Juniper Networks Design—WAN

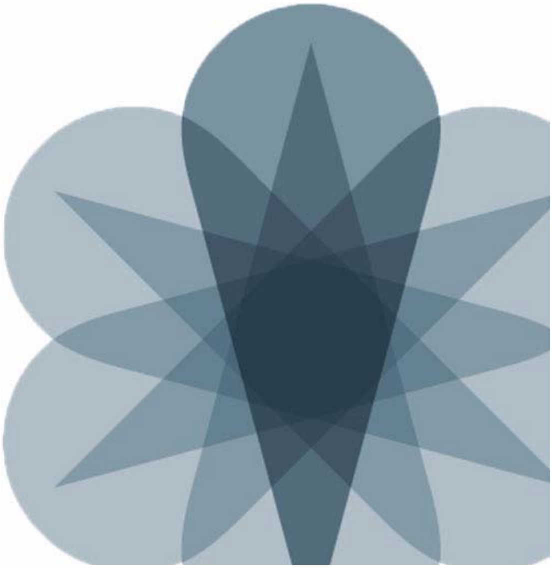
16.a

Student Guide
Volume 1 of 3

Worldwide Education Services

1133 Innovation Way
Sunnyvale, CA 94089
USA
408-745-2000
www.juniper.net

Course Number: EDU-JUN-JND-WAN



This document is produced by Juniper Networks, Inc.

Juniper Networks Design—WAN Student Guide, Revision 16.a

Printed in USA.

Revision History:

Revision 16.a—May 2016.

The information in this document is current as of the date listed above.

The information in this document has been carefully verified and is believed to be accurate. Juniper Networks assumes no responsibilities for any inaccuracies that may appear in this document. In no event will Juniper Networks be liable for direct, indirect, special, exemplary, incidental, or consequential damages resulting from any defect or omission in this document, even if advised of the possibility of such damages.

Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

YEAR 2000 NOTICE

Juniper Networks hardware and software products do not suffer from Year 2000 problems and hence are Year 2000 compliant. The Junos operating system has no known time-related limitations through the year 2038. However, the NTP application is known to have some difficulty in the year 2036.

SOFTWARE LICENSE

The terms and conditions for using Juniper Networks software are described in the software license provided with the software, or to the extent applicable, in an agreement executed between you and Juniper Networks, or Juniper Networks agent. By using Juniper Networks software, you indicate that you understand and agree to be bound by its license terms and conditions. Generally speaking, the software license restricts the manner in which you are permitted to use the Juniper Networks software, may contain prohibitions against certain uses, and may state conditions under which the license is automatically terminated. You should consult the software license for further details.

Contents

Course Overview

This five-day course is designed to cover best practices, theory, and design principles for Wide Area Networks (WAN) design including WAN interconnects, security considerations, virtualization, and management/operations. This course covers both service provider and enterprise WAN design.

Intended Audience

This course is targeted specifically for those who have a solid understanding of operation and configuration and are looking to enhance their skill sets by learning the principles of WAN design.

Course Level

JND-WAN is an intermediate-level course.

Prerequisites

The prerequisites for this course are as follows:

- Knowledge of routing and switching architectures and protocols.
- Knowledge of Juniper Networks products and solutions.
- Understanding of infrastructure security principles.
- Completion of the *Juniper Networks Design Fundamentals* (JNDF) course.

Objectives

After successfully completing this course, you should be able to:

- Describe high level concepts about the different WAN architectures.
- Identify key features used to interconnect WANs.
- Describe key high level considerations about securing and monitoring a WAN deployment.
- Outline high level concepts for implementing WANs.
- Explain various methods of WAN connectivity.
- Describe basic MPLS concepts as they are related to WANs.
- Identify basic Ethernet concepts as they are related to WANs.
- Describe key concepts of network availability.
- Explain high availability features and protocols.
- Describe the key aspects of class of service.
- Describe how core WAN technologies are used to solve specific problems facing network designers.
- Discuss core routing requirements.
- Explain how to design a high performance MPLS WAN core.
- Define CoS requirements for the WAN core.
- Discuss BGP peering and path selection.
- Design MPLS Layer 2 and Layer 3 services.
- Design metro Ethernet networks.
- Understand role of class of service in provider edge.
- Describe next-generation MVPNs.
- Explain how enterprise WAN technologies are used to solve specific problems facing network designers.
- Outline various solutions regarding campus and branch WANs.
- Explain how data centers are interconnected through WANs.
- Identify various solutions regarding data center WAN interconnection.

- Describe the benefits and use cases for EVPN.
- Describe security concepts regarding WANs.
- Explain the differences between LAN security concepts and WAN security concepts.
- Explain VPN-related concepts regarding WANs.
- Describe methods to manage WANs.
- Discuss key concepts related to WAN management.
- Explain how virtualization and SDN can be leveraged in the WAN.
- Describe various SDN products and how they are used in the WAN.
- Describe MX, SRX, T, PTX, ACX, QFX, EX, and NFX Series devices and the basics of how they relate to WAN solutions.

Course Agenda

Day 1

Chapter 1: Course Introduction

Chapter 2: Overview of WAN Design

Chapter 3: WAN Connectivity

Chapter 4: Network Availability and Traffic Prioritization

Lab: Network Availability and CoS Design

Day 2

Chapter 5: Service Provider Core WAN

Lab: WAN Core Design

Day 3

Chapter 6: Service Provider Edge WAN

Lab: Service Provider Edge—VPN Design

Lab: Service Provider Edge—Services Design

Day 4

Chapter 7: Enterprise WAN

Lab: Enterprise WAN Design

Chapter 8: Data Center WAN

Lab: Data Center WAN Design

Chapter 9: WAN Security

Lab: WAN Security Design

Day 5

Chapter 10: WAN Management

Lab: WAN Management Design

Chapter 11: WAN Virtualization and SDN

Lab: SDN Design

Chapter 12: WAN Device Portfolio

Document Conventions

CLI and GUI Text

Frequently throughout this course, we refer to text that appears in a command-line interface (CLI) or a graphical user interface (GUI). To make the language of these documents easier to read, we distinguish GUI and CLI text from standard text according to the following table.

Style	Description	Usage Example
Franklin Gothic	Normal text.	Most of what you read in the Lab Guide and Student Guide.

`Courier New`	Console text:
	• Screen captures
	• Noncommand-related syntax
	GUI text elements:
	• Menu names
	• Text field entry
		`commit complete`
		`Exiting configuration mode`
		Select `File > Open`, and then click `Configuration.conf` in the `Filename` text box.

Input Text Versus Output Text

You will also frequently see cases where you must enter input text yourself. Often these instances will be shown in the context of where you must enter them. We use bold style to distinguish text that is input versus text that is simply displayed.

Style	Description	Usage Example

`Normal CLI`
`Normal GUI`
	No distinguishing variant.	`Physical interface:fxp0, Enabled`
		View configuration history by clicking `Configuration > History`.

`CLI Input`
`GUI Input`
	Text that you must enter.	`lab@San_Jose>` **`show route`**
		Select `File > Save`, and type **`config.ini`** in the `Filename` field.

Defined and Undefined Syntax Variables

Finally, this course distinguishes between regular text and syntax variables, and it also distinguishes between syntax variables where the value is already assigned (defined variables) and syntax variables where you must assign the value (undefined variables). Note that these styles can be combined with the input style as well.

Style	Description	Usage Example

`CLI Variable`
`GUI Variable`
	Text where variable value is already assigned.	`policy` *`my-peers`*
		Click *`my-peers`* in the dialog.

`CLI Undefined`
`GUI Undefined`
	Text where the variable's value is the user's discretion or text where the variable's value as shown in the lab guide might differ from the value the user must input according to the lab topology.	Type **`set policy`** ***`policy-name`***.
		`ping 10.0.`***`x.y`***
		Select `File > Save`, and type ***`filename`*** in the `Filename` field.

Additional Information

Education Services Offerings

You can obtain information on the latest Education Services offerings, course dates, and class locations from the World Wide Web by pointing your Web browser to: http://www.juniper.net/training/education/.

About This Publication

The *Juniper Networks Design—WAN Student Guide* is written and maintained by the Juniper Networks Education Services development team. Please send questions and suggestions for improvement to training@juniper.net.

Technical Publications

You can print technical manuals and release notes directly from the Internet in a variety of formats:

- Go to http://www.juniper.net/techpubs/.
- Locate the specific software or hardware release and title you need, and choose the format in which you want to view or print the document.

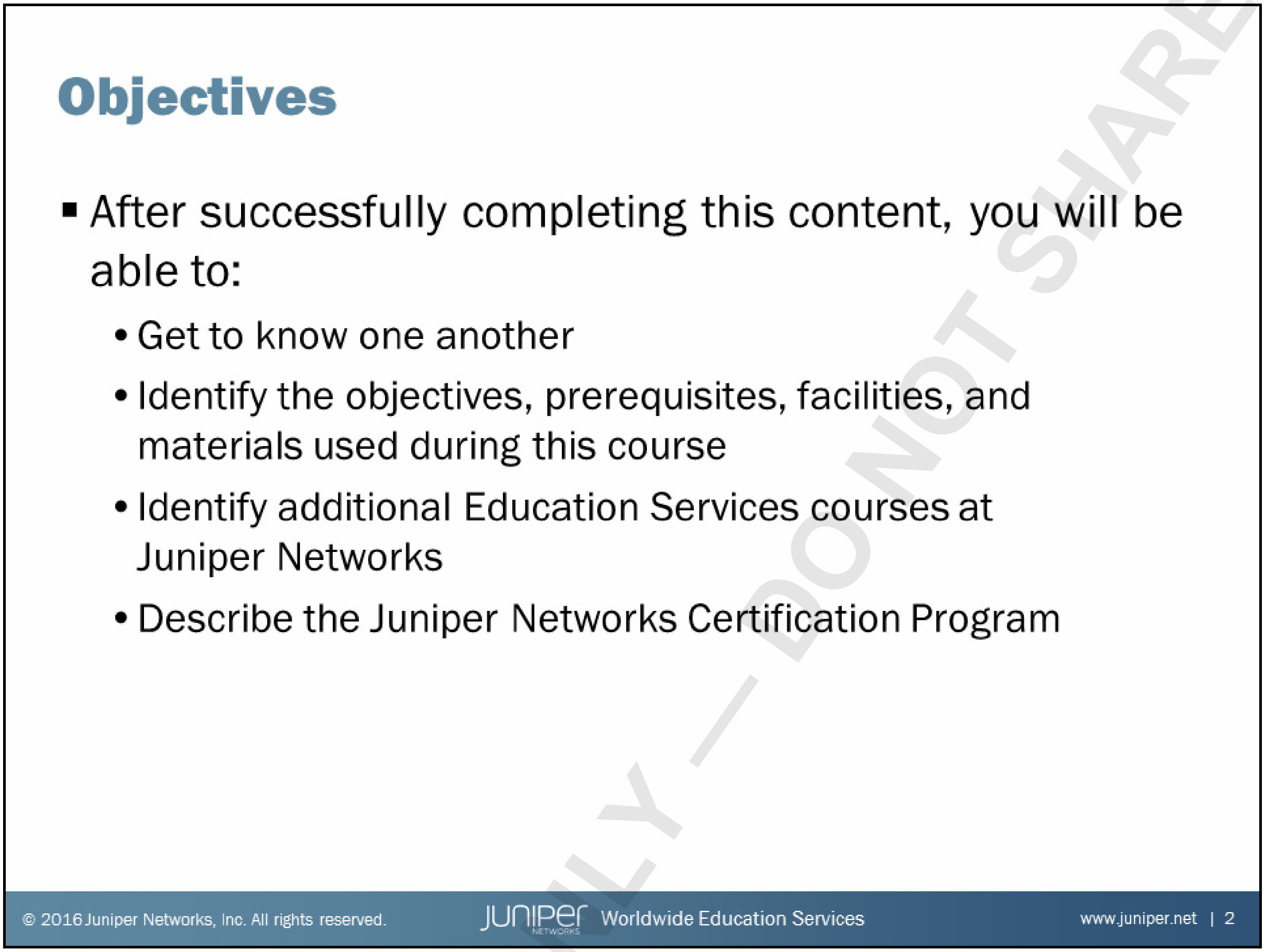
Documentation sets and CDs are available through your local Juniper Networks sales office or account representative.

Juniper Networks Support

For technical support, contact Juniper Networks at http://www.juniper.net/customers/support/, or at 1-888-314-JTAC (within the United States) or 408-745-2121 (outside the United States).

Juniper Networks Design—WAN

Chapter 1: Course Introduction

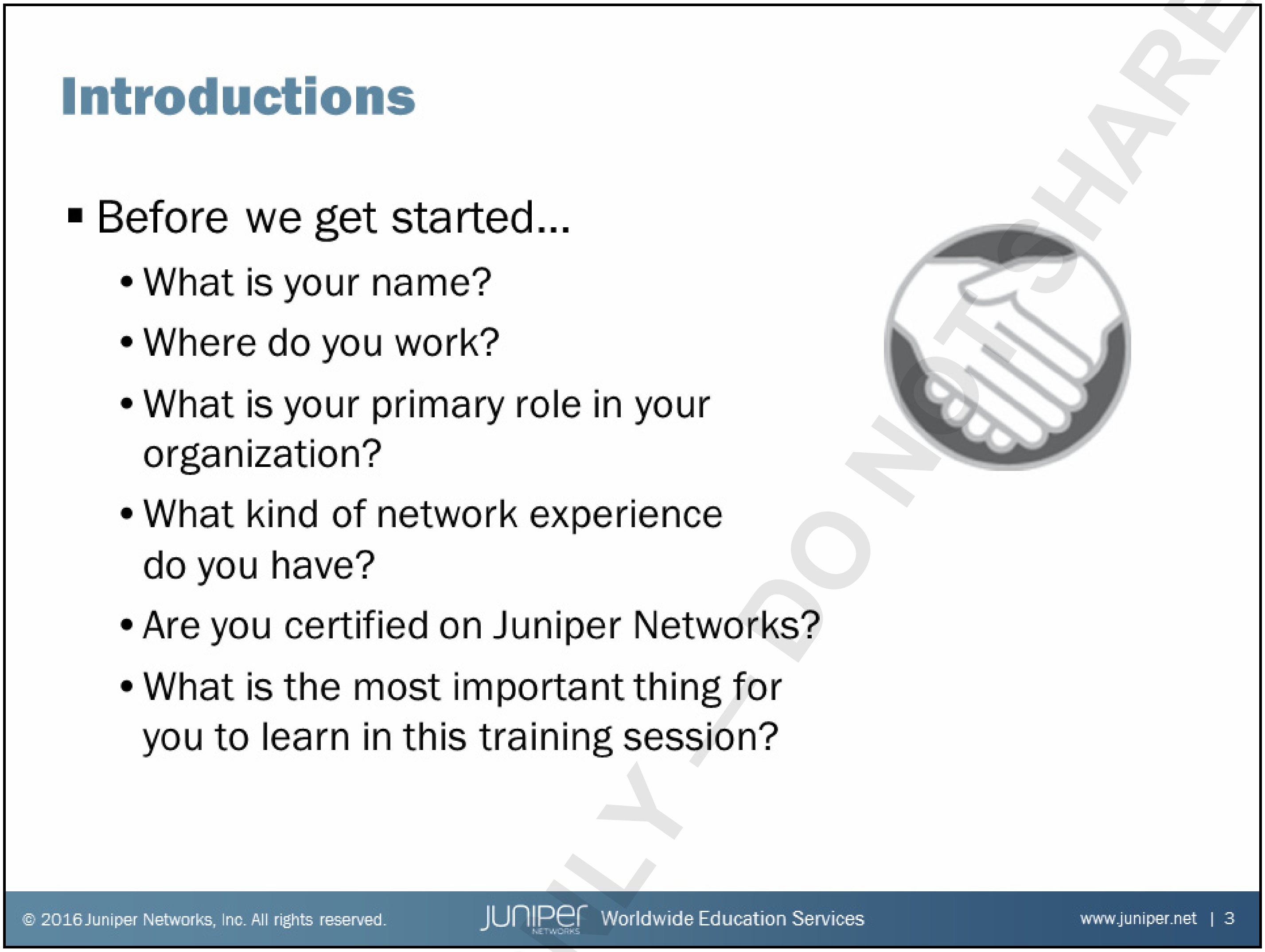


Objectives

- After successfully completing this content, you will be able to:
  - Get to know one another
  - Identify the objectives, prerequisites, facilities, and materials used during this course
  - Identify additional Education Services courses at Juniper Networks
  - Describe the Juniper Networks Certification Program

We Will Discuss:

- Objectives and course content information;
- Additional Juniper Networks, Inc. courses; and
- The Juniper Networks Certification Program.



Introductions

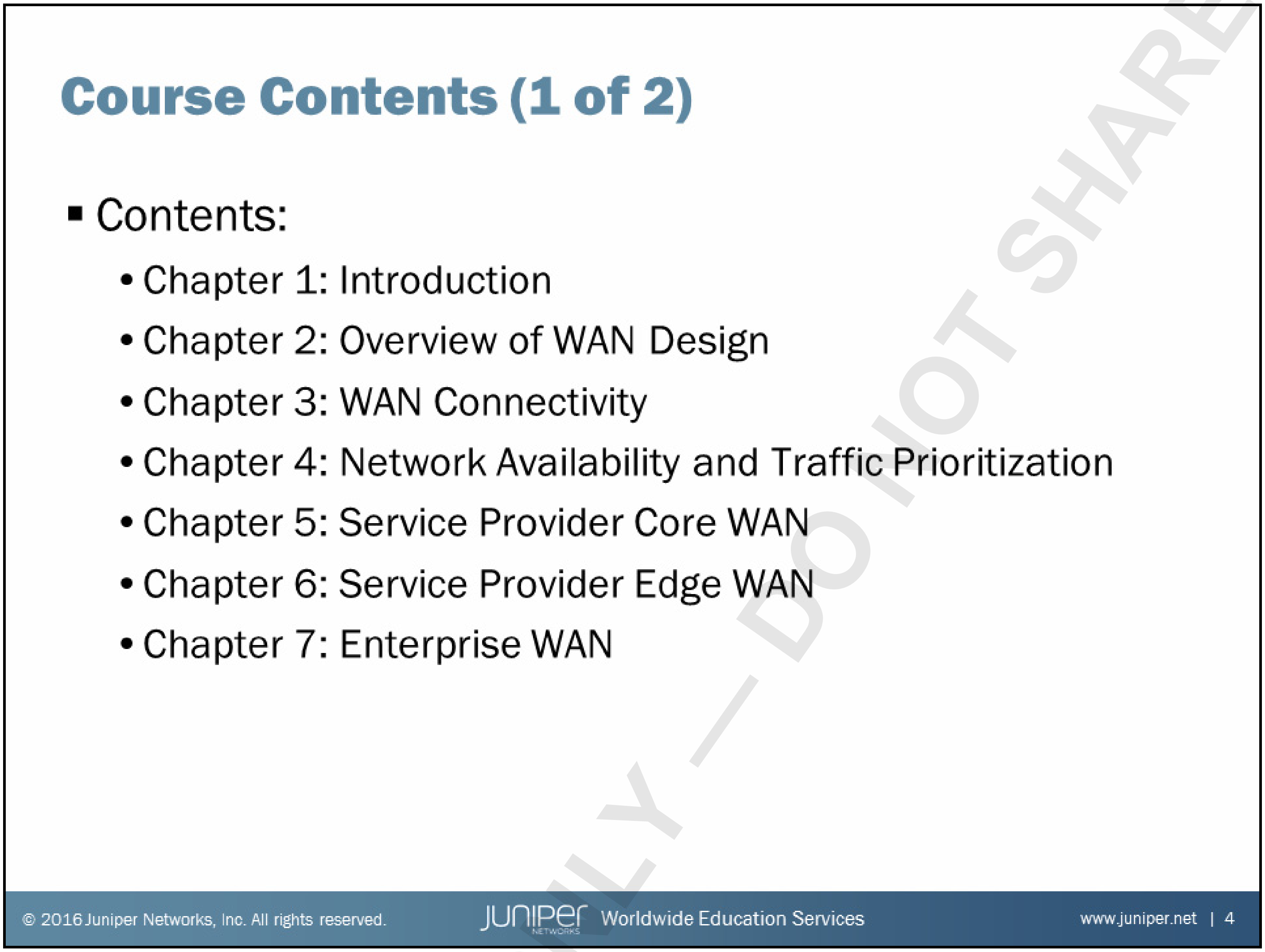
- Before we get started...
  - What is your name?
  - Where do you work?
  - What is your primary role in your organization?
  - What kind of network experience do you have?
  - Are you certified on Juniper Networks?
  - What is the most important thing for you to learn in this training session?

Worldwide Education Services

www.juniper.net | 3

Introductions

The slide asks several questions for you to answer during class introductions.



Course Contents (1 of 2)

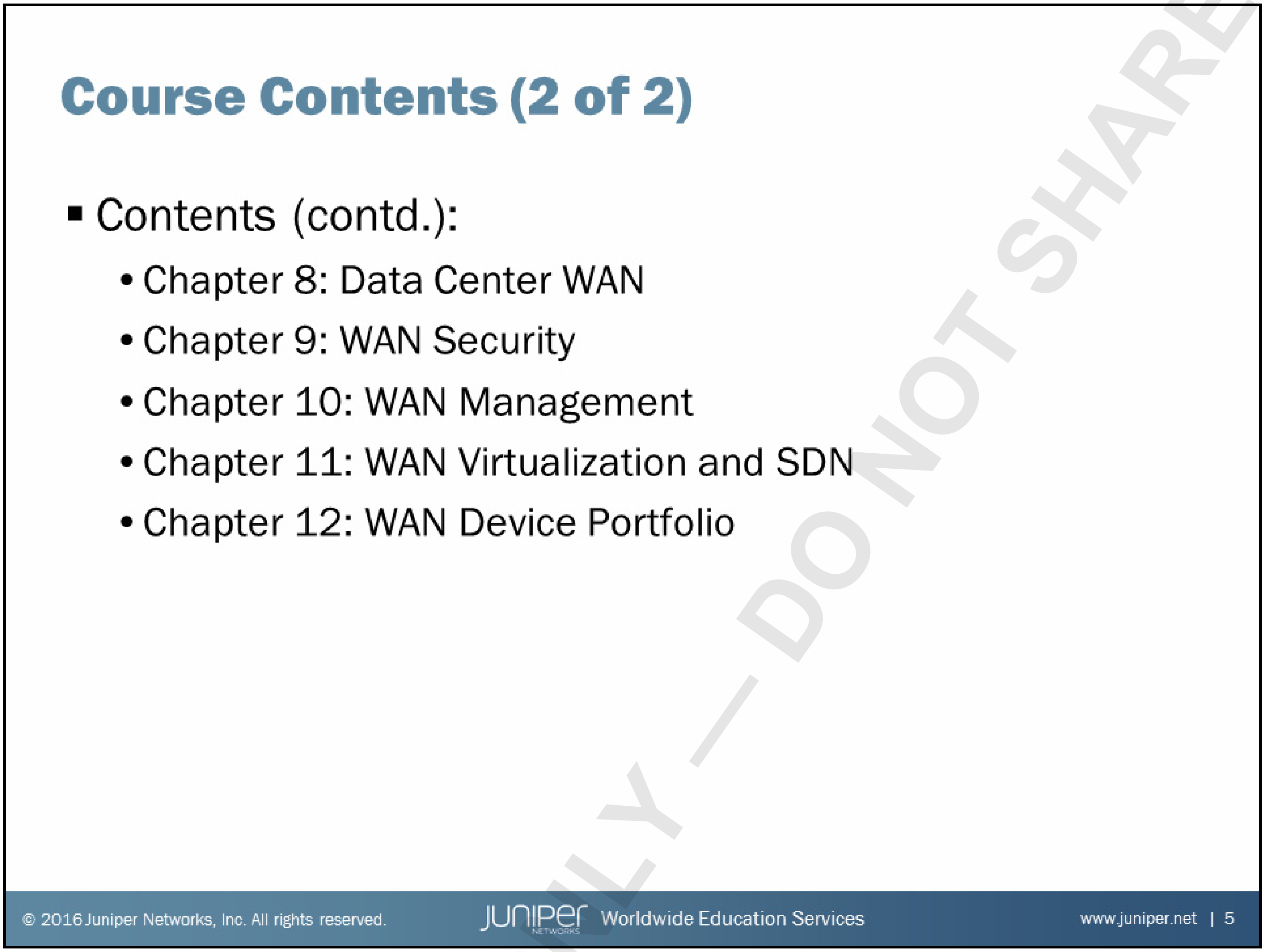
- Contents:
  - Chapter 1: Introduction
  - Chapter 2: Overview of WAN Design
  - Chapter 3: WAN Connectivity
  - Chapter 4: Network Availability and Traffic Prioritization
  - Chapter 5: Service Provider Core WAN
  - Chapter 6: Service Provider Edge WAN
  - Chapter 7: Enterprise WAN

Worldwide Education Services

www.juniper.net | 4

Course Contents: Part 1

The slide lists the topics we discuss in this course.



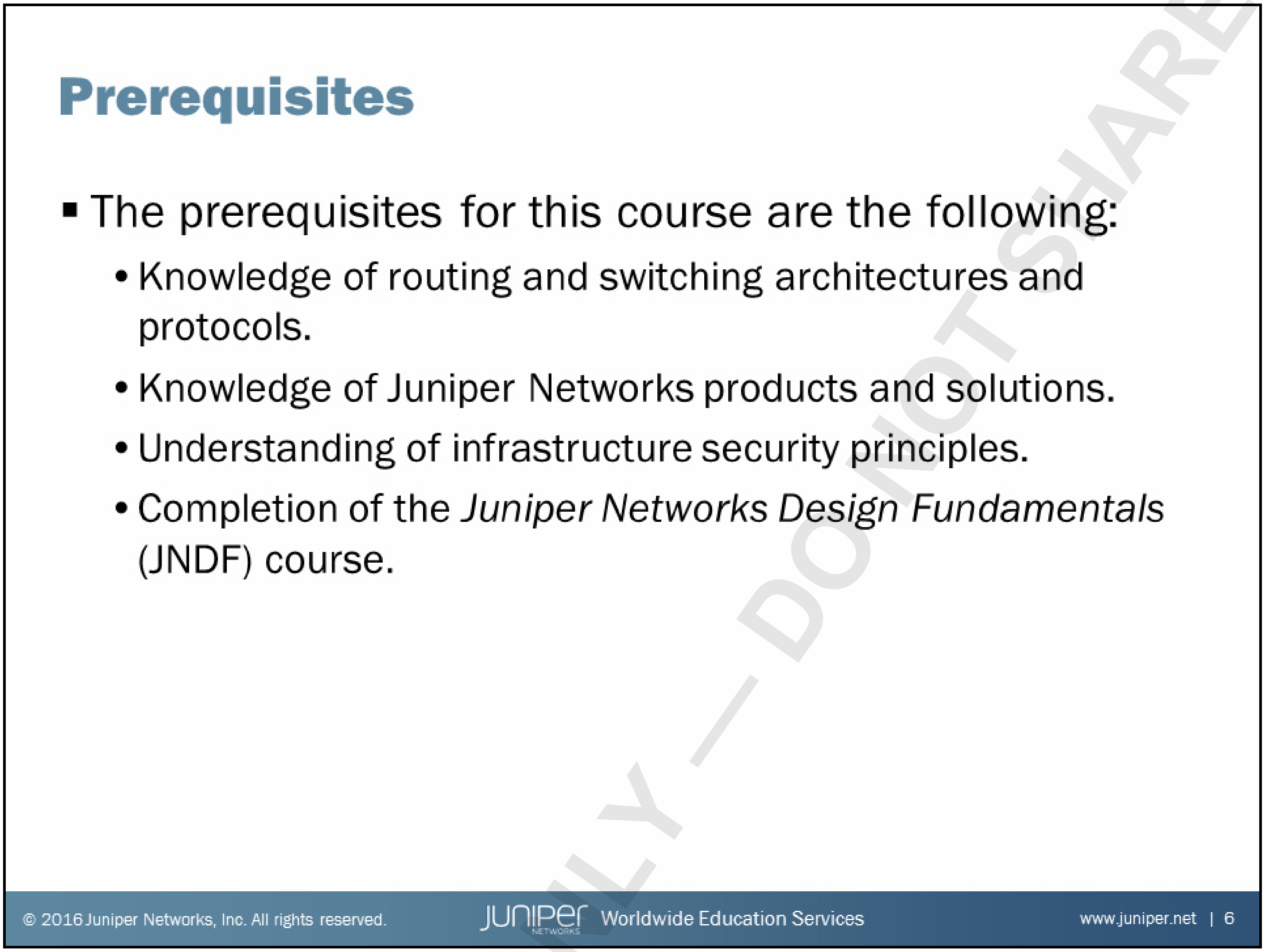
Course Contents (2 of 2)

- Contents (contd.):
  - Chapter 8: Data Center WAN
  - Chapter 9: WAN Security
  - Chapter 10: WAN Management
  - Chapter 11: WAN Virtualization and SDN
  - Chapter 12: WAN Device Portfolio

Worldwide Education Services

Course Contents: Part 2

The slide lists the remainder of the topics we discuss in this course.



Prerequisites

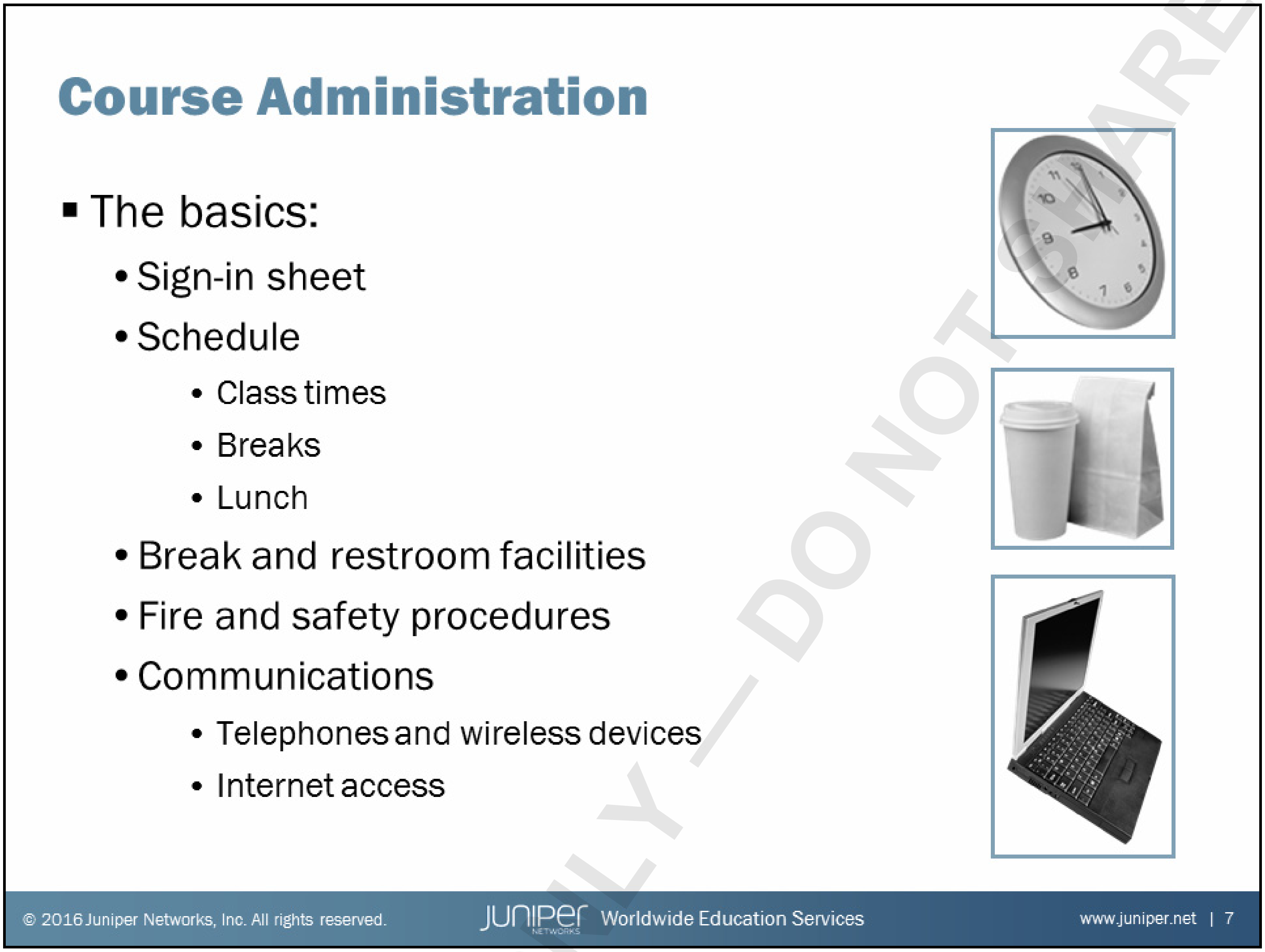
- The prerequisites for this course are the following:
  - Knowledge of routing and switching architectures and protocols.
  - Knowledge of Juniper Networks products and solutions.
  - Understanding of infrastructure security principles.
  - Completion of the *Juniper Networks Design Fundamentals* (JNDF) course.

Worldwide Education Services

www.juniper.net | 6

Prerequisites

The slide lists the prerequisites for this course.



Course Administration

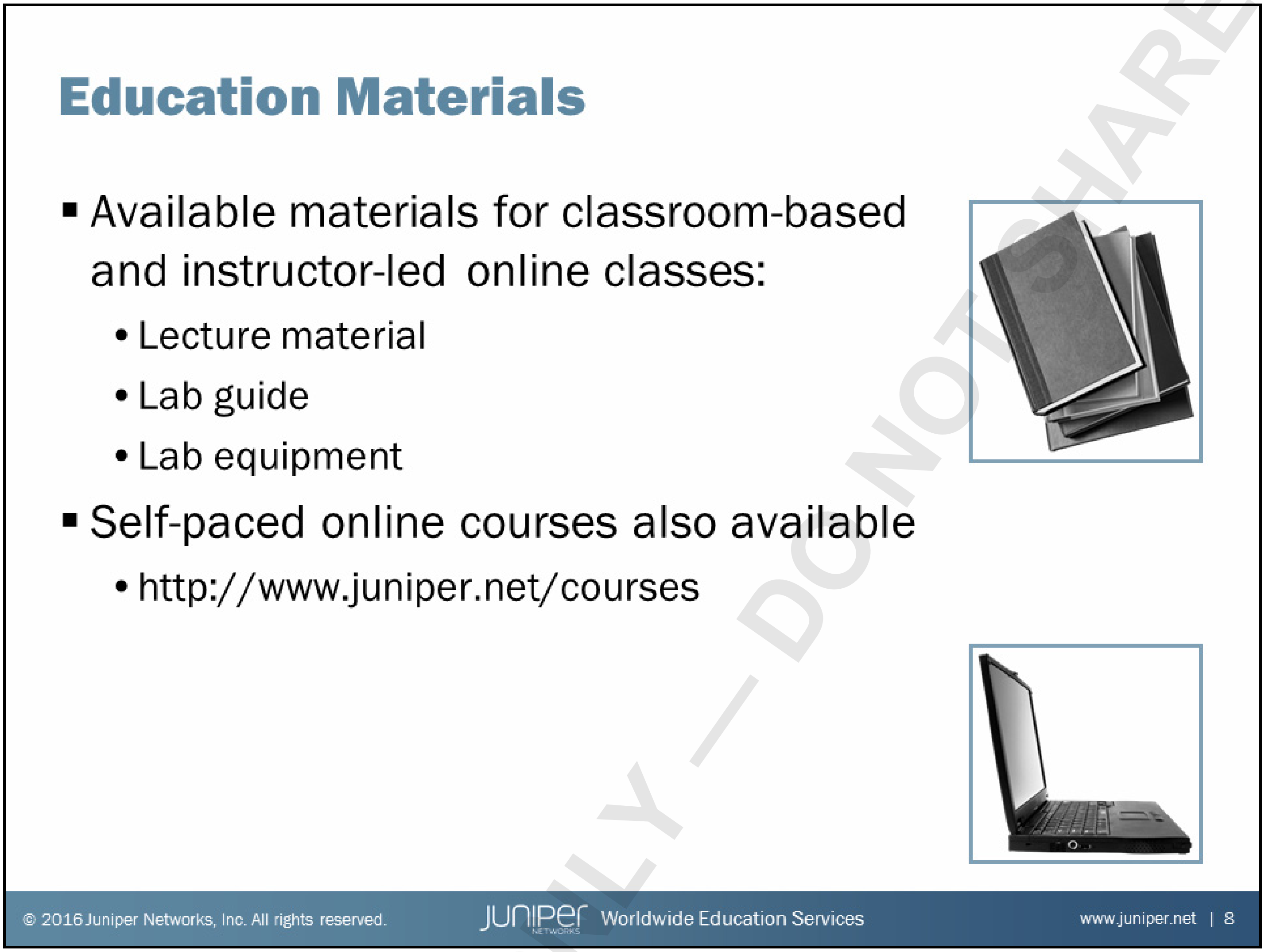
- The basics:
  - Sign-in sheet
  - Schedule
    - Class times
    - Breaks
    - Lunch
  - Break and restroom facilities
  - Fire and safety procedures
  - Communications
    - Telephones and wireless devices
    - Internet access

Worldwide Education Services

www.juniper.net | 7

General Course Administration

The slide documents general aspects of classroom administration.



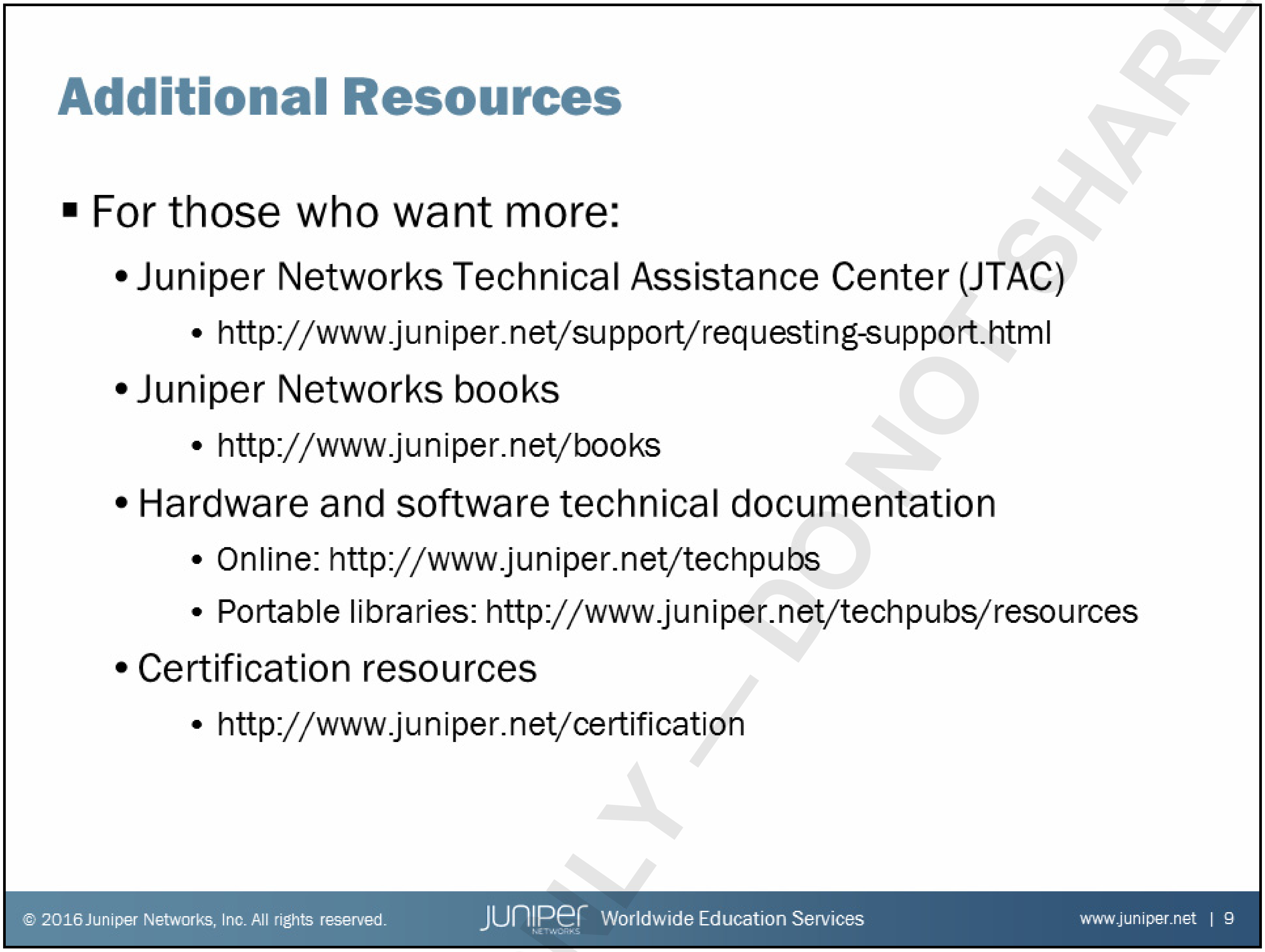
Education Materials

- Available materials for classroom-based and instructor-led online classes:
  - Lecture material
  - Lab guide
  - Lab equipment
- Self-paced online courses also available
  - http://www.juniper.net/courses

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 8

Training and Study Materials

The slide describes Education Services materials that are available for reference both in the classroom and online.



Additional Resources

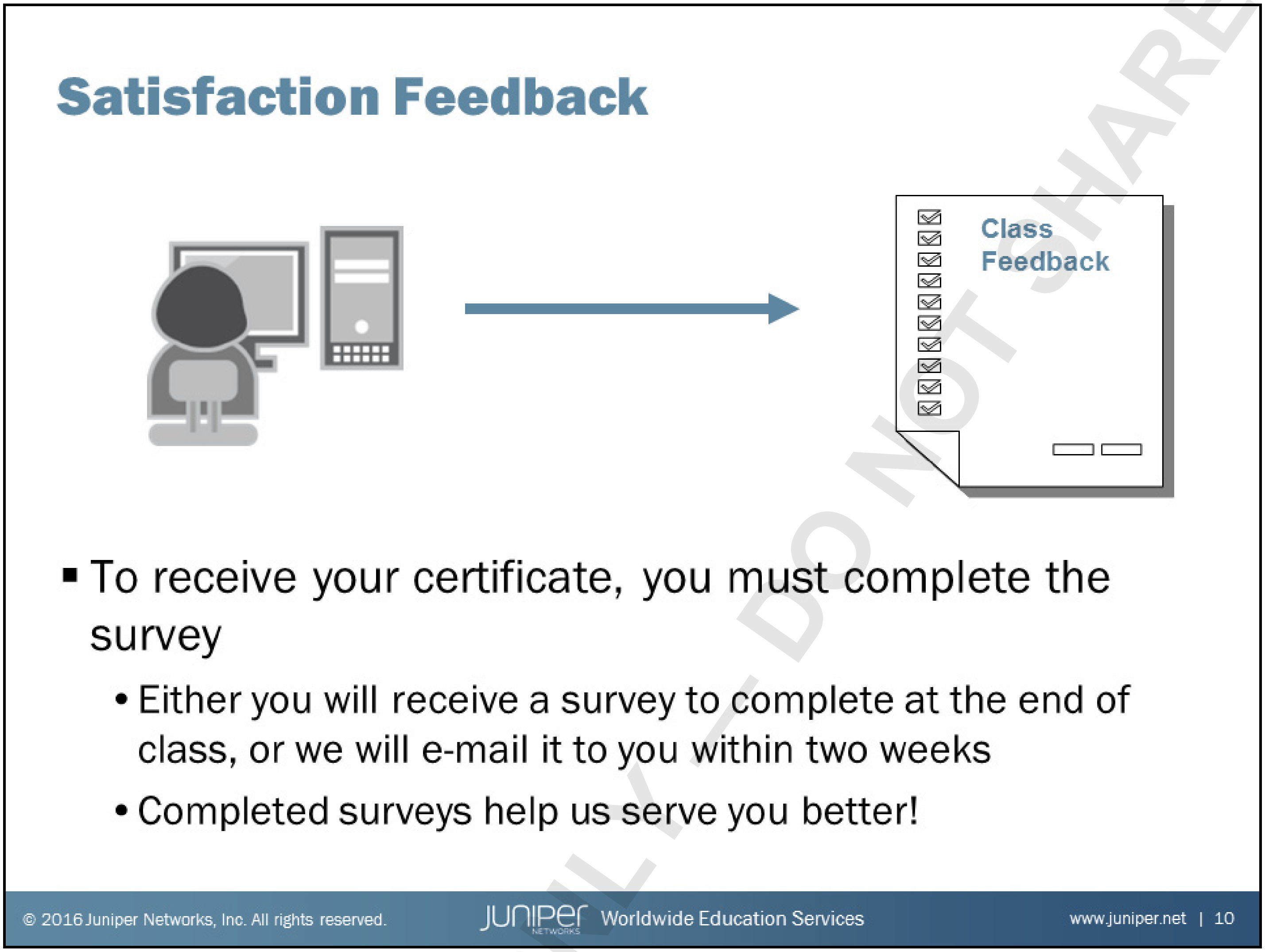
- For those who want more:
  - Juniper Networks Technical Assistance Center (JTAC)
    - http://www.juniper.net/support/requesting-support.html
  - Juniper Networks books
    - http://www.juniper.net/books
  - Hardware and software technical documentation
    - Online: http://www.juniper.net/techpubs
    - Portable libraries: http://www.juniper.net/techpubs/resources
  - Certification resources
    - http://www.juniper.net/certification

Juniper Networks Worldwide Education Services

www.juniper.net | 9

Additional Resources

The slide provides links to additional resources available to assist you in the installation, configuration, and operation of Juniper Networks products.



Satisfaction Feedback

Class
Feedback

- To receive your certificate, you must complete the survey
  - Either you will receive a survey to complete at the end of class, or we will e-mail it to you within two weeks
  - Completed surveys help us serve you better!

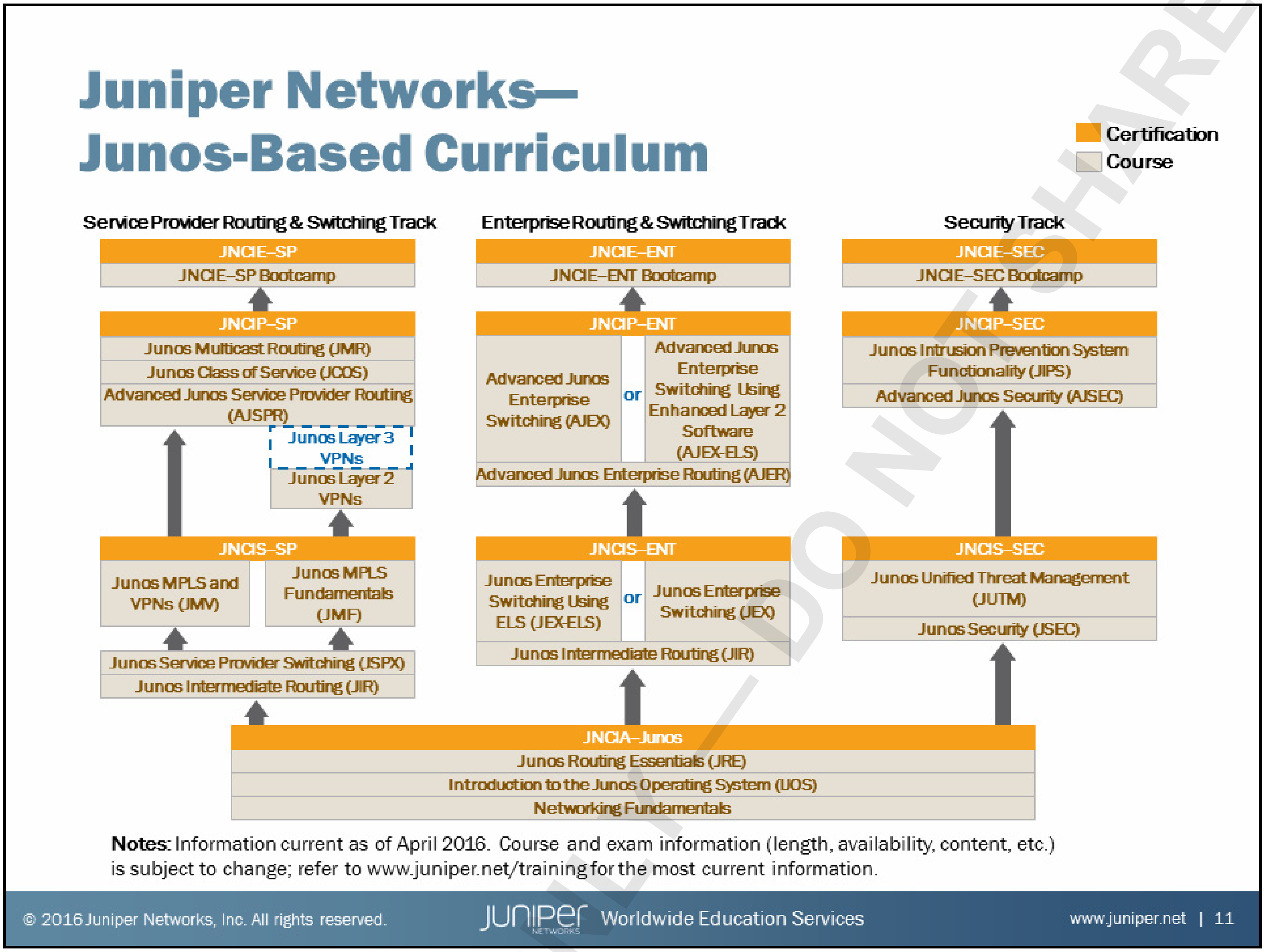
Worldwide Education Services

www.juniper.net | 10

Satisfaction Feedback

Juniper Networks uses an electronic survey system to collect and analyze your comments and feedback. Depending on the class you are taking, please complete the survey at the end of the class, or be sure to look for an e-mail about two weeks from class completion that directs you to complete an online survey form. (Be sure to provide us with your current e-mail address.)

Submitting your feedback entitles you to a certificate of class completion. We thank you in advance for taking the time to help us improve our educational offerings.



Juniper Networks— Junos-Based Curriculum

🟧 Certification
□ Course

Service Provider Routing & Switching Track

- JNCIE–SP
- JNCIE–SP Bootcamp
- JNCIP–SP
- Junos Multicast Routing (JMR)
- Junos Class of Service (JCOS)
- Advanced Junos Service Provider Routing (AJSPR)
- Junos Layer 3 VPNs
- Junos Layer 2 VPNs
- JNCIS–SP
- Junos MPLS and VPNs (JMV)
- Junos MPLS Fundamentals (JMF)
- Junos Service Provider Switching (JSPX)
- Junos Intermediate Routing (JIR)

Enterprise Routing & Switching Track

- JNCIE–ENT
- JNCIE–ENT Bootcamp
- JNCIP–ENT
- Advanced Junos Enterprise Switching (AJEX) or Advanced Junos Enterprise Switching Using Enhanced Layer 2 Software (AJEX-ELS)
- Advanced Junos Enterprise Routing (AJER)
- JNCIS–ENT
- Junos Enterprise Switching Using ELS (JEX-ELS) or Junos Enterprise Switching (JEX)
- Junos Intermediate Routing (JIR)

Security Track

- JNCIE–SEC
- JNCIE–SEC Bootcamp
- JNCIP–SEC
- Junos Intrusion Prevention System Functionality (JIPS)
- Advanced Junos Security (AJSEC)
- JNCIS–SEC
- Junos Unified Threat Management (JUTM)
- Junos Security (JSEC)

JNCIA–Junos

- Junos Routing Essentials (JRE)
- Introduction to the Junos Operating System (IJOS)
- Networking Fundamentals

Notes: Information current as of April 2016. Course and exam information (length, availability, content, etc.) is subject to change; refer to www.juniper.net/training for the most current information.

Juniper Networks Worldwide Education Services

www.juniper.net | 11

Juniper Networks Education Services Curriculum

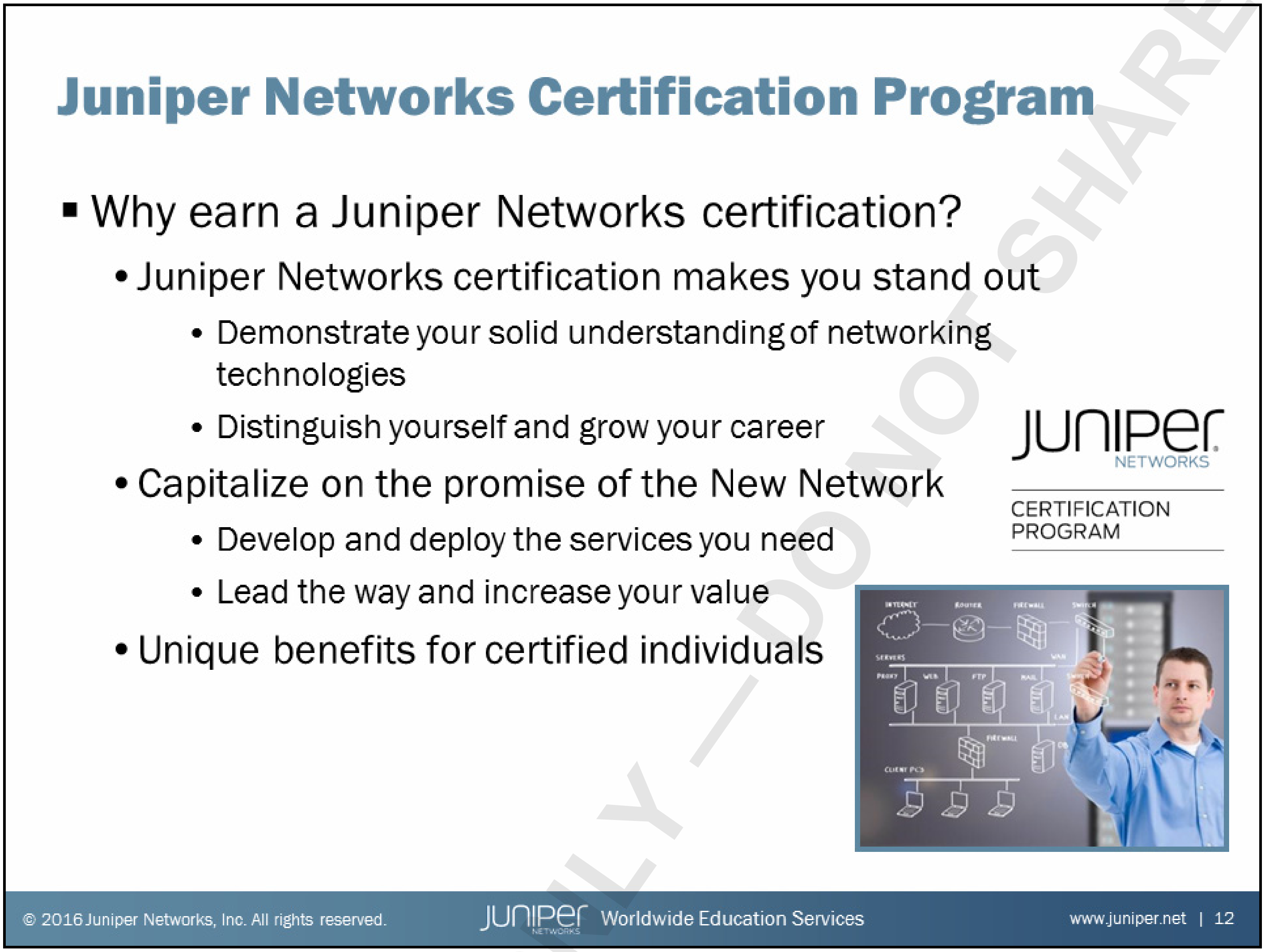
Juniper Networks Education Services can help ensure that you have the knowledge and skills to deploy and maintain cost-effective, high-performance networks for both enterprise and service provider environments. We have expert training staff with deep technical and industry knowledge, providing you with instructor-led hands-on courses in the classroom and online, as well as convenient, self-paced eLearning courses. In addition to the courses shown on the slide, Education Services offers training in automation, E-Series, firewall/VPN, IDP, network design, QFabric, support, and wireless LAN.

Courses

Juniper Networks courses are available in the following formats:

- Classroom-based instructor-led technical courses
- Online instructor-led technical courses
- Hardware installation eLearning courses as well as technical eLearning courses
- Learning bytes: Short, topic-specific, video-based lessons covering Juniper products and technologies

Find the latest Education Services offerings covering a wide range of platforms at http://www.juniper.net/training/technical_education/.



Juniper Networks Certification Program

- Why earn a Juniper Networks certification?
  - Juniper Networks certification makes you stand out
    - Demonstrate your solid understanding of networking technologies
    - Distinguish yourself and grow your career
  - Capitalize on the promise of the New Network
    - Develop and deploy the services you need
    - Lead the way and increase your value
  - Unique benefits for certified individuals

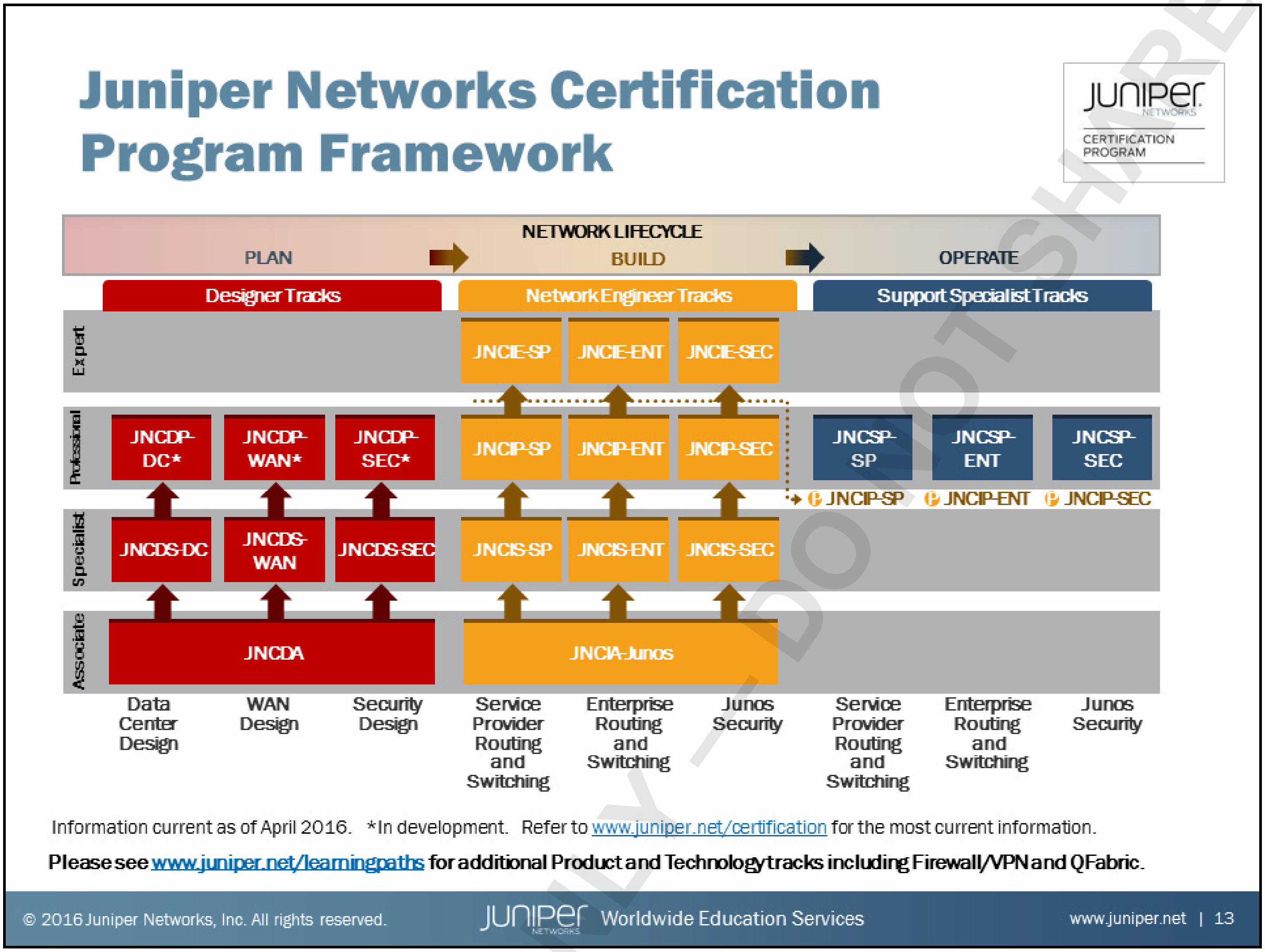
CERTIFICATION PROGRAM

Worldwide Education Services

www.juniper.net | 12

Juniper Networks Certification Program

A Juniper Networks certification is the benchmark of skills and competence on Juniper Networks technologies.



Juniper Networks Certification Program Framework

NETWORK LIFECYCLE

PLAN → BUILD → OPERATE

	Designer Tracks			Network Engineer Tracks			Support Specialist Tracks		
Expert				JNCIE-SP	JNCIE-ENT	JNCIE-SEC			
Professional	JNCDP-DC*	JNCDP-WAN*	JNCDP-SEC*	JNCIP-SP	JNCIP-ENT	JNCIP-SEC	JNCSP-SP JNCIP-SP	JNCSP-ENT JNCIP-ENT	JNCSP-SEC JNCIP-SEC
Specialist	JNCDS-DC	JNCDS-WAN	JNCDS-SEC	JNCIS-SP	JNCIS-ENT	JNCIS-SEC			
Associate	JNCDA			JNCIA-Junos					
	Data Center Design	WAN Design	Security Design	Service Provider Routing and Switching	Enterprise Routing and Switching	Junos Security	Service Provider Routing and Switching	Enterprise Routing and Switching	Junos Security

Information current as of April 2016. *In development. Refer to www.juniper.net/certification for the most current information.

Please see www.juniper.net/learningpaths for additional Product and Technology tracks including Firewall/VPN and QFabric.

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 13

Juniper Networks Certification Program Overview

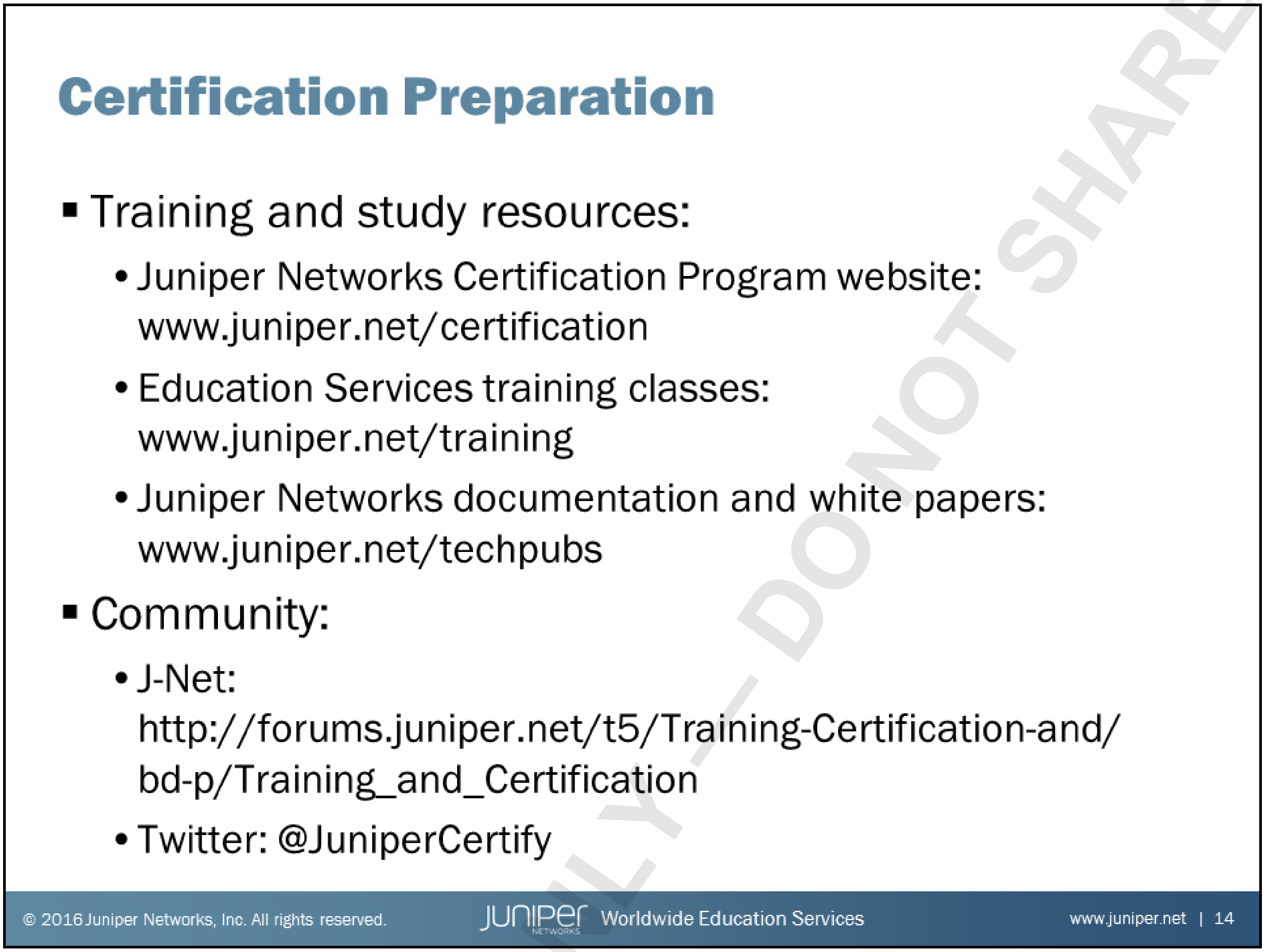
The Juniper Networks Certification Program (JNCP) consists of platform-specific, multitiered tracks that enable participants to demonstrate competence with Juniper Networks technology through a combination of written proficiency exams and hands-on configuration and troubleshooting exams. Successful candidates demonstrate a thorough understanding of Internet and security technologies and Juniper Networks platform configuration and troubleshooting skills.

The JNCP offers the following features:

- Multiple tracks;
- Multiple certification levels;
- Written proficiency exams; and
- Hands-on configuration and troubleshooting exams.

Each JNCP track has one to four certification levels—Associate-level, Specialist-level, Professional-level, and Expert-level. The Associate-level, Specialist-level, and Professional-level exams are computer-based exams composed of multiple choice questions administered at Pearson VUE testing centers worldwide.

Expert-level exams are composed of hands-on lab exercises administered at select Juniper Networks testing centers. Please visit the JNCP website at http://www.juniper.net/certification for detailed exam information, exam pricing, and exam registration.



Certification Preparation

- Training and study resources:
  - Juniper Networks Certification Program website: www.juniper.net/certification
  - Education Services training classes: www.juniper.net/training
  - Juniper Networks documentation and white papers: www.juniper.net/techpubs
- Community:
  - J-Net: http://forums.juniper.net/t5/Training-Certification-and/bd-p/Training_and_Certification
  - Twitter: @JuniperCertify

Worldwide Education Services

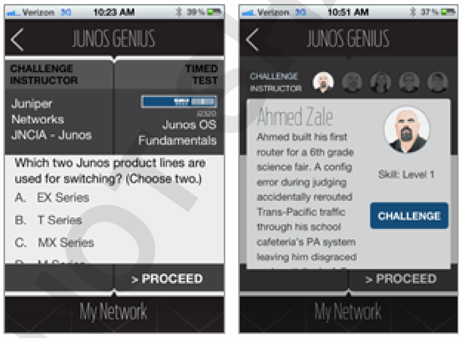
www.juniper.net | 14

Preparing and Studying

The slide lists some options for those interested in preparing for Juniper Networks certification.

Junos Genius: Certification Preparation App
Unlock your Genius...

- Practice for multiple exams in *Study Mode*
  - Hundreds of multiple choice questions and answer explanations, many with CLI snapshots
- Simulate an exam in *Time Challenge Mode*
- Earn device achievements by winning in *Instructor Challenge Mode*
- Build a virtual network with device achievements
- Track your results in the app and Game Center; share your network through Facebook and Twitter



Verizon 3G 10:23 AM 39%

JUNOS GENIUS

CHALLENGE INSTRUCTOR

TIMED TEST

Juniper Networks JNCIA - Junos

Junos OS Fundamentals

Which two Junos product lines are used for switching? (Choose two.)

A. EX Series

B. T Series

C. MX Series

> PROCEED

My Network

Verizon 3G 10:51 AM 37%

JUNOS GENIUS

CHALLENGE INSTRUCTOR

Ahmed Zale

Ahmed built his first router for a 6th grade science fair. A config error during judging accidentally rerouted Trans-Pacific traffic through his school cafeteria's PA system leaving him disgraced

Skill: Level 1

CHALLENGE

> PROCEED

My Network

ANDROID APP ON Google play

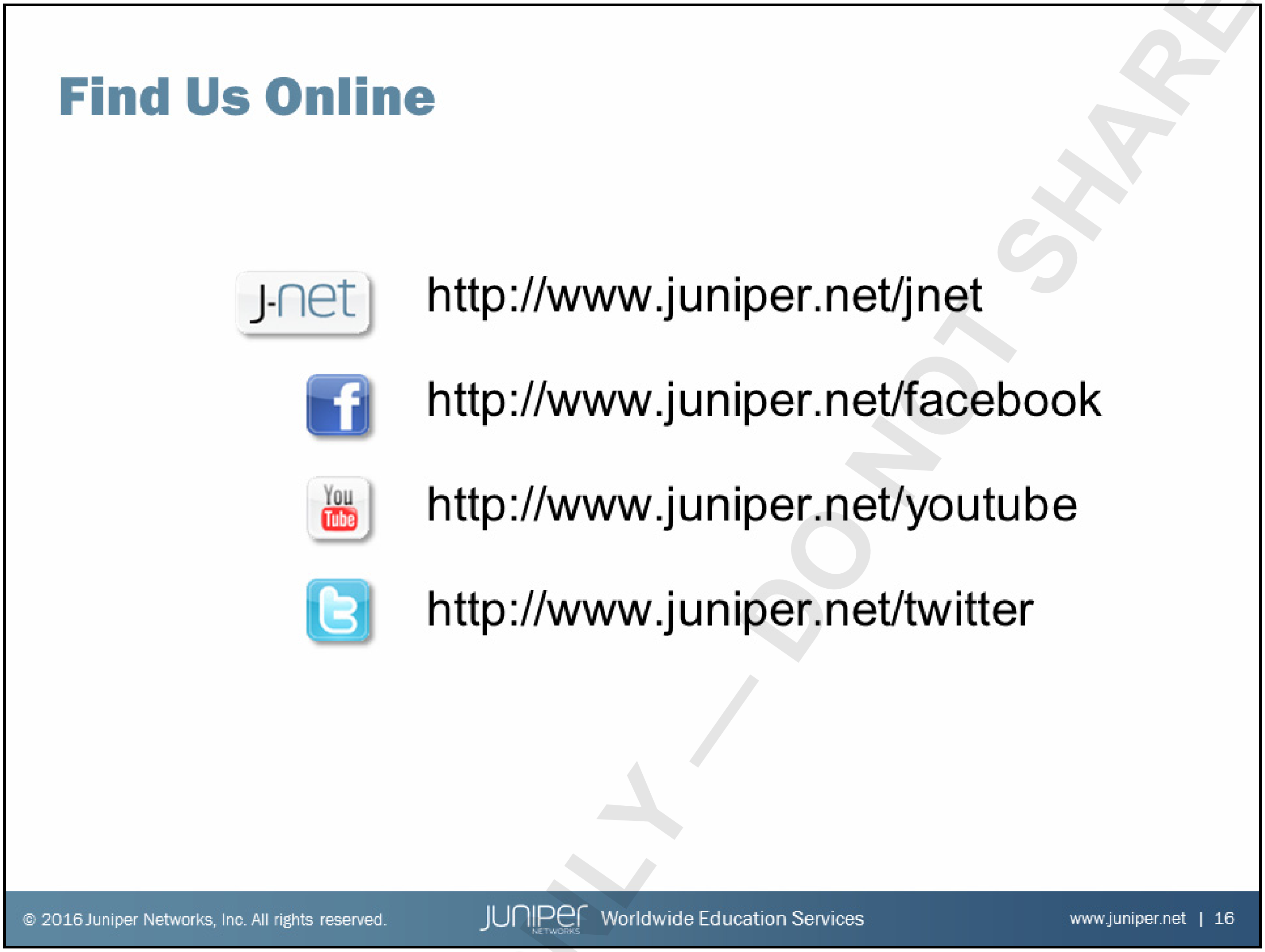
Available on the App Store

JUNOS GENIUS

www.juniper.net/junosgenius

Junos Genius

The Junos Genius application takes certification exam preparation to a new level. With Junos Genius you can practice for your exam with flashcards, simulate a live exam in a timed challenge, and even build a virtual network with device achievements earned by challenging Juniper instructors. Download the app now and *Unlock your Genius today!*



Find Us Online

J-net http://www.juniper.net/jnet

f http://www.juniper.net/facebook

You Tube http://www.juniper.net/youtube

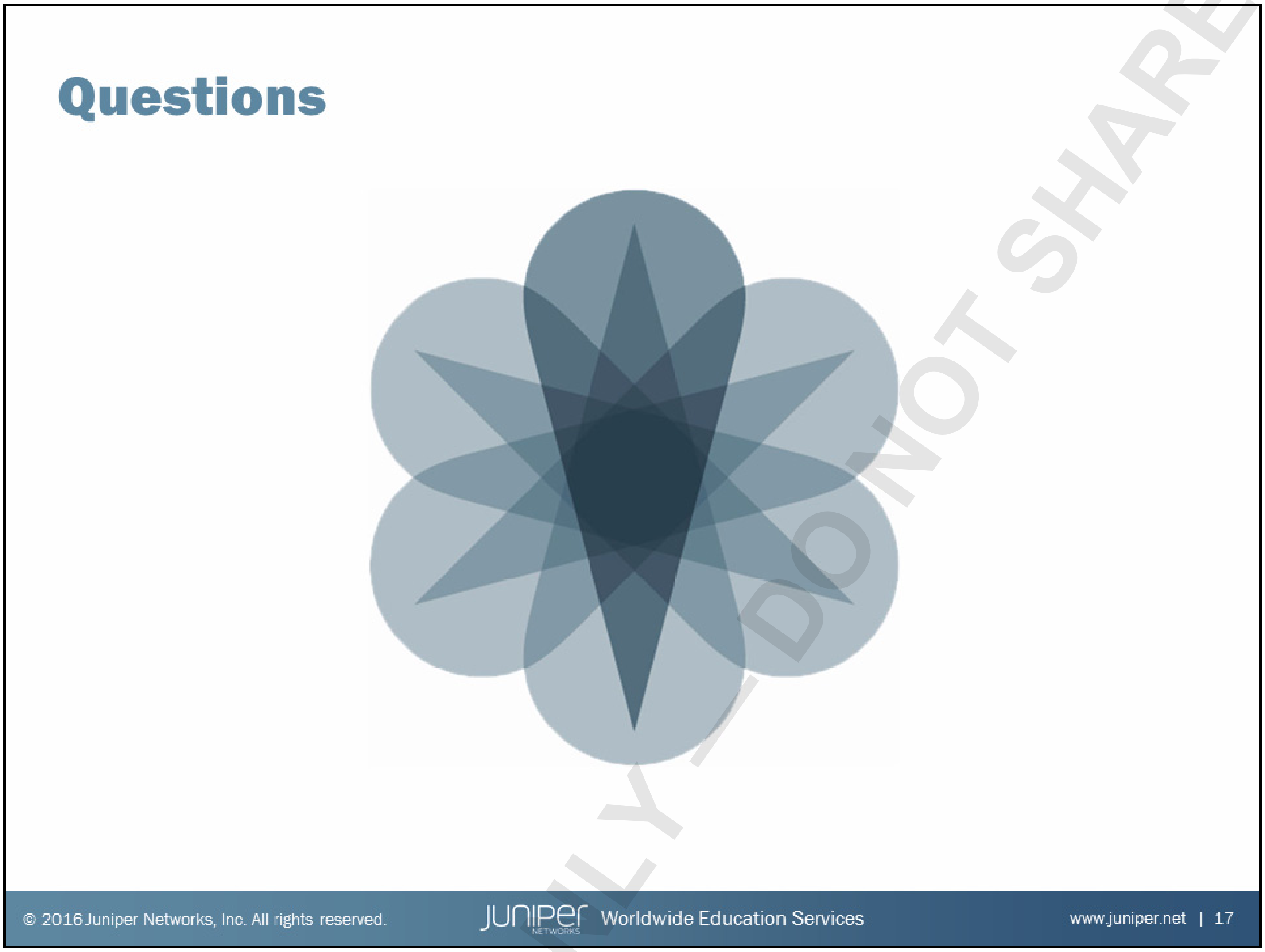
http://www.juniper.net/twitter

JUNIPER NETWORKS Worldwide Education Services

www.juniper.net | 16

Find Us Online

The slide lists some online resources to learn and share information about Juniper Networks.



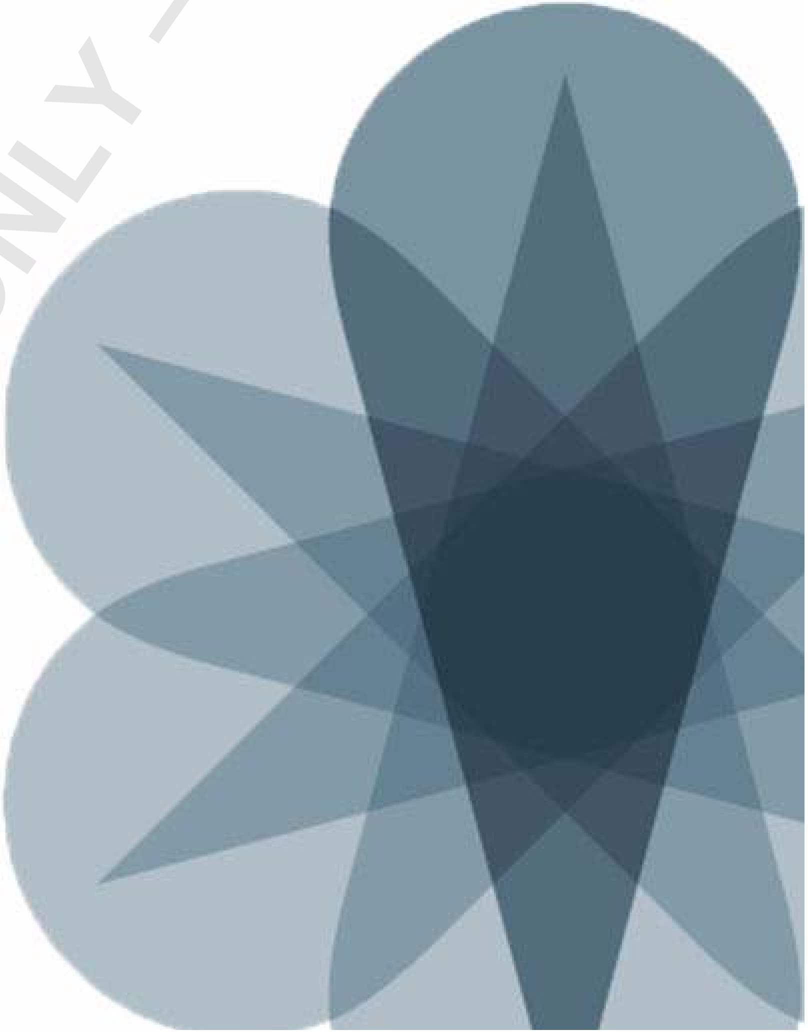
Questions

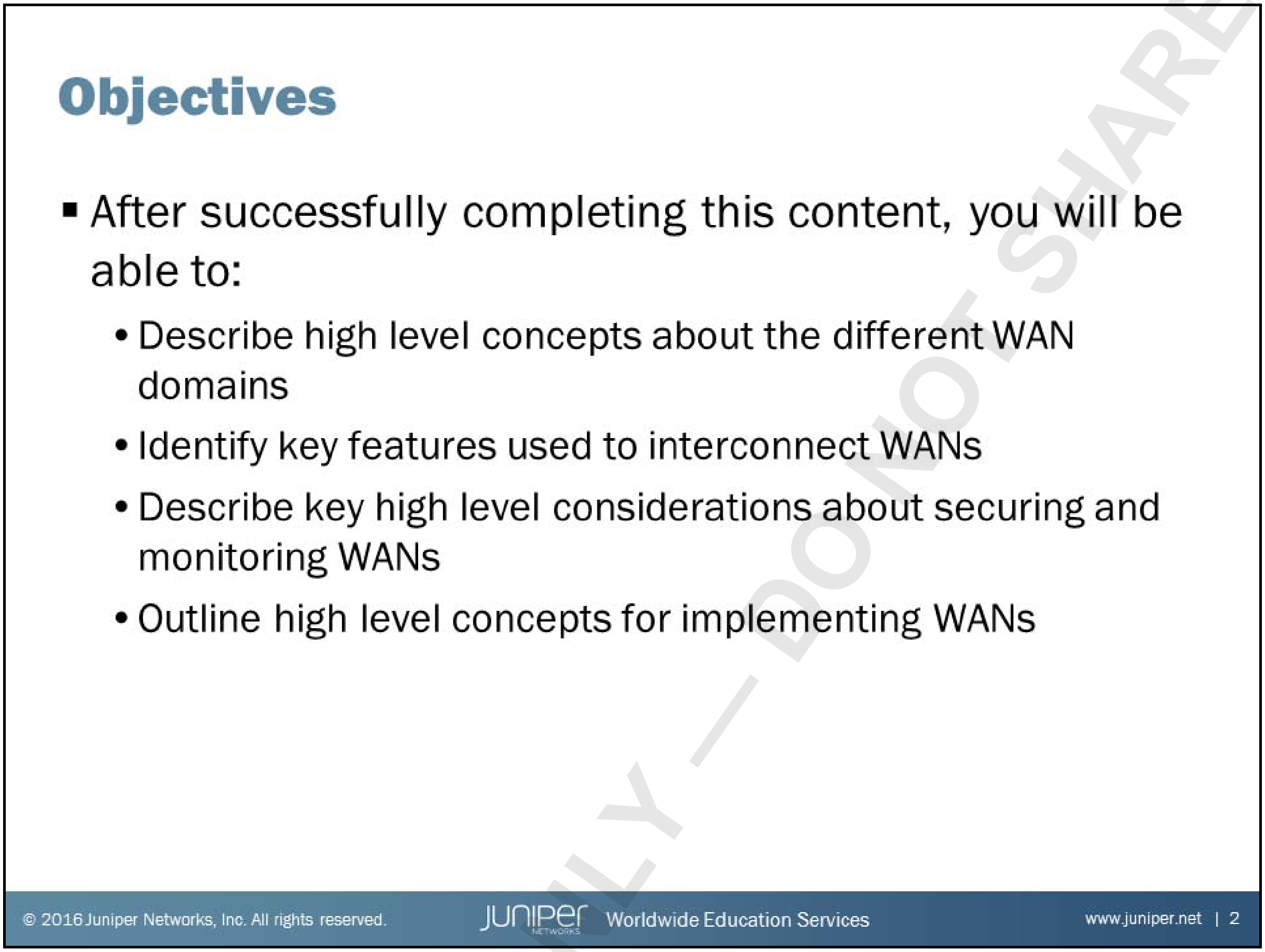
Any Questions?

If you have any questions or concerns about the class you are attending, we suggest that you voice them now so that your instructor can best address your needs during class.

Juniper Networks Design—WAN

Chapter 2: Overview of WAN Design





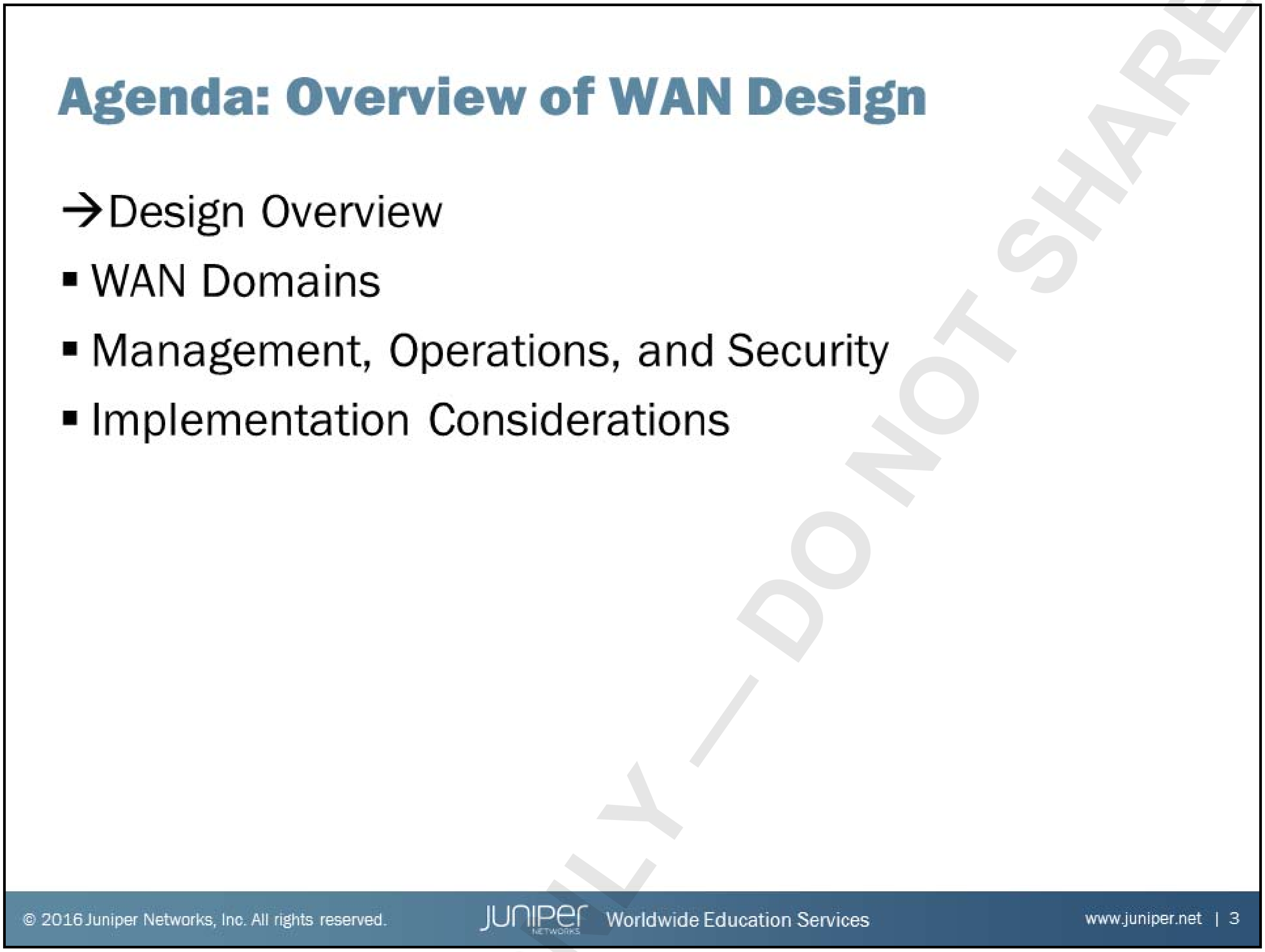
Objectives

- After successfully completing this content, you will be able to:
  - Describe high level concepts about the different WAN domains
  - Identify key features used to interconnect WANs
  - Describe key high level considerations about securing and monitoring WANs
  - Outline high level concepts for implementing WANs

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 2

We Will Discuss:

- High level concepts about the different WAN domains;
- Features used to interconnect WANs;
- Key high level considerations about securing and monitoring WANs; and
- High level concepts for implementing WANs.



Agenda: Overview of WAN Design

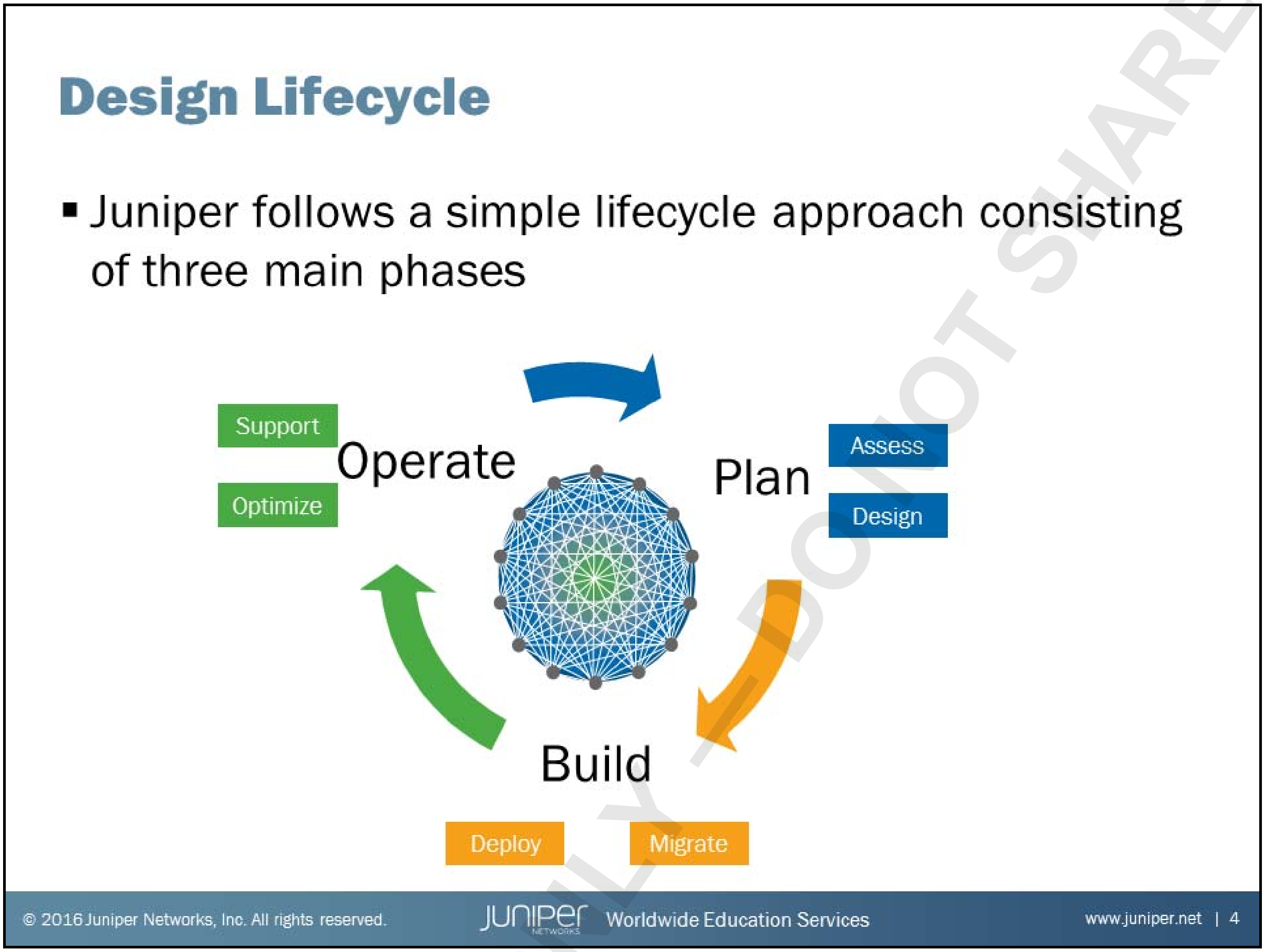
- →Design Overview
- WAN Domains
- Management, Operations, and Security
- Implementation Considerations

Worldwide Education Services

www.juniper.net | 3

Design Overview

The slide lists the topics we will discuss. We discuss the highlighted topic first.



Design Lifecycle

- Juniper follows a simple lifecycle approach consisting of three main phases

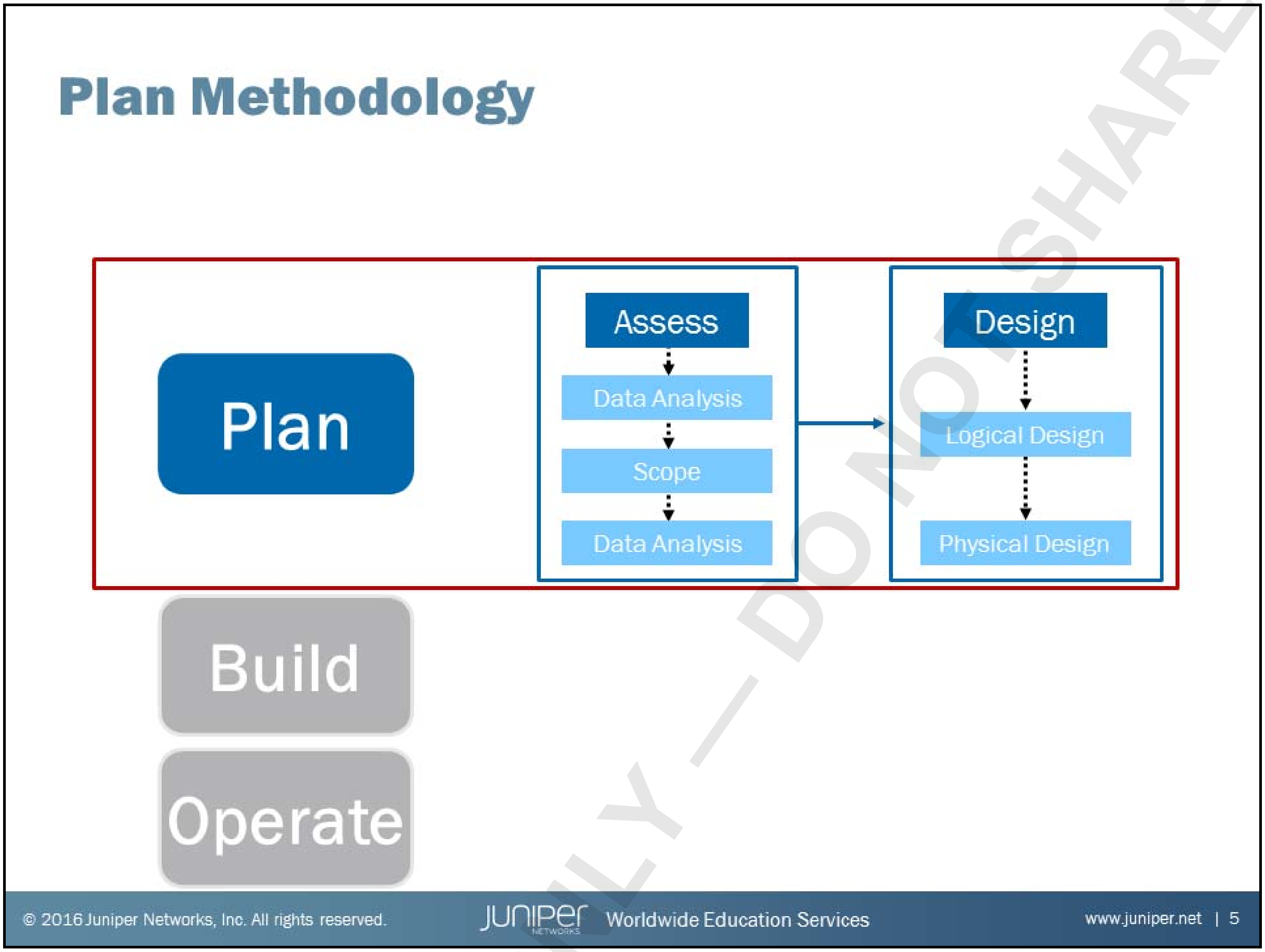
Juniper Worldwide Education Services

www.juniper.net | 4

Design Lifecycle

Juniper follows a simple lifecycle approach when designing a network for a customer. This approach consists of three main phases and is often cyclical through the lifetime of the design:

1. *Plan*:
   - a. *Assess* the current environment and its ability to satisfy the customer's business and technology requirements.
   - b. *Design* high-level architectural plans, as well as low-level detailed plans of network devices, configurations, and interconnections.
2. *Build*:
   - a. *Deploy* the design in both test and production environments.
   - b. *Migrate* from the existing environment to the new environment. Provide installation and configuration, system testing, and system enablement.
3. *Operate:*
   - a. *Support* the customer by focusing on the most effective use of the solution. Provide support for any issues and faults, as well as proactive maintenance.
   - b. *Optimize* the network as more systems and users come online, and as system usage increases.



Plan Methodology

Plan Methodology

In this course, we focus primarily on the design—or *Plan*—phase of the service lifecycle. This phase consists of two main sub-phases:

1. *Assess* the current environment and its ability to satisfy the customer's business and technology requirements.

   a. Identify the technology shortfalls that need to be addressed as well as develop a core technology roadmap that will achieve the customer's required end-game environment. Evaluate what is necessary for migrating successfully from one environment to the other.

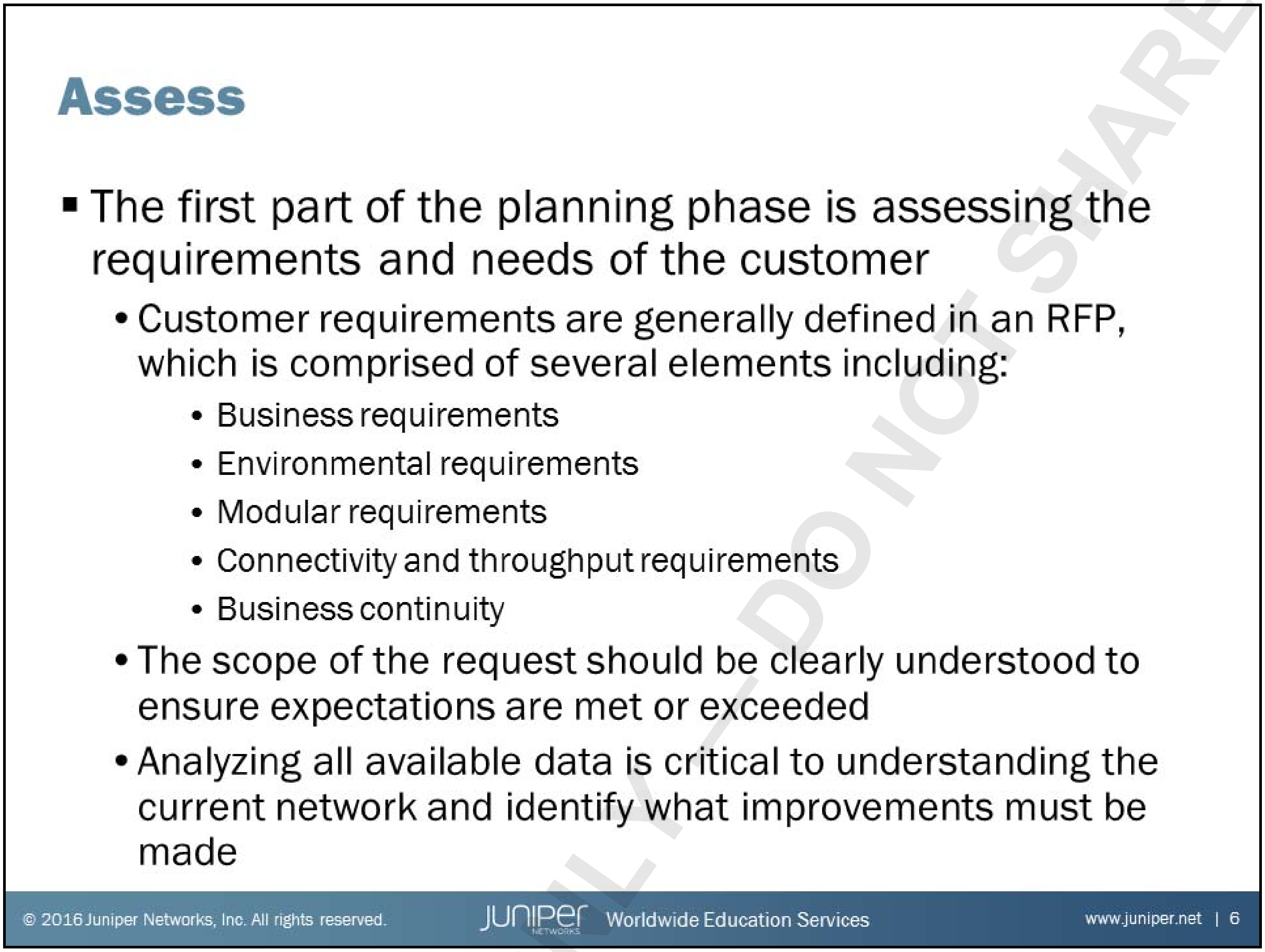
   b. Determine the scope of the design project. For example, the customer might need a design as small as a network segment, or as large as an entire enterprise network. Is the network a simple upgrade from an existing environment, or will you be creating an entirely new network?

   c. Perform a data analysis to determine the condition of the current network and what improvements need to be made based on customer requirements and scope of the design. How does the business drive the data used on the network? How many users access the network internally and externally?

2. *Design* high-level architectural plans, as well as low-level detailed plans of network devices, configurations, and interconnections. Create a project plan. Evaluate and detail responsibilities, timelines, and dependencies.

   a. The high-level design typically is the logical topology that identifies protocols used, network addressing, security, and naming conventions. The design might also include WAN and service provider access.

   b. The low-level design is the physical design and consists of physical devices that will be used in the design, cabling and wiring considerations. Service provider access should be determined by this point.



Assess

- The first part of the planning phase is assessing the requirements and needs of the customer
  - Customer requirements are generally defined in an RFP, which is comprised of several elements including:
    - Business requirements
    - Environmental requirements
    - Modular requirements
    - Connectivity and throughput requirements
    - Business continuity
  - The scope of the request should be clearly understood to ensure expectations are met or exceeded
  - Analyzing all available data is critical to understanding the current network and identify what improvements must be made

Juniper Networks Worldwide Education Services www.juniper.net | 6

Assess

Before any network can be designed, you must first assess the customer's environment. When assessing a customer's environment, you must determine what the customer requirements and scope are for the design project, and you must analyze the data provided to you to build a scalable network that will last well into the future.

