

MidoNet リファレンスアーキテクチャ

2015.06-SNAPSHOT (2015-10-15 05:20 UTC)

DRAFT

MidoNet リファレンスアーキテクチャ

2015.06-SNAPSHOT (2015-10-15 05:20 UTC)

製作著作 © 2015 Midokura SARL All rights reserved.

概要

MidoNetは、Infrastructure-as-a-Service (IaaS)のためのネットワーク仮想化ソフトウェアです。

これにより、ネットワークハードウェアとIaaSクラウドを切り離すことができ、ホストと物理ネットワークの間に、インテリジェントなソフトウェア抽象レイヤーを作成することができます。

この文書には、推奨されるハードウェアを含め、MidoNetのネットワーク仮想化インフラストラクチャを準備するのに便利な情報が含まれています。特に、この文書では次について説明します。

- MidoNetの概要。
- OpenStack®およびその他のクラウドコントローラに対してMidoNetネットワーク仮想化を構成するのに必要なハードウェアとオペレーティングシステムソフトウェアの概要。
- ボーダーゲートウェイプロトコル（BGP）の設定とMidoNetネットワークアーキテクチャの一般的な概要。



注意

この文書はドラフトです。それは、関連する情報が欠落しているか、テストされていない情報が含まれていることができる。ご自身の責任でそれを使用してください。



注記

援助を必要とする場合は、 [MidoNetメーリングリスト](#)や[チャット](#) までご連絡ください。

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

目次

はじめに	vi
表記規則	vi
1. MidoNetの概要	1
MidoNetの主な特長	1
推奨されるハードウェア	2
インストールの要件	2
OpenStackの統合	3
2. MidoNetのネットワークアーキテクチャ	4
内部と上位のネットワーク	4
下層ネットワーク	5
BGPの設定とレイヤー3のトポロジ	6
仮想ルーター	7
プロバイダルータ	7
コンピュータホストエージェント	8
ブリッジ	8
メタデータサーバー	8
3. ソリューションコンポーネント	9
状態管理	9
4. MidoNetゲートウェイノード	12
ゲートウェイノードの設定条件	12
ゲートウェイノードの接続	12
5. Midolman	13
推奨するインストールノード	13
構成ガイドライン	13
アクセスに関する注意事項	13
6. MidoNet API	15
推奨されるインストールノード	15
耐障害性の構成ガイドライン	15
アクセスに関する注意事項	15
7. MidoNetコマンドラインインターフェイス	16

