Nokia Deepfield Cloud Genome

Mapping the internet services supply chain in the cloud era

Nokia Deepfield Cloud Genome® maps the global internet to provide unprecedented visibility for service providers, webscale companies and large digital enterprises.

Deepfield Cloud Genome is a cloud-based data feed that continuously probes and tracks billions of IPv4 and IPv6 addresses on the internet, maps them to Domain Name Server (DNS) names, and employs advanced machine learning (ML) rules to tag the addresses into different types and categories. The result is a global, cloud-based data feed about internet content, applications and services: a dynamic supply map of the internet.

When correlated with the information obtained from the network, Cloud Genome allows full network and service visibility down to the service/application, content delivery network (CDN), peering/transit router, backbone node, access system and device/subscriber.

Cloud Genome is a component of the Nokia Deepfield Genome, a set of two complementary and proprietary data feeds that employ Nokia patented technology:

• Cloud Genome: Provides full visibility of internet content, applications and services.
• Secure Genome: Provides full visibility of internet security-related data.

Deepfield Genome is an intrinsic part of the Nokia Deepfield portfolio of IP network intelligence, analytics and distributed denial of service (DDoS) security applications for service providers (cable providers, multiple system operators [MSOs], telecommunications service providers), webscale companies and large digital enterprises.

Deepfield applications ingest data from many sources in your network, then process and correlate the data to provide a real-time, multidimensional view of your network, services and IP flows.

Deepfield applications are further enhanced by Deepfield Genome to deliver better network insight and facilitate improved performance and security.

Features

• Maps and tracks more than 4 billion IPv4 and 1 billion IPv6 endpoints
• Maintains hundreds of thousands of DNS mappings
• Crawls more than 1 billion internet IP addresses daily
• Keeps track of the most relevant (150-200 million) IP addresses for internet content, applications and services and tags them against CDNs, servers and hosting companies
• Employs more than 100 advanced ML rules for automatic classification of internet traffic
• Uses several external API-accessible databases to enhance accuracy.
Benefits
• Provides an up-to-date snapshot of the supply map of the internet
• Precisely allocates applications and flows into traffic types and categories
• Allows full matching of network information with the larger internet context and end-to-end visibility of internet services and application flows.

How Cloud Genome works
The internet is continuously growing—both in size and in its variety of content and services. Service providers, webscale companies and large digital enterprises want better insights about their networks and the services offered to subscribers. Today, most of these services are flowing into and across the network from the large application and content domains and CDNs. To understand how services are delivered, you need full insight into your network’s traffic flows and the ability to correlate the insights in an end-to-end manner to the internet—from the originating domains to end systems and subscribers.

Network analytics solutions that ingest only flow-based (e.g., NetFlow) and Border Gateway Protocol (BGP) data, relying on just IP addresses and BGP routes, don’t have the ability to fully understand and categorize internet traffic and are not up to the task of detailed traffic analysis and reporting that is required in the cloud era.

Legacy approaches with dedicated hardware probes and using deep packet inspection (DPI) technology are also inefficient when dealing with extreme volumes of traffic which may be encrypted. These legacy approaches may not be scalable or may not be able to provide the required end-to-end service visibility.

By contrast, Cloud Genome continuously crawls the internet and maps billions of IP addresses and millions of cloud applications and services. Cloud Genome categorizes traffic across a large number of distinct traffic categories to provide a rich and detailed understanding of the global internet services supply chain (see Figure 1), including:
• CDN infrastructure
• Major internet providers
• Autonomous System Numbers (ASNs)
• Hosting companies
• Internet of Things devices
• GeoIP information (countries/regions/cities).

Cloud Genome relies on a number of distributed, cloud-based, data-mining agents. These agents constantly examine and interact with internet endpoints to learn about the services and the traffic patterns, and then build a dynamic map of the entire internet. Cloud Genome is used to complement and augment network information and provide a level of visibility into internet applications and services that was previously unattainable.

Cloud Genome enables an end-to-end view across an entire infrastructure and even of encrypted traffic without using a single network probe or DPI. With a holistic perspective of the whole network, you can accurately plan your network capacity, monitor end-to-end service delivery, and ensure optimal quality of experience (QoE) for customers.

Figure 1. Deepfield Genome
Providing context to the Deepfield applications

Cloud Genome provides context to the Deepfield analytics and security applications:

- Nokia Deepfield Cloud Intelligence: Delivers end-to-end network visibility for traffic engineering, capacity planning and service optimization.
- Nokia Deepfield Service Intelligence: Monitors QoE for over-the-top (OTT) video services in real time.
- Nokia Deepfield Subscriber Intelligence: Provides service visibility with end-user and subscriber-level granularity.

An example of how Deepfield Genome knowledge enhances legacy flow-based information is shown in Figure 2.

Figure 2. Enhancing legacy flow-based information with Deepfield Genome knowledge

<table>
<thead>
<tr>
<th>Ingress interface</th>
<th>Egress interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source address</td>
<td>Server type</td>
</tr>
<tr>
<td></td>
<td>Device type</td>
</tr>
<tr>
<td>Reflector type</td>
<td></td>
</tr>
<tr>
<td>CDN</td>
<td>Hosting provider</td>
</tr>
<tr>
<td>Protocol</td>
<td></td>
</tr>
<tr>
<td>Source port</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td></td>
</tr>
<tr>
<td>Site/service</td>
<td></td>
</tr>
<tr>
<td>Geographical info</td>
<td></td>
</tr>
<tr>
<td>Geographical info</td>
<td></td>
</tr>
<tr>
<td>Fully QDNI</td>
<td></td>
</tr>
<tr>
<td>Trusted</td>
<td>Untrusted</td>
</tr>
<tr>
<td>Number of bytes</td>
<td></td>
</tr>
<tr>
<td>Number of packets</td>
<td></td>
</tr>
</tbody>
</table>

Legacy flow-based information | Enhanced by Cloud Genome | Enhanced by Secure Genome
Cloud Genome identifies the endpoints for all network traffic traversing the internet, providing visibility down to the CDN, site and service. This level of detail gives Deepfield Cloud Intelligence the ability to accurately classify all traffic across your entire infrastructure, arming network operators with knowledge on how to make the best decisions about where to, for example, add bandwidth or add caches to ensure the best network performance.

Because Cloud Genome is a data feed with a holistic view of all internet applications and services, it allows Deepfield Service Intelligence to have a complete, real-time view of OTT video services and to monitor the aggregate number of streams and average bit rates (ABRs) delivered to any part of the network or to any location or subscriber.

Cloud Genome also empowers Deepfield Subscriber Intelligence to allow full visibility into how services are delivered to the access layer with subscriber-level granularity and facilitates detailed insights on service consumption patterns and customer experience.

Cloud Genome is instrumental in enabling diverse use cases for Deepfield applications, including:

- Peering analytics and optimization
- Backbone engineering
- Traffic engineering and capacity planning
- Content delivery network analytics
- Traffic analysis by access type
- OTT video insights for improved customer QoE
- Per-service plan/subscriber traffic analytics and content consumption insight
- Customer care.

To learn more about the Deepfield solution, visit the Deepfield web page.