The Request for Proposal (RFP) is the process the customer typically used to solicit potential vendors for network design proposals. No two RFPs are identical, as each customer has unique requirements. The RFP can be very short and concise, or it can contain several pages filled with concise requests. As the architect, your job is to ensure you have thoroughly read and understood the RFP. Although they can differ in appearance and requirements, every RFP normally includes the following:

- A list of design requirements including business goals, environmental requirements, connectivity and throughput needs, scope, and information on the existing network environment.
- The types of solutions that the design must include, such as wireless, high availability (HA), security, and so forth. The wording for these requirements can be generalized or very specific.
- Warranty requirements for the products you offer as well as any legal terms attached to the solution.

As you assess the company infrastructure, you must understand the entire scope of the design project. A network design might seem small initially, with only a few hundred users, but that network has the potential to grow significantly over time. Your design must be able to accommodate any foreseeable change in capacity. Designing with modularity in mind will help you accommodate any network the customer has asked you to design.

Continued on the next page.

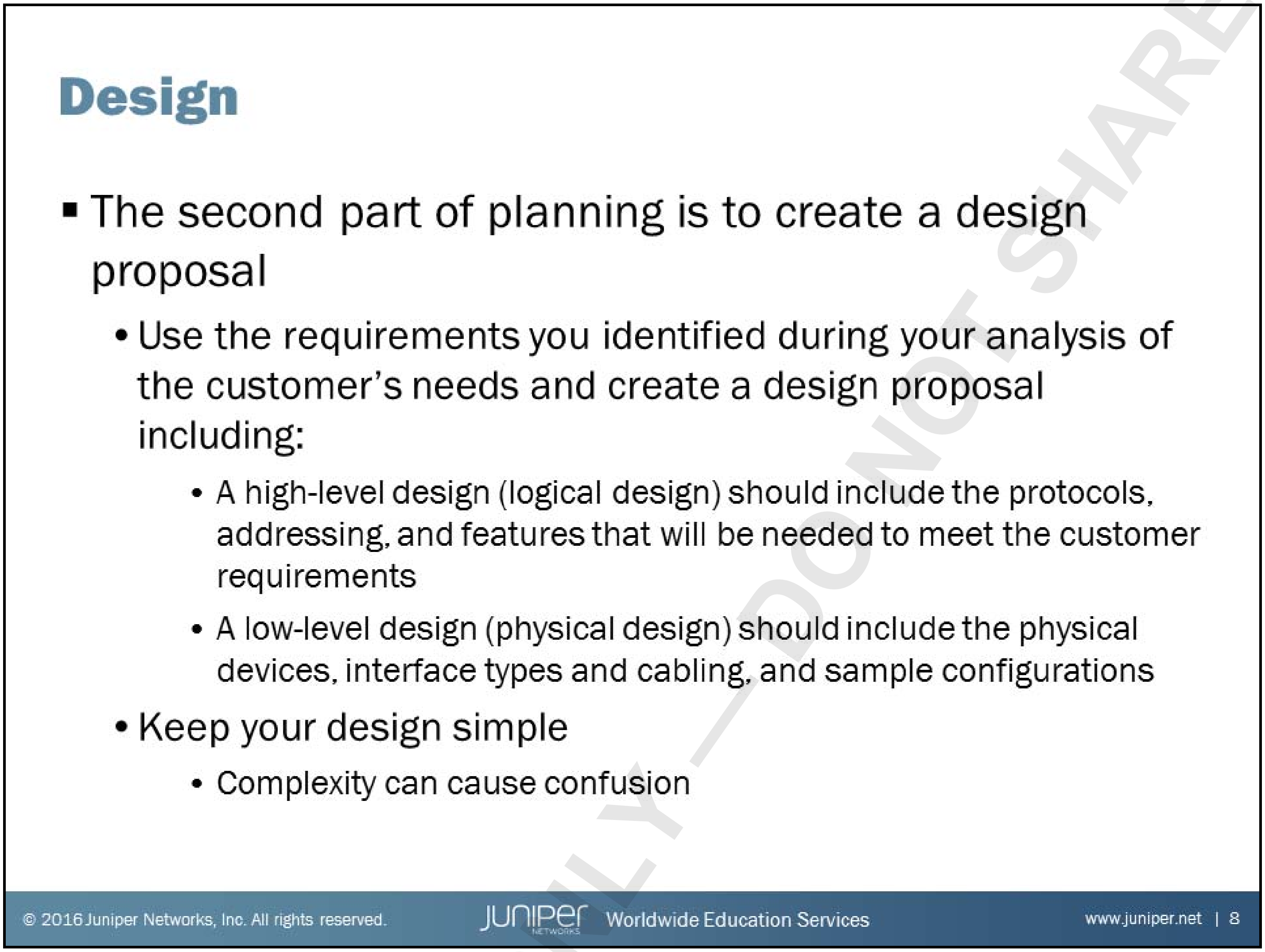
Assess (contd.)

There likely will be a lot of data to examine and decipher, so the question must be asked, *Do I understand what is needed?* If the answer is *no*, then you will need to refer back to the customer and ask more questions and collect more data.

This cycle continues until you feel like you do understand what the customer is asking for. You should then determine, *Do I know how I can help?* If the answer is *yes*, assemble your team of subject matter experts and begin planning your proposal. If you don't know how you can help, you will need to consult with the subject matter experts that will help you answer questions and help you with a design. In every case where you do not feel comfortable with a certain concept, you must not only consult with individuals who are experts, but truly become the expert yourself. There are not many issues that only come up once. Become an expert and you will put yourself ahead of the crowd on future projects.

In many cases, after you validate the data and consult with your team, new questions might arise where you must refer back to the customer. At this point, you should ask yourself, *Do I need more information?* If the answer is *yes*, then you will again find yourself cycling through customer data to fill in the blanks for your RFP response.

Once you have all the information you need and are your proposal is ready, you can respond to the RFP.



Design

- The second part of planning is to create a design proposal
  - Use the requirements you identified during your analysis of the customer's needs and create a design proposal including:
    - A high-level design (logical design) should include the protocols, addressing, and features that will be needed to meet the customer requirements
    - A low-level design (physical design) should include the physical devices, interface types and cabling, and sample configurations
  - Keep your design simple
    - Complexity can cause confusion

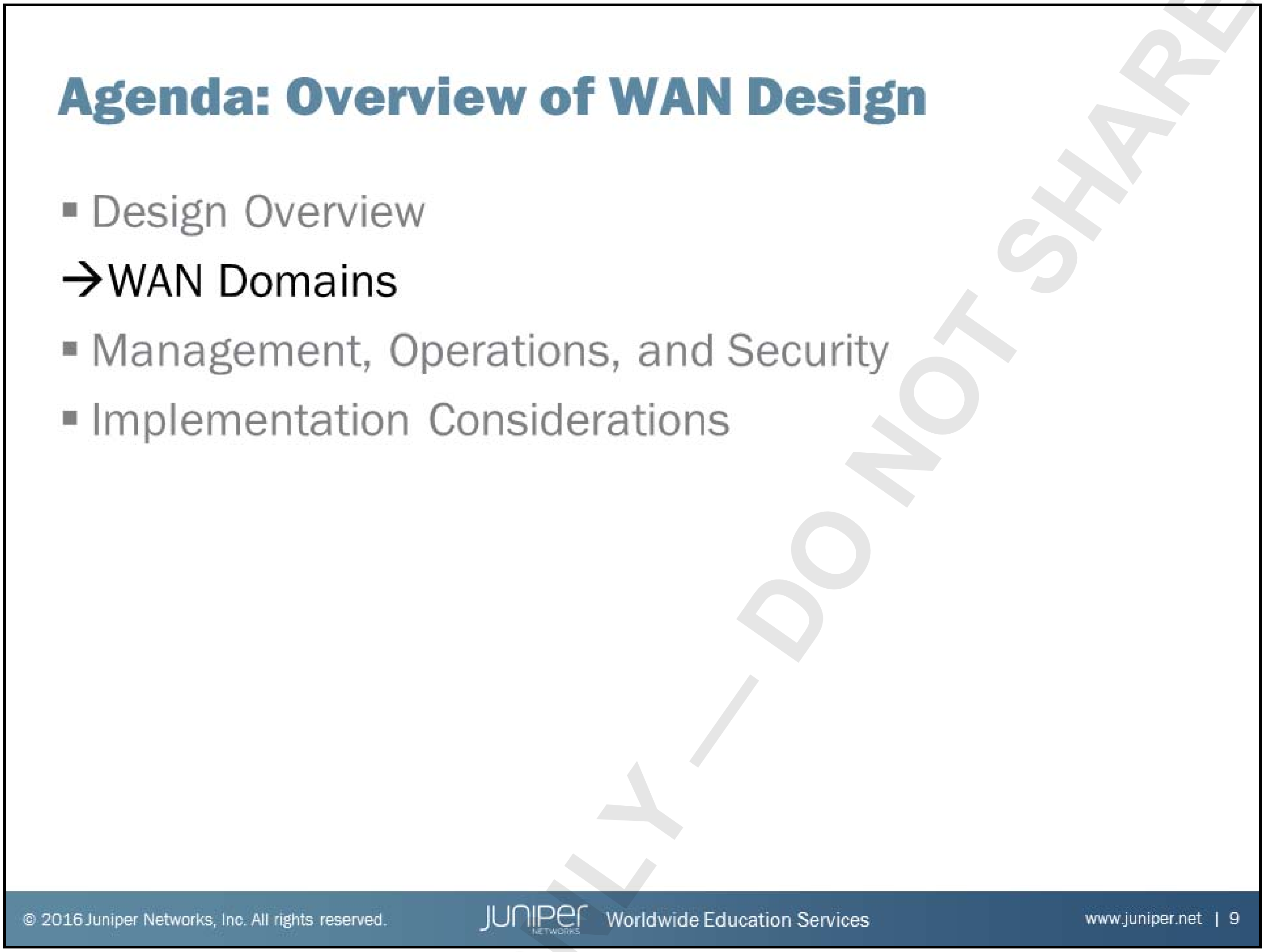
Juniper Networks Worldwide Education Services www.juniper.net | 8

Design

The next step in the plan phase is creating your design proposal. As you create your design proposal, you will need to keep the tips listed on the slide in mind. You've spent a lot of time collecting data from the customer and organizing it into a proposal you believe in. Now is the time to pay attention to the details so ensure your customer actually pays attention to your proposal.

A design is typically broken down into two levels. The high-level design generally contains the logical design components including routing protocols, addressing, device naming convention, CoS requirements, and management network needs. The low-level design contains a lot more detailed information including device choices, interface types, cabling, and configurations.

Keep the customer in mind at all times as you create your design. Some designs can be so complicated that they will not only be difficult to implement, they will be difficult for the customer to comprehend. Document your design well and keep it simple for the customer to understand. You might believe that your design is the customer's best option, but if the document is difficult to read, the customer might perceive that the design is too complicated. Some customers will equate a complicated design as being too time consuming and too expensive to implement.



Agenda: Overview of WAN Design

- Design Overview

→WAN Domains

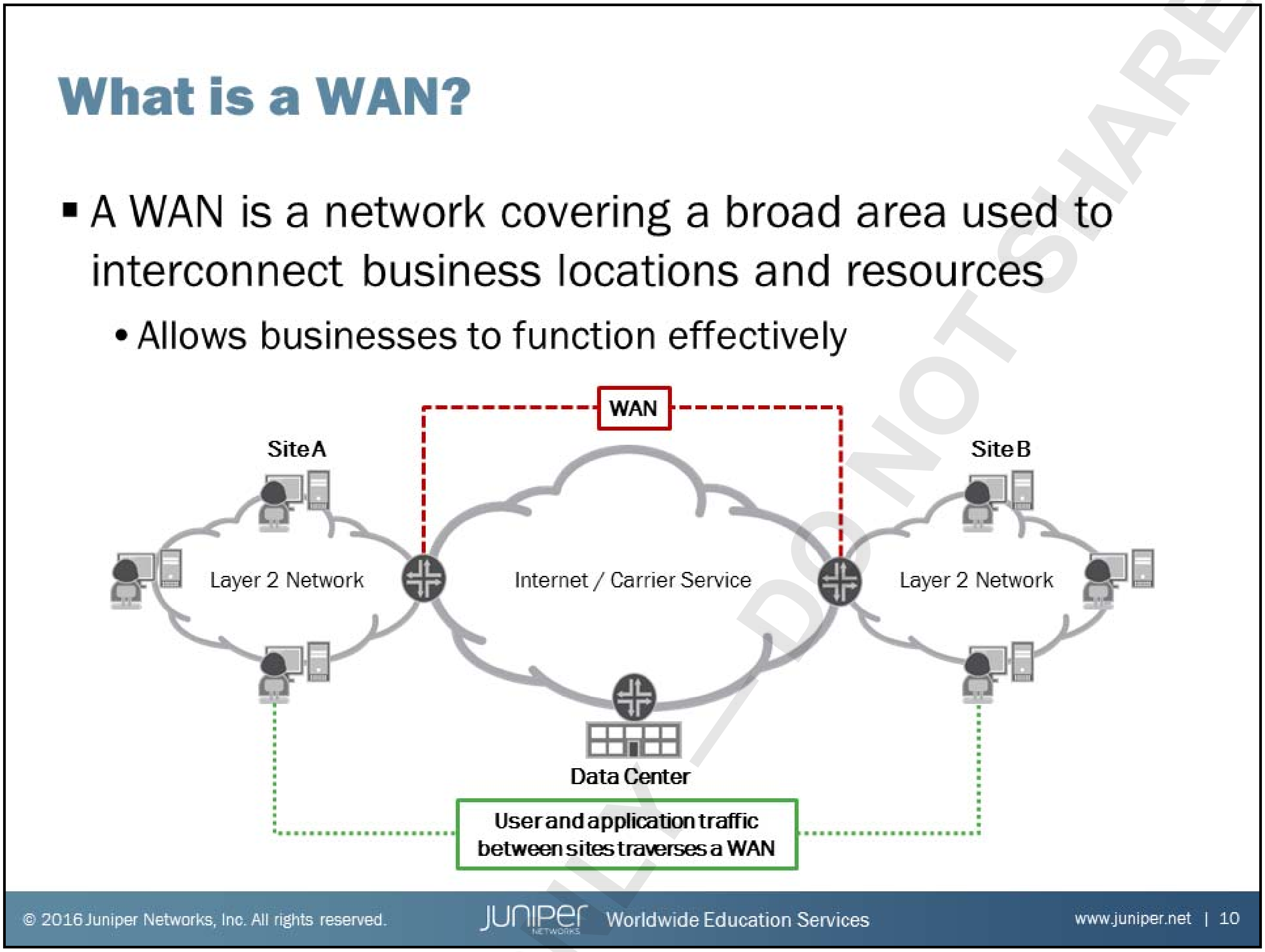
- Management, Operations, and Security
- Implementation Considerations

Juniper NETWORKS Worldwide Education Services

www.juniper.net | 9

WAN Domains

The slide highlights the topic we discuss next.



What is a WAN?

- A WAN is a network covering a broad area used to interconnect business locations and resources
  - Allows businesses to function effectively

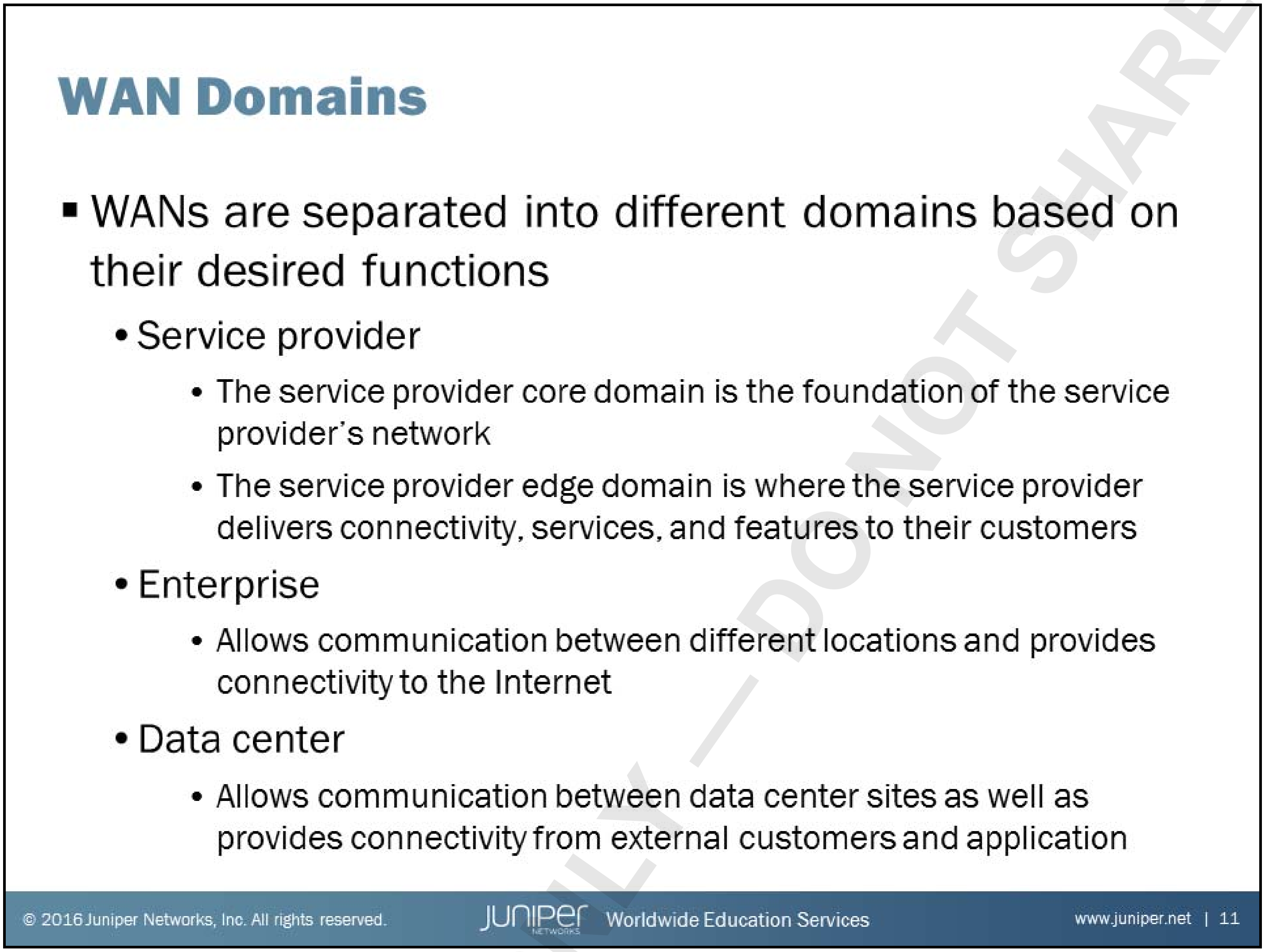
WAN
Site A
Site B
Layer 2 Network
Internet / Carrier Service
Layer 2 Network
Data Center
User and application traffic between sites traverses a WAN

Worldwide Education Services

www.juniper.net | 10

What Is a WAN?

A WAN is a network covering a broad and geographically disperse area that is used to interconnect business locations and resources. The WAN is considered a key component of today's business and allows companies to operate and function effectively. It is through the WAN that key application traffic, critical for business operations, can pass between remote locations. We describe key components of the WAN along with design considerations and options throughout this chapter.



WAN Domains

- WANs are separated into different domains based on their desired functions
  - Service provider
    - The service provider core domain is the foundation of the service provider's network
    - The service provider edge domain is where the service provider delivers connectivity, services, and features to their customers
  - Enterprise
    - Allows communication between different locations and provides connectivity to the Internet
  - Data center
    - Allows communication between data center sites as well as provides connectivity from external customers and application

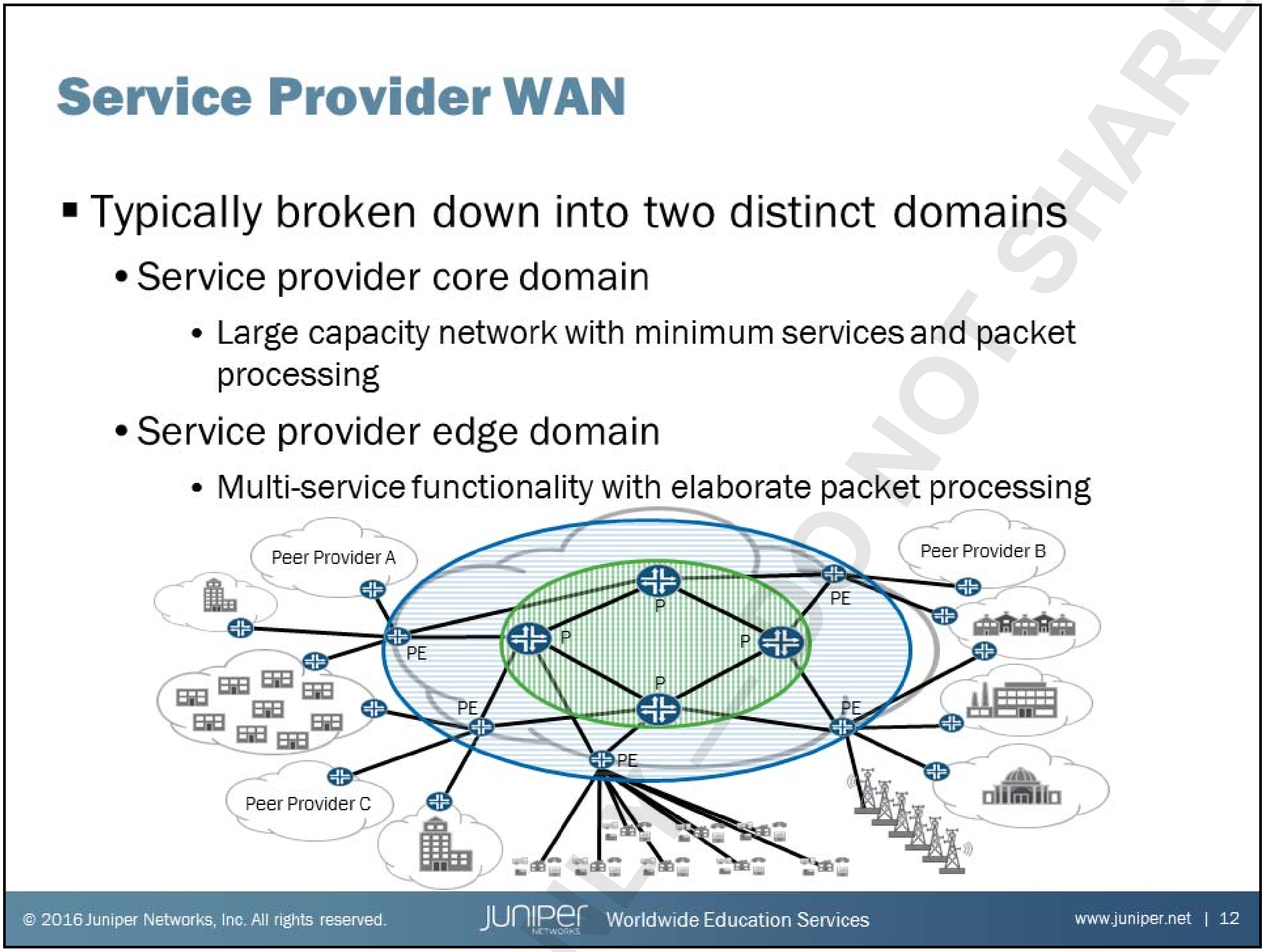
Worldwide Education Services

www.juniper.net | 11

WAN Domains

When designing WANs it is important to understand the different types of WAN domains. Each WAN domain has different requirements and network functions.

- Service providers deal exclusively in WANs. Their entire network is in essence a WAN used by enterprises and data centers to connect remote sites and locations. Service provider WANs can be categorized into two domains because of the unique requirements and design needs.
  - The service provider core network is considered the foundation or backbone of the network. The core serves as an aggregation point for all traffic entering edge devices including traffic transiting the service providers network from other service providers. It is often viewed as a cost center because the core is carrying constantly increasing amounts of traffic that service providers do not receive revenue from.
  - The service provider edge is generally where service providers generate revenue. The edge is where connectivity, services and features are provided to customers. The edge is also where service providers connect to other upstream providers for extended reachability.
- Enterprise's use WANs to interconnect remote site to enable application and internal traffic communication. An enterprise WAN can also refer to their Internet connect to allow internal users to reach and use external resources.
- Data center WANs are used to connect remote data center sites together and allow communication between internal resources. The WAN connection is also used to allow external users to access resources inside the data center.



Service Provider WAN

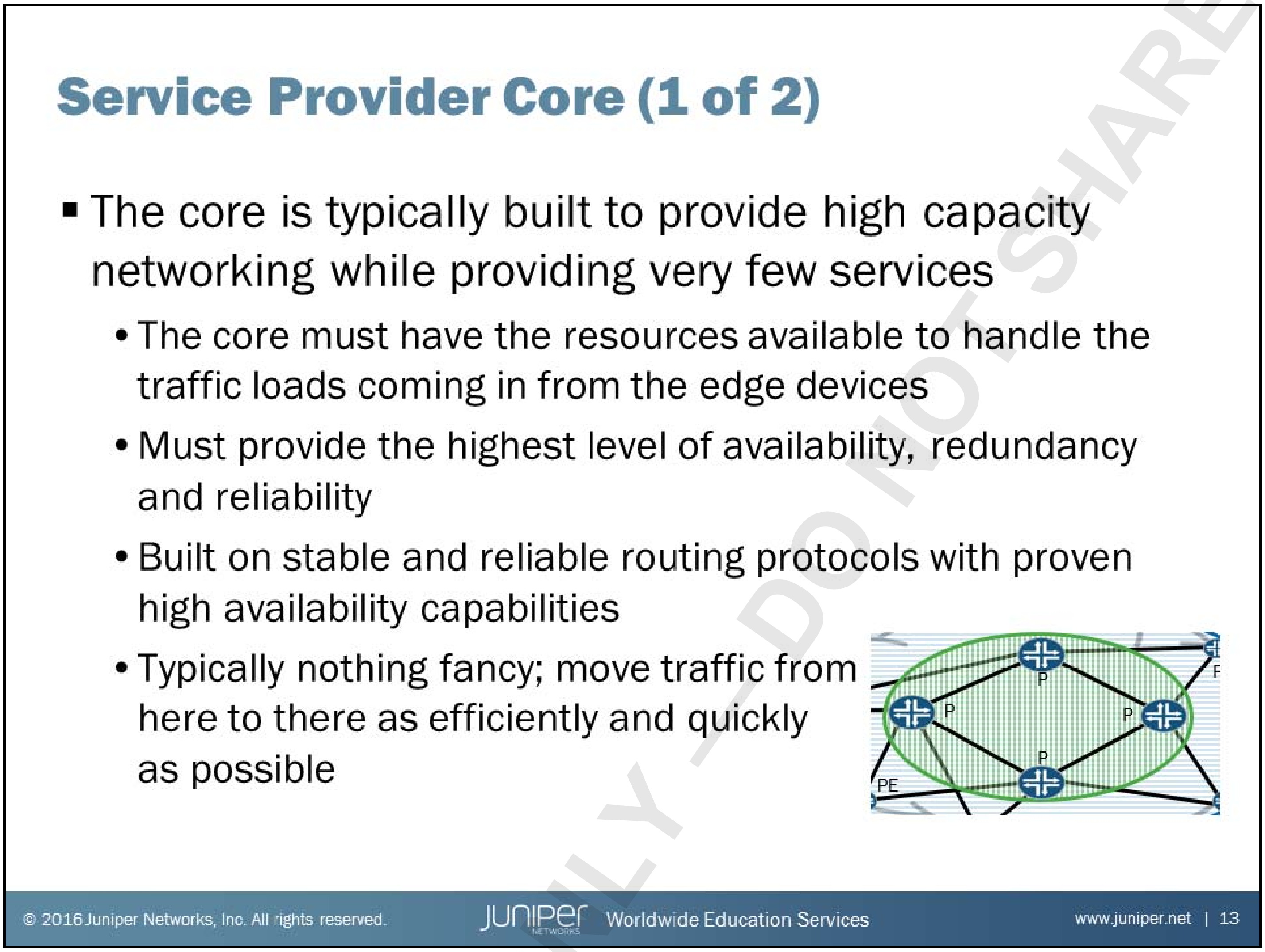
- Typically broken down into two distinct domains
  - Service provider core domain
    - Large capacity network with minimum services and packet processing
  - Service provider edge domain
    - Multi-service functionality with elaborate packet processing

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 12

Service Provider WAN

As mentioned previously service providers typically design their network in two unique domains. The service provider core is built with speed and reliability in mind. The core has to be able to efficiently handle the large amount of traffic coming in from all direction. Since the core is the backbone for a service providers network it must be very resilient and stable.

The service provider edge must provide many different features and service to customers. The edge is where new features and services are implemented. While resiliency and stability are still desired at the edge, they are not as important as they are in the core. The edge of the network is where the majority of packet processing happens including rate limiting, stateless firewalls, encryption, packet inspection, and anything else that requires extensive device processing.



Service Provider Core (1 of 2)

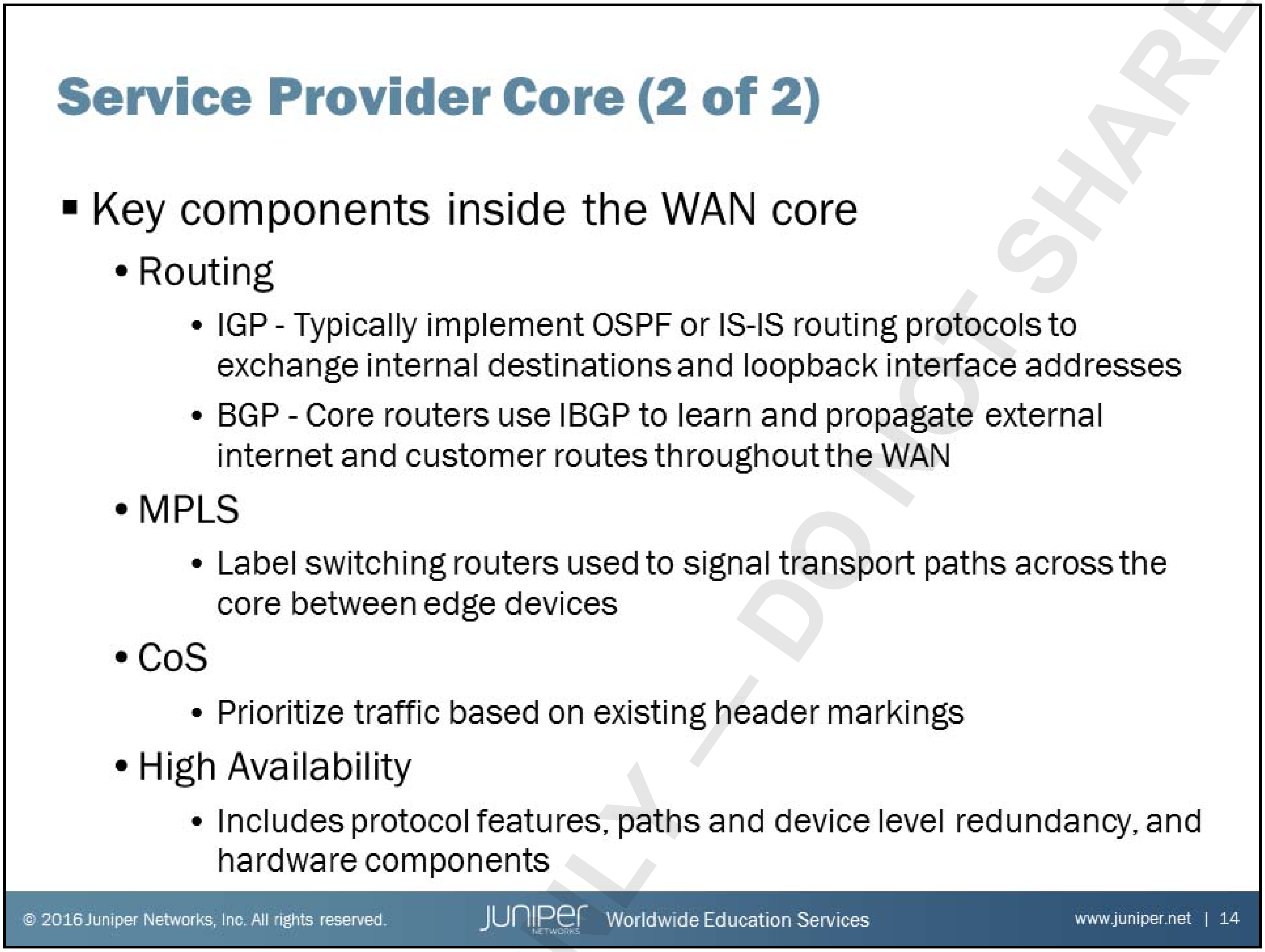
- The core is typically built to provide high capacity networking while providing very few services
  - The core must have the resources available to handle the traffic loads coming in from the edge devices
  - Must provide the highest level of availability, redundancy and reliability
  - Built on stable and reliable routing protocols with proven high availability capabilities
  - Typically nothing fancy; move traffic from here to there as efficiently and quickly as possible

Service Provider Core: Part 1

Responsibilities for service provider core routers are different than those for edge devices including the quantity and type of interfaces used. As we will discuss, edge routers are required to provide connectivity for potentially hundreds of clients while core routers typically connect all the edge devices together.

Because the core must aggregate all the traffic coming in from the edge devices it must be robust and be able to handle the load efficiently and effectively while remaining cost effective since revenue is generally not realized in this part of the network. Service providers expect that the core is a stable, highly available network. Another expectation of a core network is that it is resilient, which implies fast convergence after some unforeseen outage, or in general, a non-blocking design for traffic flowing in all directions, referred to north, south, east, and west traffic.

Because the core must be reliable, service providers don’t want to implement protocols and features that have not been proven to be stable. The end goal of the core network is to take as much traffic as possible and move it from point A to point B as fast and cost effective as possible.



Service Provider Core (2 of 2)

- Key components inside the WAN core
  - Routing
    - IGP - Typically implement OSPF or IS-IS routing protocols to exchange internal destinations and loopback interface addresses
    - BGP - Core routers use IBGP to learn and propagate external internet and customer routes throughout the WAN
  - MPLS
    - Label switching routers used to signal transport paths across the core between edge devices
  - CoS
    - Prioritize traffic based on existing header markings
  - High Availability
    - Includes protocol features, paths and device level redundancy, and hardware components

Worldwide Education Services

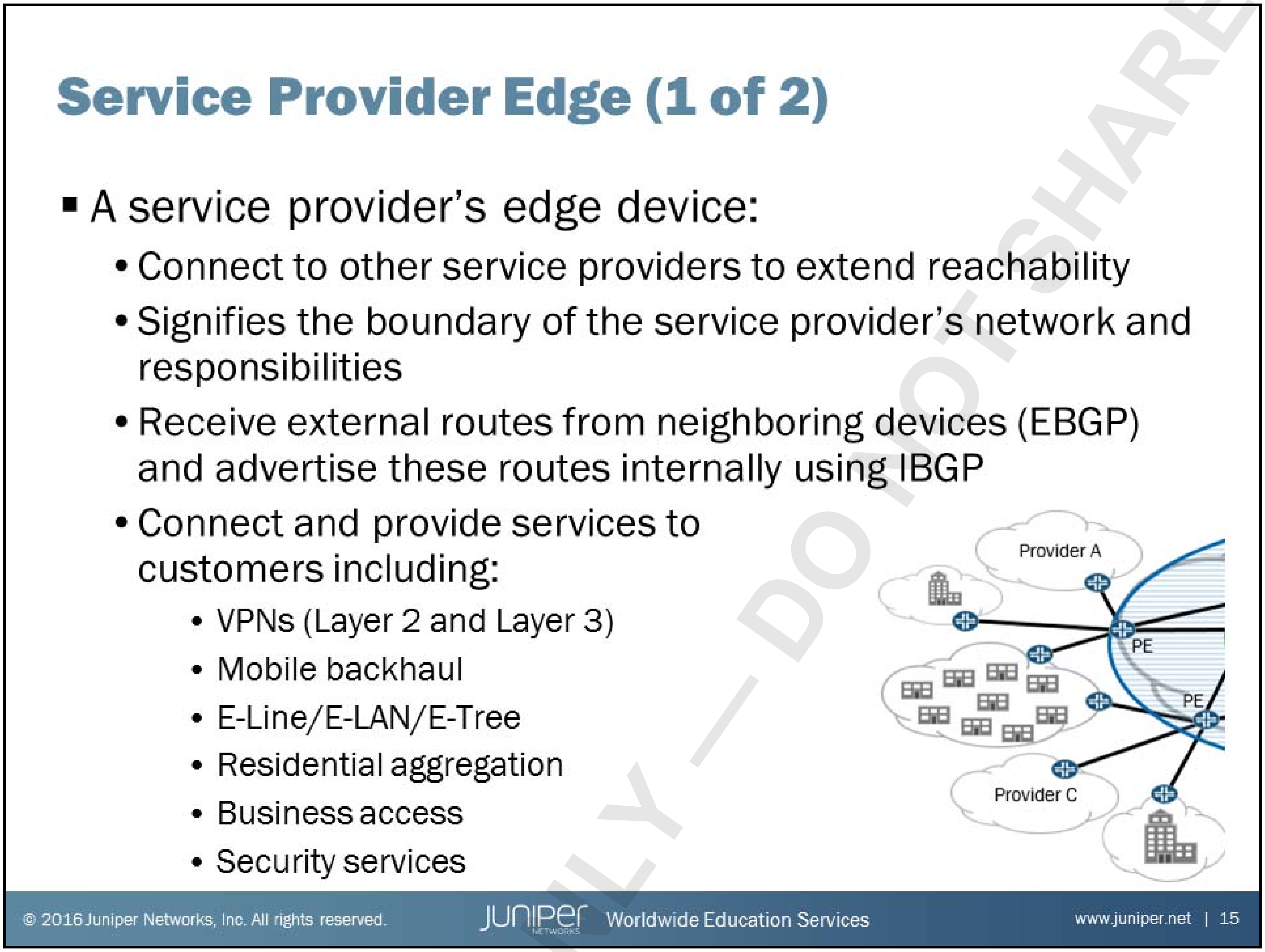
www.juniper.net | 14

Service Provider Core: Part 2

Core routers typically run OSPFv2, OSPFv3, or IS-IS to exchange internal destinations and share traffic engineering information, which will be used by MPLS to signal optimized label-switched paths (LSPs). In addition to an interior gateway protocol (IGP), core routers will be using IBGP to exchange external routes learned from neighboring devices outside of their administration.

Class of service is a key consideration within the core service provider network. Packet classification in the core is typically based on the header markings that are set on the edge devices which takes significantly less processing. If you implement the same settings on all devices and interfaces in the core you can ensure consistent handling of prioritized packets.

Because service providers expect the core to always be available, a good core design incorporates as many high availability features as possible without impacting the speed and cost effectiveness of the network.



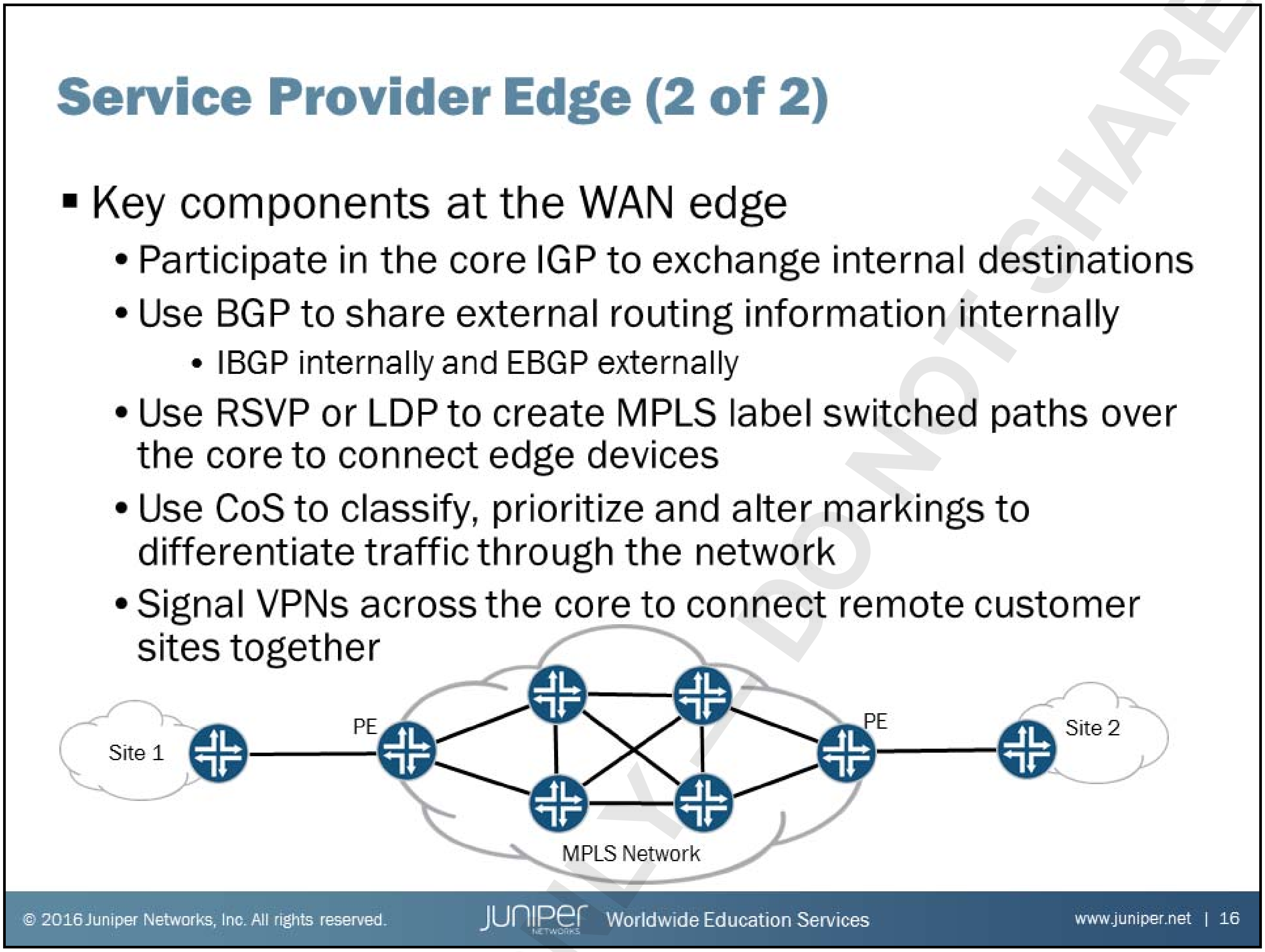
Service Provider Edge (1 of 2)

- A service provider's edge device:
  - Connect to other service providers to extend reachability
  - Signifies the boundary of the service provider's network and responsibilities
  - Receive external routes from neighboring devices (EBGP) and advertise these routes internally using IBGP
  - Connect and provide services to customers including:
    - VPNs (Layer 2 and Layer 3)
    - Mobile backhaul
    - E-Line/E-LAN/E-Tree
    - Residential aggregation
    - Business access
    - Security services

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Worldwide Education Services www.juniper.net | 15

Service Provider Edge: Part 1

The service provider edge is where service providers connect to their customers and other peer service providers. These devices signify the edge of the service providers network and typically provide custom services to their end customers. The edge devices generally learn routes from upstream service providers using EBGP and then propagated these external routes to internal neighbors using IBGP. As a service provider, there are a few common services that are offered to their customers including Layer 2 VPNs, Layer 3 VPNs, mobile backhaul, business access, and security services. The more services a provider can offer at the edge, the more opportunities they have to generate income for the company.



Service Provider Edge (2 of 2)

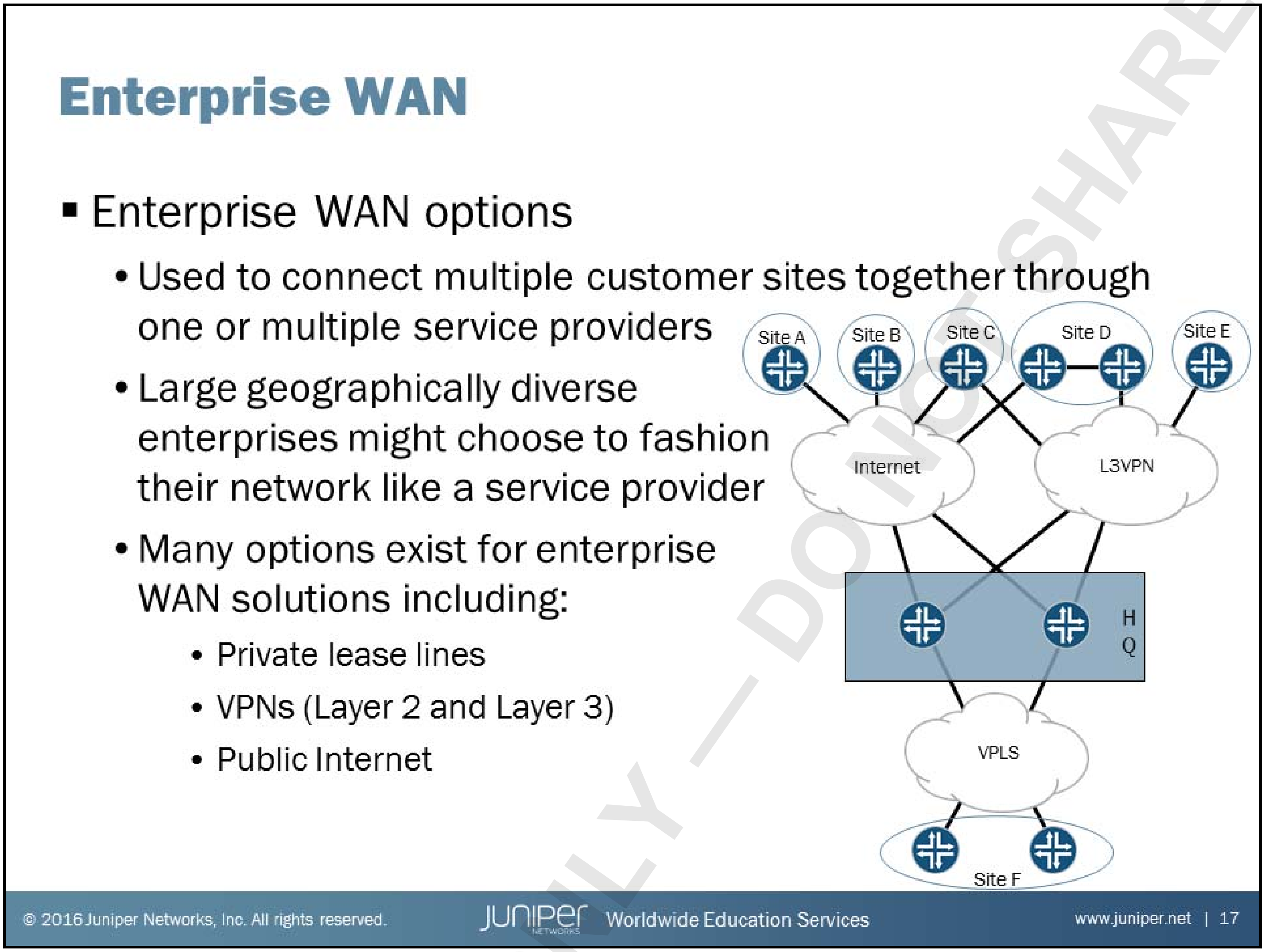
- Key components at the WAN edge
  - Participate in the core IGP to exchange internal destinations
  - Use BGP to share external routing information internally
    - IBGP internally and EBGP externally
  - Use RSVP or LDP to create MPLS label switched paths over the core to connect edge devices
  - Use CoS to classify, prioritize and alter markings to differentiate traffic through the network
  - Signal VPNs across the core to connect remote customer sites together

Worldwide Education Services

www.juniper.net | 16

Service Provider Edge: Part 2

Edge devices can participate in the core IGP. This ensures that traffic engineering information is shared out to the edge of the network and can be effectively used to optimize the MPLS LSPs through the core to other edge devices. As mentioned earlier, service providers use EBGP to learn external routes from upstream neighbors as well as customers. They will then propagate this information internally using IBGP to ensure all internal devices can reach these destinations. Another key aspect of a service providers network is using VPNs to connect customer locations. There are many different types of VPNs and since the core is built using MPLS, they are all typically offered to customers as a solution.



Enterprise WAN

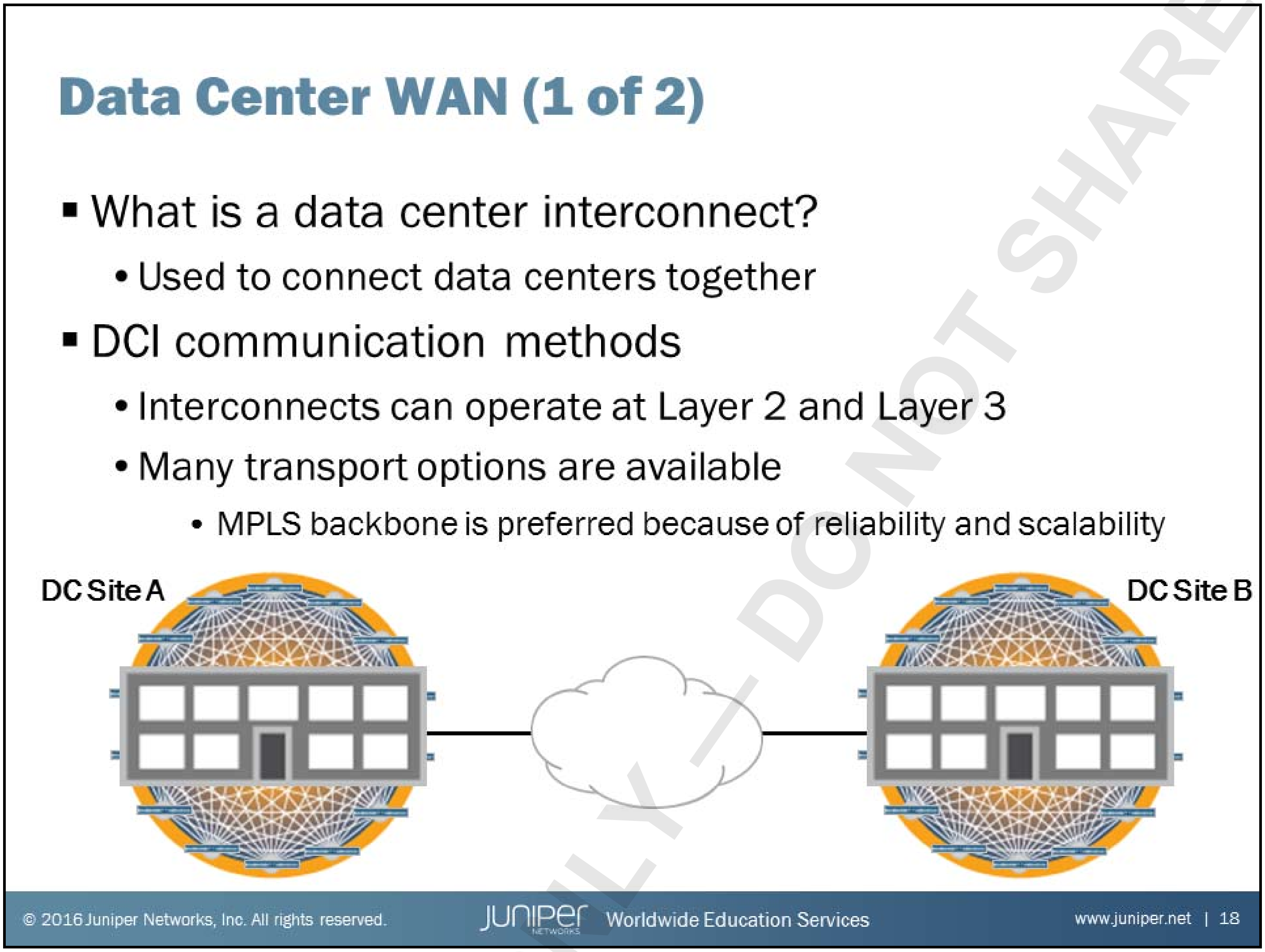
- Enterprise WAN options
  - Used to connect multiple customer sites together through one or multiple service providers
  - Large geographically diverse enterprises might choose to fashion their network like a service provider
  - Many options exist for enterprise WAN solutions including:
    - Private lease lines
    - VPNs (Layer 2 and Layer 3)
    - Public Internet

Site A
Site B
Site C
Site D
Site E
Internet
L3VPN
HQ
VPLS
Site F

Enterprise WAN

An Enterprise WAN can mean connecting multiple sites together for a single enterprise customer across multiple service providers or it can be as simple as connecting to the Internet using a single service provider. Some large service providers might have applications and services that require the network be built similar to a service providers. Meaning that they run their MPLS transport services over the top of whatever service they are using from their service providers. For example, a very large enterprise network might be running VPLS internally to connect different departments over top of a VPLS instance being provided by their provider used to connect their remote sites.

There are many choices when choosing a WAN transport and those choices can be mixed depending on topology or regional coverage. The example diagram above shows a single HQ site attached to the Internet, a L3VPN cloud, and a VPLS cloud. In the end it does not really matter as the HQ routers will process all the traffic and send it on its way, re-encapsulating the packet as needed. Optimally, the fewer transport providers the less complex the network would be but that may be a trade off in reliability.



Data Center WAN (1 of 2)

- What is a data center interconnect?
  - Used to connect data centers together
- DCI communication methods
  - Interconnects can operate at Layer 2 and Layer 3
  - Many transport options are available
    - MPLS backbone is preferred because of reliability and scalability

Juniper NETWORKS Worldwide Education Services

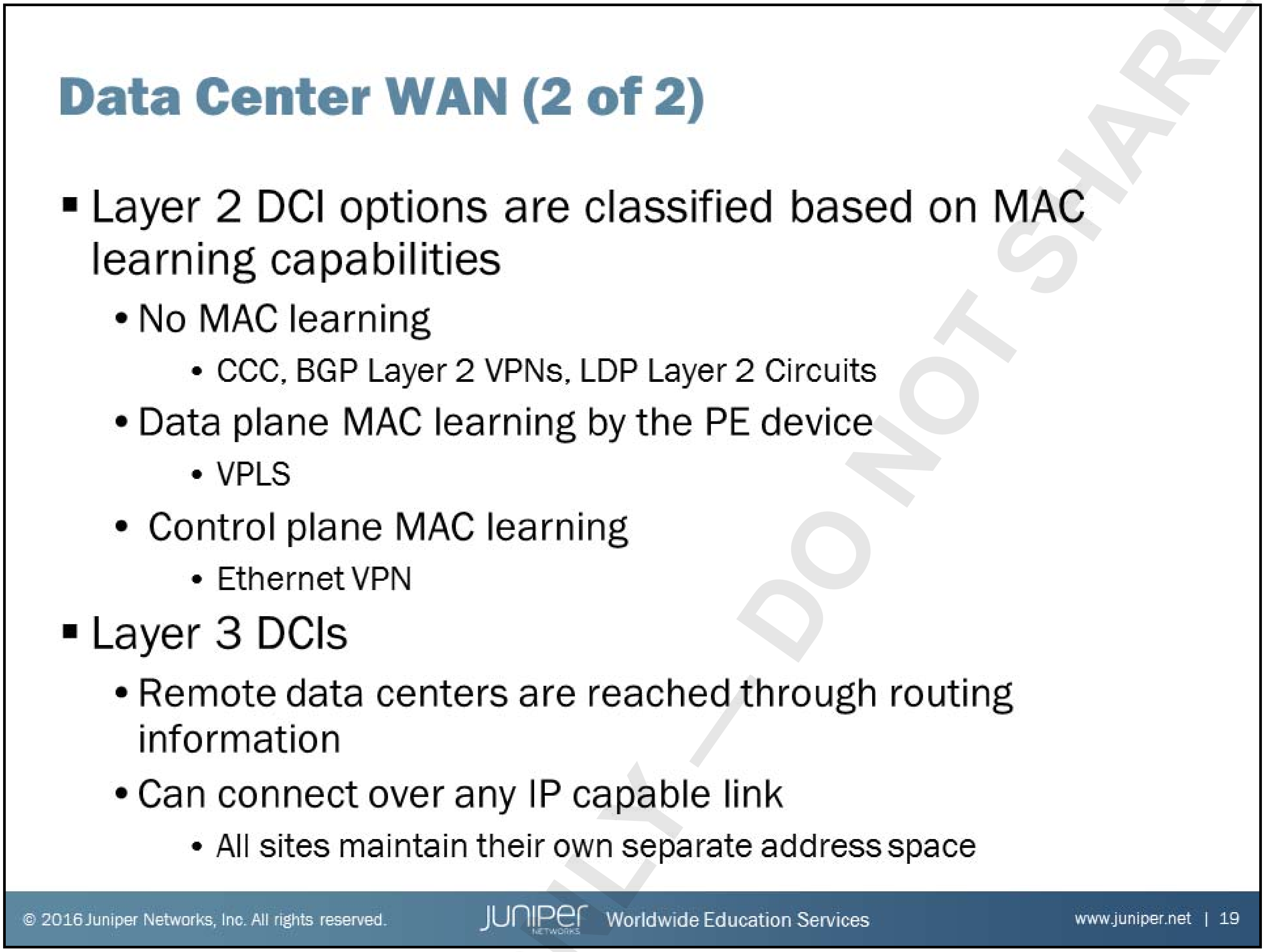
www.juniper.net | 18

What Is a Data Center Interconnect?

Data center interconnect (DCI) is basically a method to connect multiple data centers together. As the name implies, a Layer 3 DCI uses IP routing between data centers while a Layer 2 DCI extends the Layer 2 network (VLANs) from one data center to another.

DCI Communication Methods

Many of the DCI communication options rely on an MPLS network to transport frames between data centers. Although in most cases an MPLS network can be substituted with an IP network (i.e., by encapsulating MPLS in GRE), there are several advantages to using an MPLS network including availability, cost, fast failover, traffic engineering, and scalable VPN options.



Data Center WAN (2 of 2)

- Layer 2 DCI options are classified based on MAC learning capabilities
  - No MAC learning
    - CCC, BGP Layer 2 VPNs, LDP Layer 2 Circuits
  - Data plane MAC learning by the PE device
    - VPLS
  - Control plane MAC learning
    - Ethernet VPN
- Layer 3 DCIs
  - Remote data centers are reached through routing information
  - Can connect over any IP capable link
    - All sites maintain their own separate address space

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 19

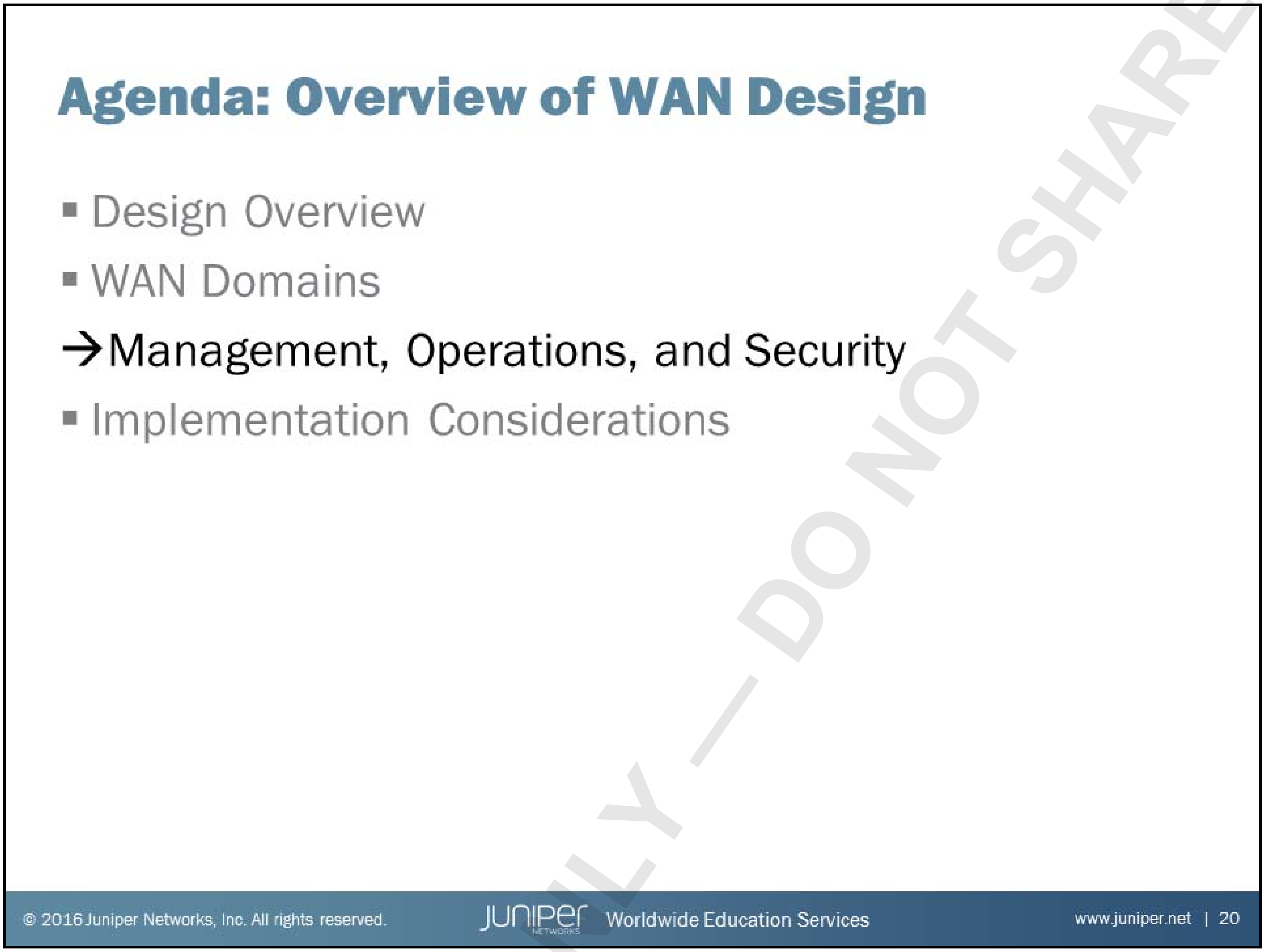
Layer 2 Options

Three classifications exist for Layer 2 DCIs:

1. No MAC learning by the Provider Edge (PE) device: This type of layer 2 DCI does not require that the PE devices learn MAC addresses.
2. Data plane MAC learning by the PE device: This type of DCI requires that the PE device learns the MAC addresses of both the local data center as well as the remote data centers.
3. Control plane MAC learning - This type of DCI requires that a local PE learn the local MAC addresses using the control plane and then distribute these learned MAC addressed to the remote PEs.

Layer 3 Options

A Layer 3 DCI uses routing to interconnect data centers. Each data center must maintain a unique IP address space. A Layer 3 DCI can be established using just about any IP capable link.



Agenda: Overview of WAN Design

- Design Overview
- WAN Domains

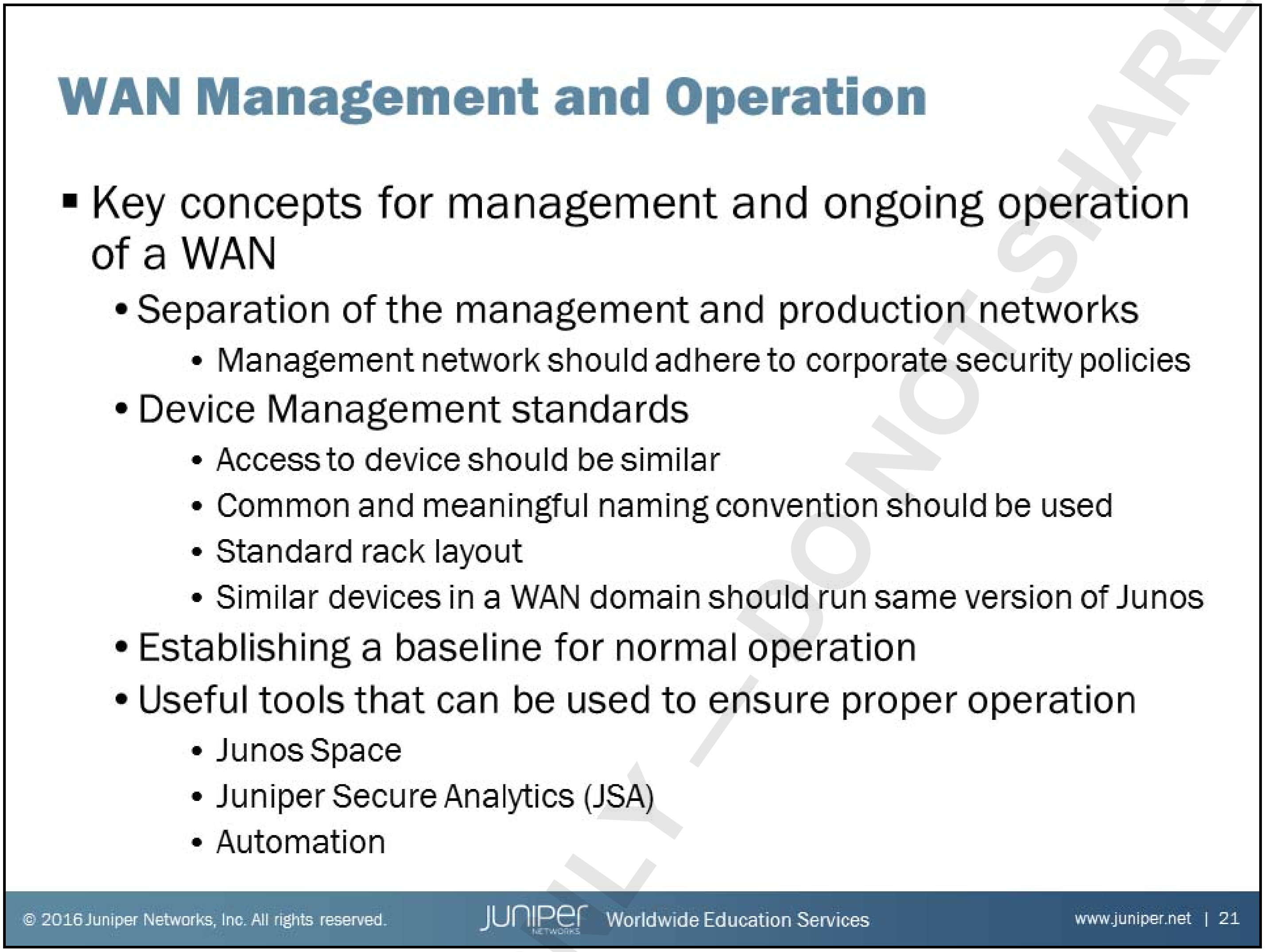
→Management, Operations, and Security

- Implementation Considerations

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 20

Management, Operations and Security

The slide highlights the topic we discuss next.



WAN Management and Operation

- Key concepts for management and ongoing operation of a WAN
  - Separation of the management and production networks
    - Management network should adhere to corporate security policies
  - Device Management standards
    - Access to device should be similar
    - Common and meaningful naming convention should be used
    - Standard rack layout
    - Similar devices in a WAN domain should run same version of Junos
  - Establishing a baseline for normal operation
  - Useful tools that can be used to ensure proper operation
    - Junos Space
    - Juniper Secure Analytics (JSA)
    - Automation

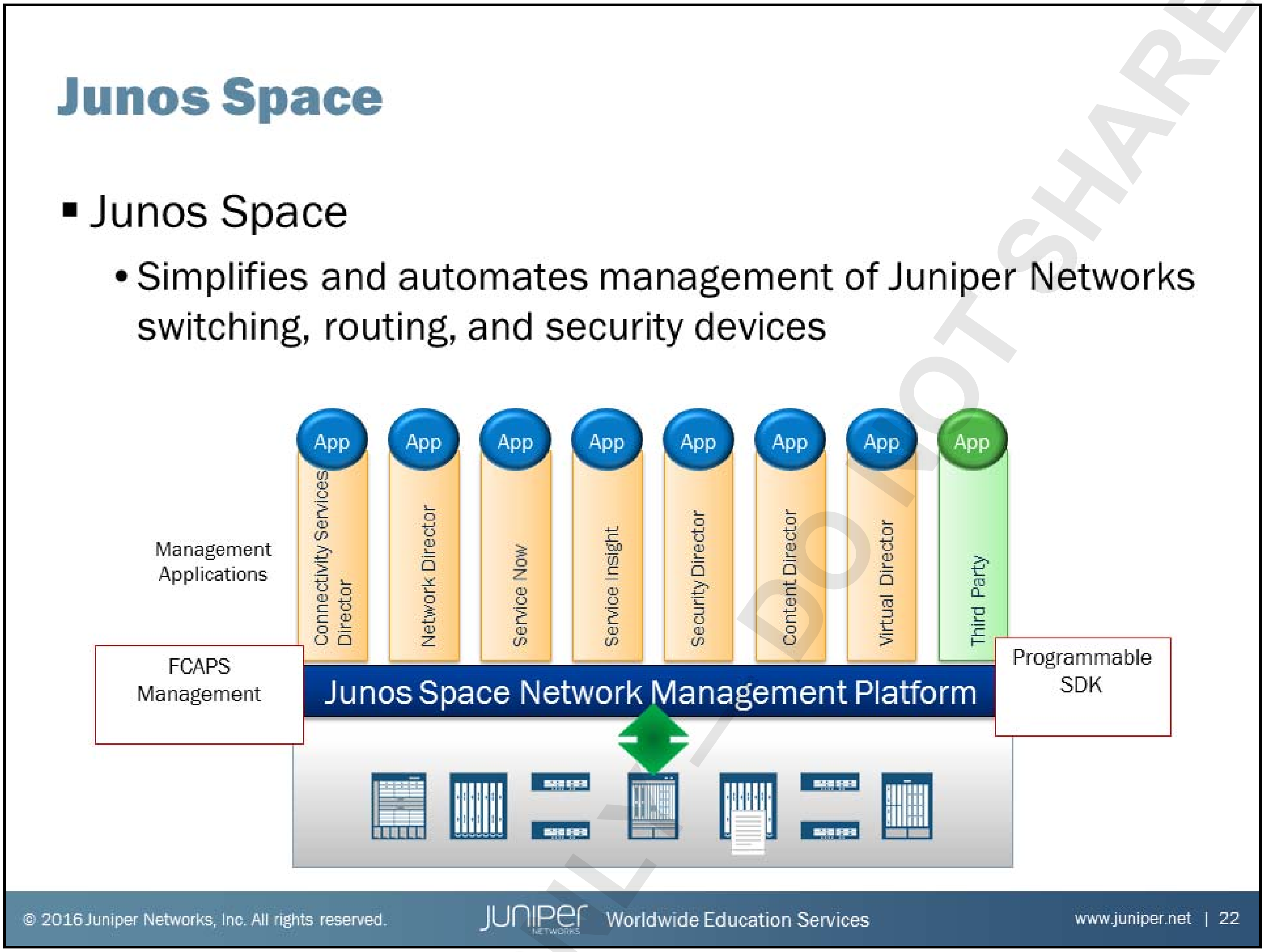
Worldwide Education Services

www.juniper.net | 21

WAN Management and Operation

An important part of designing a good WAN is considering the ongoing management and operation of its devices and health. To efficiently manage your data center, you should define and incorporate network standards including separating the management network from the production network, device access, device naming conventions, rack layouts, and Junos versions. This will make the ongoing tasks associated with ensuring the data center is healthy and operating at peak performance much easier.

Another important point to consider is, how do I know what is normal for my WAN? This is not an easy question to answer and is different for all WANs. It is important to know what normal is in order to determine if something is abnormal. Once the WAN is in production you can determine what the baseline is by using tools designed to monitor you network’s health including Junos Space, Juniper Secure Analytics and automation scripts. You might need to establish a new baseline multiple times throughout the life cycle of a WAN as new services or devices are added.



Junos Space

- Junos Space
  - Simplifies and automates management of Juniper Networks switching, routing, and security devices

Management Applications

App — Connectivity Services Director

App — Network Director

App — Service Now

App — Service Insight

App — Security Director

App — Content Director

App — Virtual Director

App — Third Party

FCAPS Management

Junos Space Network Management Platform

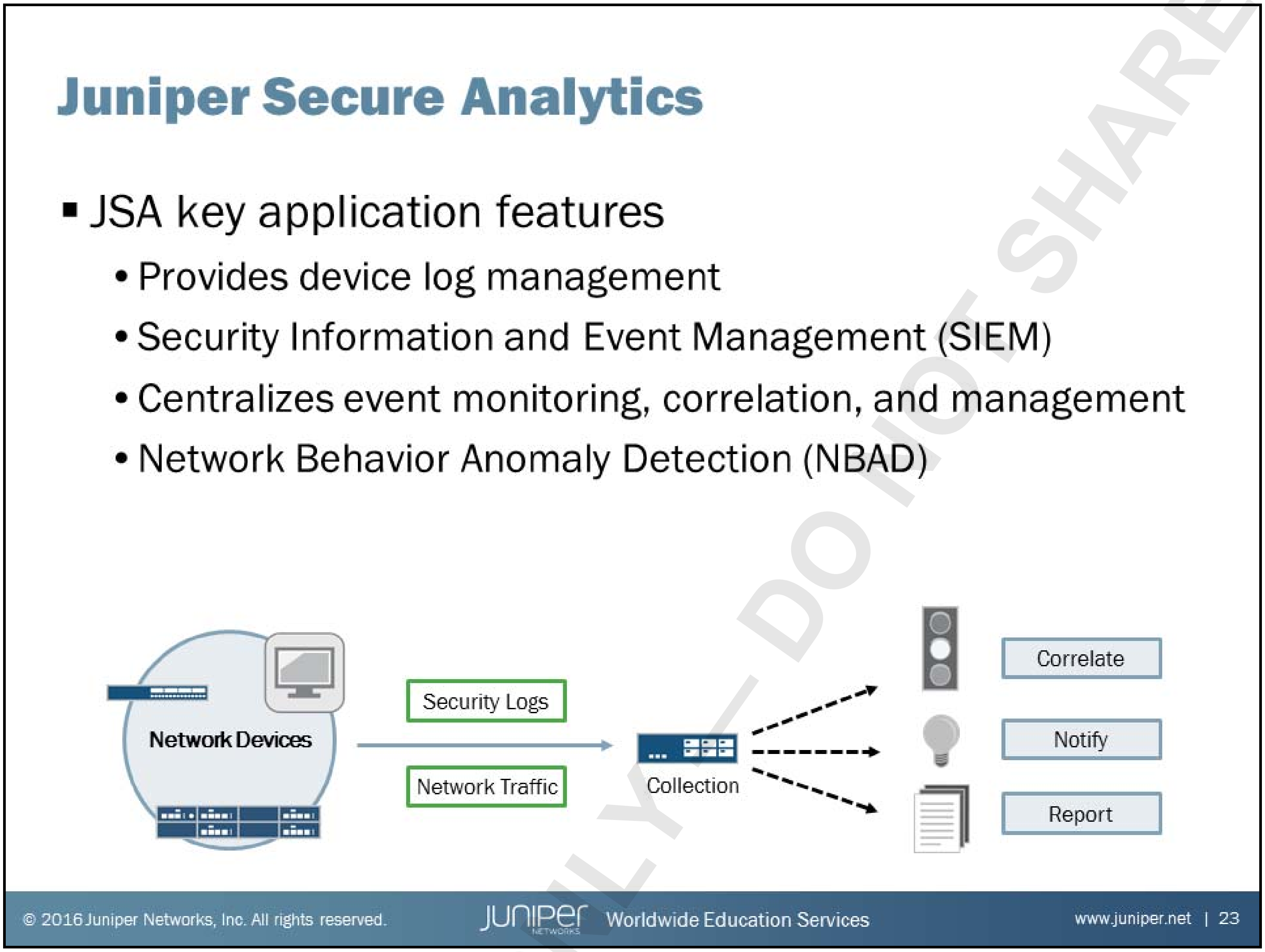
Programmable SDK

Juniper Networks Worldwide Education Services

www.juniper.net | 22

Junos Space

WAN operation can be greatly simplified by using Junos Space. Junos Space is a comprehensive network management solution that simplifies and automates management of Juniper Networks switching, routing, and security devices. With all of its components working together, Junos Space offers a unified network management and orchestration solution to help you more efficiently manage a WAN.



Juniper Secure Analytics

- JSA key application features
  - Provides device log management
  - Security Information and Event Management (SIEM)
  - Centralizes event monitoring, correlation, and management
  - Network Behavior Anomaly Detection (NBAD)

Network Devices

Security Logs

Network Traffic

Collection

Correlate

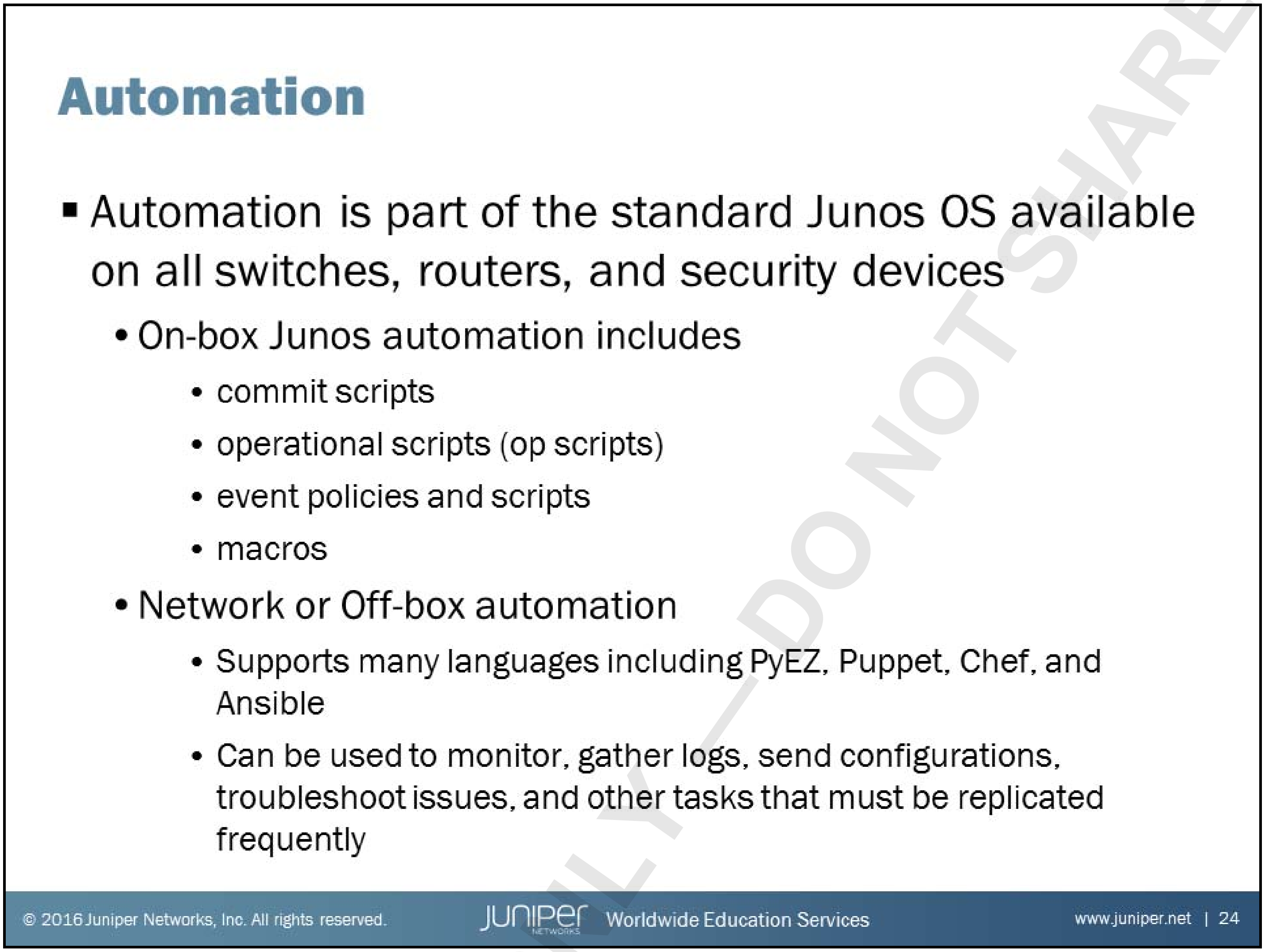
Notify

Report

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 23

Juniper Secure Analytics

Juniper Secure Analytics (JSA) is a unique log management, security event management, and network anomaly behavior detection (NBAD) solution combined into one device. JSA can be used to effectively identify and mitigate security threats in a timely and efficient manner.



Automation

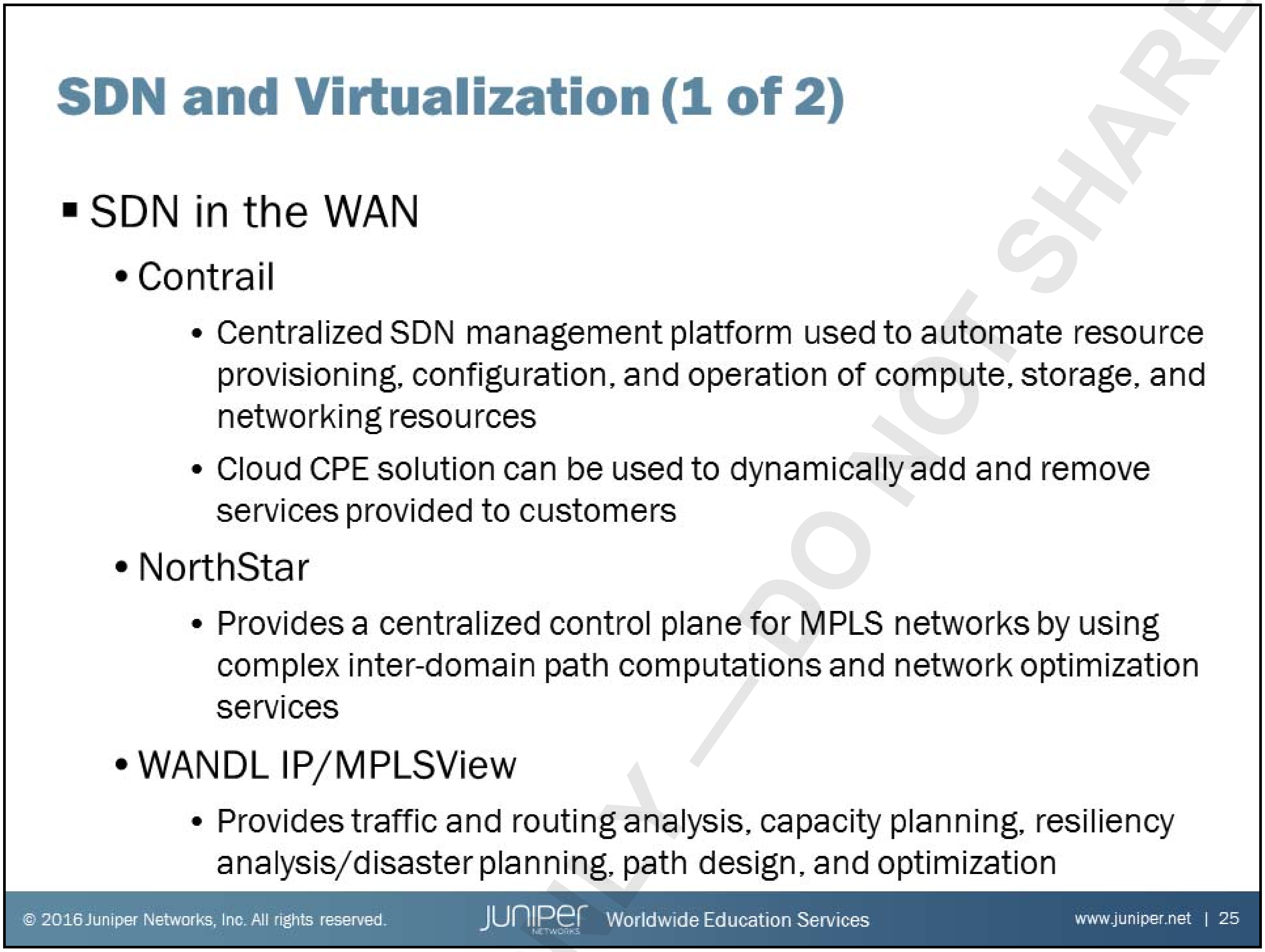
- Automation is part of the standard Junos OS available on all switches, routers, and security devices
  - On-box Junos automation includes
    - commit scripts
    - operational scripts (op scripts)
    - event policies and scripts
    - macros
  - Network or Off-box automation
    - Supports many languages including PyEZ, Puppet, Chef, and Ansible
    - Can be used to monitor, gather logs, send configurations, troubleshoot issues, and other tasks that must be replicated frequently

Juniper Networks Worldwide Education Services

www.juniper.net | 24

Automation

Junos automation is part of the standard Junos OS available on all switches, routers, and security devices running Junos OS. Junos automation can be used to automate operational and configuration tasks on a network's Junos devices. The slide highlights both on-box and off-box automation capabilities. including support for multiple scripting languages.



SDN and Virtualization (1 of 2)

- SDN in the WAN
  - Contrail
    - Centralized SDN management platform used to automate resource provisioning, configuration, and operation of compute, storage, and networking resources
    - Cloud CPE solution can be used to dynamically add and remove services provided to customers
  - NorthStar
    - Provides a centralized control plane for MPLS networks by using complex inter-domain path computations and network optimization services
  - WANDL IP/MPLSView
    - Provides traffic and routing analysis, capacity planning, resiliency analysis/disaster planning, path design, and optimization

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 25

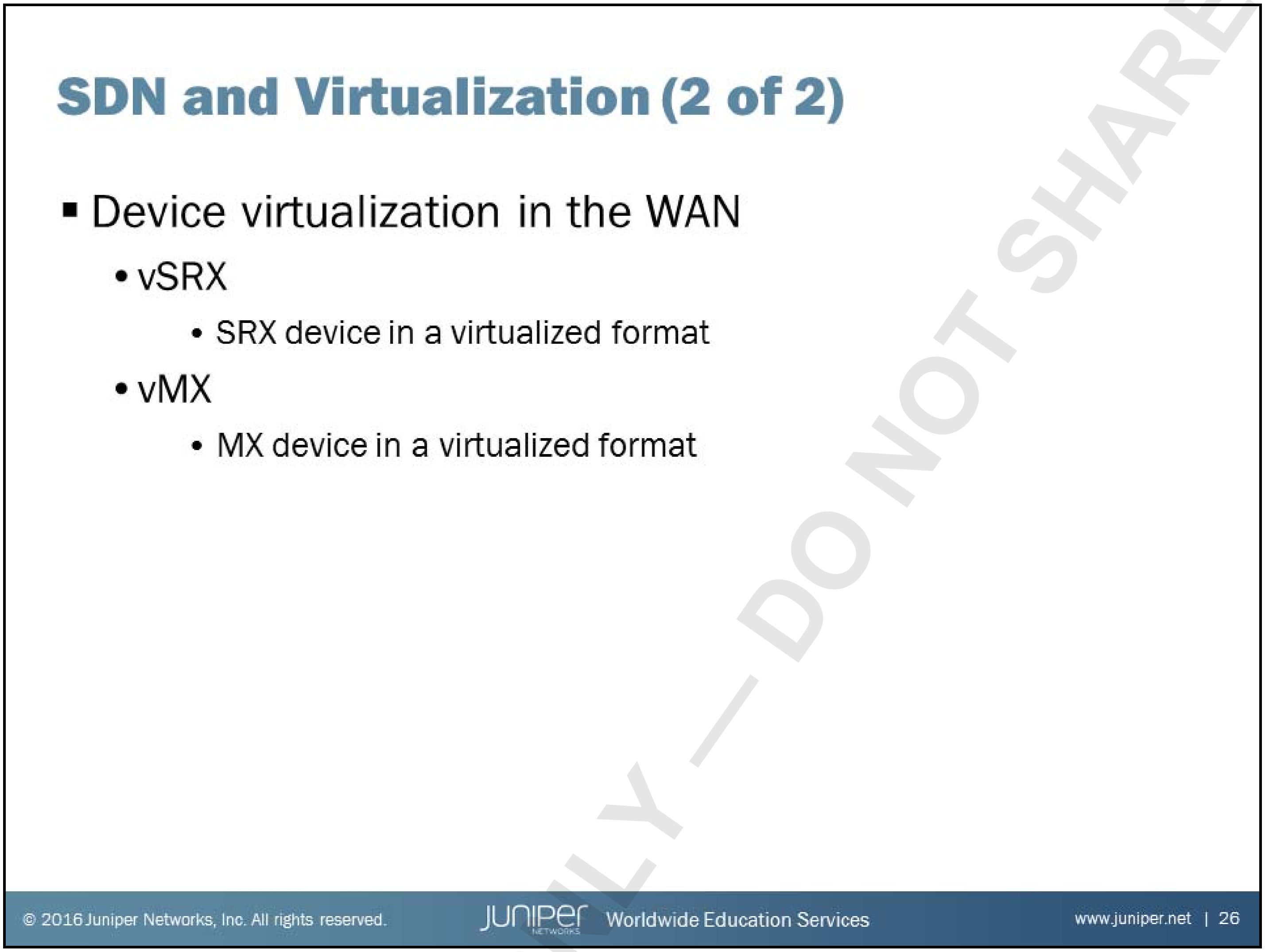
SDN and Virtualization: Part 1

Juniper has a very complete SDN strategy which is described as the 6-4-1 SDN strategy. This 6-4-1 SDN strategy consists of six general principles (Separate, Centralize, Use the cloud, Common platform, Standardize, and Apply Broadly), four steps (Centralize management, Extract services, Centralize controller and Optimize the hardware), and one licensing model.

Juniper's Contrail is a simple, open, and agile SDN solution that automates and orchestrates the creation of highly scalable virtual networks. These virtual networks let you harness the power of the cloud—for new services, increased business agility, and revenue growth.

Juniper's NorthStar Controller is another example of software defined networking. It provides a centralized control plane for MPLS domains. The NorthStar Controller is a WAN SDN controller that automates the discovery and creation of traffic-engineered label switched paths across a service provider or large enterprise network. NorthStar Controller provides granular visibility into, and control over, IP/MPLS flows. The NorthStar Controller provides complex inter-domain path computation and network optimization services allowing networks to fully utilize available bandwidth.

The WANDL IP/MPLSView solution addresses major areas of network planning including traffic and routing analysis, capacity planning, resiliency analysis/disaster planning, path design, and optimization. From a set of network configuration files and other optional data, an intelligent multivendor analyzer within IP/MPLSView constructs an accurate network topology, fully aware of multiple protocols, layers, autonomous systems (AS), areas, VPNS, etc. Alternatively, the user can manually construct any network topology using the IP/MPLSView advanced graphical interface.



SDN and Virtualization (2 of 2)

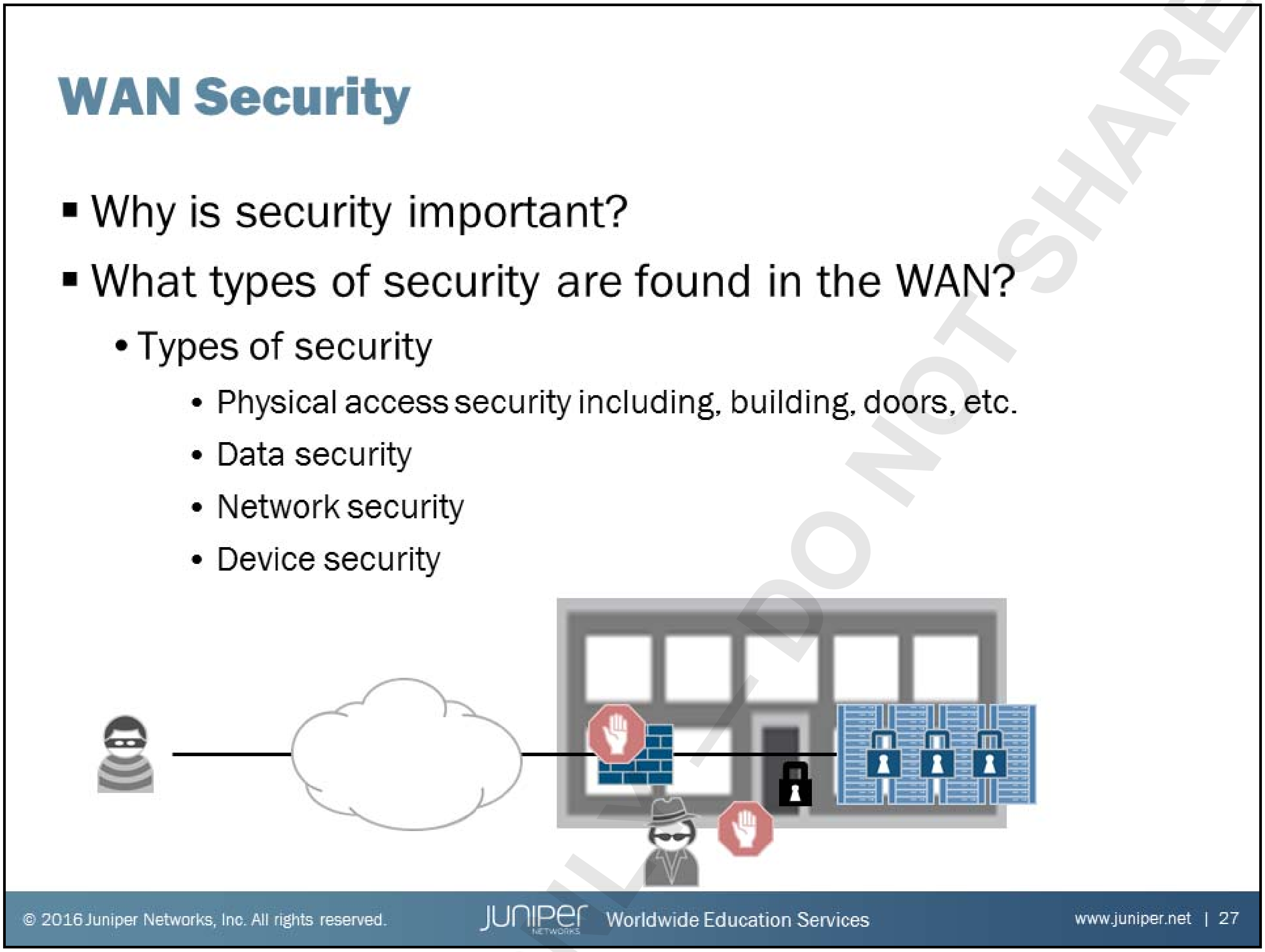
- Device virtualization in the WAN
  - vSRX
    - SRX device in a virtualized format
  - vMX
    - MX device in a virtualized format

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 26

SDN and Virtualization: Part 2

The Junos OS can also run as a virtual machine (VM) using either VMware or KVM as the host software. Two products are currently available—virtual SRX (vSRX) and virtual MX (vMX).

- vSRX: The vSRX Services Gateway (formerly known as Firefly Perimeter) delivers a complete virtual firewall solution, including advanced security, robust networking, and automated VM life-cycle management capabilities for service providers and enterprises. vSRX empowers security professionals to deploy and scale firewall protection in highly dynamic environments. Based on the SRX Series Services Gateways, vSRX extends the SRX Series capabilities to virtualized and cloud environments. The vSRX's automated provisioning capabilities, enabled through Junos Space Virtual Director, allow network and security administrators to quickly and efficiently provision and scale firewall protection to meet the dynamic needs of virtualized and cloud environments. By combining the vSRX's provisioning application with the power of Junos Space Security Director, administrators can significantly improve policy configuration, management, and visibility into both physical and virtual assets from a common, centralized platform.
- vMX: The virtual MX Series 3D Universal Edge Router extends over 15 years of Juniper Networks edge routing expertise to the virtual realm. The vMX is a full-featured, carrier-grade router with complete control, forwarding, and management planes. It runs the Junos OS, and supports vTrio packet handling and forwarding by compiling the programmable Junos Trio chipset microcode for x86 chipsets. With its granular, 'pay as you grow' licensing model, the vMX reduces the risk associated with new market entry and service innovation and allows you to start small, move fast, and stay profitable. Not only is it an ideal platform for markets and applications that are difficult to serve with traditional hardware routers, it is also a great option for proof of concept validation, lab testing, and feature and release certification.



WAN Security

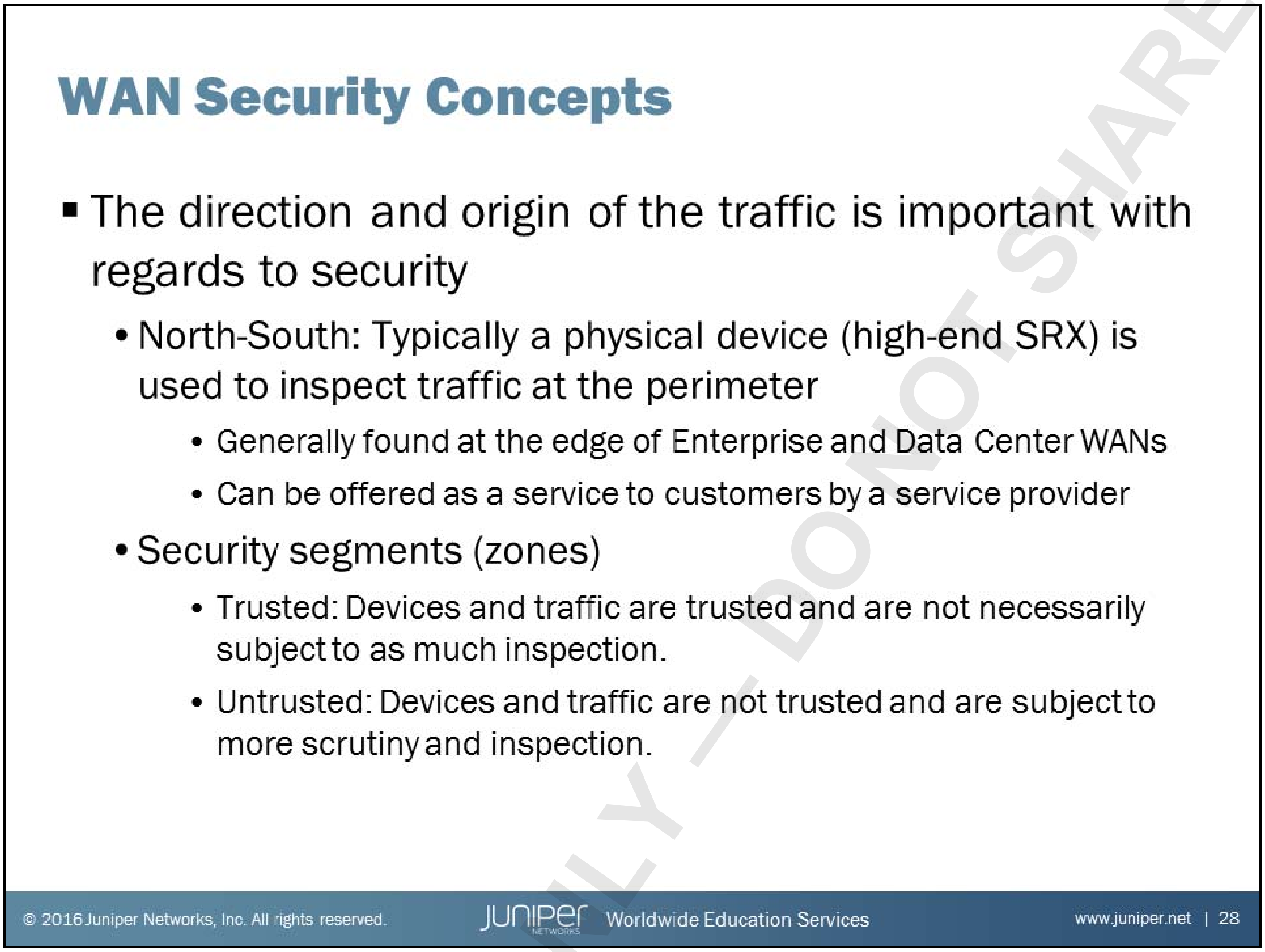
- Why is security important?
- What types of security are found in the WAN?
  - Types of security
    - Physical access security including, building, doors, etc.
    - Data security
    - Network security
    - Device security

JUNIPER NETWORKS Worldwide Education Services

www.juniper.net | 27

Why Is Security Important?

WAN security is an integral part of a design proposal and is extremely important because you are protecting the physical devices, personal information, and intellectual property of the customer. When dealing with security there are many aspect that must be considered. The physical security of the data center can seem unimportant but it is your first line of defense. The building and network should be secured and access should be limited to a few trust worthy employees. This might not be part of your design proposal, but should be discussed with the customer. In the WAN, it is important to ensure all aspects are protected including the network, devices, and data being passed without significantly impacting performance.



