P4 VMへの透過的なパケット転送を可能にするAF_XDP拡張

田代 滉太[1]　岩井 正輝[1]　光来 健一[1]

1. はじめに

近年，プログラム可能なネットワークスイッチにおいて，P4 言語を用いた柔軟なパケット転送制御が可能となっている．例えば，トラフィック制御や計測，セキュリティ機能の高度化などに用いることができる．仮想マシン（VM）が接続される仮想ネットワークでも，P4 プログラムを実行する仮想 P4 スイッチが開発されている．クラウドのユーザが自身の P4 プログラムを仮想スイッチに適用できれば，VM ごとに異なる転送制御を行える．さらに，パケットのヘッダ情報だけでなく，ユーザの VM 内の情報も利用できれば，VM 内のシステムやアプリケーションの状態に応じた制御が可能になる．

一方，クラウドの仮想スイッチとユーザの P4 プログラムは互いに信頼できるとは限らない．仮想スイッチが P4 プログラムや VM 内の情報を不正に取得する恐れがある．逆に，ユーザの P4 プログラムが仮想スイッチの処理を妨げ，他の VM の通信に影響を与える恐れもある．先行研究の P4 Shield では，P4 プログラムを専用の P4 VM に分離して実行することで両者を隔離する [1]．しかし，パケットを P4 VM へ渡し，処理結果を受け取るために仮想スイッチの内部実装を変更する必要がある．この方法は特定の仮想スイッチに依存し，他の仮想スイッチへの適用や既存実装の保守を難しくする．

本研究では，AF_XDP を拡張することによって仮想スイッチを変更せずに P4 VM にパケット転送を行うことを可能にする TransP4 を提案する．

2. TransP4

図 1 に TransP4 の構成を示す．TransP4 は，送受信するパケットの情報を OS カーネル内で P4 VM に転送して P4 プログラムの実行結果を利用することにより，ユーザ空間で動作する仮想スイッチへの変更を不要にする．カーネル内でパケットを扱えるようにするために，AF_XDP ソケット API を用いる仮想スイッチを対象とする．AF_XDP はカーネルのネットワークスタックをバイパスし，XDP を用いてユーザ空間とデバイスドライバ間で高速なパケットのやり取りを実現する機構である．TransP4 はデバイスドライバ層においてパケットの情報を P4 VM に転送する．

AF_XDP ソケット API においては，パケットデータは UMEM と呼ばれるユーザ空間のメモリ領域に格納され，TX リングおよび RX リングと呼ばれるリングバッファを介してユーザ空間とデバイスドライバ間でのやり取りを行う．仮想スイッチがパケットを送信する際には，パケットデータを UMEM に格納し，そのデータを指すディスクリプタを TX リングに書き込む．一方，仮想スイッチがパケットを受信する際には，RX リングからディスクリプタを取り出し，そのディスクリプタが指す UMEM 上のパケットデータを取得する．

TransP4 は AF_XDP ソケット API を変更しないようにすることで，既存の仮想スイッチとの互換性を維持する．送信時には，デバイスドライバが TX リングから取り出したディスクリプタに基づいて P4 VM にパケット情報を転送する．P4 VM は UMEM 上のパケットデータを用いて P4 プログラムを実行し，パケット送信が許可された場合には，NIC が DMA を用いて UMEM 上のパケットデータを送信する．受信時には，デバイスドライバが P4 VM にパケット情報を転送する．P4 VM は UMEM 上のパケットデータを用いて P4 プログラムを実行し，パケット受信が許可された場合には，デバイスドライバが RX リングにディスクリプタを書き込む．

