary capability models in more detail.

Version (2) improves on version (1), using \(\bowtie\) as described above and adding two other features.

\[
\begin{align*}
\ast \text{bank} : & \quad @\text{ks}[\text{av us bmon} \rightarrow !] \bowtie[\text{us}] [\text{bmon us exts} \rightarrow !] \\
\text{UC5} : & \quad \sim <
\end{align*}
\]

The first feature introduces the \(\sim\) operator to express that evidence produced by \(C\) is to be passed along to be processed by function \(D\). The second feature introduces the signature operator “!” to express that the measurements from \(C_1\) and \(C_2\) are to be separately signed and returned to bank.

## 5 Remote Attestation in Programmable Dataplanes

We first convert the example shown in Fig. 2 into Copland to show the language being used in a simple network setting. We then extend Copland to express UC1-UC5 from §2.

We start with the out-of-band variant from Fig. 2: here the evidence is sent from the switch directly to the appraiser for certification. The switch first hashes (the # operator) and then signs (! operator) this evidence. The first relying party, RP1, receives direct evidence of this appraisal and certification. RP2 can later retrieve this evidence from the appraiser.
Note that both expressions are bound by \( n \), a nonce parameter following Helble et al. [14]. This parameter is negotiated separately by \( RP1 \) and \( RP2 \). Example (3) below uses simplified syntax to reduce clutter by eliding obvious place details, and steps ①-③ from Fig. 2 are shown in blue to distinguish them from Copland syntax. Example (3) is described by two expressions that are evaluated in parallel:

\[
\begin{align*}
+&RP1, n : \langle @Switch \langle \text{attest (Hardware} \sim \text{Program)} \rangle \to \# \\
&\quad \text{①: Claim} \\
&\quad \to ![+] & \& \& ③: \text{Evidence} \\
&\quad \to ![+] & \& \& ③: \text{Evidence} \\
&\quad \langle @Appraiser \langle \text{appraise} \to \text{certify}(n) \rangle \to ![+] & \& \& ③: \text{Result} \\
&\quad \to ![+] & \& \& ③: \text{Result} \\
+&RP2, n : \langle @Appraiser \langle \text{retrieval}(n) \rangle \rangle ④: \text{Result}
\end{align*}
\]

The *in-band evidence* variant (4) is similar except that the evidence first reaches \( RP2 \) who then makes the appraisal request. Since \( RP2 \) learns the result of the appraisal directly, it does not need to separately enquire for a certificate as in the out-of-band variant. In this variant, we do not need a nonce to link the requests by the two Relying Parties (though a nonce can be used for freshness), and there is no need to store a certificate for later retrieval by another party. Unlike the previous example, this setting is described by a single expression. At the end of this process, both \( RP2 \) and \( RP1 \) would have received a signed certificate from the appraiser.

\[
\begin{align*}
+&RP1 : \langle \text{attribution} (\text{Hardware} \sim \text{Program}) \rangle \to # \\
&\quad \text{①: Claim} \\
&\quad \to ![+] & \& \& ③: \text{Evidence} \\
&\quad \to ![+] & \& \& ③: \text{Evidence} \\
&\quad \langle @Appraiser \langle \text{appraise} \to \text{certify}(n) \rangle \to ![+] & \& \& ③: \text{Result} \\
&\quad \to ![+] & \& \& ③: \text{Result} \\
+&RP2 : \langle @Appraiser \langle \text{retrieval}(n) \rangle \rangle ④: \text{Result}
\end{align*}
\]

### 5.1 Network-aware Copland

Copland expressions incorporate network topology details which restricts applicability to network settings where we do not have full topological visibility into the network. Even if they have knowledge of the network’s topology, the forwarding path between two peers is typically chosen outside their control, and the path might change without warning due to routing changes. In the banking example from §4.2, neither the client nor the bank is likely to have a complete picture of how their traffic is being forwarded. The banking service might be running in a cloud, the internal topology of which is not disclosed to the bank. The client app might be running behind a NAT which hides details from the rest of the Internet.

We sketch an extension of Copland that incorporates features of NetKAT [2] to describe attestation specifications over networks that can include programmable dataplanes.