WAN Security Concepts

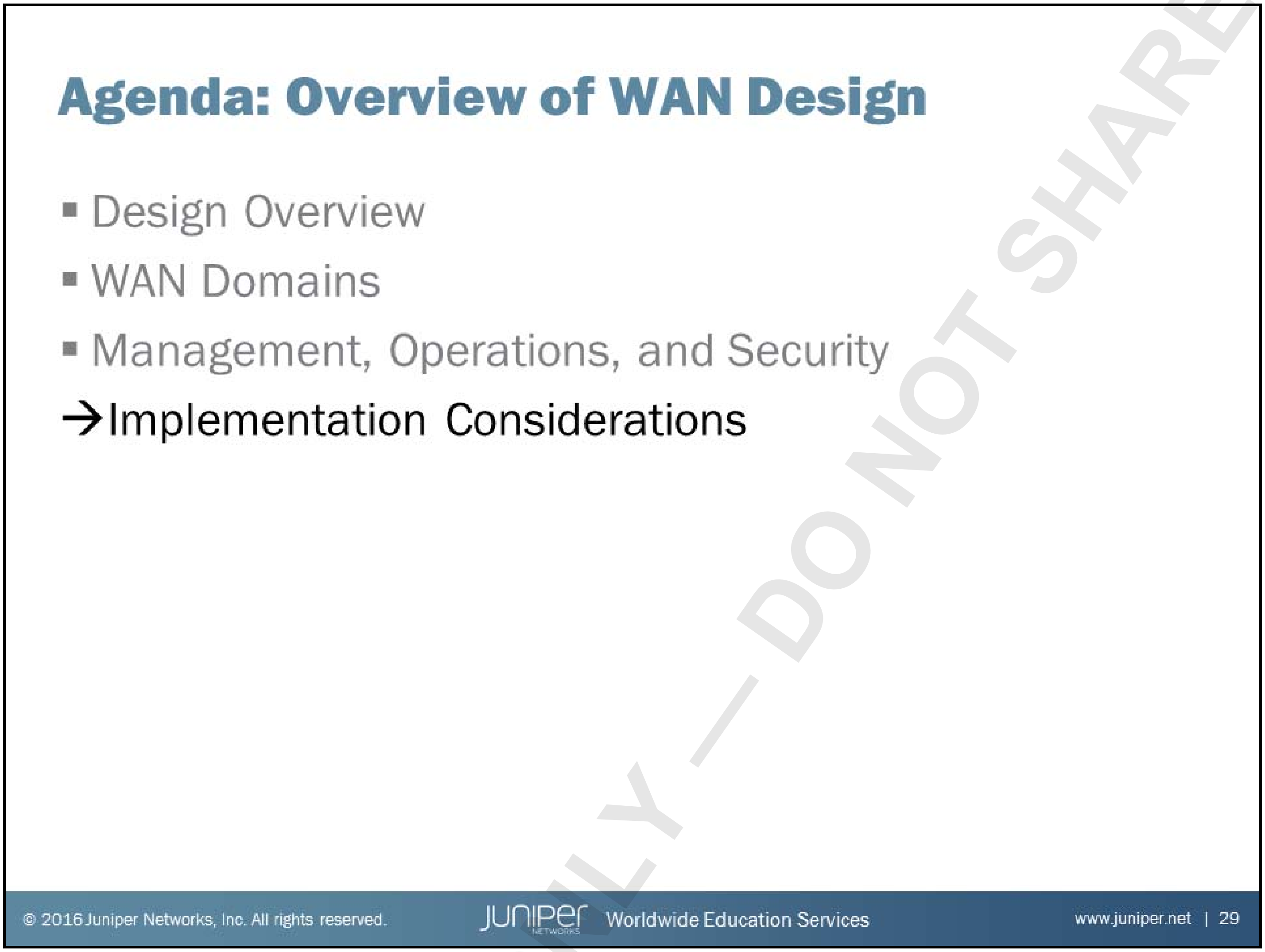
- The direction and origin of the traffic is important with regards to security
  - North-South: Typically a physical device (high-end SRX) is used to inspect traffic at the perimeter
    - Generally found at the edge of Enterprise and Data Center WANs
    - Can be offered as a service to customers by a service provider
  - Security segments (zones)
    - Trusted: Devices and traffic are trusted and are not necessarily subject to as much inspection.
    - Untrusted: Devices and traffic are not trusted and are subject to more scrutiny and inspection.

Juniper Networks Worldwide Education Services www.juniper.net | 28

Traffic Patterns

Understanding how traffic flows through and within your WAN is key to knowing how and where your security devices need to be placed. If you are designing a service provider WAN, you will not typically be dealing with security devices unless you are offering security as a service to your customers. You might have security devices incorporated in your management network depending on the corporate security policy. Security for enterprise and Data Center WANs is typically implemented as close to the edge as possible to avoid necessary access into the network.

Another key aspect of your security design is creating logical separation between areas in the data center by defining and implementing security zones. Then policies can be used to control traffic that is allowed into one zone but not into another. Basically, you take a trusted versus untrusted approach to traffic flowing through your network. Another method of segmentation is using virtual routing instances to create logical separation in your network. This approach allows you to continue to apply security policies while completely separating traffic using the firewall instead of other devices in the network.



Agenda: Overview of WAN Design

- Design Overview
- WAN Domains
- Management, Operations, and Security

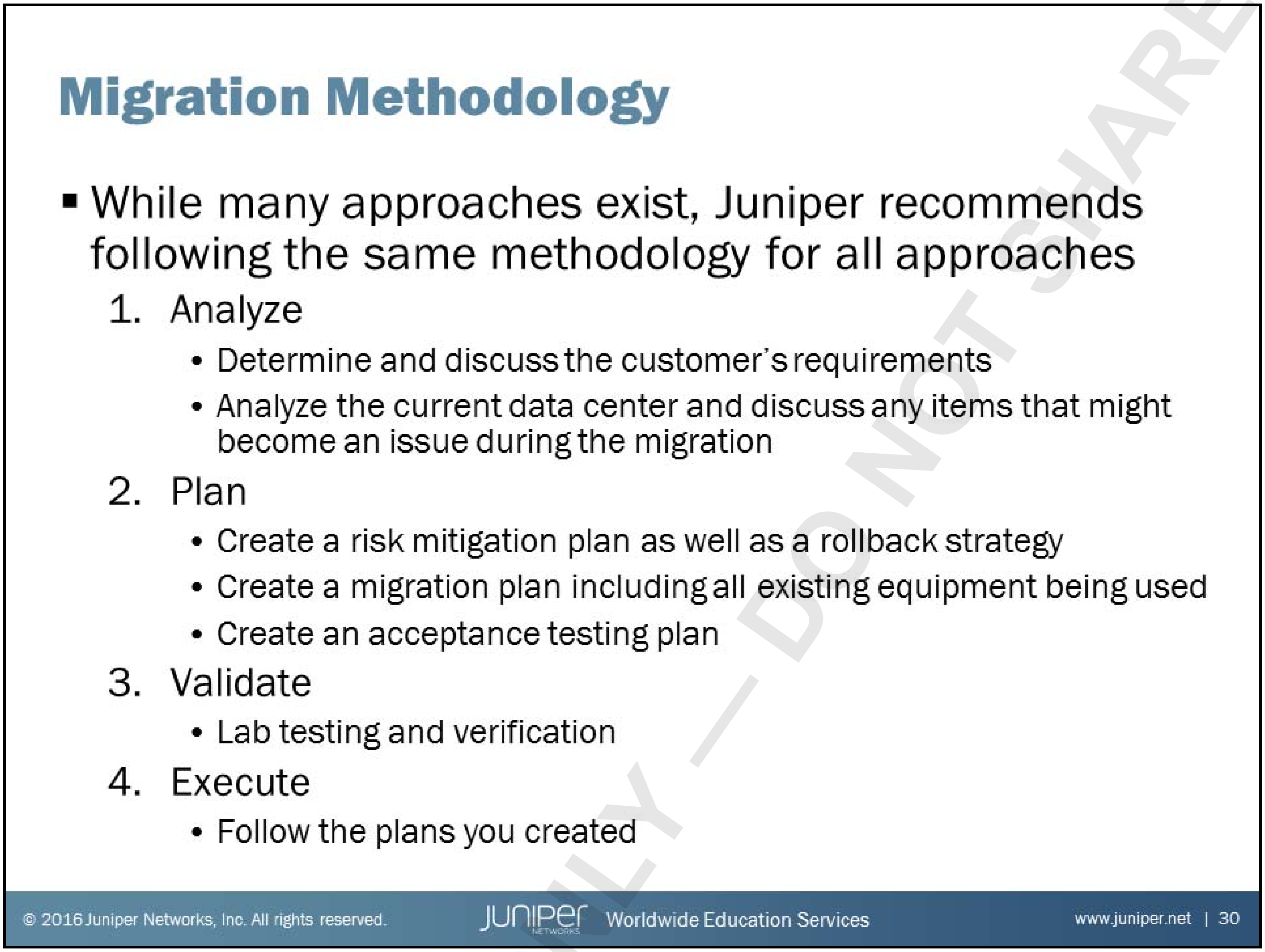
→Implementation Considerations

Worldwide Education Services

www.juniper.net | 29

Implementation Considerations

The slide highlights the topic we discuss next.

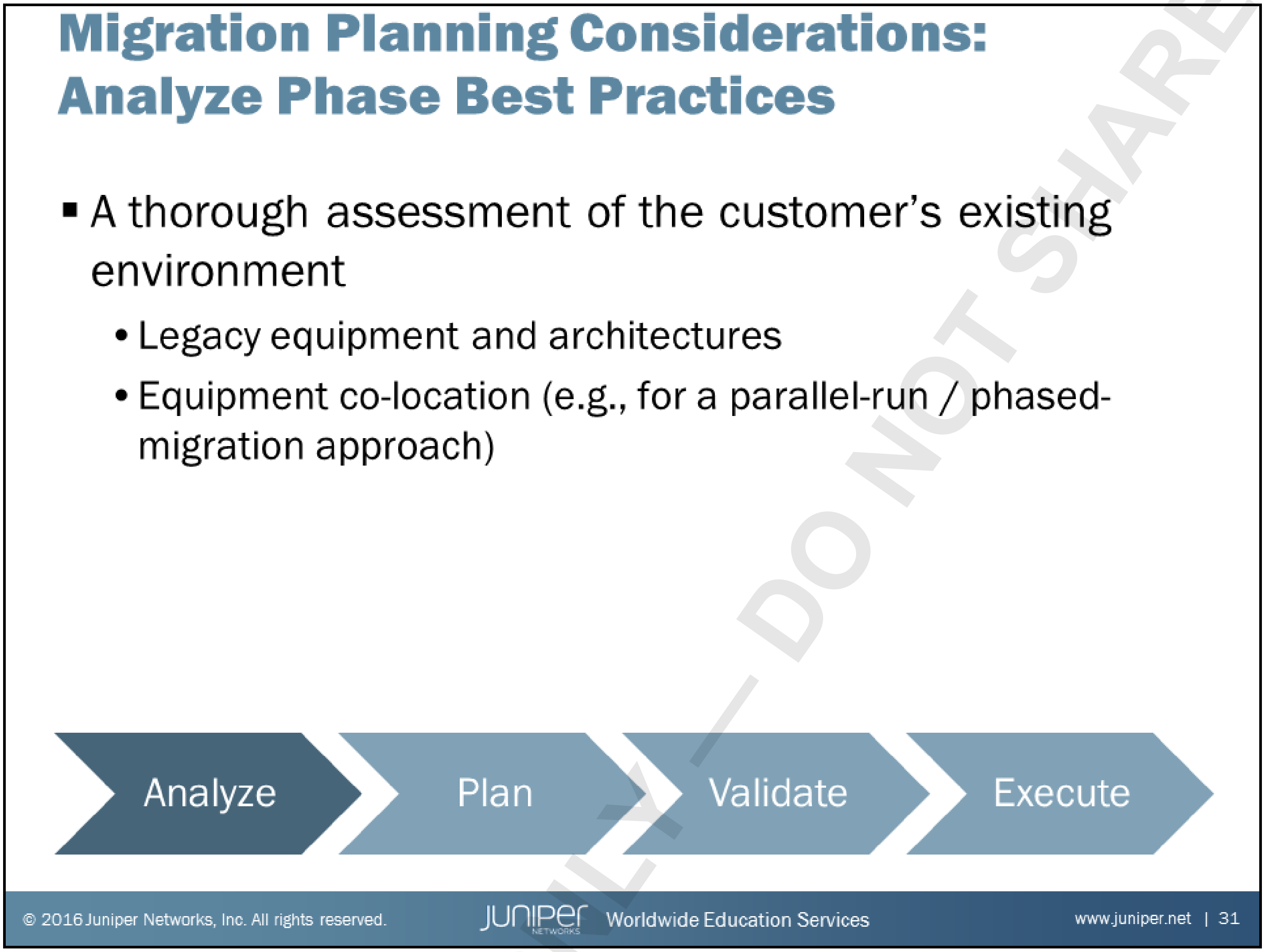


Migration Methodology

- While many approaches exist, Juniper recommends following the same methodology for all approaches
  1. Analyze
     - Determine and discuss the customer's requirements
     - Analyze the current data center and discuss any items that might become an issue during the migration
  2. Plan
     - Create a risk mitigation plan as well as a rollback strategy
     - Create a migration plan including all existing equipment being used
     - Create an acceptance testing plan
  3. Validate
     - Lab testing and verification
  4. Execute
     - Follow the plans you created

Common Migration Methodology

Juniper Networks follows a common methodology when performing a migration, regardless of the size or scope of the project. The methodology requires that you not only understand the current state of the customer's network, but that you also understand what the customer desires when the project is complete. The slide highlight a few of the common tasks associated with each of the four major steps in any migration process.



Migration Planning Considerations: Analyze Phase Best Practices

- A thorough assessment of the customer's existing environment
  - Legacy equipment and architectures
  - Equipment co-location (e.g., for a parallel-run / phased-migration approach)

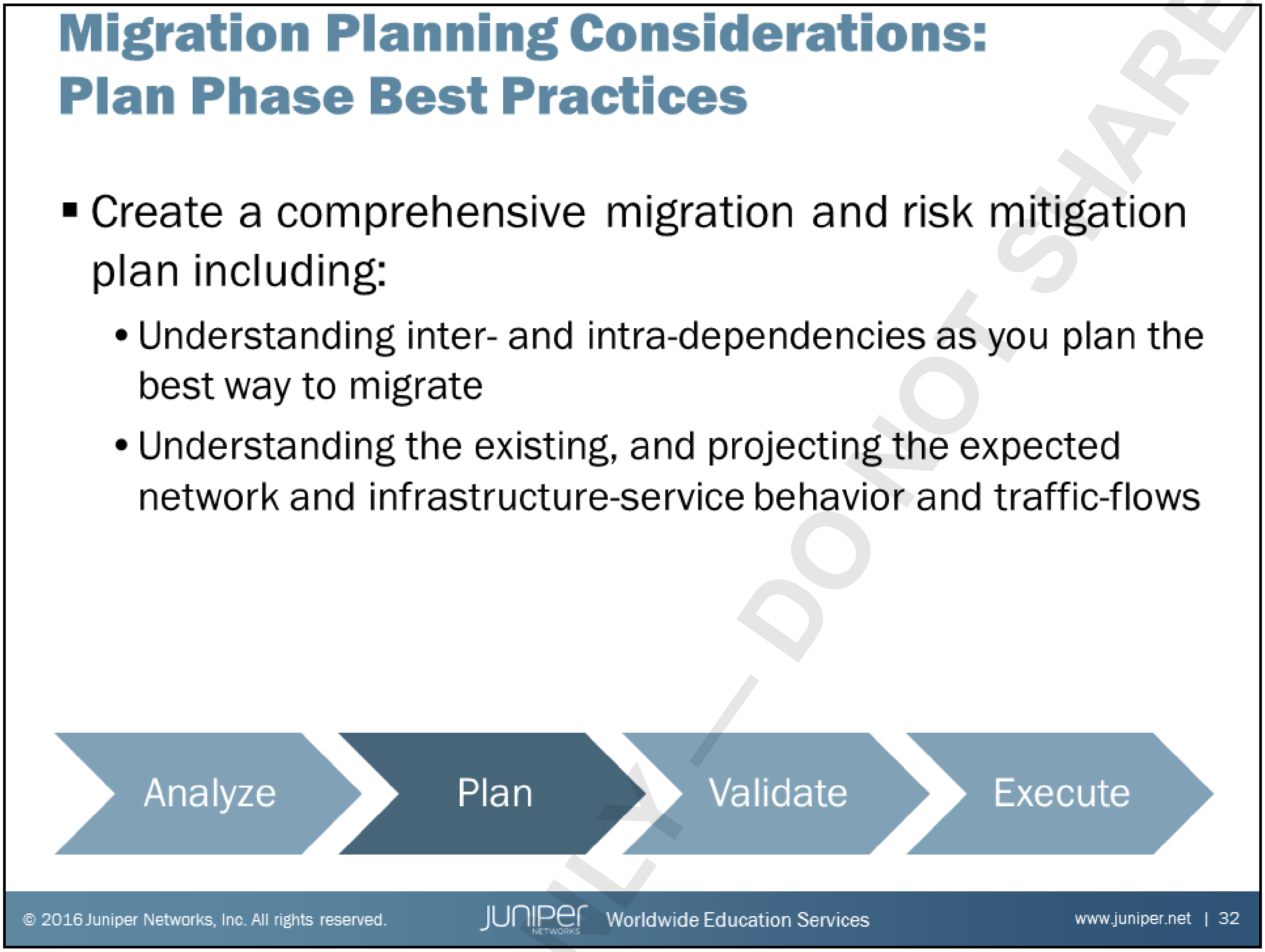
Analyze → Plan → Validate → Execute

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 31

Analyze Phase: Best Practices

As part of the design phase, take special care to get a thorough assessment of the customer's current network including:

- Legacy equipment and architectures
  - What legacy equipment & architectures have to be retained and integrated into the new network?
  - What legacy protocols have to be accommodated?
  - What interoperability requirements have to be planned for and validated?
- Equipment co-location for a parallel-run / phased-migration approach
  - Physical space requirements: Determine if there is enough rack and stack space, cable and patch-panel capacity to accommodate both old and new network equipment.
  - Power / cooling requirement: Determine if there is enough power and cooling capabilities to accommodate the equipment.



Migration Planning Considerations: Plan Phase Best Practices

- Create a comprehensive migration and risk mitigation plan including:
  - Understanding inter- and intra-dependencies as you plan the best way to migrate
  - Understanding the existing, and projecting the expected network and infrastructure-service behavior and traffic-flows

Analyze
Plan
Validate
Execute

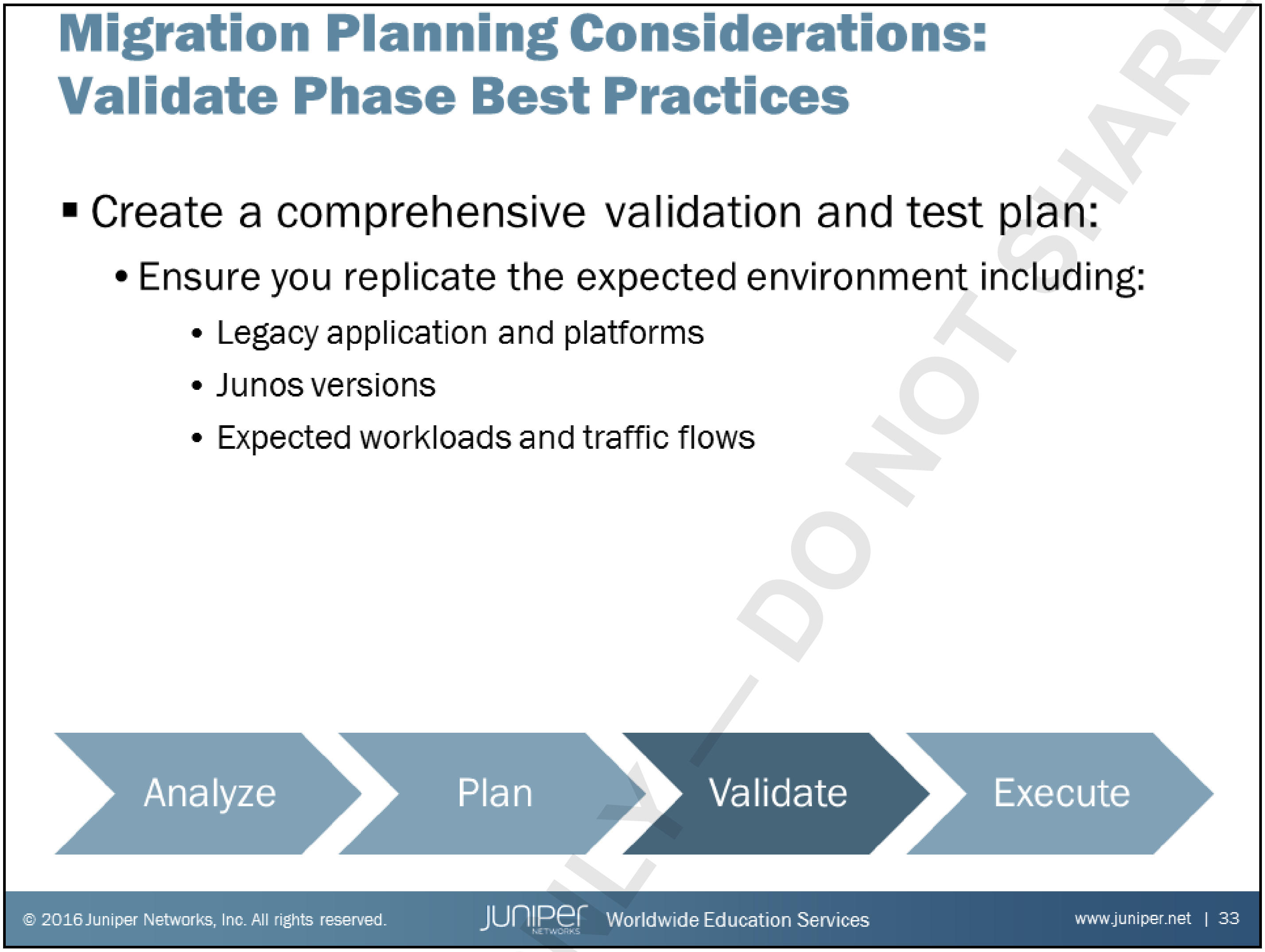
Worldwide Education Services

www.juniper.net | 32

Plan Phase: Best Practices

A comprehensive migration and risk mitigation plan should include:

- A detailed understanding of the inter- and intra- dependencies as we plan the best way to migrate:
  - Connections to upstream providers
  - Customer connections and services
  - Other functional blocks like out-of-band (OOB) management infrastructure
- A detailed understanding of the existing, and projected network and infrastructure-service behavior and traffic-flows.
  - In the existing domain
  - During the migration (interim environment in case of linkage and parallel-run)
  - After the migration



Migration Planning Considerations: Validate Phase Best Practices

- Create a comprehensive validation and test plan:
  - Ensure you replicate the expected environment including:
    - Legacy application and platforms
    - Junos versions
    - Expected workloads and traffic flows

Analyze
Plan
Validate
Execute

Worldwide Education Services

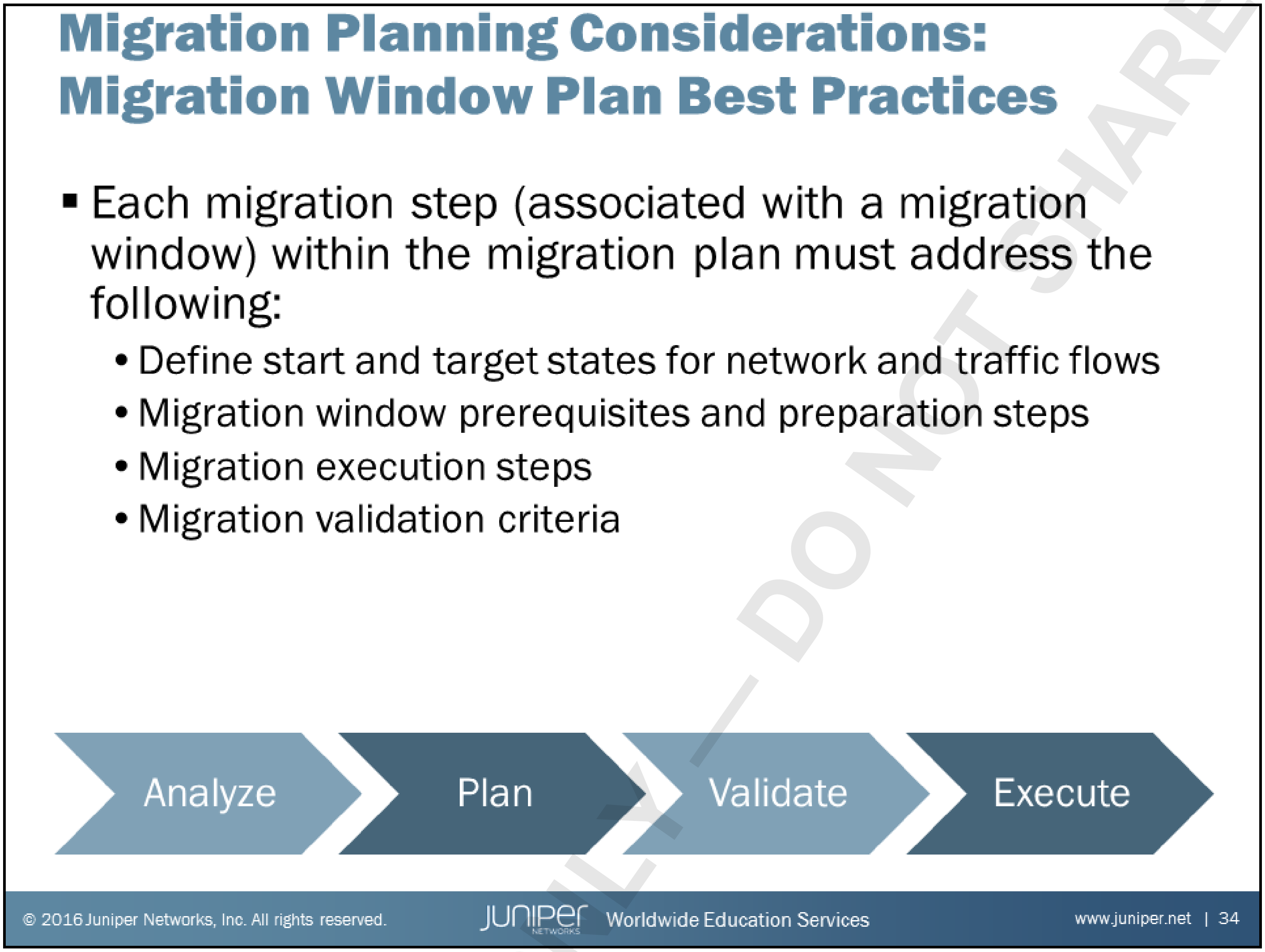
www.juniper.net | 33

Validate Phase: Best Practices

A comprehensive validation test plan should include:

- Detailed information about device and services that must be provided by the proposed network.
  - Legacy devices that must be incorporated as well as any third-party devices that should be included.
  - Ensure you match any unique hardware requirements that are included in the proposal.
- Junos versions
  - A list of features that need to be verified as a result of the product issue impact review (PIIR).
  - A list of potential problems or customer concerns that need to be tested.
- Traffic and unique applications that need to be verified.

A good testing plan will match the expected final network design. This phase is where you can relieve any customer concerns about performance, potential problems, and over all functionality. This phase will help you identify the process for implementation and learn what works and what does not. It also allows you time and flexibility for fixing any issues you might come across without an impact to production traffic and customer SLAs.



Migration Planning Considerations: Migration Window Plan Best Practices

- Each migration step (associated with a migration window) within the migration plan must address the following:
  - Define start and target states for network and traffic flows
  - Migration window prerequisites and preparation steps
  - Migration execution steps
  - Migration validation criteria

Analyze → Plan → Validate → Execute

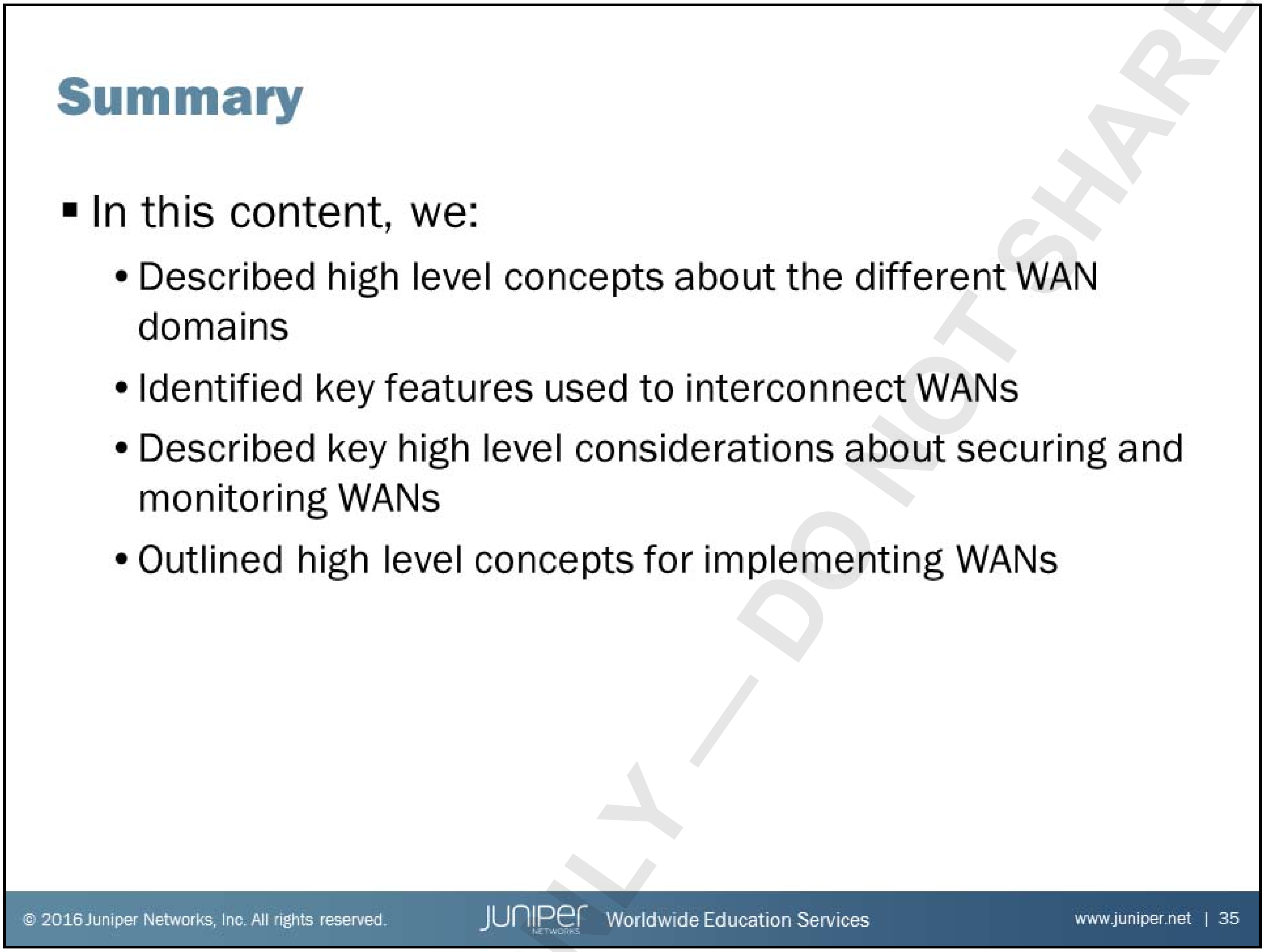
Juniper Networks Worldwide Education Services

www.juniper.net | 34

Migration Window Plan: Best Practices

Each migration step (associated with a migration window) within the migration plan needs to address, as a minimum, the following:

- The starting state of the network and traffic flows.
- The target state of the network and traffic flows.
- The migration window prerequisites and preparation steps;
  - Expected maintenance and downtime provision required.
- The migration execution steps;
  - Showing start and end of expected service disruption.
- The migration validation criteria including:
  - Defined migration rollback steps (in case migration validation fails); and
  - Detailed post-migration observation plan (in case migration validation is successful).



Summary

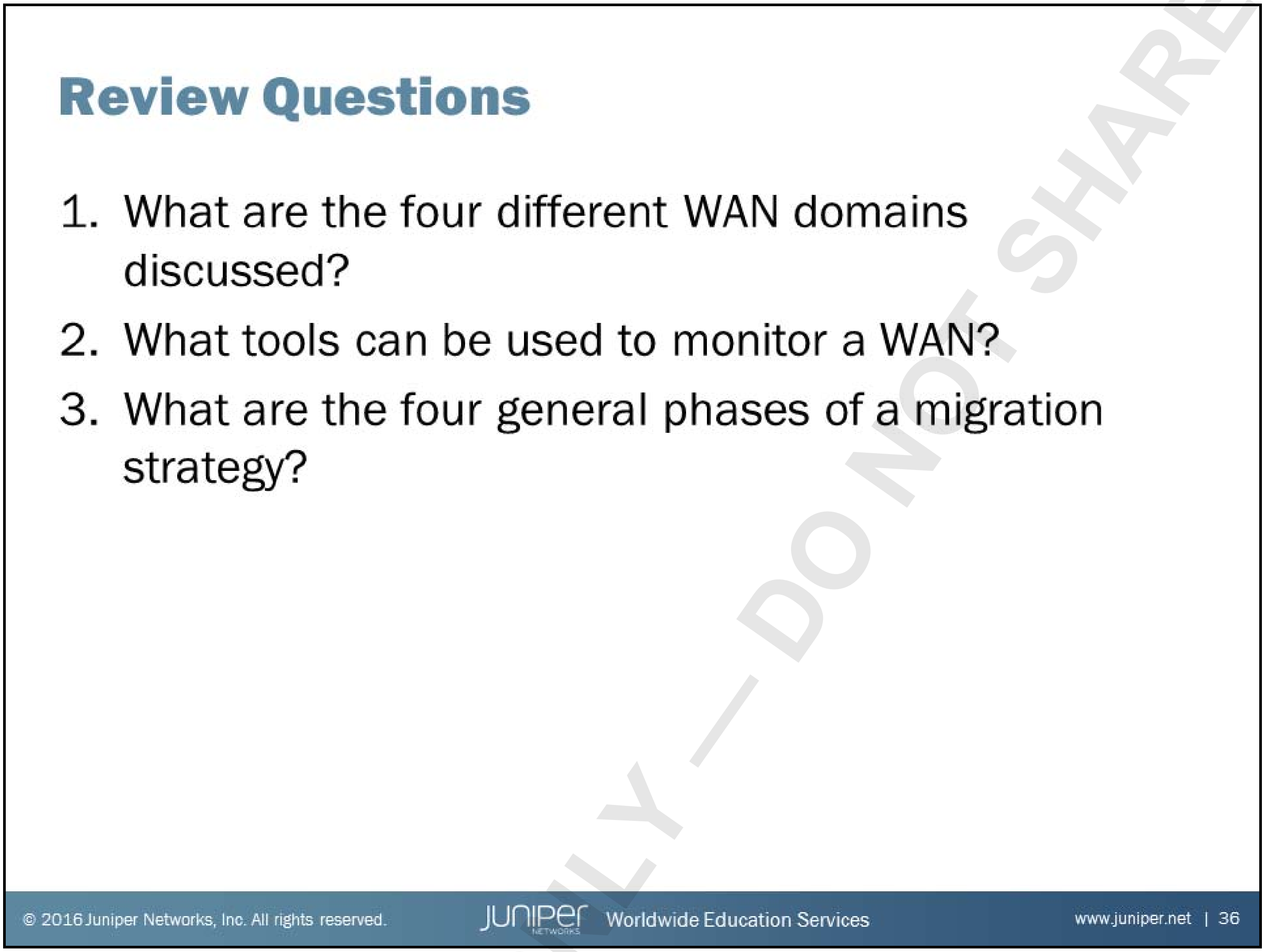
- In this content, we:
  - Described high level concepts about the different WAN domains
  - Identified key features used to interconnect WANs
  - Described key high level considerations about securing and monitoring WANs
  - Outlined high level concepts for implementing WANs

Worldwide Education Services

www.juniper.net | 35

We Discussed:

- High level concepts about the different WAN domains;
- Features used to interconnect WANs;
- Key high level considerations about securing and monitoring WANs; and
- High level concepts for implementing WANs.



Review Questions

1. What are the four different WAN domains discussed?
2. What tools can be used to monitor a WAN?
3. What are the four general phases of a migration strategy?

Worldwide Education Services

Review Questions

1.

2.

3.

Answers to Review Questions

1.

The four WAN domains discussed are service provider core, service provider edge, enterprise WAN, and data center WAN.

2.

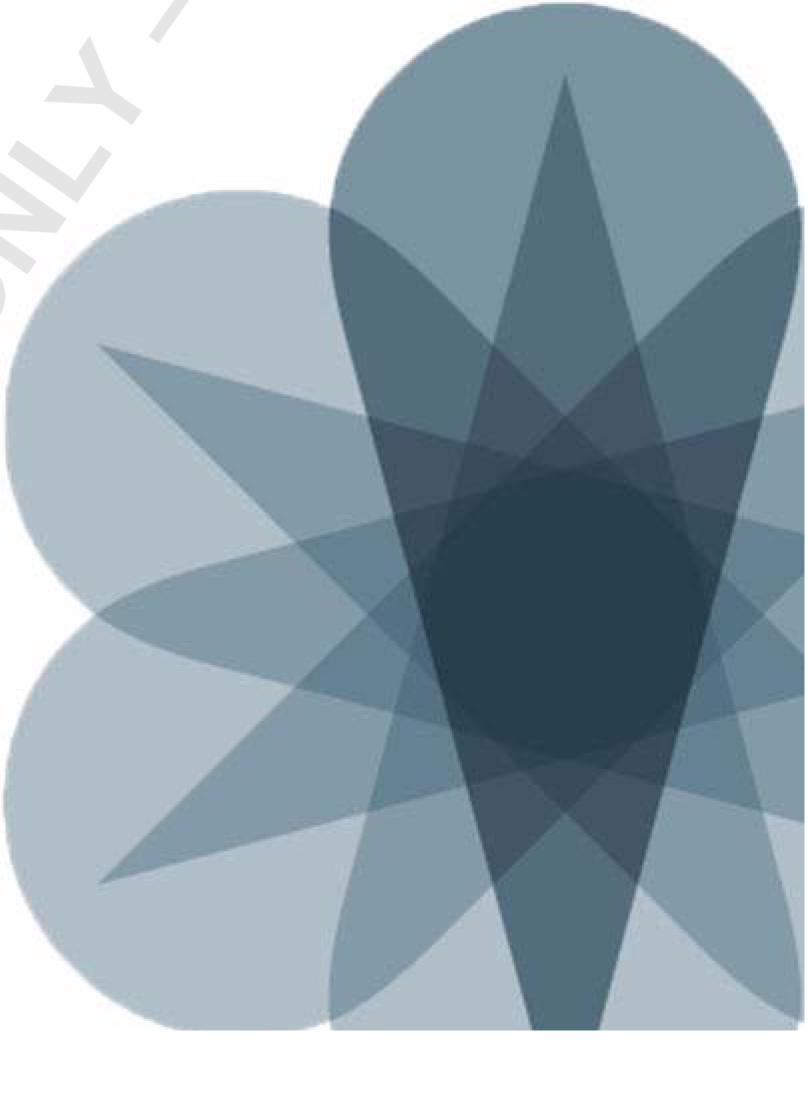
There are a few tools that can be used to manage a WAN including WANDL, NorthStar, Junos Space, Juniper Secure Analytics, and automation.

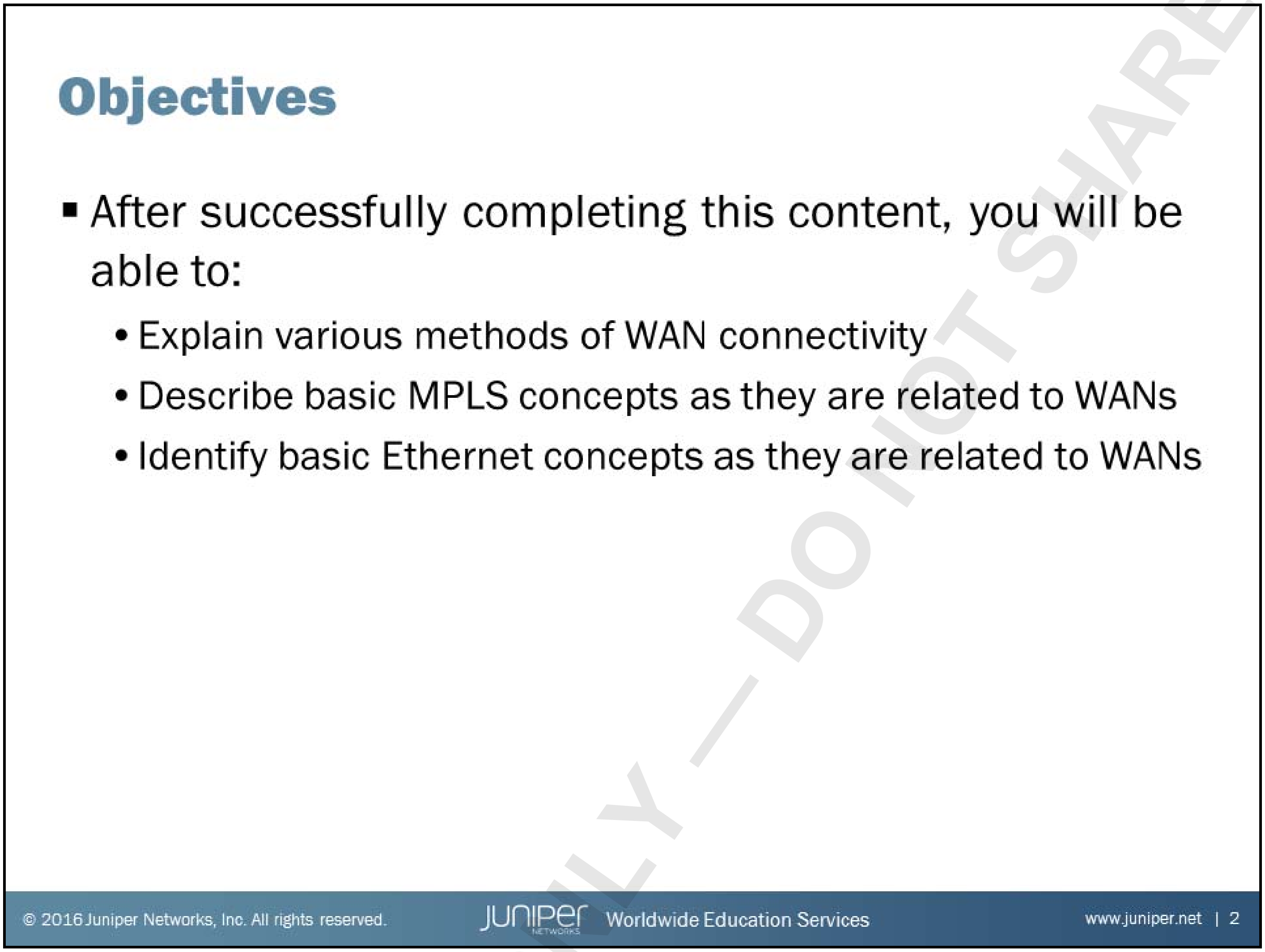
3.

The four phases of a migration strategy are analyze, plan, validate and execute.

Juniper Networks Design—WAN

Chapter 3: WAN Connectivity





Objectives

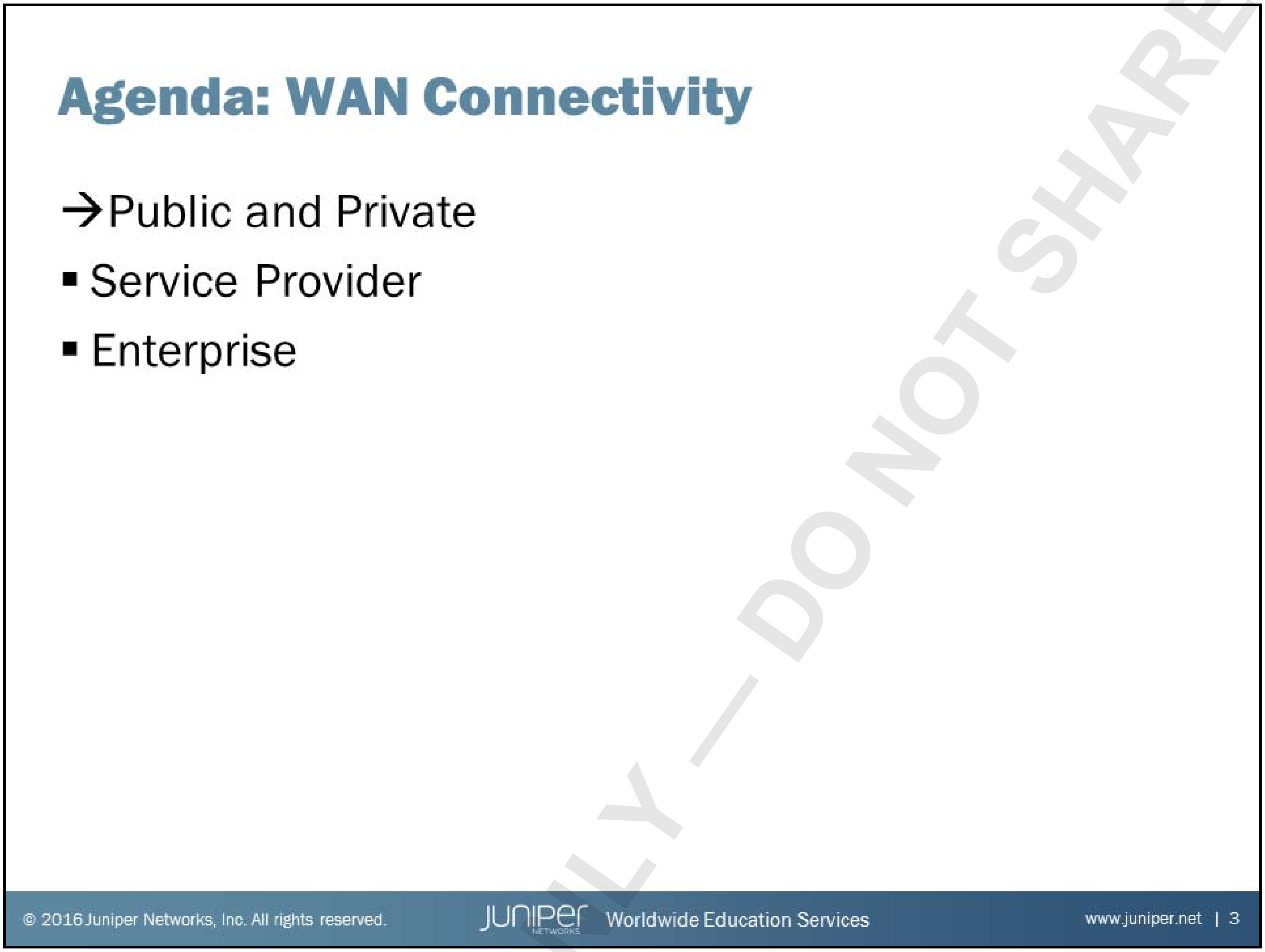
- After successfully completing this content, you will be able to:
  - Explain various methods of WAN connectivity
  - Describe basic MPLS concepts as they are related to WANs
  - Identify basic Ethernet concepts as they are related to WANs

Juniper NETWORKS Worldwide Education Services

www.juniper.net | 2

We Will Discuss:

- Several methods of WAN connectivity;
- How MPLS works and its relation to WANs; and
- Basic Ethernet concepts and how they apply to WANs.



Agenda: WAN Connectivity

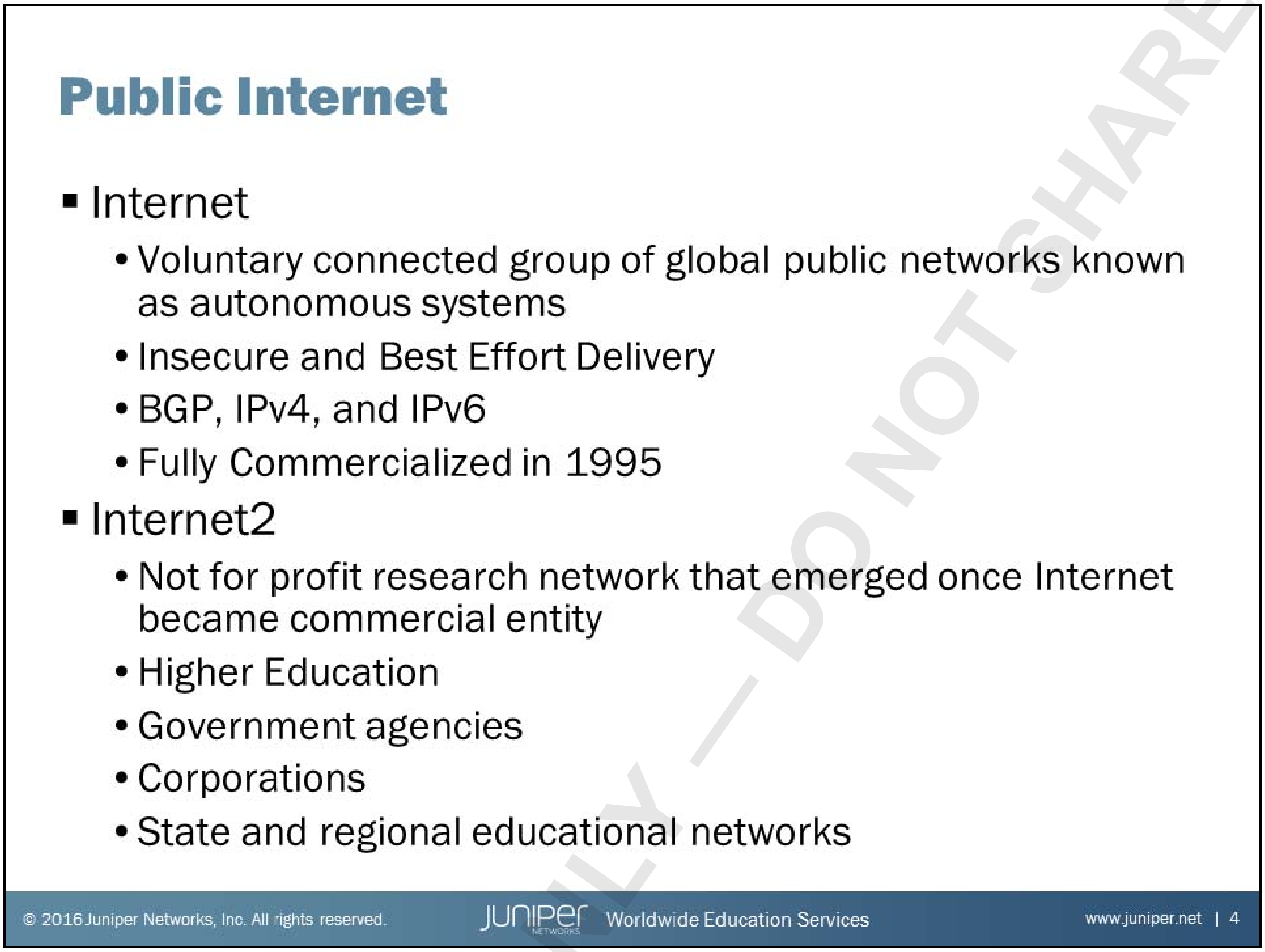
- →Public and Private
- Service Provider
- Enterprise

Juniper Networks Worldwide Education Services

www.juniper.net | 3

SDN Overview

The slide lists the topics we will discuss. We discuss the highlighted topic first.



Public Internet

- Internet
  - Voluntary connected group of global public networks known as autonomous systems
  - Insecure and Best Effort Delivery
  - BGP, IPv4, and IPv6
  - Fully Commercialized in 1995
- Internet2
  - Not for profit research network that emerged once Internet became commercial entity
  - Higher Education
  - Government agencies
  - Corporations
  - State and regional educational networks

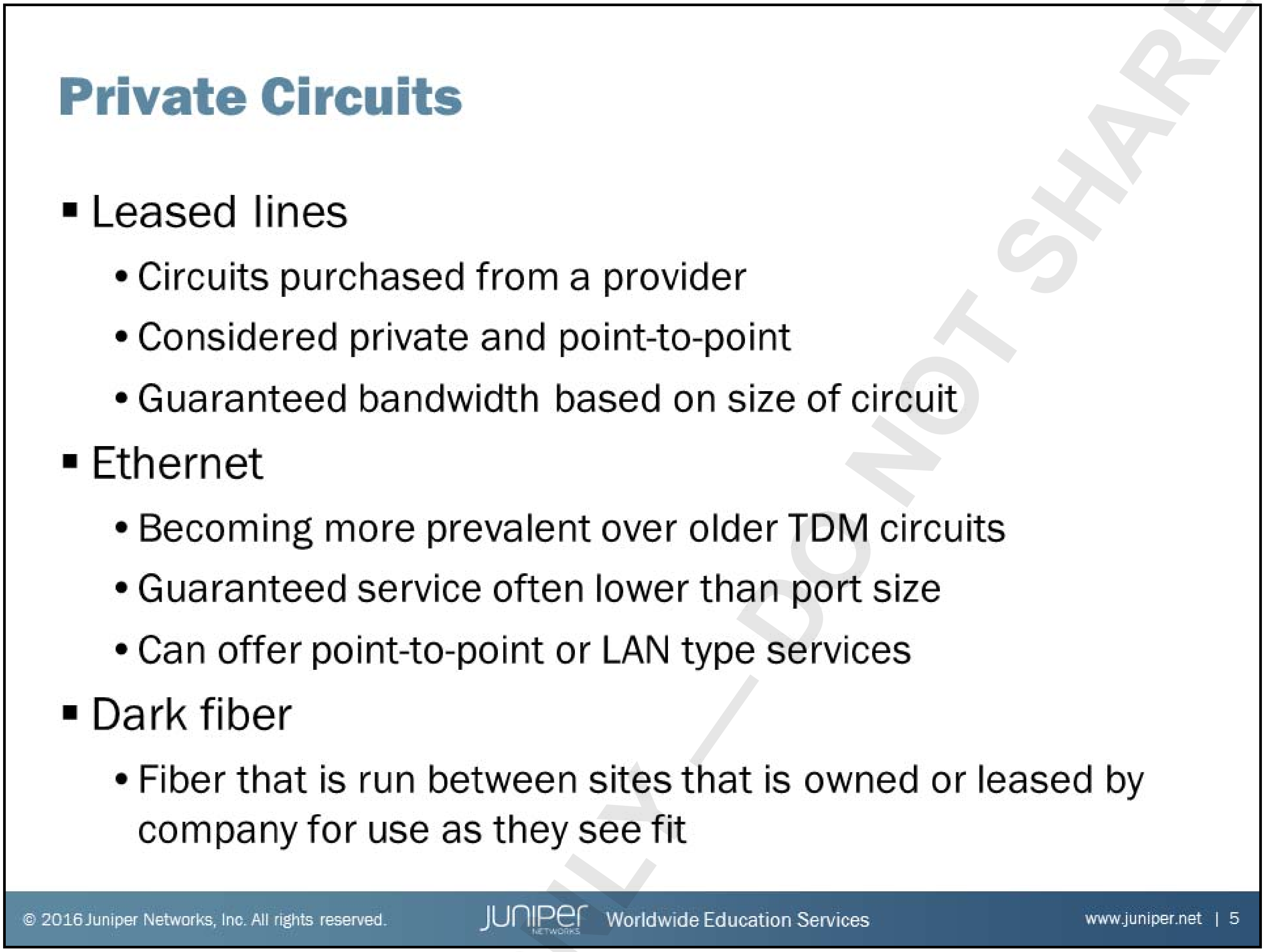
Juniper Networks Worldwide Education Services www.juniper.net | 4

Internet

The Internet started out as a research network that became fully commercialized in 1995, connecting billions of users and millions of networks together globally. The Internet is now a public network running on IP protocols that serve commercial, government, non-profit, and research institutions around the world. The Internet is inherently an insecure network and many protocols have been adopted to overlay the Internet’s IP network to increase security. The Internet is also categorized by best effort delivery, offering no real quality of service guarantees.

Internet2

Internet2 is another large IP network primarily focused on providing access to research institutions. Similar in nature to what the original Internet was when it was first created. It allows participating members to communicate with each other on a dedicated IP network that is not quite as open as the commercial Internet. Internet2 has members and a board of trustees. While not as common the commercial Internet, it is not unusual to find Internet2 connections in networks today.



Private Circuits

- Leased lines
  - Circuits purchased from a provider
  - Considered private and point-to-point
  - Guaranteed bandwidth based on size of circuit
- Ethernet
  - Becoming more prevalent over older TDM circuits
  - Guaranteed service often lower than port size
  - Can offer point-to-point or LAN type services
- Dark fiber
  - Fiber that is run between sites that is owned or leased by company for use as they see fit

Worldwide Education Services

www.juniper.net | 5

Leased Lines

Leased lines are dedicated circuits provisioned by a service provider for connectivity between sites. Generally speaking, a customer will order a circuit to terminate at two sites and the service provider will provision that circuit through their network. Once provisioned, the expectation would be the bandwidth of that circuit is always available between the two sites and the customer is responsible for attaching networking devices and managing the traffic on that circuit.

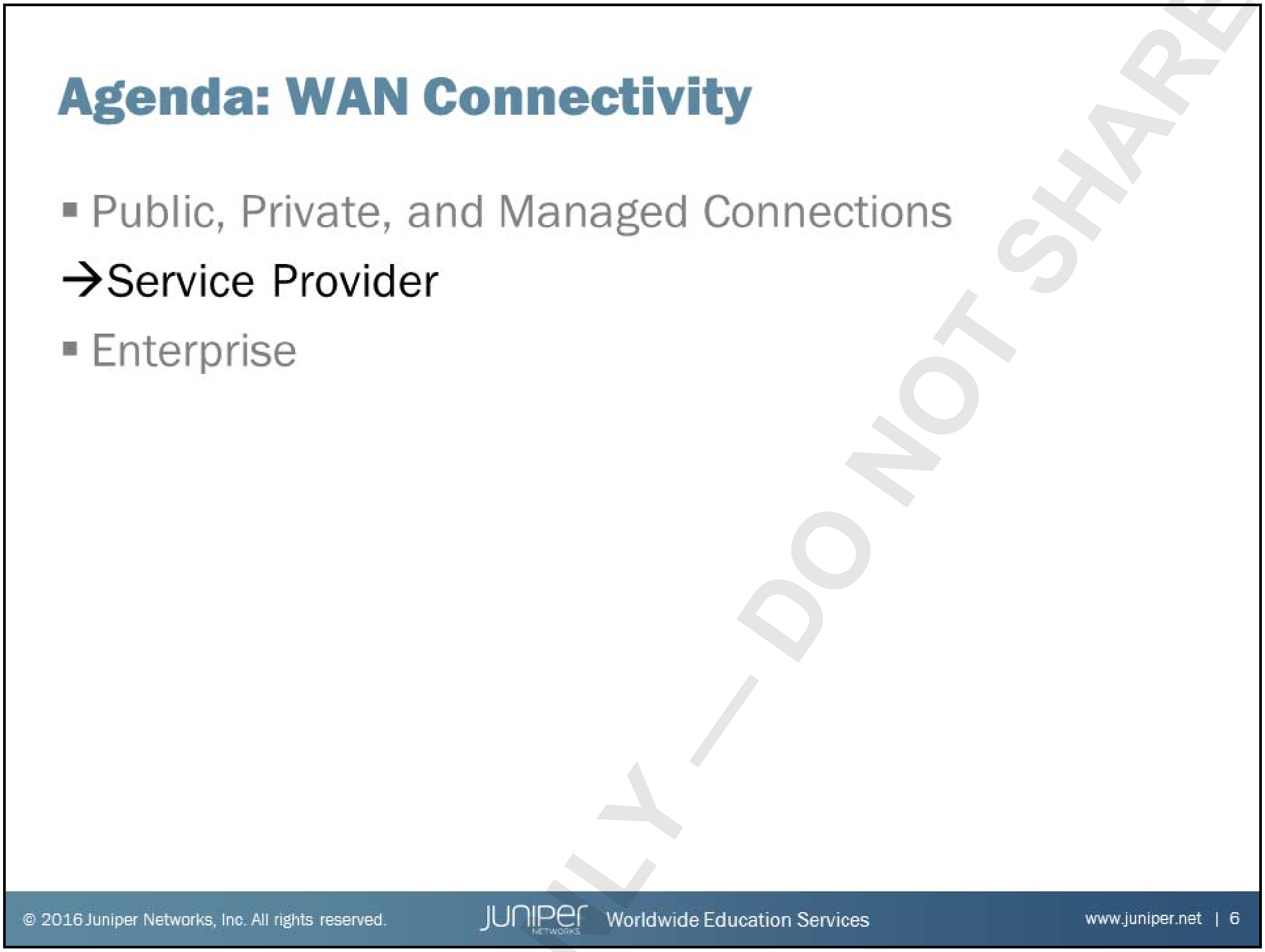
Ethernet

As Ethernet has grown in popularity, it is replacing dedicated leased lines as means of WAN transport. In this case, the circuit is provisioned with a port speed and a guaranteed rate. Port speed is just how fast you connect to the provider's network. They can provision any sized Ethernet port that is available on the network (10M, 100M, 1G, 10G, etc.). The actual guaranteed throughput can be well below the actual port speed. This allows service providers to offer up lower cost options to customers. When attaching to the network, customers need to shape their traffic to the actual provisioned circuit size (or the provider will do it for them by dropping excess traffic).

Ethernet services can be deployed as point-to-point or as a LAN type service. Generally for WAN deployments point-to-point is used but in metro areas LAN might be preferred.

Dark Fiber

Dark fiber is the term for unlit fiber between two sites. Here a customer can own fiber that runs between buildings or even between sites within a metropolitan area. This fiber can be used as the customer sees fit and just needs the correct optical connections to light the fiber and make it work.



Agenda: WAN Connectivity

- Public, Private, and Managed Connections

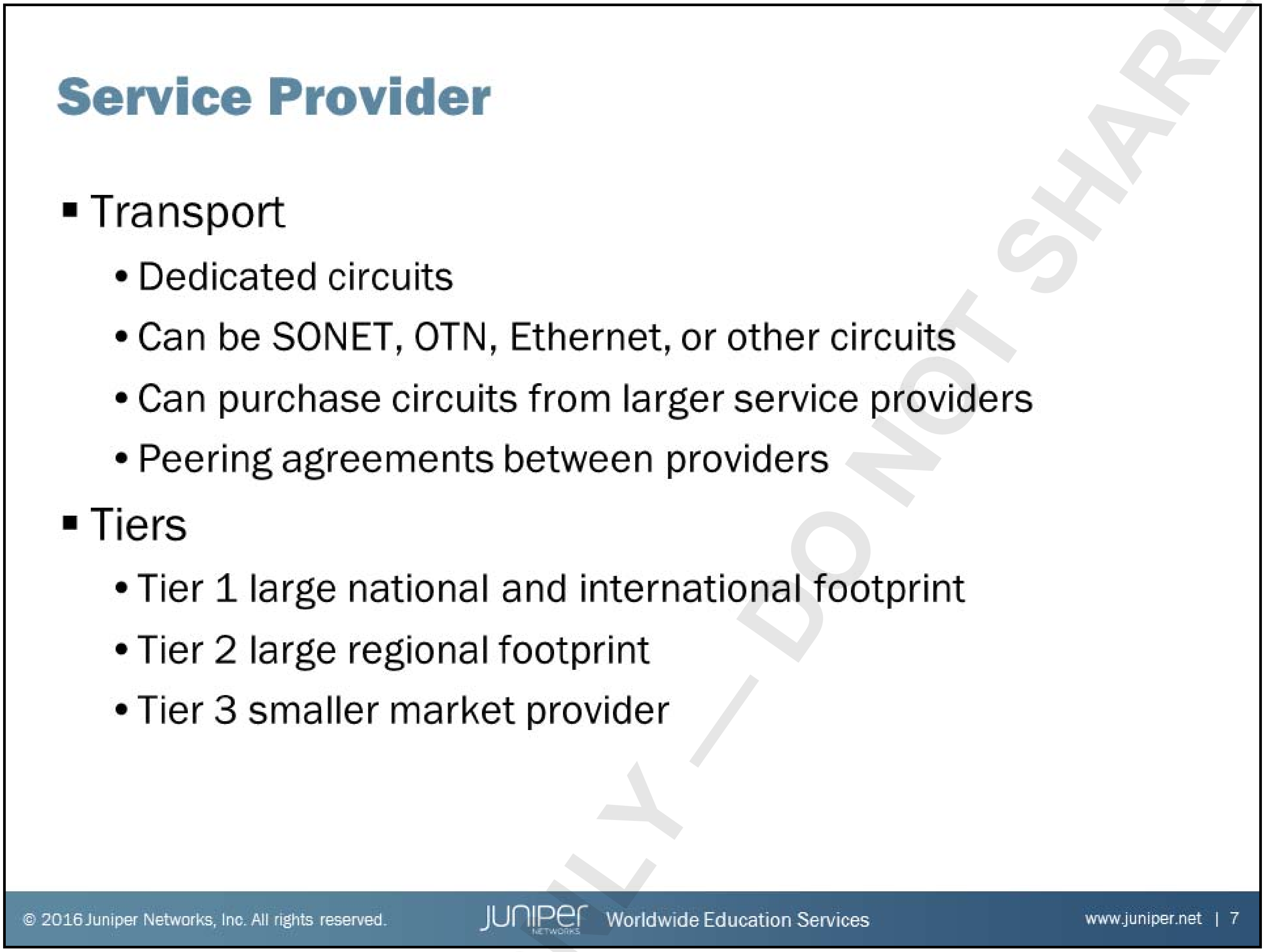
→Service Provider

- Enterprise

Worldwide Education Services www.juniper.net | 6

Service Provider

The slide highlights the topic we discuss next.



Service Provider

- Transport
  - Dedicated circuits
  - Can be SONET, OTN, Ethernet, or other circuits
  - Can purchase circuits from larger service providers
  - Peering agreements between providers
- Tiers
  - Tier 1 large national and international footprint
  - Tier 2 large regional footprint
  - Tier 3 smaller market provider

Worldwide Education Services

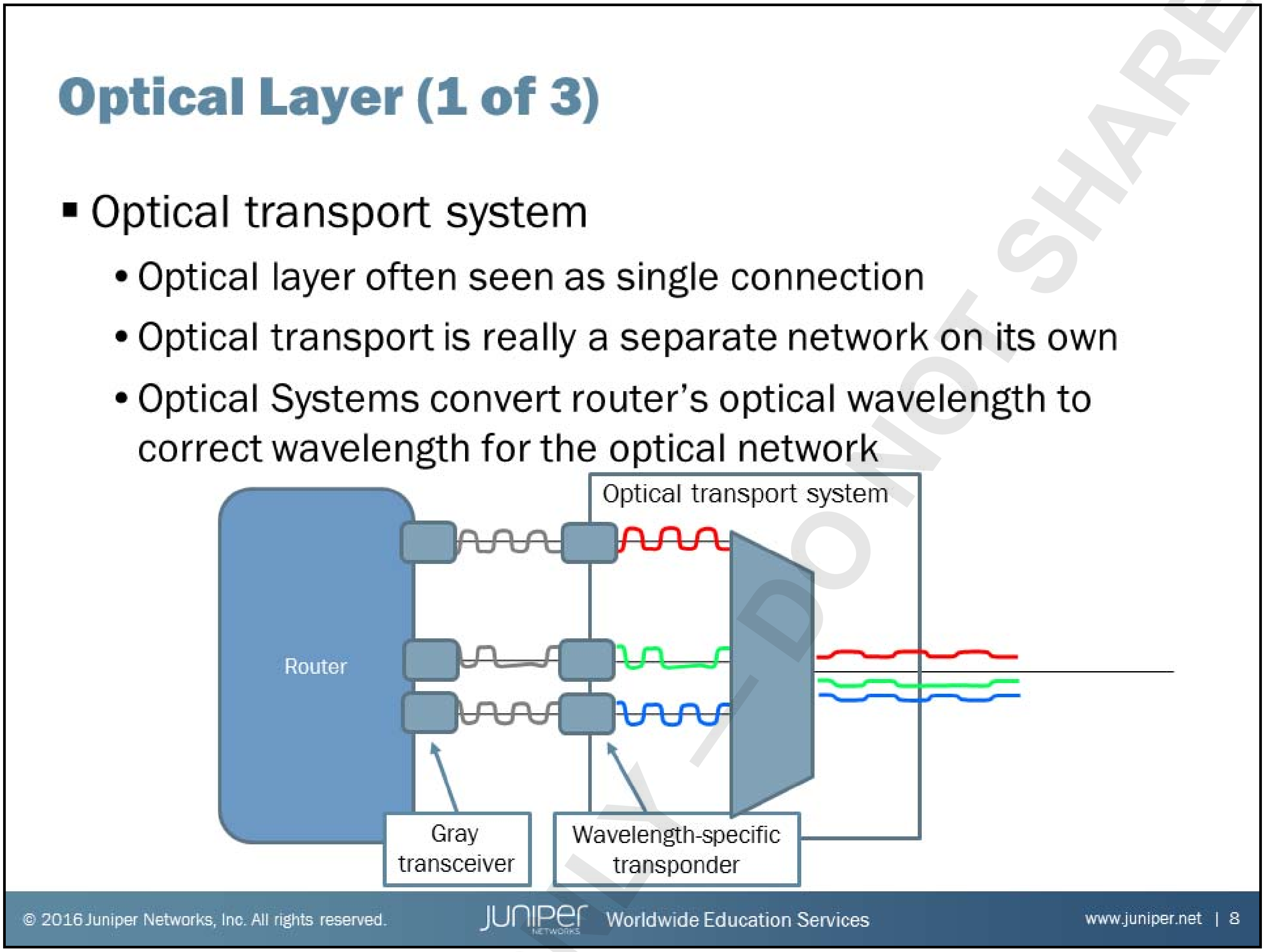
www.juniper.net | 7

Transport

Service providers generally use dedicated circuits to build their network. This can be SONET, Ethernet, or other legacy transport. They also try and optimize their networks to have enough bandwidth available to sustain outages and still maintain their service level agreements (SLAs). Generally this means provisioning a lot more bandwidth than they actually use when everything is operating properly. Some providers might purchase circuits from other providers and have peering agreements between providers.

Service Provider Tiers

Often you will hear providers referred to as Tier 1, Tier 2, or Tier 3 providers. There is no official definition on what constitutes these tiers but generally Tier 1 providers are national or global networks who provision their own circuits and wholesale services to other smaller providers. Tier 2 providers are large regional or multi-regional players. They can also have their own circuits but probably purchase many of their circuits from Tier 1 providers. Tier 3 providers are smaller providers who cover metro or regional areas and almost always purchase circuits from larger providers. They might also wholesale services for larger providers as well.



Optical Layer (1 of 3)

- Optical transport system
  - Optical layer often seen as single connection
  - Optical transport is really a separate network on its own
  - Optical Systems convert router's optical wavelength to correct wavelength for the optical network

Router

Optical transport system

Gray transceiver

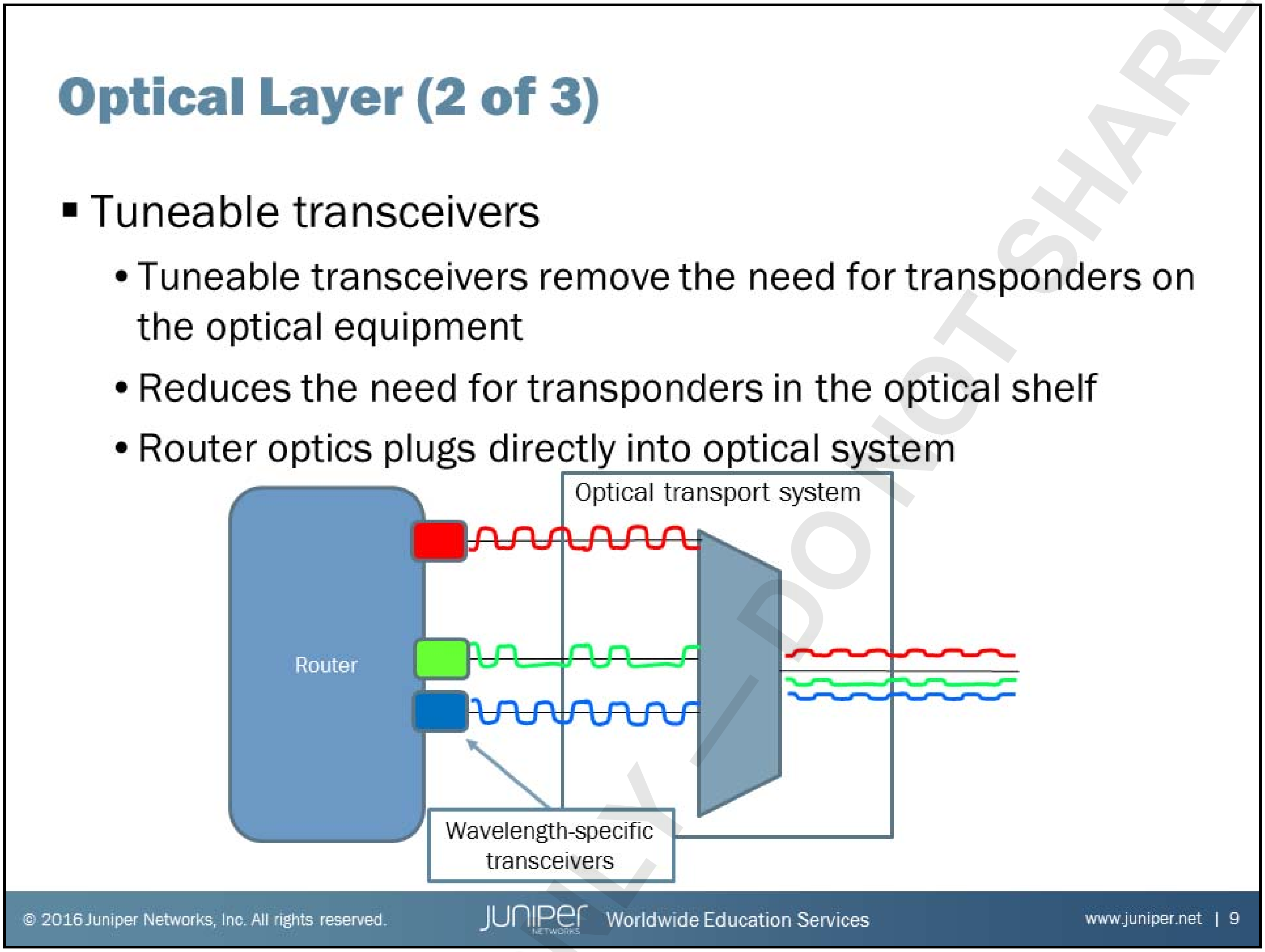
Wavelength-specific transponder

Worldwide Education Services

www.juniper.net | 8

Optical Layer: Part 1

With the optical layer, routers historically use what is known as a gray interface to plug into Optical Transport System (OTS), which maps that optical signal to the correct wavelength required by circuit in the optical network. The OTS can multiplex a bunch of different wavelengths on a single fiber, or can send the wavelength to another fiber as well. This is the entry point to the optical network. The router's optic is cabled to a wavelength-specific transponder on the optical system, typically housed on a shelf. Gray optics are fixed to a specific wavelength (for example,1310nm on SMF).



Optical Layer (2 of 3)

- Tuneable transceivers
  - Tuneable transceivers remove the need for transponders on the optical equipment
  - Reduces the need for transponders in the optical shelf
  - Router optics plugs directly into optical system

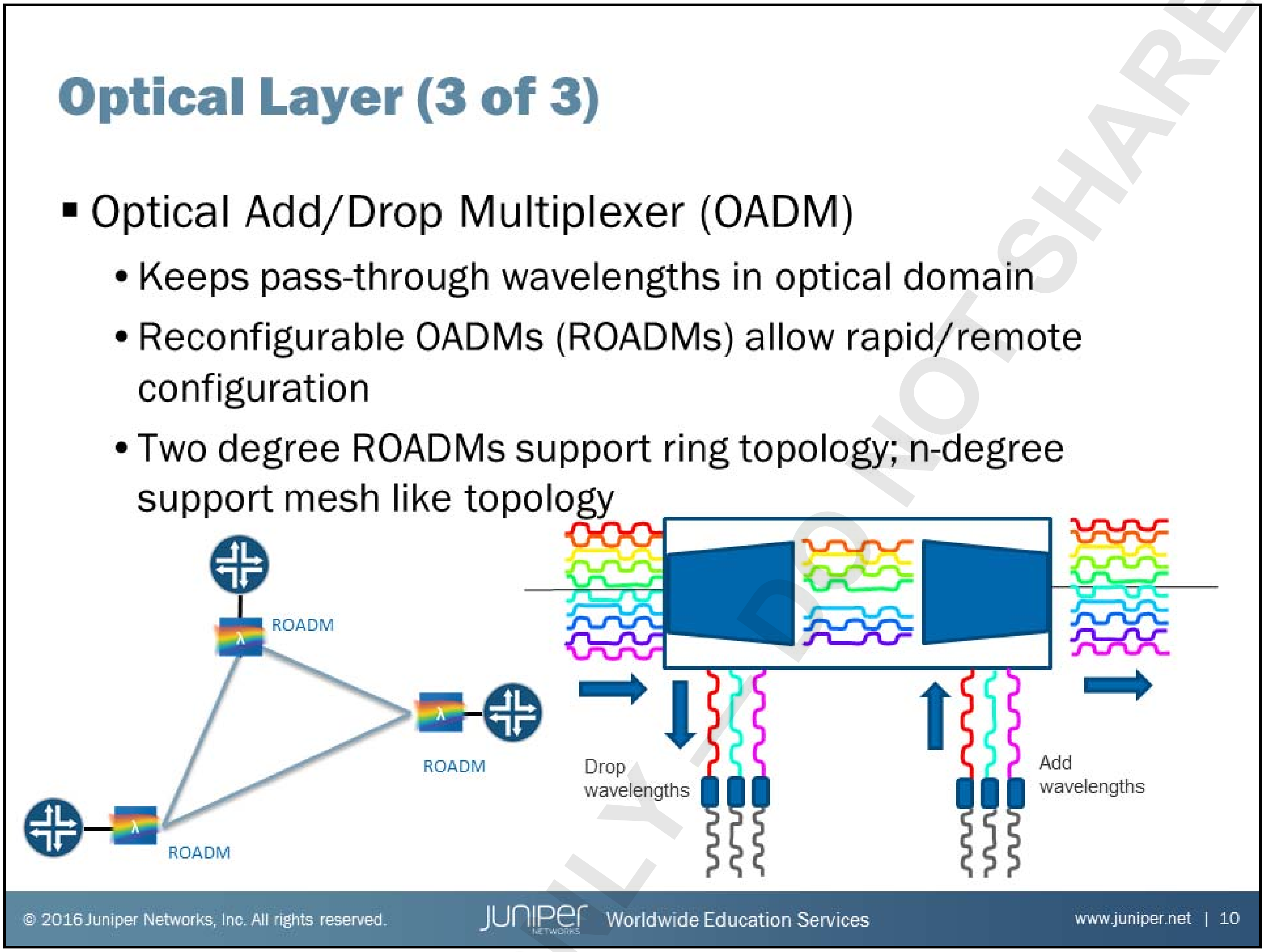
Optical transport system

Router

Wavelength-specific transceivers

Optical Layer: Part 2

In some cases, optical transceivers in the router can be tuned to a specific wavelength, removing the need for transponders on the optical shelf and removing one conversion point. This can save money and strengthen the overall signal going into the wire.



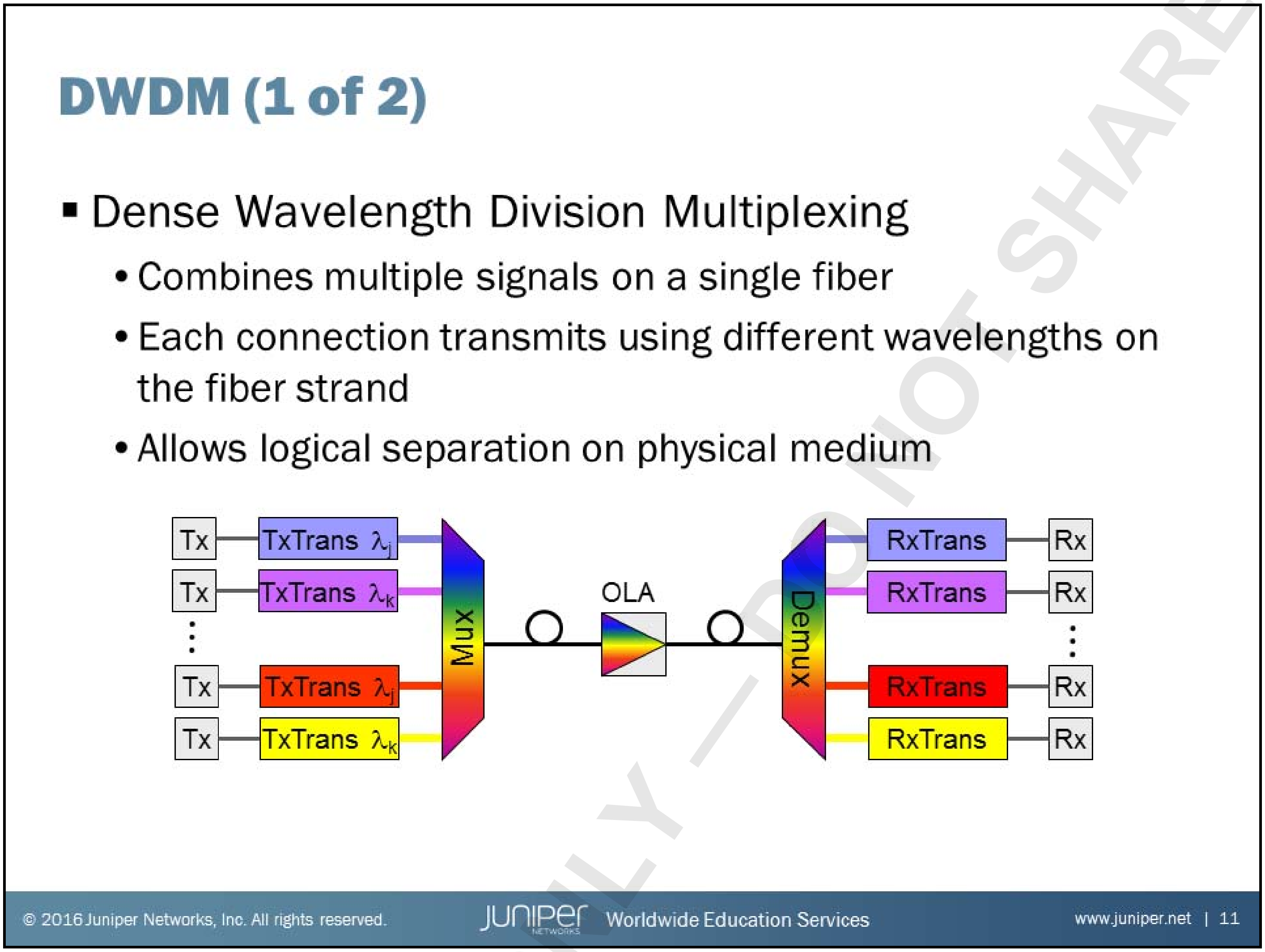
Optical Layer (3 of 3)

- Optical Add/Drop Multiplexer (OADM)
  - Keeps pass-through wavelengths in optical domain
  - Reconfigurable OADMs (ROADMs) allow rapid/remote configuration
  - Two degree ROADMs support ring topology; n-degree support mesh like topology

Juniper Networks Worldwide Education Services www.juniper.net | 10

Optical Layer: Part 3

OADMs simply are how wavelengths are added or removed from a fiber-path. ROADMs allow this configuration to be done using software. With two degree ROADMs the optical network is deployed in a ring topology. With three degree (or higher) ROADMs, more mesh like optical networks can be built out. OADMs are typically bi-directional. The diagram above only shows one direction for simplicity.



DWDM (1 of 2)

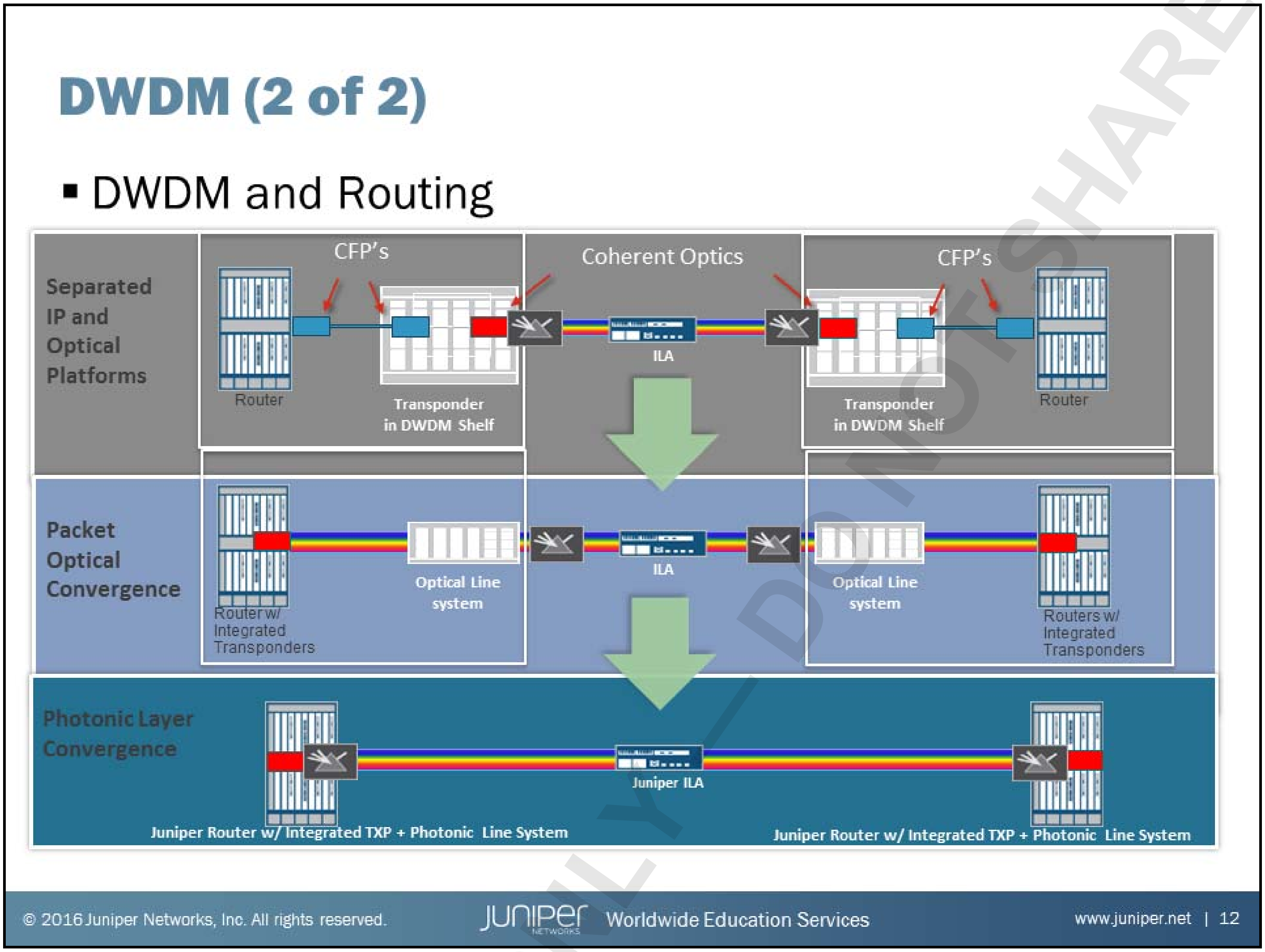
- Dense Wavelength Division Multiplexing
  - Combines multiple signals on a single fiber
  - Each connection transmits using different wavelengths on the fiber strand
  - Allows logical separation on physical medium

Worldwide Education Services

www.juniper.net | 11

Dense Wavelength Division Multiplexing

Light operates on different wavelengths and dense wavelength division multiplexing (DWDM) takes advantage of this by multiplexing multiple signals on a single fiber. Dark fiber is one circuit per fiber. DWDM it can be many circuits per fiber. How many circuits depends on many factors. The more sensitive the optics is, the more signals can be muxed onto a single fiber. DWDM is how service provider, or any organization, take advantage of the properties of light to carry more than circuit across a single fiber.



DWDM (2 of 2)

- DWDM and Routing

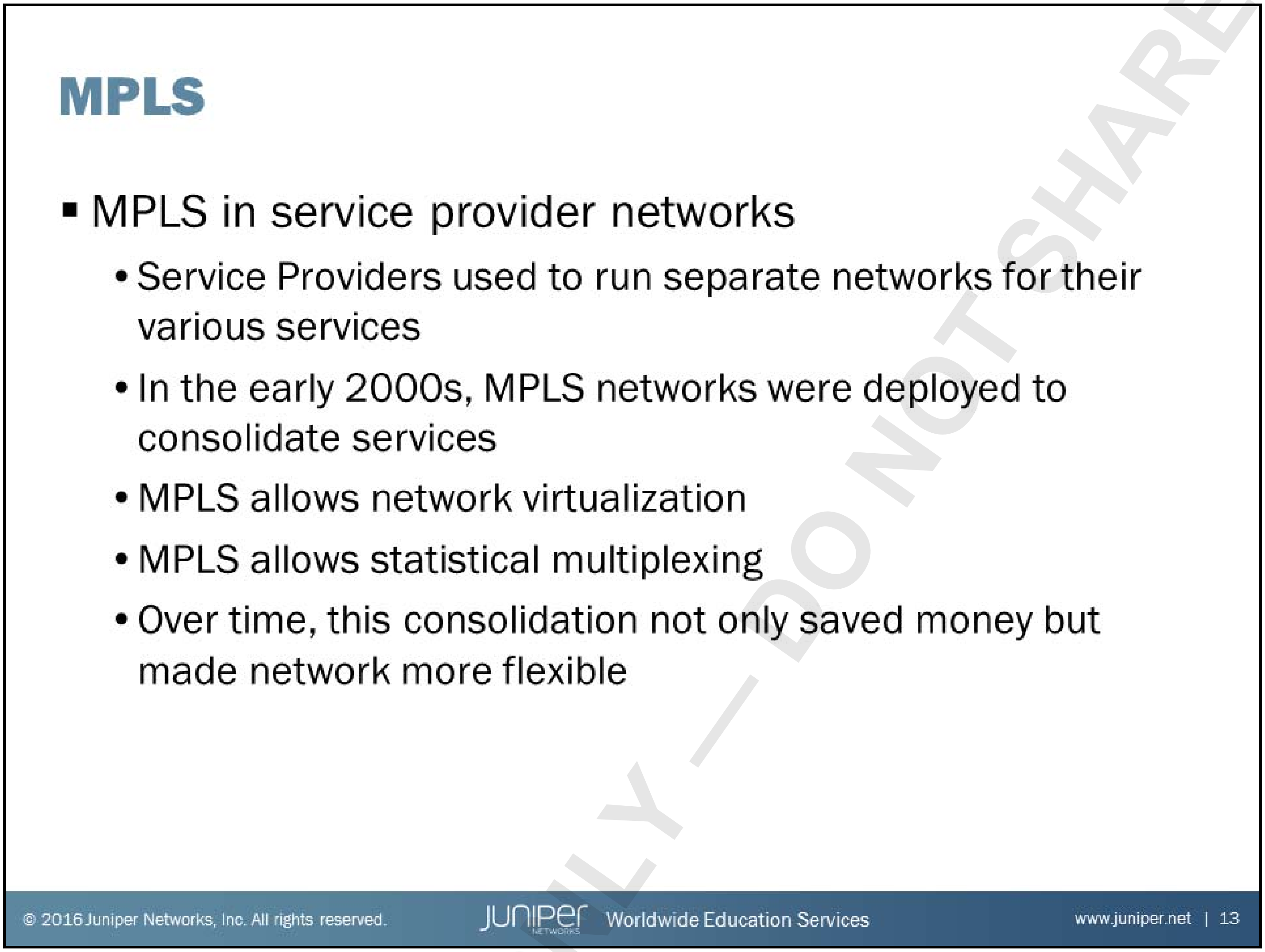
Separated IP and Optical Platforms
CFP's
Coherent Optics
CFP's
Router
Transponder in DWDM Shelf
ILA
Transponder in DWDM Shelf
Router
Packet Optical Convergence
Router w/ Integrated Transponders
Optical Line system
ILA
Optical Line system
Routers w/ Integrated Transponders
Photonic Layer Convergence
Juniper ILA
Juniper Router w/ Integrated TXP + Photonic Line System
Juniper Router w/ Integrated TXP + Photonic Line System

Juniper Networks Worldwide Education Services www.juniper.net | 12

DWDM Routing

The slide shows traditional router interaction with a DWDM system. The traditional box shows a router with a gray (non-tuneable) interface connecting to a transponder and the ultimately a DWDM switch. The transponder will take the wavelength of the gray interface and tune it to the different wavelengths on the DWDM switch allowing two or more circuits using the same wavelength to share a single fiber.

If the router has coherent optics, the wavelength can be assigned on the router itself, meaning there is no need for the intermediate transponder. The router can plug directly into the DWDM switch. Both methods are valid and both can be used in a particular network.



MPLS

- MPLS in service provider networks
  - Service Providers used to run separate networks for their various services
  - In the early 2000s, MPLS networks were deployed to consolidate services
  - MPLS allows network virtualization
  - MPLS allows statistical multiplexing
  - Over time, this consolidation not only saved money but made network more flexible

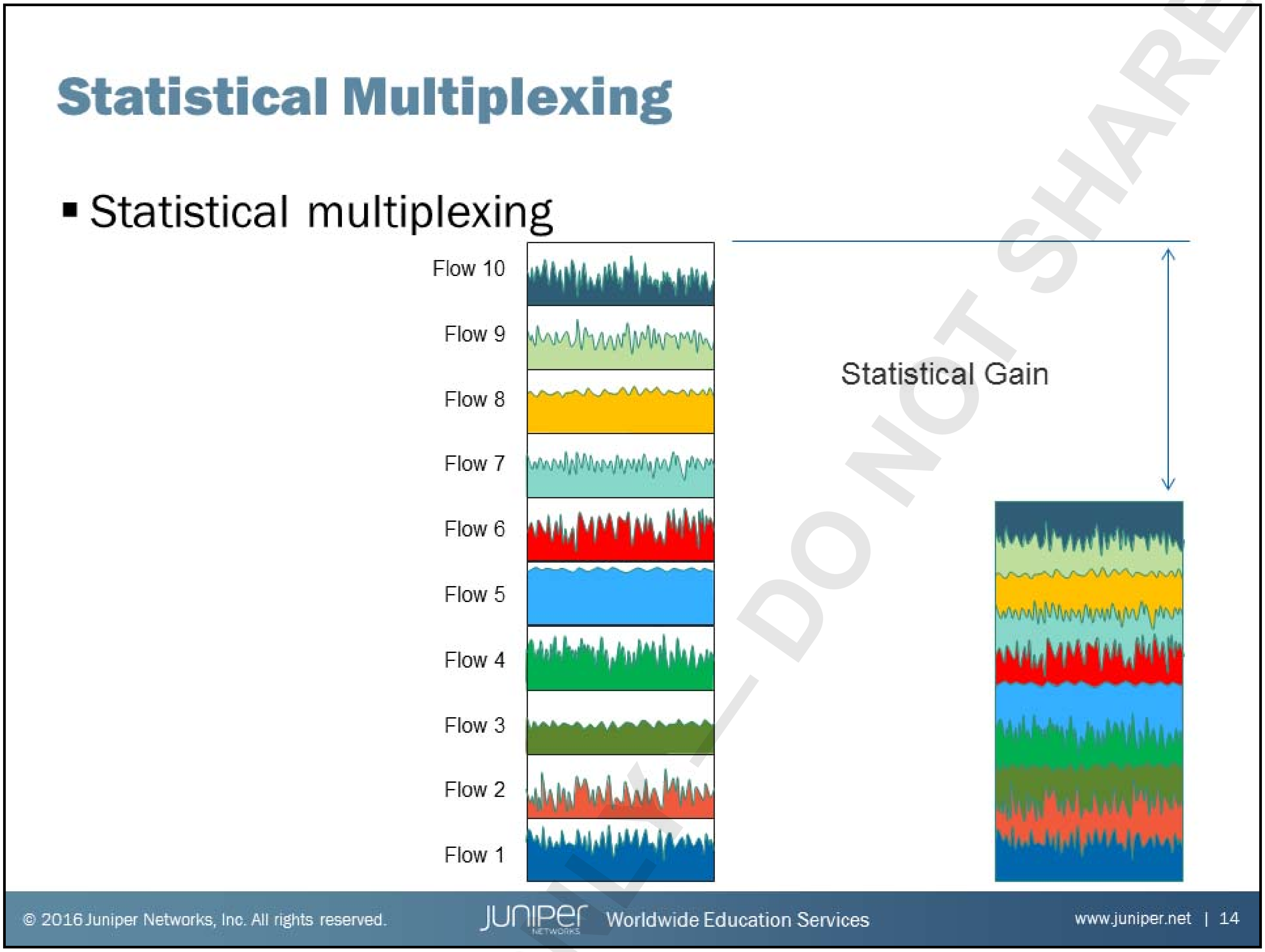
Worldwide Education Services

www.juniper.net | 13

MPLS in SP Networks

Service providers have mostly moved to IP/MPLS networks throughout the last 15 years. This technology was introduced as a way to consolidate disparate networks into a single unified network and gain the advantages of statistical multiplexing. Statistical multiplexing allows many flows of bursty traffic to share the same resources. Fewer resources need to be allocated as the bursts level out over time. The other advantage to MPLS is that it essentially virtualizes the network. What VLANs did for switching network designs in the 1990s, MPLS did for the whole network. With MPLS it has become easy to deploy new networks over shared infrastructure. Each network is identified by an MPLS label, keeping it logically separated from the other networks. All of these networks though, still share a common infrastructure.

This virtualization will carry forward into the SDN era. MPLS truly has become common practice and the technology is flexible enough to continue the virtualization trend down to the network device level.



Statistical Multiplexing

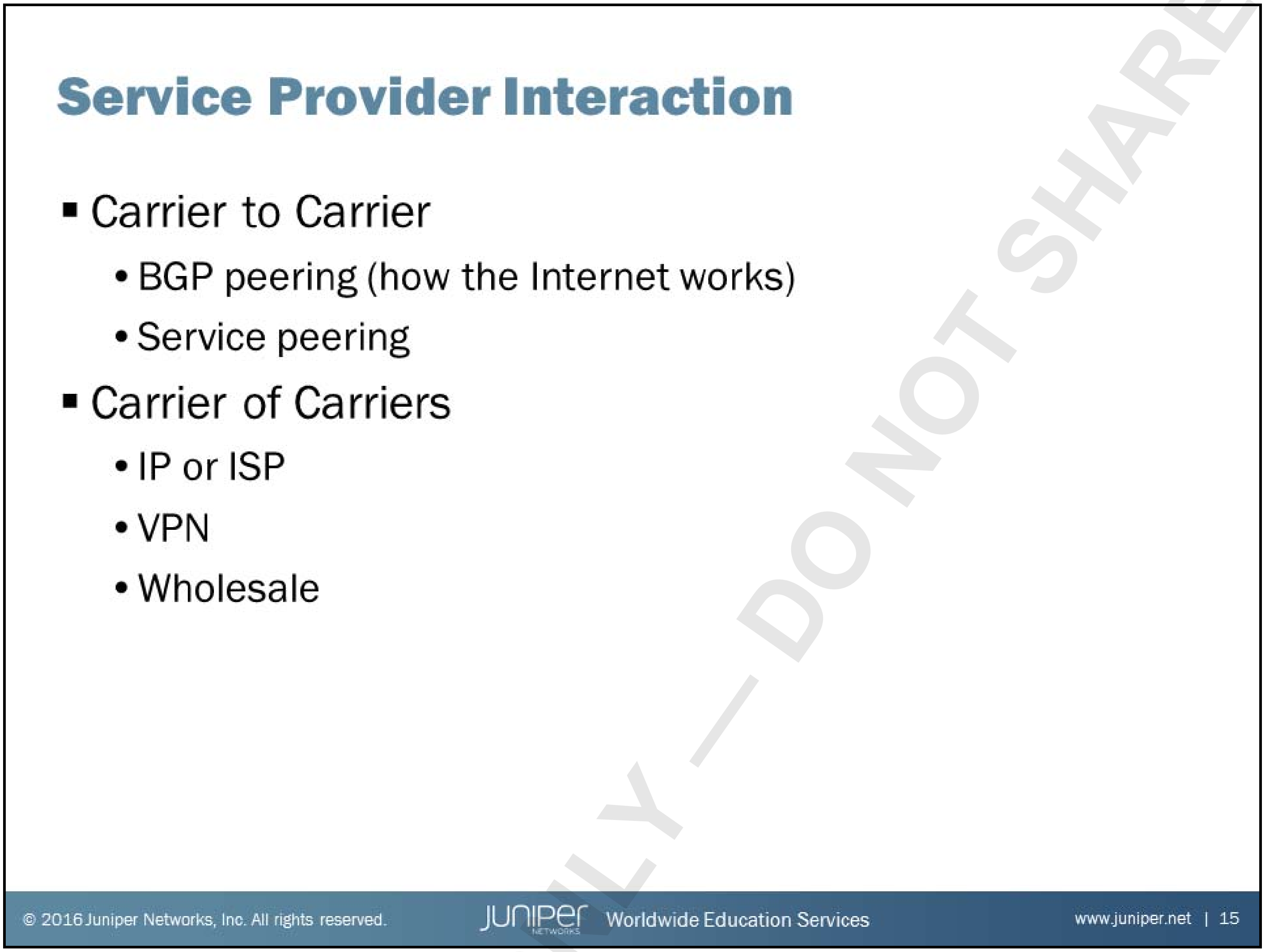
- Statistical multiplexing

Juniper Networks Worldwide Education Services

www.juniper.net | 14

Statistical Multiplexing

The slide visualizes the concept of statistical multiplexing. 10 distinct flows are shown as separate circuits or networks. The whitespace represents wasted cost because it is not being used. When these flows are combined, we are able to fully utilize the circuits. The is the advantage of virtualization with MPLS, combining disparate systems to maximize resource utilization.



