Cisco Tetration Analytics and Turbonomic Solution

Deploy intent-based networking for distributed applications.

Highlights
- Provide performance assurance for distributed applications.
- Real-time localization of workloads reduces east-west latency.
- Workload mobility accounts for application dependencies.

Challenges: Dynamic, distributed applications and increasing network complexity
Application architectures have changed dramatically over the years, but networks have remained relatively static. With the adoption of distributed architectures, containerization, microservices, and workload mobility technologies, communication patterns between application components are more dynamic and the network is more critical than ever. Now, 76 percent of data center traffic is east-west, creating new challenges that your organization must address:
- Your applications have multiple components that need to communicate to perform. How do you help ensure that your distributed applications don’t congest the network?
- Migrating applications risks downtime. How do you help ensure dependent that workloads are migrated together? How do you reduce costs in the cloud as you also do so in hybrid cloud deployments?

These are difficult problems to solve, especially as cloud workloads are projected to more than triple (grow 3.2-fold), and workload density for cloud servers is projected to increase from 7.3 to 11.9 between 2015 and 2020 (Cisco® Global Cloud Index). The challenges of managing today’s environments are too complex not to turn over to software.

The Cisco Tetration Analytics™ and Turbonomic solution
Today’s modern applications are distributed, relying on your network more than ever for performance. Interdependent workloads communicating across hosts, clusters, data centers, and clouds create new challenges in ensuring that workloads get the network, computing, and storage resources they need when they need them. This is a problem that software can solve.

The Cisco Tetration Analytics™ and Turbonomic solution supports real-time distributed application fluidity by helping assure performance while increasing efficiency. With this joint solution doing what it does best, your staff members can focus on what they do best: innovating and creative problem solving.
Architectural overview

Turbonomic accesses Cisco Tetration Analytics data through Representational State Transfer (REST) APIs. It uses telemetry data to make resource allocation decisions in real time (Figure 1).

Figure 1  Cisco Tetration Analytics and Turbonomic solution

Solution use cases

Table 1 shows common uses for the Cisco Tetration Analytics and Turbonomic solution.

Table 1  Solution use cases

<table>
<thead>
<tr>
<th>Use case 1</th>
<th>Intent-based networking for distributed applications reduces east-west latency.</th>
<th>Using Cisco Tetration Analytics telemetry data, Turbonomic reduces network latency between endpoints by localizing “chatty” workloads without risking host congestion in network, computing, or storage resources.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use case 2</td>
<td>Distributed application fluidity helps ensure performance while reducing costs in hybrid clouds.</td>
<td>Turbonomic uses Cisco Tetration Analytics application dependency mapping to fluidly migrate dependent workloads across clusters, data centers, and clouds, helping ensure application performance while reducing costs.</td>
</tr>
</tbody>
</table>
Solution overview
Figures 2 through 8 provide an overview of the solution.

Figure 2  Turbonomic maps the data center stack to its supply chain market model and uses Cisco Tetration Analytics to gain a deep understanding of the communication patterns and relationships between workloads.
Figure 3  Cisco Tetration Analytics data collection tells Turbonomic the relationship between endpoints

Figure 4  Cisco Tetration Analytics monitors the flow patterns between workloads, and Turbonomic uses that information to make the right workload placement decisions
Figure 5 Using Cisco Tetration Analytics, the Turbonomic analysis engine understands cross-host east-west traffic; in this example, utilization is 100%.

Figure 6 Understanding of the relationship between workloads and the network, Turbonomic recommends moving a specific virtual machine, tomcat2-ubuntu, to another host to prevent network congestion on the current host.
Figure 7  Recommended actions can be performed directly from the user interface by selecting the actions and clicking Apply Selected, or, if fully automated, these actions will be performed in real time.

Figure 8  By performing the actions that Turbonomic recommends, network congestion risks are avoided.
Solution components
The Cisco Tetration Analytics platform installs sensors on endpoints (such as servers) to collect telemetry information. The platform can support approximately up to 5000 unique endpoints or process up to one million unique flows per second (Figures 9 and 10).

Figure 9  Cisco Tetration Analytics dashboard

Figure 10  Cisco Tetration Analytics high-level architecture
Turbonomic deploys as a virtual machine in your environment. Working through the APIs of Cisco Tetration Analytics and the elements already in your environment (hypervisors, network, public cloud, cloud management, orchestration, storage, hyperconverged systems, etc.), it uses an agentless approach to pull the data that’s already being collected. It then analyzes that information to determine the workload placement, sizing, and provisioning that helps ensure that workloads get the resources they need when they need them. When the process is fully automated, performance is assured, with increased efficiency in real time (Figures 11 and 12).

Figure 11  Turbonomic dashboard

![Turbonomic dashboard](image)

Figure 12  Turbonomic high-level architecture

![Turbonomic high-level architecture](image)
How it works
The Cisco Tetration Analytics platform delivers pervasive visibility across the data center. The platform installs software agents on servers, which act as sensors and collects flow data between all endpoints. Turbonomic uses Cisco Tetration Analytics REST APIs to access this data, understand the communication patterns between application components, and make real-time workload placement decisions that reduce east-west latency while taking into account computing and storage needs.

Turbonomic makes the right resource allocation decisions in real time. It does this by abstracting data center and cloud environments into a supply-chain market of buyers and sellers. Workloads (applications, containers, and virtual machines) “buy” the resources (CPU, memory, I/O, I/O operations per second [IOPS], flow, etc.) they need from the infrastructure (hosts, storage, network, hyperconverged systems, etc.). The “price” of these resources is a function of their utilization: the greater the utilization, the more expensive the resource. This abstraction enables workloads to shop around for the best overall price for the multiple resources they need to perform. The result is a decision engine that enables your workloads to manage themselves in real time.

Main capabilities
Real-time workload management of computing, network, and storage resources helps ensure the performance of distributed applications.

Benefits of the solution
The Cisco Tetration Analytics and Turbonomic solution offers these benefits:

- Software manages software. Your staff can focus on innovation for the business.
- It reduces east-west latency in distributed applications.
- You can monitor real-time computing, network, and storage utilization across the stack.
- You can gain a holistic understanding of workloads and their interdependencies and relationship to the network and data center or cloud stack.
- You can fully automate workload placement decisions for real-time performance assurance.

Conclusion
The Turbonomic decision engine uses the deep insights from the Cisco Tetration Analytics platform to automate decisions about optimal workload placement and scaling in a virtualized environment. This joint solution delivers self-managing workloads that reduce latency in distributed applications. The solution will continue to address additional use cases in the future to solve emerging customer requirements.

For more information
For additional information see:

- www.cisco.com/go/dcecosystem
- www.cisco.com/go/tetration
- www.turbonomic.com