図の一覧

2.1. MidoNetのトポロジの例	4
2.2. レイヤー3のトポロジ（物理下層ネットワーク）	6
2.3. レイヤー3のトポロジ（仮想上層ネットワーク）	7

表の一覧

1.1. 推奨される導入ハードウェア 2

はじめに

表記規則

MidoNet のドキュメントは、いくつかの植字の表記方法を採用しています。

注意

注意には以下の種類があります。



注記

簡単なヒントや備忘録です。



重要

続行する前に注意する必要があるものです。



警告

データ損失やセキュリティ問題のリスクに関する致命的な情報です。

コマンドプロンプト

\$ プロンプト

root ユーザーを含むすべてのユーザーが、\$ プロンプトから始まるコマンドを実行できます。

プロンプト

root ユーザーは、# プロンプトから始まるコマンドを実行する必要があります。利用可能ならば、これらを実行するために、sudo コマンドを使用できます。

推奨されるハードウェア

このセクションでは、MidoNetの導入に推奨されるハードウェアの情報を提供します。

表1.1 推奨される導入ハードウェア

ハードウェア	要件
ネットワークステイトデータベース、APIおよびエージェントノード	CPU : 64-ビット x86、クアッドコア以上 メモリ : 32GB以上 (RAM) HDD : 30GB以上 (空き容量) NIC : 1Gbit以上x 2
GWノードx 2	CPU : 64-ビット x86、クアッドコア以上 メモリ : 32GB以上 (RAM) HDD : 30GB以上 ネットワーク : 1Gbit以上x 3
NICカード :	高性能データネットワーク: 複数のキューとMSI-XをサポートするNICを使用
トップオブブラックスイッチ	ジャンプフレームをサポートするノンブロッキングマルチレイヤースイッチ (L2/L3)
ハードディスク	理想的には、ZooKeeperのトランザクションログとCassandraデータファイルは、ホスト上の他のサービスの追加が可能な専用ディスクが必要です。ただし、小規模なPOCSやデプロイメントの場合は、Cassandraディスクを他のサービスと共有し、ZooKeeperのトランザクションログを専用にするだけでも構いません。

インストールの要件

オペレーティングシステム

MidoNetは次のオペレーティングシステムの64ビットバージョンに対応しています。

- Ubuntu 14.04 LTS
- Red Hat Enterprise Linux 7
- CentOS 7

BGPの設定要件

GWノードにBGPを設定するには、次のハードウェアと情報が必要です。

- ボーダールーターに接続したGWノード2台。一般的には、ロードバランスのために、各GWノードは異なるボーダールーターに接続します。
- 各GWノードには少なくとも2つの物理ネットワークインターフェイスが必要とされ、1つは内部ネットワーク、もう1つは上流のボーダールーターに接続します。
- 使用しているローカル（プライベート）ネットワークの自律システム（AS）番号
- リモート（パブリック）ネットワーク数（インターネットサービスプロバイダ（ISP）またはデータセンターなど）

- T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

- T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT



T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

そのため、グローバルIPアドレスを使用しないIPネットワークをお勧めします (<http://tools.ietf.org/html/rfc1918>[RFC 1918]参照)。ラック1つまたはラックなしの小規模な導入では単一障害点になってしまいますが、1つのイーサネットスイッチを内部のネットワークに使用できます。大規模な導入では階層型IPルーティングネットワークが適切です。可能であればECMP (Equal Cost Multi-Path) を使用してください。

下層ネットワーク

下層ネットワークは、MidoNetソフトウェアをホストする物理ネットワークです。

GREおよびVXLANトンネル

MidoNetでは、下層の物理ホスト間の通信にトンネルを使用します。 MidoNetは次の2つのトンネルプロトコルをサポートしています。

- GRE (General Routing Encapsulation) プロトコル (MidoNetのデフォルトのトンネルプロトコル)。GREのラッパーサイズは46バイトに固定されています。
- VxLAN (Virtual Extensible LAN) プロトコル。 VxLANでは50バイトのオーバーヘッドが追加されます。

フラグメンテーションとリアセンブルを回避するため、適切なMTU（最大転送単位）を設定してこのオーバーヘッドを許可します。

下層ネットワークに対するMTUサイズの注意事項

オーバーヘッドを許可するために、ネットワークのアップリンクと仮想マシンはMTUのデフォルトサイズ(1500)に、GREトンネルを使用する物理ネットワークデバイスのMTUは1546にします。VxLANトラフィックが機能するためには、物理ネットワークデバイスのMTUのサイズを1550にする必要があります。



重要

仮想マシンのMTUは、ボーダーゲートウェイのアップリンクインターフェイスのMTUを上回らないようにします。

上層ネットワークに対するMTUのサイズの注意事項

最適なデータリンクの設定は個々の環境によって異なります。MidoNetはイーサネットのジャンボフレームをサポートしています。ジャンボフレームのサポートを設定する場合、MidoNetネットワークのネットワークインターフェイスのMTUは、下層（物理）ネットワークのMTUのサイズを少なくとも46バイトまたは50バイト下回る必要があります（GREとVXLANをそれぞれカプセル化するため）。仮想ネットワークのトラフィックパスで発生するMTUの不一致に注意してください。このような不一致によって、IPのフラグメンテーション/デフラグメンテーションが起こり、ネットワークパフォーマンスに悪影響を及ぼす可能性があります。