Service Provider Interaction

- Carrier to Carrier
  - BGP peering (how the Internet works)
  - Service peering
- Carrier of Carriers
  - IP or ISP
  - VPN
  - Wholesale

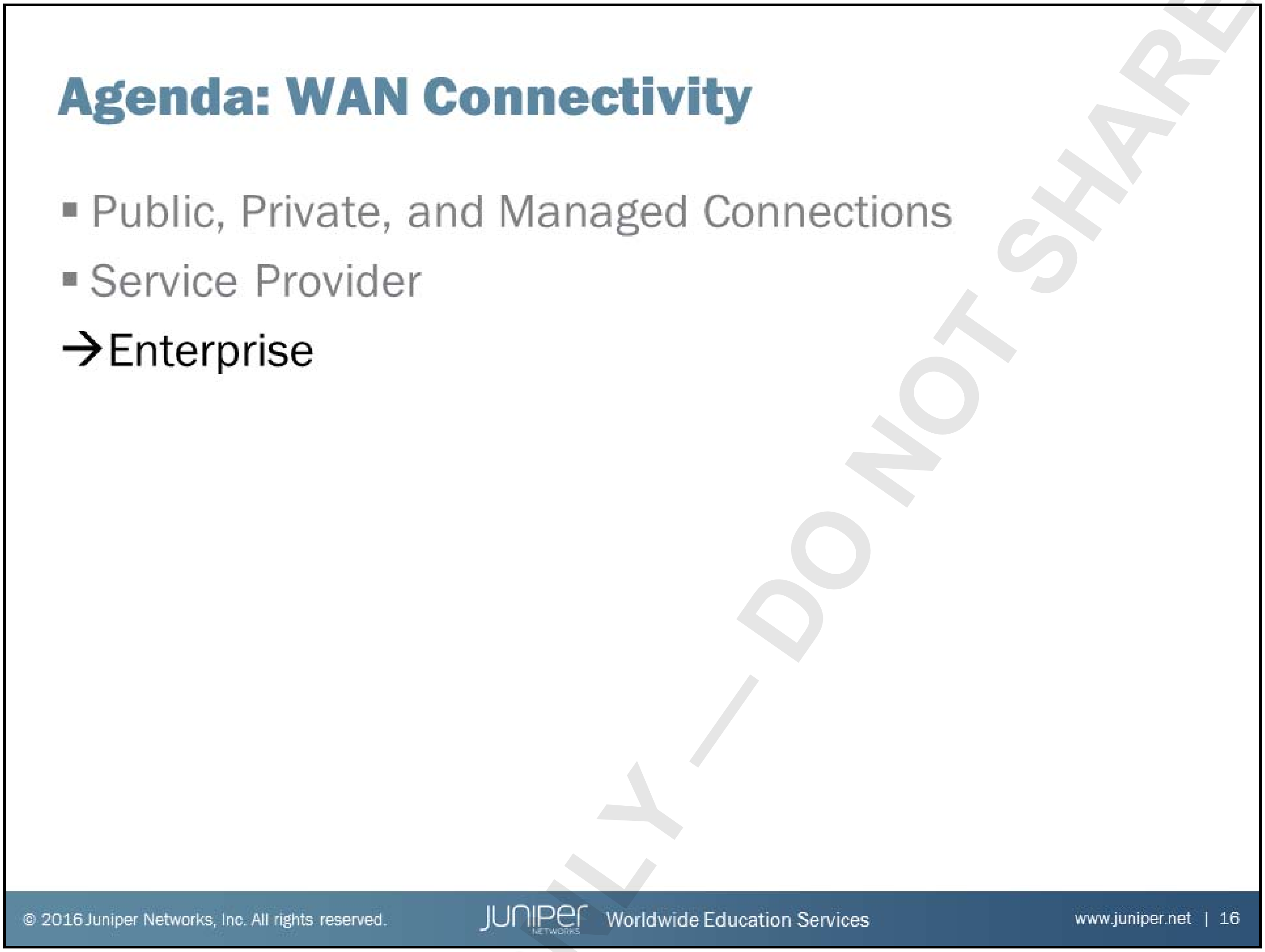
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 15

Carrier to Carrier

Carrier to Carrier is a traditional peering arrangement where two service providers agree to meet at common peering point. The most common arrangement is BGP peering. Here two providers simply exchange routes with each other. There can be financial arrangements and incentives based on how much traffic each provider is expected to exchange. From a technical standpoint, this is the simplest method of all the peering arrangements as it is basically an IPv4 and IPv6 BGP route exchange. Service peering is a little more complex, where two providers meet to provide some sort of service to their customers. In this case, a company might purchase VPN services from both providers and need them to connect.

Carrier of Carriers

Carrier of Carriers is where one service provider (Carrier A) acts as the backbone network for another service provider (Carrier B). In this case, Carrier A sets up a VPN for Carrier B. Carrier B is the service provider of record for the customer. This works well for pure IP traffic (IP/ISP) or for VPN traffic. In the IP/ISP case, Carrier B does not necessarily need to run MPLS, but Carrier A certainly does. When using VPNs, where Carrier B is selling VPN services to their customers, both Carrier B and Carrier A use MPLS. Another case is the wholesale model, where one carrier wholesales services to another provider and that provider re-sells those services. In those situations the reseller does not need to have their own infrastructure, but most often they will.



Agenda: WAN Connectivity

- Public, Private, and Managed Connections
- Service Provider

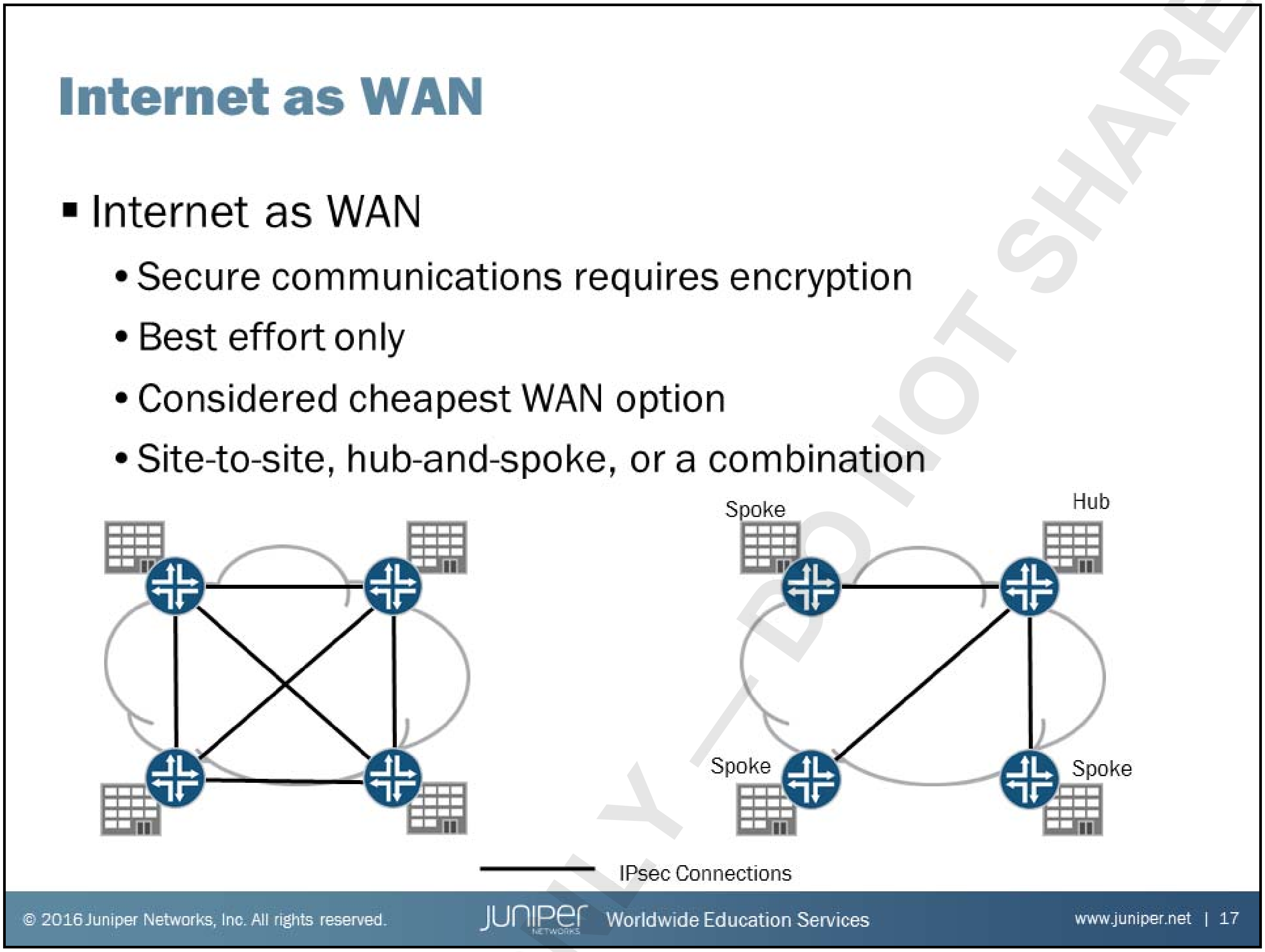
→Enterprise

Worldwide Education Services

www.juniper.net | 16

Enterprise

The slide highlights the topic we discuss next.



Internet as WAN

- Internet as WAN
  - Secure communications requires encryption
  - Best effort only
  - Considered cheapest WAN option
  - Site-to-site, hub-and-spoke, or a combination

Juniper Worldwide Education Services

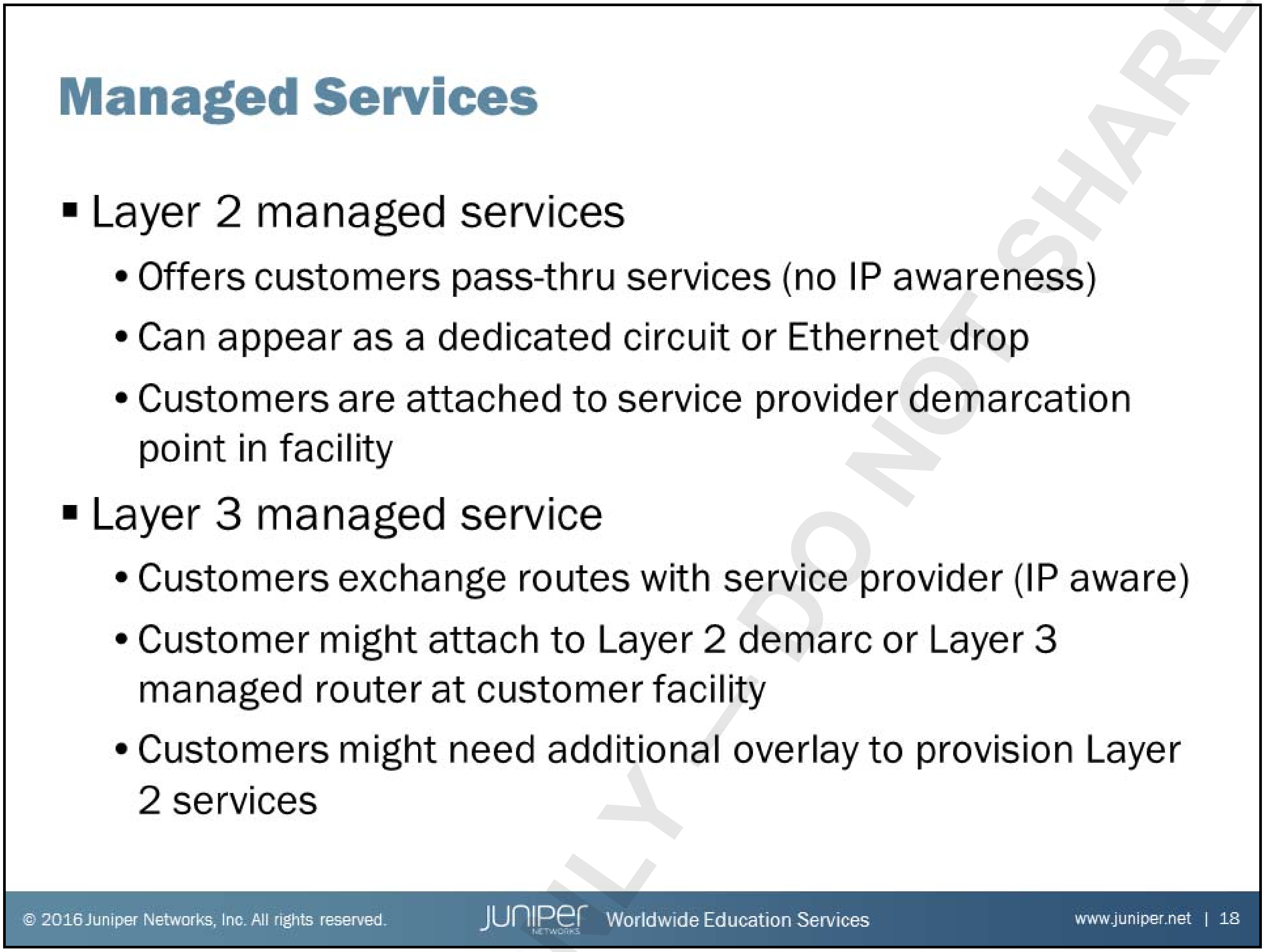
www.juniper.net | 17

Internet as WAN

Using the Internet as a WAN is considered one of the cheapest options around. However, there are two major issues with it. First, it requires encryption to be setup between sites. While this technically is not “required”, how many organizations would actually be OK with all their traffic passing through a public network unencrypted?

Secondly, there is no quality of service guarantees. The Internet is best effort. Even if your local ISP has amazing service, you will eventually pass through a network where you can see performance declines. Even if all your sites are connected the same ISP, they are more than likely prioritizing their customers which have paid for their managed services rather than ISP traffic with no service guarantees. However, this might not be a problem for many organizations and the cost savings more than justify the potential performance hit. And in some cases, there might not be a performance hit.

When setting up the encryption overlay, any topology can be created. Instead of ordering circuits you simply establish encrypted tunnels. Typically, hub and spoke or full mesh topologies are deployed but any topology that can be turned up is possible. The encrypted network is essentially another virtual networking technique where the Internet is the underlay and the encryption tunnels are the overlay.



Managed Services

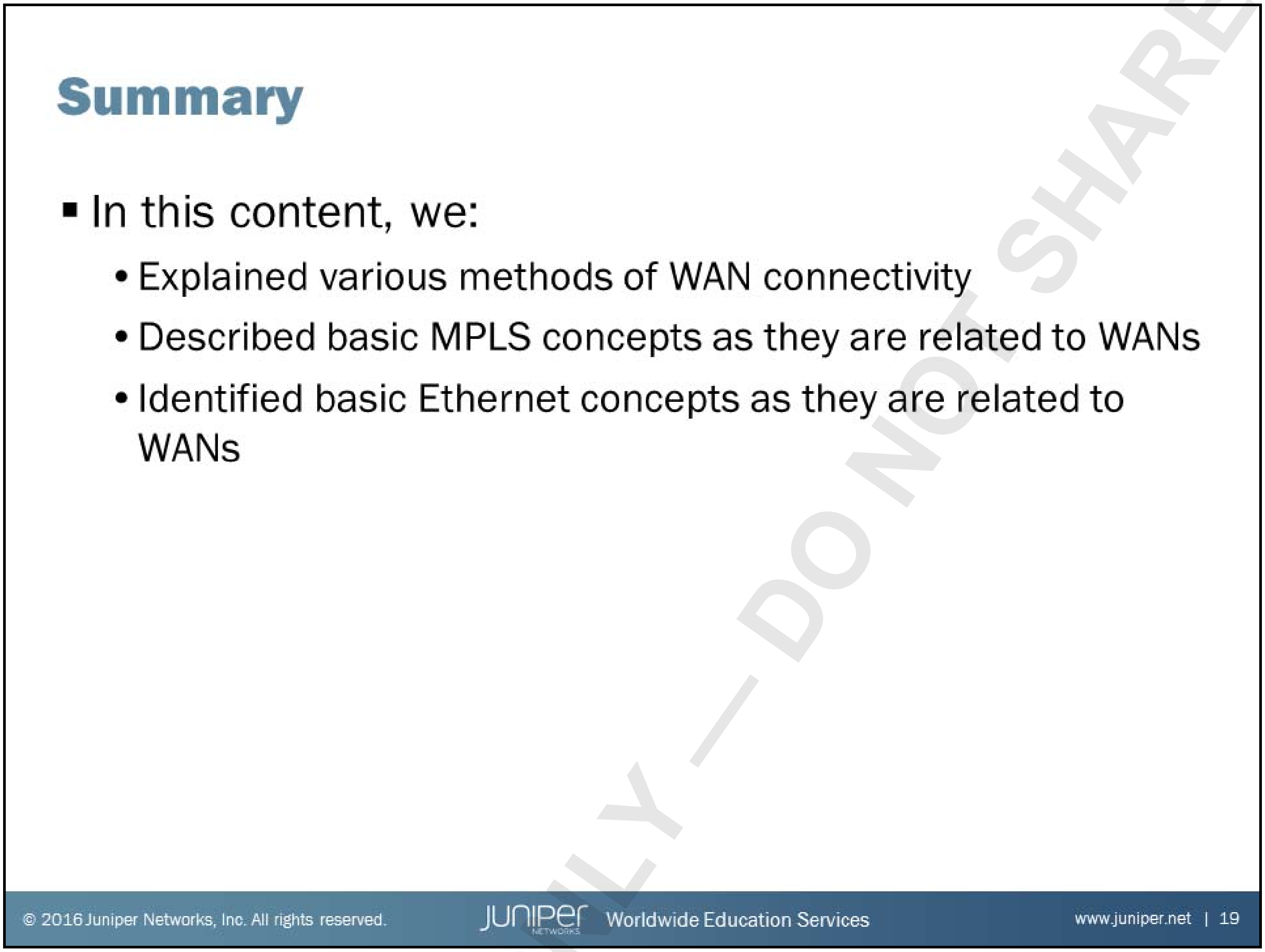
- Layer 2 managed services
  - Offers customers pass-thru services (no IP awareness)
  - Can appear as a dedicated circuit or Ethernet drop
  - Customers are attached to service provider demarcation point in facility
- Layer 3 managed service
  - Customers exchange routes with service provider (IP aware)
  - Customer might attach to Layer 2 demarc or Layer 3 managed router at customer facility
  - Customers might need additional overlay to provision Layer 2 services

Layer 2 Managed Services

With Layer 2 managed services, the customer receives a more traditional circuit type of service where the customer provides routing capabilities on top of the circuit. Increasingly, these services are Ethernet based but in some areas these circuits can be cable, DSL, or other transport protocol. The customer simply attaches their router to the provider's demarcation point and configures routing, assuming everything was provisioned correctly.

Layer 3 Managed Service

With an Layer 3 managed service, the provider is aware of IP routing. This is usually L3VPN services and the provider either has a static route to the customer site or actually performs a dynamic route exchange with the customer's router. If the provider is providing a fully managed service, it might also include service provider controlled router on the customer's premises. For IP routing, this is an easy way to quickly build a WAN, but if the customer requires Layer 2 provisioning this requires an additional layer of virtualization on the customer's part.



Summary

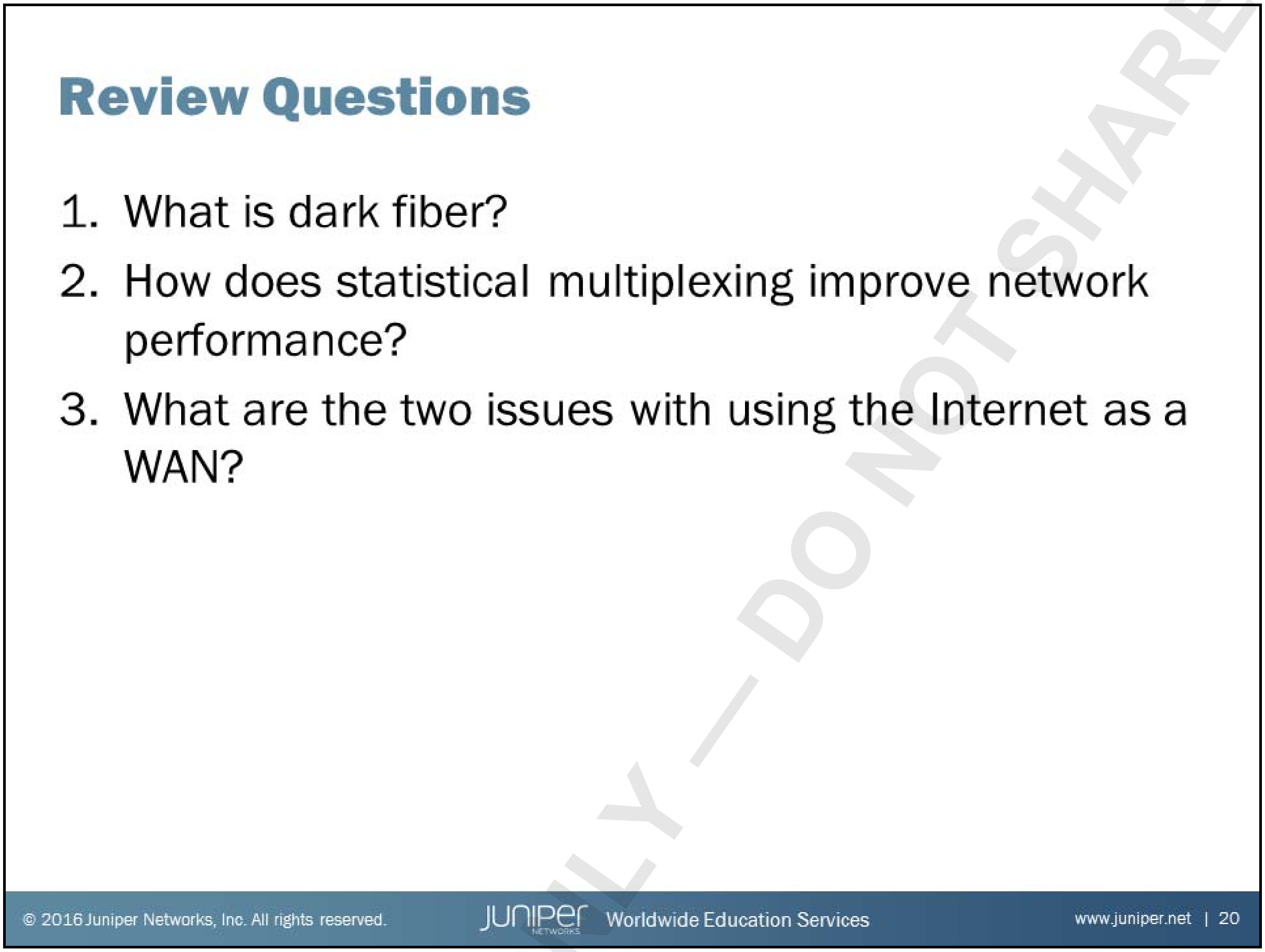
- In this content, we:
  - Explained various methods of WAN connectivity
  - Described basic MPLS concepts as they are related to WANs
  - Identified basic Ethernet concepts as they are related to WANs

Worldwide Education Services

www.juniper.net | 19

We Discussed:

- Several methods of WAN connectivity;
- How MPLS works and its relation to WANs; and
- Basic Ethernet concepts and how they apply to WANs.



Review Questions

1. What is dark fiber?
2. How does statistical multiplexing improve network performance?
3. What are the two issues with using the Internet as a WAN?

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 20

Review Questions

1.

2.

3.

Answers to Review Questions

1.

Dark fiber is unused fiber optic strands withing a fiber optic cable that can be used as the network needs grow.

2.

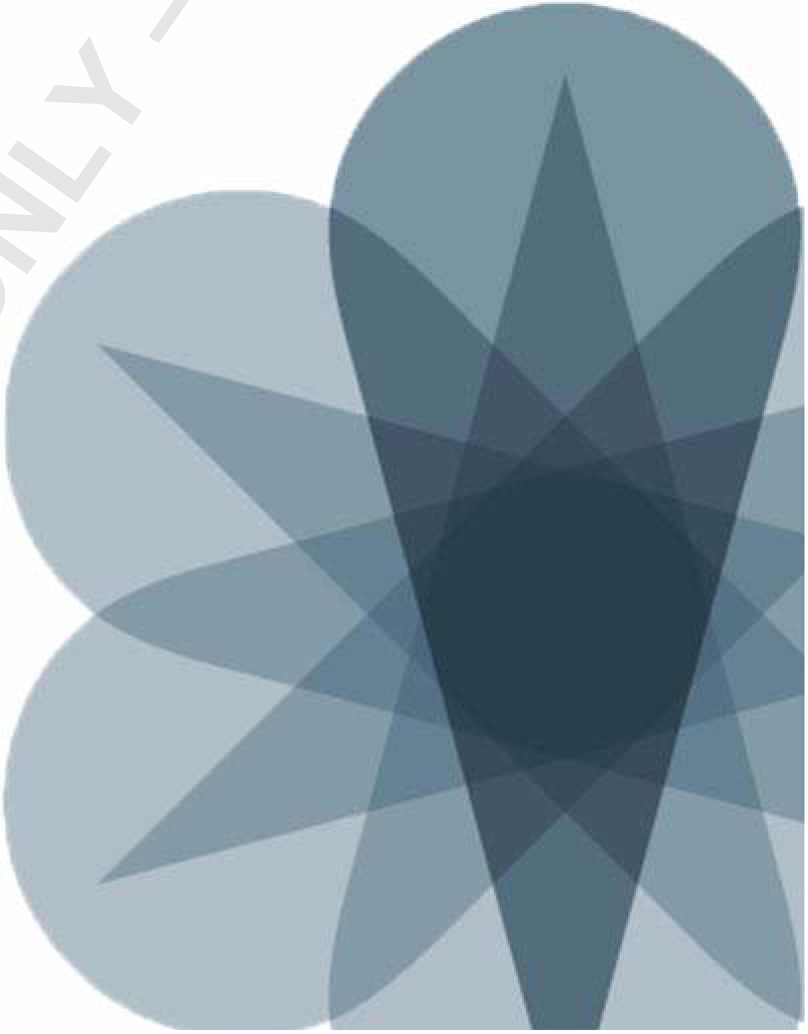
Statistical multiplexing improves network performance by because it allows for full utilization of network bandwidth.

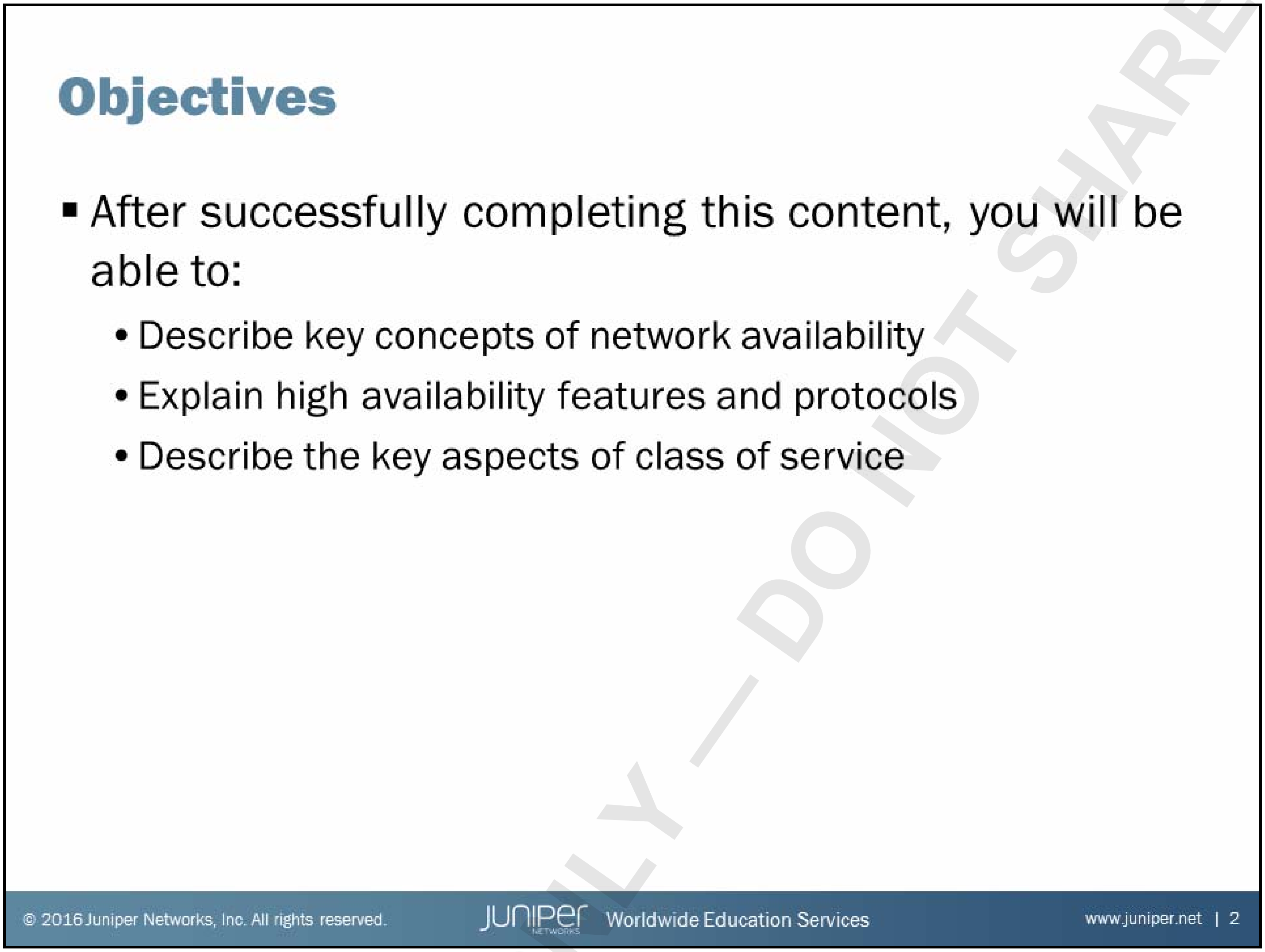
3.

The two issues with using the Internet as a WAN is that the traffic needs to be encrypted and there is no bandwidth guarantee.

Juniper Networks Design—WAN

Chapter 4: Network Availability and Traffic Prioritization





Objectives

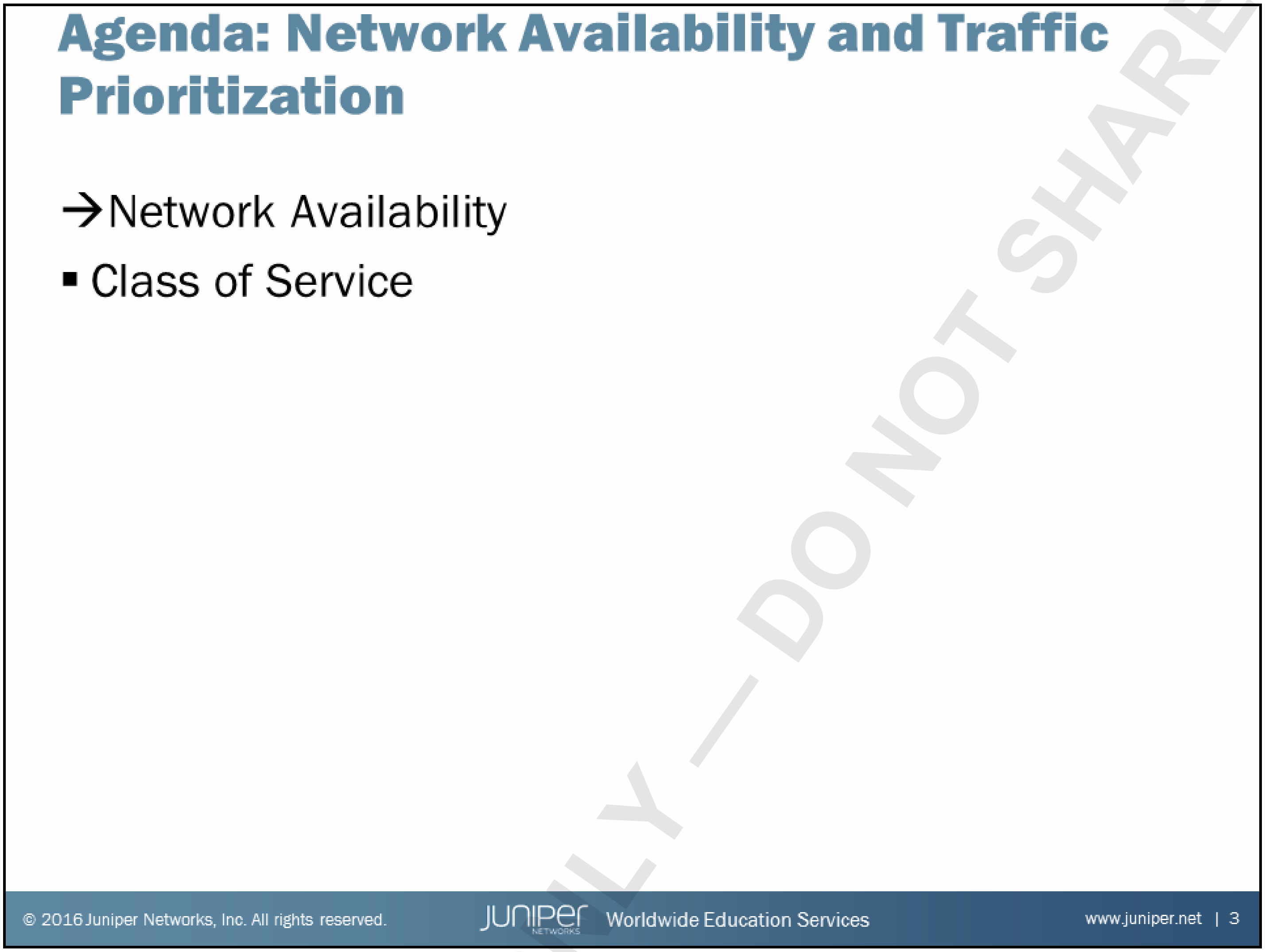
- After successfully completing this content, you will be able to:
  - Describe key concepts of network availability
  - Explain high availability features and protocols
  - Describe the key aspects of class of service

Worldwide Education Services

www.juniper.net | 2

We Will Discuss:

- Key concepts of network availability;
- High availability features and protocols; and
- Key aspects of class of service.



Agenda: Network Availability and Traffic Prioritization

→Network Availability

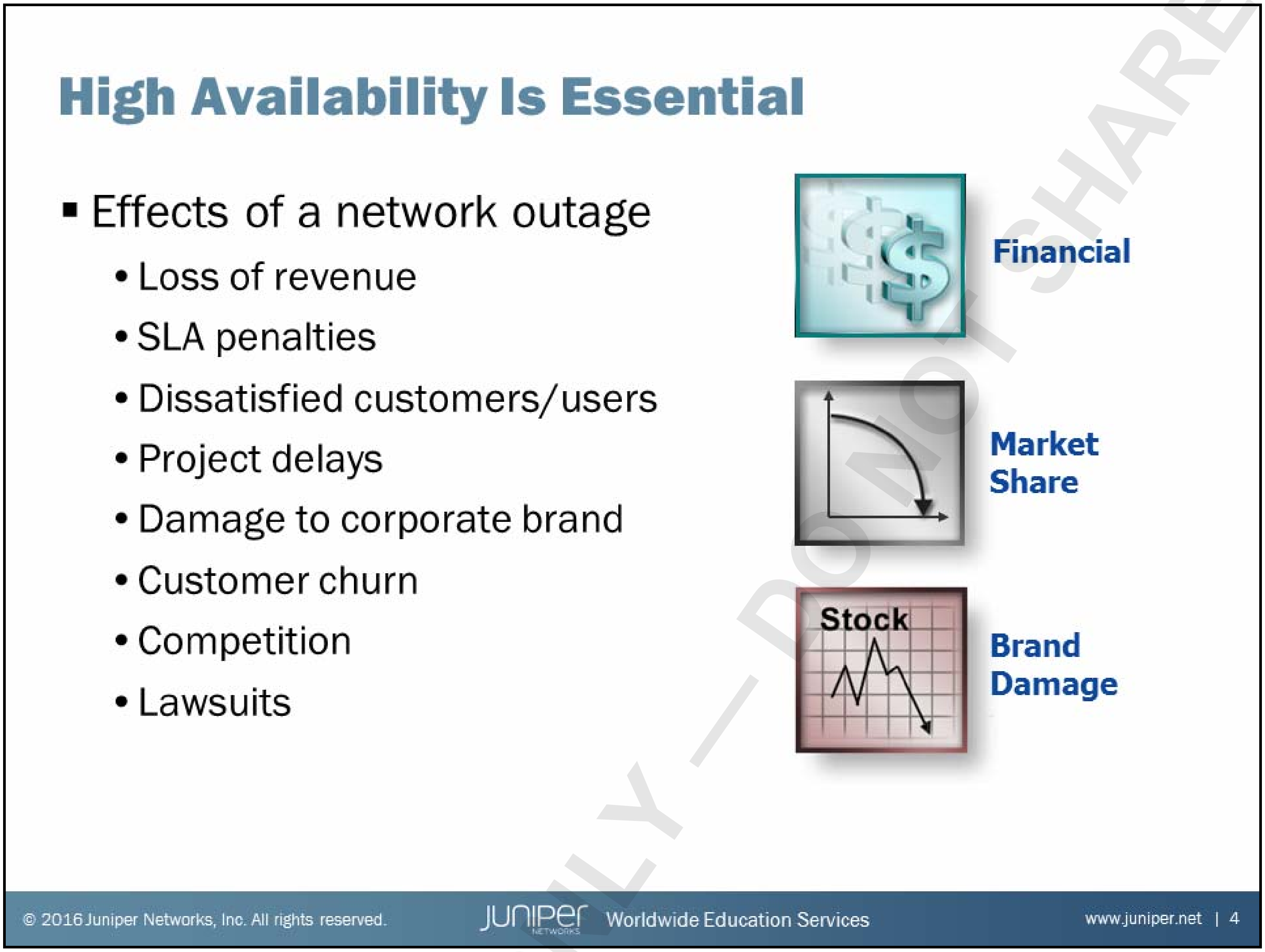
- Class of Service

Juniper Networks Worldwide Education Services

www.juniper.net | 3

Network Availability

The slide lists the topics we will discuss. We discuss the highlighted topic first.



High Availability Is Essential

- Effects of a network outage
  - Loss of revenue
  - SLA penalties
  - Dissatisfied customers/users
  - Project delays
  - Damage to corporate brand
  - Customer churn
  - Competition
  - Lawsuits

Financial

Market Share

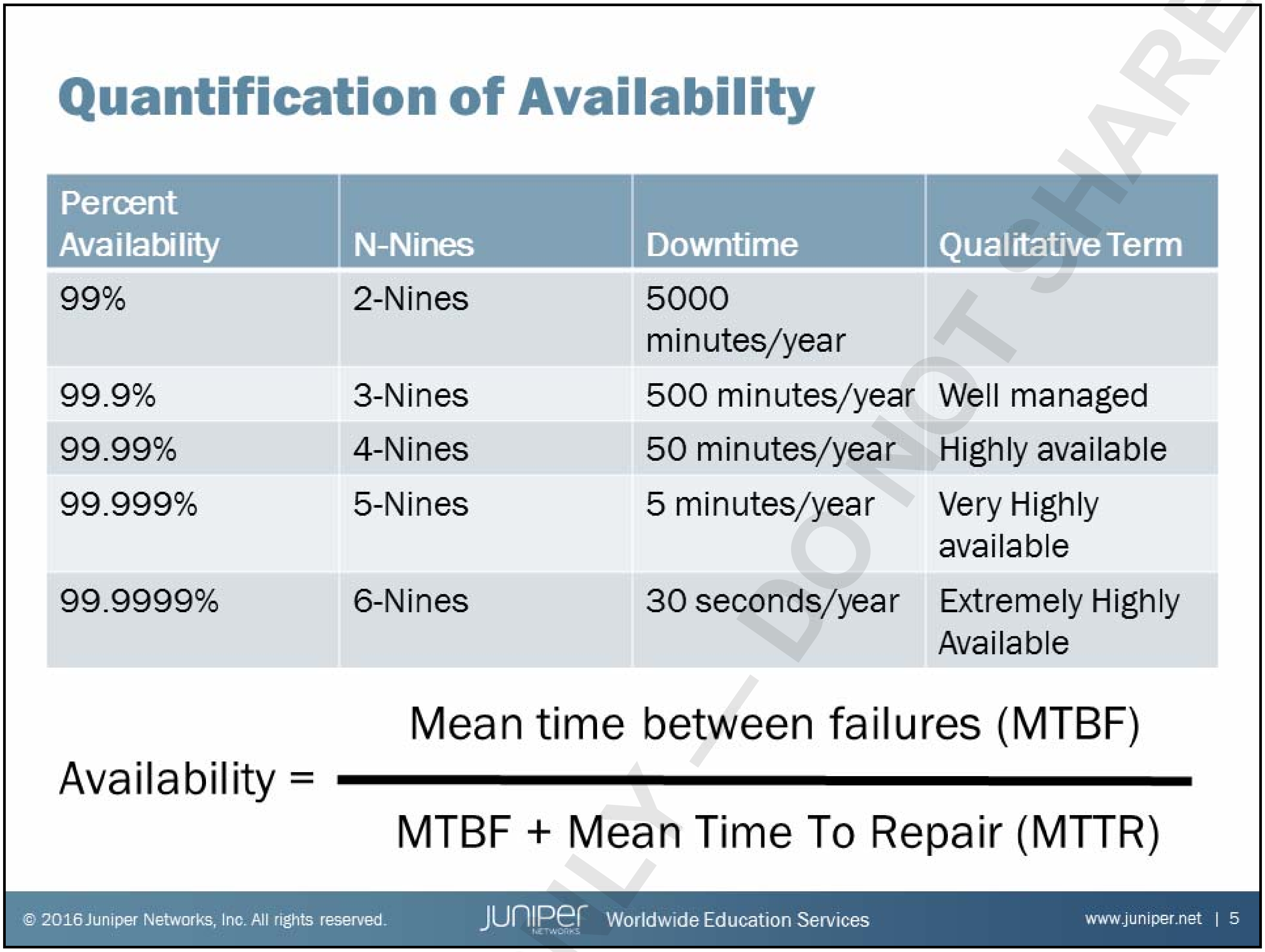
Brand Damage

Juniper Networks Worldwide Education Services

www.juniper.net | 4

Effects of Network Outage

Clearly, there are a number of potential negative impacts resulting from network or device outages. Some of the long term ramifications, such as bad press in a publication like Light Reading, or Network World are difficult to measure, but they are obviously undesirable and can have a lasting impact.



Quantification of Availability

Percent Availability	N-Nines	Downtime	Qualitative Term
99%	2-Nines	5000 minutes/year	
99.9%	3-Nines	500 minutes/year	Well managed
99.99%	4-Nines	50 minutes/year	Highly available
99.999%	5-Nines	5 minutes/year	Very Highly available
99.9999%	6-Nines	30 seconds/year	Extremely Highly Available

$$\text{Availability} = \frac{\text{Mean time between failures (MTBF)}}{\text{MTBF + Mean Time To Repair (MTTR)}}$$

Worldwide Education Services

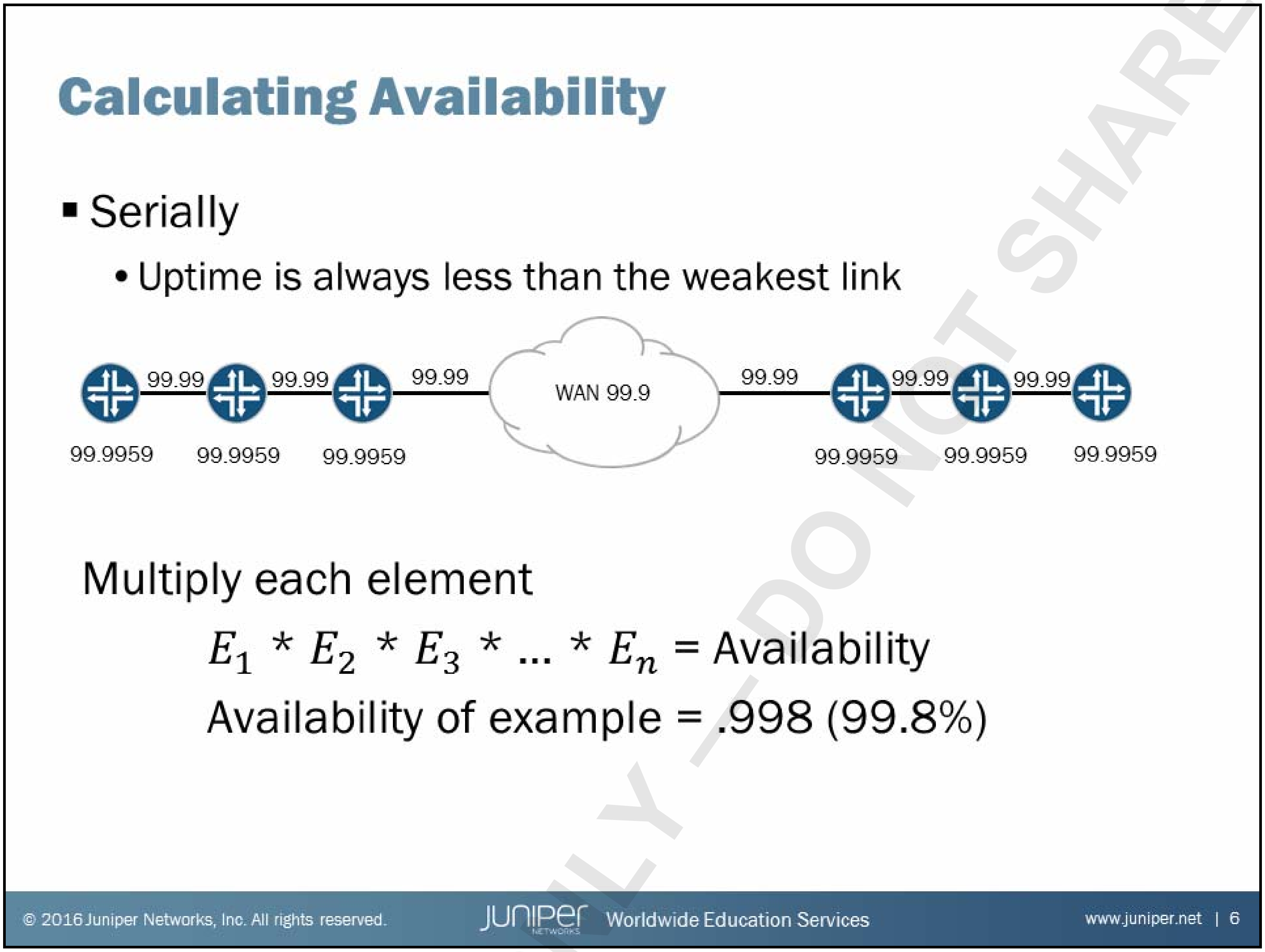
www.juniper.net | 5

Quantification of Availability

To deploy dependable networks and devices, it is important to define a mechanism for quantifying dependability. The “9’s” terminology is the most familiar to the industry and is widely used to measure specifically the availability of network devices. The 9’s imply the amount of inherent downtime. Downtime is typically specified in Telcordia requirements such as GR-1110-CORE, Broadband Switching System Generic Requirements.

The 9’s provide an operational target to which networks and devices can be managed. To determine the specific availability of a system, mean time between failures (MTBF) and mean time to repair (MTTR), the time it takes to repair a component, are useful figures. MTBF and MTTR figures provide very specific availability measurements for not only systems, but also the individual components that make up those systems.

It should be noted that MTBF figures are often conservative and the actual MTBF is larger than a calculated (or estimated) one.



Calculating Availability

- Serially
  - Uptime is always less than the weakest link

Multiply each element

$$E_1 * E_2 * E_3 * \ldots * E_n = \text{Availability}$$

Availability of example = .998 (99.8%)

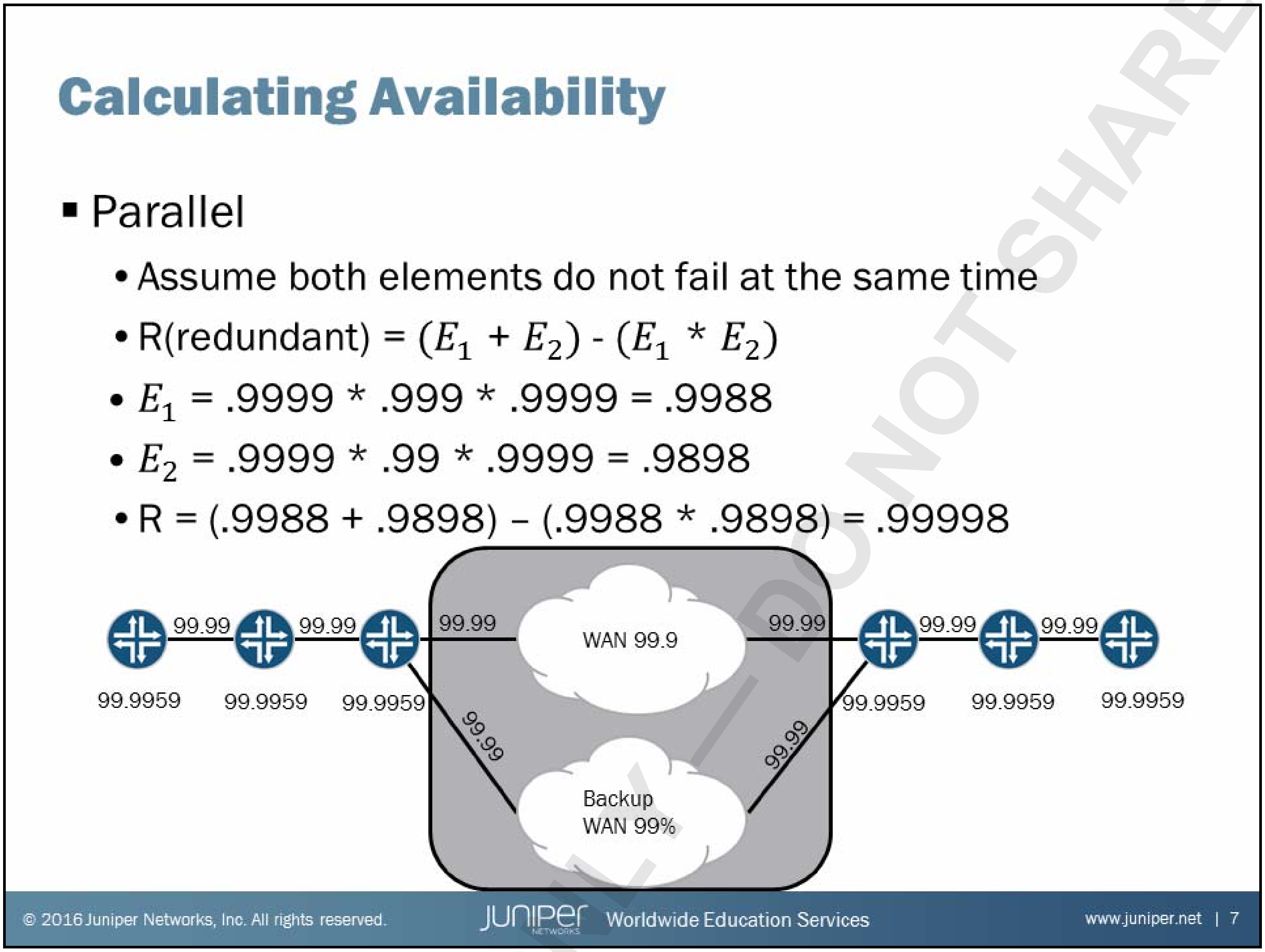
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 6

Serial Calculation

Calculating availability with a serial circuit is straightforward. Take the number from each element and multiply them together. In the example on the slide:

- .999959 * .9999 * .999959 * .9999 * .999959 * .9999 * .999 WAN Availability * .9999 * .999959 * .9999 * .999959 * .9999 * .999959 = .998 (99.8%)

Keep in mind that the availability will always be less than the weakest link. Often it is calculated as the weakest link which is not correct.



Calculating Availability

- Parallel
  - Assume both elements do not fail at the same time
  - R(redundant) = $(E_1 + E_2)$ - $(E_1 * E_2)$
  - E_1 = .9999 * .999 * .9999 = .9988
  - E_2 = .9999 * .99 * .9999 = .9898
  - R = (.9988 + .9898) – (.9988 * .9898) = .99998

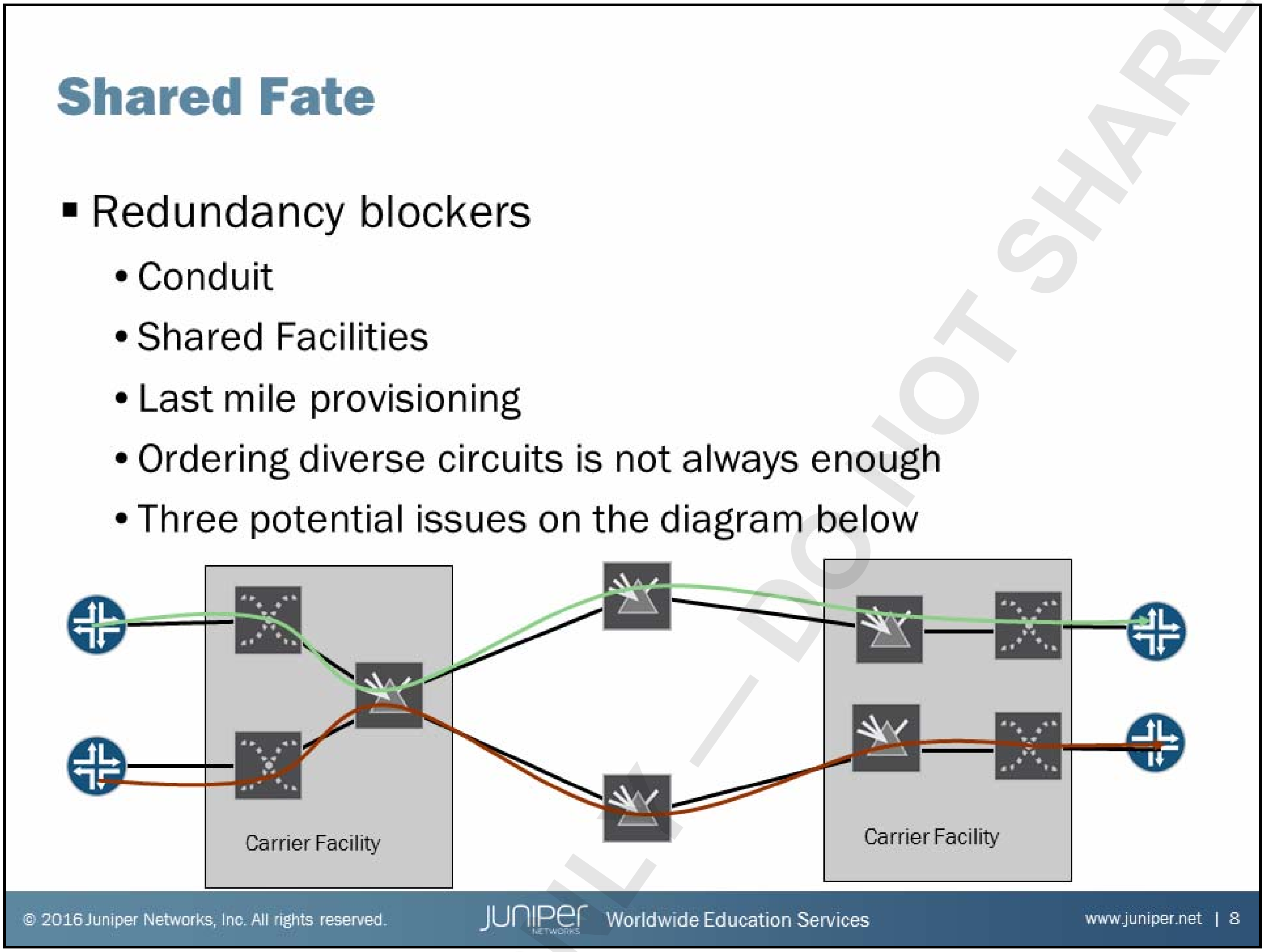
Worldwide Education Services

www.juniper.net | 7

Parallel Calculation

When calculating parallel paths, we need to find the reliability of the two parallel elements. One assumption is the two elements will not fail at the same time. In the example on the slide, a second backup WAN provider is brought in with a lower SLA (99% versus 99.9%). The calculation for each element is made serially as there are the elements involved in each path (circuit + WAN + circuit). A Redundant availability calculation is made of 99.998%. Redundancy really adds quite a bit of uptime to the equation. Using this new value, we can calculate the overall redundancy:

- .999959 * .9999 * .999959 * .9999 * .999959 * .99998 WAN Availability * .999959 * .9999 * .999959 * .9999 * .999959 = .9993 (99.93%)



Shared Fate

- Redundancy blockers
  - Conduit
  - Shared Facilities
  - Last mile provisioning
  - Ordering diverse circuits is not always enough
  - Three potential issues on the diagram below

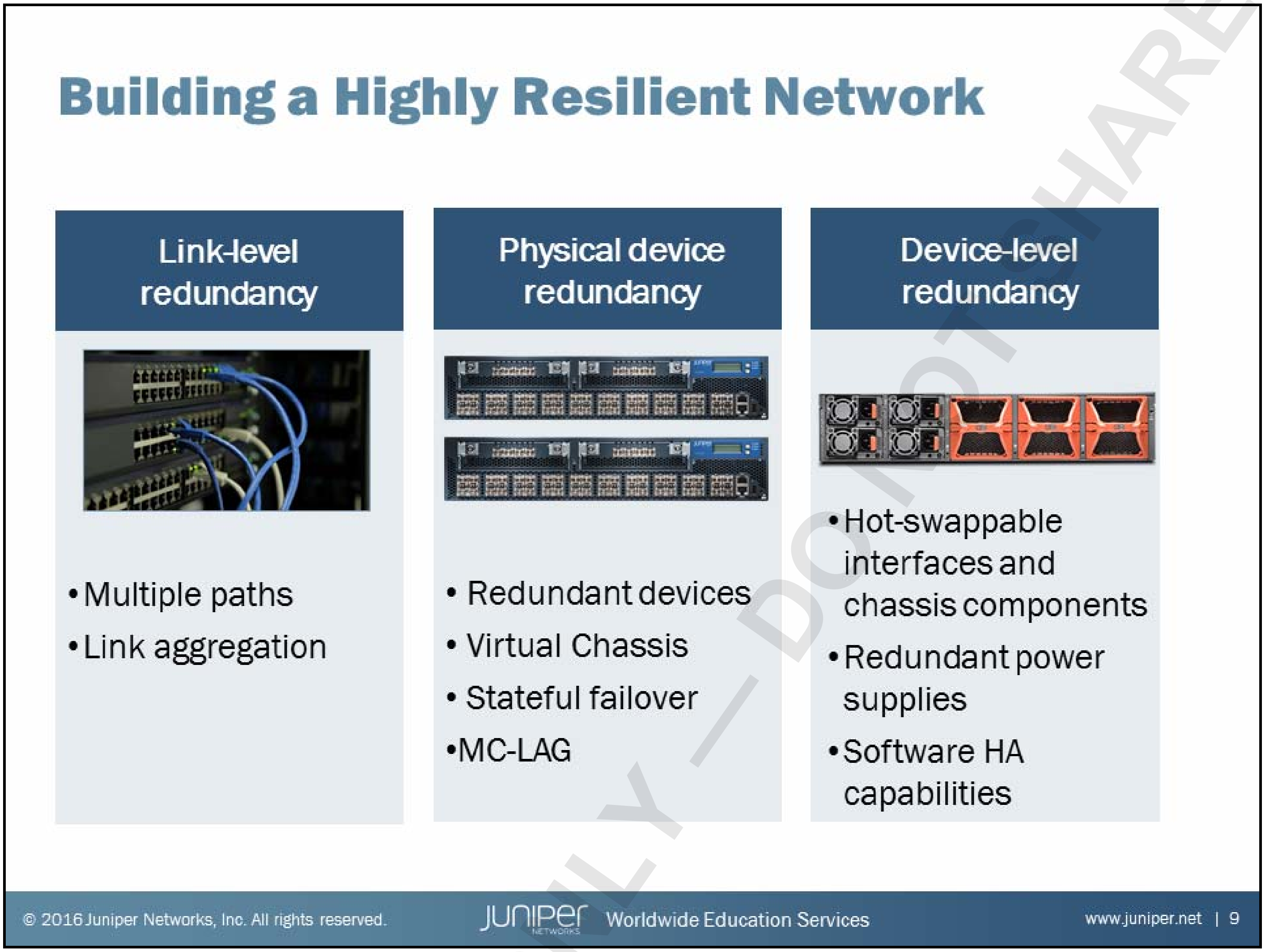
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 8

Shared Fate

Redundancy is great and necessary for maintaining very high uptimes, but sometimes it is not always possible. The concept of shared fate comes in when a network path shares the same fate as another path. This can be due to any number of reasons. Sometimes there is only one conduit path for a circuit to take. Two networking devices may share the same rack, row, or facility. Human error can creep in as well. Although not common, circuits that are ordered diverse can land on the same optical switch. Sometimes human error is not to blame but the blame falls on capacity constraints. Optical networks also have their own redundancy facilities and during an outage two diverse circuits may be re-routed over the same backup path.

The most common is the shared facility. Even if you use circuits from different carriers, it is possible they land at the same local facility. Often network designers are unaware of shared fates until something bad happens. But sometimes we do know about them. In those cases we might not be able to prevent them but we can plan accordingly. Depending on network design it can even be possible to tag network paths that are on a shared path while the control plane or SDN controller takes these into account when planning the best path. Even if there is nothing that can be done (in some areas true diversity is impossible), knowing about redundancy issues can allow your organization to plan accordingly.

In the graphic on the slide, both carrier facilities represent a potential outage point. If either facility went offline, both circuits would drop. Both circuits also traverse a single optical switch in carrier facility on the left. Having a single optical switch go offline is more likely than losing an entire facility but all three represent areas where path diversity is broken.



Building a Highly Resilient Network

Link-level redundancy

- Multiple paths
- Link aggregation

Physical device redundancy

- Redundant devices
- Virtual Chassis
- Stateful failover
- MC-LAG

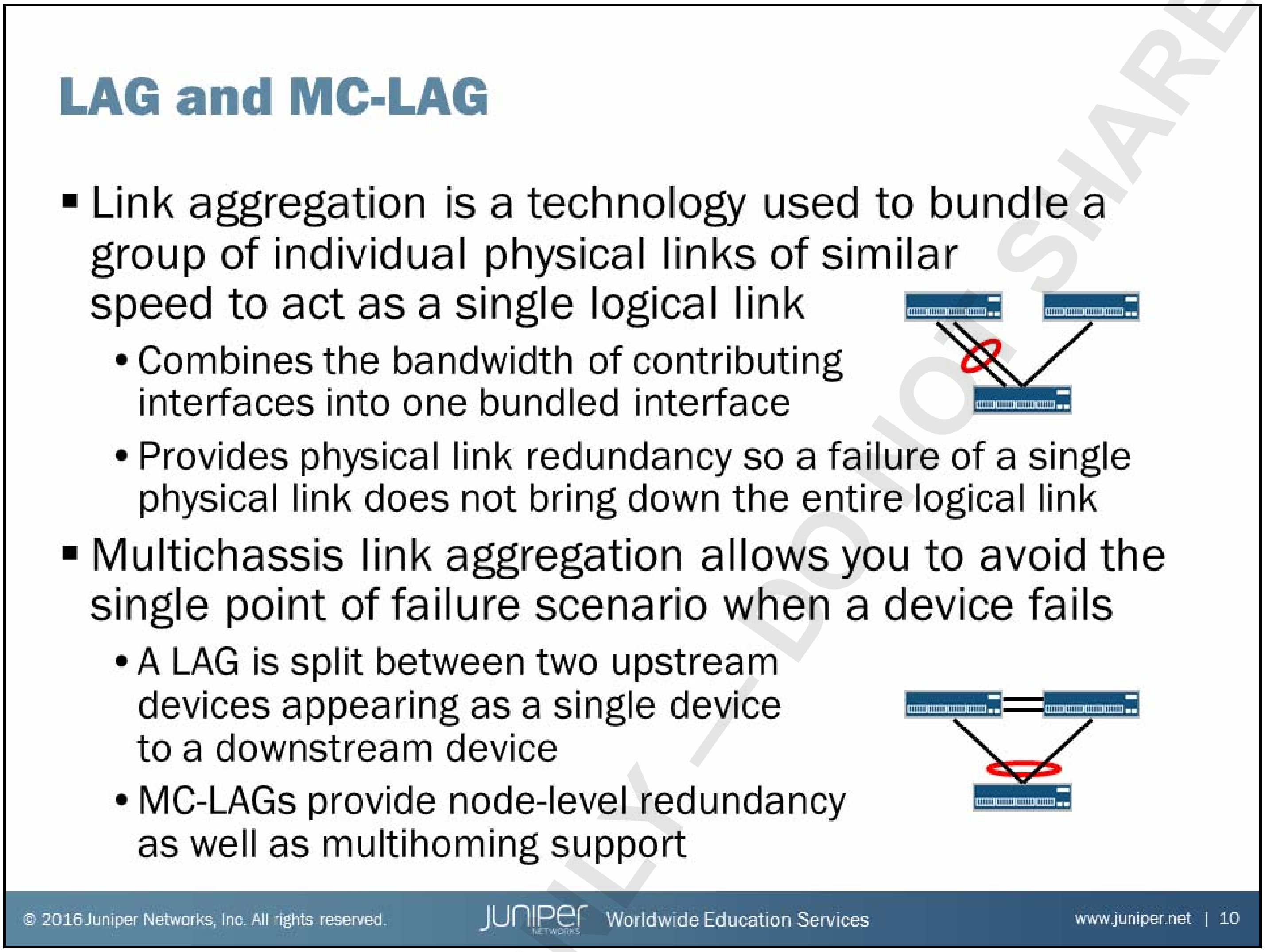
Device-level redundancy

- Hot-swappable interfaces and chassis components
- Redundant power supplies
- Software HA capabilities

Building a Highly Resilient Network

Addressing the resiliency and network availability requirements of a network can require one or more of the following:

- Link-level redundancy such as multiple paths, and link aggregation (LAG);
- Physical-device redundancy such as backup devices, high-availability (HA) clusters, virtual chassis, stateful failover for firewall platforms, and multichassis link aggregation (MC-LAG); and
- Device-level redundancy such as hot-swappable interface cards, chassis components, power supplies, and software high availability features.



LAG and MC-LAG

- Link aggregation is a technology used to bundle a group of individual physical links of similar speed to act as a single logical link
  - Combines the bandwidth of contributing interfaces into one bundled interface
  - Provides physical link redundancy so a failure of a single physical link does not bring down the entire logical link
- Multichassis link aggregation allows you to avoid the single point of failure scenario when a device fails
  - A LAG is split between two upstream devices appearing as a single device to a downstream device
  - MC-LAGs provide node-level redundancy as well as multihoming support

© 2016 Juniper Networks, Inc. All rights reserved. JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 10

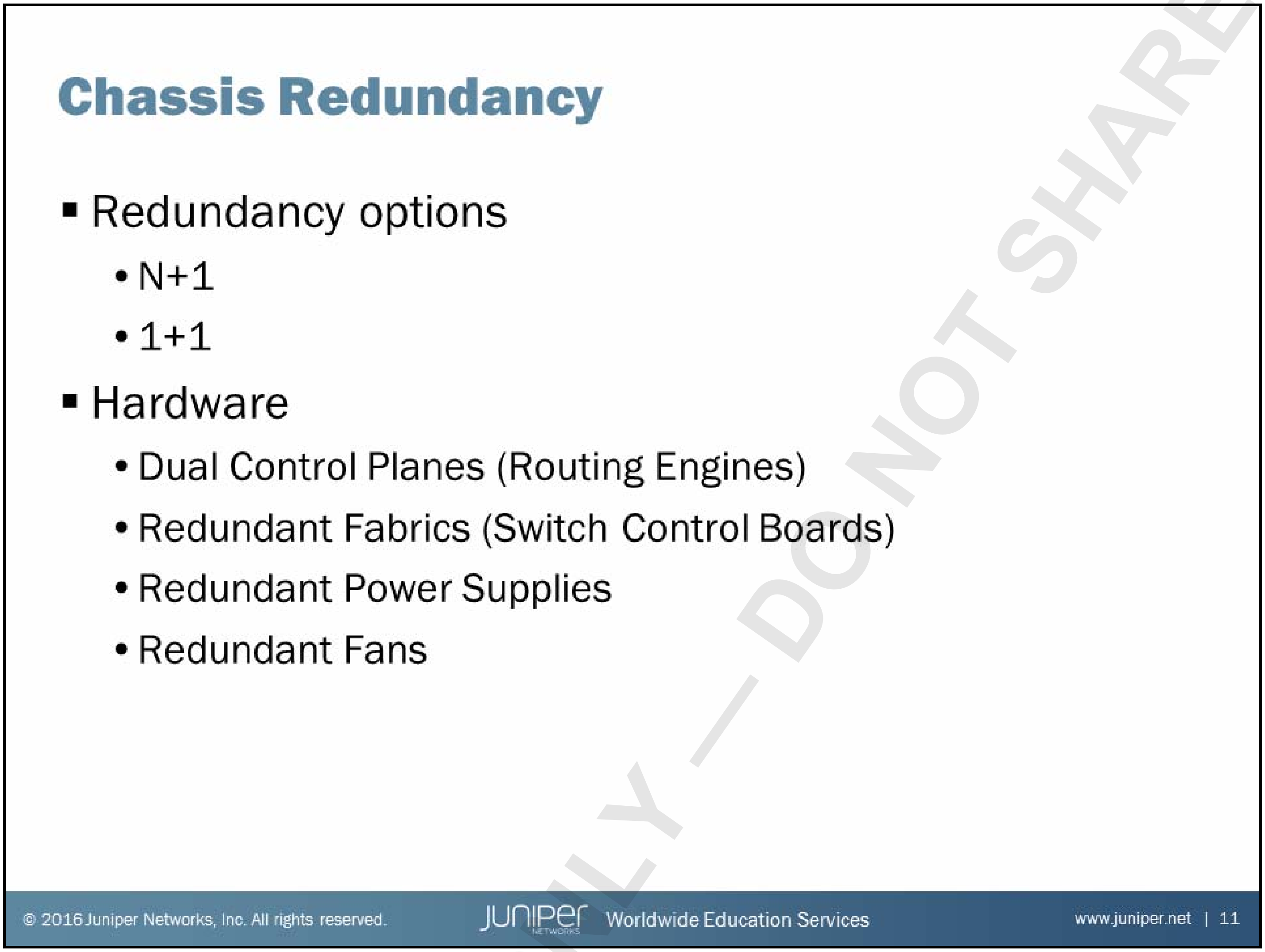
What Is LAG?

Link aggregation is a combination of multiple single physical links into one logical link. The advantage of link aggregation is twofold: it results in more aggregated bandwidth when compared to the individual link's bandwidth constituting link aggregation, and it provides the aggregate link's reliability because it consists of multiple physical links. If a Physical Layer failure of one of the link aggregation members occurs, the other members of the link will continue to forward traffic, although at a reduced bandwidth. For example, you have three 1GbE interfaces combined into a 3GbE aggregated Ethernet (AE) bundle. If one of those 1GbE interfaces fails, the interface (and bandwidth) is removed from the bundle but the remaining 2 1GbE interfaces continue to forward traffic participating in the 2GbE AE bundle. Most networking vendors support link aggregation technology (also referred to as bundling)—especially for connections between highly critical servers and network entities.

What Is MC-LAG?

An MC-LAG allows two similarly configured devices, known as MC-LAG peers, to emulate a logical LAG interface which connects to a separate device at the remote end of the LAG. The remote LAG endpoint may be a switch or router depending on the deployment scenario. The two MC-LAG peers appear to the remote endpoint connecting to the LAG as a single device.

MC-LAG builds on the standard LAG concept defined in 802.3ad and allows a LAG from one device to be spread between two upstream devices. MC-LAGs provide node-level redundancy as well as multihoming support for mission critical deployments.



Chassis Redundancy

- Redundancy options
  - N+1
  - 1+1
- Hardware
  - Dual Control Planes (Routing Engines)
  - Redundant Fabrics (Switch Control Boards)
  - Redundant Power Supplies
  - Redundant Fans

Worldwide Education Services

www.juniper.net | 11

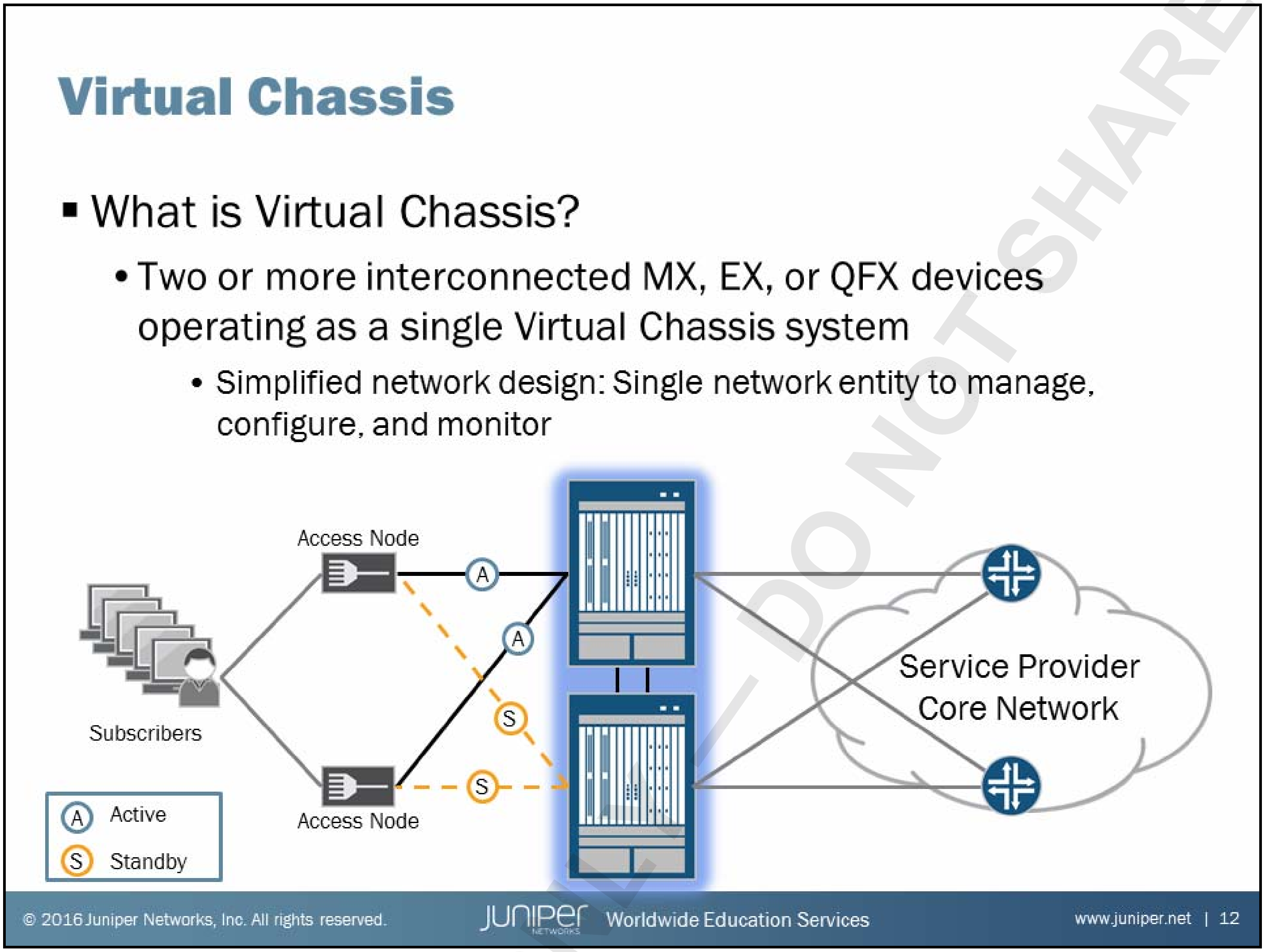
Redundancy Options

In a chassis there are components that can be built redundantly. These are often 1+1 (Primary and Backup) or N+1 (N number of something + 1 backup). In some systems it is possible to use the backup components and then run in reduced capacity mode when 1 of the fabrics fail, which might be preferable to an full outage.

Hardware

Common redundant hardware components include the control plane, switching fabrics, power supplies, and fans.

In addition, the chassis also has interface cards which are not necessarily backed up with hardware, but can potentially be backed up using software through virtualization techniques. The control plane is generally a 1+1 redundancy scheme where the backup control plane (Routing-Engine) is waiting for the primary Routing Engine to fail. Fabrics are platform specific and can be deployed 1+1, N+1, or even no backup with reduced capacity in the event of a failure. Redundant power supplies and fans also vary by device and chassis type.



Virtual Chassis

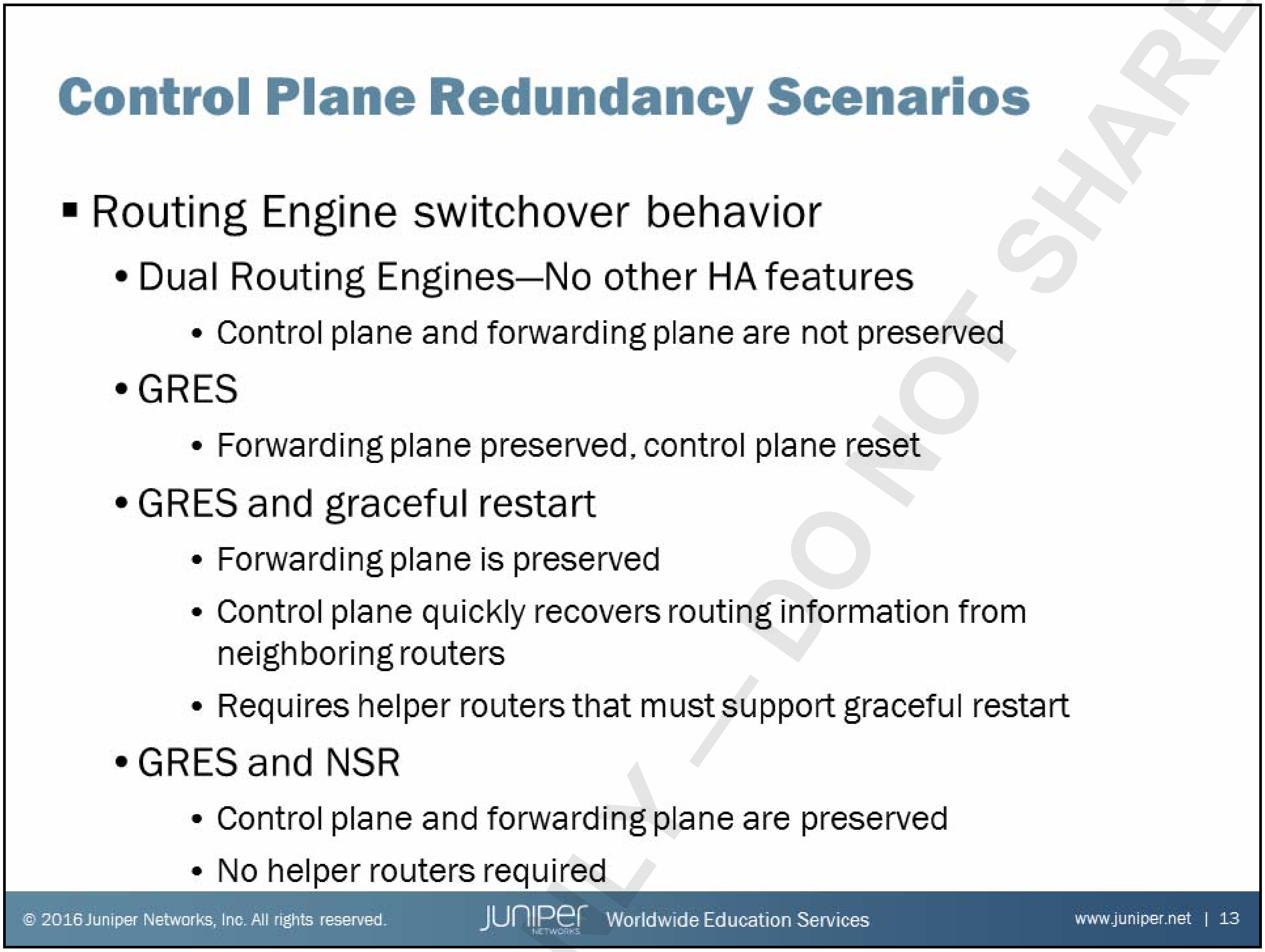
- What is Virtual Chassis?
  - Two or more interconnected MX, EX, or QFX devices operating as a single Virtual Chassis system
    - Simplified network design: Single network entity to manage, configure, and monitor

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 12

What Is a Virtual Chassis?

You can connect two or more like devices together to form one unit and manage the unit as a single chassis, called a Virtual Chassis. There are few different switch and router models that can be used to create a Virtual Chassis. Virtual Chassis can be used to provide inter-chassis redundancy for MX Series 3D Universal Edge Routers. As more high-priority voice and video traffic is carried on the network, inter-chassis redundancy has become a baseline requirement for providing stateful redundancy on broadband subscriber management equipment such as broadband services routers, broadband network gateways, and broadband remote access servers. A Virtual Chassis interconnects two MX, EX or QFX Series devices into a logical system that you can manage as a single network element.

The member devices in a Virtual Chassis are interconnected by means of Virtual Chassis ports that you configure on each member device. Comparatively speaking, managing a Virtual Chassis system is much simpler than managing many individual devices.



Control Plane Redundancy Scenarios

- Routing Engine switchover behavior
  - Dual Routing Engines—No other HA features
    - Control plane and forwarding plane are not preserved
  - GRES
    - Forwarding plane preserved, control plane reset
  - GRES and graceful restart
    - Forwarding plane is preserved
    - Control plane quickly recovers routing information from neighboring routers
    - Requires helper routers that must support graceful restart
  - GRES and NSR
    - Control plane and forwarding plane are preserved
    - No helper routers required

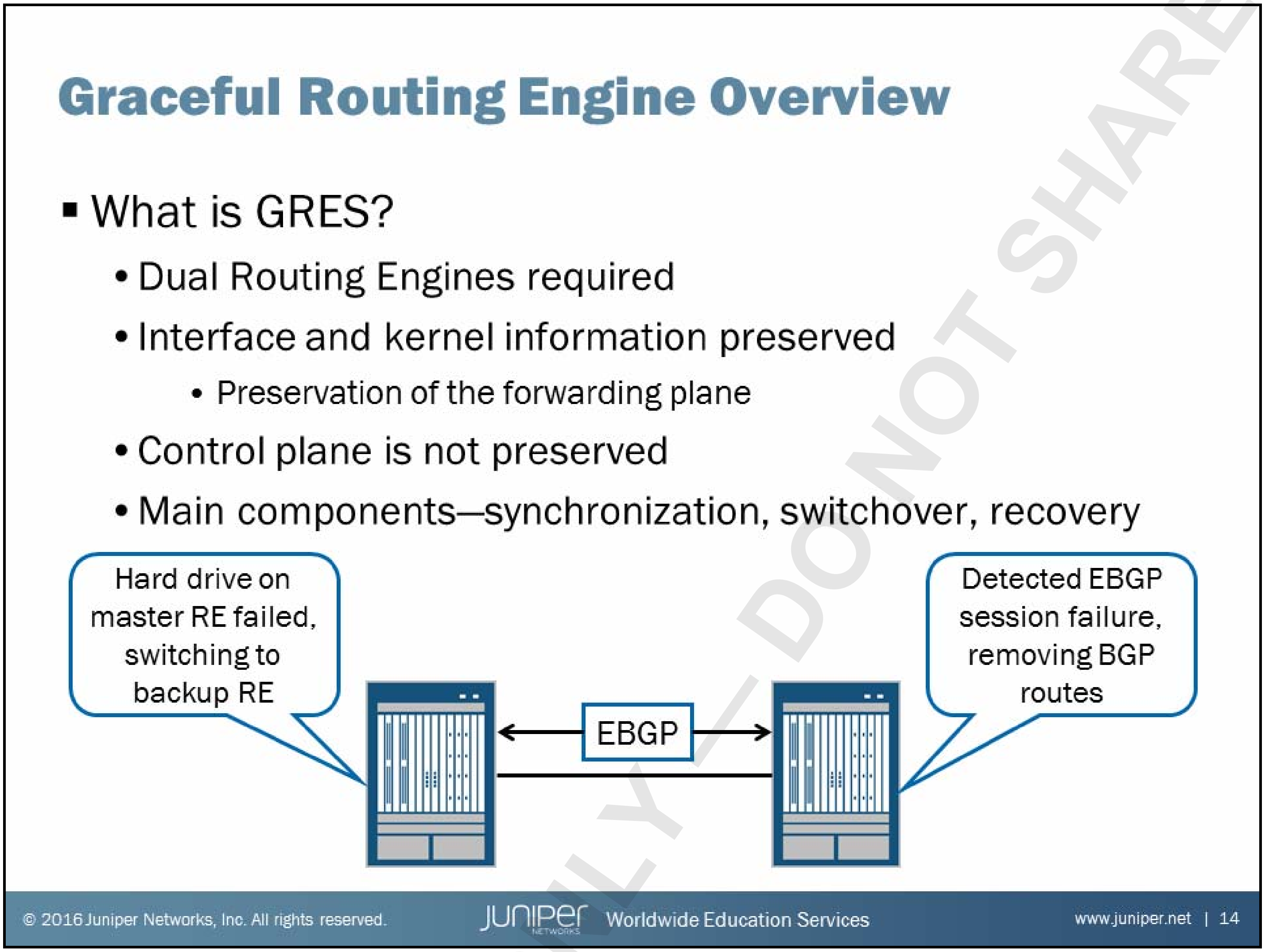
Juniper Networks Worldwide Education Services www.juniper.net | 13

Control Plane Redundancy Scenarios

The slide describes the effects of a Routing Engine switchover when no high availability (HA) features are enabled and when graceful Routing Engine switchover (GRES), graceful restart, and nonstop routing (NSR) features are enabled.

- Dual Routing Engines (no other HA features): When the switchover to the new master Routing Engine is complete, routing convergence takes place and traffic is resumed. All physical interfaces are taken offline, PFEs restart, the backup Routing Engine restarts the rpd process, and all hardware and interfaces are discovered by the new master Routing Engine. The switchover takes several minutes and all of the router's adjacencies are aware of the hardware and routing change.
- GRES only: During the switchover, interface and kernel information is preserved. The switchover is faster because the PFEs are not restarted. The new master Routing Engine restarts the rpd process. All hardware and interfaces are acquired by a process that is similar to a warm restart. All adjacencies are aware of the router's change in state.
- GRES and graceful restart: Traffic is not interrupted during the switchover. Interface and kernel information is preserved. Graceful restart protocol extensions quickly collect and restore routing information from the neighboring routers. Neighbors are required to support graceful restart and a wait interval is required. The rpd process restarts. For certain protocols, a significant change in the network can cause graceful restart to stop.
- GRES and NSR: Traffic is not interrupted during the switchover. Interface, kernel, and routing protocol information is preserved. Unsupported protocols must be refreshed using the normal recovery mechanisms inherent in each protocol.

When given the choice, we recommend that you always deploy NSR to provide control plane HA, and only employ graceful restart when you have to.



Graceful Routing Engine Overview

- What is GRES?
  - Dual Routing Engines required
  - Interface and kernel information preserved
    - Preservation of the forwarding plane
  - Control plane is not preserved
  - Main components—synchronization, switchover, recovery

Hard drive on master RE failed, switching to backup RE

Detected EBGP session failure, removing BGP routes

Juniper Networks Worldwide Education Services

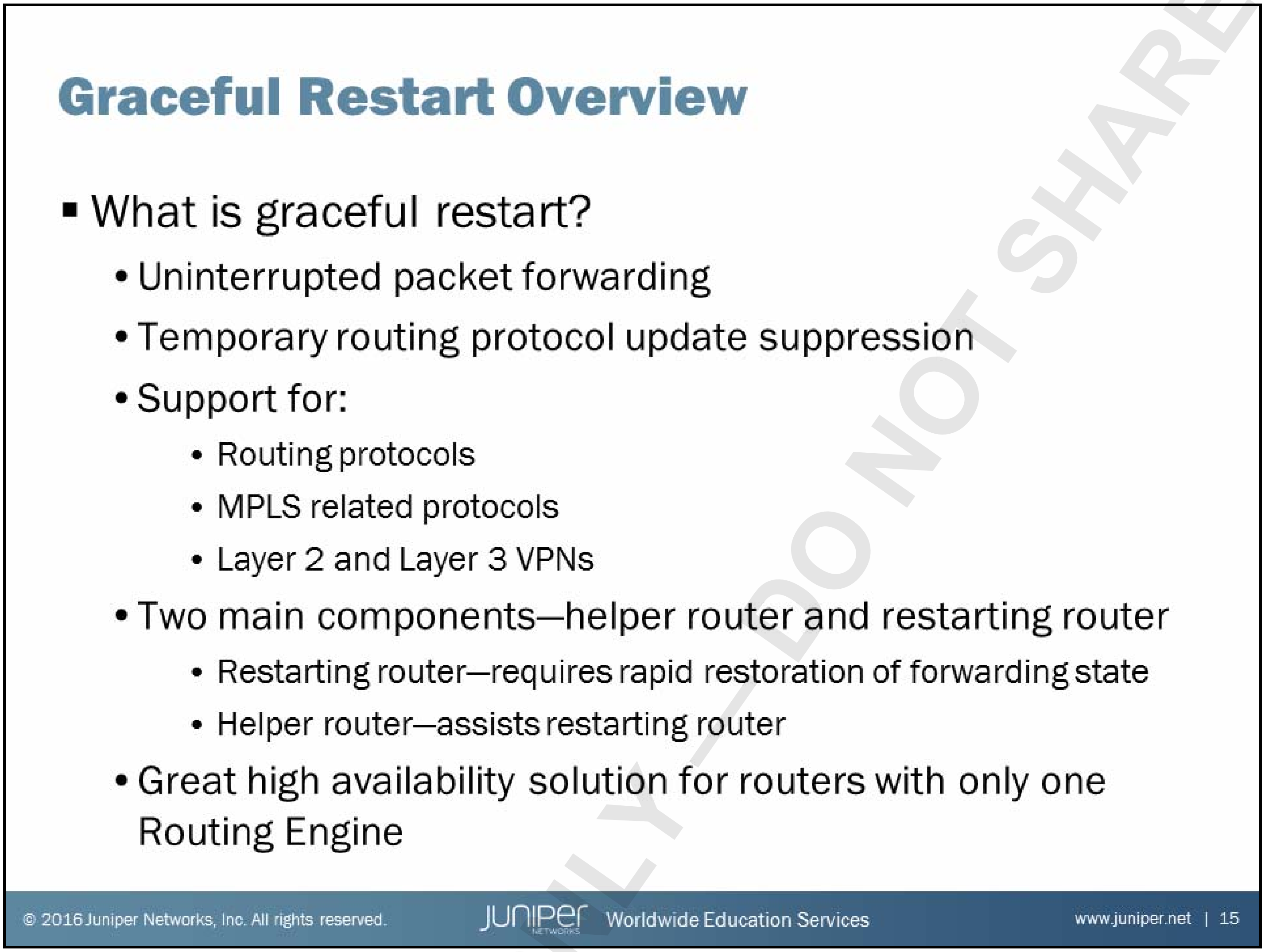
www.juniper.net | 14

What Is GRES?

The graceful Routing Engine switchover (GRES) feature in the Junos OS enables a routing platform with redundant Routing Engines to continue forwarding packets, even if one Routing Engine fails. GRES preserves interface and kernel information. Traffic is not interrupted, however, GRES does not preserve the control plane. Neighboring routers detect that the router has experienced a restart event and react to the event in a manner prescribed by individual routing protocol specifications. To preserve routing during a switchover, GRES must be combined with either graceful restart protocol extensions or NSR. Any updates to the master Routing Engine are replicated to the backup Routing Engine as soon as they occur. If the kernel on the master Routing Engine stops operating, the master Routing Engine experiences a hardware failure, or the administrator initiates a manual switchover, mastership switches to the backup Routing Engine.

If the backup Routing Engine does not receive a keepalive from the master Routing Engine after 2 seconds, it determines that the master Routing Engine has failed and takes mastership. The Packet Forwarding Engine (PFE) seamlessly disconnects from the old master Routing Engine and reconnects to the new master Routing Engine. The PFE does not reboot, and traffic is not interrupted. The new master Routing Engine and the PFE then become synchronized. If the new master Routing Engine detects that the PFE state is not up to date, it resends state update messages.

Successive Routing Engine switchover events must be a minimum of 240 seconds (4 minutes) apart after both Routing Engines have come up. If the router displays a warning message similar to “Standby Routing Engine is not ready for graceful switchover. Packet Forwarding Engines that are not ready for graceful switchover might be reset. Do not attempt the switchover.” If you choose to proceed with the switchover, the PFEs that were not ready for graceful switchover are reset. We recommend that you wait until the warning no longer appears and then proceed with the switchover.



Graceful Restart Overview

- What is graceful restart?
  - Uninterrupted packet forwarding
  - Temporary routing protocol update suppression
  - Support for:
    - Routing protocols
    - MPLS related protocols
    - Layer 2 and Layer 3 VPNs
  - Two main components—helper router and restarting router
    - Restarting router—requires rapid restoration of forwarding state
    - Helper router—assists restarting router
  - Great high availability solution for routers with only one Routing Engine

Worldwide Education Services www.juniper.net | 15

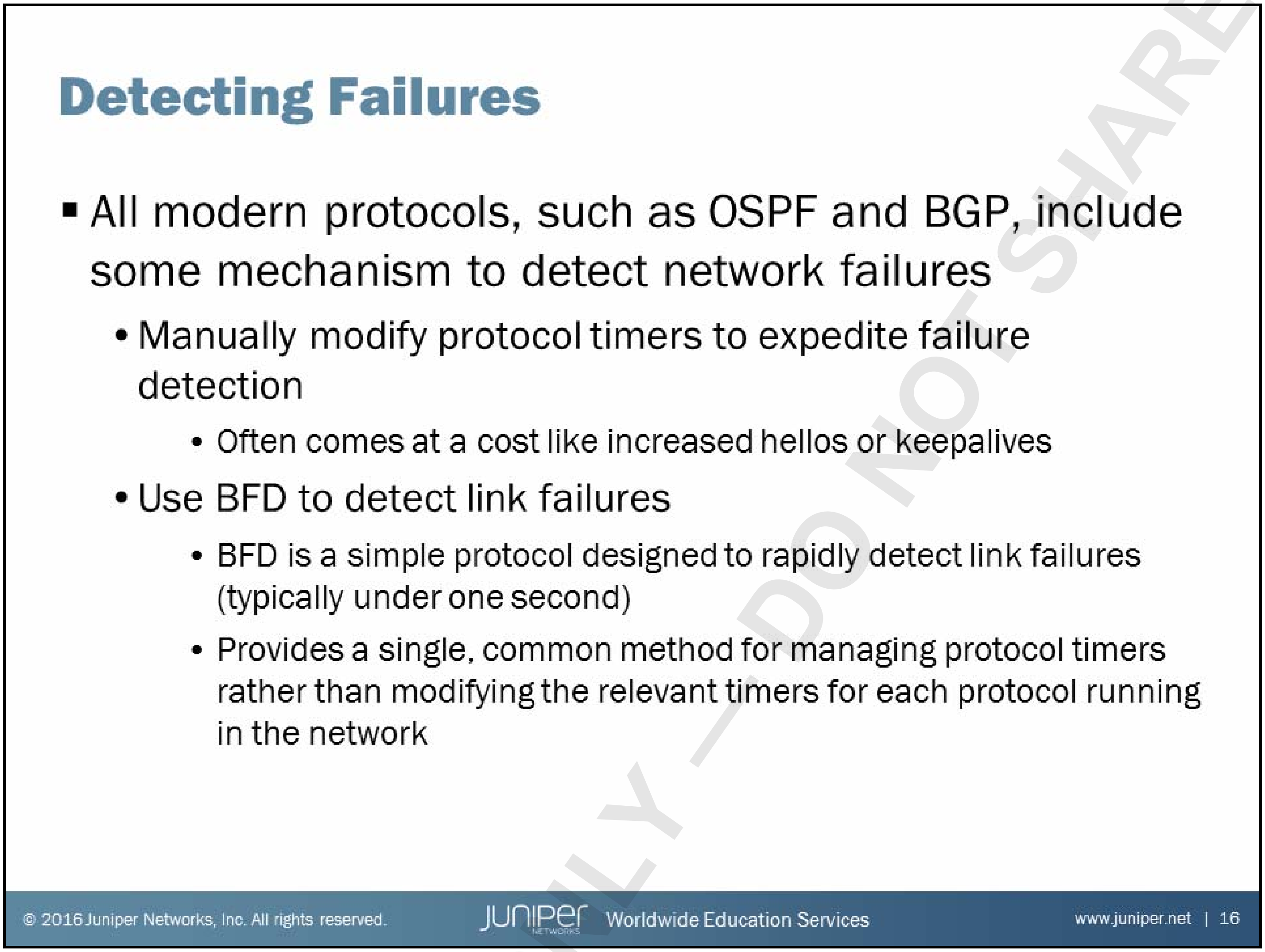
What Is Graceful Restart?

With routing protocols, any service interruption requires that an affected router recalculate adjacencies with neighboring routers, restore routing table entries, and update other protocol-specific information. An unprotected restart of a router can result in forwarding delays, route flapping, wait times stemming from protocol reconvergence, and even dropped packets. The main benefits of graceful restart are uninterrupted packet forwarding and temporary suppression of all routing protocol updates. Graceful restart enables a router to pass through intermediate convergence states that are hidden from the rest of the network. Three main types of graceful restart are available on Juniper Networks routing platforms:

- Graceful restart for aggregate and static routes and for routing protocols—Provides protection for aggregate and static routes and for Border Gateway Protocol (BGP), End System-to-Intermediate System (ES-IS), Intermediate System-to-Intermediate System (IS-IS), Open Shortest Path First (OSPF), Routing Information Protocol (RIP), next-generation RIP (RIPng), and Protocol Independent Multicast (PIM) sparse mode routing protocols.
- Graceful restart for MPLS-related protocols—Provides protection for Label Distribution Protocol (LDP), Resource Reservation Protocol (RSVP), circuit cross-connect (CCC), and translational cross-connect (TCC).
- Graceful restart for virtual private networks (VPNs)—Provides protection for Layer 2 and Layer 3 VPNs.

Graceful restart works similarly for routing protocols and MPLS protocols and combines components of these protocol types to enable graceful restart in VPNs.

Most graceful restart implementations define two types of routers—the restarting router and the helper router. The restarting router requires rapid restoration of forwarding state information so it can resume the forwarding of network traffic. The helper router assists the restarting router in this process. Graceful restart configuration statements typically affect either the restarting router or the helper router. Graceful restart not only helps provide HA in routing platforms with dual routing engines, but you can also use it as a method to provide HA for routing platforms with only a single routing engine.



Detecting Failures

- All modern protocols, such as OSPF and BGP, include some mechanism to detect network failures
  - Manually modify protocol timers to expedite failure detection
    - Often comes at a cost like increased hellos or keepalives
  - Use BFD to detect link failures
    - BFD is a simple protocol designed to rapidly detect link failures (typically under one second)
    - Provides a single, common method for managing protocol timers rather than modifying the relevant timers for each protocol running in the network

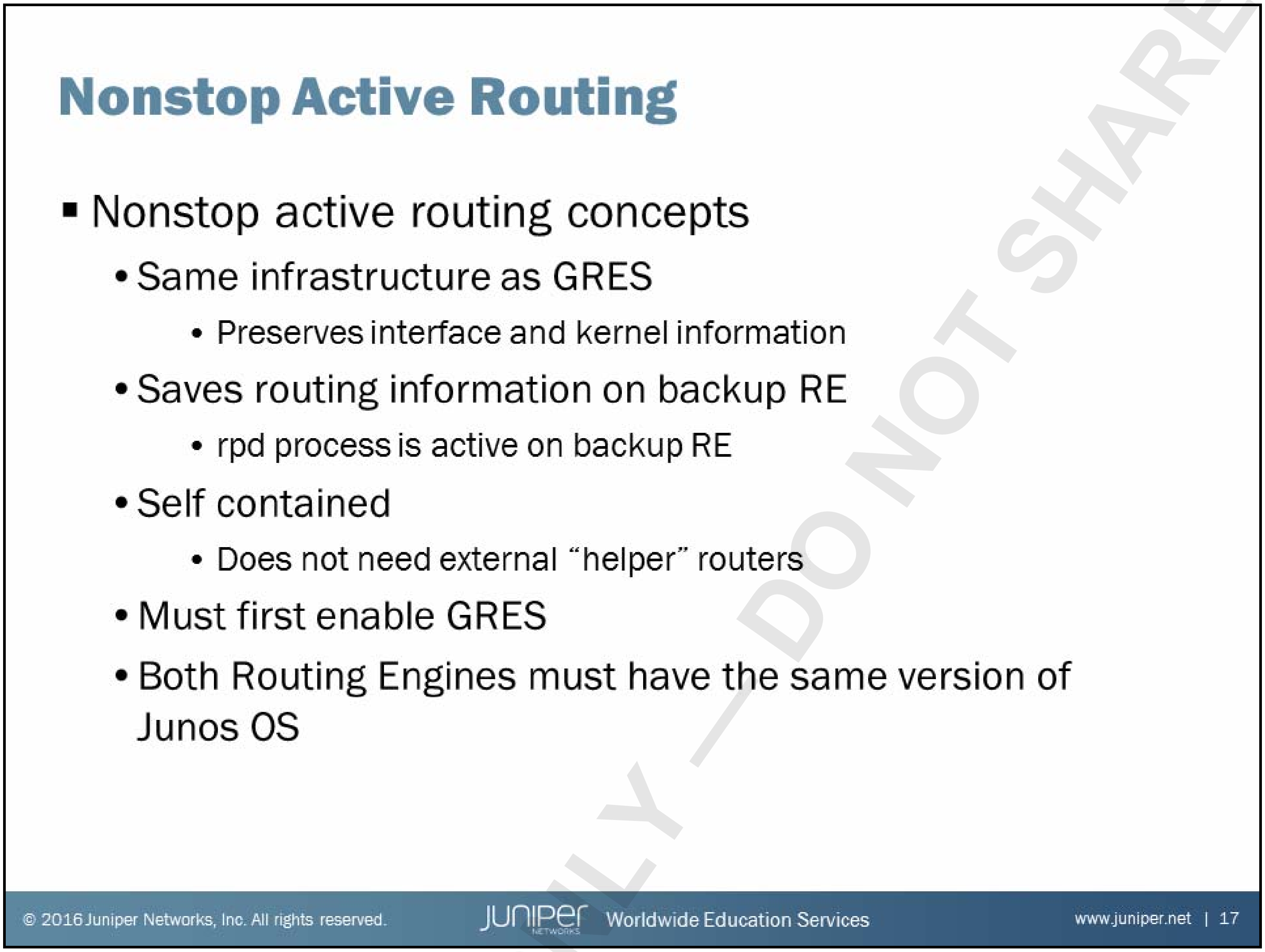
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 16

Detecting Network Failures

All modern protocols, such as OSPF and BGP, include some mechanism to detect network failures. One problem with protocol failure detection mechanisms is that they can be slow (especially with the default timers). For example, OSPF using its default timers, can take up to 40 seconds before a neighbor is declared dead.

