**Network Segmentation Analyzer**User Guide for Reading Network Diagrams

*Version 3.0*

**Prepared by: PruTech Network Security Team**

# Table of Contents

1. Introduction

2. Types of Diagrams

3. Reading the Diagrams

3.1 Shapes and Visual Elements

3.2 Colors and Security Zones

3.3 Line Types and Patterns

4. Understanding Data Sources

5. Upstream vs Downstream Dependencies

6. Circle Sizes and Visual Indicators

7. Unknown Connections Explained

8. Interactive Features (HTML Diagrams)

9. Using Diagrams for Security Analysis

10. Frequently Asked Questions

# 1. Introduction

Welcome to the Network Segmentation Analyzer user guide. This document will help you understand and interpret the network topology diagrams generated by our analysis system.

## What You Will Learn

* How to read tier-based network architecture diagrams
* Understanding upstream and downstream dependencies
* Interpreting data source attribution (observed vs predicted)
* Using diagrams for security analysis and segmentation planning

## Who Should Read This

This guide is intended for:

* Network administrators planning security segmentation
* Security architects designing zero-trust architectures
* Application owners understanding their dependencies
* Compliance teams assessing regulatory requirements (PCI-DSS, HIPAA, etc.)

# 2. Types of Diagrams

The Network Segmentation Analyzer generates two main types of diagrams for each application:

## 2.1 Standard Tier-Based Diagrams

These diagrams show your application organized by network tiers (Web, App, Data, etc.) on the left side, with external dependencies (databases, queues, downstream applications) on the right side.

**Files:** {APP}\_diagram.html, {APP}\_diagram.mmd, {APP}\_diagram.png

## 2.2 Application Data Flow Diagrams

These diagrams show internal application architecture with component grouping by type (databases, caches, queues) and application-level data flows.

**Files:** {APP}\_application\_diagram.html, {APP}\_application\_diagram.mmd, {APP}\_application\_diagram.png

# 3. Reading the Diagrams

## 3.1 Shapes and Visual Elements

### Shape Guide:

* BOXES (Rectangles): Network tiers/zones containing multiple servers
* CIRCLES: Services, APIs, or application endpoints
* RECTANGLES: Data stores (databases), caches, message queues
* CYLINDERS: Database systems (specialized data store shape)
* ROUNDED BOXES: External systems or downstream applications

**Important:** The SIZE of circles does NOT indicate importance or traffic volume. All shapes are sized for readability. Traffic volume is indicated by LINE THICKNESS.

## 3.2 Colors and Security Zones

Background colors indicate security zones and risk levels:

* RED/PINK (#ffcccc): Web Tier - High Risk (internet-facing)
* BLUE (#cce5ff): Application Tier - Medium Risk (business logic)
* ORANGE (#ff9966): Data Tier - High Value (databases, sensitive data)
* LIGHT ORANGE (#ffcc99): Cache Tier (Redis, Memcache)
* PURPLE (#cc99ff): Messaging Tier (Kafka, RabbitMQ)
* YELLOW (#ffff99): Management Tier (infrastructure, monitoring)
* GRAY (#e0e0e0): Unknown - Could not be classified from network data

**Security Implication:** Higher risk zones (Red, Orange) should have stricter firewall rules and cannot communicate directly with each other without passing through the Application Tier.

## 3.3 Line Types and Patterns

Lines connecting components show network traffic flows:

* THICK SOLID LINES (=====>) : High volume (>100 network flows)
* SOLID LINES (--->) : Medium volume (10-100 flows)
* DASHED LINES (-.-.-> ) : Low volume (<10 flows) or Predicted connections

**Arrow Direction:** Arrows point in the direction of traffic flow. A → B means "A initiates connections to B".

# 4. Understanding Data Sources

The diagrams distinguish between three types of connections based on data source:

## 4.1 Black Solid Lines = Observed Data (ExtraHop)

**Source:** Network flow data captured from ExtraHop monitoring

**Reliability:** HIGHEST - These connections were actually observed on your network

**Usage:** Use these for primary segmentation decisions and firewall rules

## 4.2 Blue Dashed Lines = ML Predictions

**Source:** Machine Learning type inference or Markov chain predictions based on usage patterns

**Reliability:** MEDIUM - Predicted based on similar applications and network patterns

**Usage:** Consider these as "likely dependencies" that should be verified with application teams before implementing firewall rules

## 4.3 Gray Dashed Lines = Unknown/Unclassified

**Source:** Connections that could not be classified from available network data

**Reliability:** LOWEST - Requires manual investigation

**Usage:** Flag these for application owners to provide details on the service type and purpose

**IMPORTANT:** Always verify ML predictions with application teams before implementing restrictive firewall rules. False positives could break application functionality.

# 5. Upstream vs Downstream Dependencies

## Understanding Terminology

**UPSTREAM:** Services that YOUR application DEPENDS ON. If upstream services fail, your application may fail.  
Examples: Databases, caches, authentication services, third-party APIs

**DOWNSTREAM:** Services that DEPEND ON your application. If your application fails, downstream services may fail.  
Examples: Web frontends calling your API, reporting systems pulling your data

## How to Identify in Diagrams

In our tier-based diagrams:

* LEFT SIDE: Your application tiers (Web, App, Data, etc.)
* RIGHT SIDE: External dependencies grouped by type
* ARROWS POINTING RIGHT (→): Your app depends on these (UPSTREAM)
* ARROWS POINTING LEFT (←): These depend on your app (DOWNSTREAM - less common)

## Segmentation Implications

For micro-segmentation planning:

1. UPSTREAM dependencies: Must allow OUTBOUND traffic from your app tier to these services (e.g., App Tier → Database Tier on port 3306)
2. DOWNSTREAM dependencies: Must allow INBOUND traffic from these services to your app tier (e.g., Web Tier → App Tier on port 8080)
3. DENY ALL other traffic: Use "default deny" rules for any connections not shown in diagrams

# 6. Circle Sizes and Visual Indicators

## Do Circle Sizes Matter?

**Short Answer: NO**

Circle and box sizes in the diagrams are automatically sized for READABILITY, not for indicating traffic volume, importance, or server count.