This Copland+NetKAT hybrid has the following features:

**(Prim1)** The \( \Rightarrow \) operator is based on NetKAT’s Kleene star operator and provides path abstraction: the phrase on the left of this operator can hold for zero or more hops along the path. **(Prim2)** The \( \lor \) operator provides place abstraction by relaxing the requirement to explicitly name places at the time of writing a policy. **(Prim3)** Reachability testing is provided through a combination of the \( \Rightarrow \) operator and path abstraction. The \( \Rightarrow \) operator adapts NetKAT’s Boolean test prefix, and applies a Boolean test to a device before having it produce an attestation. A node (for which a \( \Rightarrow \)-test holds) is reachable if there is a path leading to it. That node can also attest the result of the test. This test is done for two design reasons: to “fail early” and avoid the attestation effort, and to apply different attestations based on which Boolean test succeeds.

**AP1** in Table 1 is an example of UC5. It adapts the bank example we saw earlier, the original parts of which are shown in blue. Here, the bank is the relying party (RP) as before. The nonce \( n \) and property \( X \) are \( RP \)-chosen parameters. \( X \) is some property that bank wishes to be attested at each hop—such as which P4 program and tables were used for forwarding. **AP1** also serves as example of UC1 by extending the property \( X \) to include other configuration details.

Terms *hop* and *client* are abstract place names—the first will be used for each hop along a path, and the second will be used to refer to the end of the path. The phrase to the left of \( \Rightarrow \) describes the gathering of evidence from each *hop* (which must satisfy test \( K_{\text{hop}} \)). Evidence is sent to the specific Appraiser place. \( K_{\text{hop}} \) and \( K_{\text{client}} \) ensure a pre-existing relationship between bank, *client* and each *hop*. This is not necessary, and is done to strengthen the specification.

Unlike **AP1**, **AP2** has a switch being the relying party. This policy is an example of UC4. Except for the use of \( \Rightarrow \), this could have been written entirely in Copland. In this example, \( P \) is a test being made on a packet. If the test succeeds then the test result is signed and sent to the Appraiser for storing.

**AP3** is a more complex example that shows an attestation for a path that has specific functions \( (F_1, F_2) \) running in abstract places \( (p, q) \). An interesting feature of this example is that \( p \) passes its evidence to \( q \) before it reaches Appraiser, and between \( q \) and \( r \) we do not require nodes that support RA.

### 5.2 Executing RA Policies

After authoring an RA policy, how do we deploy it? The policy will be compiled by the Relying Party and serialized into an options header in the transport layer, to be evaluated along the path of traffic that it is sending out. The Relying Party will then query named places for evidence.

To interpret this policy on flows we envisage a device that we call PERA, for “PISA [7] switch Extended with RA”. This switch will have access to specialized hardware primitives that can produce and consume evidence. These primitives might be integrated into the ASIC or might be remotely invoked by the programmable switch [27].

Fig. 3 sketches this hardware, noting that evidence might be sent in-band or out-of-band, as illustrated back in Fig. 2. Here, (A) and (D) show in-band evidence being received and sent by
Peer 1 and Peer 2 includes Non-attesting Elements (NE) indices in our design space for PERA. The path between the switch. Cases (B), (C) and (E) show evidence arriving and leaving separately. For some situations, it might be adequate to expect evidence to be gathered for each packet, and to add the RA policy to each packet. But in other situations, such per-packet overhead might be cumbersome and prohibitive.

![Image of an RA-capable programmable switch.](image)

Figure 3: An RA-capable programmable switch. Evidence-handling is tuned to balance performance and security.

![Table: Examples of Attestation Policies (APs) in Network-aware Copland.](table)

Table 1: Examples of Attestation Policies (APs) in Network-aware Copland.

In addition to the specification language and execution mechanism, we envisage a configuration interface that can tune the level of detail and frequency of evidence. Fig. 4 illustrates the main configuration choices. Inertia refers to the level of variability of attestable information across time: at one extreme, the model number of the hardware will not change, at the other extreme, a packet might be completely different than those that came before it. High-inertia attestations are more easily cached since they take longer to expire.

6 CONCLUSION

While programmable networking equipment offers great flexibility, networks implemented on such equipment are at risk from attacks that dynamically modify security-critical network behavior. This paper has sketched an approach for using RA techniques to enable dynamic assessment of network security characteristics. By extending the Copland RA policy language with elements of the NetKAT SDN programming language, we can define RA policies for programmable networks that specify the generation, collection, and evaluation of evidence of network program integrity. By implementing such RA policies, network dataplanes can participate in the process of proving their own trustworthiness.

ACKNOWLEDGMENTS

We thank the anonymous HotNets reviewers for their feedback. This material is based upon work supported by a Google Research Award, by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-19-C-0106, and by the National Science Foundation (NSF) under Grant No. ITE 2226443. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of funders.
Remote Attestation in Programmable Dataplanes

REFERENCES