You can adjust the default timers to lower the time it takes a protocol to detect failures and declare a neighbor or peer dead. In fact, you can adjust the dead timers for some interior gateway protocols (IGPs), such as OSPF and IS-IS, to detect failures in about 1 second. There is, however, a cost associated with lowering protocol timers. Lowering a protocol's timers often equates to an increase of hellos or keepalives and, consequently, more processing overhead for the related protocol and routing process. Increasing the load on a protocol or the routing process can potentially cause undesirable results and adversely affect a router's overall performance, especially in large networks.

Instead of tuning individual protocol parameters, network designers commonly use the Bi-directional Forwarding Detection (BFD) protocol to rapidly detect neighbor failure. The BFD protocol is a simple protocol designed to rapidly detect link failures. Once two devices negotiate and establish a BFD session, BFD continuously sends hellos to monitor the associated link. If BFD stops receiving hellos from its neighbor, it takes down the session and notifies the system that a communication problem exists. BFD can detect link failures in less than a second, which means hellos are exchanged and processed frequently and efficiently. BFD can provide a number of key benefits within a high availability network. BFD relieves protocols from being required to provide fast-failure detection; in fact, routing protocol timers for hellos or keepalives can be left at their default values or even be increased to reduce the associated processing. BFD also provides a single, common method for managing protocol timers. Rather than modifying the relevant timers for each protocol running in the network, you can leave the protocol timers at their default settings and simply implement BFD to use consistent timer values for all protocols. Another key benefit provided by BFD is that it provides a failure detection mechanism for static routes, which, unlike modern routing protocols, do not have such a mechanism otherwise.



Nonstop Active Routing

- Nonstop active routing concepts
  - Same infrastructure as GRES
    - Preserves interface and kernel information
  - Saves routing information on backup RE
    - rpd process is active on backup RE
  - Self contained
    - Does not need external “helper” routers
  - Must first enable GRES
  - Both Routing Engines must have the same version of Junos OS

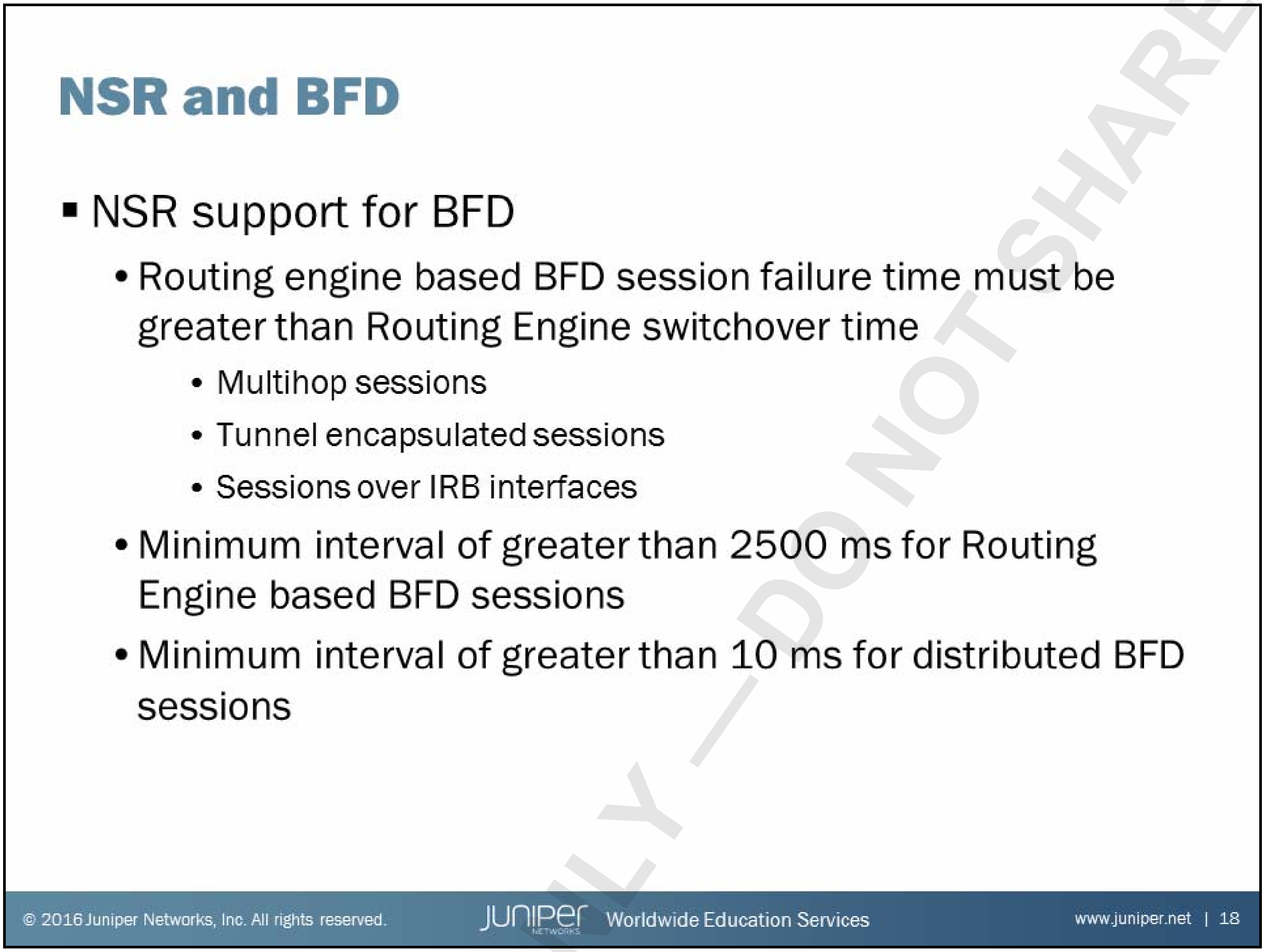
Juniper NETWORKS Worldwide Education Services

www.juniper.net | 17

Nonstop Active Routing Concepts

Nonstop Active Routing (NSR) uses the same infrastructure as GRES to preserve interface and kernel information. However, NSR also saves routing protocol information by running the rpd process on the backup Routing Engine. By saving this additional information, NSR is self-contained and does not rely on helper routers to assist the routing platform in restoring routing protocol information. NSR is advantageous in networks where neighbor routers do not support graceful restart protocol extensions. As a result of this enhanced functionality, NSR is a natural replacement for graceful restart.

Note that to use NSR, you must first enable GRES on your routing platform and both Routing Engines must be on the same Junos OS version.



NSR and BFD

- NSR support for BFD
  - Routing engine based BFD session failure time must be greater than Routing Engine switchover time
    - Multihop sessions
    - Tunnel encapsulated sessions
    - Sessions over IRB interfaces
  - Minimum interval of greater than 2500 ms for Routing Engine based BFD sessions
  - Minimum interval of greater than 10 ms for distributed BFD sessions

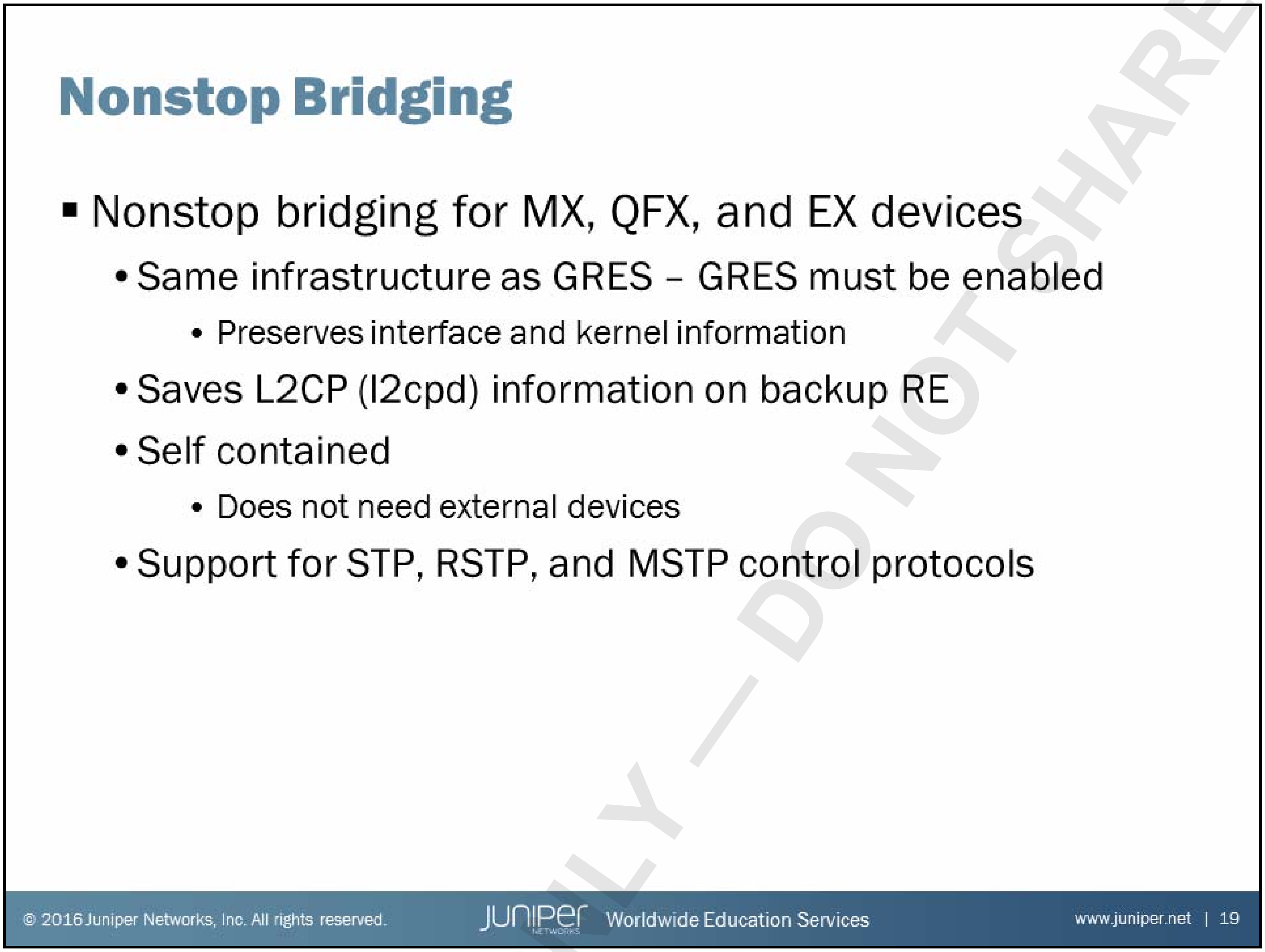
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 18

NSR Support for BFD

NSR supports the BFD protocol, which uses the topology discovered by routing protocols to monitor neighbors. As mentioned earlier, the BFD protocol is a simple hello mechanism that detects failures in a network. Because BFD is streamlined to be efficient at fast liveness detection, when it is used in conjunction with routing protocols, routing recovery times are improved. With NSR enabled, BFD session states are not restarted when a Routing Engine switchover occurs.

When a BFD session is distributed to the PFE, BFD packets continue to be sent during a Routing Engine switchover. If nondistributed, or Routing Engine based, BFD sessions are to be kept alive during a switchover, you must ensure that the session failure detection time is greater than the Routing Engine switchover time, which in most situations is 2 seconds. The following BFD sessions are not distributed to the PFE: multihop sessions, tunnel-encapsulated sessions, and sessions over integrated routing and bridging (IRB) interfaces. An example of a multihop session is a standard IBGP peering or a multihop EBGP peering.

For BFD sessions to remain up during a Routing Engine switchover event when NSR is configured, specify a minimum interval of 2500 ms or greater for Routing Engine based sessions. For distributed BFD sessions with NSR configured, we recommend that you set the a minimum interval of 10 ms or greater.



Nonstop Bridging

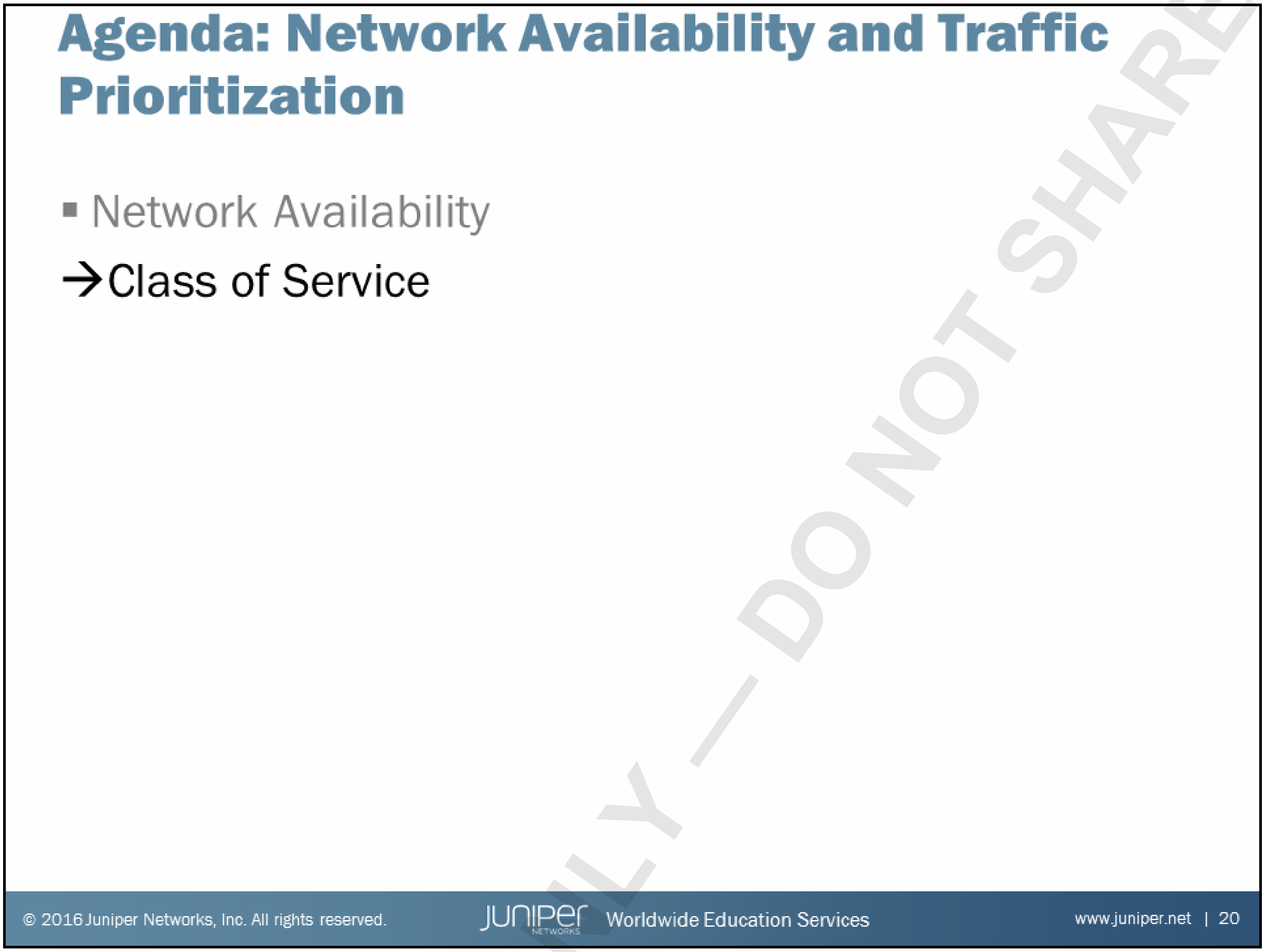
- Nonstop bridging for MX, QFX, and EX devices
  - Same infrastructure as GRES – GRES must be enabled
    - Preserves interface and kernel information
  - Saves L2CP (l2cpd) information on backup RE
  - Self contained
    - Does not need external devices
  - Support for STP, RSTP, and MSTP control protocols

Worldwide Education Services

www.juniper.net | 19

Nonstop Bridging Concepts

Nonstop bridging (NSB) uses the same infrastructure as GRES to preserve interface and kernel information. This reliance on GRES requires that GRES is enabled before you can enable NSB. NSB saves Layer 2 Control Protocol (L2CP) information by running the Layer 2 Control Protocol process daemon (`l2cpd`) on the backup Routing Engine. By running the `l2cpd` process on the backup Routing Engine, the router is able to maintain the state of the Spanning Tree Protocol (STP), Rapid Spanning Tree Protocol (RSTP), and Multiple Spanning Tree Protocol (MSTP) control protocols through a Routing Engine switchover without the help of an external device.



Agenda: Network Availability and Traffic Prioritization

- Network Availability

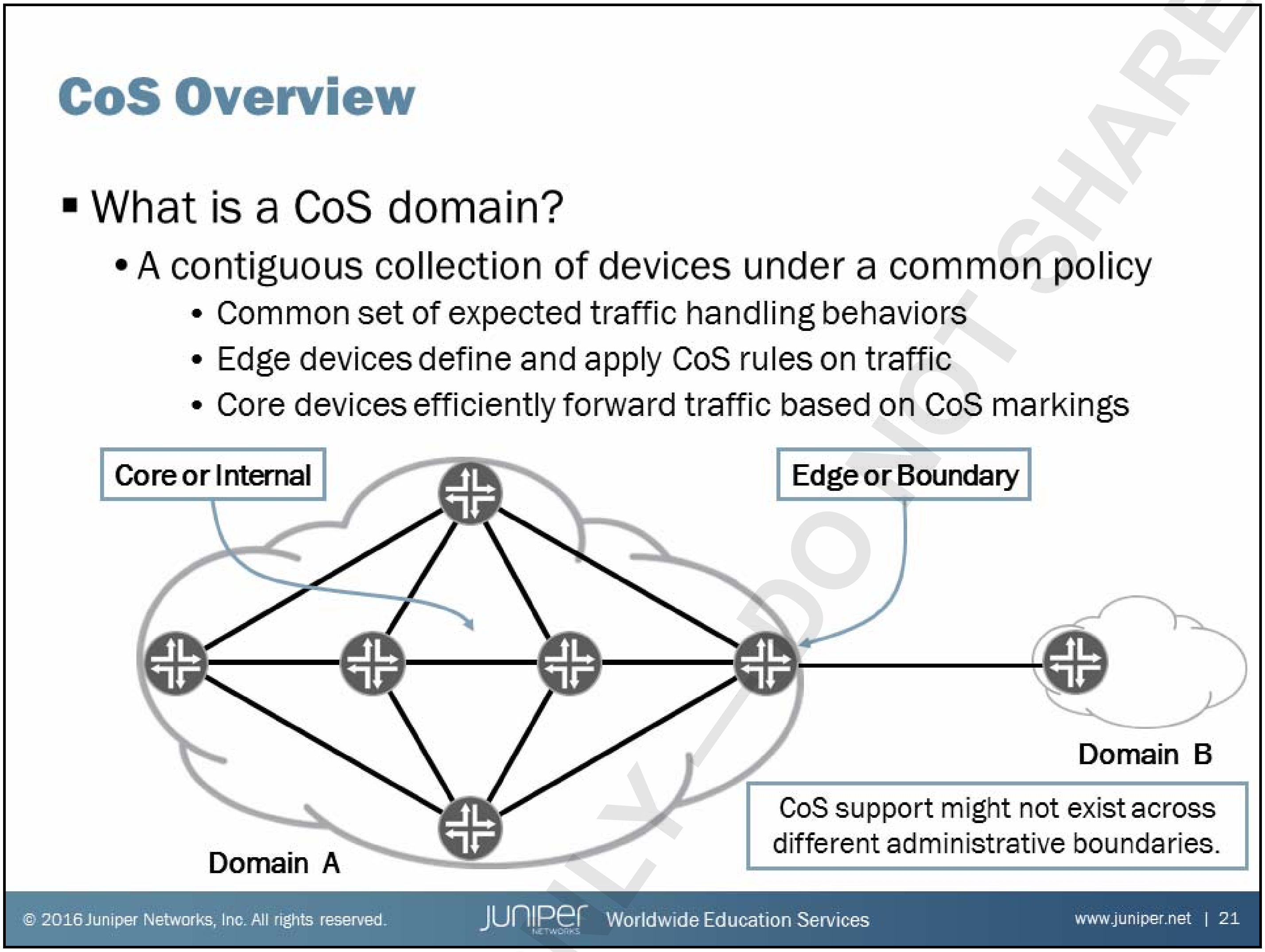
→Class of Service

Juniper Networks Worldwide Education Services

www.juniper.net | 20

Class of Service

The slide highlights the topic we discuss next.



CoS Overview

- What is a CoS domain?
  - A contiguous collection of devices under a common policy
    - Common set of expected traffic handling behaviors
    - Edge devices define and apply CoS rules on traffic
    - Core devices efficiently forward traffic based on CoS markings

Core or Internal

Edge or Boundary

Domain B

Domain A

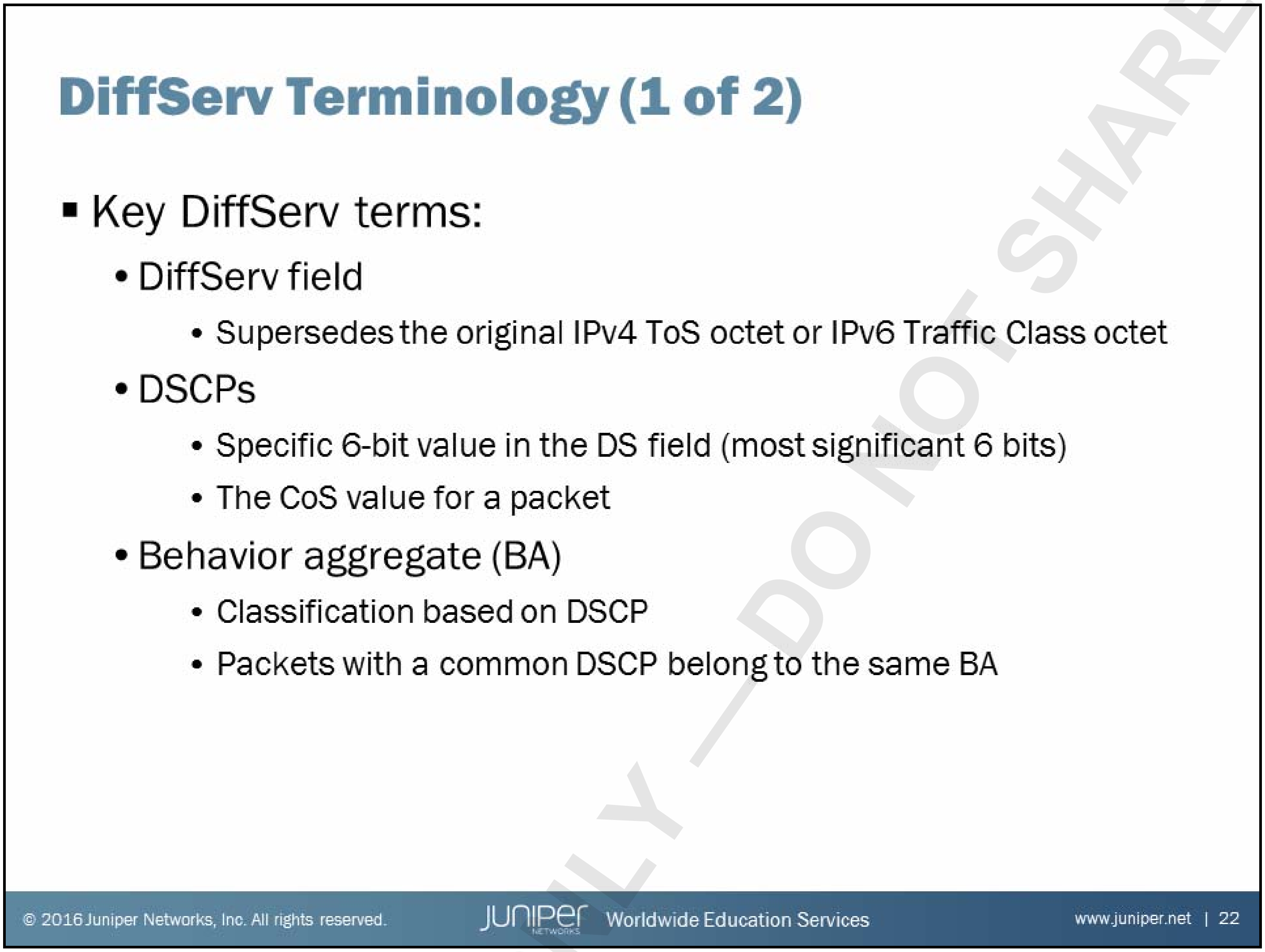
CoS support might not exist across different administrative boundaries.

Juniper Networks Worldwide Education Services

www.juniper.net | 21

CoS Domain

A CoS domain is a collection of devices under a common administrative control with a common expectation for traffic handling. A domain is made up of edge, or boundary, devices as well as internal (core) devices. This distinction is a critical point because the role of a given device varies based on its designation. In general terms, edge devices tend to have more complex data handling and manipulation functions when compared to core devices. For example, policing, shaping, metering (accounting), and complex multifield classification functions are normally performed by devices that are attached to customers or other domains. In contrast, core devices normally perform only behavior aggregate (BA) based classification and transmission scheduling based on defined CoS rules. Because all devices in a CoS domain are under common control and make use of a common set of defined CoS rules, expecting that end-to-end performance and service levels can be predicted and met is reasonable.



DiffServ Terminology (1 of 2)

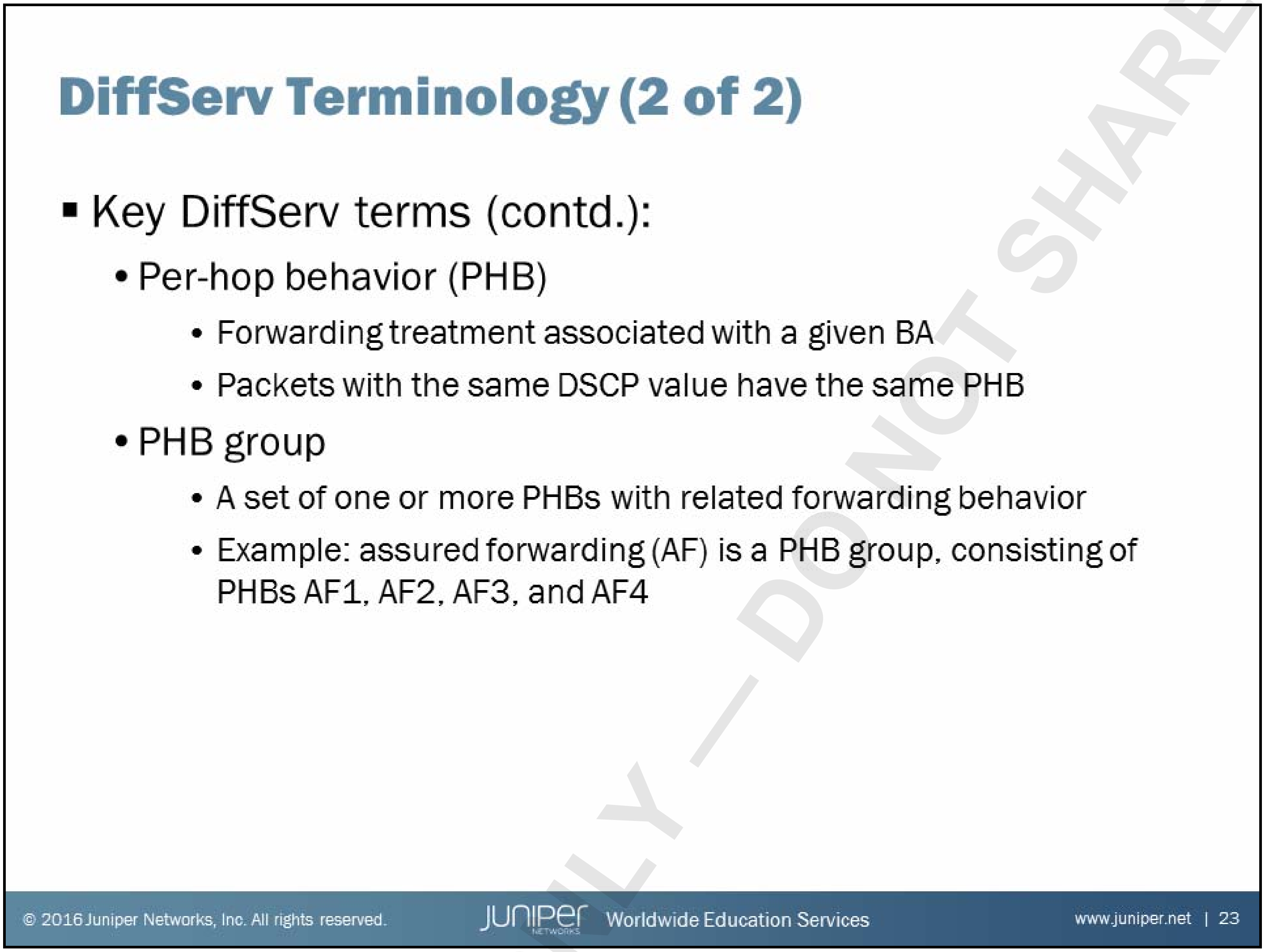
- Key DiffServ terms:
  - DiffServ field
    - Supersedes the original IPv4 ToS octet or IPv6 Traffic Class octet
  - DSCPs
    - Specific 6-bit value in the DS field (most significant 6 bits)
    - The CoS value for a packet
  - Behavior aggregate (BA)
    - Classification based on DSCP
    - Packets with a common DSCP belong to the same BA

Juniper Networks Worldwide Education Services www.juniper.net | 22

DiffServ Terms: Part 1

The slide defines several terms that are critical to a basic understanding of DiffServ:

- DiffServ field (DS field): The DS field refers to the original IPv4 ToS field that is redefined to carry DSCPs;
- DSCPs: 6-bit CoS values; and
- Behavior aggregate (BA): A BA describes the logical grouping of traffic flows into an aggregate flow that requires similar handling and treatment.



DiffServ Terminology (2 of 2)

- Key DiffServ terms (contd.):
  - Per-hop behavior (PHB)
    - Forwarding treatment associated with a given BA
    - Packets with the same DSCP value have the same PHB
  - PHB group
    - A set of one or more PHBs with related forwarding behavior
    - Example: assured forwarding (AF) is a PHB group, consisting of PHBs AF1, AF2, AF3, and AF4

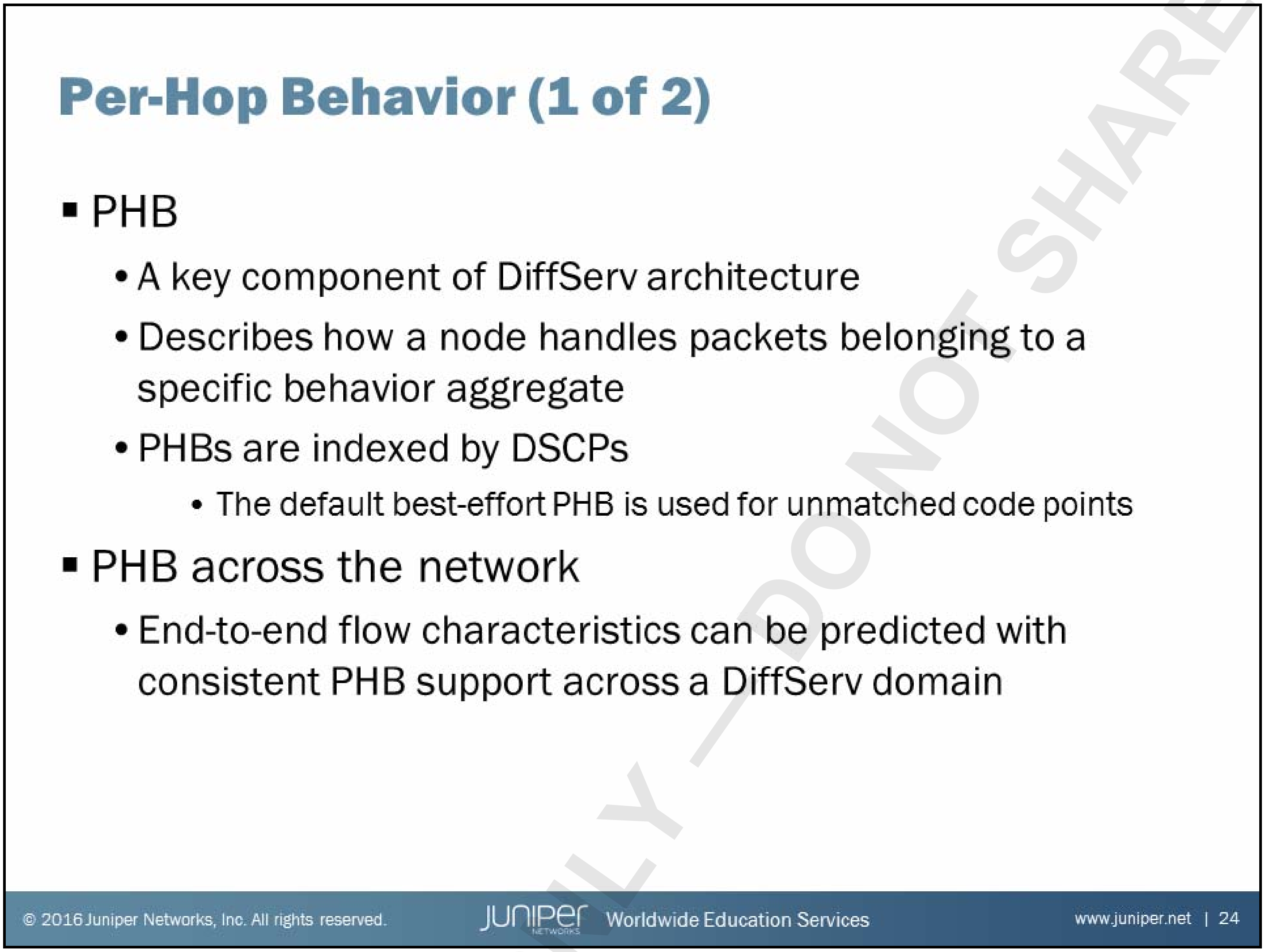
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 23

DiffServ Terms: Part 2

This list is a continuation from the previous page:

- Per-hop behavior (PHB): A device's PHB describes the externally visible manner in which packets are handled and forwarded for a given BA. For example, a PHB could result in a device marking all traffic associated with the best-effort class with a common DSCP.
- PHB group: A per-hop behavior group defines a set of related PHBs. For example, the assured forwarding (AF) group consists of four separate classes, AF1, AF2, AF3, and AF4, each with three possible drop profiles. The AF1 class defines a particular PHB, while the AF category defines a per-hop grouping.

For more detailed information, refer to section 1.2 of RFC 2475.



Per-Hop Behavior (1 of 2)

- PHB
  - A key component of DiffServ architecture
  - Describes how a node handles packets belonging to a specific behavior aggregate
  - PHBs are indexed by DSCPs
    - The default best-effort PHB is used for unmatched code points
- PHB across the network
  - End-to-end flow characteristics can be predicted with consistent PHB support across a DiffServ domain

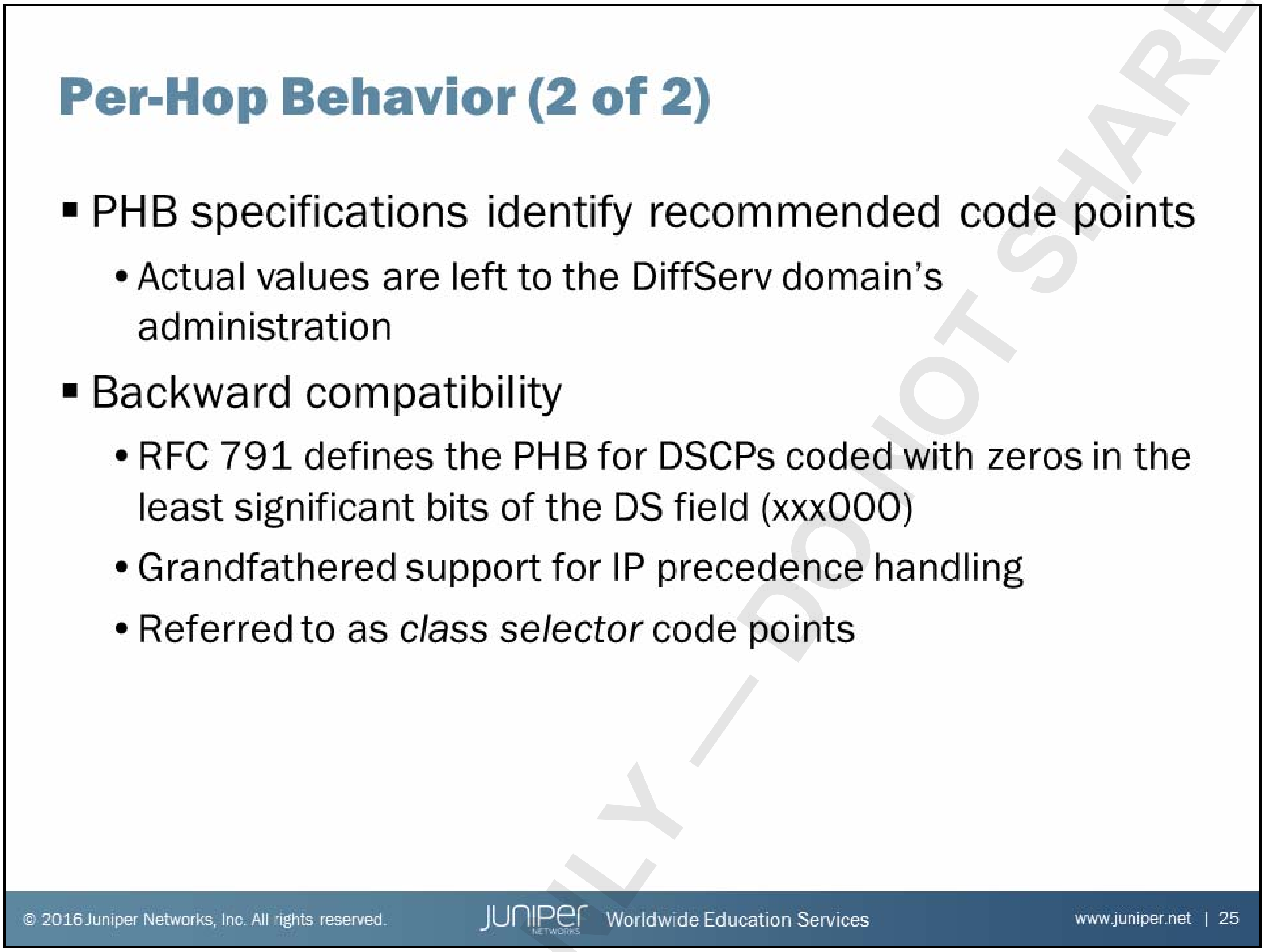
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 24

Per-Hop Behavior

A key component of the DiffServ architecture is the concept of a PHB that describes the externally visible way in which a DiffServ node handles traffic belonging to a given forwarding class (or BA). The particular PHB applied to a packet is a function of ingress classification using the DSCP. A DiffServ node must have a default PHB available for use when handling unclassified traffic. The default PHB is equivalent to conventional best-effort forwarding.

PHB Across the Network

Because all the devices in a DiffServ domain implement a common set of PHBs, the end-to-end performance of the network can be accurately modeled. This modeling normally assumes that the network has appropriate protection in place to guard against unusual traffic patterns that could negate capacity planning assumptions and jeopardize service-level agreements (SLAs).



Per-Hop Behavior (2 of 2)

- PHB specifications identify recommended code points
  - Actual values are left to the DiffServ domain's administration
- Backward compatibility
  - RFC 791 defines the PHB for DSCPs coded with zeros in the least significant bits of the DS field (xxx000)
  - Grandfathered support for IP precedence handling
  - Referred to as *class selector* code points

Worldwide Education Services

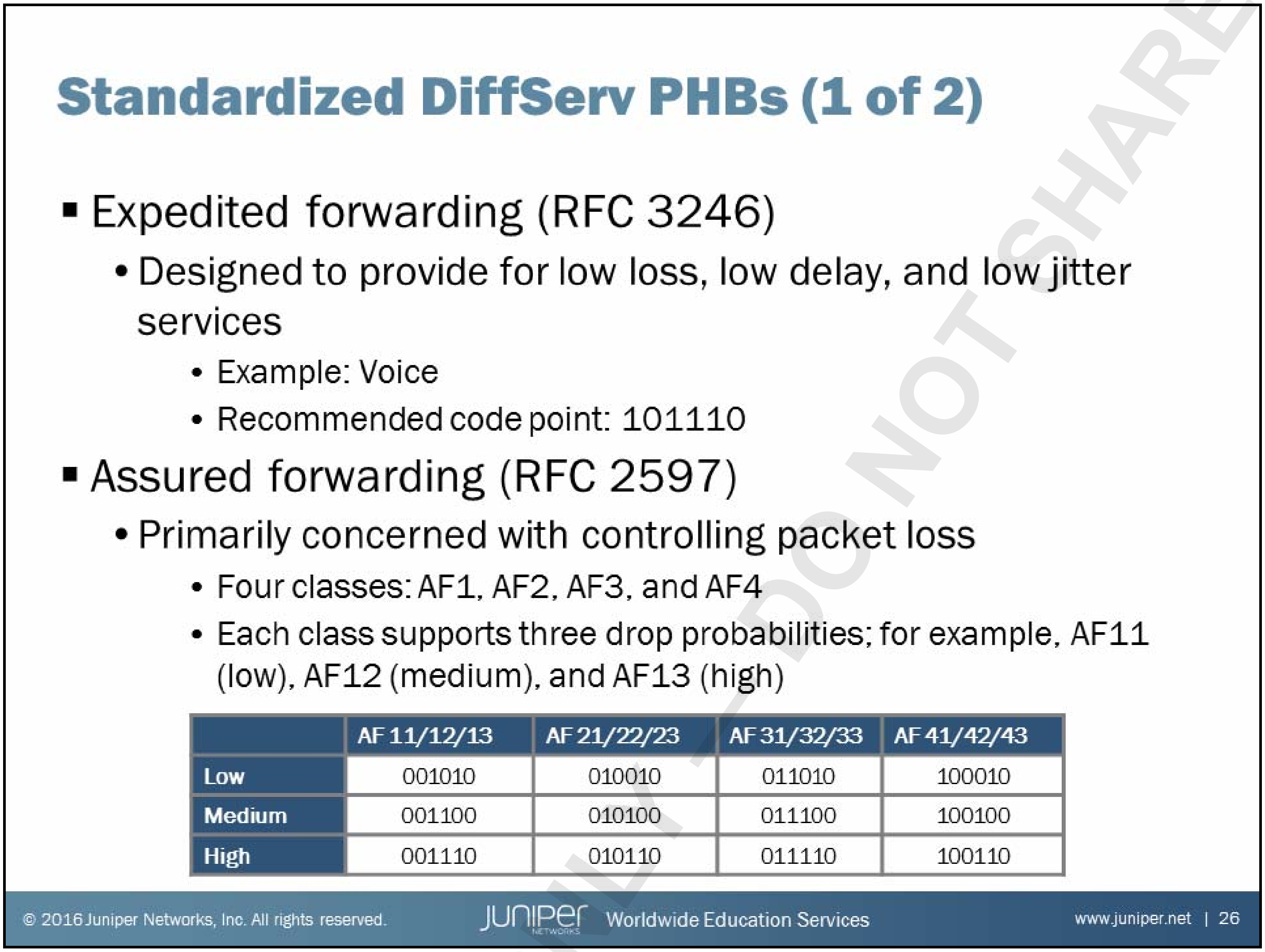
www.juniper.net | 25

PHB Specifications Identify Recommended Code Points

A PHB specification normally includes one or more recommendations detailing the DSCPs associated with that PHB. Note that these specifications are only recommendations and that the actual code points used to identify a given BA are left to the administrative authority for that domain.

Backward Compatibility with IP Precedence

The PHB for DSCPs with zeros in the three least-significant bits of the DS field is defined as being compatible with the historic treatment of IP precedence, as outlined in RFC 791. The eight code points associated with IP precedence compatibility are not officially given a forwarding class designation; they are simply referred to as class selector (CS) code points.



Standardized DiffServ PHBs (1 of 2)

- Expedited forwarding (RFC 3246)
  - Designed to provide for low loss, low delay, and low jitter services
    - Example: Voice
    - Recommended code point: 101110
- Assured forwarding (RFC 2597)
  - Primarily concerned with controlling packet loss
    - Four classes: AF1, AF2, AF3, and AF4
    - Each class supports three drop probabilities; for example, AF11 (low), AF12 (medium), and AF13 (high)

	AF 11/12/13	AF 21/22/23	AF 31/32/33	AF 41/42/43
Low	001010	010010	011010	100010
Medium	001100	010100	011100	100100
High	001110	010110	011110	100110

Worldwide Education Services

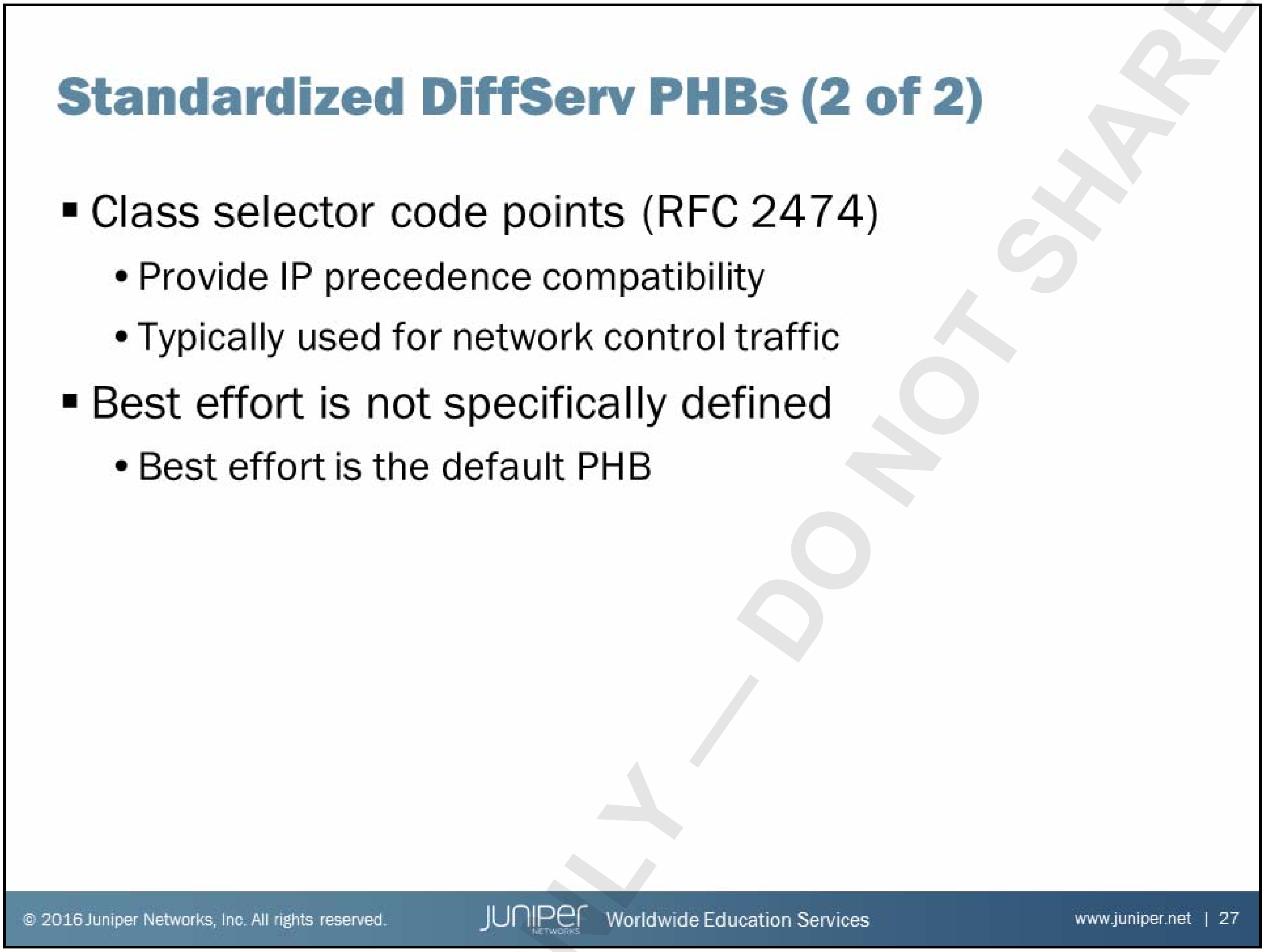
www.juniper.net | 26

Expedited Forwarding

RFC 3246 defines the expedited-forwarding (EF) PHB. This PHB is designed to provide low loss, low latency, and low jitter services to support delay-sensitive and loss-sensitive applications such as voice or video conferencing.

Assured Forwarding

RFC 2597 defines the AF PHB. This PHB is primarily concerned with packet loss, because no delay-related parameters are defined. Within the AF category, four classes exist: AF1, AF2, AF3, and AF4. Within each category, three classes are defined that differ based upon their loss probability. Put another way, AF11 should have a lower percentage of packet drops when compared to AF43.



Standardized DiffServ PHBs (2 of 2)

- Class selector code points (RFC 2474)
  - Provide IP precedence compatibility
  - Typically used for network control traffic
- Best effort is not specifically defined
  - Best effort is the default PHB

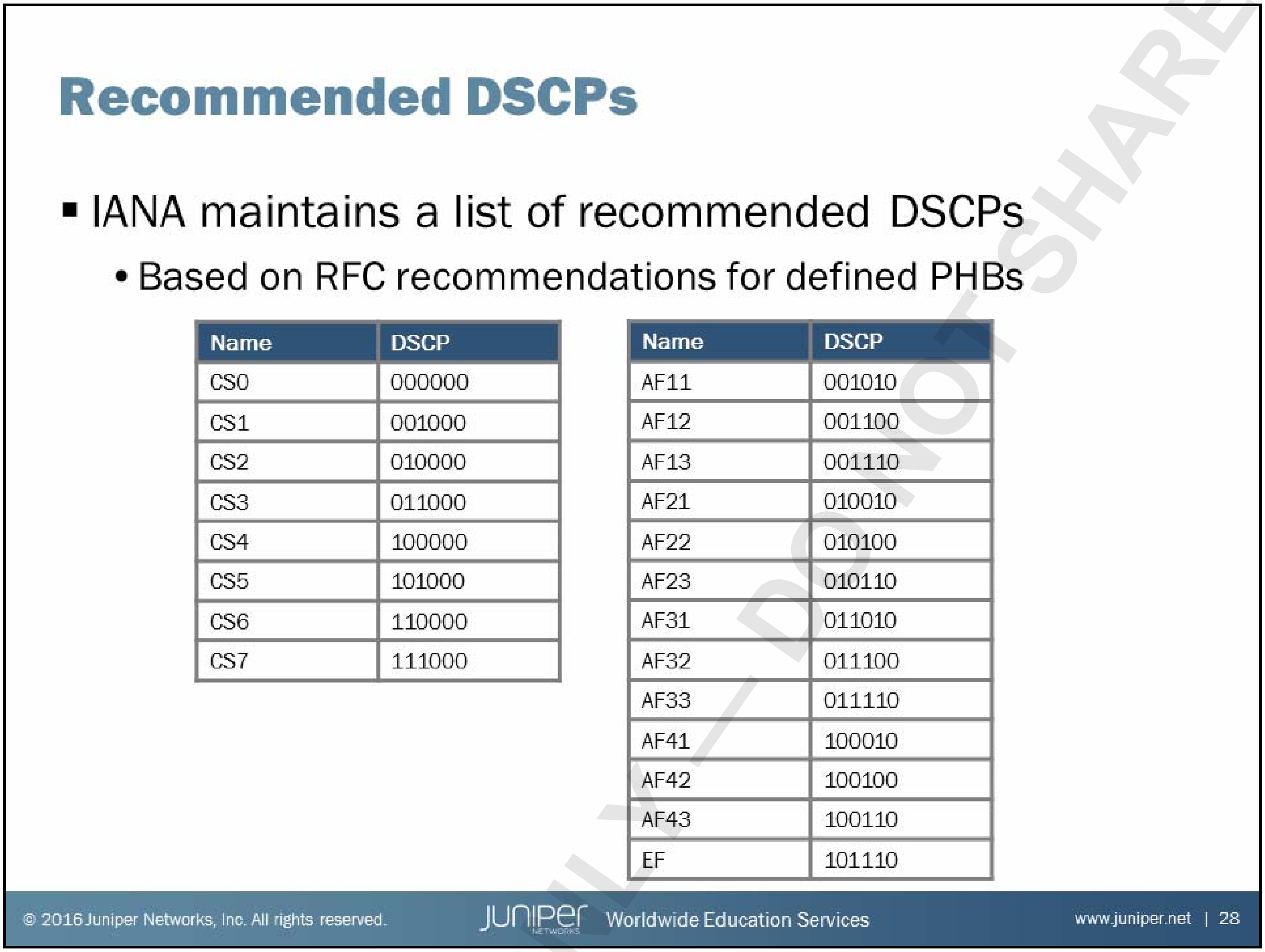
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 27

Class Selector Code Points

As mentioned earlier, a set of CS code points are designated to provide backward compatibility with the historic use of the IP precedence field. While not mandated, the CS code points are normally used to support the network control class, because historically, IP precedence was used to minimize packet drops for control traffic. RFC 2474 defines the CS code space.

Best Effort Is Undefined

A PHB for the best-effort (BE) class is not defined, because the PHB for the BE class, or for traffic that cannot be classified, should equate to conventional (that is, no CoS) packet handling as defined in RFC 1812.



Recommended DSCPs

- IANA maintains a list of recommended DSCPs
  - Based on RFC recommendations for defined PHBs

Name	DSCP
CS0	000000
CS1	001000
CS2	010000
CS3	011000
CS4	100000
CS5	101000
CS6	110000
CS7	111000

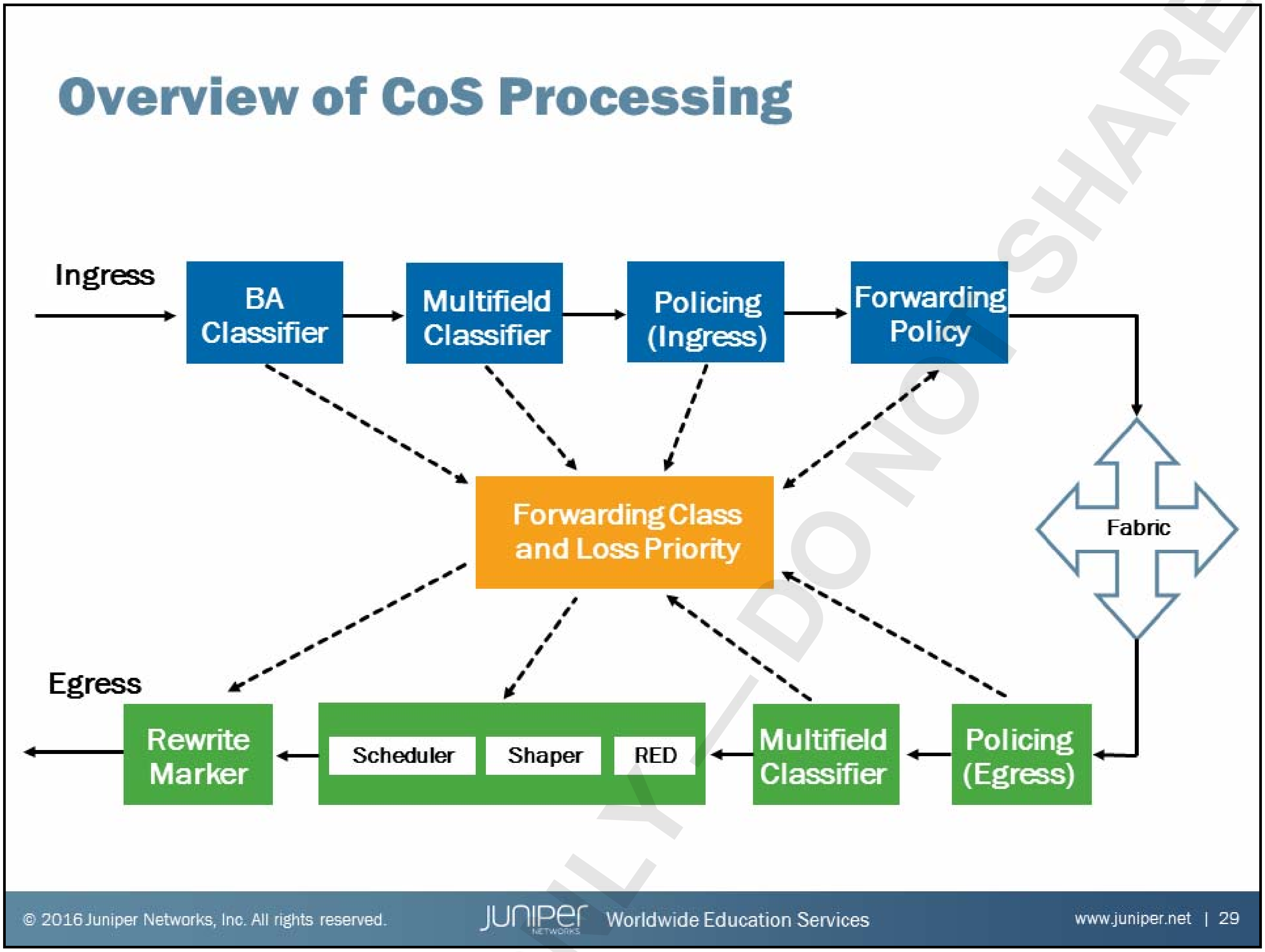
Name	DSCP
AF11	001010
AF12	001100
AF13	001110
AF21	010010
AF22	010100
AF23	010110
AF31	011010
AF32	011100
AF33	011110
AF41	100010
AF42	100100
AF43	100110
EF	101110

Juniper Networks Worldwide Education Services

www.juniper.net | 28

Recommended DSCPs

The specifications that define the known PHBs include one or more recommended DSCPs that identify the PHBs. The Internet Assigned Numbers Authority (IANA) maintains a list of recommended DSCP values at: http://www.iana.org/assignments/dscp-registry. The slide shows these recommended values along with the associated PHB specification. Note that these values are only recommendations; the actual DSCP-to-forwarding-class mappings used within a domain are left to that domain's administrative authority.



Overview of CoS Processing

Ingress

BA Classifier

Multifield Classifier

Policing (Ingress)

Forwarding Policy

Forwarding Class and Loss Priority

Fabric

Egress

Rewrite Marker

Scheduler

Shaper

RED

Multifield Classifier

Policing (Egress)

JUNIPer NETWORKS Worldwide Education Services

www.juniper.net | 29

CoS Processing on Junos Devices

The slide displays the primary CoS processing stages in Juniper Networks M Series Multiservice Edge Routers, T Series Core Routers, and MX Series 3D Universal Edge Routers. The stages are as follows:

- Code point classifier: The first CoS processing stage occurs at ingress when traffic is classified according to a BA code point value in the form of IP precedence, DSCPs, MPLS EXP bits, or IEEE 802.1P priority values.
- Multifield classifier: This processing stage provides multifield classification capabilities based on a firewall filter. The net result of traffic classification is the association of a forwarding class and loss-priority value for a particular packet.
- Policing: Ingress policing limits the amount of traffic that can ingress the router, while egress policing shapes and limits the traffic volume that leaves the router. In most cases, ingress policing is deployed only on customer-facing edge routers. Policers can alter the packet’s forwarding class or loss-priority settings when the policer’s traffic profile is exceeded.
- Forwarding policy: Junos policy can alter the forwarding next hop for a particular packet based on its associated forwarding class and route-filter type match criteria. This capability enables class-of-service-based forwarding (CBF).

Continued on the next page.

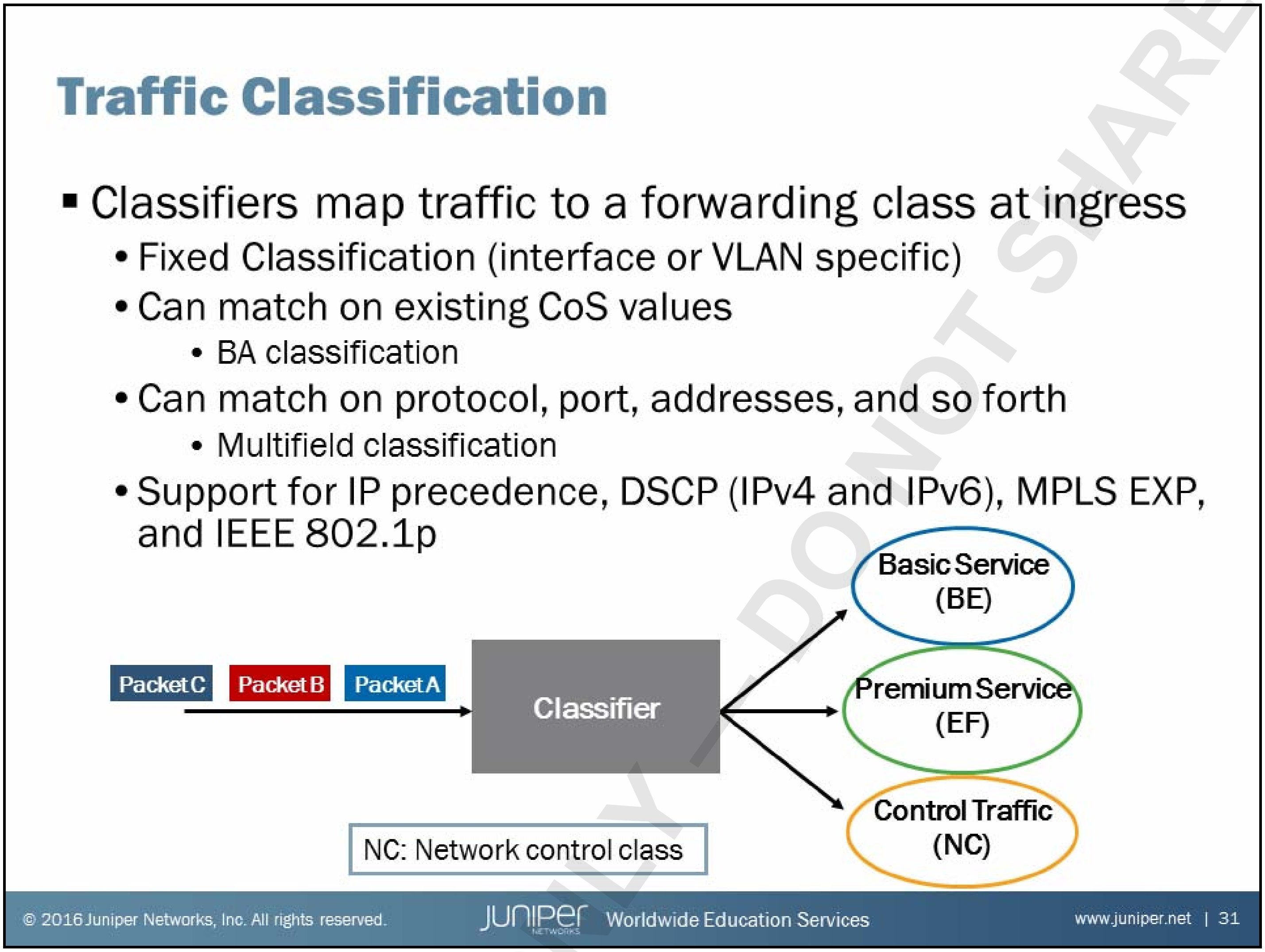
CoS Processing on Junos Devices (contd.)

- Egress policing: After the route lookup, a packet begins its journey toward the selected egress interface. The first egress CoS processing state is output policing, which is again based on either a firewall or an interface-level policer. Once again, excess traffic can be discarded or marked with a loss priority for later discard in the event of congestion.
- Egress multifield classifier: This processing stage provides multifield classification capabilities based on a firewall filter applied to outgoing traffic. The net result of traffic classification is the association of a forwarding class and loss-priority value for a particular outbound packet.
- Scheduling and weighted random early detection (WRED): Schedulers are used to service the queues associated with each forwarding class. Schedulers make use of weighted round-robin (WRR) techniques to service each queue based on priority level. Congestion avoidance through a WRED mechanism is also performed at this stage.
- Rewrite marker: The final CoS stage involves rewriting CoS fields in the packet header so that the next node can act solely based on exiting CoS values.

The middle box (forwarding class and loss priority) represents two data values that can either be inputs to or outputs of the process components. The arrows with the dotted lines indicate inputs and outputs (or settings and actions based on settings). For example, the multifield classifier sets the forwarding class and loss priority of incoming packets. This means that the forwarding class and loss priority are outputs of the classifier; thus, the arrow points away from the classifier. The scheduler receives the forwarding class and loss priority settings, and queues the outgoing packet based on those settings. This means that the forwarding class and loss priority are inputs to the scheduler; thus, the arrow points to the scheduler.

The CoS processing described on the slide represents the concept of egress CoS processing, which is similarly supported on most Junos OS devices. With the appropriate hardware installed, an MX Series 3D Universal Edge Router can also perform ingress CoS processing. Ingress CoS processing is not covered in this course.

Typically, only a combination of some components (not all) is used to define a CoS service offering.



Traffic Classification

- Classifiers map traffic to a forwarding class at ingress
  - Fixed Classification (interface or VLAN specific)
  - Can match on existing CoS values
    - BA classification
  - Can match on protocol, port, addresses, and so forth
    - Multifield classification
  - Support for IP precedence, DSCP (IPv4 and IPv6), MPLS EXP, and IEEE 802.1p

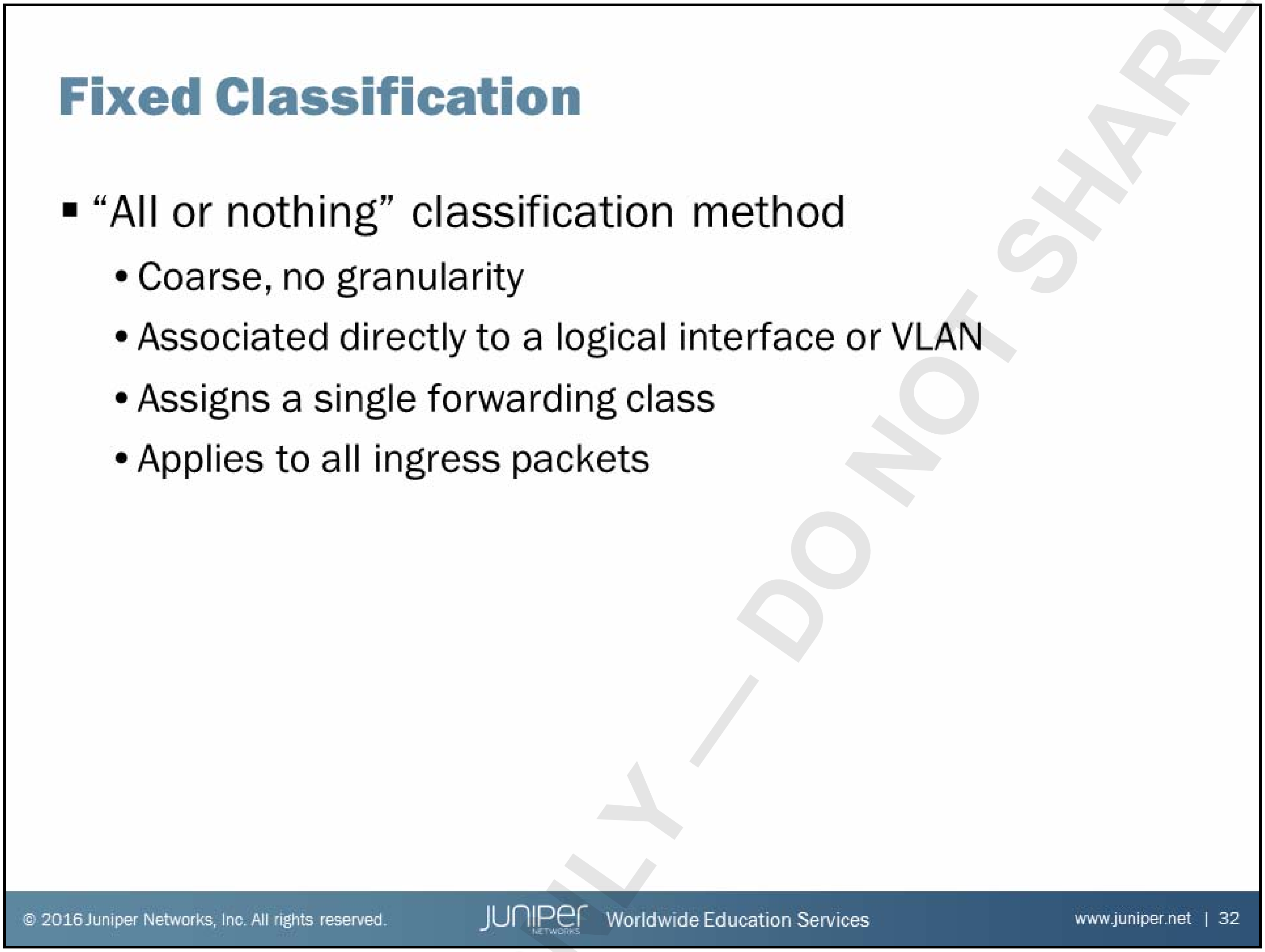
Packet C
Packet B
Packet A
Classifier
Basic Service (BE)
Premium Service (EF)
Control Traffic (NC)
NC: Network control class

Worldwide Education Services

Classifiers Map Traffic to a Forwarding Class

As traffic arrives at the device, it is classified as belonging to one of the forwarding classes defined on that device. The simplest way to classify incoming packets is to use fixed classification. With fixed classification, you assign a single forwarding class to a logical interface or VLAN. The device assigns all traffic arriving at that interface to the defined forwarding class, and by extension to the related queue.

The device can also match traffic based on existing CoS values using BA classification, or it can match on a variety of fields in a packet's header (IP address, protocol, port, and so forth) using multifield classification. Junos classifiers support a variety of protocol types, as shown on the slide.



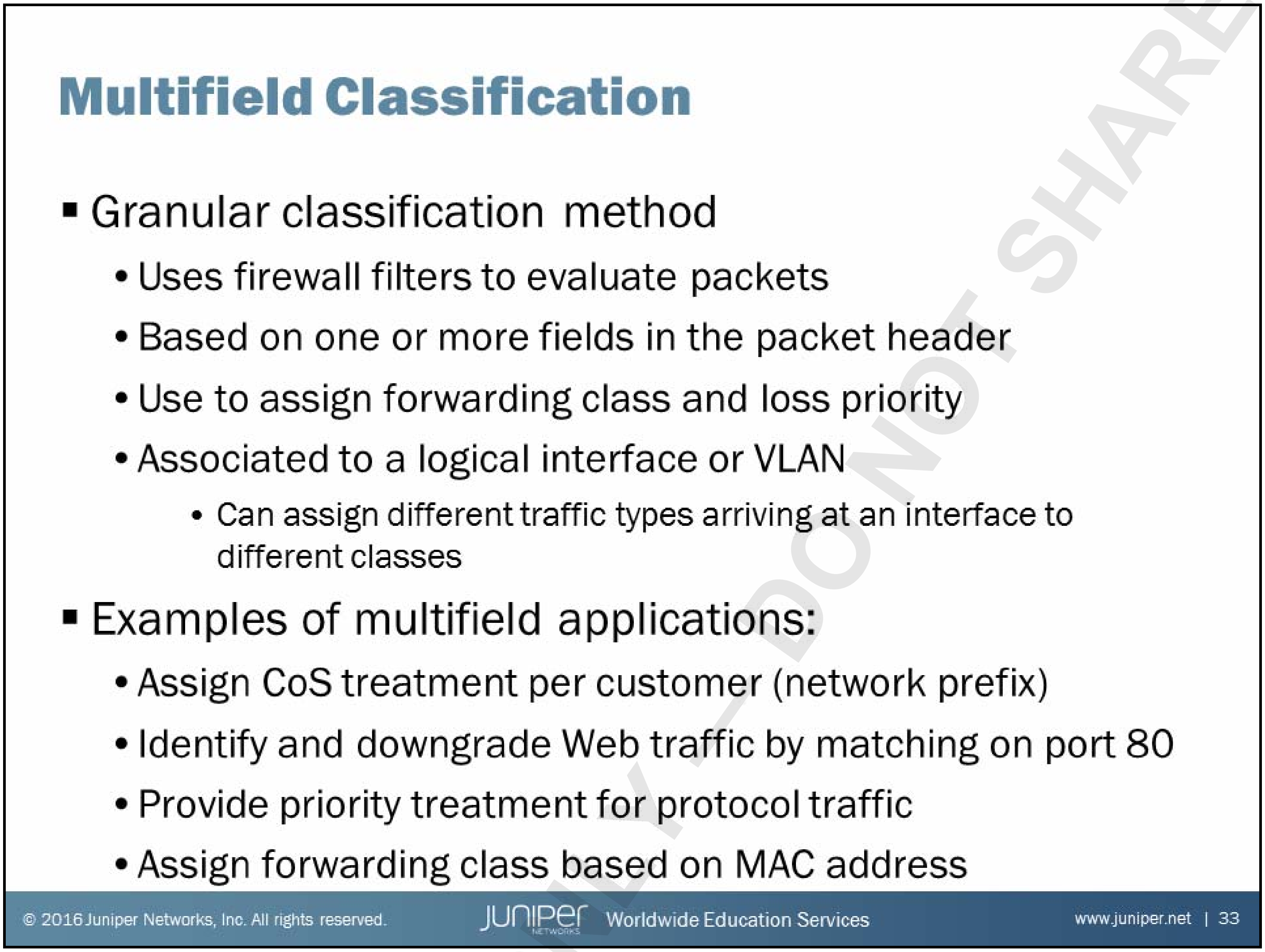
Fixed Classification

- “All or nothing” classification method
  - Coarse, no granularity
  - Associated directly to a logical interface or VLAN
  - Assigns a single forwarding class
  - Applies to all ingress packets

JUNIPER NETWORKS
Worldwide Education Services
www.juniper.net | 32

Fixed Classification—“All or Nothing” Method

As mentioned previously, fixed classification is the simplest way to classify incoming packets. With fixed classification, you assign a single forwarding class to a logical interface or VLAN. The device assigns all traffic arriving at that interface to the defined forwarding class, and by extension to the related queue. Fixed classification can be a good approach when you specifically want to assign all inbound traffic from a neighbor to a specific forwarding class and queue. For example, perhaps you have a customer attached to a given port, and all of that customer’s traffic should be treated in a specific way. Fixed classification provides the easiest way to complete this task. While fixed classification is simple and efficient, it has no granularity. If you require any differentiation of traffic ingressing the interface, fixed classification will not meet your needs.



Multifield Classification

- Granular classification method
  - Uses firewall filters to evaluate packets
  - Based on one or more fields in the packet header
  - Use to assign forwarding class and loss priority
  - Associated to a logical interface or VLAN
    - Can assign different traffic types arriving at an interface to different classes
- Examples of multifield applications:
  - Assign CoS treatment per customer (network prefix)
  - Identify and downgrade Web traffic by matching on port 80
  - Provide priority treatment for protocol traffic
  - Assign forwarding class based on MAC address

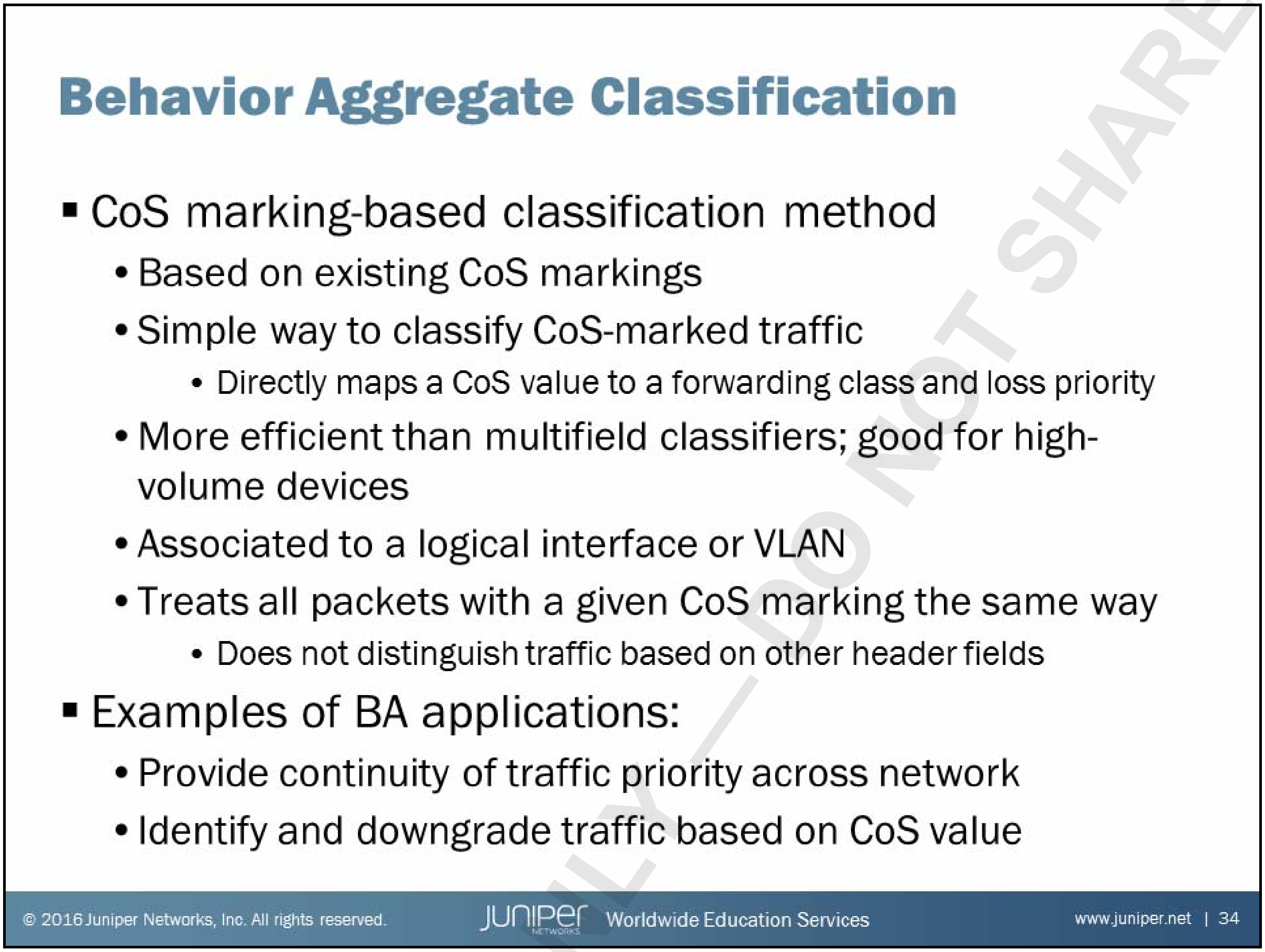
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 33

Multifield Classification—Granular Method

When you need granular control to apply CoS values to inbound traffic, use MF classification. MF classification allows you to use standard firewall filters in order to match against a variety of fields in a packet header—a source IP address, for example. When packets arriving on a given logical interface or VLAN match against the desired parameters, the device can assign a specific forwarding class and PLP value.

Multifield Application Examples

The slide provides some examples of where multifield classification can be helpful.



Behavior Aggregate Classification

- CoS marking-based classification method
  - Based on existing CoS markings
  - Simple way to classify CoS-marked traffic
    - Directly maps a CoS value to a forwarding class and loss priority
  - More efficient than multifield classifiers; good for high-volume devices
  - Associated to a logical interface or VLAN
  - Treats all packets with a given CoS marking the same way
    - Does not distinguish traffic based on other header fields
- Examples of BA applications:
  - Provide continuity of traffic priority across network
  - Identify and downgrade traffic based on CoS value

Worldwide Education Services www.juniper.net | 34

Behavior Aggregate Classification—CoS Marking-Based Method

When traffic coming from a neighboring node already has CoS markings, you can use BA classification. BA classification can be applied per logical interface or VLAN, and it provides a simple way to directly map a marked packet to a forwarding class and PLP value. BA classification is more efficient than multifield classification, because it requires less packet analysis. The efficiency benefit makes BA classification a good choice for devices with high traffic volumes, such as routers in a network core.

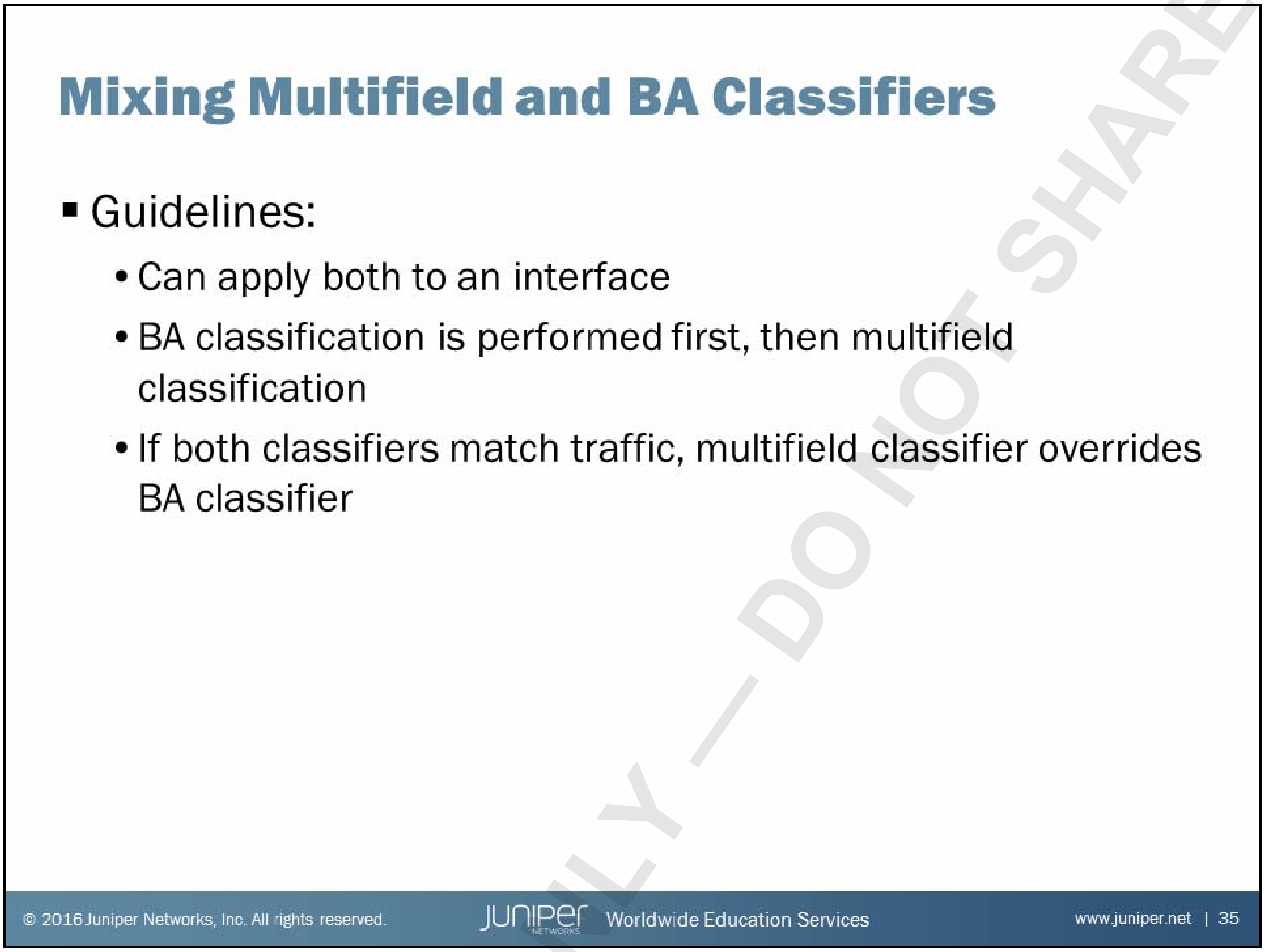
BA classification is based entirely on existing CoS markings. It treats all traffic with a given CoS value in the same way; that is, the Junos device assigns all inbound traffic with a given CoS marking to the same forwarding class and queue.

BA classifiers can match against the following incoming CoS markings:

- IPv4 DSCP
- IPv6 DSCP
- IP precedence bits
- MPLS EXP bits
- IEEE 802.1p CoS bits
- IEEE 802.1ad drop eligible indicator (DEI) bit

BA Application Examples

The slide provides some examples of where BA classification can be helpful.



Mixing Multifield and BA Classifiers

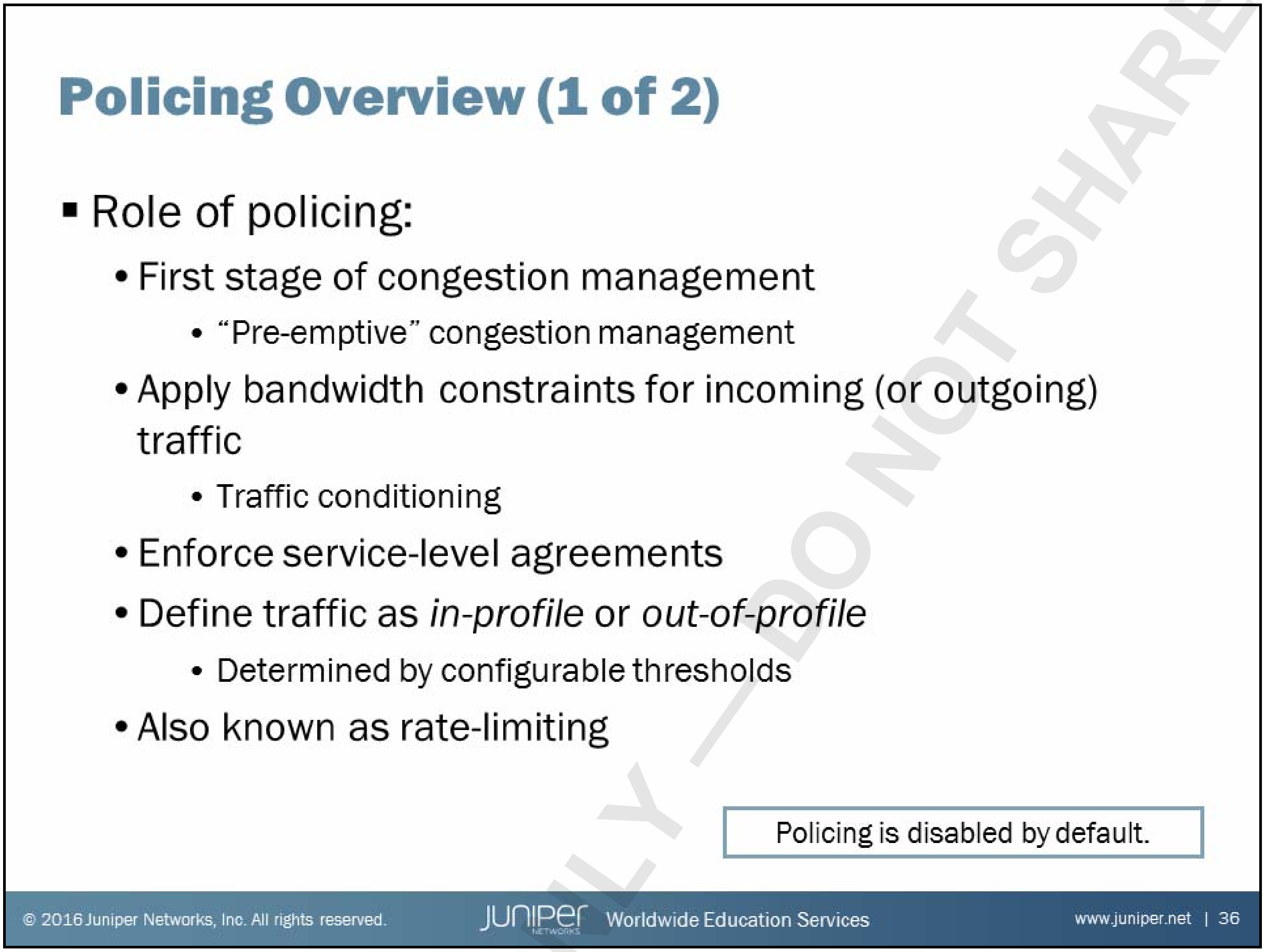
- Guidelines:
  - Can apply both to an interface
  - BA classification is performed first, then multifield classification
  - If both classifiers match traffic, multifield classifier overrides BA classifier

Juniper Networks Worldwide Education Services

www.juniper.net | 35

Guidelines When Mixing Multifield and BA Classifiers

You can apply both multifield and BA classifiers to a logical interface. Because BA classification is performed before multifield classification, the latter overrides the former if a conflict occurs.



Policing Overview (1 of 2)

- Role of policing:
  - First stage of congestion management
    - “Pre-emptive” congestion management
  - Apply bandwidth constraints for incoming (or outgoing) traffic
    - Traffic conditioning
  - Enforce service-level agreements
  - Define traffic as *in-profile* or *out-of-profile*
    - Determined by configurable thresholds
  - Also known as rate-limiting

Policing is disabled by default.

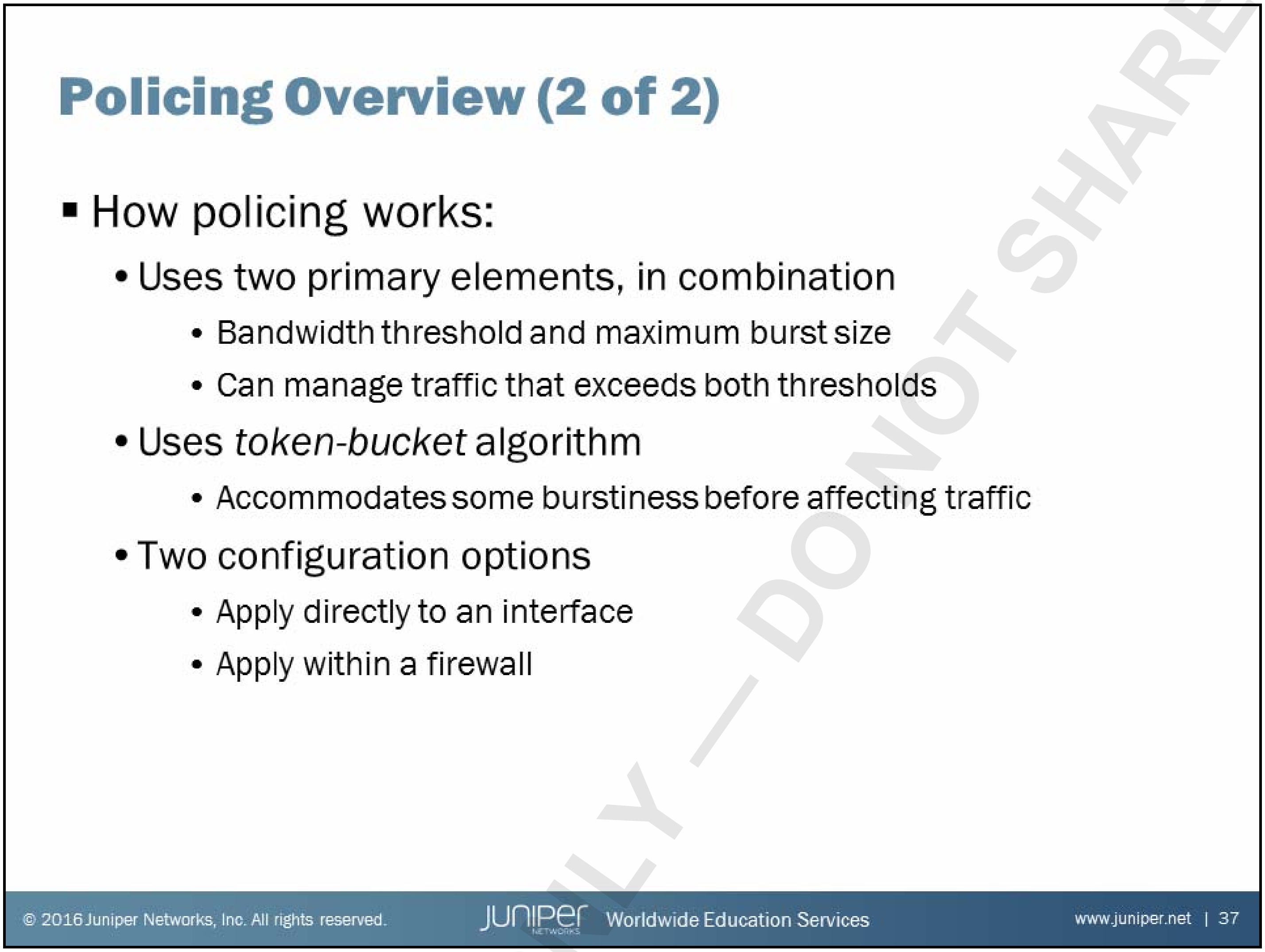
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 36

Role of Policing

The primary function of policing is to provide a first stage of congestion management. With policing, you can condition incoming or outgoing traffic by applying bandwidth constraints. Policing is an excellent way to control the amount of traffic entering your network, and allows you to enforce service-level agreements (SLAs).

Policers provide configurable thresholds that allow you to create two categories of traffic. Traffic that is within the threshold limits is referred to as in-profile; traffic that exceeds the threshold limits is referred to as out-of-profile.

Policing is sometimes also known as rate-limiting.



Policing Overview (2 of 2)

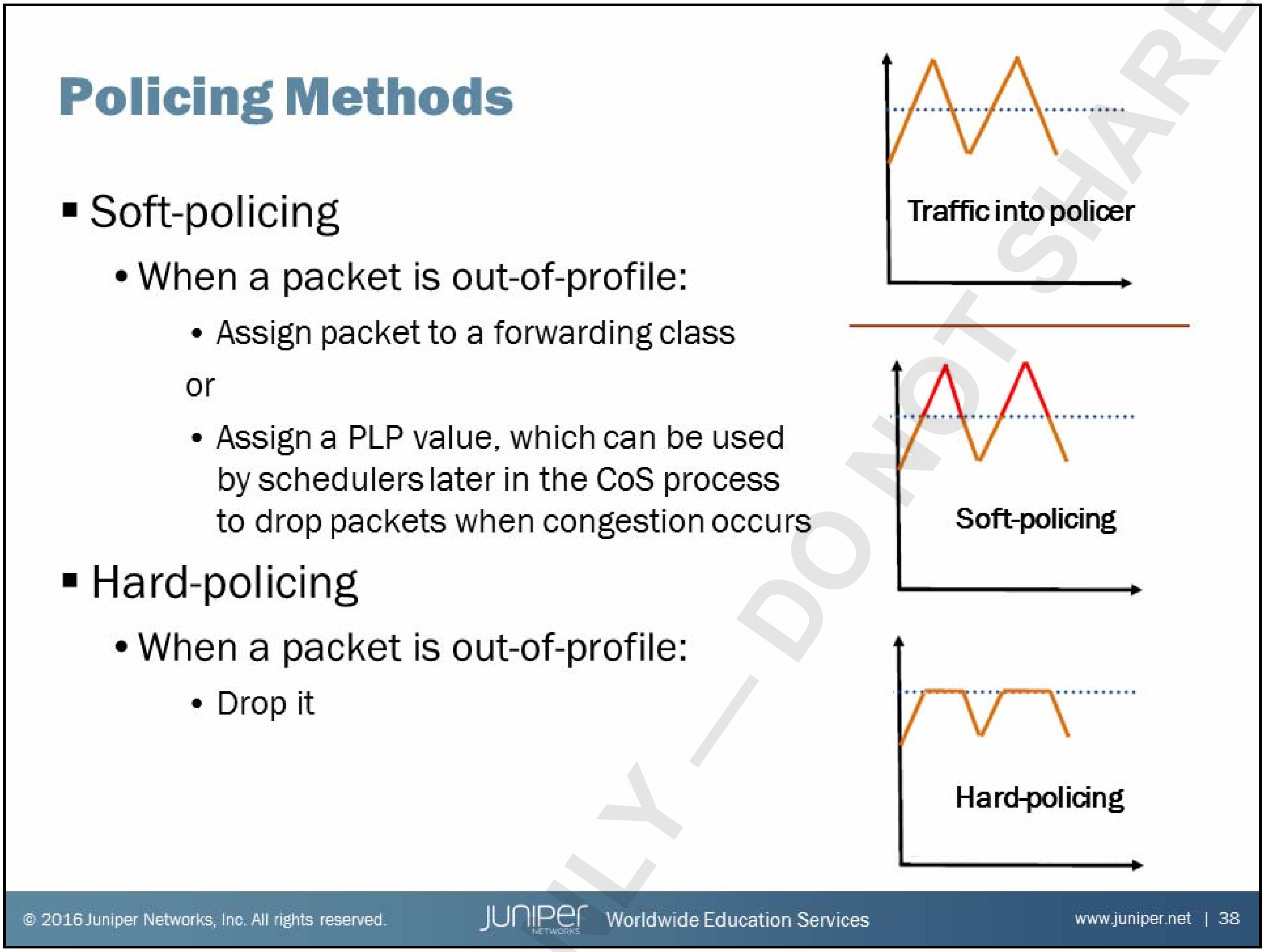
- How policing works:
  - Uses two primary elements, in combination
    - Bandwidth threshold and maximum burst size
    - Can manage traffic that exceeds both thresholds
  - Uses *token-bucket* algorithm
    - Accommodates some burstiness before affecting traffic
  - Two configuration options
    - Apply directly to an interface
    - Apply within a firewall

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 37

How Policing Works

Policing is based on a token-bucket algorithm. Unlike a leaky-bucket algorithm, the token-bucket method allows for a certain amount of burstiness before taking action on traffic that exceeds configured thresholds. Policers on Junos devices generally consist of two primary elements: a bandwidth threshold and a maximum burst size. The bandwidth threshold defines the acceptable rate for a given flow of traffic. When traffic exceeds this rate, it is also allowed to burst beyond the threshold for a certain period of time. When a sustained burst of traffic exceeds both thresholds, the device takes action on the out-of-profile traffic.

Two general configuration options exist for policers on Junos devices. You can create a policer and apply it directly to an interface, or you create a policer, use it as an action within a firewall filter, and apply the filter to an interface.



Policing Methods

- Soft-policing
  - When a packet is out-of-profile:
    - Assign packet to a forwarding class

    or

    - Assign a PLP value, which can be used by schedulers later in the CoS process to drop packets when congestion occurs
- Hard-policing
  - When a packet is out-of-profile:
    - Drop it

Traffic into policer

Soft-policing

Hard-policing

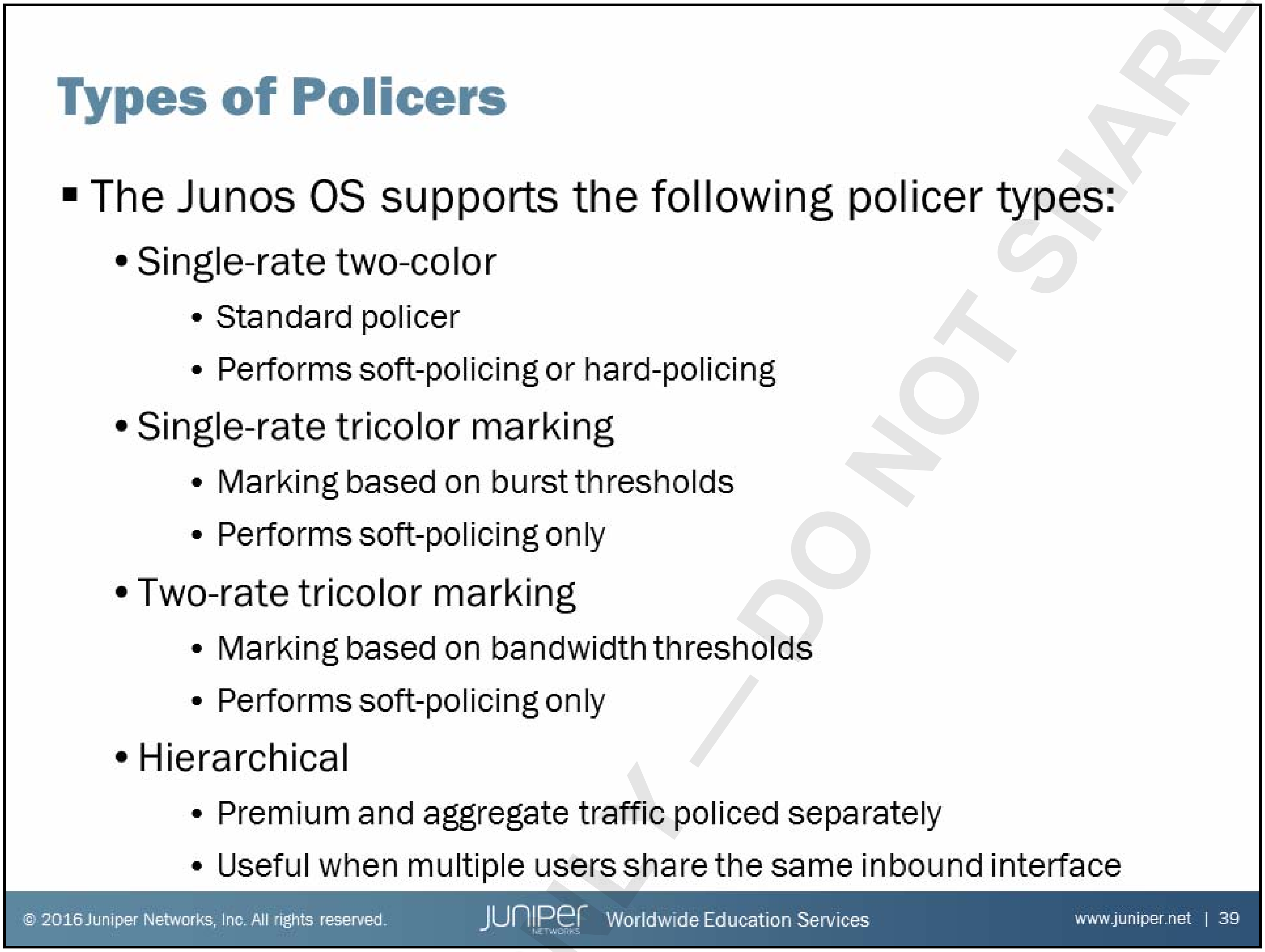
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 38

Soft-Policing

You can configure the Junos OS to take action on packets that are out-of-profile. One option is to allow the traffic to enter the device and assign it to a specific (usually less preferable) forwarding class. Another option is to allow the traffic and give it a specific (usually less preferable) PLP value. These are examples of soft-policing, because they simply mark and reclassify the packet.