## What DOES Indicate Importance?

1. LINE THICKNESS: Thicker lines = higher traffic volume = more critical dependency
2. LINE COLOR: Black solid = actually observed (most reliable), Blue dashed = predicted (verify first)
3. SERVER COUNT: Look for "X server(s)" in tier boxes - more servers may indicate higher availability needs
4. ZONE COLORS: Red/Orange zones contain higher-risk or higher-value assets requiring stricter controls

**Example:** A small circle with a thick black line is MORE important than a large circle with a thin dashed line.

# 7. Unknown Connections Explained

You may see some connections labeled as "Unknown" or shown in gray. This section explains why.

## What Are Unknown Connections?

Unknown connections are network flows that could not be definitively classified based on available ExtraHop network flow data.

## Common Causes

1. MISSING SERVICE INDICATORS: Destination endpoints do not have clear service type indicators in their network signatures (e.g., non-standard ports, encrypted traffic)
2. INSUFFICIENT CONTEXT: Flow data lacks sufficient context to determine the application protocol (e.g., generic TCP connections without payload inspection)
3. CUSTOM SERVICES: Connections involve custom or proprietary services without standard port/protocol patterns (e.g., internal RPC frameworks, custom APIs)

## Recommended Actions

**Step 1:** Identify the IP addresses involved in Unknown connections

**Step 2:** Contact application owners to determine the service type and purpose

**Step 3:** Update firewall documentation with correct service classifications

**Step 4:** If these are test/development connections, consider excluding them from production segmentation rules

# 8. Interactive Features (HTML Diagrams)

HTML diagram files (\*.html) provide interactive features for easier navigation:

## Navigation Controls

* MOUSE WHEEL: Zoom in/out
* CLICK & DRAG: Pan around the diagram
* 4-WAY ARROWS: Pan in specific directions
* FIT TO SCREEN: Auto-resize diagram to fit your screen
* RESET VIEW: Return to original zoom and position

## Legend Controls

**NEW:** The legend can now be hidden/shown using the X button to save screen space. Click "Show Legend" to bring it back.

## Exporting and Sharing

Three formats are available for each diagram:

* HTML: Interactive version for analysis (best for detailed review)
* PNG: Static image for presentations and documentation
* MMD: Mermaid source code for editing or import into other tools

# 9. Using Diagrams for Security Analysis

## 9.1 Micro-Segmentation Planning

Follow these steps to use diagrams for zero-trust segmentation:

1. IDENTIFY TIERS: Note which tiers exist (Web, App, Data, etc.)
2. MAP ALLOWED FLOWS: For each thick black line, create an "ALLOW" firewall rule
3. VERIFY PREDICTIONS: For blue dashed lines, verify with app teams before allowing
4. DENY ALL ELSE: Implement "default deny" for any traffic not explicitly shown
5. MONITOR UNKNOWN: Flag Unknown connections for investigation before production deployment

## 9.2 Compliance Requirements

These diagrams help satisfy regulatory requirements:

* PCI-DSS 1.2.1: Network diagrams showing cardholder data flows
* HIPAA Security Rule: Technical safeguards documentation (§164.312)
* SOX IT Controls: Documentation of critical financial system dependencies
* GDPR Article 32: Documentation of security measures for personal data

## 9.3 Incident Response

During security incidents, use diagrams to:

* IDENTIFY BLAST RADIUS: Which systems could be affected if this server is compromised?
* LATERAL MOVEMENT PATHS: What connections would an attacker use to move to high-value systems?
* CONTAINMENT POINTS: Where should firewall rules be tightened to isolate the incident?
* DEPENDENCY IMPACT: If we isolate this system, what downstream services will be affected?

# 10. Frequently Asked Questions

### Q: Why are some IPs shown instead of hostnames?

A: This can occur when: (1) DNS reverse lookup failed, (2) The IP is in a demo/test environment without proper DNS registration, or (3) The system could not resolve the hostname. In these cases, the raw IP address is displayed. For real hostnames, you will see "IP - Hostname" format.

### Q: Can I edit the diagrams?

A: Yes! The .mmd (Mermaid) files can be edited with any text editor or imported into tools like Lucidchart, Draw.io, or Mermaid Live Editor (mermaid.live). Make changes and regenerate the diagrams.

### Q: How often should diagrams be regenerated?

A: Recommended frequency: (1) After major application deployments or architecture changes, (2) Quarterly for stable environments, (3) Before security audits or compliance reviews, (4) After discovering unauthorized network connections in monitoring.

### Q: What if I see connections that should not exist?

A: This is a SECURITY FINDING. Black solid lines represent actual observed traffic. If you see a connection that violates your security policy (e.g., Web Tier directly connecting to Database Tier), investigate immediately for: (1) Misconfigured applications, (2) Shadow IT, (3) Potential compromise, (4) Incomplete firewall rules.

### Q: Can the ML predictions be improved?

A: Yes! The ML Predictions Dashboard (if available) provides a feedback system. Mark predictions as Correct or Wrong, and the model will improve over time. More training data = better predictions.

### Q: Why do some applications have no Unknown connections while others do?

A: Applications using standard technologies (MySQL, Redis, Kafka) on standard ports are easier to classify automatically. Custom in-house services or applications using non-standard ports/protocols are more likely to appear as Unknown until manually classified.