下層ネットワークで1500バイト以上のイーサネットフレームがサポートされない場合、MidoNetネットワークのMTUを1454または1450バイトに設定します（GREとVLANをそれぞれカプセル化するため）。この設定によって、仮想マシンのネットワークインターフェイスに適切なMTUのサイズを設定することができます。

L3ゲートウェイアップリンクのNICのオフロード

NICへのオフロードは仲介ホストではなく、エンドホストに対して実行されます。ゲートウェイアップリンクのNICはルーターのNICと同様に扱う必要があります。

L3ゲートウェイアップリンクのNICでLR0が有効な場合、NICは受信したTCPパケットと結合して、宛先のMTUより大きいパケットをMidoNetに渡します。MidoNetではサイズの大きいセグメントはオフロード（LS0-転送前にサイズの大きいTCPを分割）されず、IPフラグメンテーションもサポートされないため、パケットはドロップします。そのため、L3ゲートウェイアップリンクのNICへのオフロードを無効にする必要があります。アップリンクのNICで次を実行してください。

```
# ethtool -K p2p1 lro off
```

または、下記をネットワークスクリプトファイルに追加してください。

```
ETHTOOL_OPTS="lro off"
```

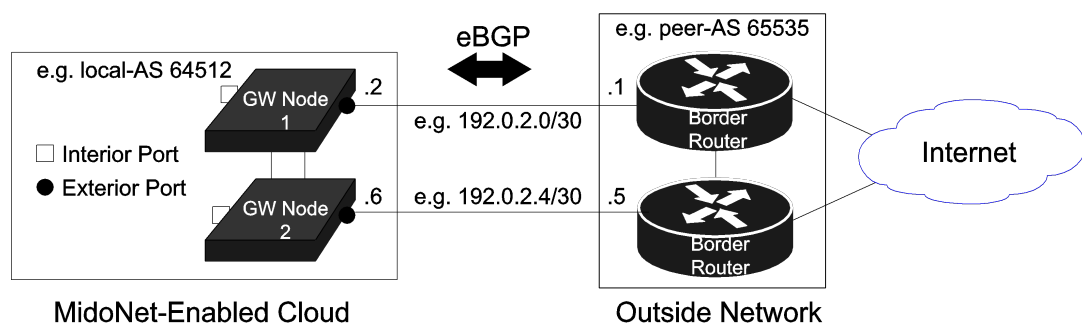
BGPの設定とレイヤー3のトポロジ

このセクションでは、BGPの設定とレイヤー3のトポロジに関する図と情報を提供します。

図2.2「レイヤー3のトポロジ（物理下層ネットワーク）」 [6] は、基盤となるネットワークインフラストラクチャの例です。

図2.3「レイヤー3のトポロジ（仮想上層ネットワーク）」[7]は、MidoNet仮想化ネットワークを基盤となるネットワークアーキテクチャの上位に配置した例と、BGPルートアドバタイズメントの情報です。

図2.2 レイヤー3のトポロジ（物理下層ネットワーク）



The diagram illustrates a MidoNet-Enabled Cloud architecture. On the left, a large blue cloud represents the cloud environment, labeled "MidoNet-Enabled Cloud". Inside this cloud, there are two "Project" containers. "Project 1" contains three VMs (VM0, VM1, VM2) connected to "Subnet 1" and "Subnet 2", which are then connected to a "Virtual Router" (router1). "Project 2" contains three VMs (VM0, VM1, VM2) and one VM3, connected to "Subnet 3" and "Subnet 4", which are then connected to a "Virtual Router" (router2). These virtual routers are connected to a "Cloud-Provider (Virtual) Router" (router0). The cloud is labeled "e.g. local-AS 64512".