Hard-Policing

A more harsh policing option is to drop any traffic that exceeds the configured thresholds. This approach is an example of hard-policing.

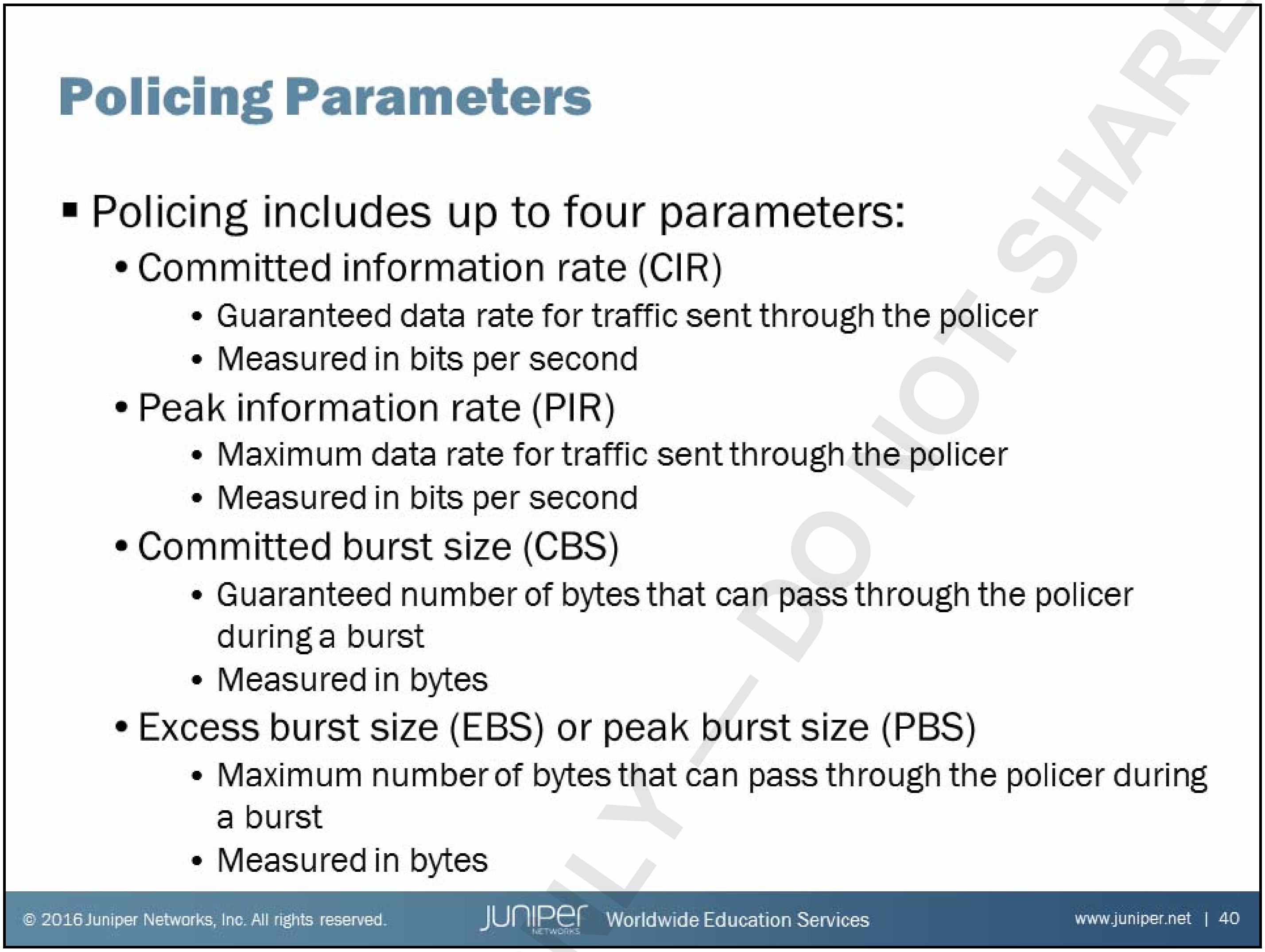


Types of Policers

- The Junos OS supports the following policer types:
  - Single-rate two-color
    - Standard policer
    - Performs soft-policing or hard-policing
  - Single-rate tricolor marking
    - Marking based on burst thresholds
    - Performs soft-policing only
  - Two-rate tricolor marking
    - Marking based on bandwidth thresholds
    - Performs soft-policing only
  - Hierarchical
    - Premium and aggregate traffic policed separately
    - Useful when multiple users share the same inbound interface

The Junos OS Supports Several Policer Types

The slide lists the policer types supported by Junos devices. Note that not all Junos devices support tricolor marking.

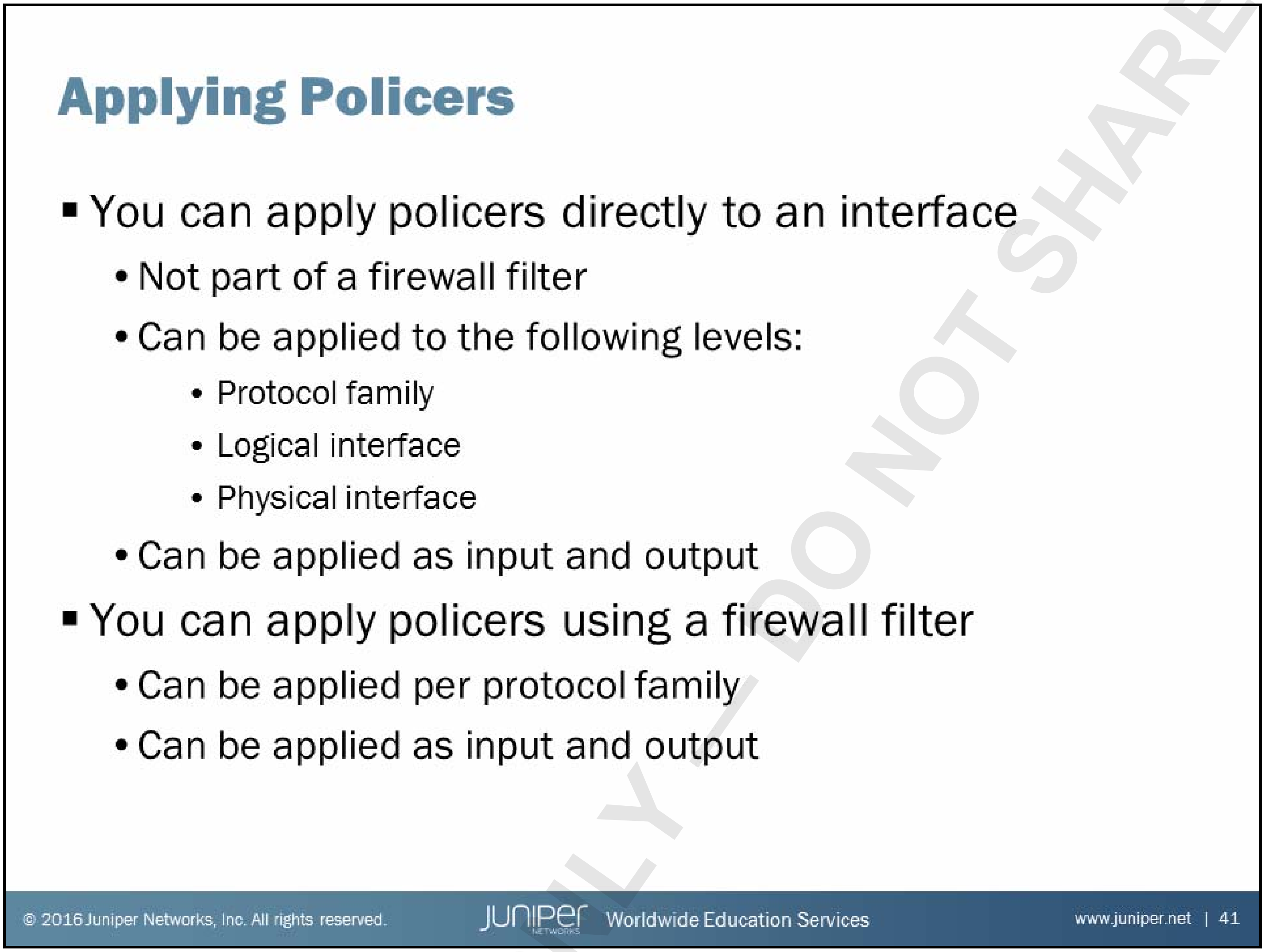


Policing Parameters

- Policing includes up to four parameters:
  - Committed information rate (CIR)
    - Guaranteed data rate for traffic sent through the policer
    - Measured in bits per second
  - Peak information rate (PIR)
    - Maximum data rate for traffic sent through the policer
    - Measured in bits per second
  - Committed burst size (CBS)
    - Guaranteed number of bytes that can pass through the policer during a burst
    - Measured in bytes
  - Excess burst size (EBS) or peak burst size (PBS)
    - Maximum number of bytes that can pass through the policer during a burst
    - Measured in bytes

Four Parameters for Policing

Policing involved up to four parameters, depending on the policer being used. The committed information rate (CIR) and peak information rate (PIR) relate to bandwidth thresholds, measured in bits per second. The committed burst size (CBS), excess burst size (EBS), and peak burst size (PBS) relate to burst thresholds, measured in bytes.



Applying Policers

- You can apply policers directly to an interface
  - Not part of a firewall filter
  - Can be applied to the following levels:
    - Protocol family
    - Logical interface
    - Physical interface
  - Can be applied as input and output
- You can apply policers using a firewall filter
  - Can be applied per protocol family
  - Can be applied as input and output

Worldwide Education Services

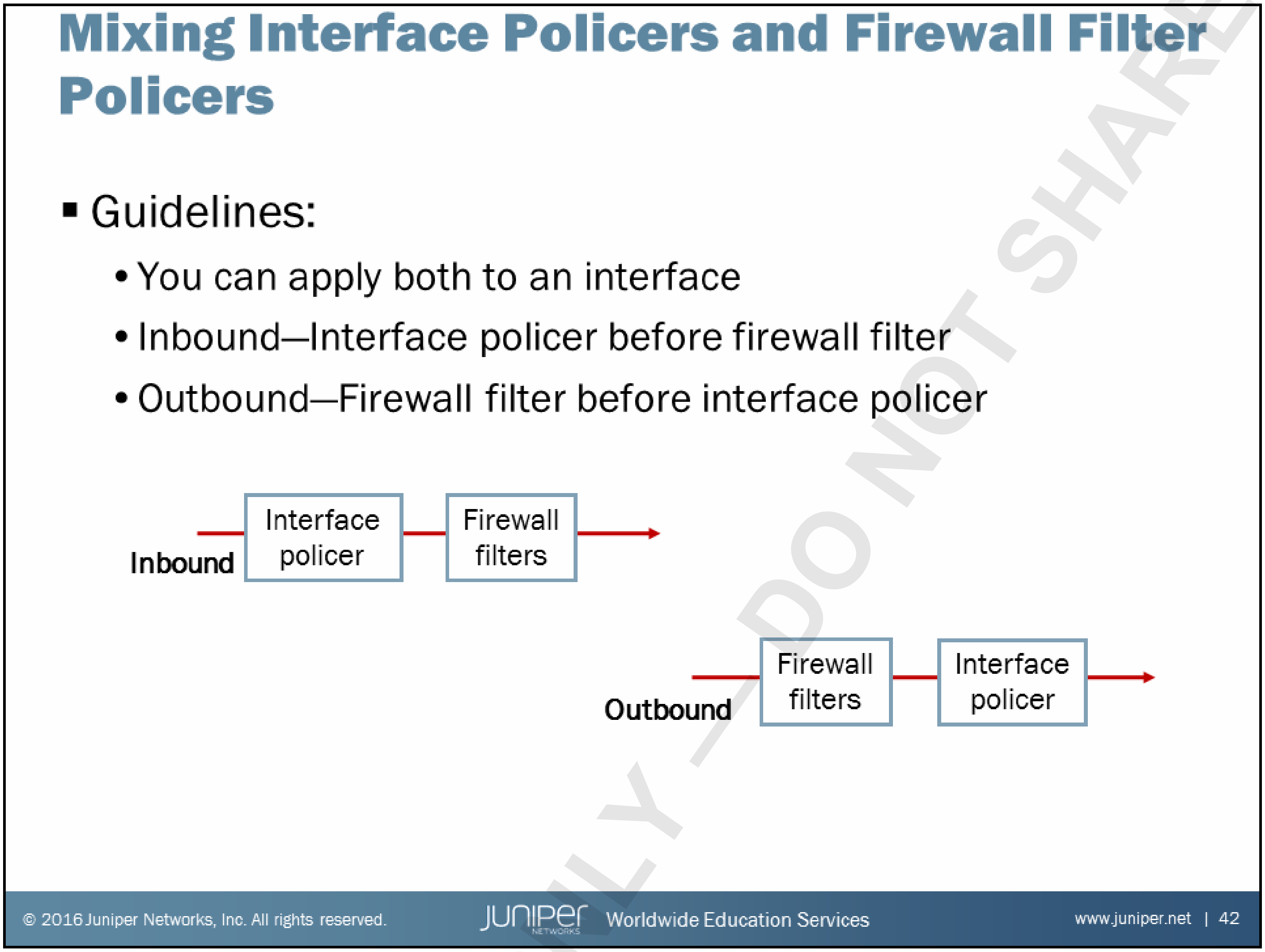
www.juniper.net | 41

Applying Policers Directly to an Interface

One method of using policers is to apply them directly to an interface. You can apply policers per protocol family, logical interface, or physical interface. You can also apply policers in the inbound or outbound direction.

Applying Policers Using a Firewall Filter

Another method of using policers is to reference them within a firewall filter, and then apply the filter to an interface. You can apply policers per protocol family within a filter, and apply the filter to an interface in the inbound or outbound direction.



Mixing Interface Policers and Firewall Filter Policers

- Guidelines:
  - You can apply both to an interface
  - Inbound—Interface policer before firewall filter
  - Outbound—Firewall filter before interface policer

Inbound

Interface policer

Firewall filters

Outbound

Firewall filters

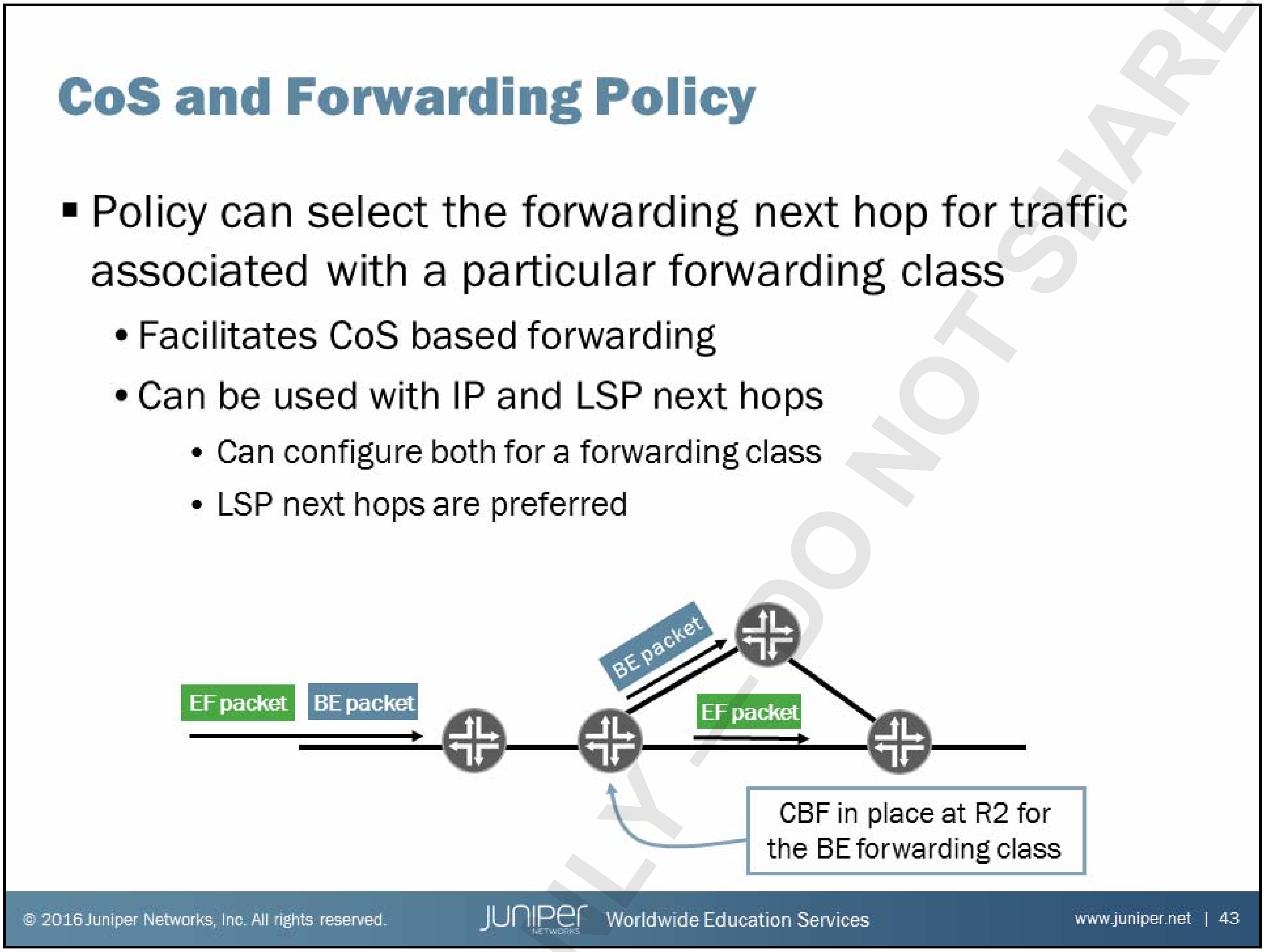
Interface policer

Worldwide Education Services

www.juniper.net | 42

Guidelines for Mixing Policers

You can apply both interface and firewall filter policers to an interface. The system evaluates input interface policers before any input firewall filters, whereas it evaluates output interface policers after any output firewall filters.



CoS and Forwarding Policy

- Policy can select the forwarding next hop for traffic associated with a particular forwarding class
  - Facilitates CoS based forwarding
  - Can be used with IP and LSP next hops
    - Can configure both for a forwarding class
    - LSP next hops are preferred

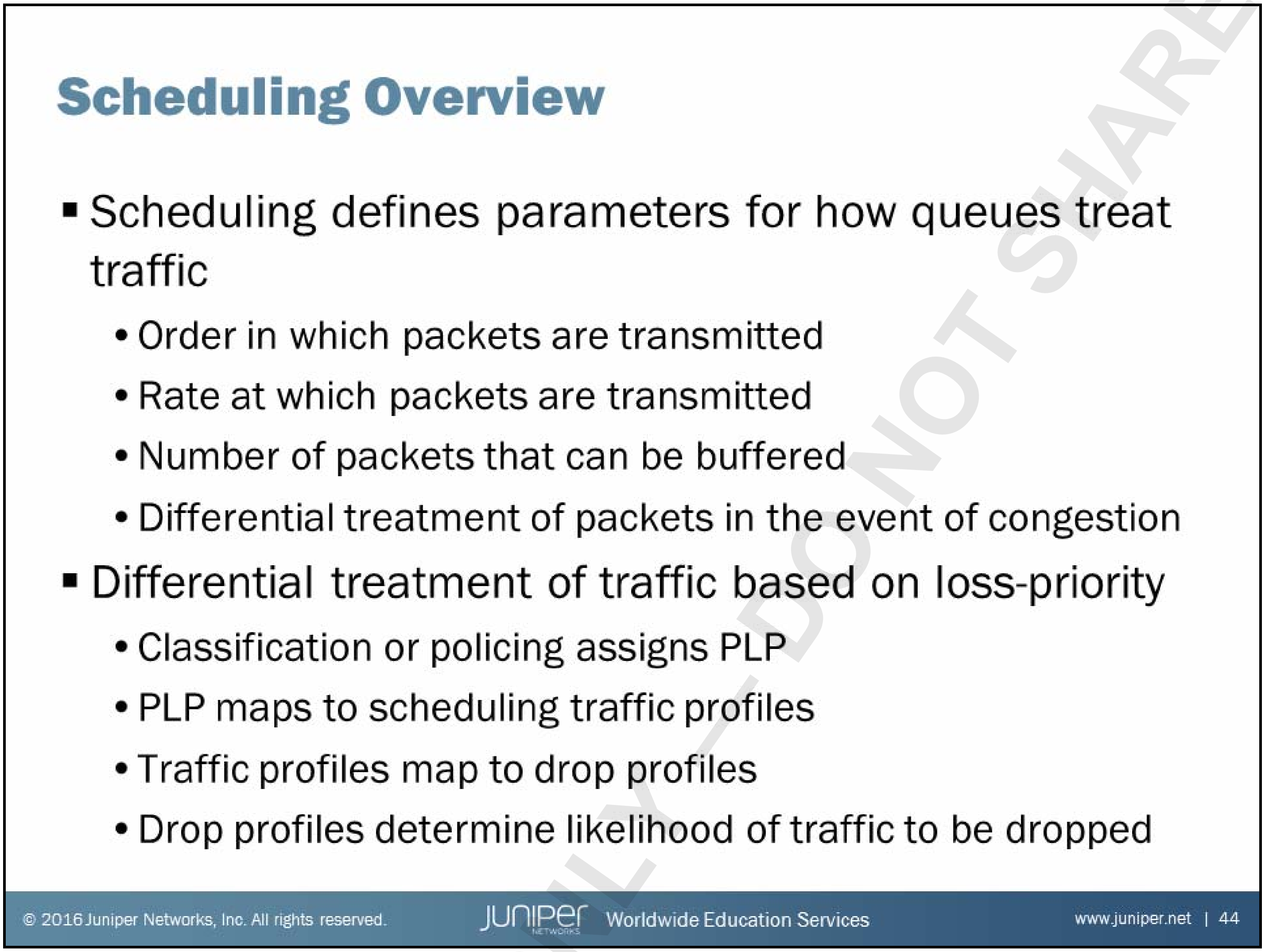
CBF in place at R2 for the BE forwarding class

Juniper Networks Worldwide Education Services

www.juniper.net | 43

CoS and Forwarding Policy

You can use routing policy to affect the forwarding next hop associated with a given forwarding class. The slide shows an example of CoS based forwarding (CBF), where traffic associated with the BE class is directed along a forwarding path that differs from the interior gateway protocol's (IGP's) preferred route to the destination. CBF forwarding classes can be assigned an IP next hop, MPLS next hop, or both. When a forwarding class is identified with both LSP next hops and IP next hops, the LSP next hops are preferred.



Scheduling Overview

- Scheduling defines parameters for how queues treat traffic
  - Order in which packets are transmitted
  - Rate at which packets are transmitted
  - Number of packets that can be buffered
  - Differential treatment of packets in the event of congestion
- Differential treatment of traffic based on loss-priority
  - Classification or policing assigns PLP
  - PLP maps to scheduling traffic profiles
  - Traffic profiles map to drop profiles
  - Drop profiles determine likelihood of traffic to be dropped

Worldwide Education Services

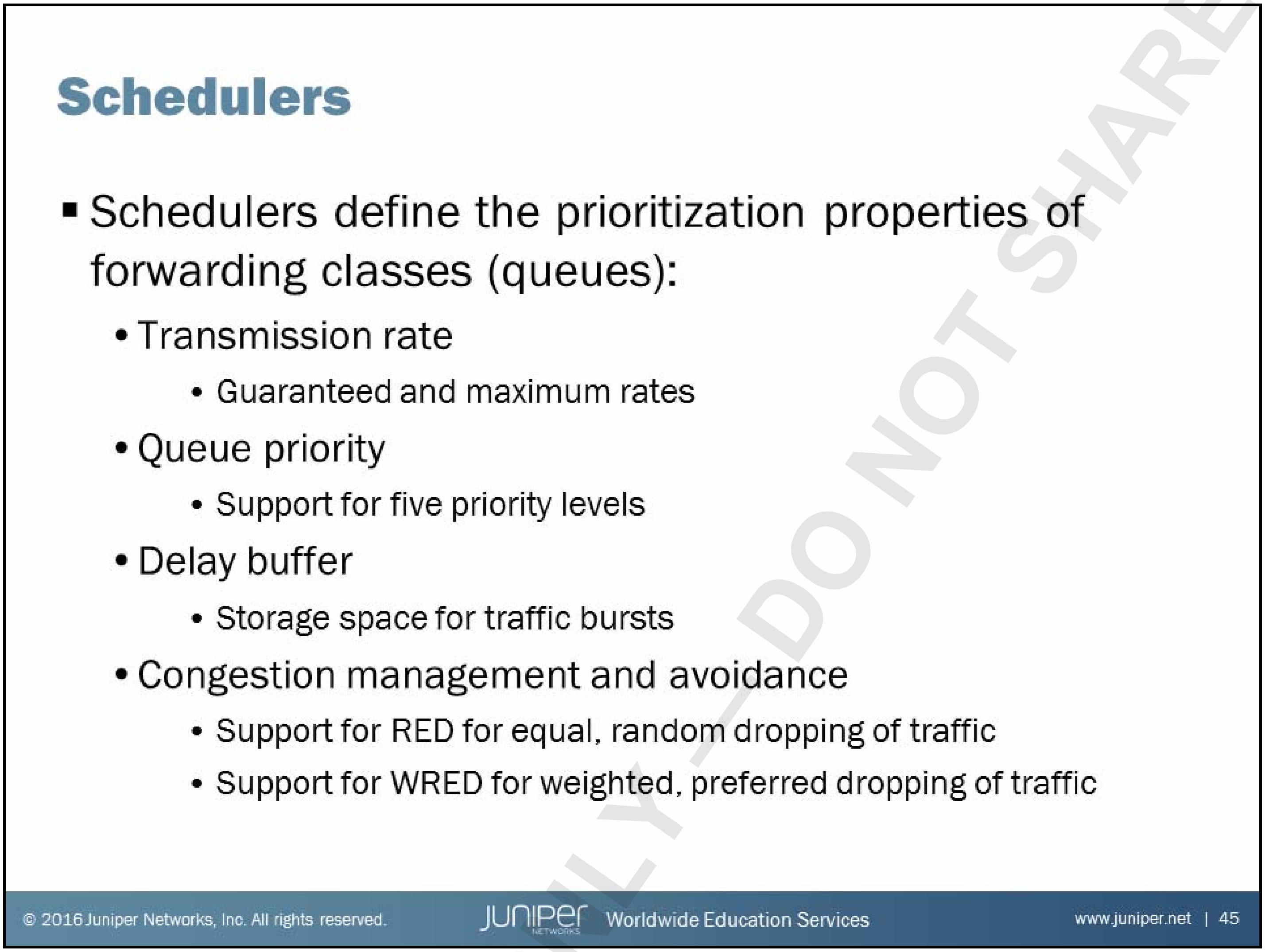
www.juniper.net | 44

Role of Scheduling

The primary function of scheduling is to define the parameters for how queues treat traffic. Scheduling determines the order in which packets are transmitted, the rate at which packets are transmitted, the number of packets the system can buffer, and the differential treatment of packets in the event of congestion.

Differential Treatment of Traffic Based on Loss-Priority

Back at the beginning of the CoS process, classification and policing assigned packet loss priority (PLP) values to packets. Now, at the scheduling stage, you can map the PLP values to scheduling traffic profiles. You can then map the traffic profiles to drop profiles, which determine the likelihood of dropping a specific traffic type. The mapping of PLP values to drop profiles gives PLP meaning and purpose.



Schedulers

- Schedulers define the prioritization properties of forwarding classes (queues):
  - Transmission rate
    - Guaranteed and maximum rates
  - Queue priority
    - Support for five priority levels
  - Delay buffer
    - Storage space for traffic bursts
  - Congestion management and avoidance
    - Support for RED for equal, random dropping of traffic
    - Support for WRED for weighted, preferred dropping of traffic

Worldwide Education Services

www.juniper.net | 45

Schedulers Define the Prioritization Properties of Forwarding Classes

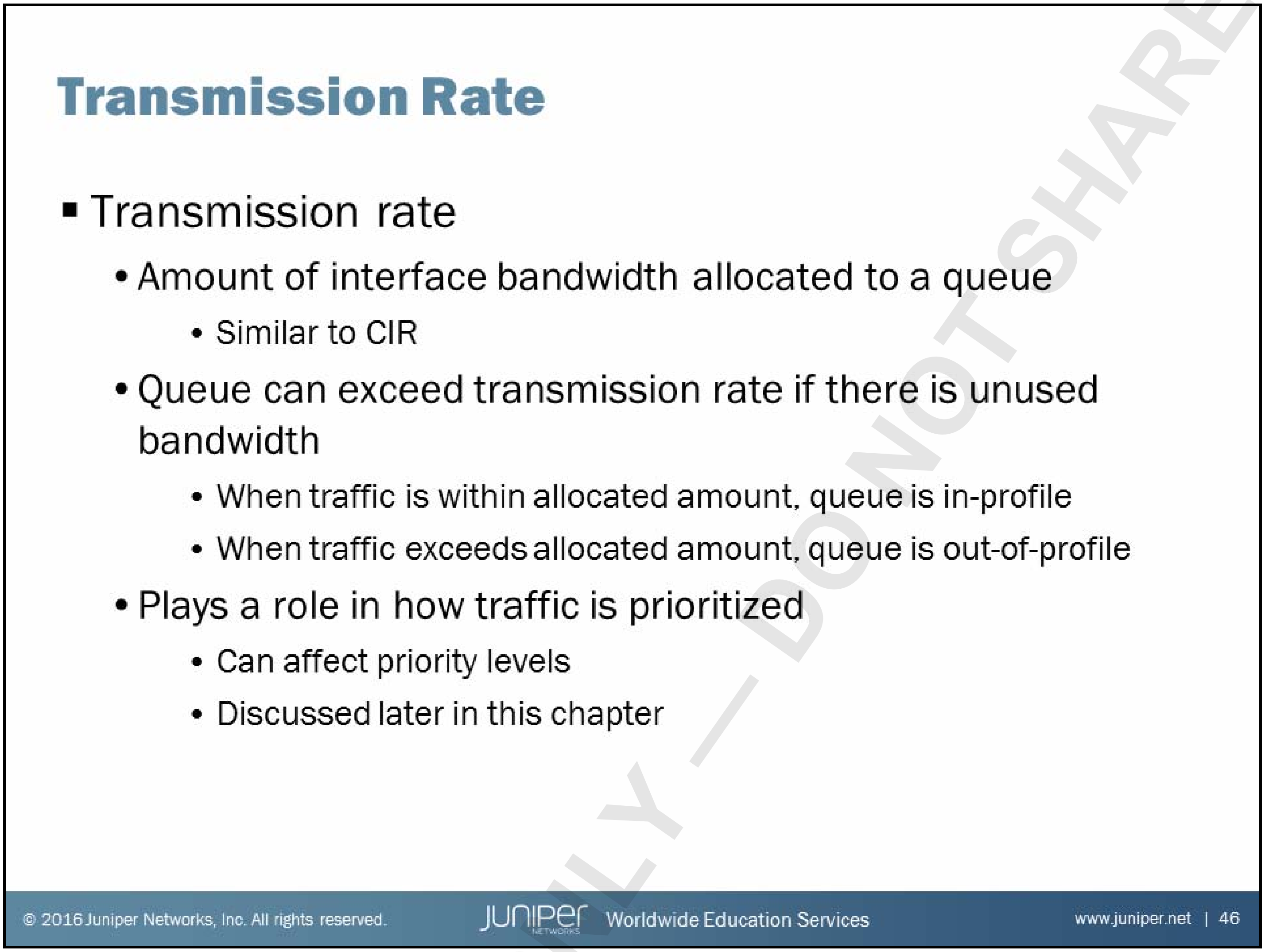
A scheduler defines a set of parameters, including transmission rate, queue priority, delay buffers, and congestion management and avoidance, for a specific forwarding class.

You measure the transmission rate in either bits per second or as a percentage of interface speed.

You can specify a priority setting that influences how the WRR algorithm services the queue for that forwarding class. In other words, the device services a high-priority queue before a low-priority queue.

You can set the buffer depth using either a percentage or a maximum temporal value.

The random early detection (RED) algorithm works to control congestion by performing packet drops when congestion is detected. WRED algorithms can factor criteria such as traffic type or loss priority into the discard decision. The goal of congestion avoidance is to prevent global synchronization of TCP sessions, a condition where multiple sources begin retransmitting and backing off in unison, which in turn leads to oscillations of either too much or too little data.



Transmission Rate

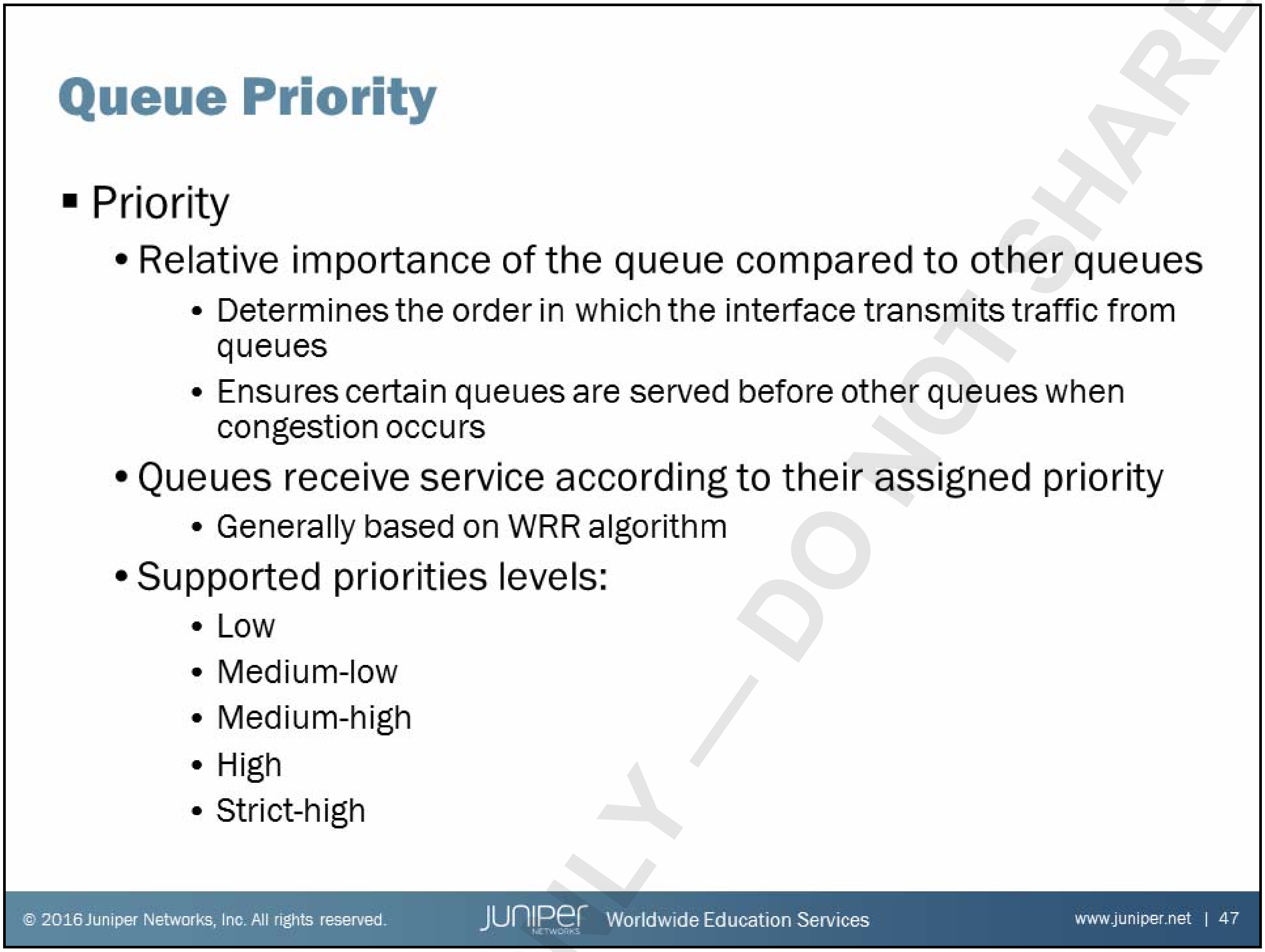
- Transmission rate
  - Amount of interface bandwidth allocated to a queue
    - Similar to CIR
  - Queue can exceed transmission rate if there is unused bandwidth
    - When traffic is within allocated amount, queue is in-profile
    - When traffic exceeds allocated amount, queue is out-of-profile
  - Plays a role in how traffic is prioritized
    - Can affect priority levels
    - Discussed later in this chapter

Juniper Networks Worldwide Education Services www.juniper.net | 46

Transmission Rate

The amount of bandwidth allocated to a queue is called its transmission rate. You can think of a queue's transmission rate as a committed information rate (CIR).

By default, a queue can exceed its configured transmission rate if unused interface bandwidth is available. When traffic is flowing within a queue's transmission rate, the traffic is referred to as "in-profile"; when the traffic exceeds the allocated amount, the traffic is referred to as "out-of-profile." Transmission rate plays a role in how a device transmits traffic out of an interface. Specifically, it can affect the precedence of priority levels.



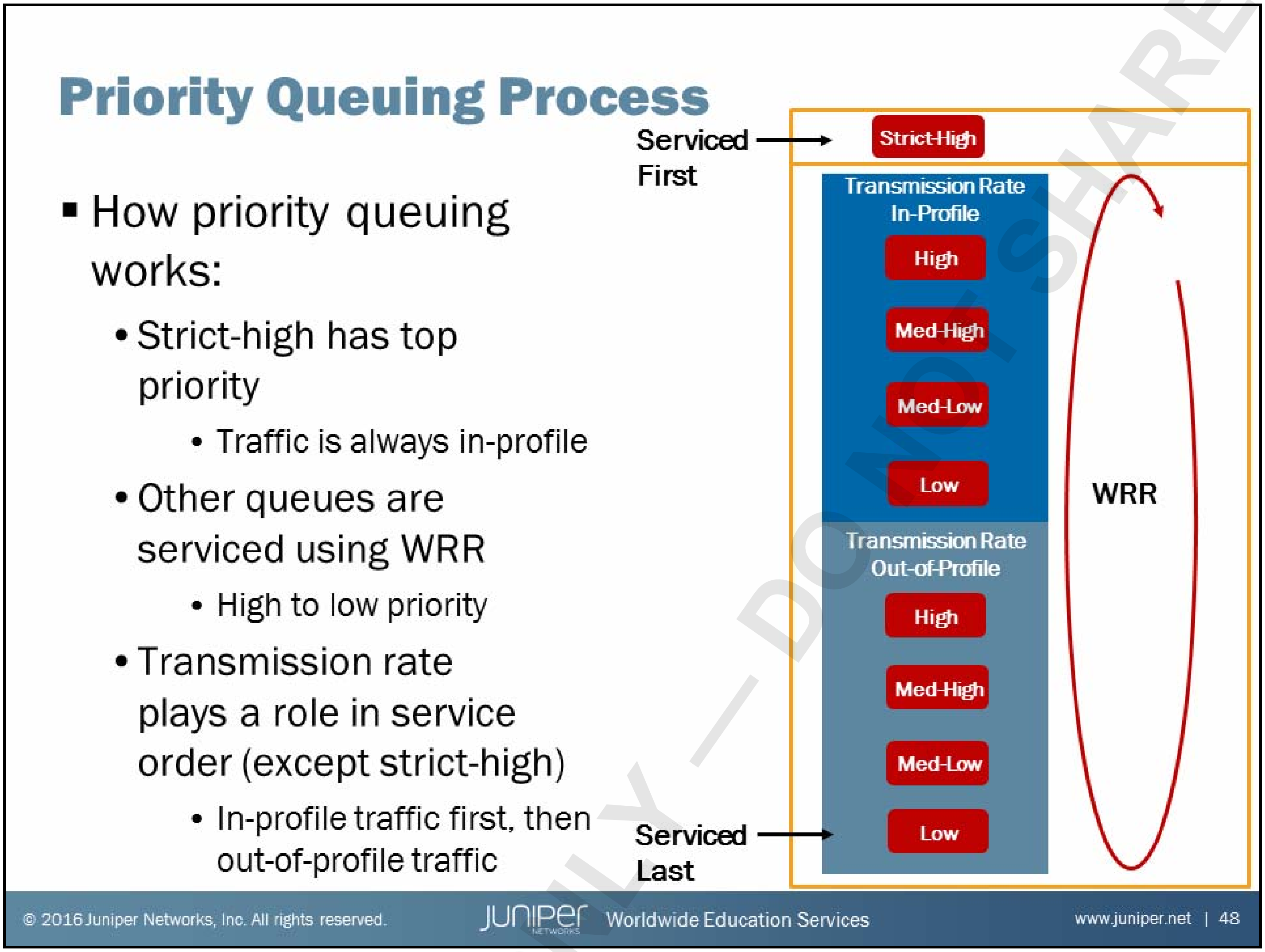
Queue Priority

- Priority
  - Relative importance of the queue compared to other queues
    - Determines the order in which the interface transmits traffic from queues
    - Ensures certain queues are served before other queues when congestion occurs
  - Queues receive service according to their assigned priority
    - Generally based on WRR algorithm
  - Supported priorities levels:
    - Low
    - Medium-low
    - Medium-high
    - High
    - Strict-high

Priority

Queue priority dictates a queue’s relative importance in comparison to other queues. It determines the order in which queues can transmit their packets, and ensures that certain queues get higher precedence when congestion occurs. You can configure queues with differing priority values. When this configuration occurs, the queues receive service according to their priority based on a WRR algorithm. The slide lists the priority levels available for schedulers.

A note about software priority versus hardware priority: Keep in mind that the available priority level combinations can vary between different Juniper Networks platforms. The priority levels listed on the slide are Junos software priorities. Software priorities map to hardware priority levels; however, these mappings depend on the Flexible PIC Concentrator (FPC) or modular port concentrator (MPC) type in which the PIC or modular interface card (MIC) is installed. In some cases, each software priority maps to a dedicated hardware priority. In other cases, multiple software priorities map to a single hardware priority. An example of the latter is described on the next page, where strict-high and high share the same hardware priority. Although mappings vary, one constant guideline is that low and high software priorities are always different hardware priorities. See the *Junos Class of Service Configuration Guide* for more detailed information on mappings between software and hardware priorities.



Priority Queuing Process

- How priority queuing works:
  - Strict-high has top priority
    - Traffic is always in-profile
  - Other queues are serviced using WRR
    - High to low priority
  - Transmission rate plays a role in service order (except strict-high)
    - In-profile traffic first, then out-of-profile traffic

Serviced First

Strict-High

Transmission Rate In-Profile

High

Med-High

Med-Low

Low

Transmission Rate Out-of-Profile

High

Med-High

Med-Low

Low

Serviced Last

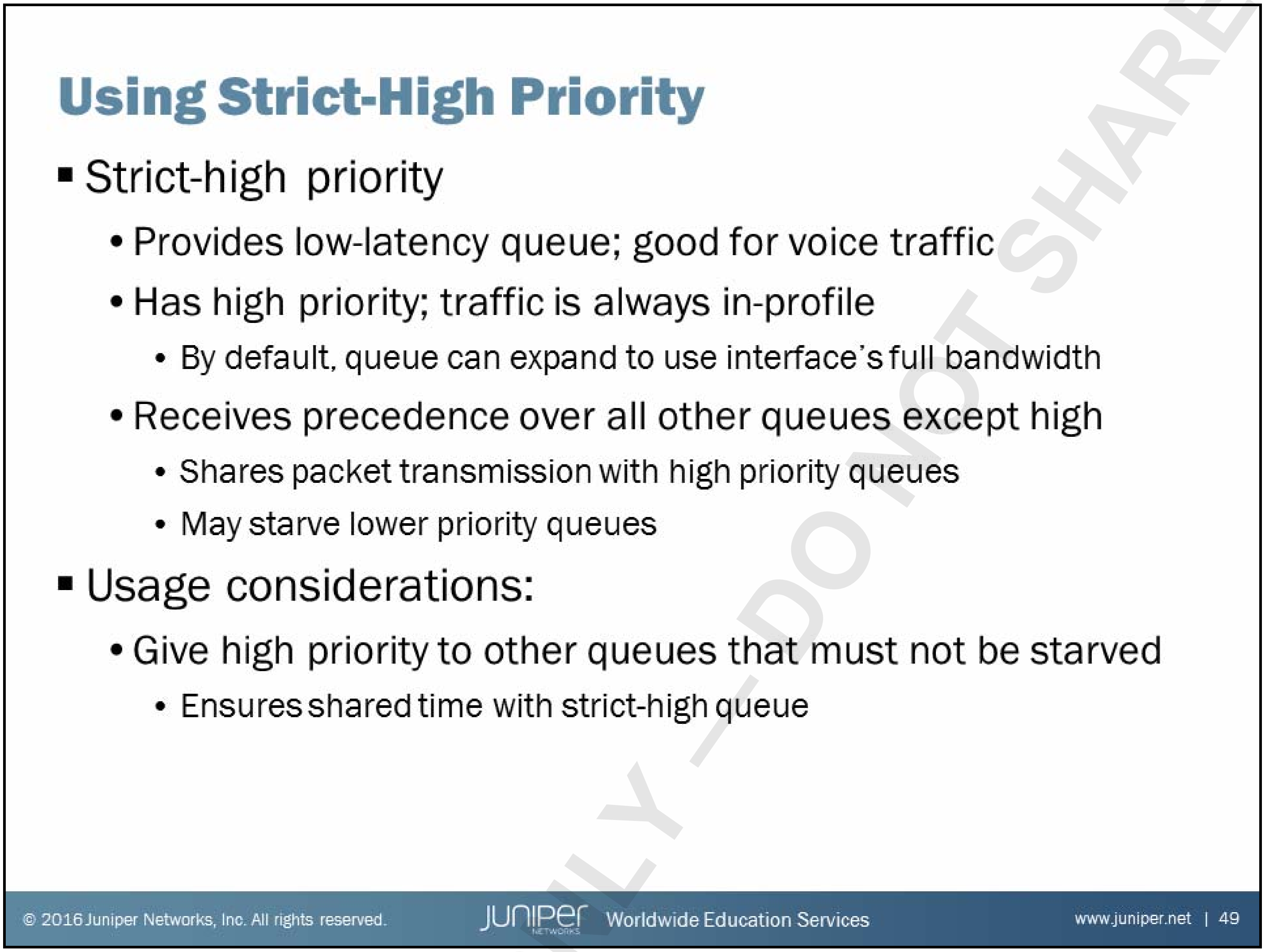
WRR

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 48

How Priority Queuing Works

Queues can transmit packets based on their priority. Strict-high priority is a special setting that always has highest precedence and provides a queue with unlimited access to the interface’s bandwidth.

The other queues use WRR to allocate bandwidth. Under this model, high priority queues have highest precedence, followed by medium-high priority queues, and so on. A critical factor that comes into play as part of this process is the queue’s transmission rate, and whether a packet is in-profile or out-of-profile. Queues with packets that are in-profile get precedence over queues with packets that are out-of-profile, even if the latter queue has a higher priority value. For example, a queue with medium-low priority and in-profile traffic gets precedence over a queue with medium-high priority and out-of-profile traffic.



Using Strict-High Priority

- Strict-high priority
  - Provides low-latency queue; good for voice traffic
  - Has high priority; traffic is always in-profile
    - By default, queue can expand to use interface's full bandwidth
  - Receives precedence over all other queues except high
    - Shares packet transmission with high priority queues
    - May starve lower priority queues
- Usage considerations:
  - Give high priority to other queues that must not be starved
    - Ensures shared time with strict-high queue

JUNIPER NETWORKS Worldwide Education Services

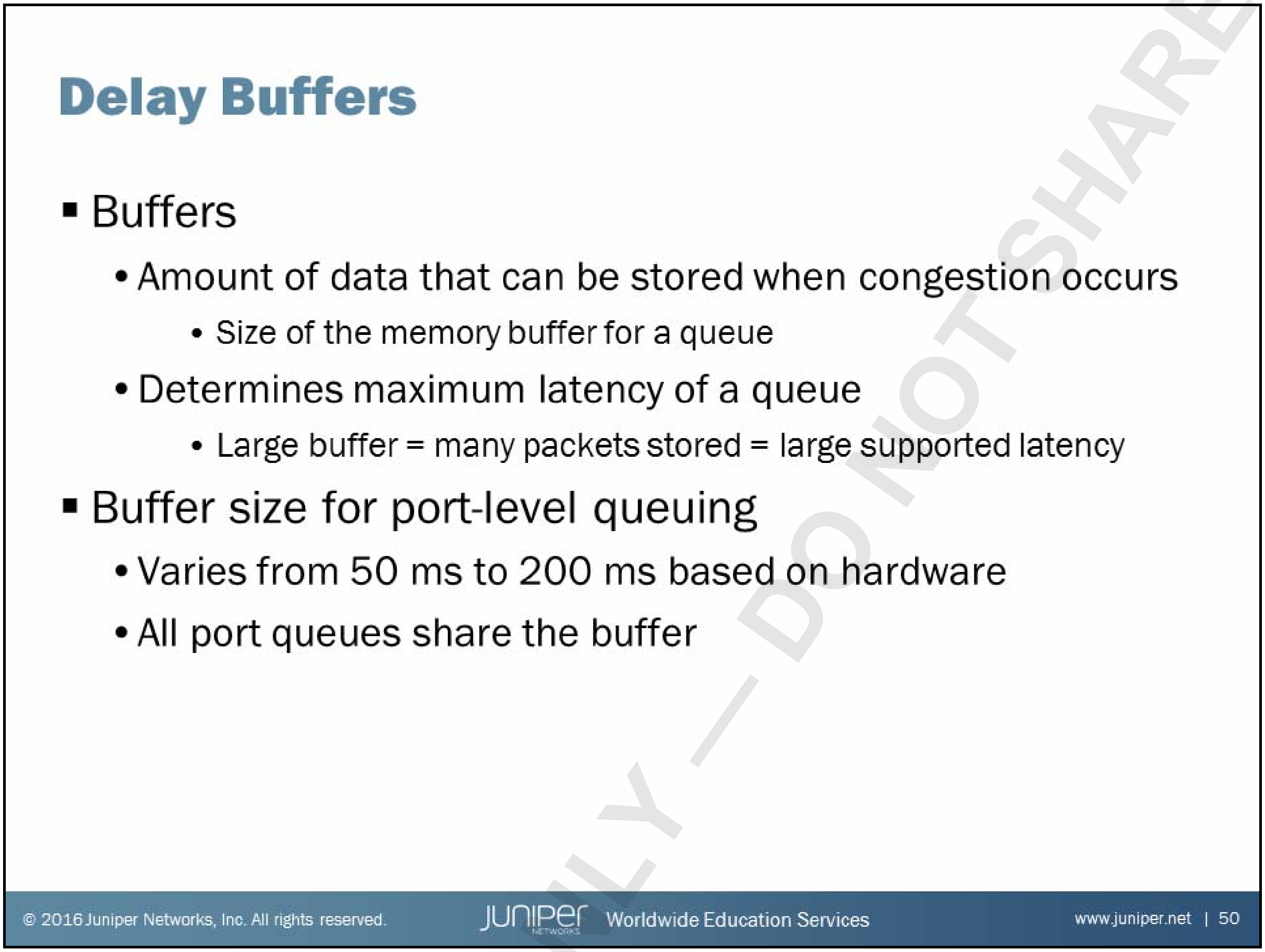
www.juniper.net | 49

Strict-High Priority

Strict-high priority is a special priority level that allows the Junos OS to designate a queue as low-latency. It has the highest precedence and the queue can use up to the full interface's bandwidth, that is, traffic is always considered in-profile. A strict-high queue is always picked ahead of queues with lower priority values, with one exception. Strict-high and high priority queues share an underlying hardware queue, which means queues marked strict-high or high actually share precedence to transmit packets. Note that the unlimited precedence of strict-high can cause starvation of lower priority queues.

Usage Considerations

Because a queue with strict-high priority gets unlimited access to an interface's bandwidth, the transmit-rate statement has no effect, regardless of whether you include it. However, you can put controls on a strict-high queue. A good practice is to give queues with important traffic high priority. Because a strict-high queue shares packet transmission with high priority queues, you can give important traffic high priority, which ensures that it will not be starved by the strict-high queue.



Delay Buffers

- Buffers
  - Amount of data that can be stored when congestion occurs
    - Size of the memory buffer for a queue
  - Determines maximum latency of a queue
    - Large buffer = many packets stored = large supported latency
- Buffer size for port-level queuing
  - Varies from 50 ms to 200 ms based on hardware
  - All port queues share the buffer

Worldwide Education Services www.juniper.net | 50

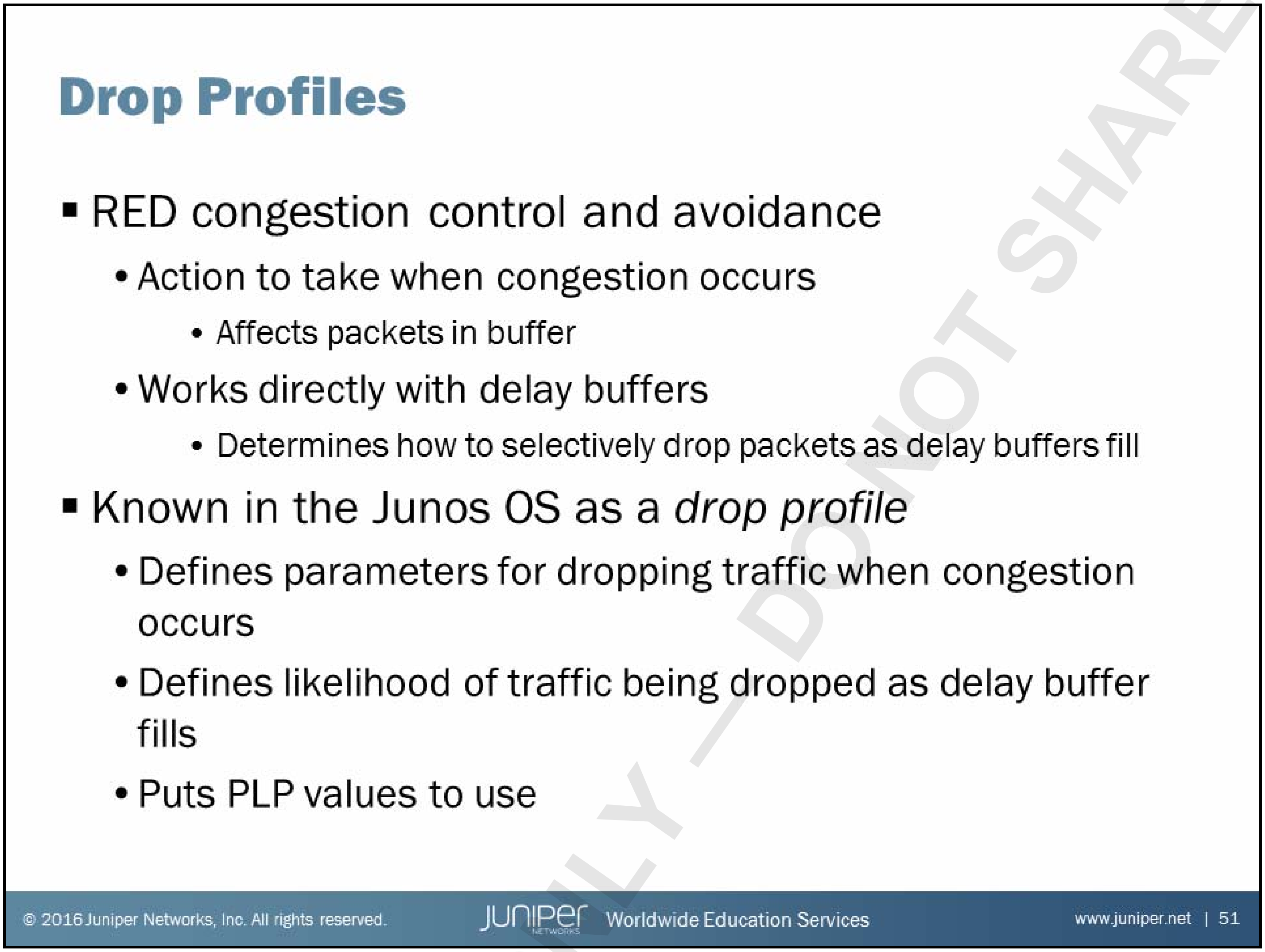
Buffers

Delay buffers define the amount of data that can be stored when congestion occurs. The configured value effectively determines the maximum latency of the queue. The larger the defined buffer size, the more packets that can be stored, and thus the larger the supported latency.

Buffer Size for Port-Level Queuing

The buffer size for port-level queuing ranges from 50 ms to 200 ms, depending on the hardware platform. As with transmission rate, all queues on a port share the available buffer. Buffer size is ultimately a factor of port bandwidth.

A note about large buffer sizes: Congestion and packet dropping occur when large bursts of traffic are received by slower interfaces, which happens when faster interfaces pass traffic to slower interfaces—which often is the case when edge devices receive traffic from the core of the network. T1, E1, and NxDS0 interfaces and data-link connection identifiers (DLCIs) configured on channelized Intelligent Queuing (IQ) PICs are limited to 100,000 microseconds of delay buffer, which might not be sufficient to support these types of traffic flows. For these interfaces, configuring a larger buffer size to prevent congestion and packet dropping might be necessary. You configure large buffers on channelized IQ, 4-port E3 IQ, and Gigabit Ethernet IQ and IQ2 PICs. See the Junos Class of Service Configuration Guide for detailed information on enabling large buffers for these interfaces.



Drop Profiles

- RED congestion control and avoidance
  - Action to take when congestion occurs
    - Affects packets in buffer
  - Works directly with delay buffers
    - Determines how to selectively drop packets as delay buffers fill
- Known in the Junos OS as a *drop profile*
  - Defines parameters for dropping traffic when congestion occurs
  - Defines likelihood of traffic being dropped as delay buffer fills
  - Puts PLP values to use

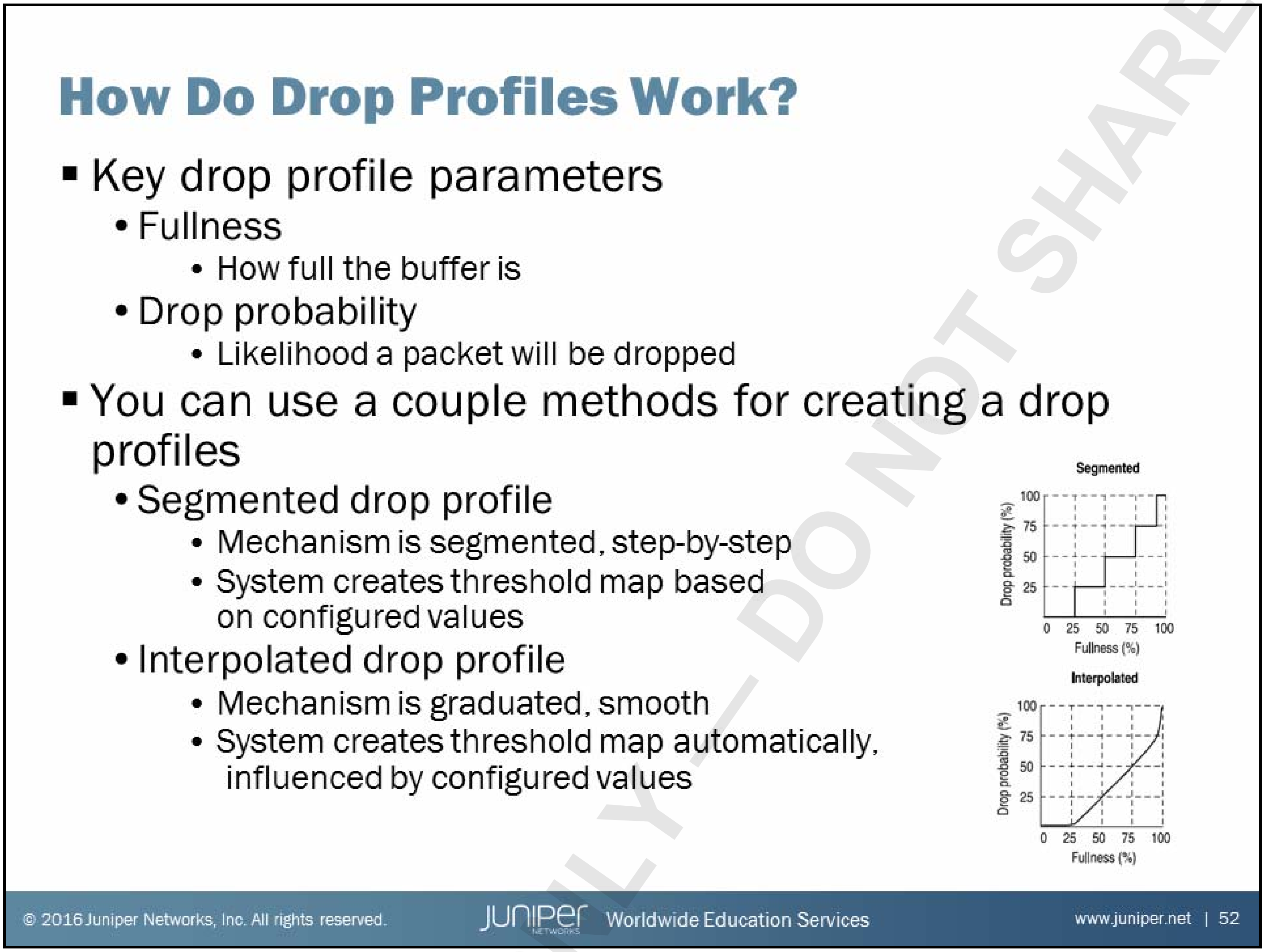
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 51

RED Congestion Control and Avoidance

RED is a mechanism that helps determine actions to take when congestion occurs within a device. RED works directly with delay buffers by determining how to selectively drop packets as the delay buffers fill.

Drop Profile

In the Junos OS, drop profiles perform RED by specifying the parameters for dropping traffic when congestion occurs. Specifically, drop profiles define the likelihood of a packet being dropped based on the fullness of its related buffer. In practical terms, drop profiles put PLP values (set on a packet at ingress) to use.



How Do Drop Profiles Work?

- Key drop profile parameters
  - Fullness
    - How full the buffer is
  - Drop probability
    - Likelihood a packet will be dropped
- You can use a couple methods for creating a drop profiles
  - Segmented drop profile
    - Mechanism is segmented, step-by-step
    - System creates threshold map based on configured values
  - Interpolated drop profile
    - Mechanism is graduated, smooth
    - System creates threshold map automatically, influenced by configured values

JUNIPer NETWORKS Worldwide Education Services www.juniper.net | 52

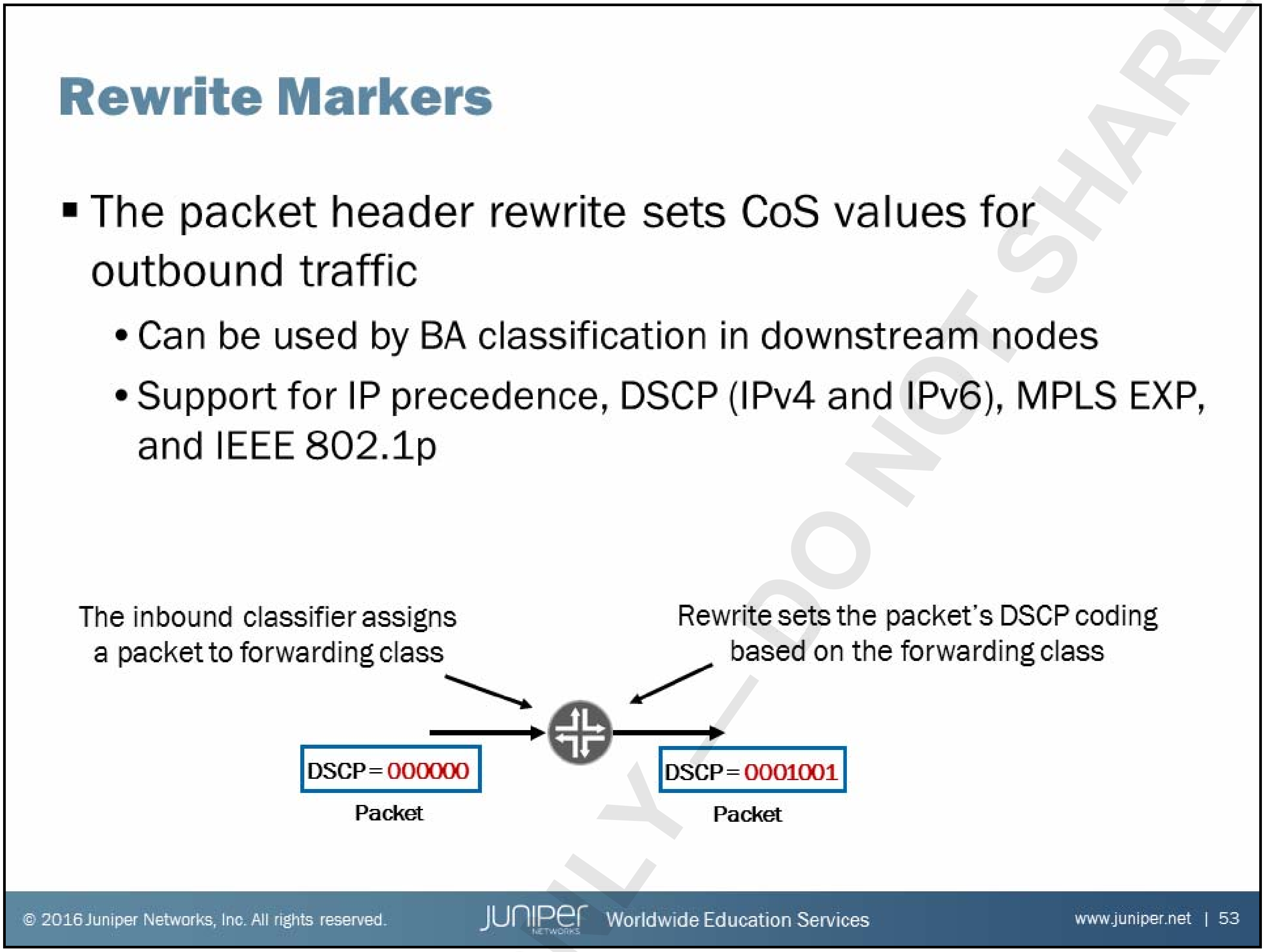
Key Drop Profile Parameters

A drop profile is essentially a dual-threshold mechanism. One threshold is queue fullness, which specifies how full the buffer is. The other threshold is drop probability, which is the likelihood a packet will be dropped. When you configure a drop profile, you are essentially configuring a series of “if-then” statements. For example, you could specify that if the buffer reaches 50% of its capacity (fullness), then a 50% chance exists that a packet arriving at that moment will be dropped (drop probability).

Drop Profile Options

There are a couple methods that can be used for creating drop profiles:

- Segmented drop profile: Functions in a sort of step-by-step manner, with the threshold map based only on your configured values. To create the drop profile’s threshold map, the Junos OS begins at the bottom-left corner in the graph shown on the slide, representing a 0% fill level and a 0% drop probability. Referring to the example on the slide, the threshold map remains at 0% drop probability (the line on the graph moves directly to the right) until it reaches the first defined fill level of 25%. The system immediately raises the drop probability (the line on the graph moves vertically) to the related configured value, 25%. The process repeats for all of the defined fill levels and drop probabilities, ending with the implicit 100% fill level and 100% drop probability. The line on the graph moves to the top-right corner as the buffer becomes completely full and drops all traffic.
- Interpolated drop profile: Creates a more gradual threshold map, because it creates the threshold map automatically. Instead of using the step-wise approach of the segmented method, an interpolated drop profile uses 64 system-defined threshold points, including your configured values, to create a threshold map. The result is a smoother threshold map with graduated data points from point (0,0) and (100,100).



Rewrite Markers

- The packet header rewrite sets CoS values for outbound traffic
  - Can be used by BA classification in downstream nodes
  - Support for IP precedence, DSCP (IPv4 and IPv6), MPLS EXP, and IEEE 802.1p

The inbound classifier assigns a packet to forwarding class

Rewrite sets the packet's DSCP coding based on the forwarding class

Worldwide Education Services

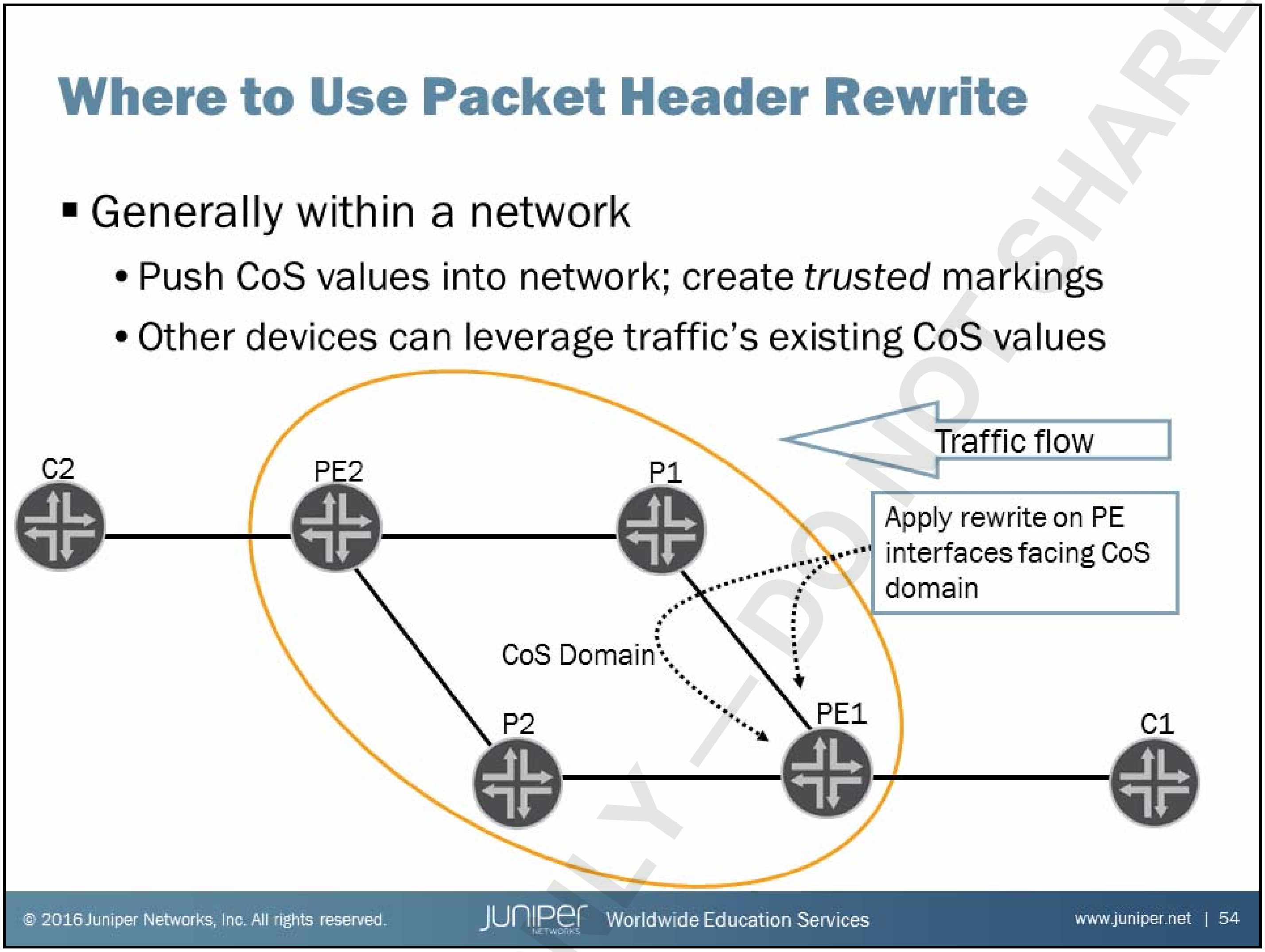
www.juniper.net | 53

Rewrite Markers

Marker rewrite is a key edge device function. The goal is to mark traffic in a consistent manner at the network edge to facilitate BA classification in core devices.

The slide shows an example of DSCP-based marking by an edge router. In this case, traffic arriving from the customer has an all-zeros DS field. The multifield classification actions of the edge router classify the traffic, and the packet travels through the device. Before the device transmits the packet, it writes a value into the packet's CoS field according to the packet's forwarding class and loss priority.

The Junos OS supports packet header rewrite actions for several protocols, as shown on the slide.



Where to Use Packet Header Rewrite

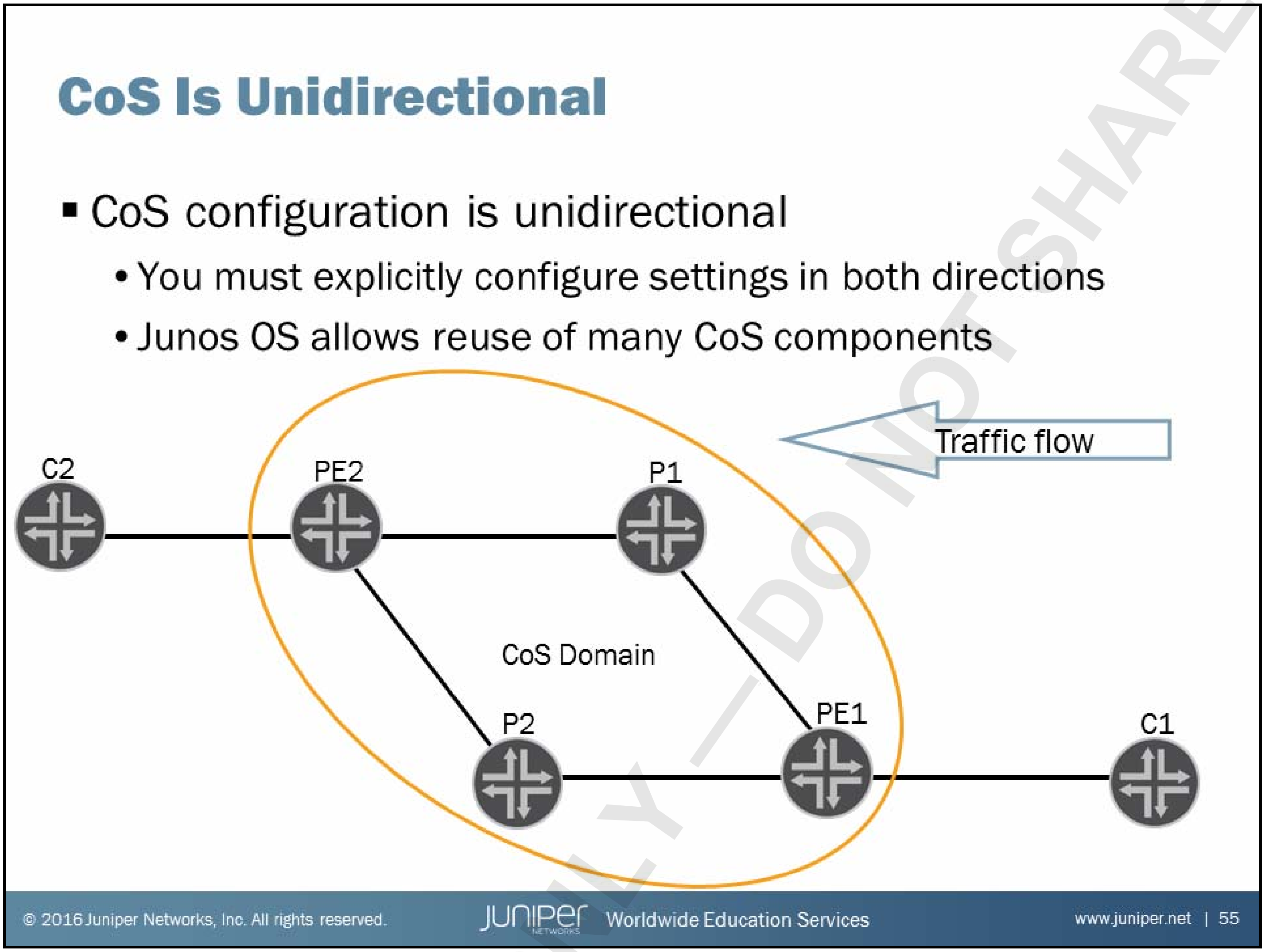
- Generally within a network
  - Push CoS values into network; create *trusted* markings
  - Other devices can leverage traffic's existing CoS values

Traffic flow

Apply rewrite on PE interfaces facing CoS domain

Apply Packet Header Rewrite Within a Network

In general, the best place to use rewrite tables is within a network. Because traffic has already passed through your edge network device, you can push CoS markings onto packets and consider them trusted. Furthermore, other nodes in the network can leverage the traffic's existing CoS values and use the more efficient BA classification method at ingress to minimize the CoS processing workload.



CoS Is Unidirectional

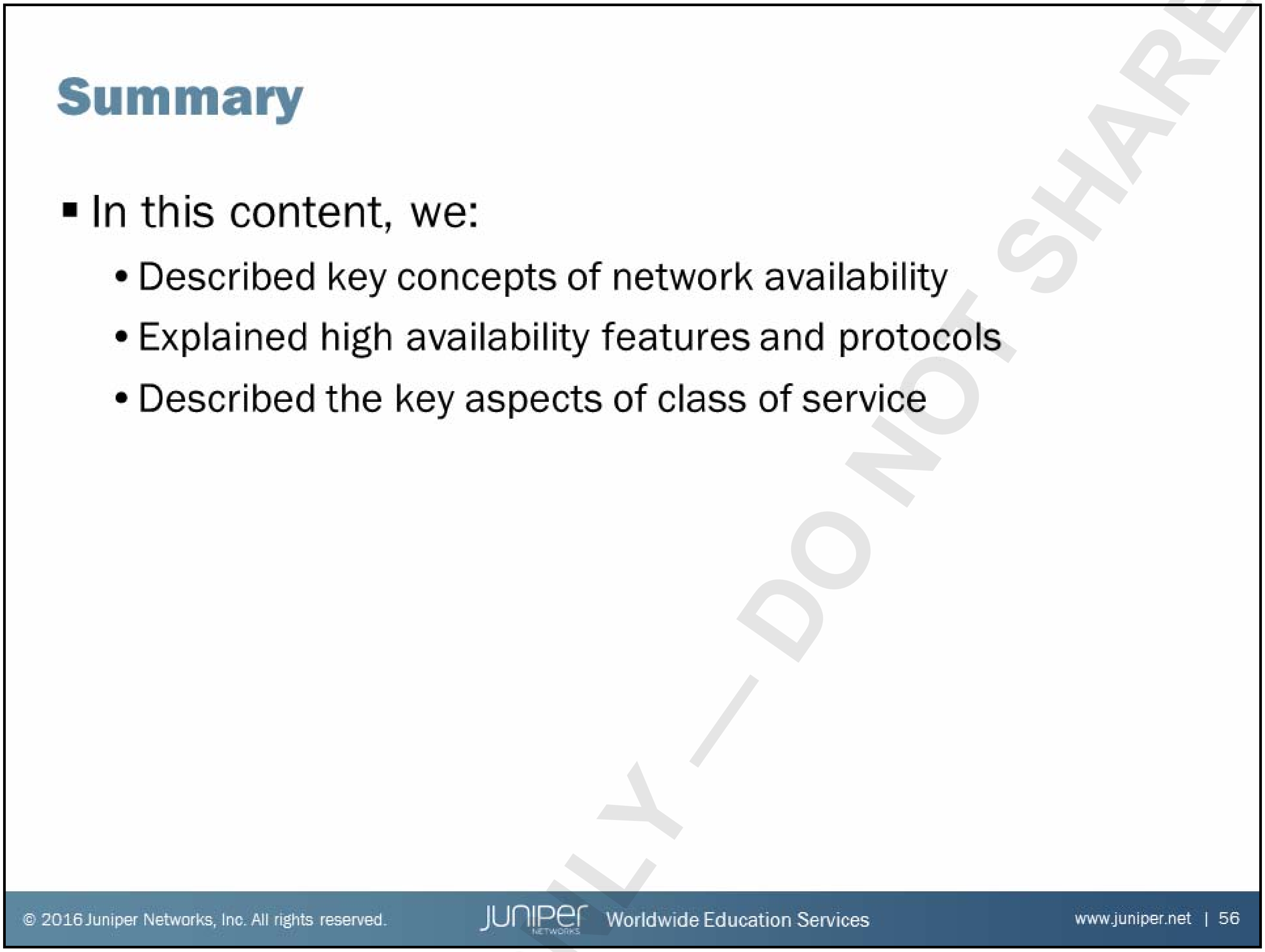
- CoS configuration is unidirectional
  - You must explicitly configure settings in both directions
  - Junos OS allows reuse of many CoS components

Worldwide Education Services

www.juniper.net | 55

CoS Configuration Is Unidirectional

Keep in mind that when you complete a CoS configuration across a network, you have generally configured it in only one direction. To support CoS in both directions, you must configure the nodes in the other direction as well. Fortunately, CoS configuration in the Junos OS is modular and template-based, so you can reuse much of the configuration you originally created.



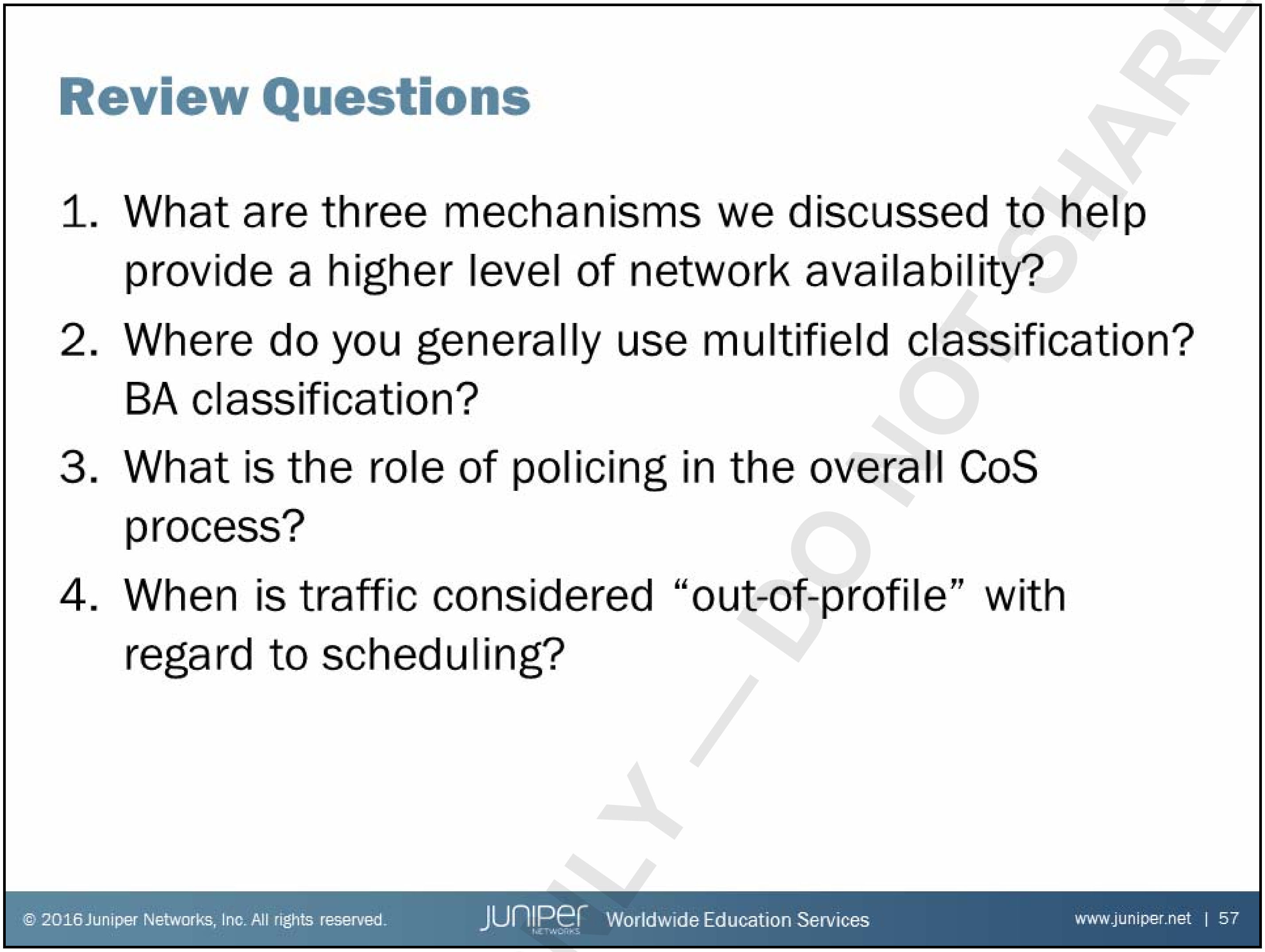
Summary

- In this content, we:
  - Described key concepts of network availability
  - Explained high availability features and protocols
  - Described the key aspects of class of service

JUNIPer NETWORKS
Worldwide Education Services
www.juniper.net | 56

We Discussed:

- Key concepts of network availability;
- High availability features and protocols and
- Key aspects of class of service.



Review Questions

1. What are three mechanisms we discussed to help provide a higher level of network availability?
2. Where do you generally use multifield classification? BA classification?
3. What is the role of policing in the overall CoS process?
4. When is traffic considered “out-of-profile” with regard to scheduling?

Worldwide Education Services

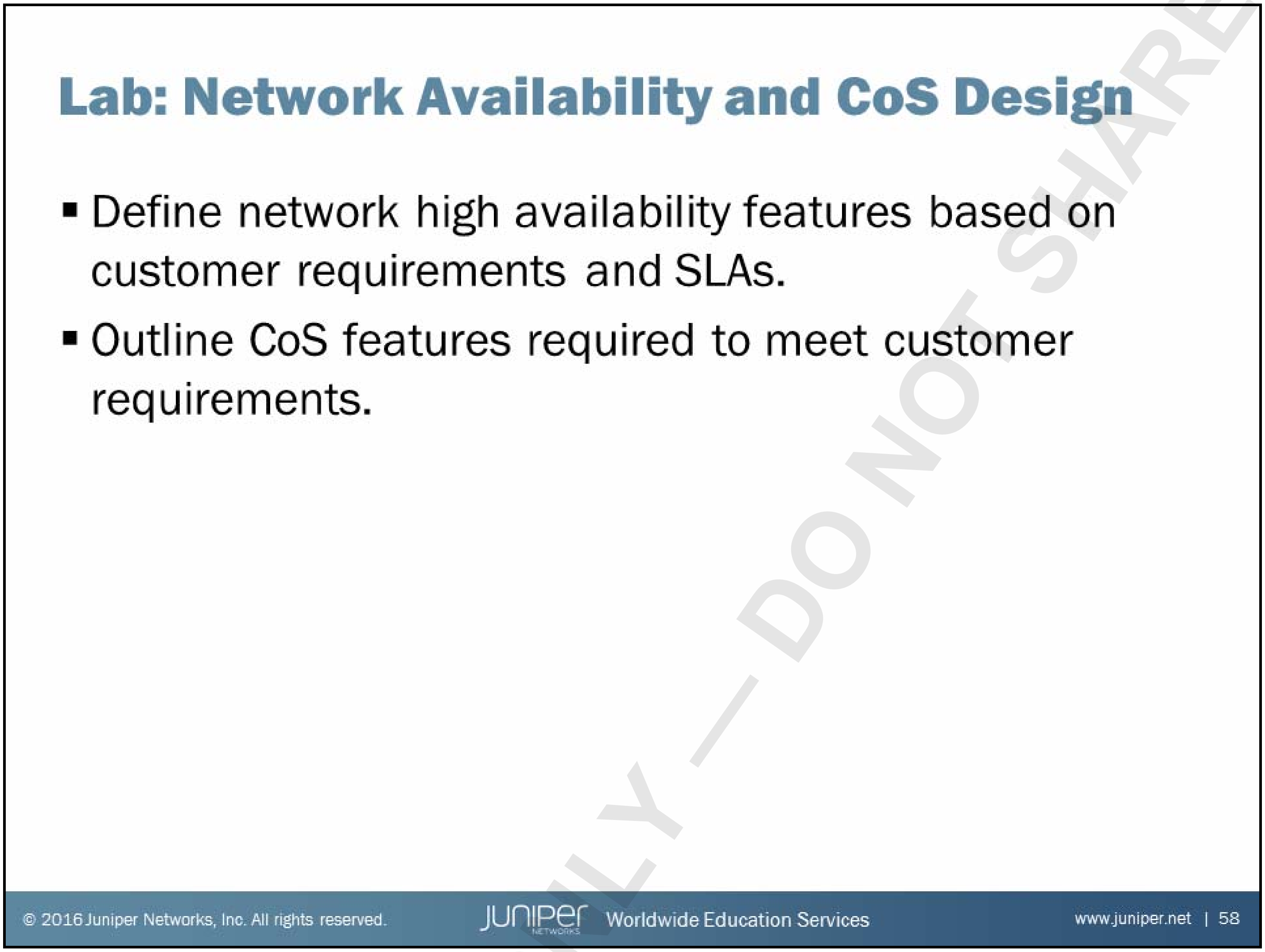
Review Questions

1.

2.

3.

4.



Lab: Network Availability and CoS Design

- Define network high availability features based on customer requirements and SLAs.
- Outline CoS features required to meet customer requirements.

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 58

Lab: Network Availability and CoS Design

The slide provides the objective for this lab.

Answers to Review Questions

1.

We discussed a few different mechanisms for network HA including, LAG, MC-LAG, hardware redundancy, Virtual Chassis, NSR, BFD, NSB, GRES, and graceful restart.

2.

You generally use multifield classification at the ingress point of the network. You generally use BA classification within the network through the core.

3.

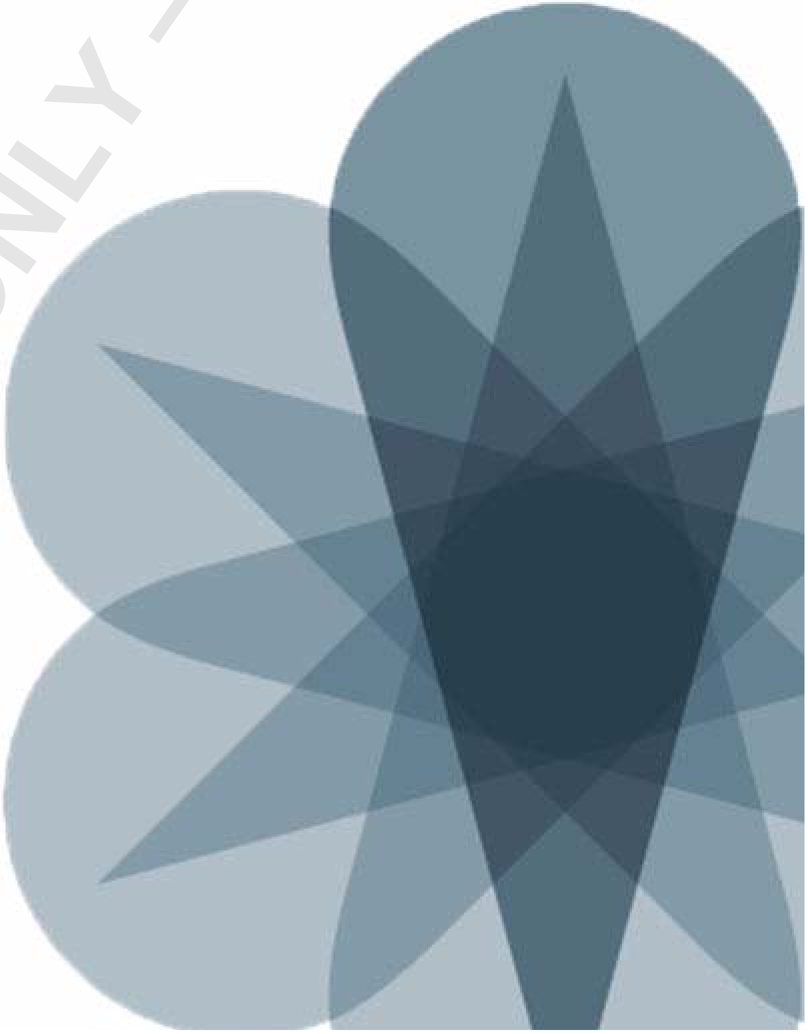
. The primary function of policing is to provide a first stage of congestion management. With policing, you can condition incoming or outgoing traffic by applying bandwidth constraints.

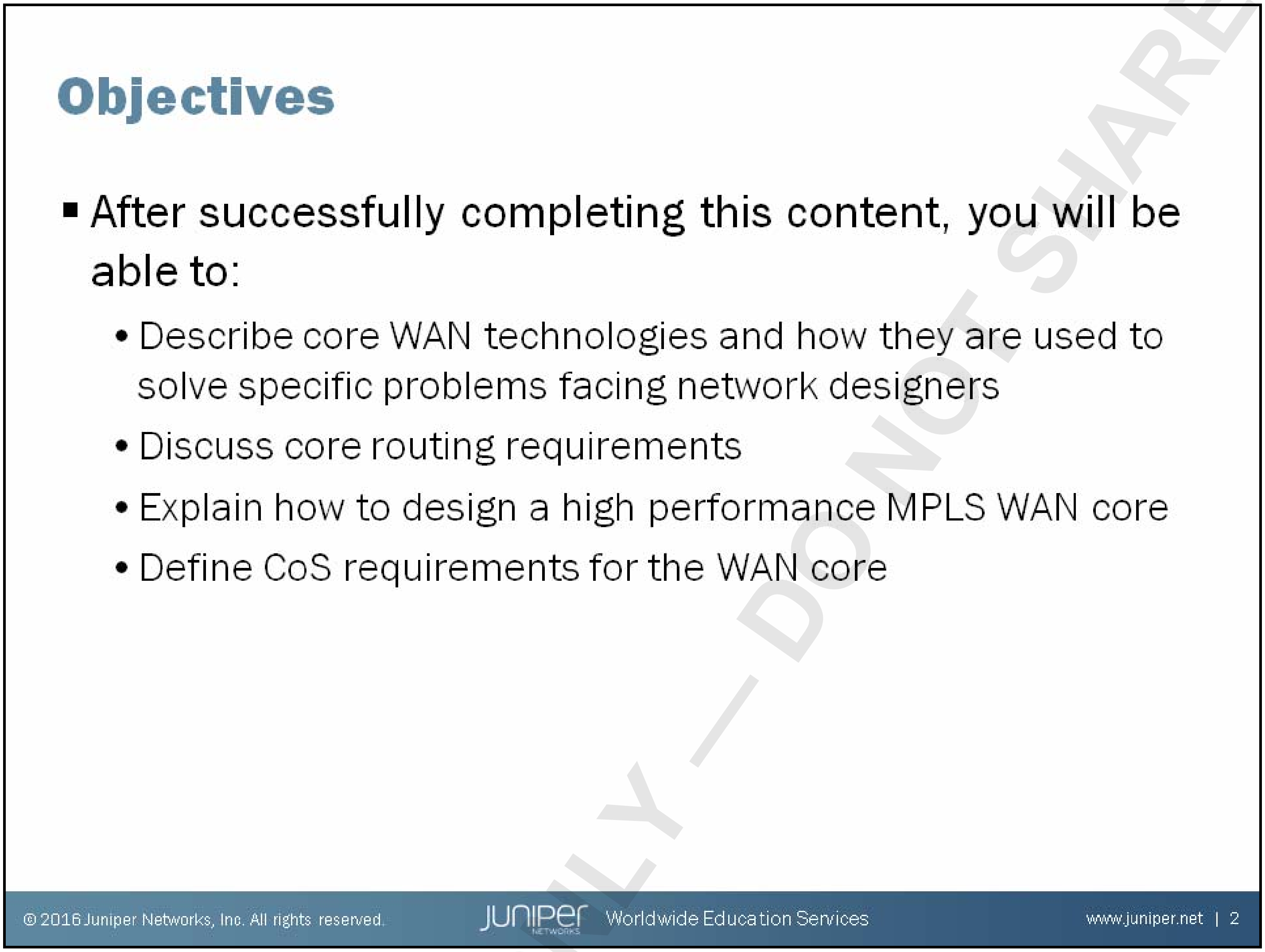
4.

Traffic is “out-of-profile” when it exceeds the queue’s allocated transmission rate.

Juniper Networks Design—WAN

Chapter 5: Service Provider Core WAN





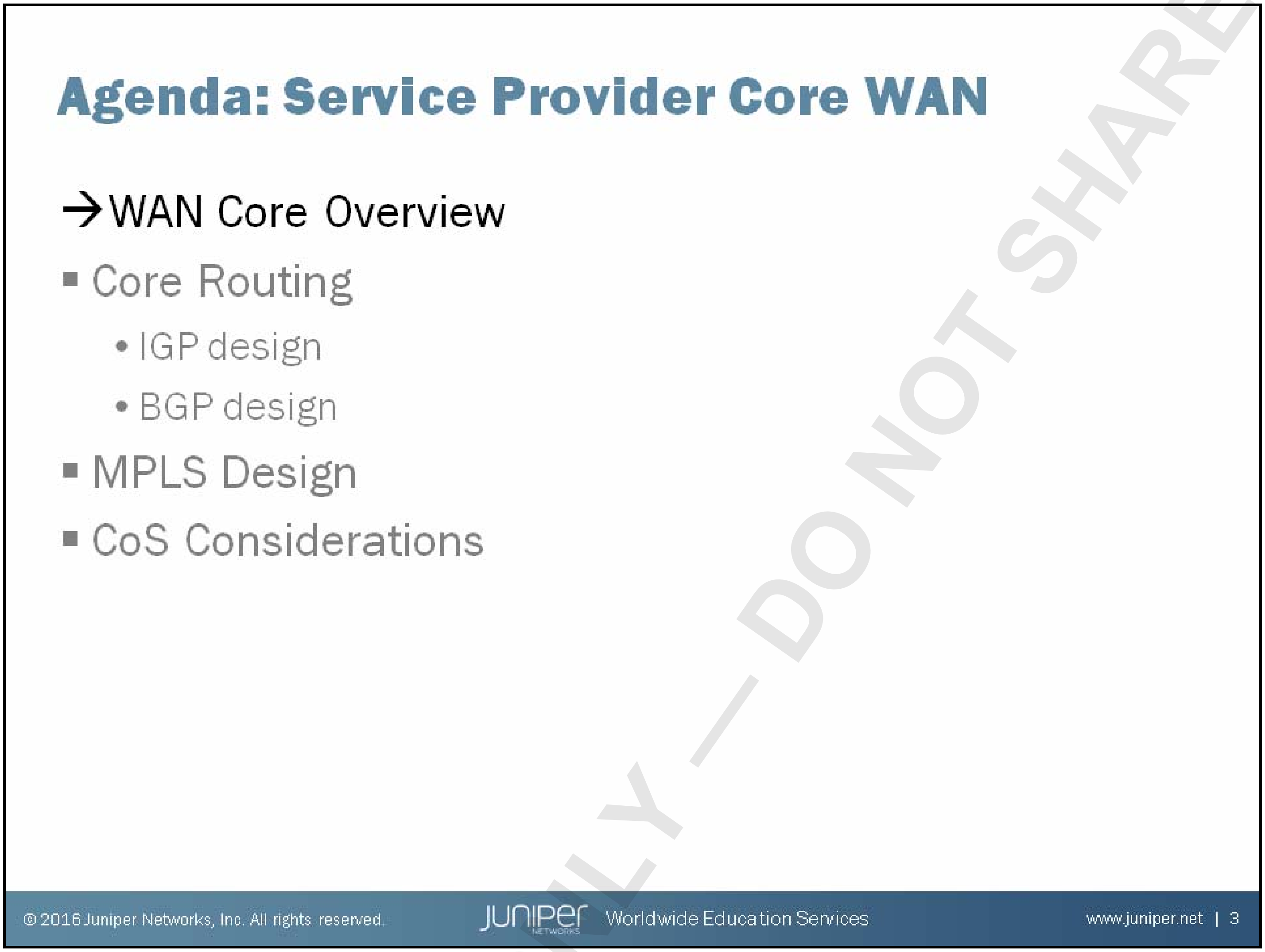
Objectives

- After successfully completing this content, you will be able to:
  - Describe core WAN technologies and how they are used to solve specific problems facing network designers
  - Discuss core routing requirements
  - Explain how to design a high performance MPLS WAN core
  - Define CoS requirements for the WAN core

Juniper Networks Worldwide Education Services www.juniper.net | 2

We Will Discuss:

- Core WAN technologies and how they are used to solve specific problems facing network designers;
- Core routing requirements;
- How to design a high performance MPLS WAN core; and
- CoS requirements for the WAN core.