3. 実験

TransP4 を用いた際の通信性能を調べる実験を行った．外

[1] 九州工業大学
Kyushu Institute of Technology

© 1959 Information Processing Society of Japan

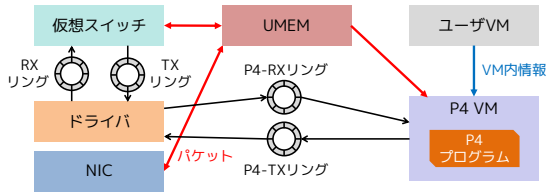


図 1 TransP4 の構成

表 1 実験に用いた 4 条件

	TransP4	SNP
plain	–	–
plain-snp	–	○
transp4	○	–
transp4-snp	○	○

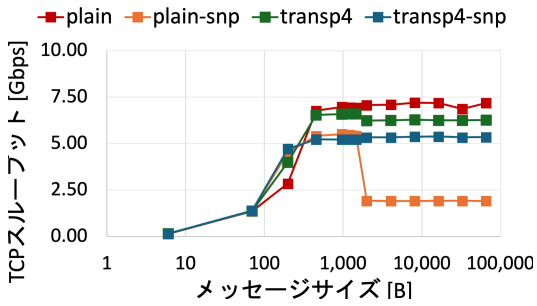


図 2 TCP 受信スループット

部クライアントからユーザ VM への通信について，TransP4 を適用するかどうか，AMD SEV-SNP を適用するかどうかで，表 1 の 4 条件を比較した．TransP4 を適用する場合には P4 プログラムを P4 VM で実行した．実験には，AMD EPYC 9174F の CPU，128 GiB のメモリ，Intel 10 GbE の NIC を搭載したサーバを用いた．ホスト OS として Linux 6.12.113，仮想スイッチとして Open vSwitch 3.5.0 を動作させ，AF_XDP native zero-copy を用いた．QEMU-KVM 9.2.4 を用いて P4 VM とユーザ VM を作成し，それぞれ 2 個の仮想 CPU，4 GiB のメモリを割り当て，virtio-net を用いた．

通信性能の測定には netperf 2.7.0 を用い，TCP/UDP STREAM は異なるメッセージサイズを用いて 5 フローを 10 秒間送信した．無負荷時の TCP/UDP 往復遅延は単一フローの RR 試験を 10 秒間実施した．各条件で 10 回測定し，5 フローの受信スループットの合計と 90 パーセンタイル遅延の平均を比較した．

3.1 受信スループット

図 2 と図 3 に TCP と UDP の受信スループットを示す．SEV-SNP を適用した場合，TransP4 を用いることで，メッセージサイズが 1460 B の時の TCP スループットは 3.7%減少し，UDP スループットは 0.1%減少した．SEV-SNP を適用しなかった場合の性能低下は TCP で 4.6%，UDP で 13.0%であったため，SEV-SNP 適用時には TransP4 のオーバヘッドが相対的に減少する．また，SEV-SNP を適用することにより，TransP4 を用いる場合には TCP スループットが 21.7%，UDP スループットが 38.3%，TransP4 を用いない場合でもそれぞれ 20.9%，29.1%低下した．このことから，SEV-SNP をユーザ VM に適用する影響の方が，P4 VM に適用する影響よりも大きいことが分かる．

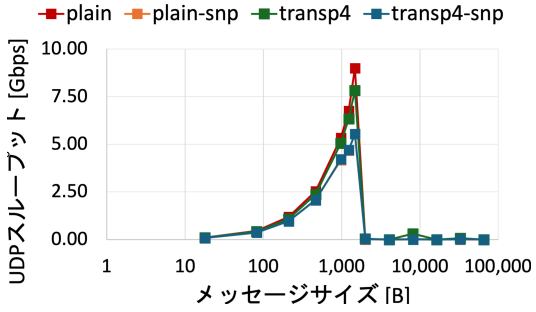


図 3 UDP 受信スループット

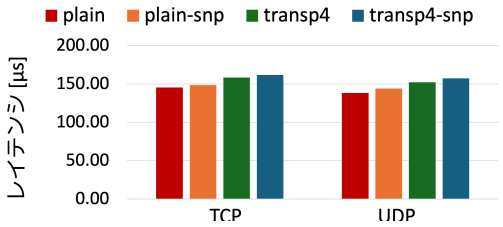


図 4 TCP と UDP の P90 遅延

3.2 往復遅延

図 4 は TCP と UDP の P90 遅延を示す．TransP4 を用いることにより，SEV-SNP を適用しない場合には，TCP で 13.2 μs，UDP で 13.8 μs 増加した．SEV-SNP を適用しても，遅延の増加はそれぞれ 13.3 μs，UDP で 13.1 μs であり，ほぼ同じであった．一方，SEV-SNP を適用することにより，TransP4 を用いる場合には TCP で 3.4 μs，UDP で 5.0 μs 増加し，TransP4 を用いない場合の増加もそれぞれ 3.3 μs，UDP 5.7 μs とほぼ同じであった．

4. まとめ

本研究では，AF_XDP を拡張することによって仮想スイッチを変更せずに P4 VM にパケット転送を行うことを可能にする TransP4 を提案した．今後の課題は，ユーザ VM と P4 VM の処理負荷を分けて測定し，性能低下の要因を調べることである．

参考文献

[1] M. Iwai and K. Kourai, “P4 Shield: Secure Execution of Users’ P4 Programs with In-VM Information inside Clouds,” IEEE CCWC, pp.1092–1099, 2026.

© 1959 Information Processing Society of Japan