On the right, an "Outside Network" is shown, containing two "Border Router" instances. The top border router is labeled "e.g. peer-AS 65535" and has an IP address ".1". The bottom border router has an IP address ".5". The "Cloud-Provider (Virtual) Router" (router0) is connected to the top border router via "port0 .2" and to the bottom border router via "port1 .6". The IP addresses for these connections are "e.g. 192.0.2.0/30" and "e.g. 192.0.2.4/30" respectively. The outside network is connected to the "Internet".

At the top, a double-headed arrow labeled "eBGP" indicates the connection between the local AS (64512) and the peer AS (65535). Text explains the advertisement process: "AS 65535 advertises 0.0.0.0/0 to AS 64512 for each peer." and "AS 64512 advertises 192.0.2.128/25 to AS 65535 for each peer."

物理ルーターと同様に、仮想ルーターでもネットワークインターフェイス（ポート）、ネットワークアップリンク（上流デバイス、通常、物理ルーターとブリッジに接続する物理イーサネットインタフェイスに接続するポート）を設定できます。また、仮想ルーターは他のルーターとブリッジに接続できます。

外部ネットワークの作成の詳細については、次のリンクを参照してください：
<http://docs.openstack.org/havana/install-guide/install/apt/content/install-neutron.configure-networks.html>

コンピュータホスト上のMidoNetエージェントは、クラウドの横方向のトラフィックの大半と上位の送信トラフィックを処理します。現在はKMVハイパーバイザをサポートしています。

ブリッジ

ブリッジはMidoNetのL2転送エレメントです。

ブリッジには仮想ポートを作成できます。また、ブリッジの仮想ポートにVMを接続できます。同じポートに接続しているすべてのVMにはL2の接続を通してアクセスできます。仮想ブリッジのポートを他の仮想デバイスや物理マシンに接続することはできません。

ブリッジには、受信したフレームを送信するネットワークデバイスのMACアドレスとそれを受信するブリッジポート間のマッピングが保存されます。

ブリッジによって、ソースのMACアドレスとブリッジポートのテーブルが自動的に作成されます。ブリッジではこのテーブルを使用して、フレームがネットワークデバイスから適切なブリッジポートに送信されます。MACテーブルは消去できます。

メタデータサーバー

メタデータサーバーはインスタンスVMの構成情報、たとえば、認証情報やVMのカスタマイズスクリプトなどを保存するのに使用されます。

メタデータサーバーにはVMの構成の値のテーブルが保存されます。

MidoNetではCassandraの耐久性、耐障害性、有効期限、低遅延の読み込み/書き込みなどの利点を活用していますが、Cassandraは一次データソースではなくバックアップとしてのみ使用します。

必要なソフトウェア

CassandraにはJavaランタイム環境（JRE）が必要です。

Linuxディストリビューションの多くで提供され、公式インストールガイドを使用してインストールできることから、OpenJDK 7をお勧めしています（<http://openjdk.java.net/>でインストールに関する情報をご覧ください）。 == Fault-tolerant configuration guidelines

推奨されるCassandraの最小構成は、3ノードのクラスタと3のレプリケーションファクタ（N）です。

MidoNetエージェント（Midolman）では、QUORUM（ $N/2 + 1$ ）を整合性ポリシーとして使用していますが、これは提案した構成では2になります。

アクセスに関する注意事項

Cassandraは2つのIPアドレスを使用します。1つはクラスタ内の通信（listen_addressパラメータ）、もう1つはリモートプロシージャコール（RPC）によるクライアントの接続です（rpc_address）。

MidoNetエージェントではDNS (Domain Name System) を使用して、ホスト名と下層ネットワークアドレスが変換されます。 MidoNetエージェントをインストールした各サーバーのホスト名が解決可能であることを確認してください。

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

[illegible]

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT

T - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT - DRAFT -

第7章 MidoNetコマンドラインインターフェイス

MidoNet CLIはMidoNetの仮想トポロジを検査および編集するためのコマンドラインインターフェイスです。