Agenda: Service Provider Core WAN

→WAN Core Overview

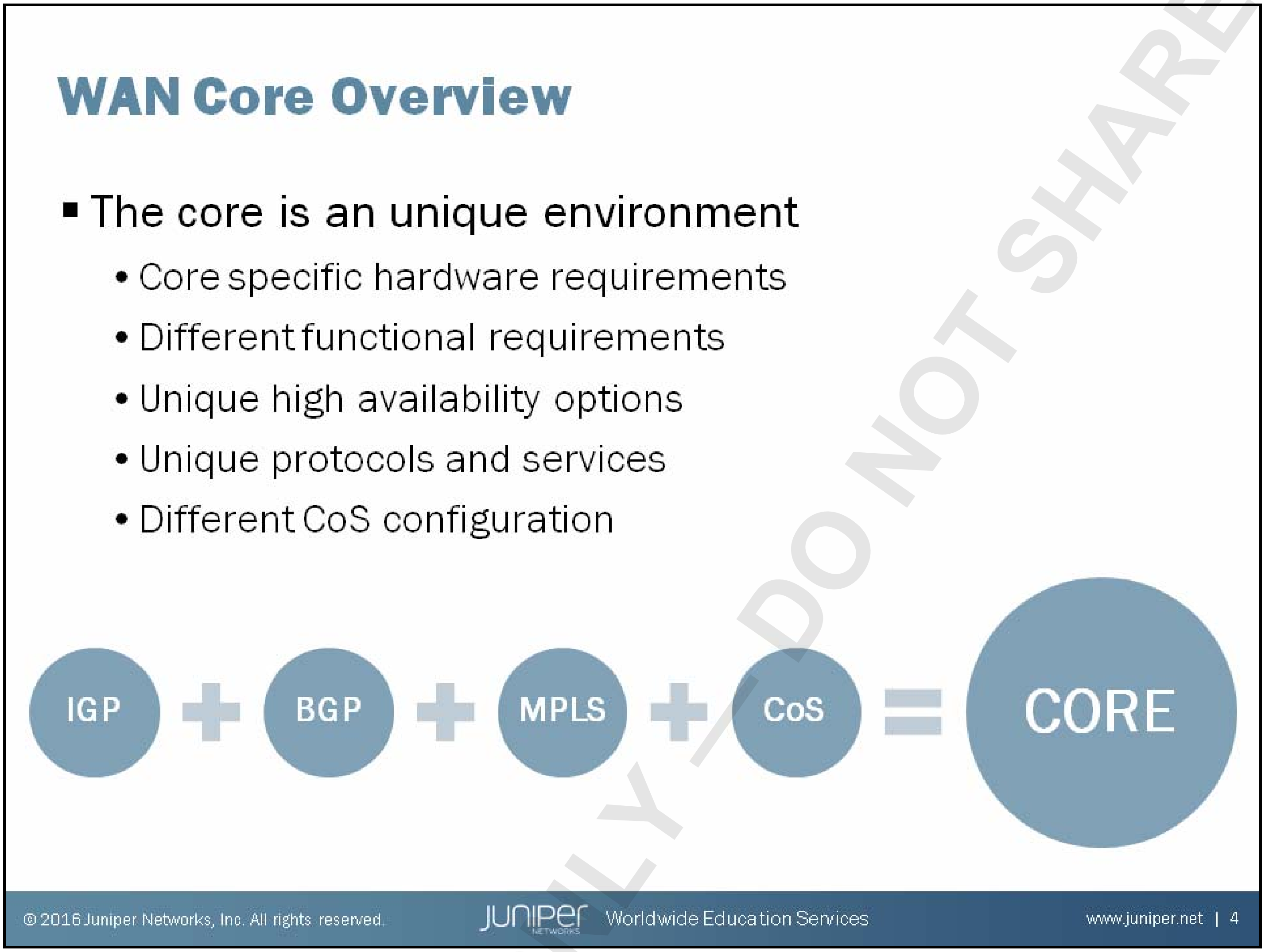
- Core Routing
  - IGP design
  - BGP design
- MPLS Design
- CoS Considerations

Juniper Networks Worldwide Education Services

www.juniper.net | 3

WAN Core Overview

The slide lists the topics we will discuss. We discuss the highlighted topic first.



WAN Core Overview

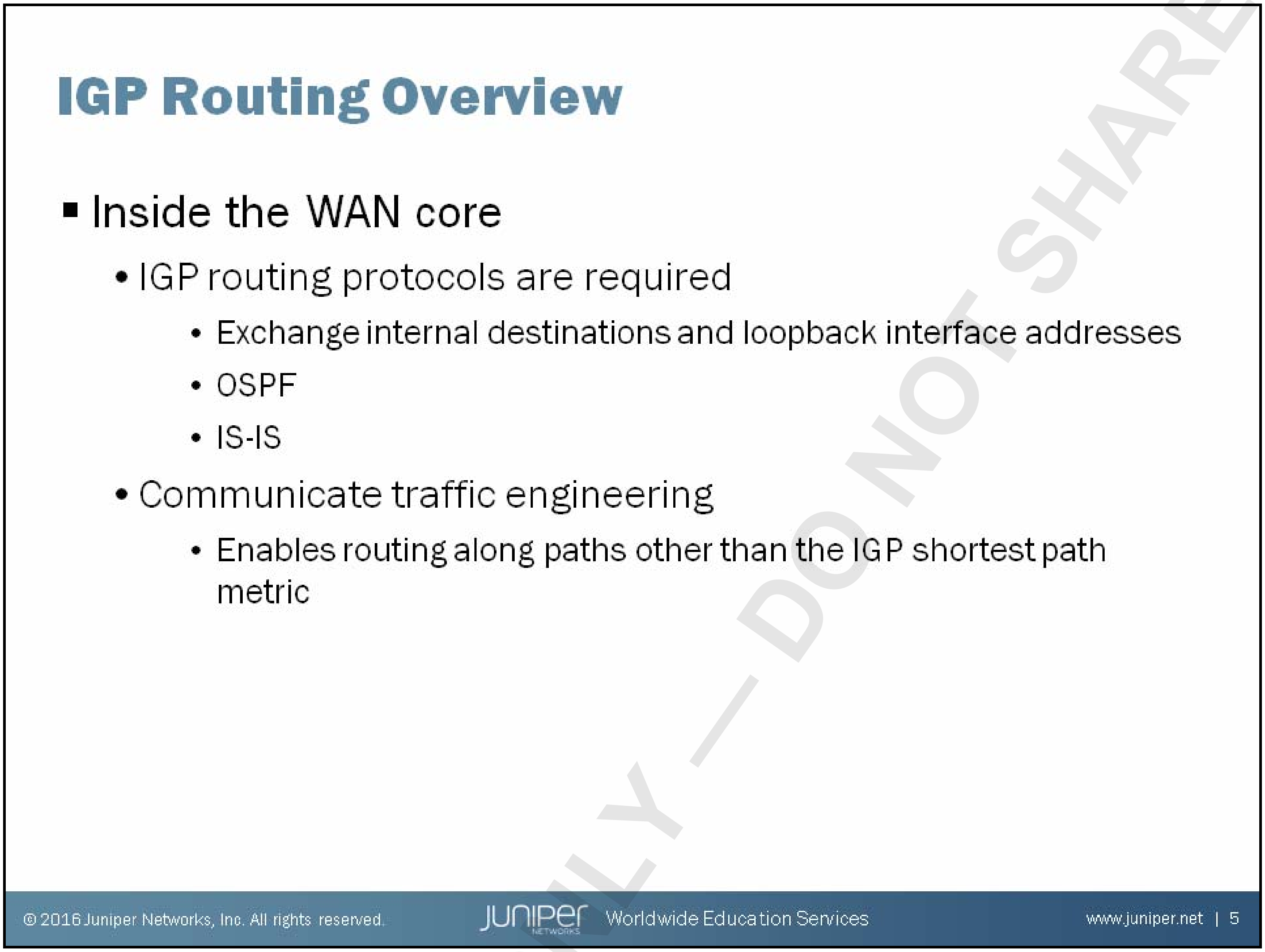
- The core is an unique environment
  - Core specific hardware requirements
  - Different functional requirements
  - Unique high availability options
  - Unique protocols and services
  - Different CoS configuration

IGP + BGP + MPLS + CoS = CORE

JUNIPER NETWORKS
Worldwide Education Services
www.juniper.net | 4

WAN Core Overview

You might find yourself saying, “It’s all the same, right?”, “It’s just a bunch of routers all linked together, like always.” Well, sort of. Responsibilities for core routers are different than those for edge devices. The quantity and type of interfaces used on edge devices are different than devices in the core. Edge routers are required to provide connectivity for hundreds of clients while core routers offer services that are not customer facing. Core routers run OSPF or IS-IS to exchange internal destinations and traffic engineering information. Core routers utilize RSVP or LDP to support an edge to edge mesh of label switched paths across the WAN core. Class of service is different in the core because core routers use a different process for traffic classification. Because of the different device responsibilities WAN designers should consider buying the right device for the two different roles.



IGP Routing Overview

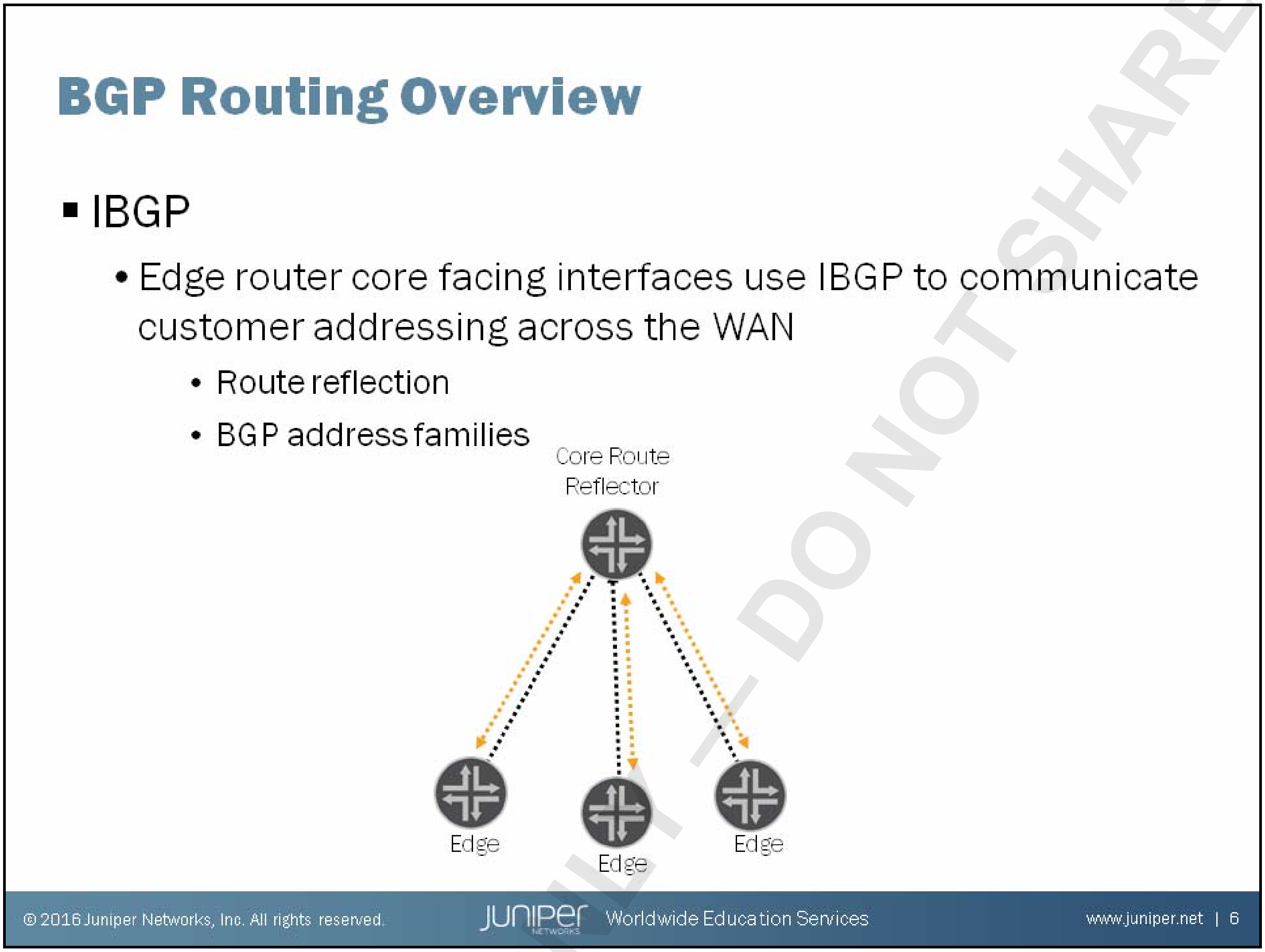
- Inside the WAN core
  - IGP routing protocols are required
    - Exchange internal destinations and loopback interface addresses
    - OSPF
    - IS-IS
  - Communicate traffic engineering
    - Enables routing along paths other than the IGP shortest path metric

Worldwide Education Services

www.juniper.net | 5

Inside the WAN Core

Core routers communicate internal network addressing information including the loopback interface addresses used for internal BGP (IBGP) peering. Interior gateway protocols (IGPs) are also configured to communicate traffic engineering data. Traffic engineering data enables complex computations for the forwarding paths created by MPLS. OSPF and IS-IS are two popular options for core IGP routing protocols.



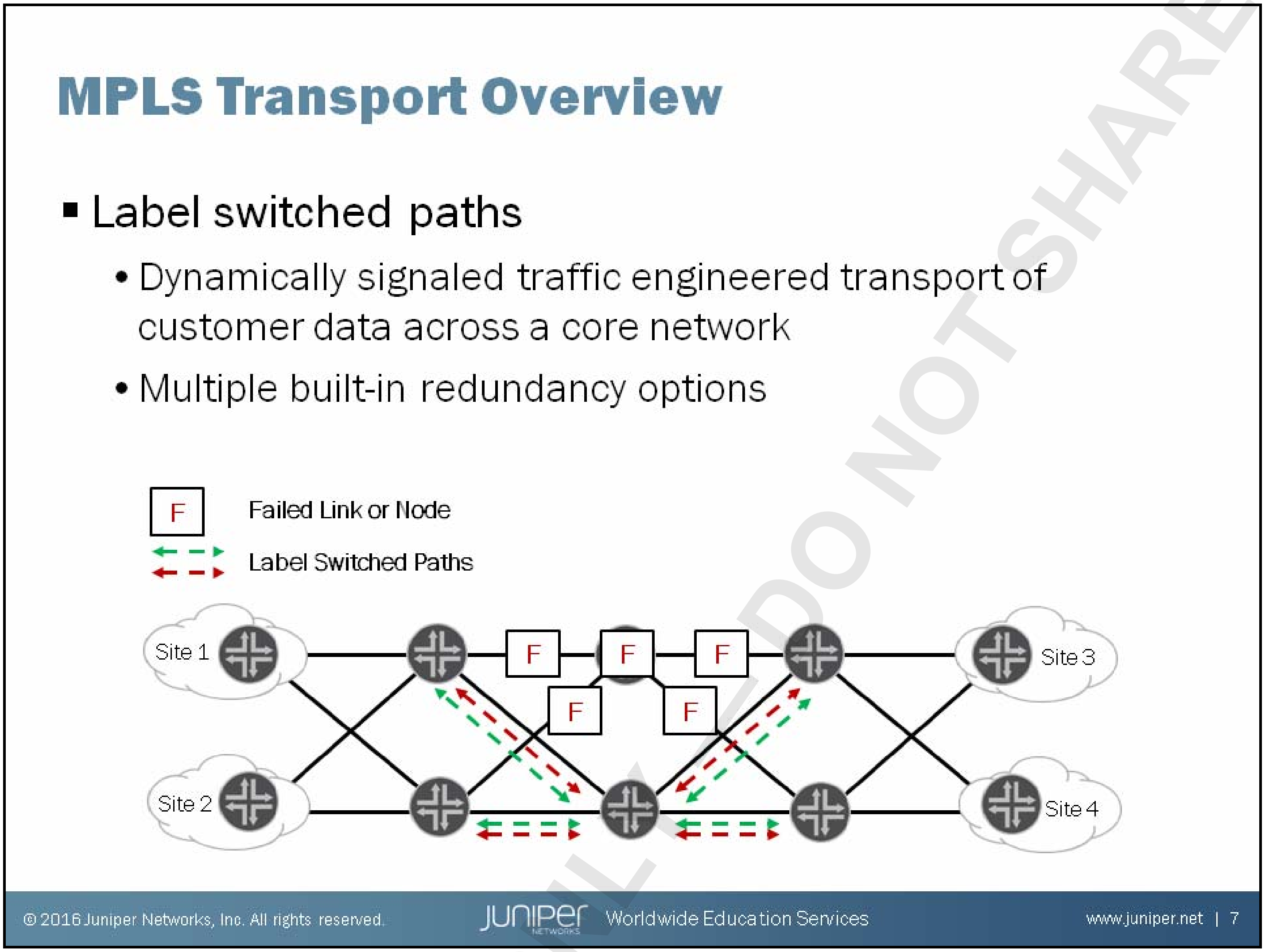
BGP Routing Overview

- IBGP
  - Edge router core facing interfaces use IBGP to communicate customer addressing across the WAN
    - Route reflection
    - BGP address families

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 6

IBGP Routing

Customer addresses are distributed across the core network using route reflectors and IBGP peering. For redundancy purposes multiple route reflectors are configured. Customer addresses are communicated using different BGP address families depending on the type of Layer 2 or Layer 3 VPN services being offered.



MPLS Transport Overview

- Label switched paths
  - Dynamically signaled traffic engineered transport of customer data across a core network
  - Multiple built-in redundancy options

F Failed Link or Node

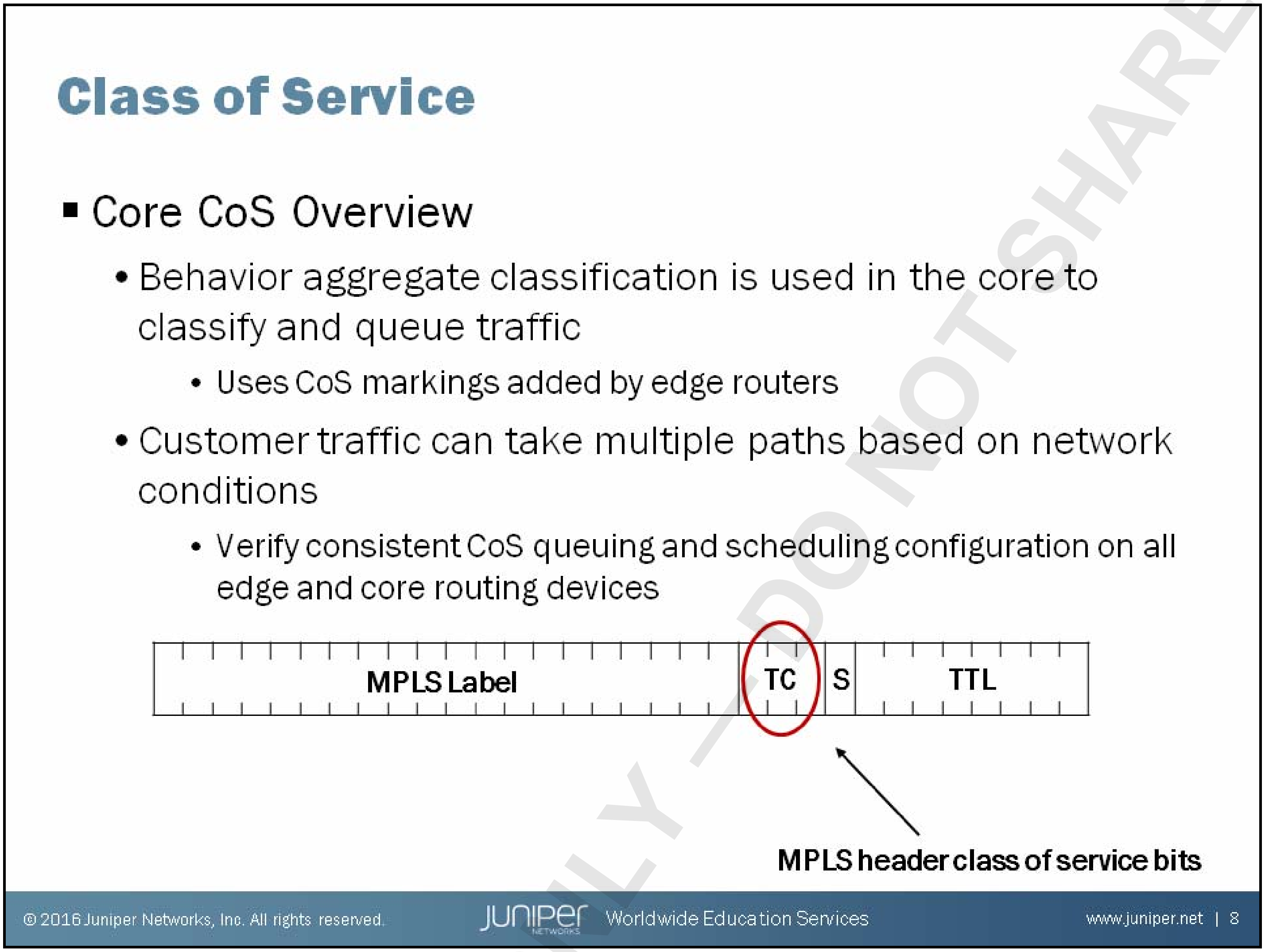
Label Switched Paths

Worldwide Education Services

www.juniper.net | 7

MPLS Label Switched Paths

MPLS label-switched paths (LSPs) are dynamically signaled across the service provider core network. RSVP signaled LSPs can include signaling requirements such as bandwidth, explicit route objects, and administrative groups. These LSP signaling requirements can be used by RSVP to signal a LSP to use a path that differs from the IGP shortest path metric.



Class of Service

- Core CoS Overview
  - Behavior aggregate classification is used in the core to classify and queue traffic
    - Uses CoS markings added by edge routers
  - Customer traffic can take multiple paths based on network conditions
    - Verify consistent CoS queuing and scheduling configuration on all edge and core routing devices

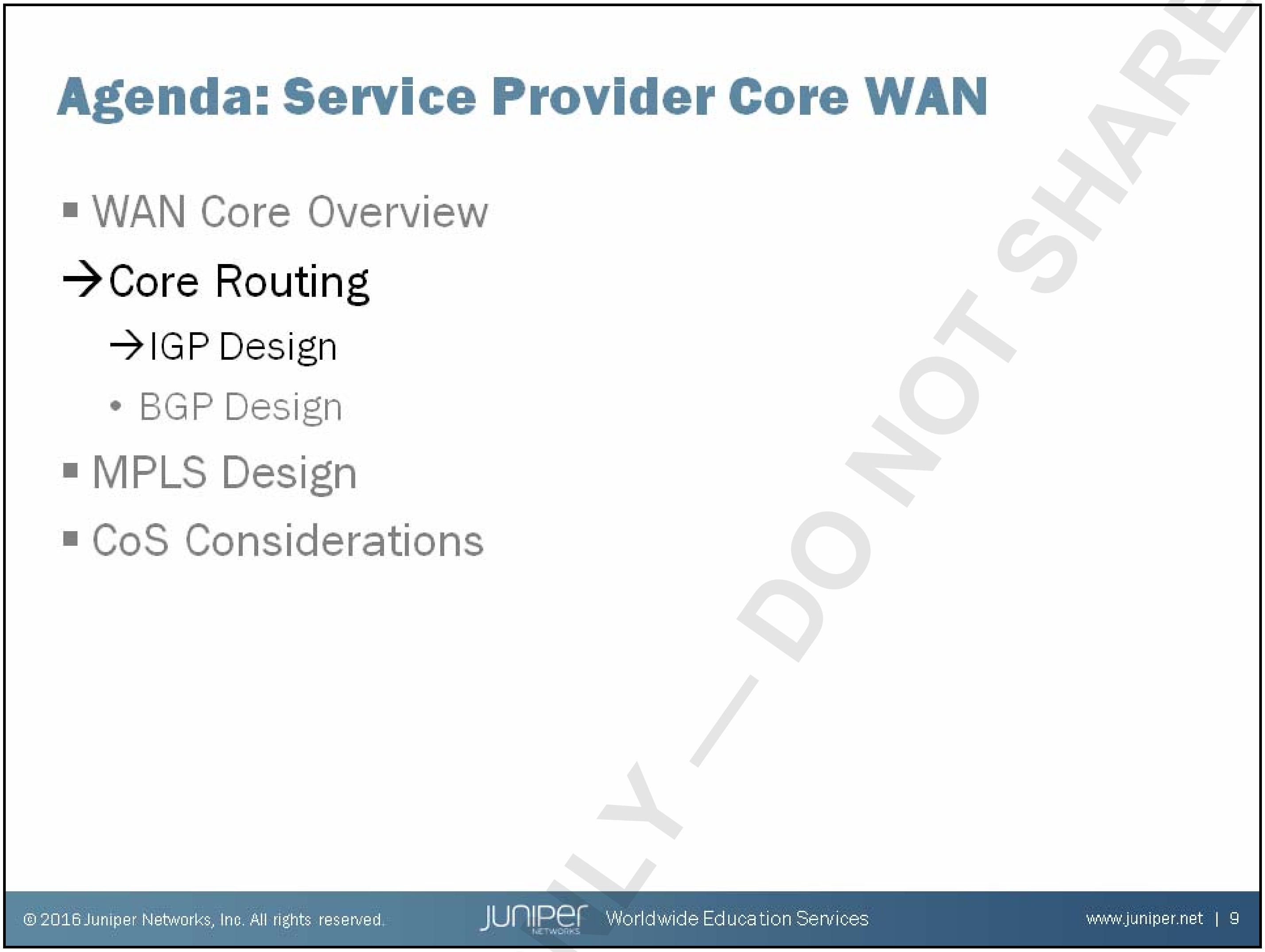
MPLS header class of service bits

Worldwide Education Services

www.juniper.net | 8

Class of Service

When customer traffic enters the service provider network, edge routers classify and queue the different traffic types by analyzing packet contents. Once the traffic is classified, the edge routers add class of service (CoS) markings to the appropriate headers and forward the traffic into the core. Core routers analyze these markings to classify and queue the data for forwarding. Edge and core routers should be configured to schedule and queue traffic similarly in order to guarantee consistent treatment of traffic across the WAN core.



Agenda: Service Provider Core WAN

- WAN Core Overview

→Core Routing

- →IGP Design
- BGP Design

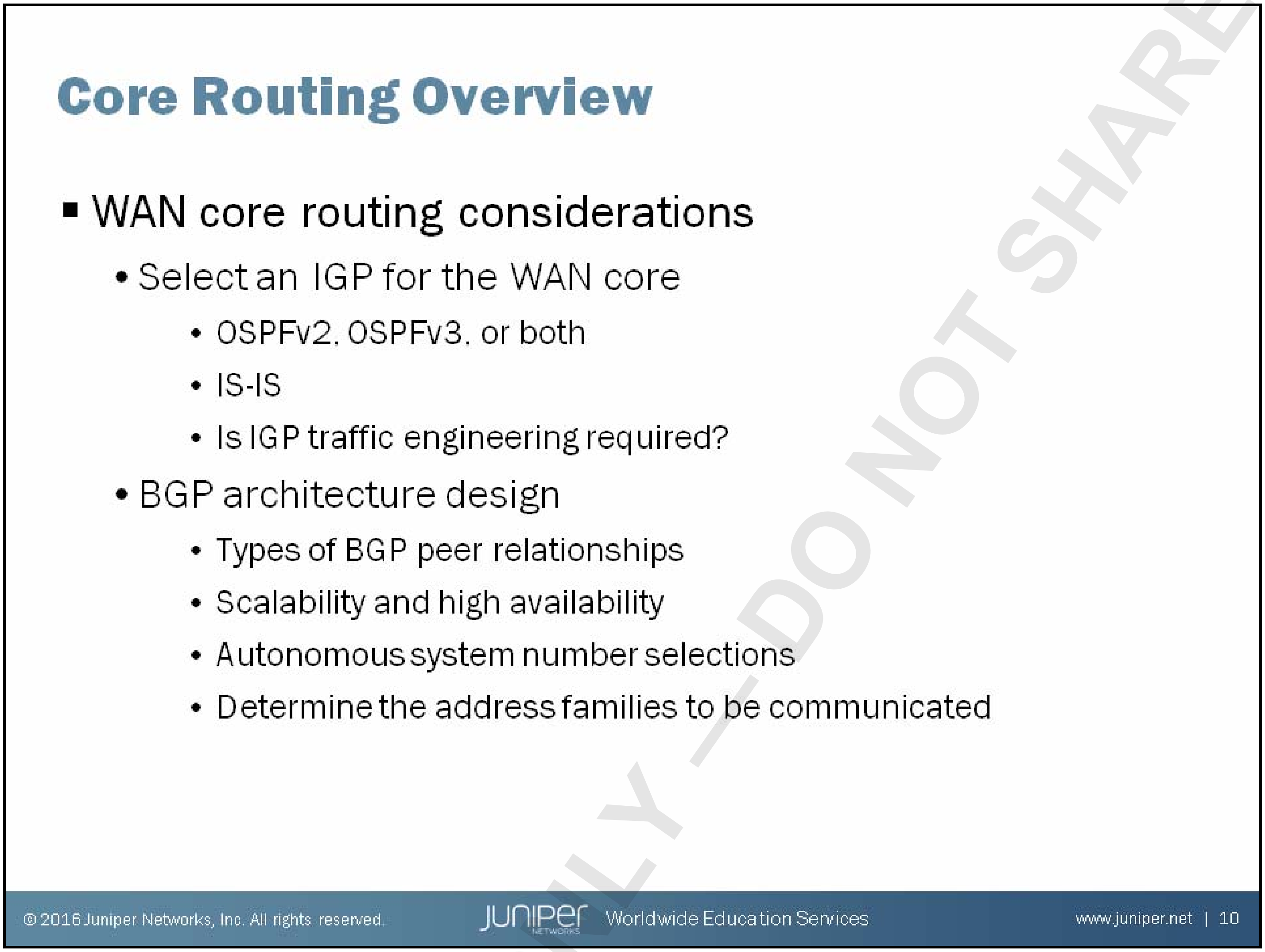
- MPLS Design
- CoS Considerations

Worldwide Education Services

www.juniper.net | 9

IGP Design

The slide highlights the topic we discuss next.



Core Routing Overview

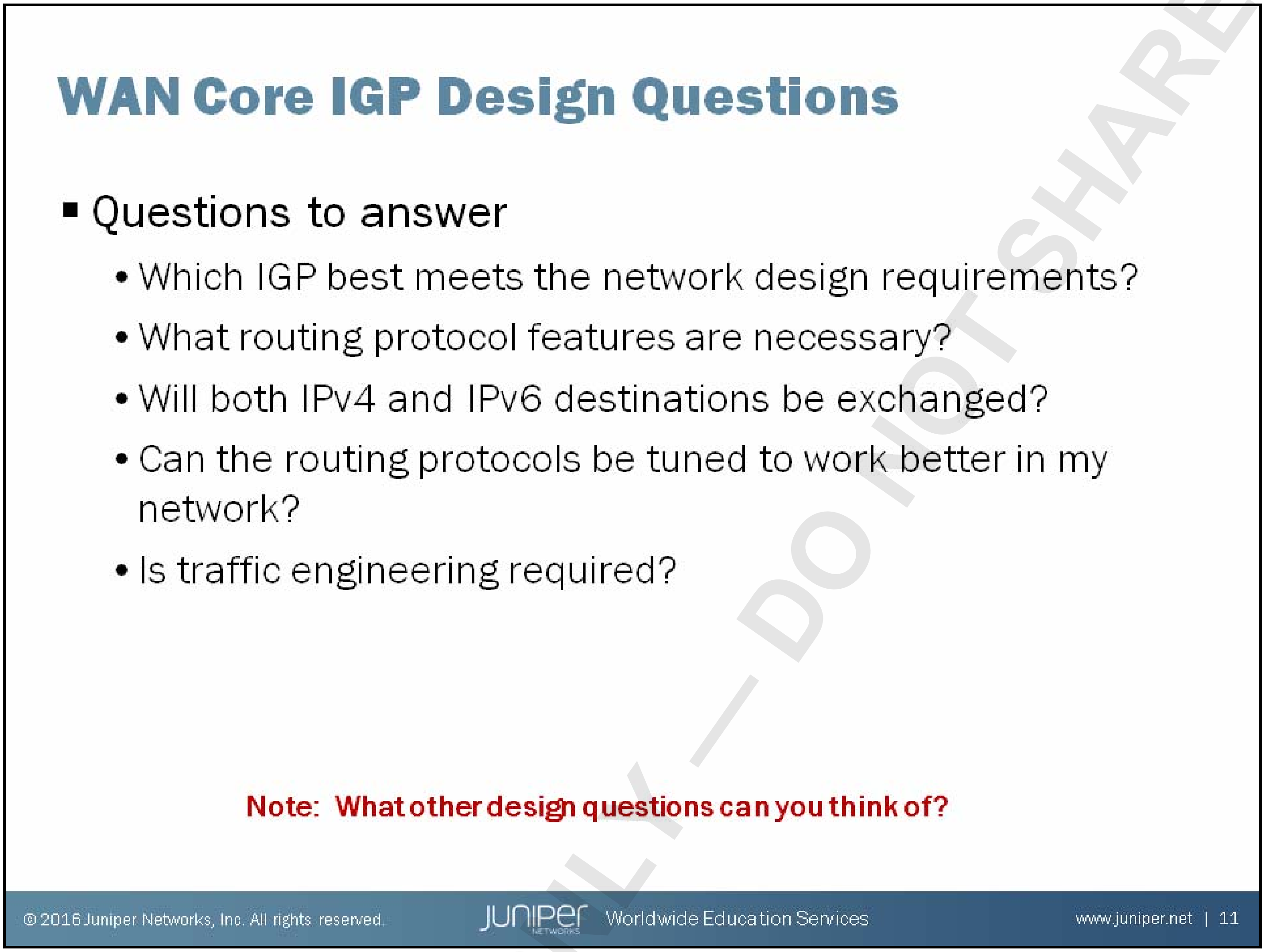
- WAN core routing considerations
  - Select an IGP for the WAN core
    - OSPFv2, OSPFv3, or both
    - IS-IS
    - Is IGP traffic engineering required?
  - BGP architecture design
    - Types of BGP peer relationships
    - Scalability and high availability
    - Autonomous system number selections
    - Determine the address families to be communicated

Worldwide Education Services

www.juniper.net | 10

Core Routing Overview

One of the first responsibilities for WAN core designers is to select an IGP that meets present and future core network design requirements. The IGP is the foundation for WAN routing and additional connectivity service offerings. Two versions of OSPF are available and one or both versions may be required depending on what type of addressing information needs to be communicated. Both OSPF and IS-IS routing protocols are capable of functioning as IGPs in the WAN core. This section examines design characteristics of both OSPF and IS-IS and how they are used in the WAN core.



WAN Core IGP Design Questions

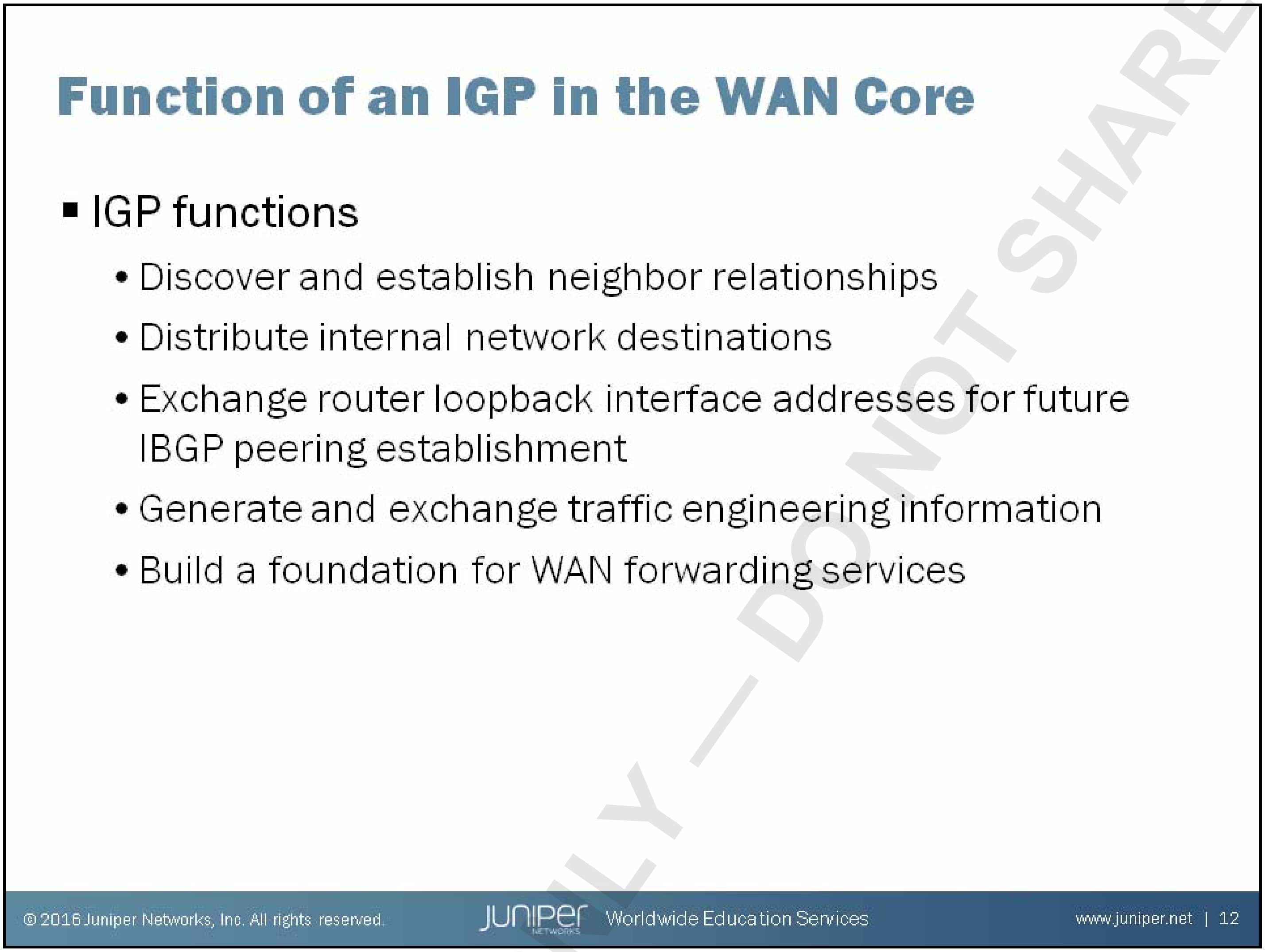
- Questions to answer
  - Which IGP best meets the network design requirements?
  - What routing protocol features are necessary?
  - Will both IPv4 and IPv6 destinations be exchanged?
  - Can the routing protocols be tuned to work better in my network?
  - Is traffic engineering required?

Note: What other design questions can you think of?

Juniper Networks Worldwide Education Services www.juniper.net | 11

Core IGP Design Questions

Congratulations, you have been tasked to lead the WAN design team for the core network. You might have a few questions bouncing around in your head. The equipment is in place and now everyone in the meeting is looking at you and your team. Which IGP will be selected as the routing foundation for the WAN? Which routing protocol has the capabilities and features necessary to offer a stable, robust, scalable solution? What addressing information needs to be communicated between IGP neighbors? Will services implemented across the WAN benefit from IGP distribution of traffic engineering information? This is a lot to think about. This section discusses IGP design concepts used by WAN network designers.



Function of an IGP in the WAN Core

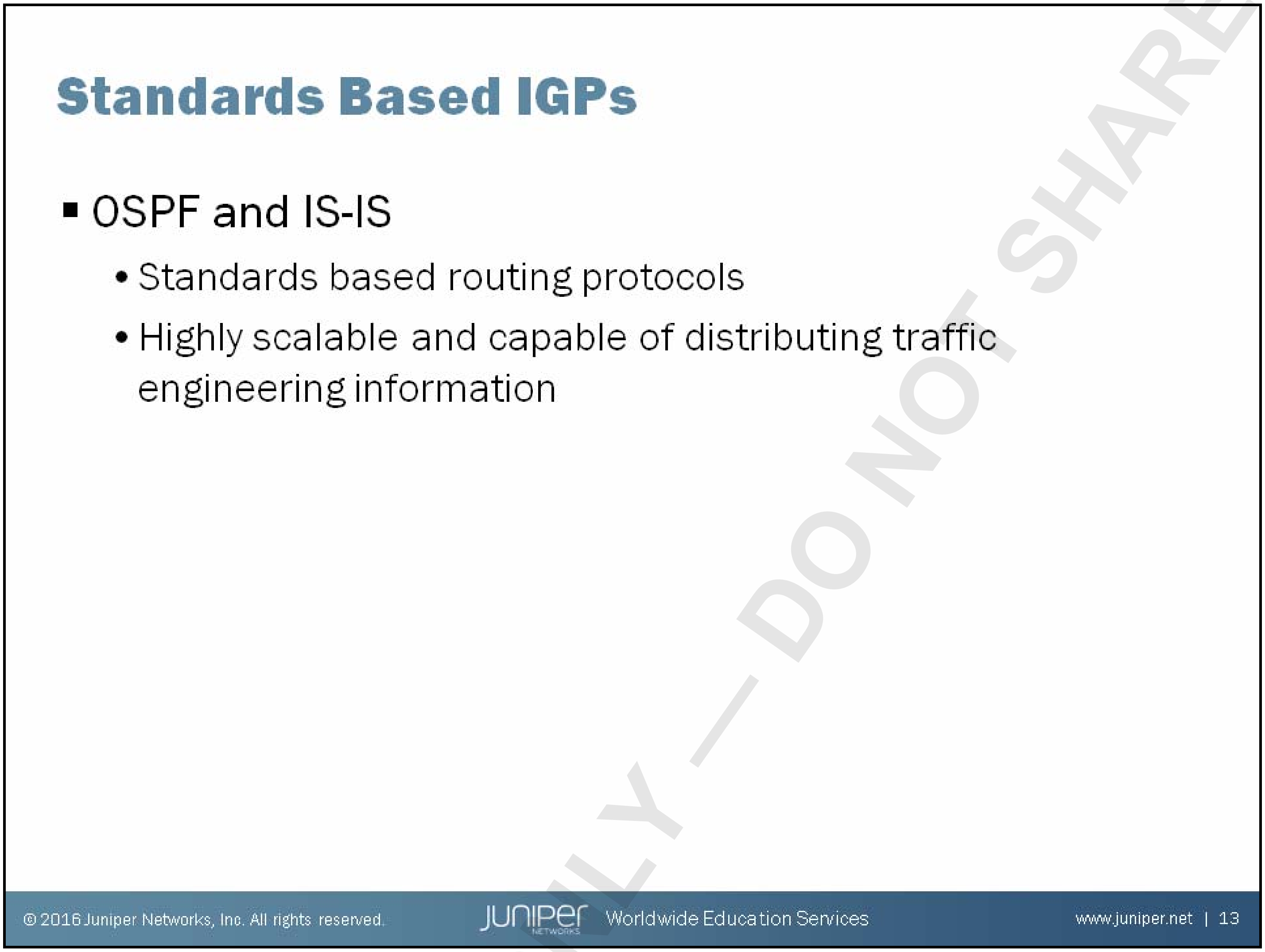
- IGP functions
  - Discover and establish neighbor relationships
  - Distribute internal network destinations
  - Exchange router loopback interface addresses for future IBGP peering establishment
  - Generate and exchange traffic engineering information
  - Build a foundation for WAN forwarding services

Worldwide Education Services

www.juniper.net | 12

WAN Core IGP Functions

IGPs are the routing foundation of the WAN core. Without proper IGP design other core services such as BGP or MPLS will not have the required foundation to be successfully implemented. BGP depends on the IGP to distribute WAN router loopback interface addresses permitting redundant IBGP peering to be configured. MPLS Layer2 and Layer3 VPNs leverage IGP traffic engineering information about the links that make up the WAN core. Traffic engineering data about core links is distributed by the IGP and stored in a database on each IGP node. A traffic engineering database gives network designers additional LSP routing options and allows granular control of MPLS traffic forwarding.



Standards Based IGPs

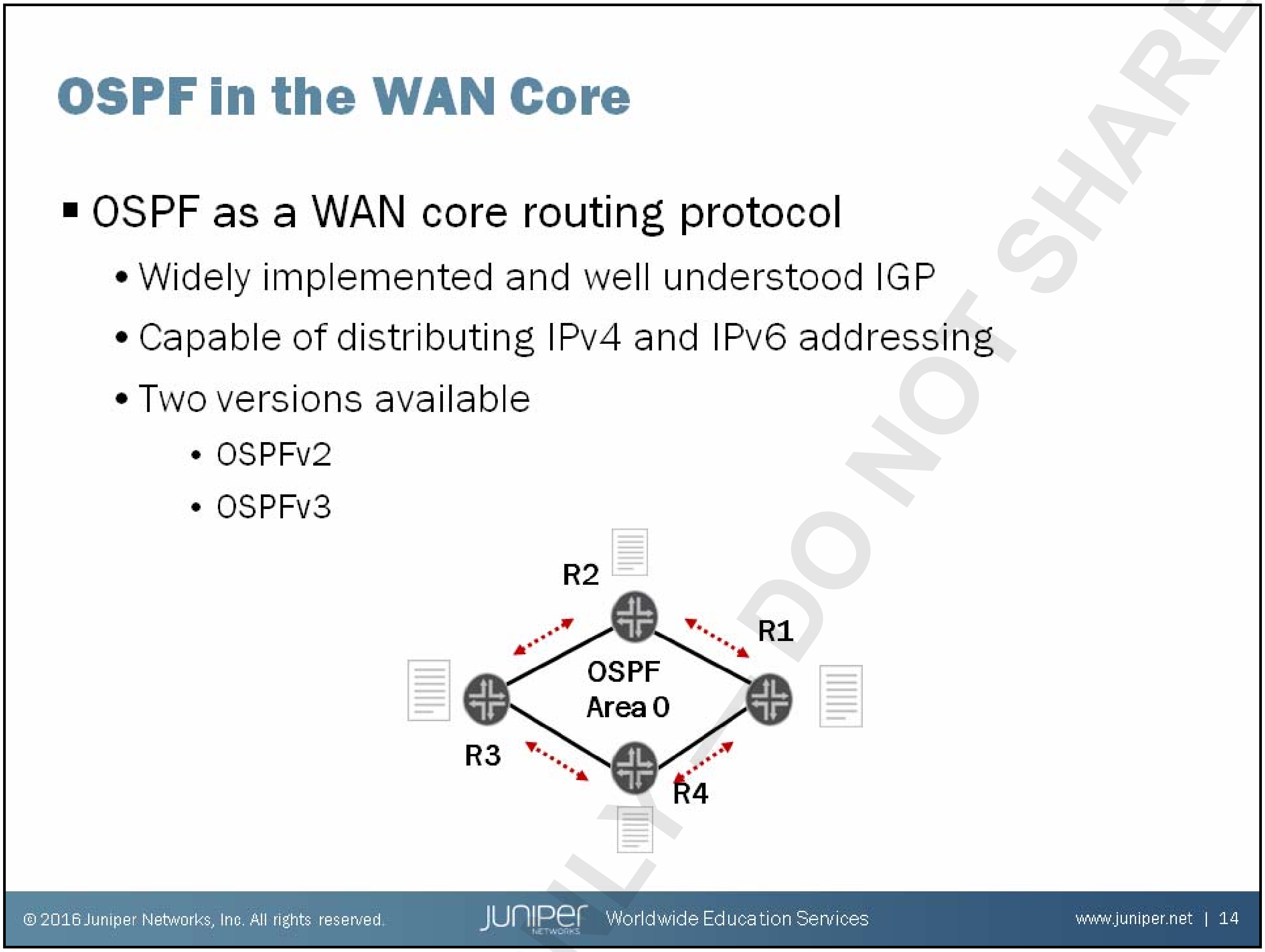
- OSPF and IS-IS
  - Standards based routing protocols
  - Highly scalable and capable of distributing traffic engineering information

© 2016 Juniper Networks, Inc. All rights reserved.

JUNIPer NETWORKS
Worldwide Education Services
www.juniper.net | 13

Standards Based IGPs

OSPF and IS-IS are two well known, widely implemented, standards based IGPs. Both routing protocols provide network designers the functionality necessary to design a successful core network.



OSPF in the WAN Core

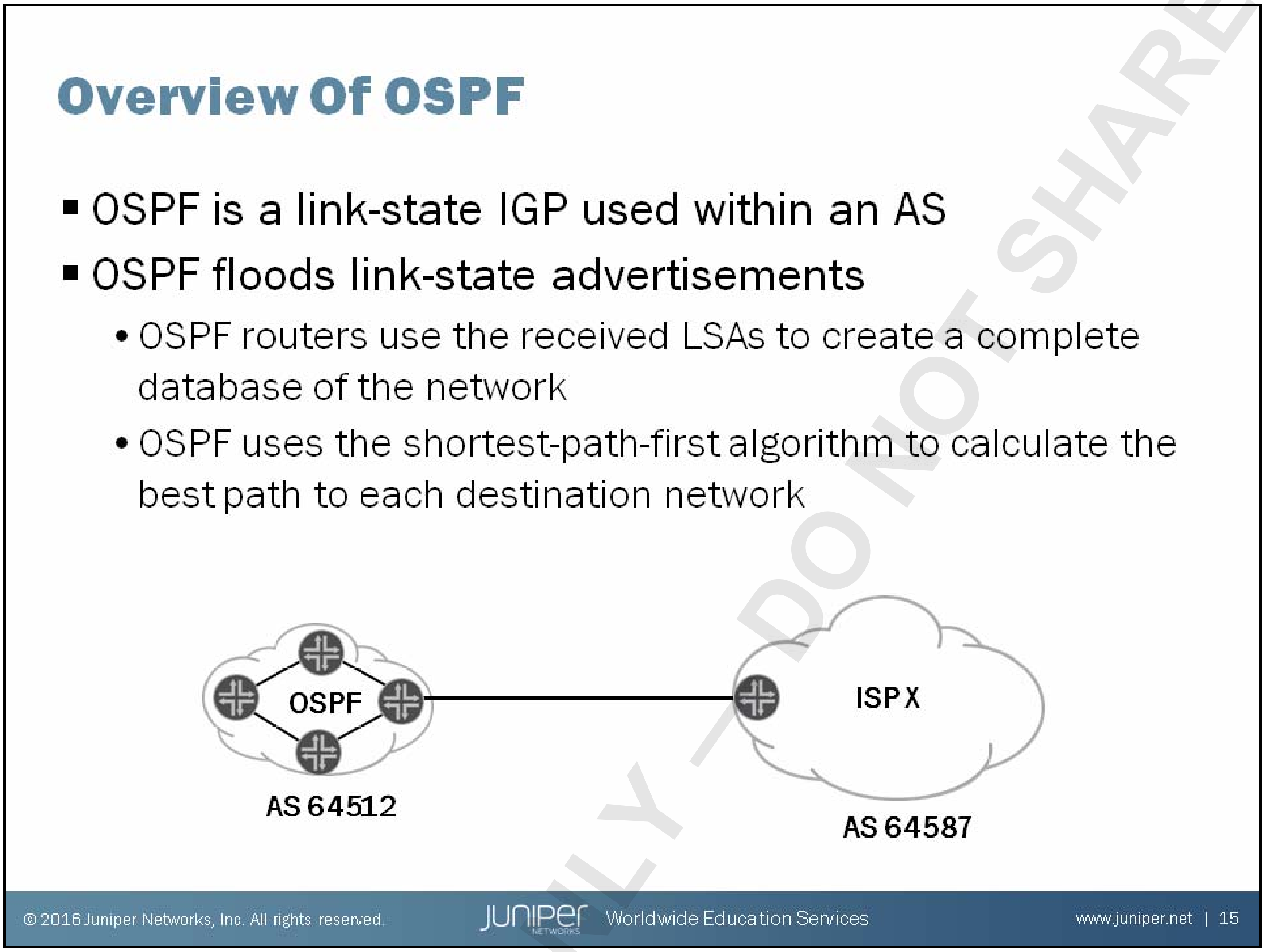
- OSPF as a WAN core routing protocol
  - Widely implemented and well understood IGP
  - Capable of distributing IPv4 and IPv6 addressing
  - Two versions available
    - OSPFv2
    - OSPFv3

JUNIPer NETWORKS Worldwide Education Services

www.juniper.net | 14

OSPF in the WAN Core

Service providers around the world implement OSPF as the core network IGP. Two versions of OSPF are available. OSPFv2 is a very popular, well understood, and widely implemented routing protocol that is capable of distributing IPv4 addressing information. OSPFv3 is a newer version of the protocol and supports communicating both IPv4 and IPv6 addressing.



Overview Of OSPF

- OSPF is a link-state IGP used within an AS
- OSPF floods link-state advertisements
  - OSPF routers use the received LSAs to create a complete database of the network
  - OSPF uses the shortest-path-first algorithm to calculate the best path to each destination network

OSPF
AS 64512
ISP X
AS 64587

Worldwide Education Services

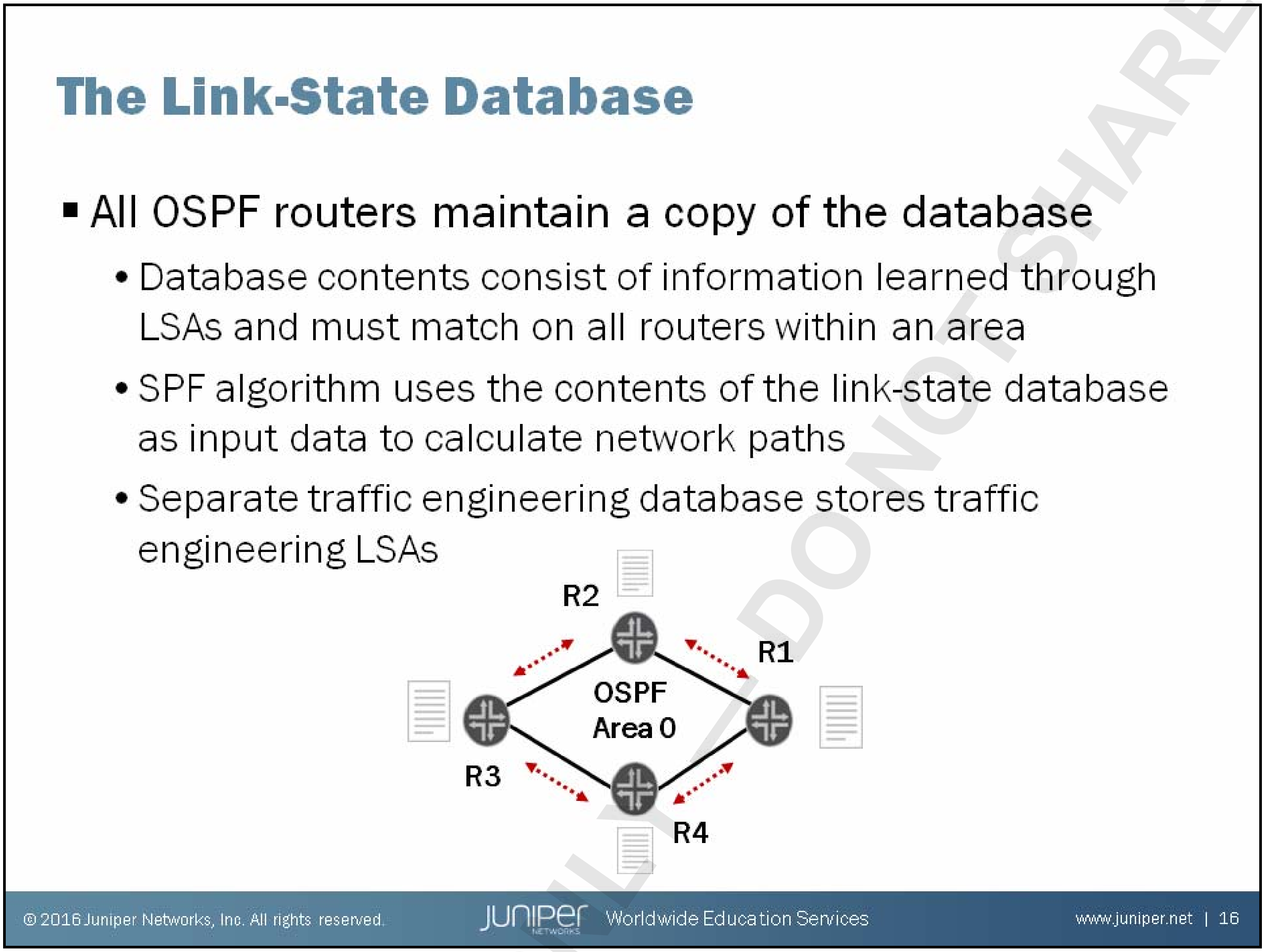
www.juniper.net | 15

OSPF Overview

OSPF is a link-state routing protocol designed for use within an autonomous system (AS). As a link-state IGP, OSPF allows for faster re-convergence, supports larger internetworks, and is less susceptible to bad routing information than distance-vector protocols.

Link State Advertisements

Once an OSPF router becomes neighbors with other OSPF routers, it can begin to share information about its attached networks. OSPF routers use link-state advertisements (LSAs) to reliably flood information about their network links and the state of those links to their neighboring OSPF routers. As link-state information is shared between OSPF routers, each OSPF router creates and maintains a link-state (or topological) database. The OSPF routers use the information provided within the LSAs as input for the shortest-path first (SPF) algorithm to calculate the best path for each destination prefix.



The Link-State Database

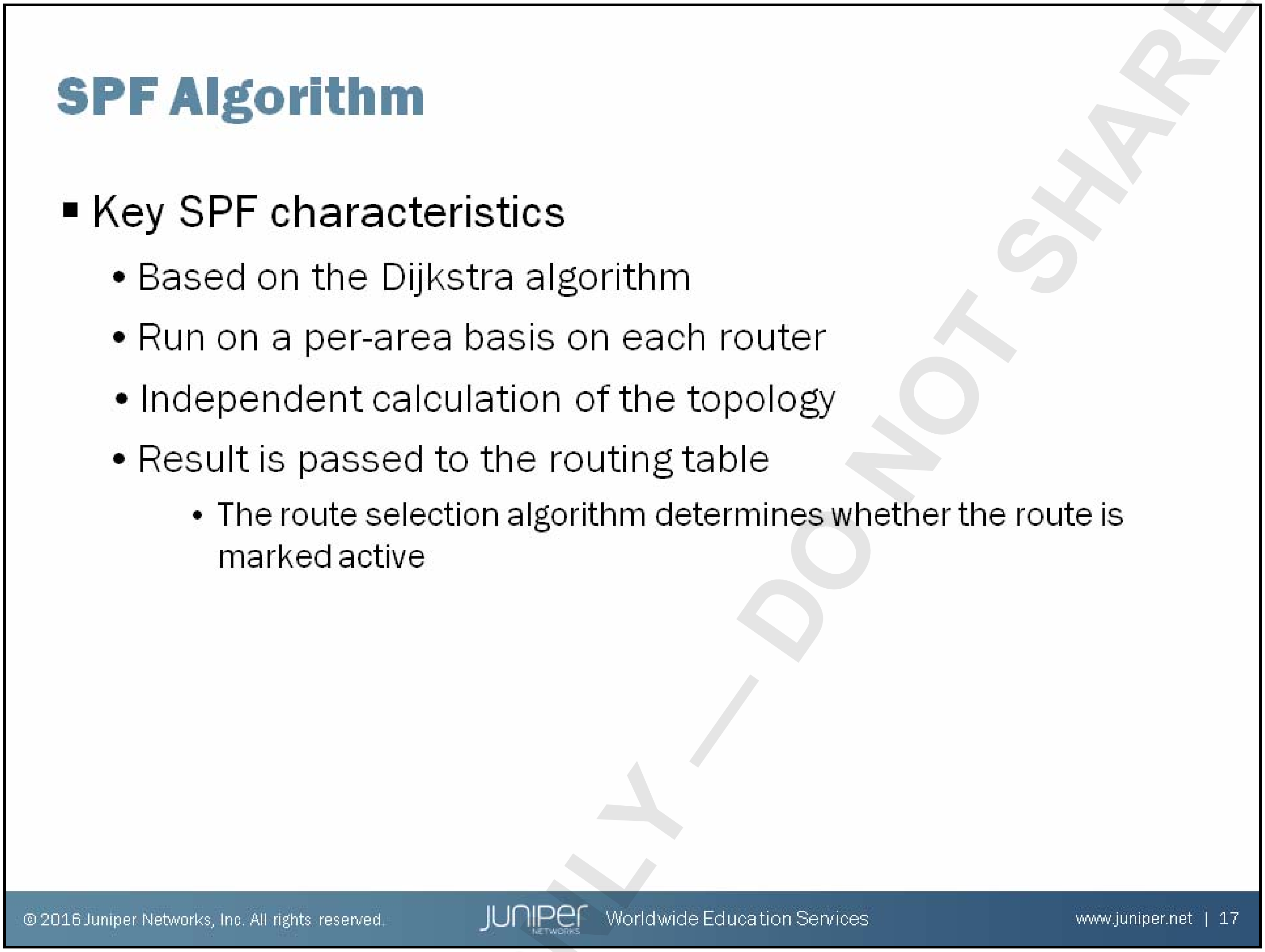
- All OSPF routers maintain a copy of the database
  - Database contents consist of information learned through LSAs and must match on all routers within an area
  - SPF algorithm uses the contents of the link-state database as input data to calculate network paths
  - Separate traffic engineering database stores traffic engineering LSAs

R2
R1
OSPF
Area 0
R3
R4

Juniper Networks Worldwide Education Services www.juniper.net | 16

The Link-State Database

In addition to discovering neighbors and flooding LSAs, a third major task of the OSPF protocol is to create and maintain the link-state database (LSDB). The link-state, or topological, database stores the LSAs as a series of records. The contents stored within the LSDB include details such as the advertising router's ID, its attached networks and neighboring routers, and the cost associated with those networks or neighbors. According to the OSPF RFC, each router in an OSPF area must have an identical LSDB to ensure accurate routing knowledge. The information recorded in the database is used as input data to calculate the shortest paths (that is, least-cost paths) for all destination prefixes within the network.



SPF Algorithm

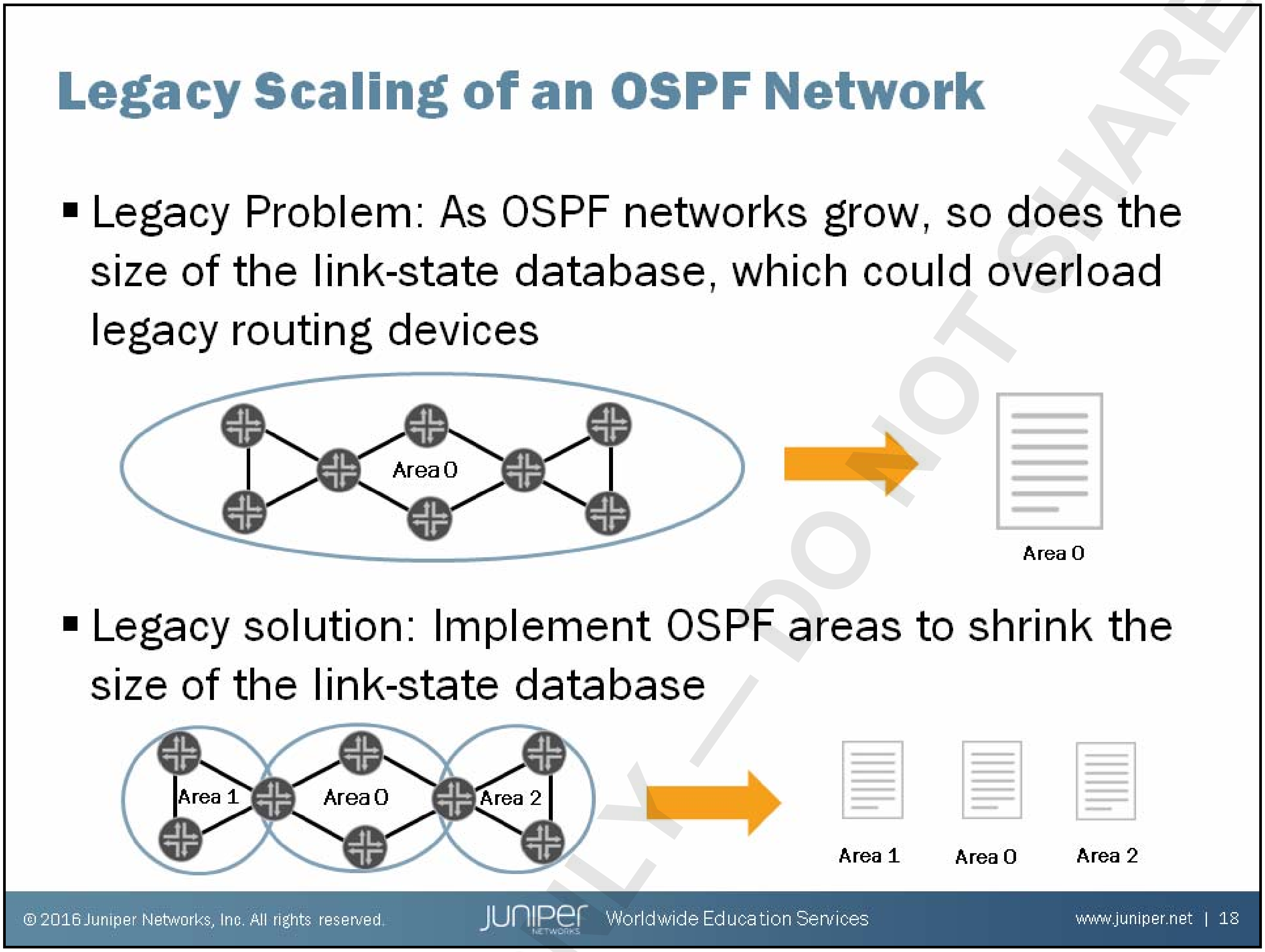
- Key SPF characteristics
  - Based on the Dijkstra algorithm
  - Run on a per-area basis on each router
  - Independent calculation of the topology
  - Result is passed to the routing table
    - The route selection algorithm determines whether the route is marked active

Worldwide Education Services

www.juniper.net | 17

SPF Algorithm

The information stored in the LSDB is used as input data to calculate the shortest paths (that is, least-cost paths) for all destination prefixes within the network. OSPF uses the SPF algorithm (also known as the Dijkstra algorithm) to calculate, all at once, the shortest paths to all destinations. It performs this calculation by creating a tree of shortest paths incrementally and picking the best candidate from that tree. The results of this calculation are then handed to the router's routing table for the actual forwarding of data packets. For the purposes of this course, the SPF and Dijkstra algorithm are considered the same.



Legacy Scaling of an OSPF Network

- Legacy Problem: As OSPF networks grow, so does the size of the link-state database, which could overload legacy routing devices

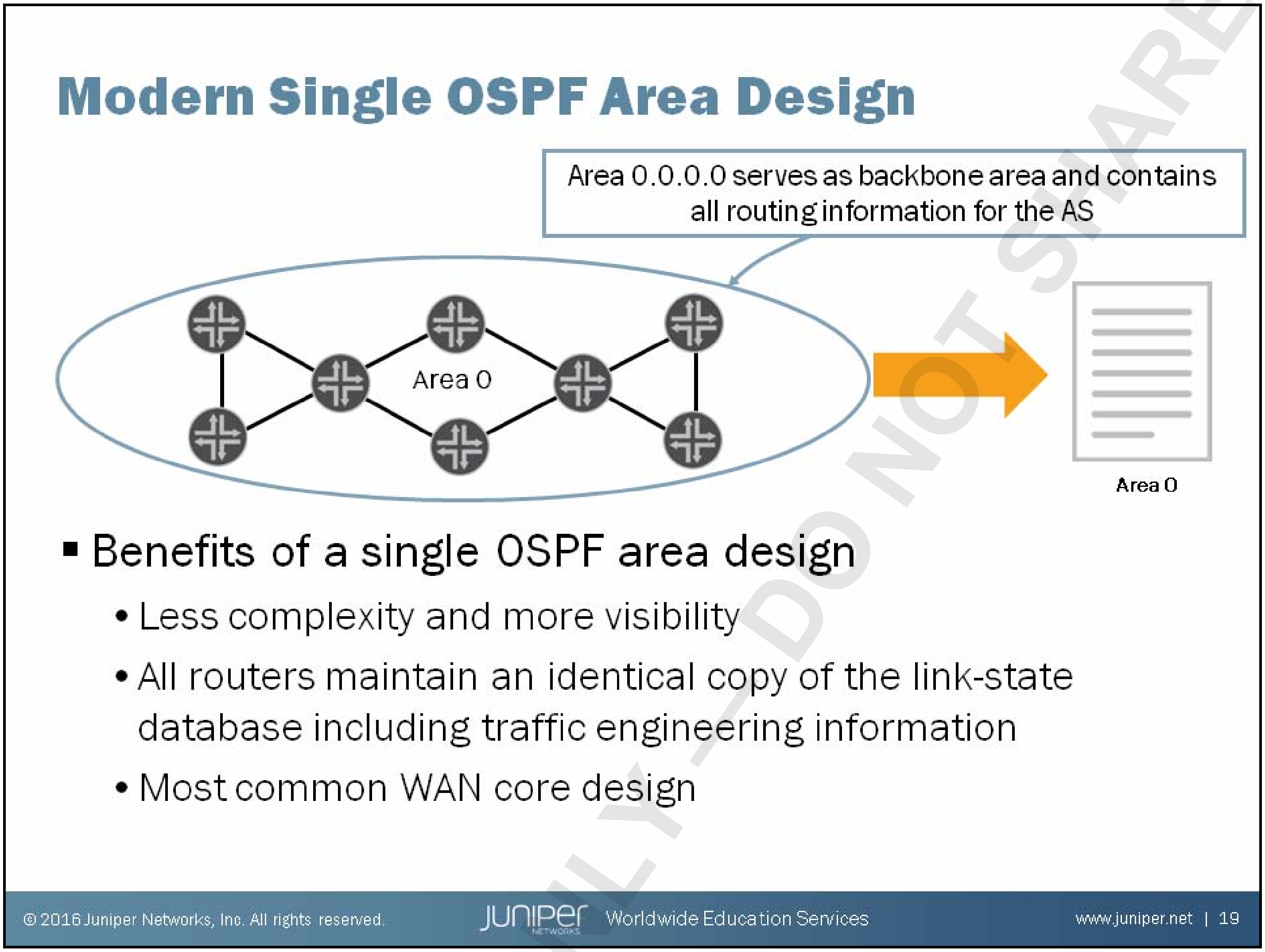
- Legacy solution: Implement OSPF areas to shrink the size of the link-state database

Legacy OSPF Scaling Problem

With a link-state protocol, flooding link-state update packets and running the SPF algorithm consume router resources. As the size of the network grows and more routers are added to the AS, this use of resources becomes a bigger and bigger issue. This issue stems from the RFC requirement that all OSPF routers share an identical LSDB. Eventually, the routers spend so much time flooding and running the SPF algorithm that they cannot route data packets. Clearly, this scenario is not optimal.

Legacy OSPF Scaling Solution

The solution to this problem (and link-state protocol scalability in general) is to reduce the size of the LSDB. You can reduce size of the LSDB using multiple OSPF areas rather than a single area that encompasses the entire AS. Note that the type of OSPF areas used is important when attempting to shrink the size of the LSDB. In addition to adding new OSPF areas that restrict LSA flooding, you can also perform route summarization on the borders between OSPF areas. Route summarization has two key benefits: 1) it reduces the size of the LSDB; and 2) it can hide some instabilities in one OSPF area from all other OSPF areas. For route summarization to be effective, you must carefully plan the addressing within your OSPF network so that subnets can be more easily summarized.



Modern Single OSPF Area Design

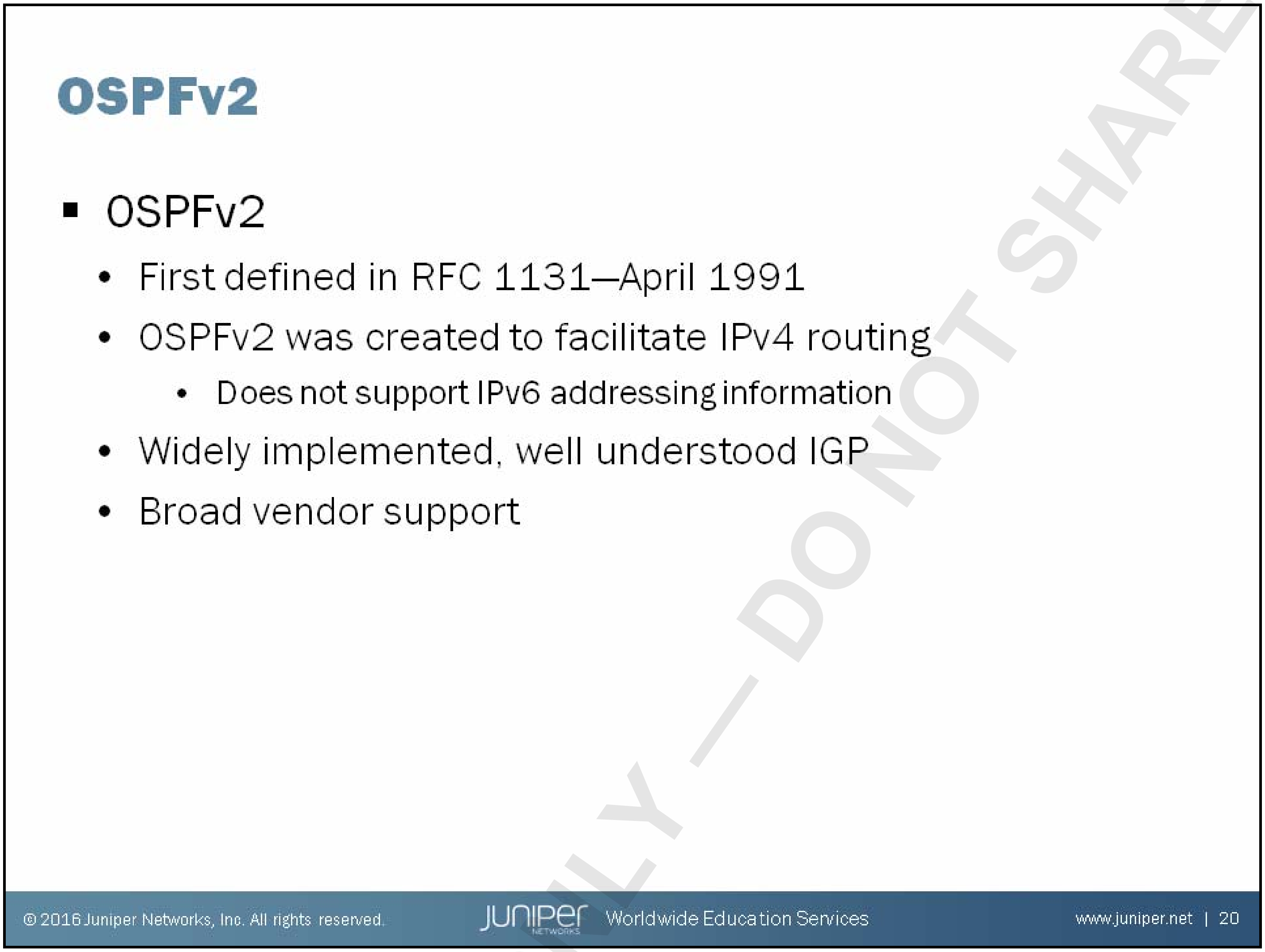
Area 0.0.0.0 serves as backbone area and contains all routing information for the AS

- Benefits of a single OSPF area design
  - Less complexity and more visibility
  - All routers maintain an identical copy of the link-state database including traffic engineering information
  - Most common WAN core design

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 19

Modern Single Area OSPF Design

Limitations on the number of routers in a single area were always a restriction imposed because of the limited amounts of memory and processing power available in legacy routing devices. Modern WAN core routers do not suffer from the same resource limitations as their historical counterparts. Most WAN providers try to limit the OSPF design to a single area. Maintaining a single OSPF area reduces network complexity and creates one LSDB that contains all routing information for the entire AS. Traffic engineering data is not flooded across area boundaries making implementing end to end traffic engineering and control more difficult. For label-switched networks, such as MPLS, most existing traffic engineering solutions require a single routing domain. These solutions do not work when a route from the ingress node to the egress node leaves the routing area or AS of the ingress node. In such cases, the path computation becomes complicated because of the unavailability of the complete routing information throughout the network. Service providers usually choose not to leak routing information beyond the routing area or AS for scalability constraints and confidentiality concerns.



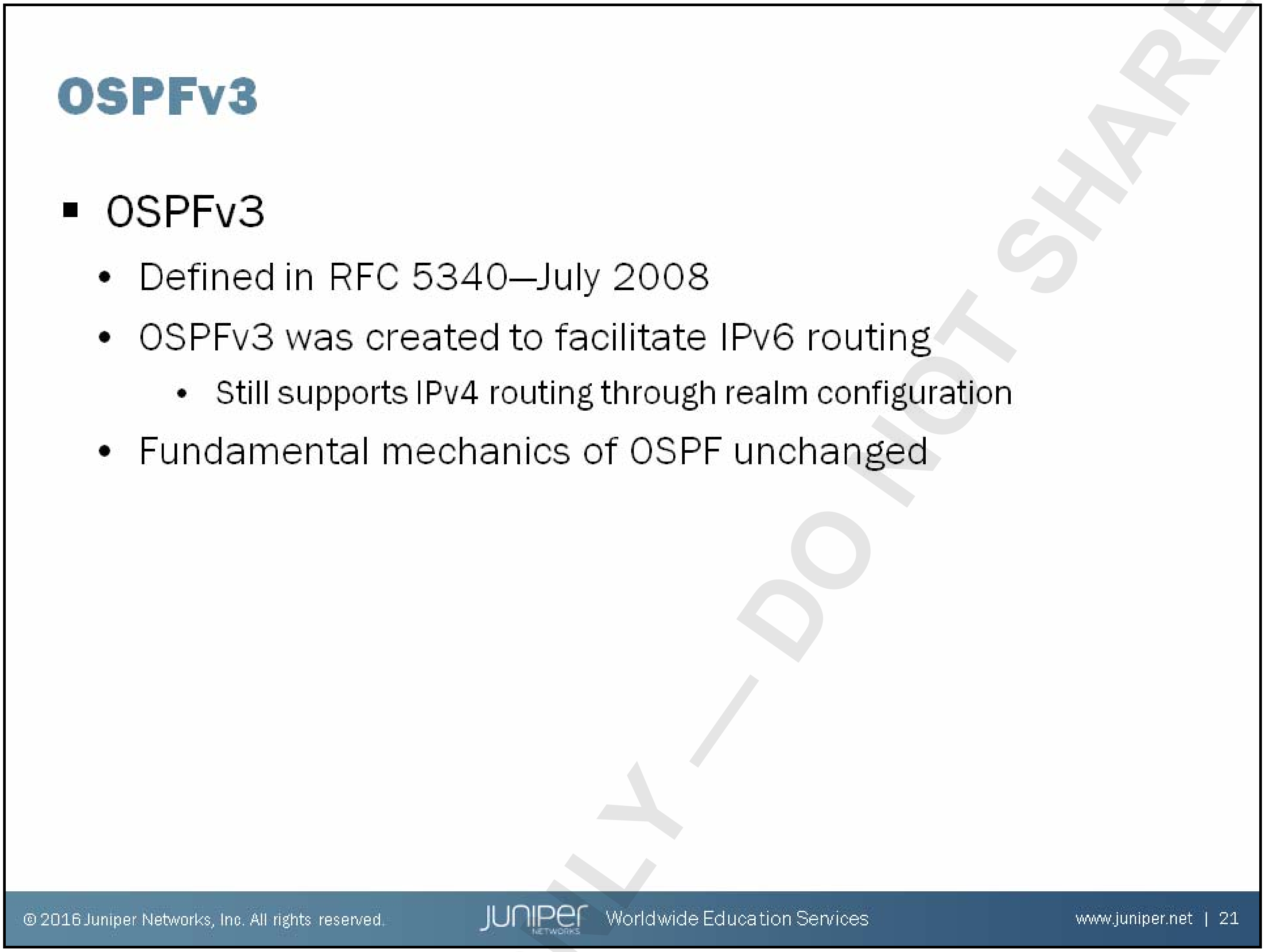
OSPFv2

- OSPFv2
  - First defined in RFC 1131—April 1991
  - OSPFv2 was created to facilitate IPv4 routing
    - Does not support IPv6 addressing information
  - Widely implemented, well understood IGP
  - Broad vendor support

JUNIPer NETWORKS
Worldwide Education Services
www.juniper.net | 20

OSPFv2

OSPFv2 was first defined in RFC 1131 which describes the mechanisms of OSPF, including LSA flooding scopes, areas, designated router election, stub areas, NSSAs, and so on. OSPFv2 is associated with IPv4 addressing. OSPFv2 does not support IPv6 addressing but it can be configured to communicate link characteristics in support of traffic engineering. OSPFv2 is one of the most widely implemented, well understood routing protocols and is supported by most routing vendors.



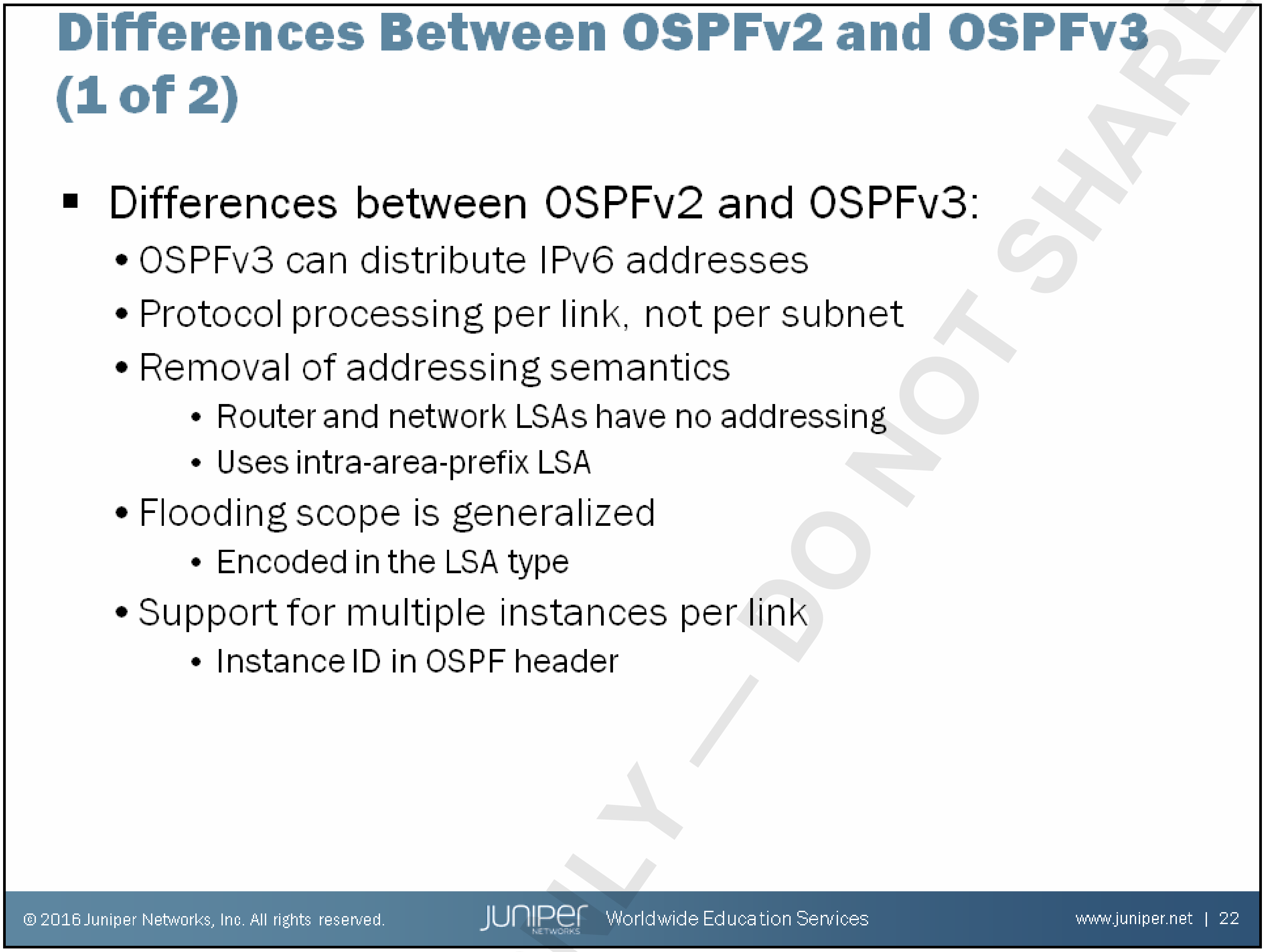
OSPFv3

- OSPFv3
  - Defined in RFC 5340—July 2008
  - OSPFv3 was created to facilitate IPv6 routing
    - Still supports IPv4 routing through realm configuration
  - Fundamental mechanics of OSPF unchanged

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 21

OSPFv3

OSPFv3 is defined in RFC 5340, with some additional features, such as grateful restart and authentication, defined in separate documents (RFC 5781—OSPFv3 Graceful Restart and RFC 4552— Authentication/Confidentiality for OSPFv3). OSPFv3 maintains the fundamental mechanisms of OSPF, including LSA flooding scopes, areas, designated router election, stub areas, NSSAs, and so on. Though OSPFv3 is often associated with IPv6 addressing, it is also completely compatible with IPv4 addressing schemes. However, realm configuration is necessary to account for the differences in IPv4 versus IPv6 addressing.



Differences Between OSPFv2 and OSPFv3 (1 of 2)

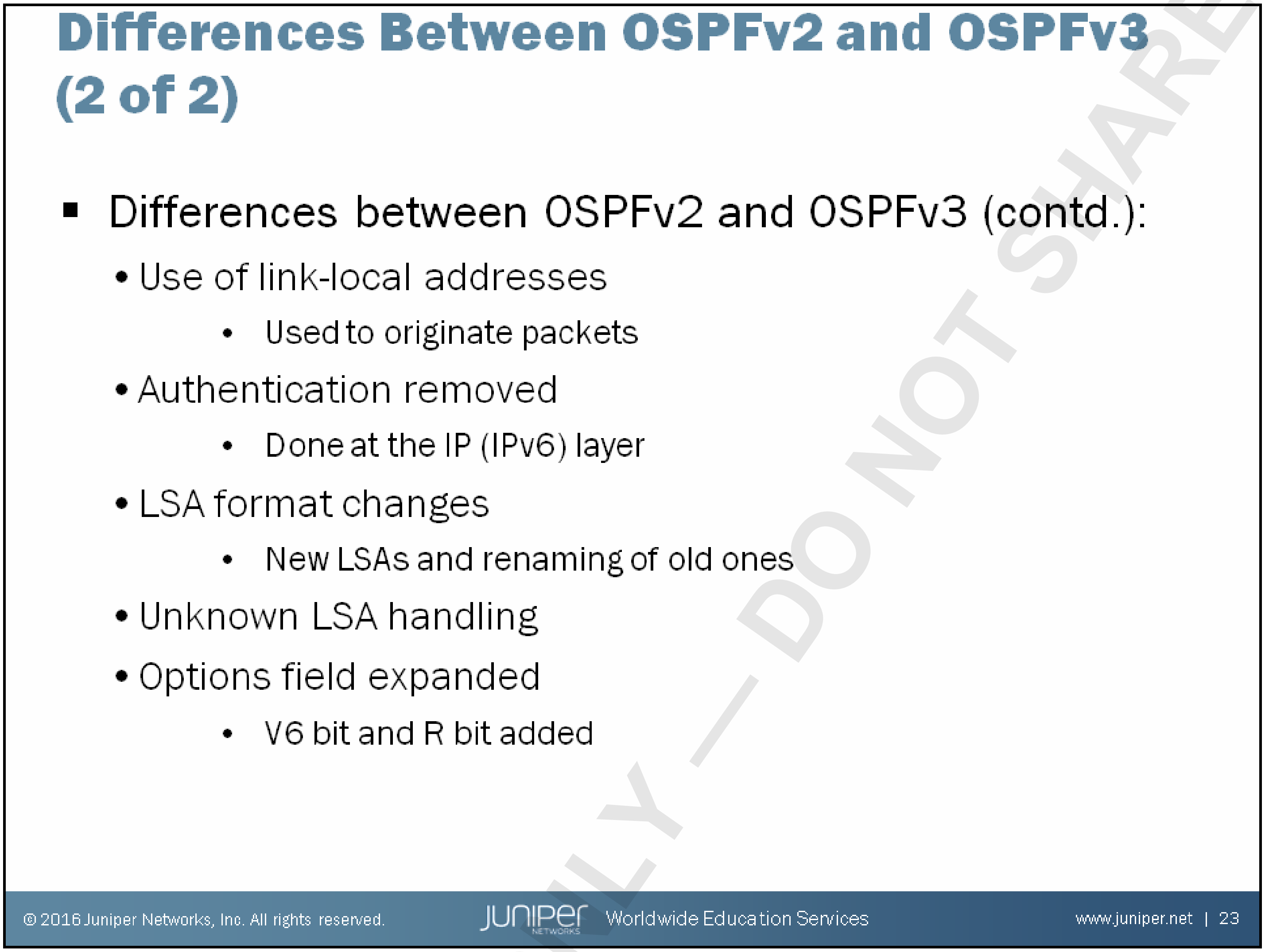
- Differences between OSPFv2 and OSPFv3:
  - OSPFv3 can distribute IPv6 addresses
  - Protocol processing per link, not per subnet
  - Removal of addressing semantics
    - Router and network LSAs have no addressing
    - Uses intra-area-prefix LSA
  - Flooding scope is generalized
    - Encoded in the LSA type
  - Support for multiple instances per link
    - Instance ID in OSPF header

Worldwide Education Services www.juniper.net | 22

Differences Between OSPFv2 and OSPFv3: Part 1

Though much remains the same, several differences exist between OSPFv2 and OSPFv3:

- Protocol processing per link, not per subnet: You need only a single adjacency per link even if multiple IPv6 subnets exist on the link. These adjacencies are formed using link-local addresses, which removes the requirement of having matching subnet and subnet mask configurations for two routers to become adjacent.
- Removal of addressing semantics: Addressing semantics were removed from OSPF headers, and no prefix information is carried in the router LSAs or network LSAs. This change makes OSPFv3 somewhat protocol independent. The router and network LSAs now express the topology in a protocol-independent fashion. To carry the equivalent information that was carried in these LSAs with OSPFv2, a new LSA called the intra-area-prefix LSA is introduced. The intra-area-prefix LSA carries the IPv6 addressing information.
- Flooding scope is generalized: Flooding is now generalized and is coded into the LS type field. An LSA can be either flooded on the local link, area, or throughout the AS. This flooding also assists in the handling of unknown LSA types as the flood scope is encoded.
- Support for multiple instances per link: This support allows multiple OSPF instances to run over a single link to support separate routing domains or to support a single link belonging to multiple areas. Previously, this functionality was achieved in OSPFv2 by tweaking authentication to hide OSPF packets from an OSPF router. This functionality is now formalized by including an instance ID in the OSPF header.



Differences Between OSPFv2 and OSPFv3 (2 of 2)

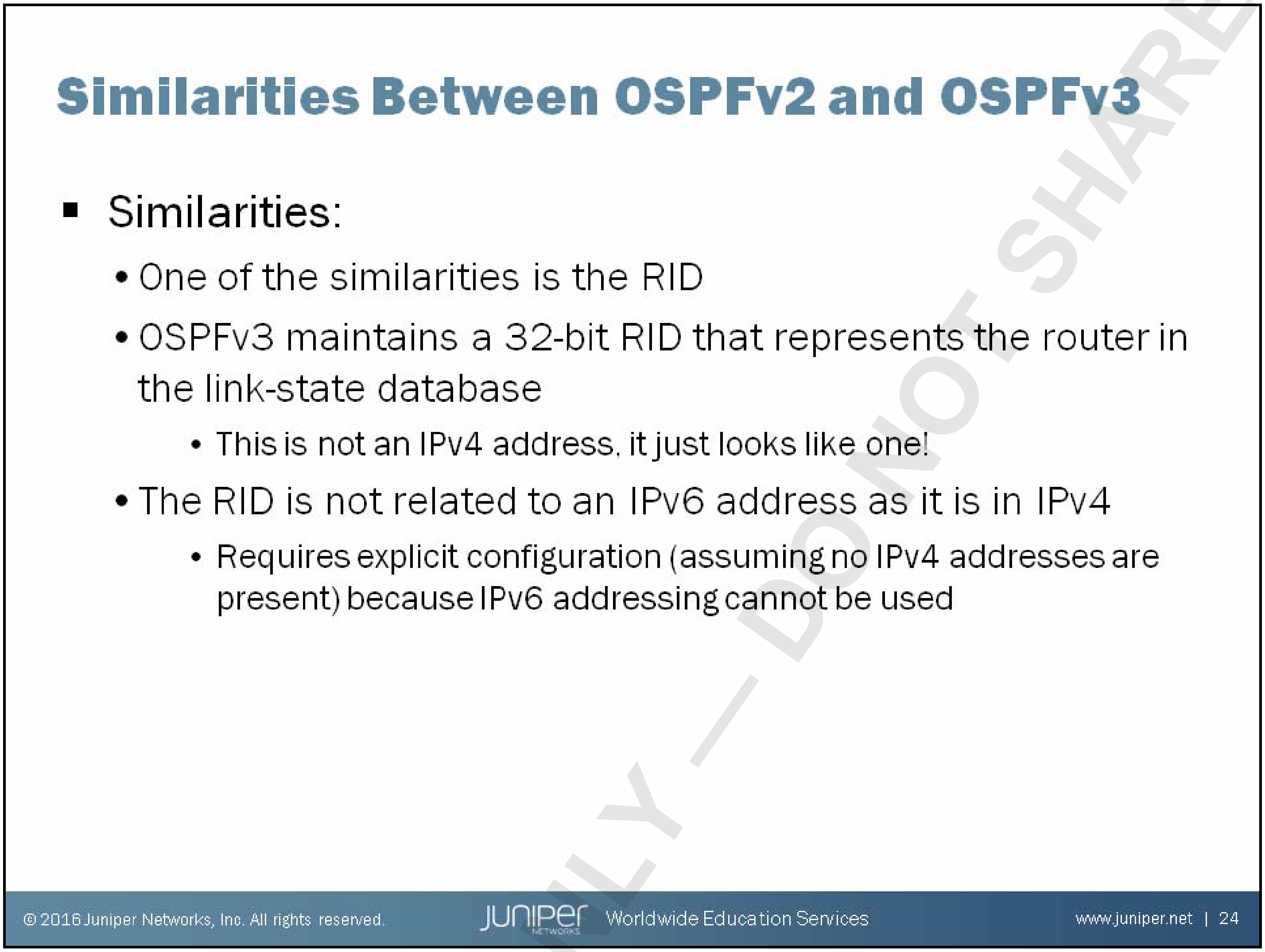
- Differences between OSPFv2 and OSPFv3 (contd.):
  - Use of link-local addresses
    - Used to originate packets
  - Authentication removed
    - Done at the IP (IPv6) layer
  - LSA format changes
    - New LSAs and renaming of old ones
  - Unknown LSA handling
  - Options field expanded
    - V6 bit and R bit added

© 2016 Juniper Networks, Inc. All rights reserved. JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 23

Differences Between OSPFv2 and OSPFv3: Part 2

Further differences between OSPFv2 and OSPFv3:

- Use of link-local addresses: OSPF packets are sent using the IPv6 link-local addressing. These link-local addresses are also used as next-hop information during forwarding.
- Authentication removed: Authentication uses the IPsec framework built into IPv6. As a result, it is not required at the application layer and was removed.
- LSA format changes: Summary LSAs are now referred to as inter-area-prefix LSAs, and ASBR summary LSAs are now inter-area-router LSAs to account for differences in address size. The intra-area-prefix LSA is also included, carrying prefix information internal to areas that were previously carried inside router and network LSAs.
- Unknown LSA handling: The old way of discarding unknown LSA types is no longer supported because a mix of capabilities in a network, especially on a single link, causes forwarding issues. OSPFv3 codes the proper handling of an unknown LSA type in the LSA handing bit. The LSA is either treated as being of local scope only or stored and forwarded as if it were understood.
- *The options field was expanded from 8 bits to 24 bits: The option field is included in* OSPF hello packets, database description packets, and certain LSAs (router LSAs, network LSAs, inter-area-router LSAs, and link LSAs). Previously defined option bits are present, as well as added support for the V6-bit and R bits. The V6-bit is used to indicate whether the route or link should be excluded from IPv6 routing calculations. The R bit is used like the IS-IS overload bit and indicates whether the originator is an active router. If the R bit is clear (that is, 0) in the OSPF options field, the advertising router can participate in OSPF without being used for transit traffic. This would be a useful setting for hosts that are multihomed but never used to forward traffic between interfaces.



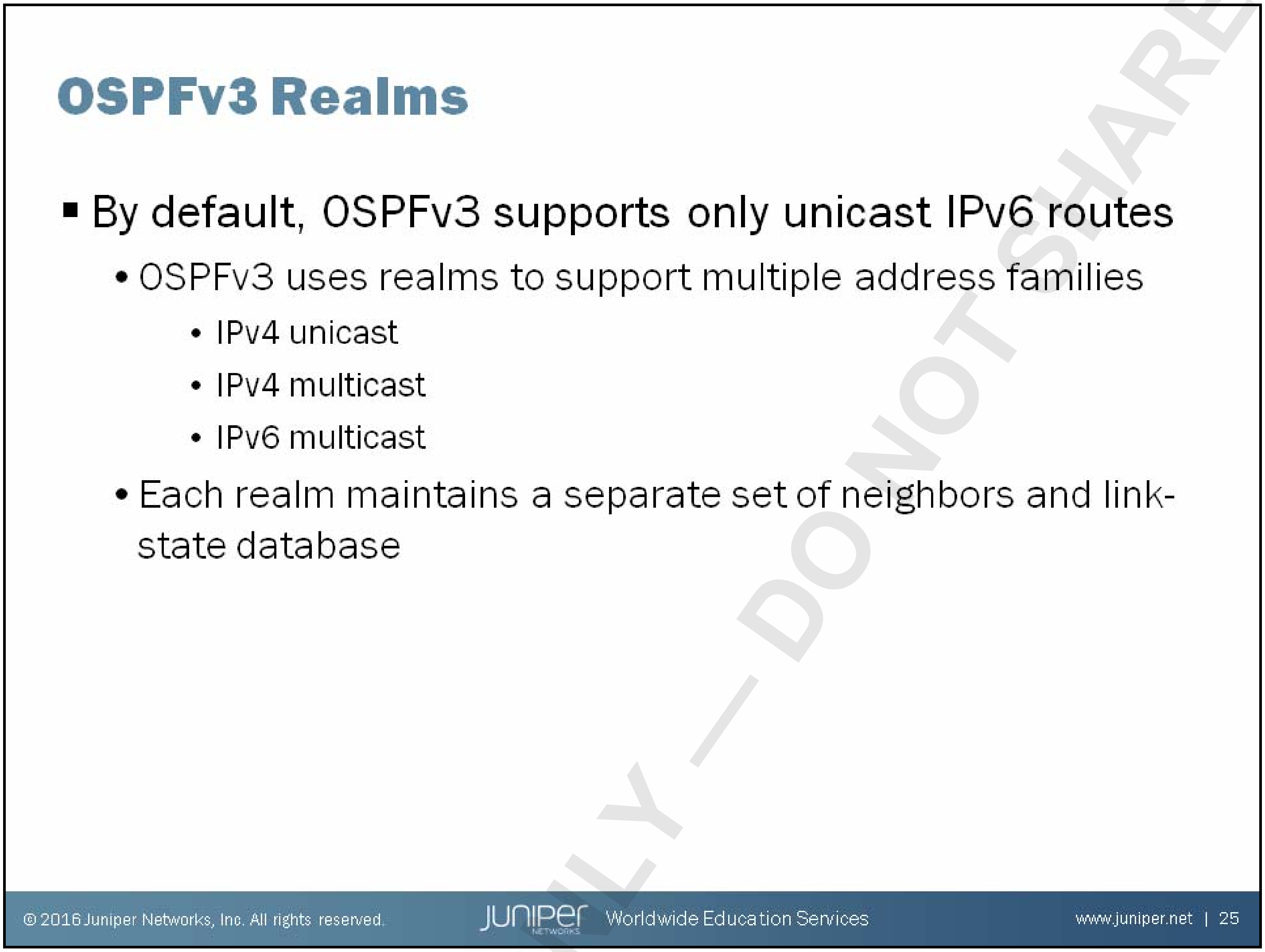
Similarities Between OSPFv2 and OSPFv3

- Similarities:
  - One of the similarities is the RID
  - OSPFv3 maintains a 32-bit RID that represents the router in the link-state database
    - This is not an IPv4 address, it just looks like one!
  - The RID is not related to an IPv6 address as it is in IPv4
    - Requires explicit configuration (assuming no IPv4 addresses are present) because IPv6 addressing cannot be used

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 24

Similarities to OSPFv2

One similarity is the preservation of the 32-bit router ID. Like OSPF, in OSPFv3 every OSPF router has a single RID; it is a 32-bit number in dotted quad notation. If the RID is not explicitly configured under routing-options, the Junos operating system (Junos OS) uses IPv4 addressing to derive a RID based upon the same rules as OSPFv2 (that is, it is likely the loopback address). However, it is conceivable that IPv6 devices be deployed without any IPv4 addresses with which to derive the RID, and so we recommend configuring the router ID explicitly.



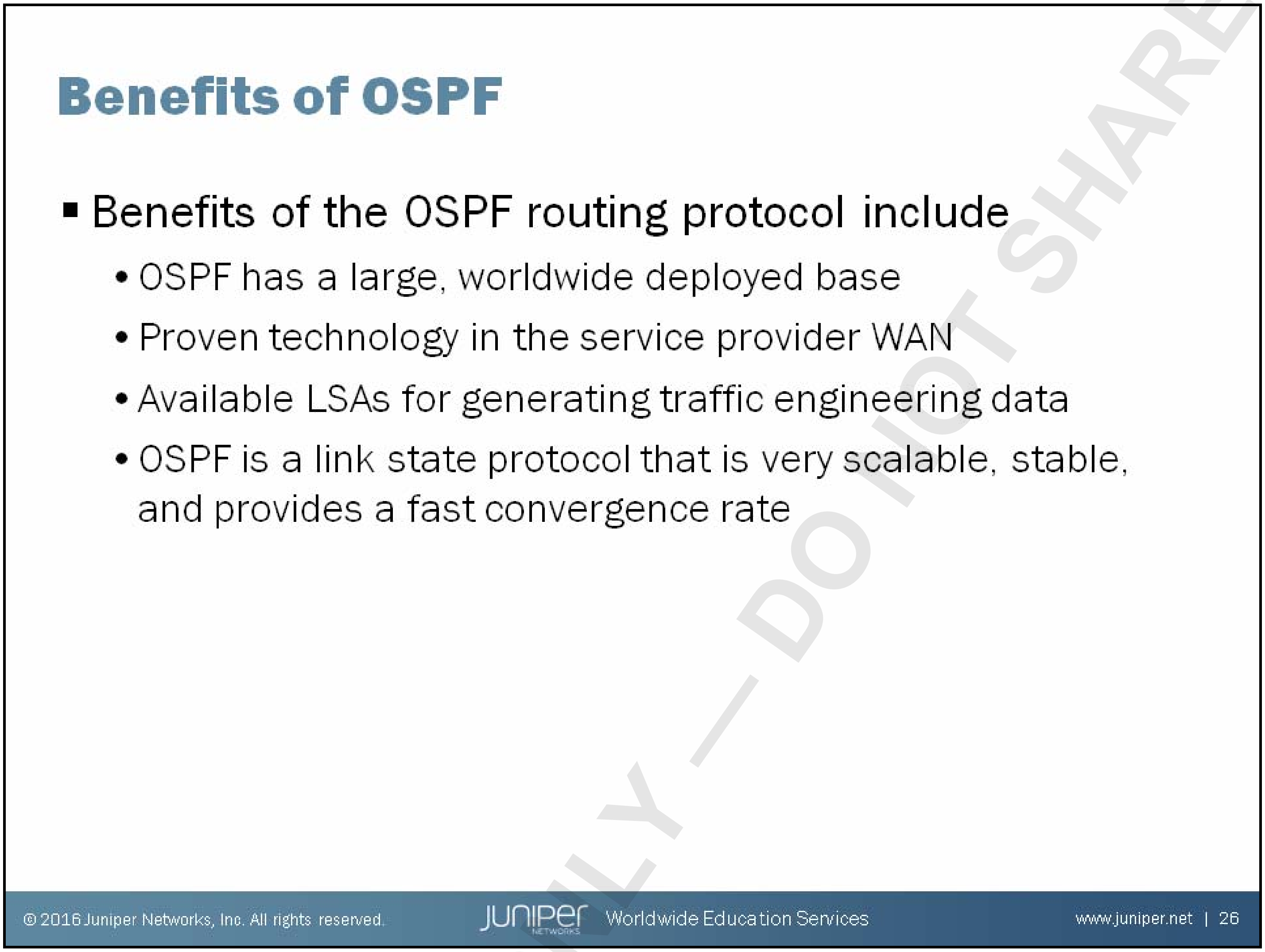
OSPFv3 Realms

- By default, OSPFv3 supports only unicast IPv6 routes
  - OSPFv3 uses realms to support multiple address families
    - IPv4 unicast
    - IPv4 multicast
    - IPv6 multicast
  - Each realm maintains a separate set of neighbors and link-state database

JUNIPER NETWORKS
Worldwide Education Services
www.juniper.net | 25

OSPFv3 Realms

By default, OSPFv3 supports only unicast IPv6 routes. However, you can configure OSPFv3 to support multiple address families, including IPv4 unicast, IPv4 multicast, and IPv6 multicast. The Junos OS maps each address family to a separate realm. Each realm maintains a separate set of neighbors and link-state database. You configure each realm independently.



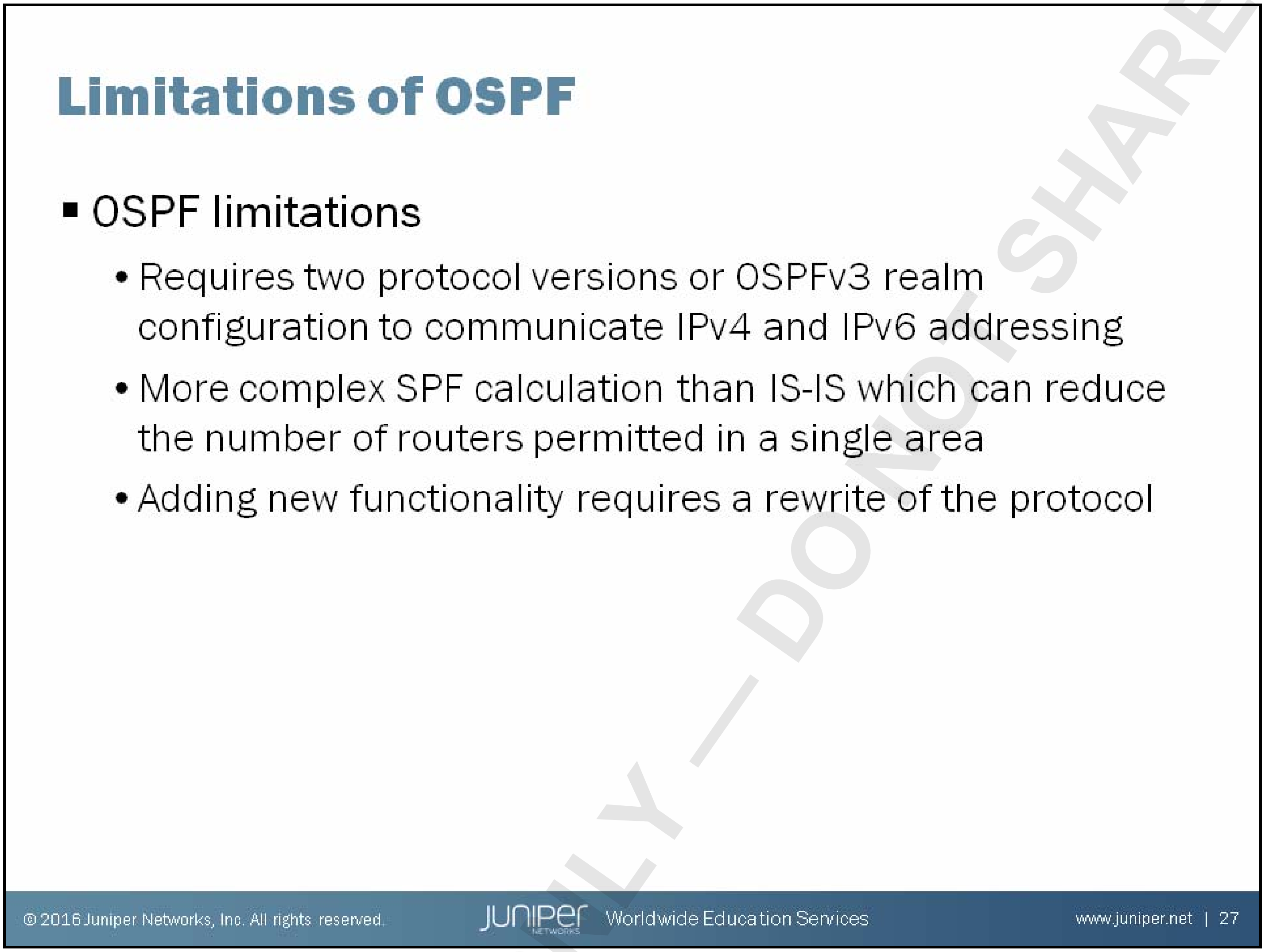
Benefits of OSPF

- Benefits of the OSPF routing protocol include
  - OSPF has a large, worldwide deployed base
  - Proven technology in the service provider WAN
  - Available LSAs for generating traffic engineering data
  - OSPF is a link state protocol that is very scalable, stable, and provides a fast convergence rate

Juniper Networks Worldwide Education Services www.juniper.net | 26

Benefits of OSPF

OSPF provides multiple benefits to WAN IGP designers. OSPF has a large worldwide deployment base in the service provider WAN core and is a well understood standards based routing protocol. OSPFv2 is capable of communicating IPv4 addresses. OSPFv3 can communicate both IPv4 and IPv6 addressing information. OSPF is a link state protocol that is scalable, stable, and provides a fast convergence rate. OSPF is capable of communicating traffic engineering data.



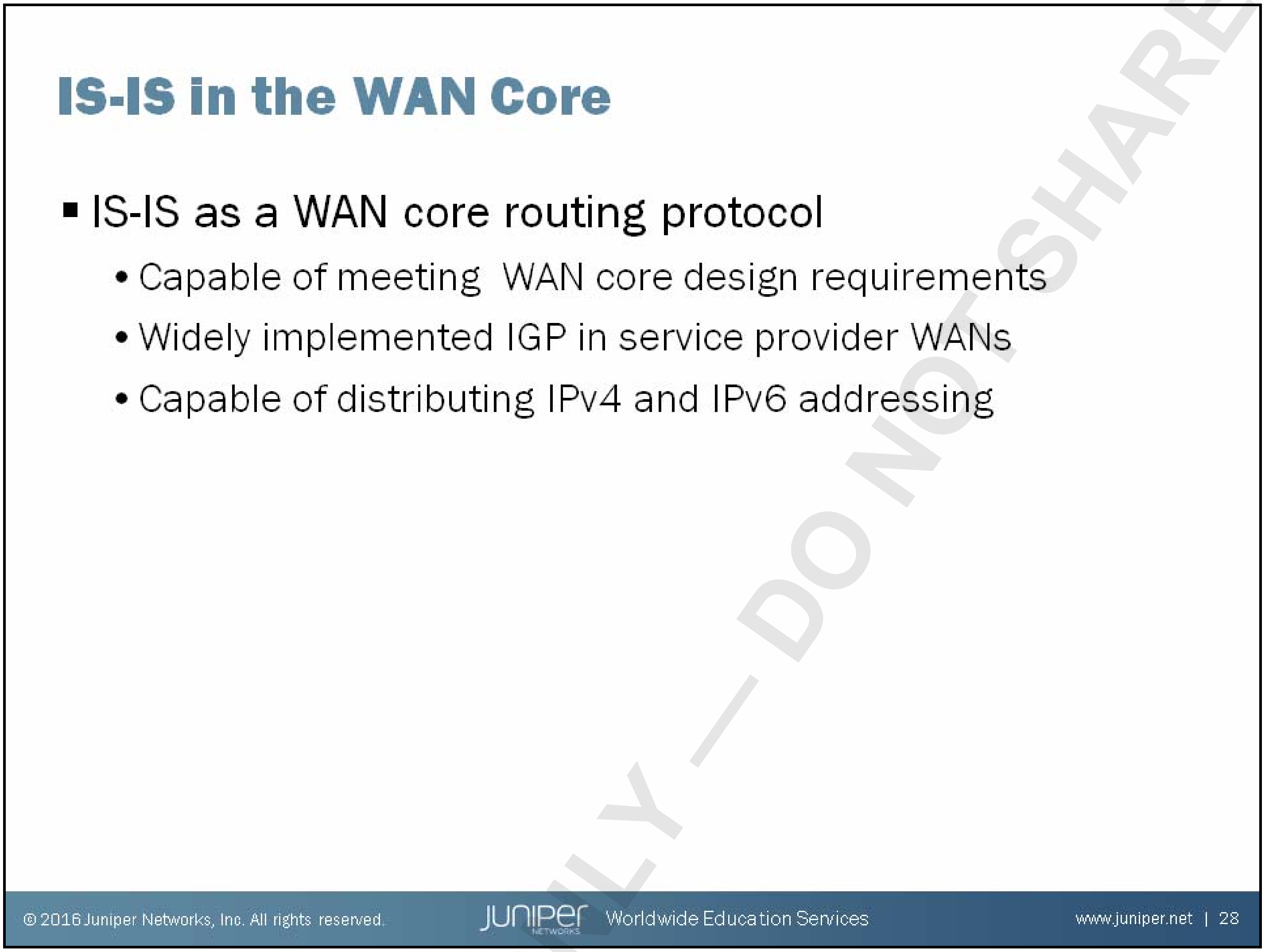
Limitations of OSPF

- OSPF limitations
  - Requires two protocol versions or OSPFv3 realm configuration to communicate IPv4 and IPv6 addressing
  - More complex SPF calculation than IS-IS which can reduce the number of routers permitted in a single area
  - Adding new functionality requires a rewrite of the protocol

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 27

OSPF Limitations

OSPFv2 is not capable of communicating IPv6 destinations. OSPFv3 can carry both IPv4 and IPv6 destinations with the configuration of realms for each protocol family. OSPF performs full SPF calculations more frequently than the IS-IS routing protocol which historically has limited the number of routers placed in a single area. OSPF lacks IS-IS TLV extensibility mechanisms requiring a protocol re-write to add new functionality.



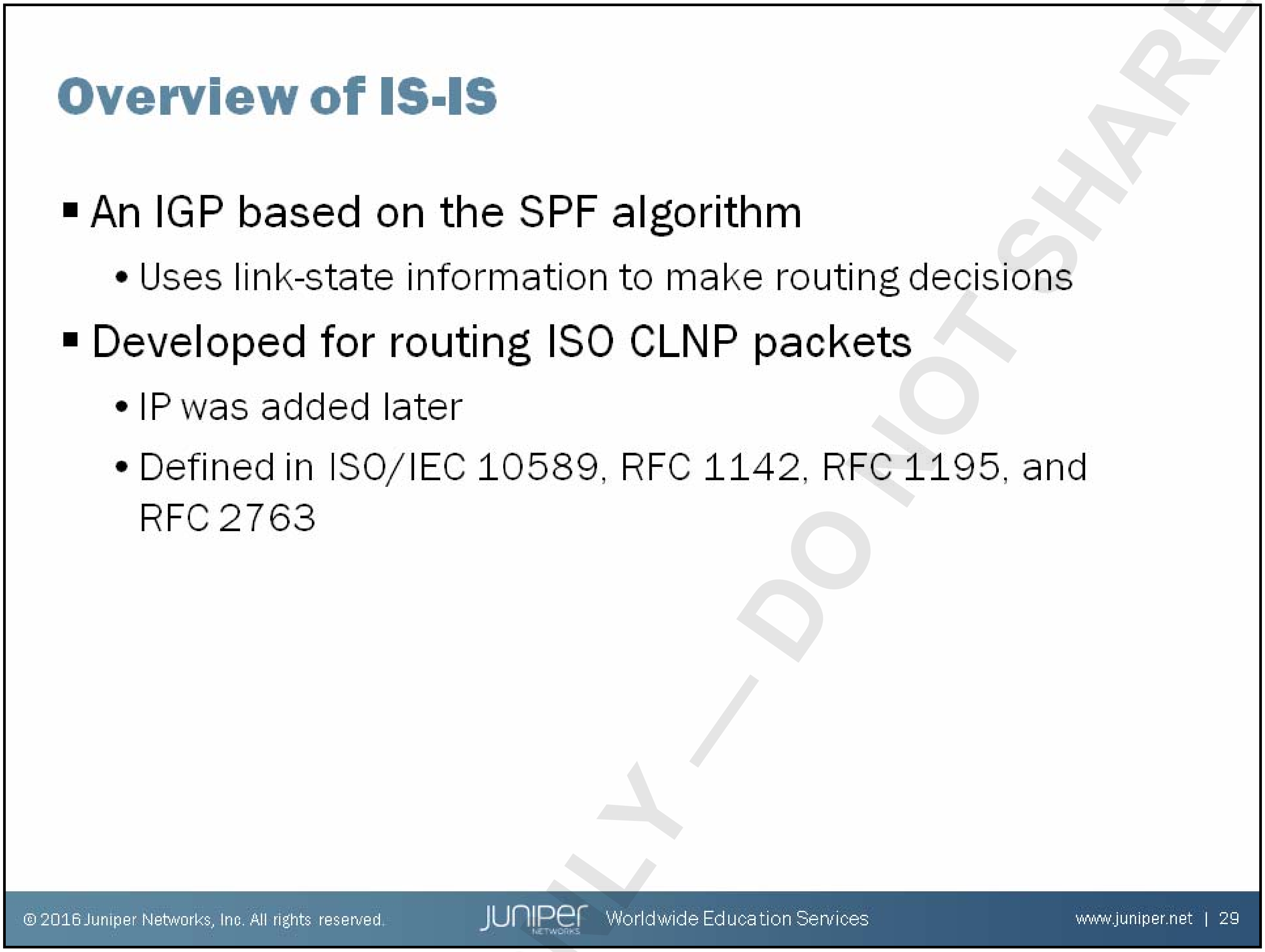
IS-IS in the WAN Core

- IS-IS as a WAN core routing protocol
  - Capable of meeting WAN core design requirements
  - Widely implemented IGP in service provider WANs
  - Capable of distributing IPv4 and IPv6 addressing

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 28

IS-IS in the WAN Core

IS-IS meets the functional and scalability requirements of the largest service provider WAN networks. Multiple WAN service providers around the world depend on IS-IS to distribute IPv4 and IPv6 addressing information. IS-IS uses provides TLVs to communicate traffic engineering data.



Overview of IS-IS

- An IGP based on the SPF algorithm
  - Uses link-state information to make routing decisions
- Developed for routing ISO CLNP packets
  - IP was added later
  - Defined in ISO/IEC 10589, RFC 1142, RFC 1195, and RFC 2763

Worldwide Education Services

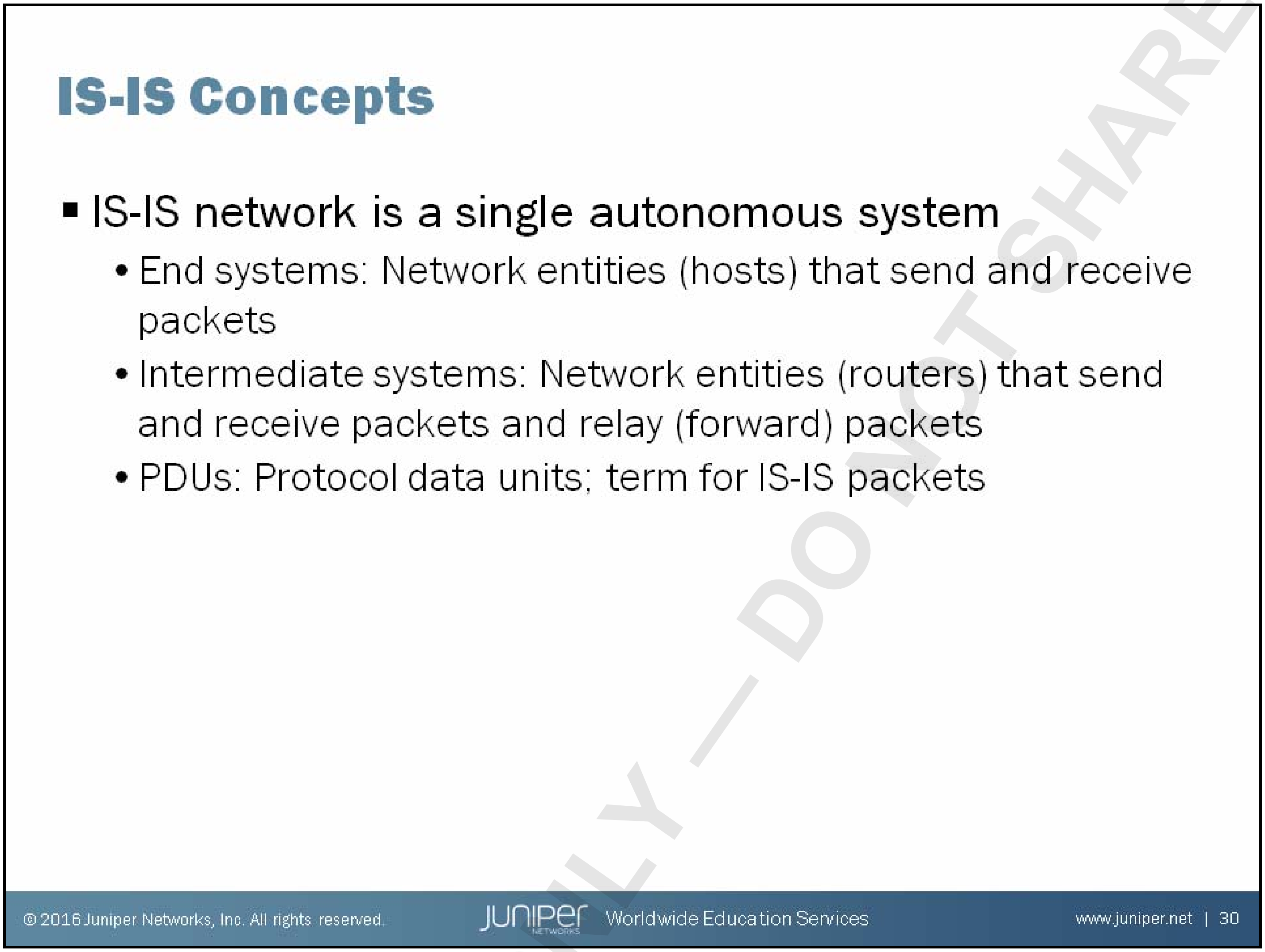
www.juniper.net | 29

Overview of IS-IS

IS-IS is an IGP that uses link-state information to make routing decisions. It also uses the SPF algorithm, similar to OSPF.

Developed by ISO

The ISO developed IS-IS to be the routing protocol for the ISO's Connectionless Network Protocol (CLNP) and is described in ISO 10589. Digital Equipment Corporation developed the protocol for its DECnet Phase V. The ISO was working on IS-IS at the same time that the Internet Advisory Board (IAB) was working on OSPF. ISO proposed that IS-IS be adopted as the routing protocol for TCP/IP in place of OSPF. This proposal was driven by the opinion that TCP/IP was an interim protocol suite that the Open Systems Interconnection (OSI) suite would eventually replace.



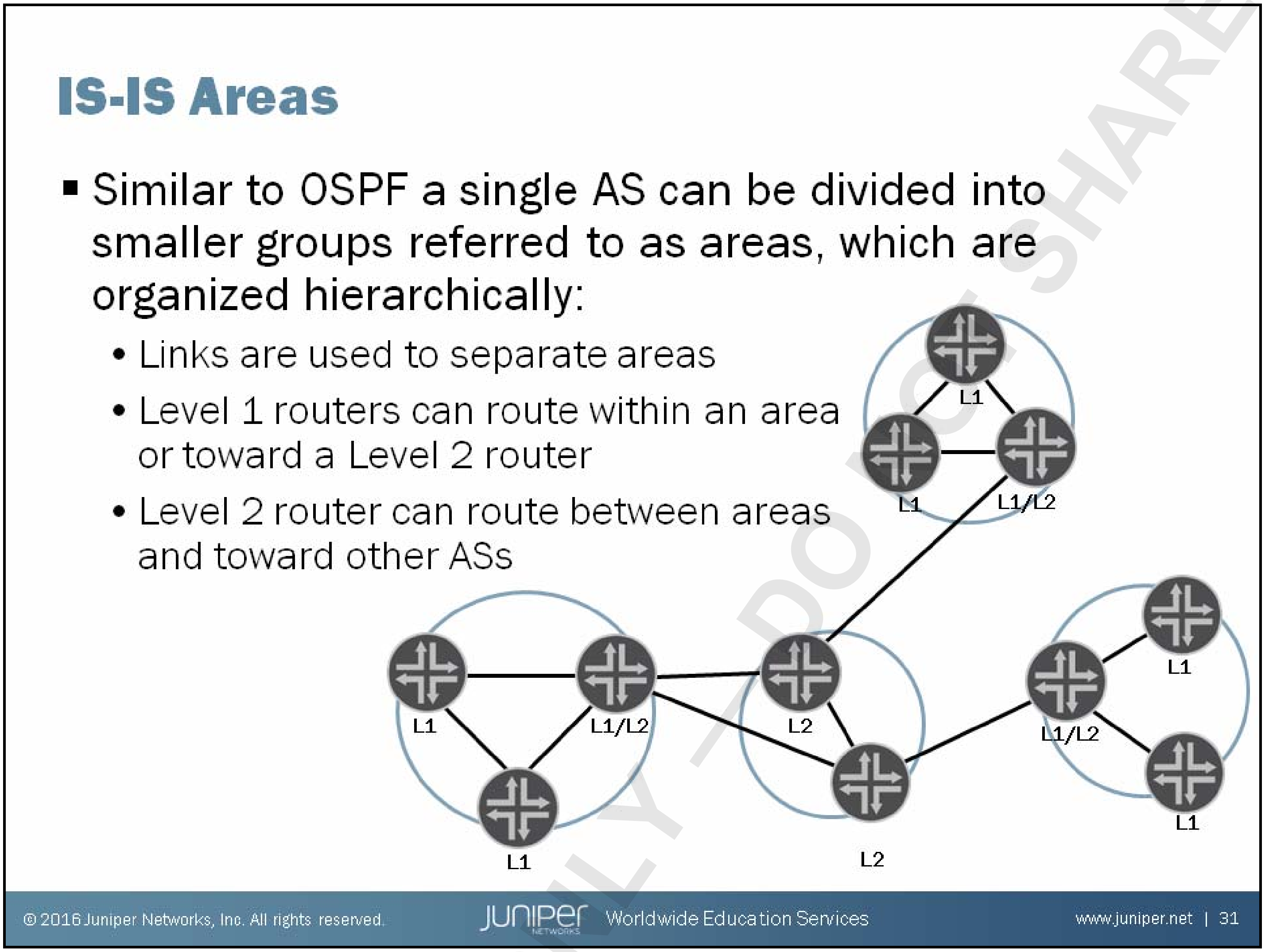
IS-IS Concepts

- IS-IS network is a single autonomous system
  - End systems: Network entities (hosts) that send and receive packets
  - Intermediate systems: Network entities (routers) that send and receive packets and relay (forward) packets
  - PDUs: Protocol data units; term for IS-IS packets

Worldwide Education Services www.juniper.net | 30

IS-IS Concepts

An IS-IS network is a single AS, also called a routing domain, that consists of end systems (ESs) and intermediate systems (ISs). End systems are network entities that send and receive packets. Intermediate systems—which is the OSI term for a router—send and receive packets and relay, or forward, packets. IS-IS packets are called protocol data units (PDUs).



IS-IS Areas

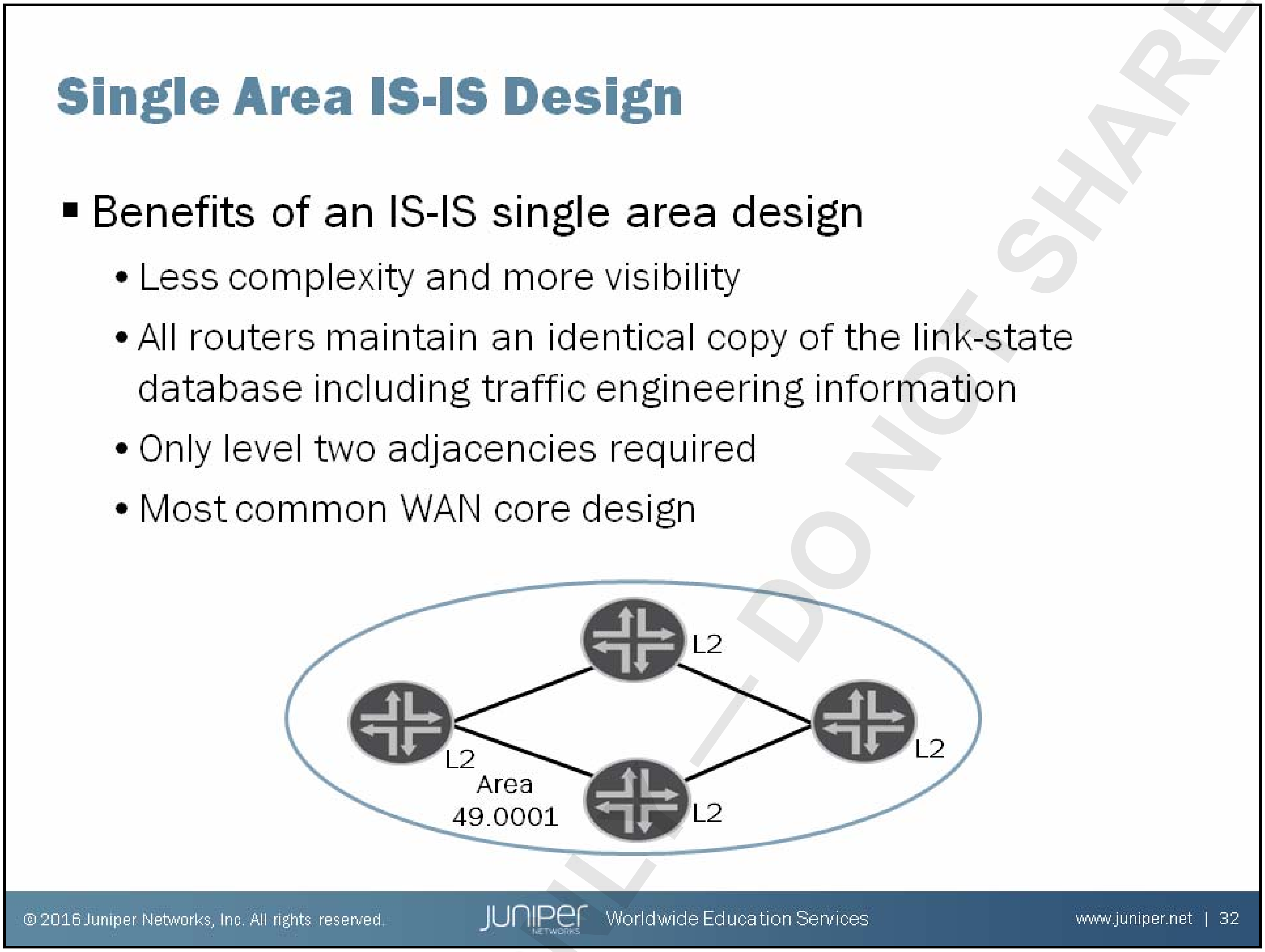
- Similar to OSPF a single AS can be divided into smaller groups referred to as areas, which are organized hierarchically:
  - Links are used to separate areas
  - Level 1 routers can route within an area or toward a Level 2 router
  - Level 2 router can route between areas and toward other ASs

JUNIPer NETWORKS Worldwide Education Services

www.juniper.net | 31

IS-IS Areas

In IS-IS, a single AS can be divided into smaller groups referred to as areas. Routing between areas is organized hierarchically, allowing a domain to be divided administratively into smaller areas. IS-IS accomplishes this organization by configuring Level 1 and Level 2 routers. Level 1 routers are used to route within an area, and Level 2 routers are used to route between areas and toward other ASs. A Level 1 and Level 2 router can route within an area on one interface and between areas on another. A Level 1 and Level 2 router sets the attached bit in the Level 1 PDUs that it generates into a Level 1 area to indicate that it is a Level 2-attached backbone router and that it can be used to reach prefixes outside the Level 1 area. Level 1 routers create a default route for interarea prefixes, which points to the closest (in terms of metrics) Level 1 and Level 2-attached router. Both IS-IS and OSPF are link-state routing protocols with many similarities. Of course, differences exist as well. In IS-IS areas, a Level 1 and Level 2 router fulfills the same purpose as an area border router in OSPF. Likewise, the collection of Level 2 routers in IS-IS is the backbone, whereas Area 0 is the backbone in OSPF. However, in IS-IS, all routers are completely within an area, and the area borders are on the links, not on the routers. The routers that connect areas are Level 2 routers, and routers that have no direct connectivity to another area are Level 1 routers. An IS can be a Level 1 router, a Level 2 router, or both (an Level 1 and Level 2 router).



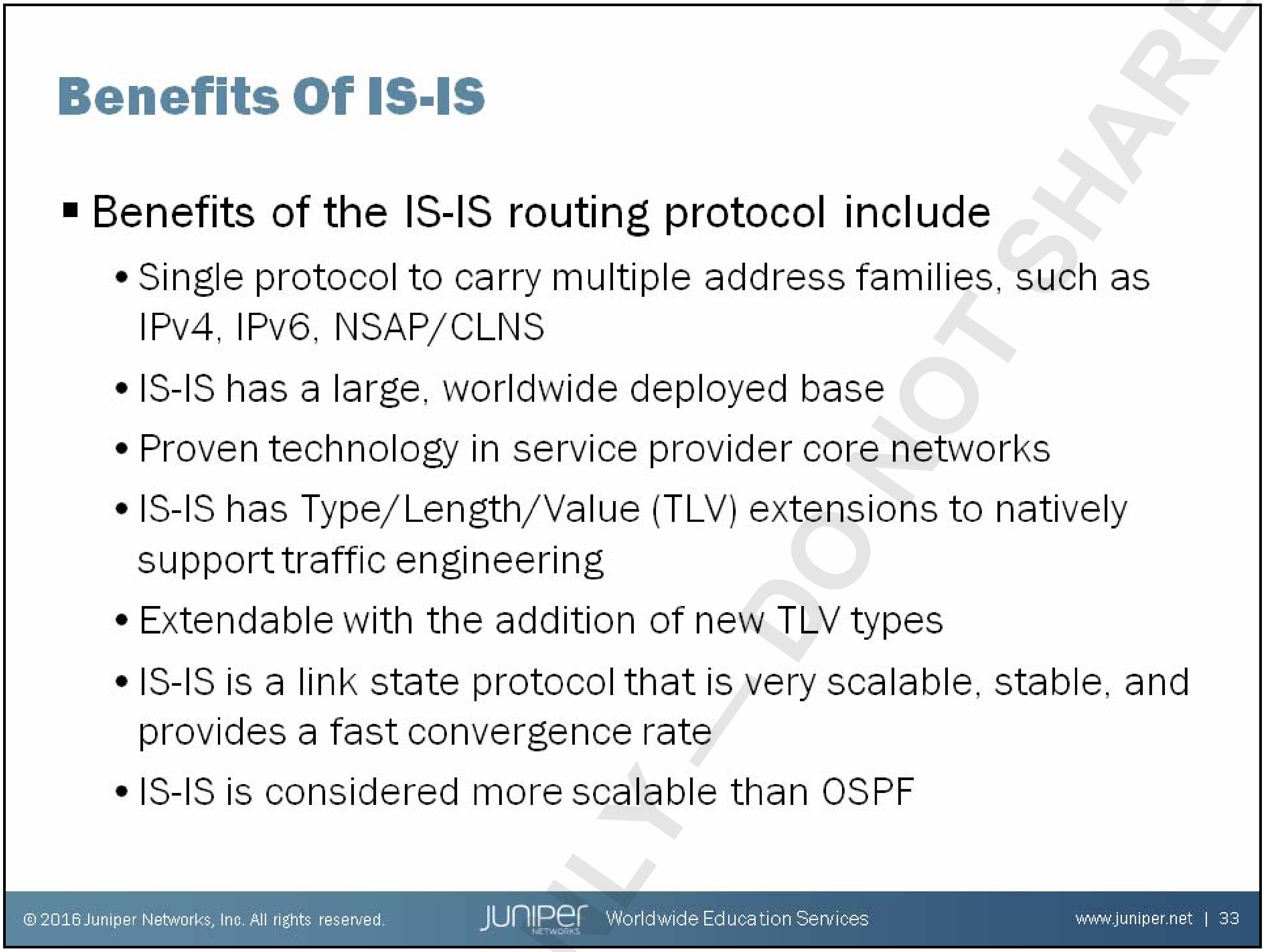
Single Area IS-IS Design

- Benefits of an IS-IS single area design
  - Less complexity and more visibility
  - All routers maintain an identical copy of the link-state database including traffic engineering information
  - Only level two adjacencies required
  - Most common WAN core design

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 32

Single Area IS-IS

Similar to OSPF, most service providers implement IS-IS as a single area AS with all routers configured to establish only level two adjacencies. Reducing the number areas reduces core network complexity. Similar to OSPF, IS-IS restricts the flow of traffic engineering data across area boundaries. Lack of traffic engineering data visibility complicates edge to edge MPLS path computations.



Benefits Of IS-IS

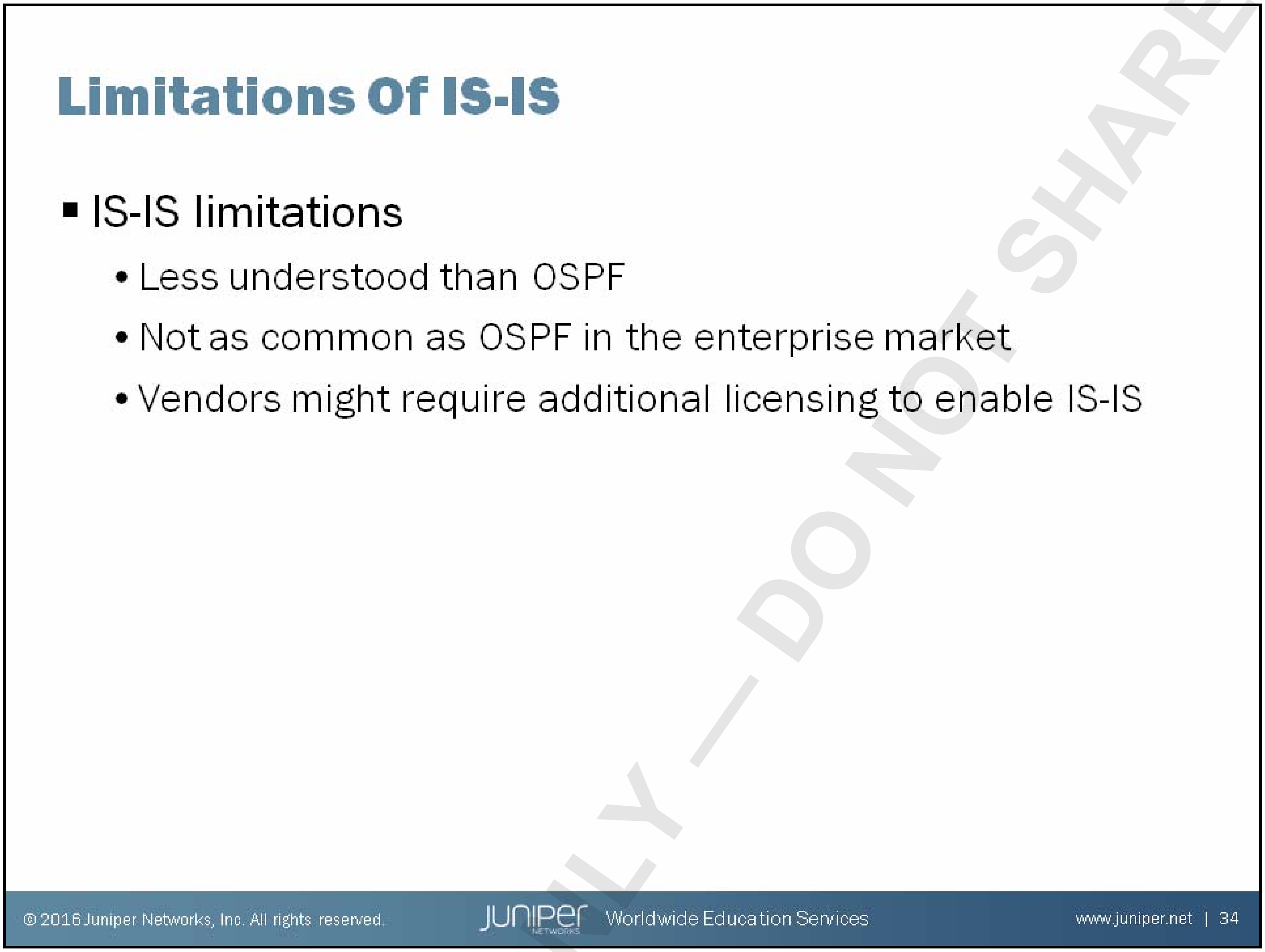
- Benefits of the IS-IS routing protocol include
  - Single protocol to carry multiple address families, such as IPv4, IPv6, NSAP/CLNS
  - IS-IS has a large, worldwide deployed base
  - Proven technology in service provider core networks
  - IS-IS has Type/Length/Value (TLV) extensions to natively support traffic engineering
  - Extendable with the addition of new TLV types
  - IS-IS is a link state protocol that is very scalable, stable, and provides a fast convergence rate
  - IS-IS is considered more scalable than OSPF

Worldwide Education Services

www.juniper.net | 33

Benefits of IS-IS

IS-IS provides multiple benefits to WAN IGP designers. IS-IS has a large worldwide deployment base in the service provider WAN. IS-IS is capable of communicating both IPv4 and IPv6 address families and has been extended to transmit traffic engineering information. One of the important benefits of IS-IS is that it’s functionality can be extended. New TLV extensions can be added allowing new information exchange without the need to upgrade the entire protocol. IS-IS is considered to be more scalable in single area environments because the SPF algorithm implemented in the protocol runs in a more efficient fashion than OSPF allowing more routers in a single area.



Limitations Of IS-IS

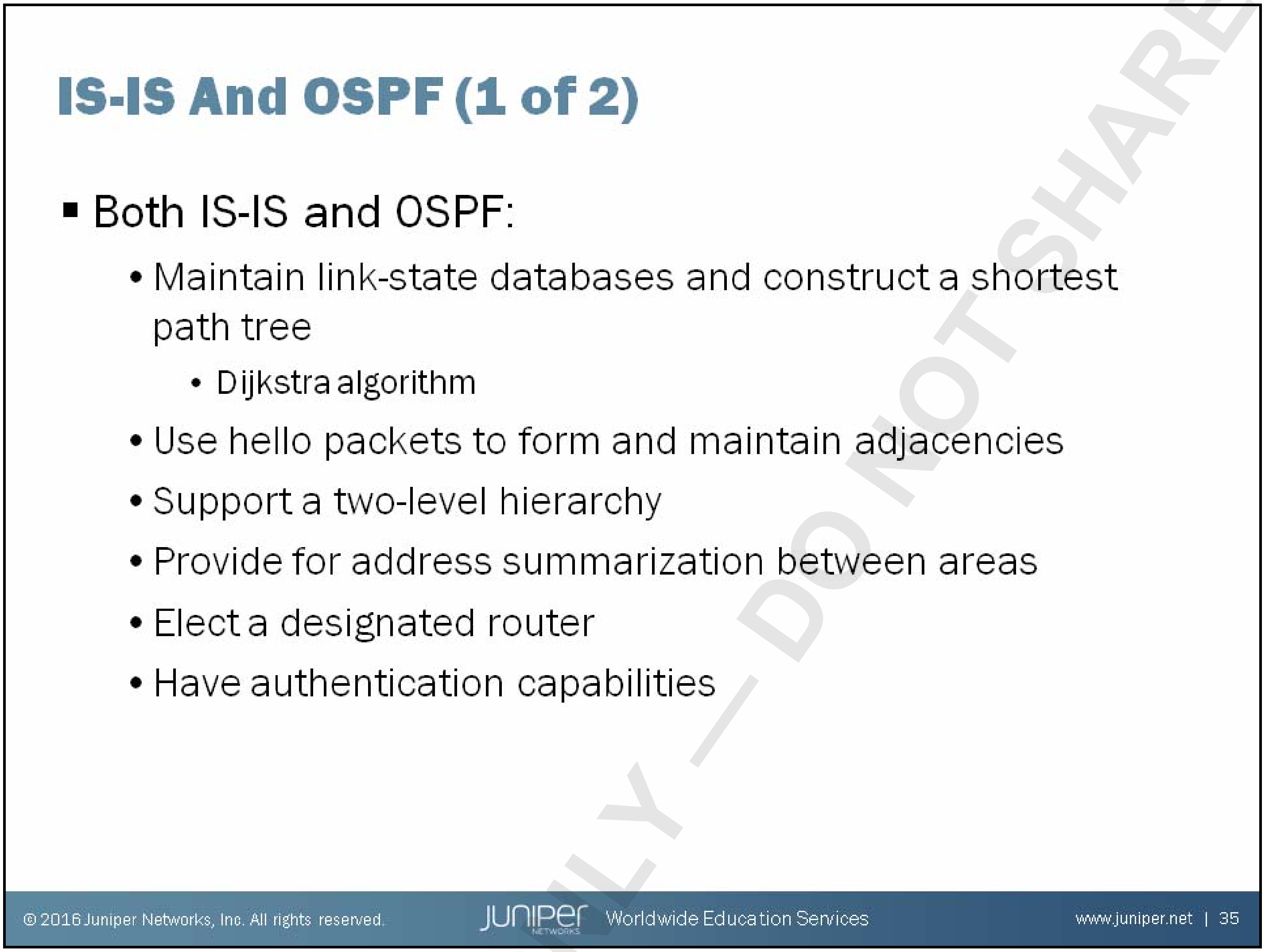
- IS-IS limitations
  - Less understood than OSPF
  - Not as common as OSPF in the enterprise market
  - Vendors might require additional licensing to enable IS-IS

Worldwide Education Services

www.juniper.net | 34

IS-IS Limitations

IS-IS is more common in the service provider world than in the enterprise market. Simply put more people have experience with OSPF than IS-IS. Some vendors may not enable the IS-IS protocol by default on their platforms. Designers should verify the network platforms they purchase for the core network support IS-IS and if enabling IS-IS requires any additional feature licensing.



IS-IS And OSPF (1 of 2)

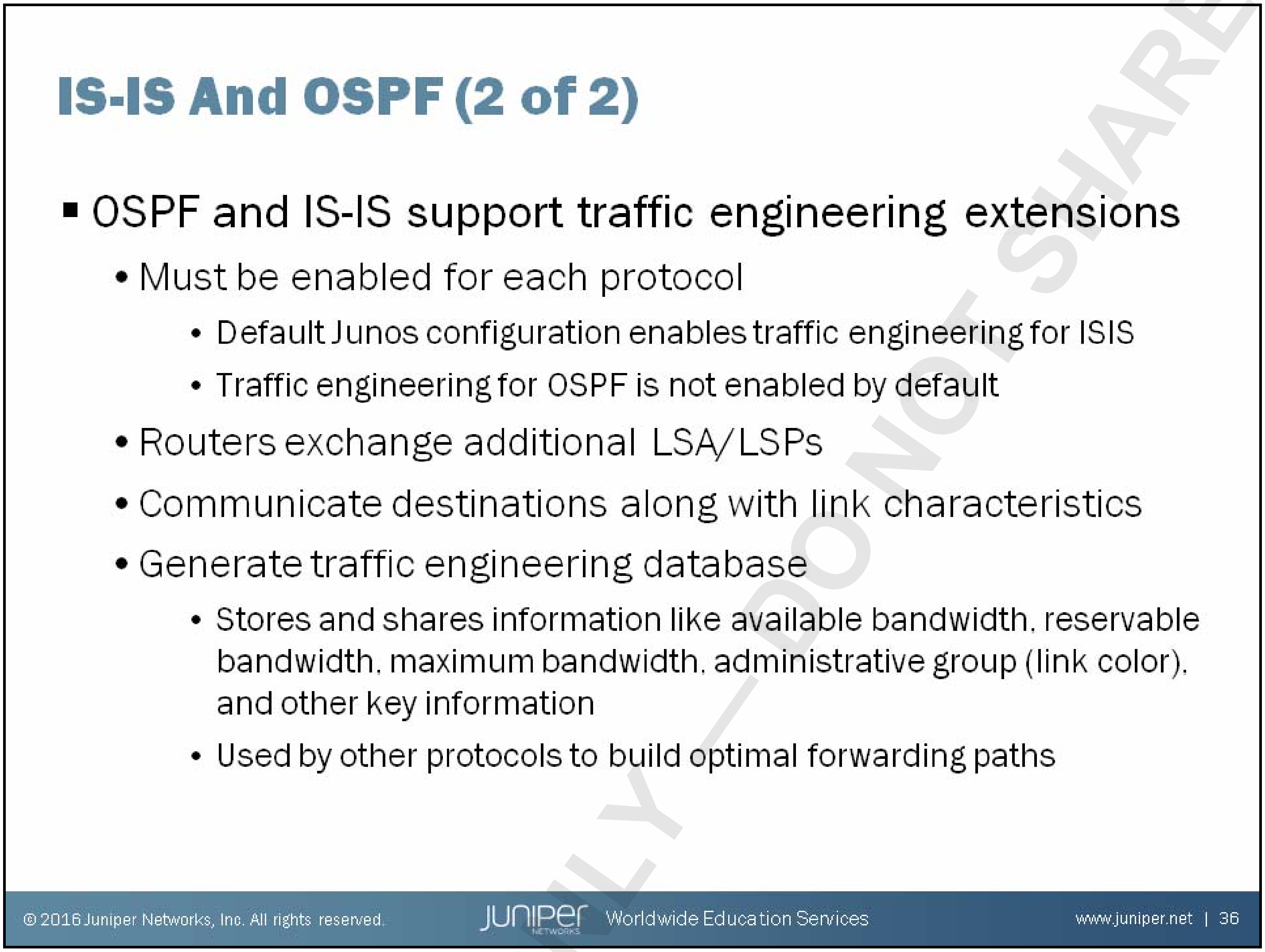
- Both IS-IS and OSPF:
  - Maintain link-state databases and construct a shortest path tree
    - Dijkstra algorithm
  - Use hello packets to form and maintain adjacencies
  - Support a two-level hierarchy
  - Provide for address summarization between areas
  - Elect a designated router
  - Have authentication capabilities

Worldwide Education Services

www.juniper.net | 35

IS-IS and OSPF Common Features

The slide lists the commonalities between IS-IS and OSPF.



IS-IS And OSPF (2 of 2)

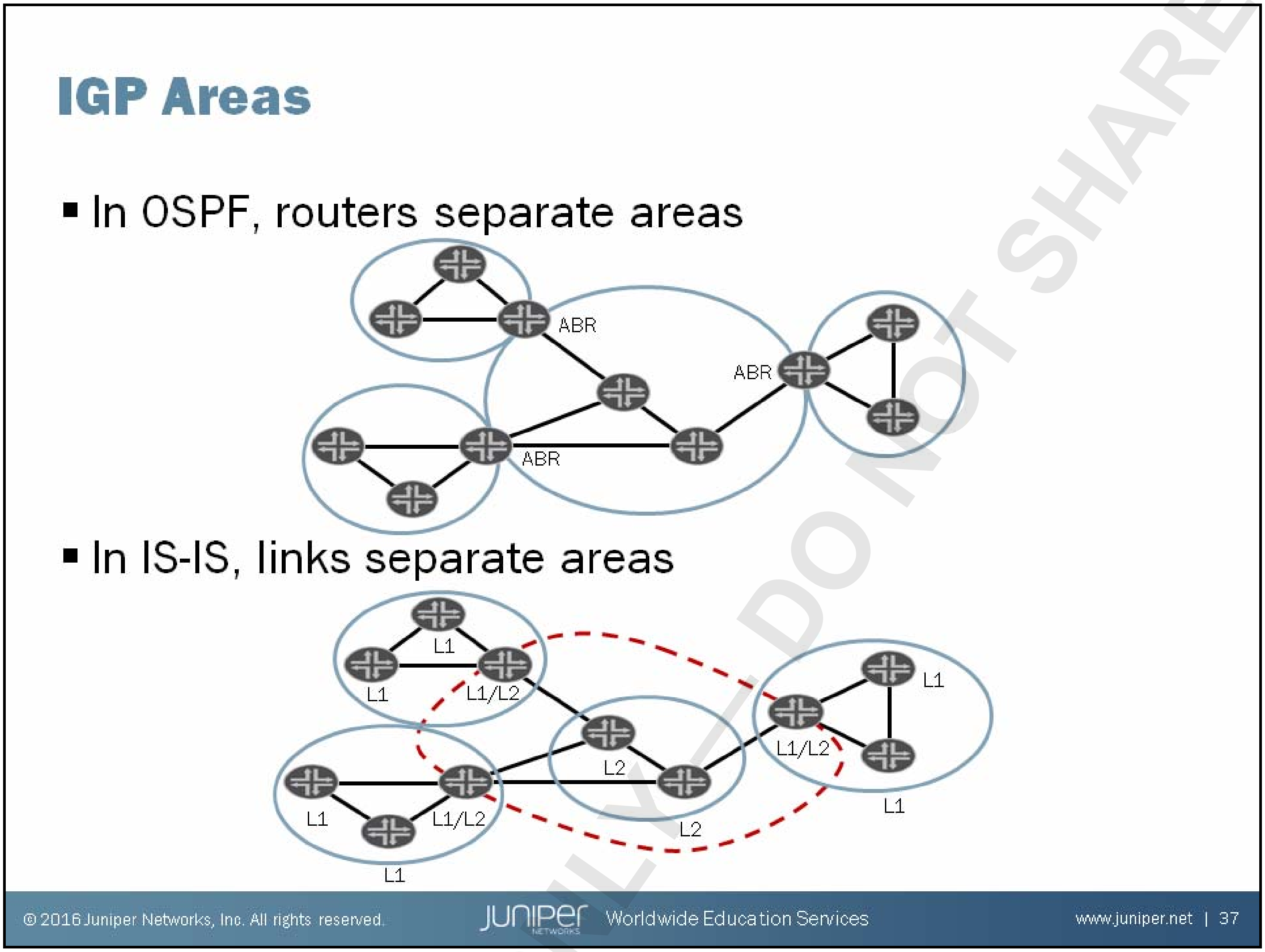
- OSPF and IS-IS support traffic engineering extensions
  - Must be enabled for each protocol
    - Default Junos configuration enables traffic engineering for ISIS
    - Traffic engineering for OSPF is not enabled by default
  - Routers exchange additional LSA/LSPs
  - Communicate destinations along with link characteristics
  - Generate traffic engineering database
    - Stores and shares information like available bandwidth, reservable bandwidth, maximum bandwidth, administrative group (link color), and other key information
    - Used by other protocols to build optimal forwarding paths

© 2016 Juniper Networks, Inc. All rights reserved. JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 36

IGP Traffic Engineering

Extensions have been added to the OSPF and IS-IS routing protocols that allow peers to exchange traffic engineering information. Additional LSAs in OSPF and link state protocol data units (LSPs) in IS-IS extend the IGPs functionality and allow information about the links that make up the WAN to be communicated among peers. Information such as available bandwidth, reservable bandwidth, maximum bandwidth, administrative group (link color) can be communicated along with network addressing. The traffic engineering information is stored in a traffic engineering database (TED) on each WAN router. OSPF and IS-IS don’t use this traffic engineering information directly. The IGP produces, distributes, and stores the information in the TED for other applications to use. Applications use TED information when computing traffic paths across the WAN. Applications that do not use the TED will only be able to route traffic based upon the IGP shortest path metric stored in the LSDB.

Most WAN providers offer MPLS services to their customers and take advantage of the granular control allowed by the information stored in the TED. Using the LSDB and the TED allow designers to specify a LSP from point A to point B that uses a path that differs from the IGP shortest path metric. MPLS can signal the desired path for customer traffic based on link characteristics stored in the TED.



IGP Areas

- In OSPF, routers separate areas

- In IS-IS, links separate areas

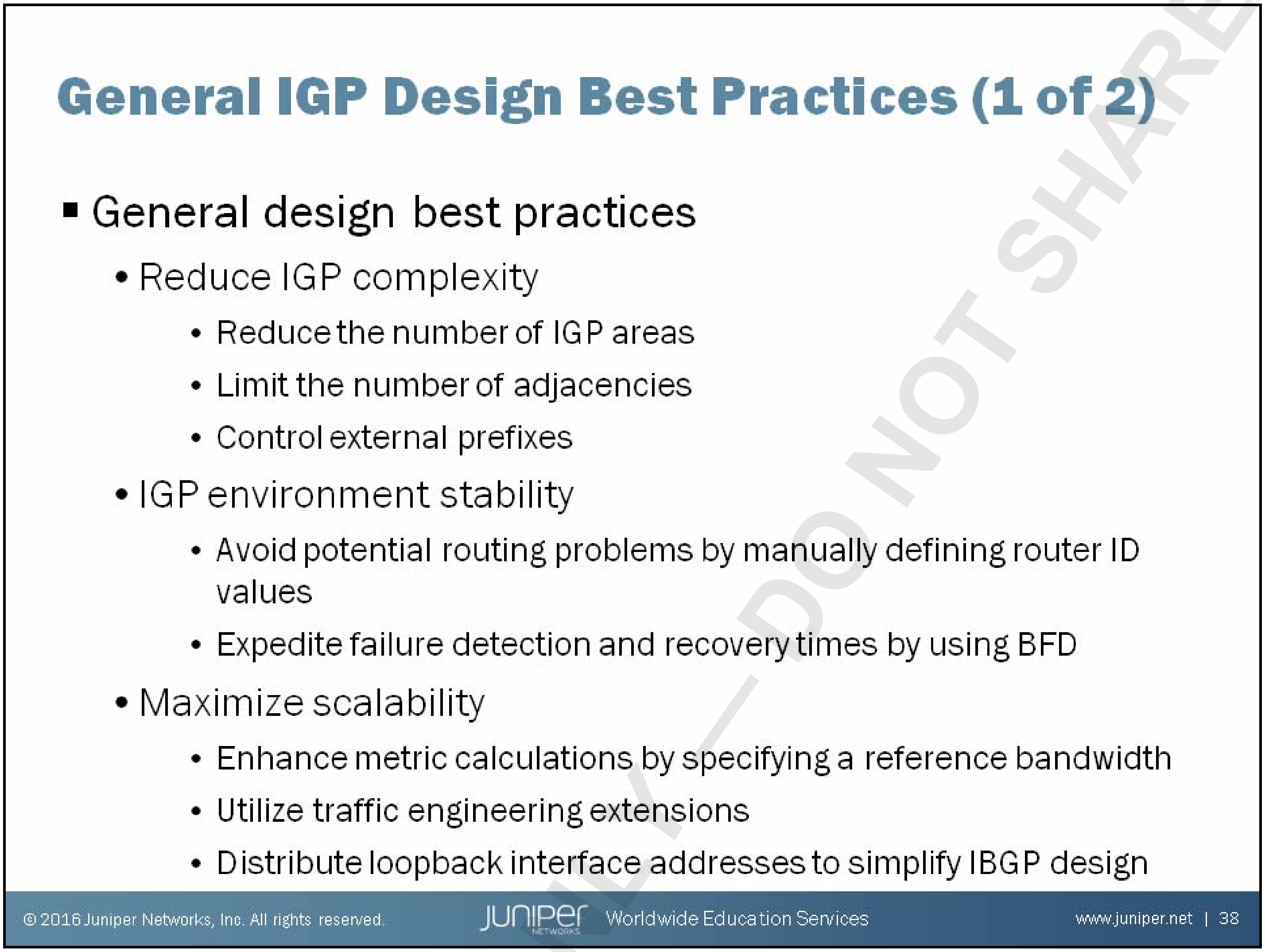
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 37

OSPF Areas

IS-IS and OSPF are link-state routing protocols with many similarities. Differences exist between them as well. Recall that, under OSPF, routers separate areas. The slide shows how a typical OSPF network might be broken up into areas. Some interfaces are in one area, and other interfaces are in another area. When an OSPF router has interfaces in more than one area, it is an area border router (ABR).

IS-IS Areas

In IS-IS areas, a Level 1/Level 2 router fulfills the same purpose as an ABR in OSPF. Likewise, the collection of Level 2 routers in IS-IS is the backbone, while Area 0 is the backbone in OSPF. However, in IS-IS, all routers are completely within an area, and the area borders are on the links, not on the routers. The routers that connect areas are Level 2 routers, and routers that have no direct connectivity to another area are Level 1 routers. An intermediate system can be a Level 1 router, a Level 2 router, or both (an L1/L2 router).



General IGP Design Best Practices (1 of 2)

- General design best practices
  - Reduce IGP complexity
    - Reduce the number of IGP areas
    - Limit the number of adjacencies
    - Control external prefixes
  - IGP environment stability
    - Avoid potential routing problems by manually defining router ID values
    - Expedite failure detection and recovery times by using BFD
  - Maximize scalability
    - Enhance metric calculations by specifying a reference bandwidth
    - Utilize traffic engineering extensions
    - Distribute loopback interface addresses to simplify IBGP design

IGP Design Best Practices

A single area design reduces complexity and increases the amount of information visible to each network node. Traffic engineering information is not flooded across area boundaries complicating path MPLS path computations across the WAN. Modern networking devices are not as limited resource wise and are less susceptible to the performance effects of SPF calculations allowing WAN designers to place more networking devices in a single area. Reducing the number of neighbor adjacencies each node maintains can reduce the number of LSA/LSPs that must be flooded. As networking information changes it must be flooded to all nodes in the area and each node must run the SPF algorithm and update the routing table. The IS-IS protocol by default establishes both a level one and level two adjacency with any discovered neighbor. In an IS-IS single area environment only level two adjacencies are necessary. Eliminating level one adjacencies improves performance of IS-IS nodes. This practice allows external network reachability. External destinations imported into the IGP create new LSA/LSPs that must be distributed to all nodes in the area. Both OSPF and IS-IS support prefix limits that control the quantity of external networks injected into the IGP.

Every network node in an IGP network is identified by a router ID value. Each network node includes its router ID value in every LSA/LSP it originates. Network administrators can use the router ID value stored with the LSA/LSPs in the LSDB to determine the origin of routing information. Loopback interface addresses are a common default value for router IDs. Changing a nodes loopback interface address changes the router ID value and causes the re-flooding of all LSA/LSPs originated by the affected node. Duplicate router ID values on two nodes causes confusion and routing failure. Designers should plan router ID values and manually define those values on each node.

Continued on the next page.

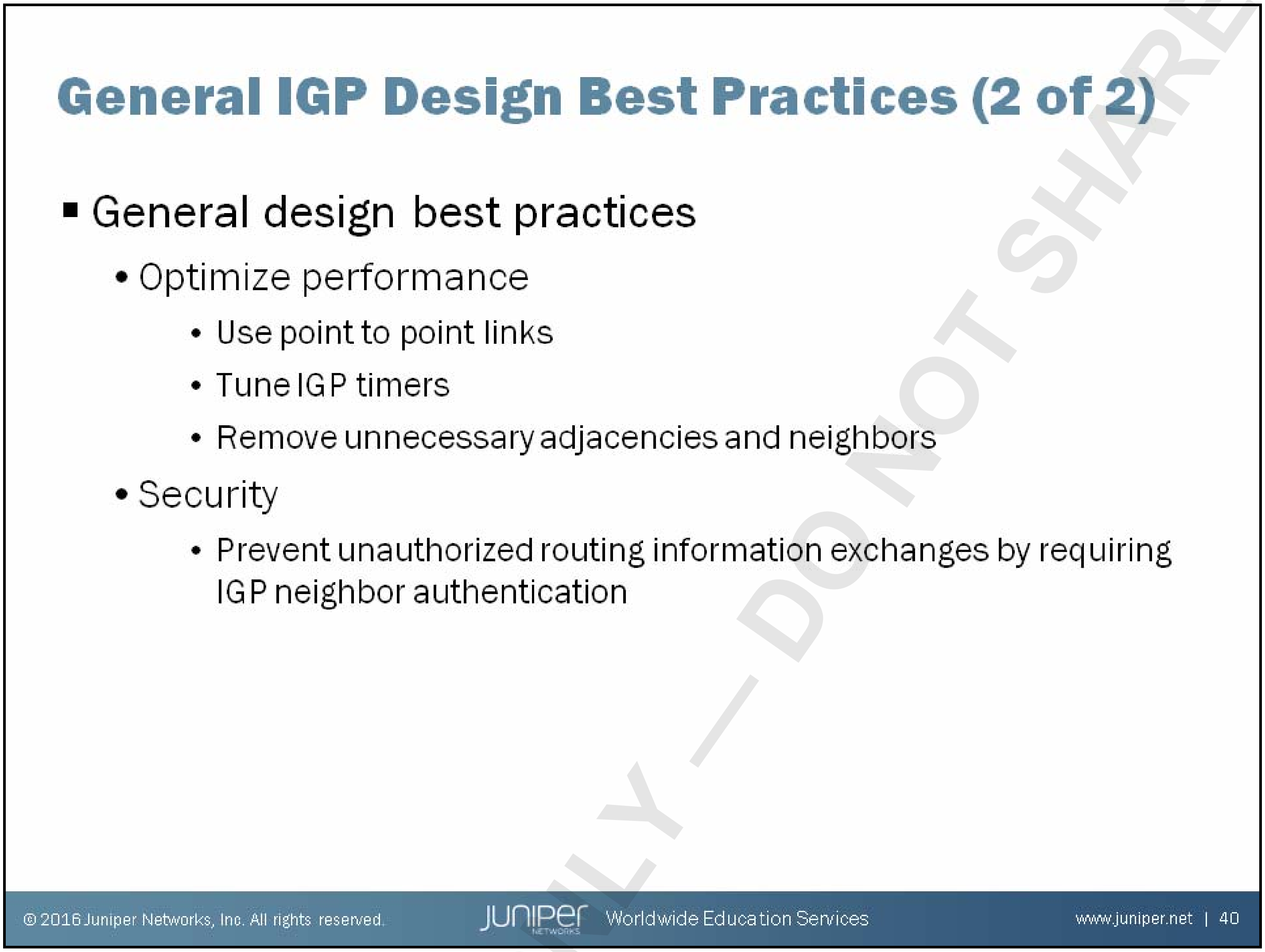
IGP Design Best Practices (contd.)

Both OSPF and IS-IS have built in mechanisms to detect the failure of a neighbor. Protocol timers can be modified to speed the time to failure detection and recovery. Instead of tuning individual IGP parameters network designers commonly use the Bi-directional Forwarding Detection (BFD) protocol to more rapidly detect neighbor failure. BFD transmits hello messages between neighbors and can be tuned to sub second notification that a neighbor is unavailable. The OSPF default timers can take up to 40 seconds to discover a faulty neighbor. BFD is compatible with IGP and exterior gateway protocol (EGP) routing protocols.

Enable traffic engineering extensions. Traffic engineering allows granular control of the path that data packets follow, bypassing the standard shortest path routing. Traffic engineering moves flows from congested links to alternate links that would not be selected by the automatically computed destination-based shortest path. Verify the protocol defaults for the IGP are used and ensure the traffic engineering extension have been enabled.

Core networks use IGPs to create a routing foundation that other protocols and services can use. All WANs will eventually use the Border Gateway Protocol (BGP) to exchange public and customer destinations. The most redundant way to establish BGP neighbors is to use loopback interface address peering. Loopback interfaces are considered to be always up and reachable through any active physical interface on the BGP neighbor. Configuring the IGP to distribute routes to each network node's loopback interface address adds redundancy and stability to BGP.

IGPs calculate metrics for the links that make up the core network. Core WAN designers can define the values used to calculate link metrics. A reference bandwidth can be defined that allows network nodes to automatically calculate metrics that take into account the different amounts of bandwidth provided by the links that make up the WAN. The reference bandwidth is divided by the interface bandwidth to calculate the metric for a particular link. Many IGP defaults set reference bandwidth to a low legacy value that doesn't take into account newer higher speed interfaces. If the reference bandwidth is set to 100M any interface with 100M bandwidth or more will calculate a metric of one. 100M, 1G, 10G, 40G, and 100G all share the same metric of one which does not offer much metric granularity for path calculations. Define a reference bandwidth value that is greater than or equal-to the bandwidth of the largest physical link in the network.



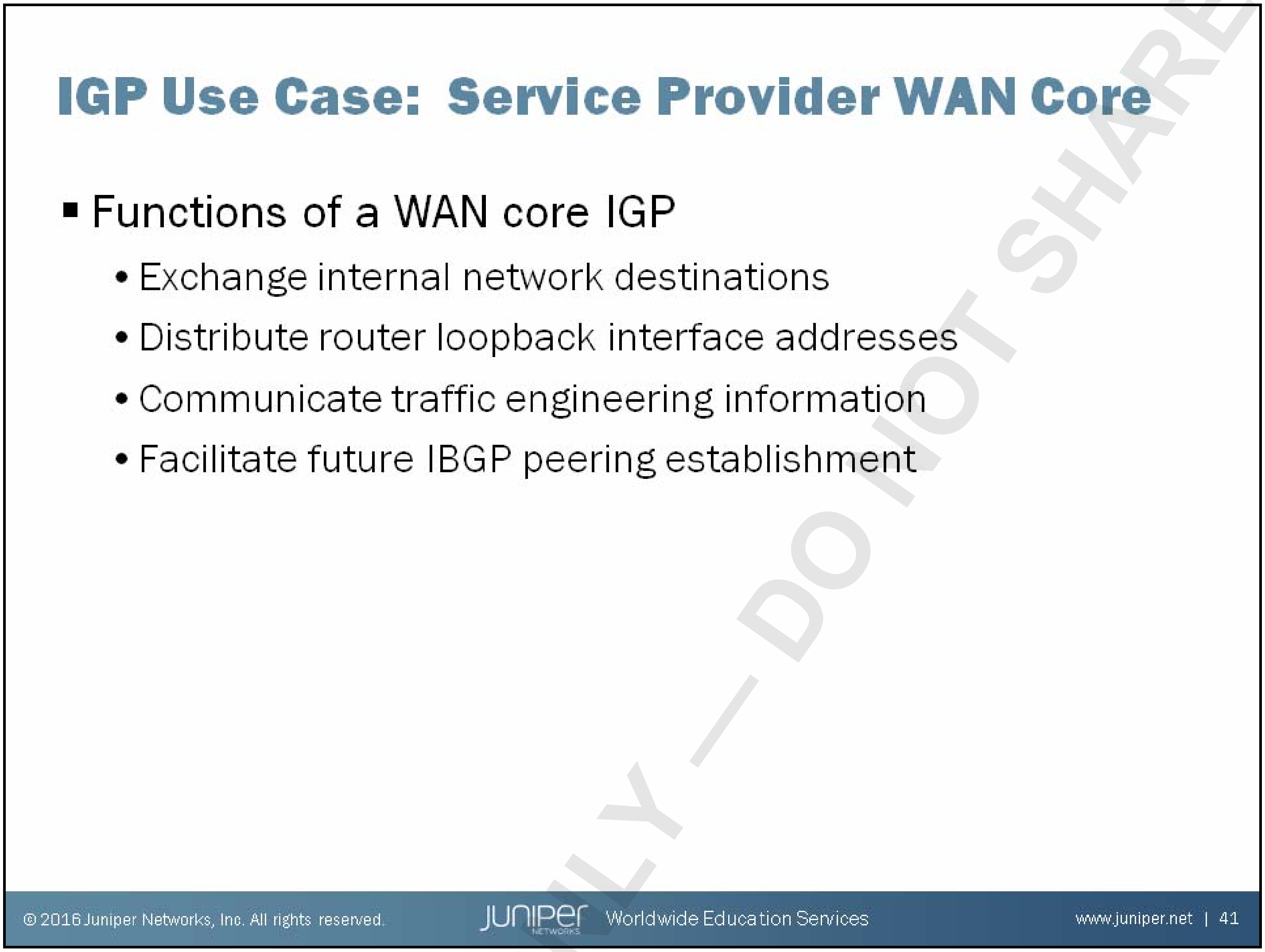
General IGP Design Best Practices (2 of 2)

- General design best practices
  - Optimize performance
    - Use point to point links
    - Tune IGP timers
    - Remove unnecessary adjacencies and neighbors
  - Security
    - Prevent unauthorized routing information exchanges by requiring IGP neighbor authentication

General IGP Design Best Practices

The defaults of OSPF allow a node to wait 40 seconds on a multi-access interfaces until adjacencies are allowed to be established. During this 40 second wait time no traffic is allowed to be forwarded over the link. OSPF and IS-IS provide design options to speed the convergence process. Defining multi-access interfaces to be point to point instead of the default of broadcast eliminates the 40 second wait time and allows adjacencies to be established immediately. Several IGP timers are available to speed time to failure notification and routing recovery. Analyze any protocol timers available in your vendors IGP implementation and design values that balance performance and stability. Reducing or disabling the number of adjacencies speeds time to a converged stable WAN.

IGPs are vulnerable to attacks that inject incorrect routing information into the LSDB. IGP neighbors can establish neighbor adjacencies without requiring authentication. This allows attackers complete visibility of WAN core network addressing and permits easy injection of malicious routing information. Network designers should require authentication before establishing any adjacencies. Several authentication methods are available for designers to choose from. Each method requires additional design, coordination, and planning.



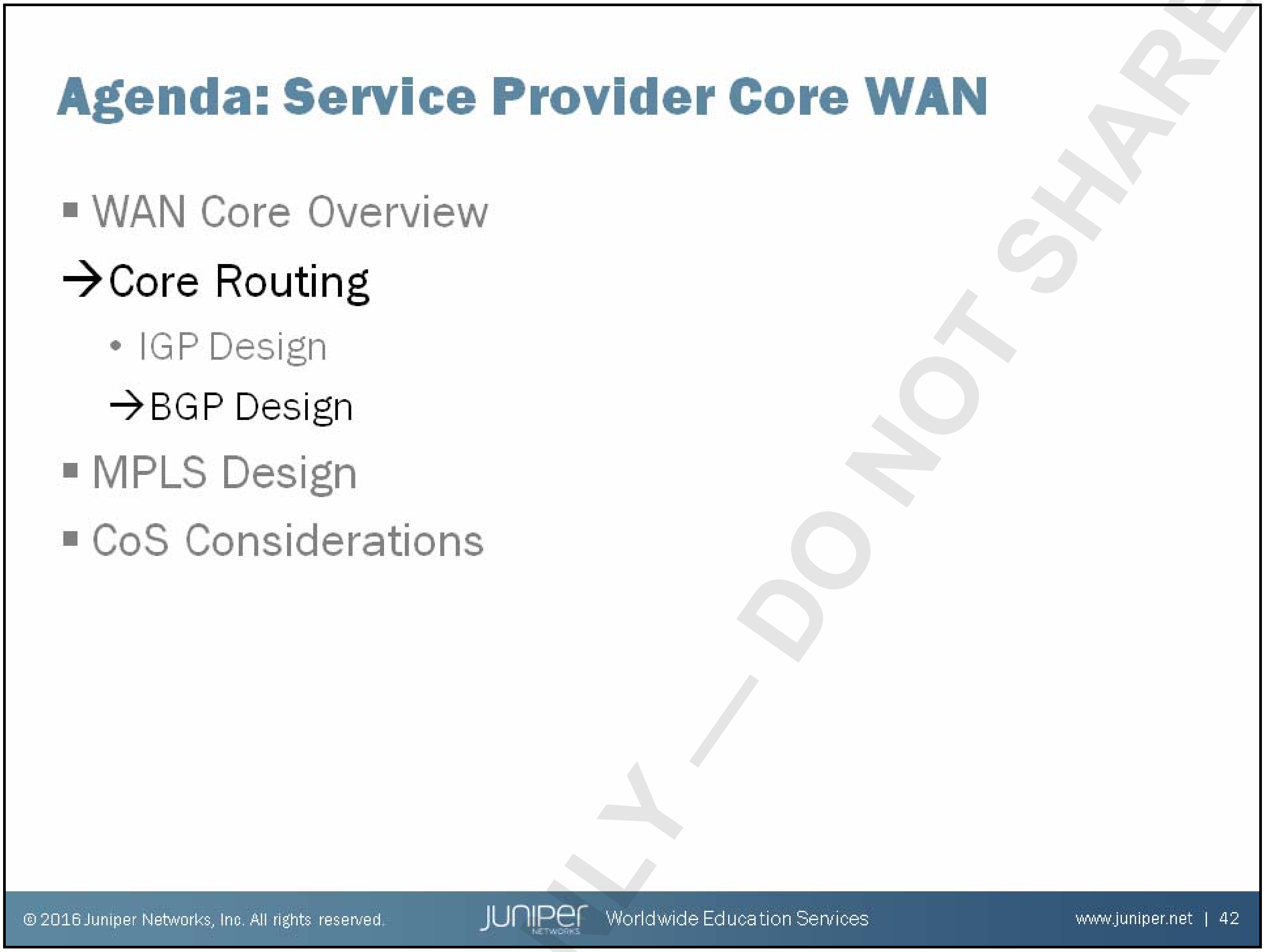
IGP Use Case: Service Provider WAN Core

- Functions of a WAN core IGP
  - Exchange internal network destinations
  - Distribute router loopback interface addresses
  - Communicate traffic engineering information
  - Facilitate future IBGP peering establishment

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 41

IGP Use Case: Service Provider Core

The slide lists the functions of a WAN IGP routing protocol.



Agenda: Service Provider Core WAN

- WAN Core Overview

→Core Routing

- IGP Design

→BGP Design

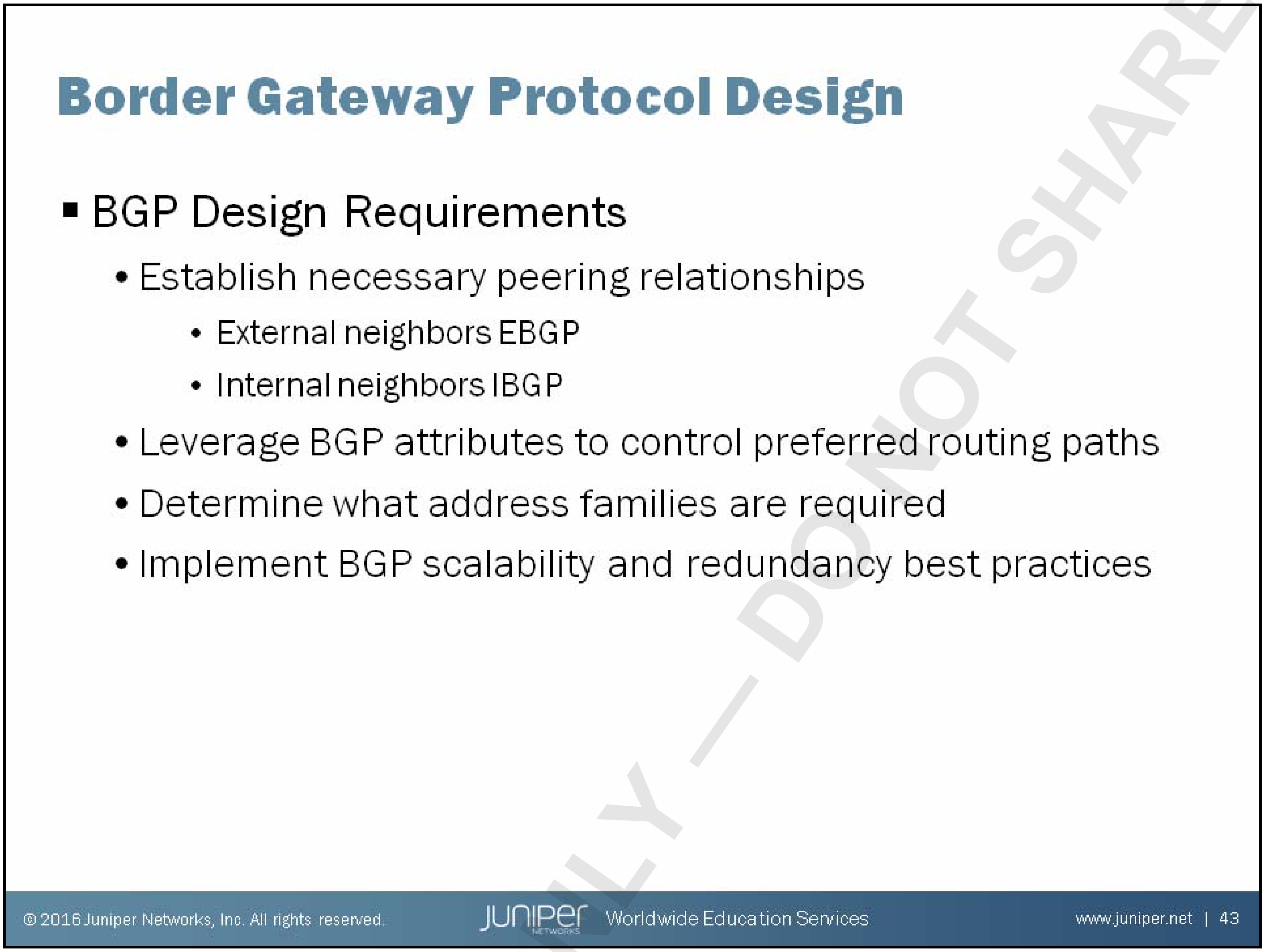
- MPLS Design
- CoS Considerations

Worldwide Education Services

www.juniper.net | 42

BGP Design

The slide highlights the topic we discuss next.

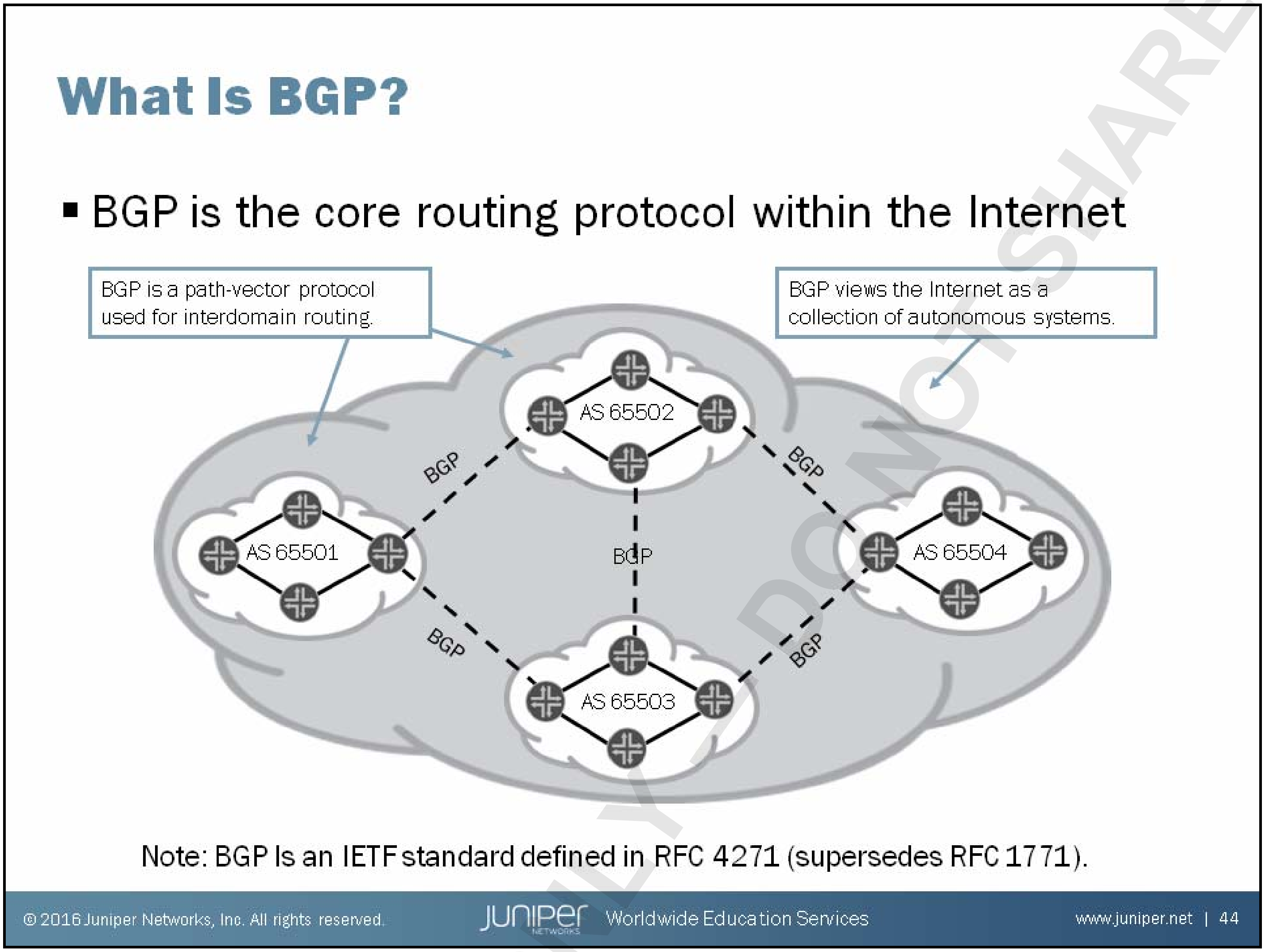


Border Gateway Protocol Design

- BGP Design Requirements
  - Establish necessary peering relationships
    - External neighbors EBGP
    - Internal neighbors IBGP
  - Leverage BGP attributes to control preferred routing paths
  - Determine what address families are required
  - Implement BGP scalability and redundancy best practices

Border Gateway Protocol Design

After the IGP design is completed attention turns to BGP. Which BGP features will be implemented to offer a stable, robust, and scalable solution? What addressing information needs to be communicated between WAN neighbors? What types of connectivity services are being offered to WAN customers? This section discusses BGP design concepts used by WAN providers.



What Is BGP?

- BGP is the core routing protocol within the Internet

BGP is a path-vector protocol used for interdomain routing.

BGP views the Internet as a collection of autonomous systems.

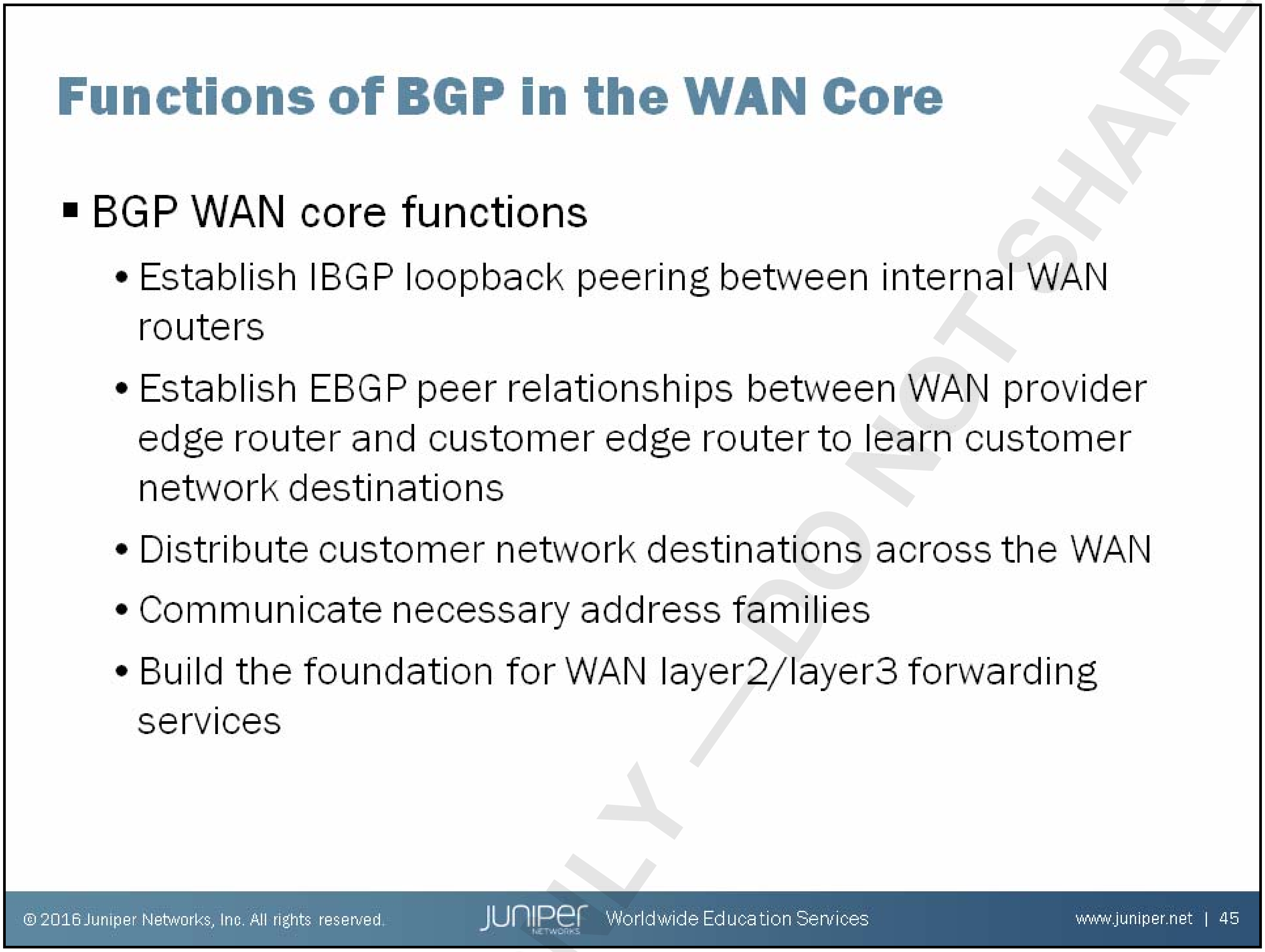
AS 65501
AS 65502
AS 65503
AS 65504
BGP
BGP
BGP
BGP
BGP

Note: BGP Is an IETF standard defined in RFC 4271 (supersedes RFC 1771).

JUNIPer NETWORKS Worldwide Education Services www.juniper.net | 44

What Is BGP?

BGP is a routing protocol used between autonomous systems (ASs) and is sometimes referred to as a path-vector routing protocol because it uses an AS path, used as a vector, to prevent inter-domain routing loops. The term path vector, in relation to BGP, means that BGP routing information includes a series of AS numbers, indicating the path that a route takes through the network. Although BGP is primarily used for inter-AS routing, BGP is also used in large networks for MPLS-based VPNs and is used to separate large OSPF domains. BGP is much more scalable and offers a greater amount of control through policy than an IGP. It exchanges routing information among ASs, which is a set of routers that operate under the same administration. BGP routing information includes the complete route to each destination. It uses the routing information to maintain an information base of network layer reachability information (NLRI), which it exchanges with other BGP systems. BGP is a classless routing protocol, that supports prefix routing, regardless of the class definitions of IPv4 addresses. BGP routers exchange routing information between peers. The peers must be connected directly for inter-AS BGP routing (unless certain configuration changes are done). The peers depend on established TCP connections, which we address later in this chapter. BGP version 4 (BGP4) is essentially the only EGP currently used in the Internet. It is defined in RFC 4271, which made the former standard of more than 10 years, RFC 1771, obsolete.



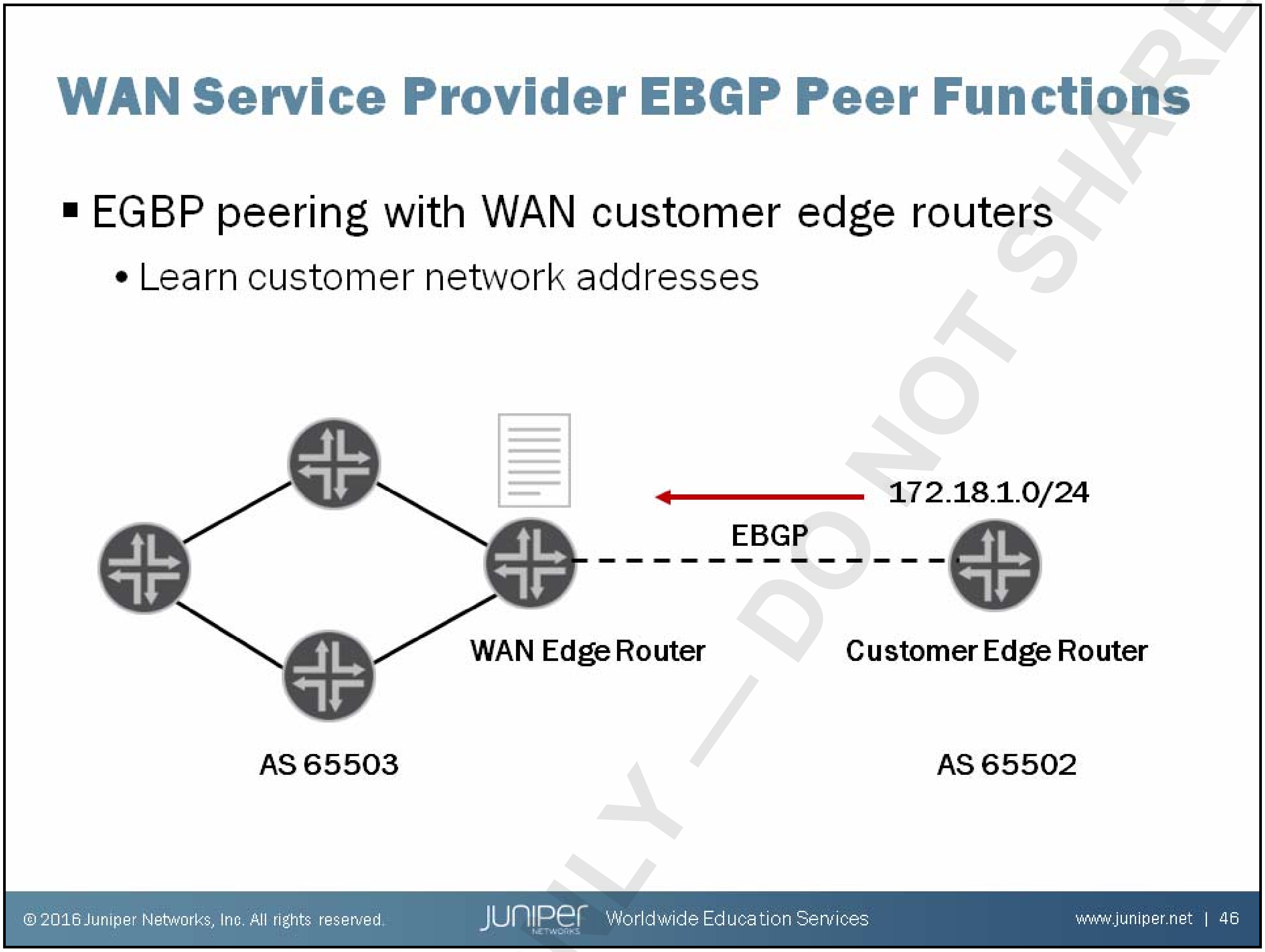
Functions of BGP in the WAN Core

- BGP WAN core functions
  - Establish IBGP loopback peering between internal WAN routers
  - Establish EBGP peer relationships between WAN provider edge router and customer edge router to learn customer network destinations
  - Distribute customer network destinations across the WAN
  - Communicate necessary address families
  - Build the foundation for WAN layer2/layer3 forwarding services

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 45

Functions Of BGP In The WAN Core

The BGP routing protocol is configured for IBGP peering between service provider core and edge routers in a full mesh or more commonly route reflector configuration. The IBGP sessions are configured to communicate the proper address families for various service provider layer 2 and layer 3 WAN connectivity offerings. The IBGP peering sessions will be used to communicate WAN customer network destinations. Service provider WAN edge routers will establish an EBGP peering relationship with WAN customer edge routers. Proper core BGP design enables service providers to offer multiple WAN connectivity solutions to customers.



WAN Service Provider EBGP Peer Functions

- EGBP peering with WAN customer edge routers
  - Learn customer network addresses

172.18.1.0/24

EBGP

WAN Edge Router

Customer Edge Router

AS 65503

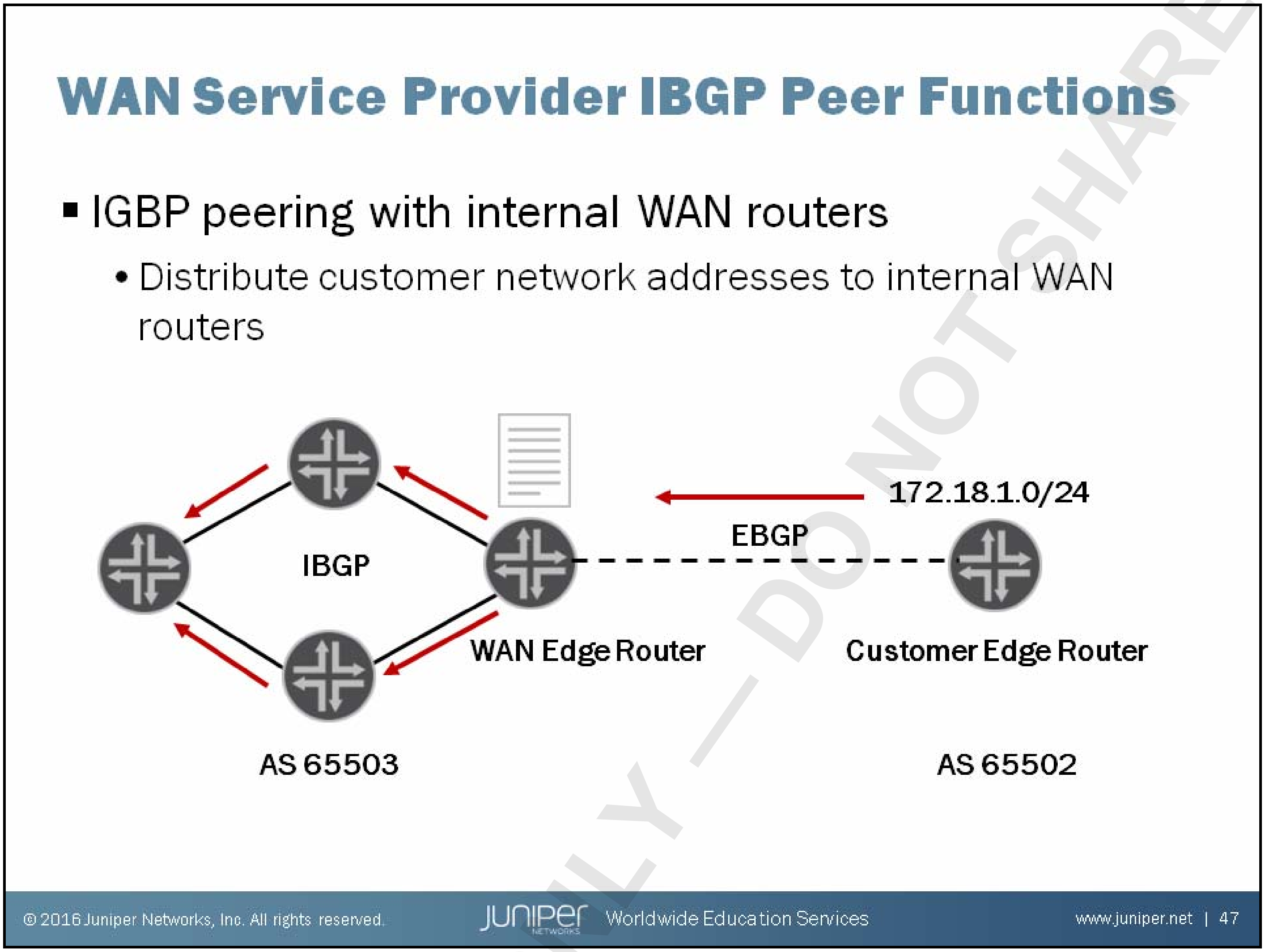
AS 65502

Worldwide Education Services

www.juniper.net | 46

WAN Service Provider EBGP Peer Functions

WAN designers define procedures for EBGP peering with customer edge routers. WAN service providers use this EBGP peer relationship to learn customer internal network addresses. Routes learned from customers are placed into separate routing instances configured on WAN provider edge routers. Organizing customer routing information into separate routing instances on the WAN provider edge routers eliminates address conflict issues and provides security and separation for customer routing information.



WAN Service Provider IBGP Peer Functions

- IGBP peering with internal WAN routers
  - Distribute customer network addresses to internal WAN routers

172.18.1.0/24

EBGP

IBGP

WAN Edge Router

Customer Edge Router

AS 65503

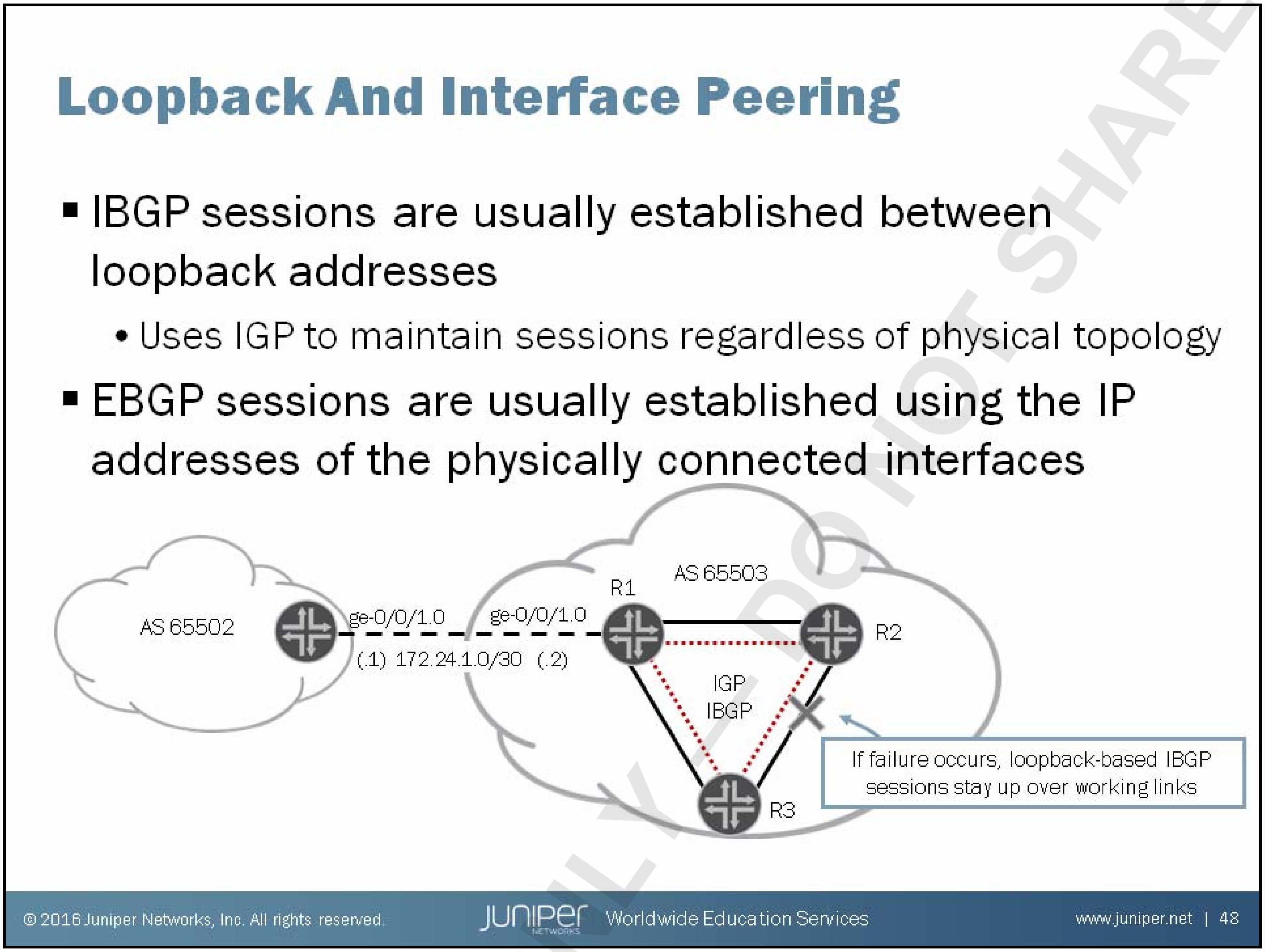
AS 65502

Juniper Worldwide Education Services

www.juniper.net | 47

WAN Service Provider IBGP Peer Functions

Core network designers define standard procedures for IBGP peering between service provider edge and core routers. WAN service providers use these IBGP peering relationships to distribute the customers internal network destinations across the WAN.



Loopback And Interface Peering

- IBGP sessions are usually established between loopback addresses
  - Uses IGP to maintain sessions regardless of physical topology
- EBGP sessions are usually established using the IP addresses of the physically connected interfaces

If failure occurs, loopback-based IBGP sessions stay up over working links

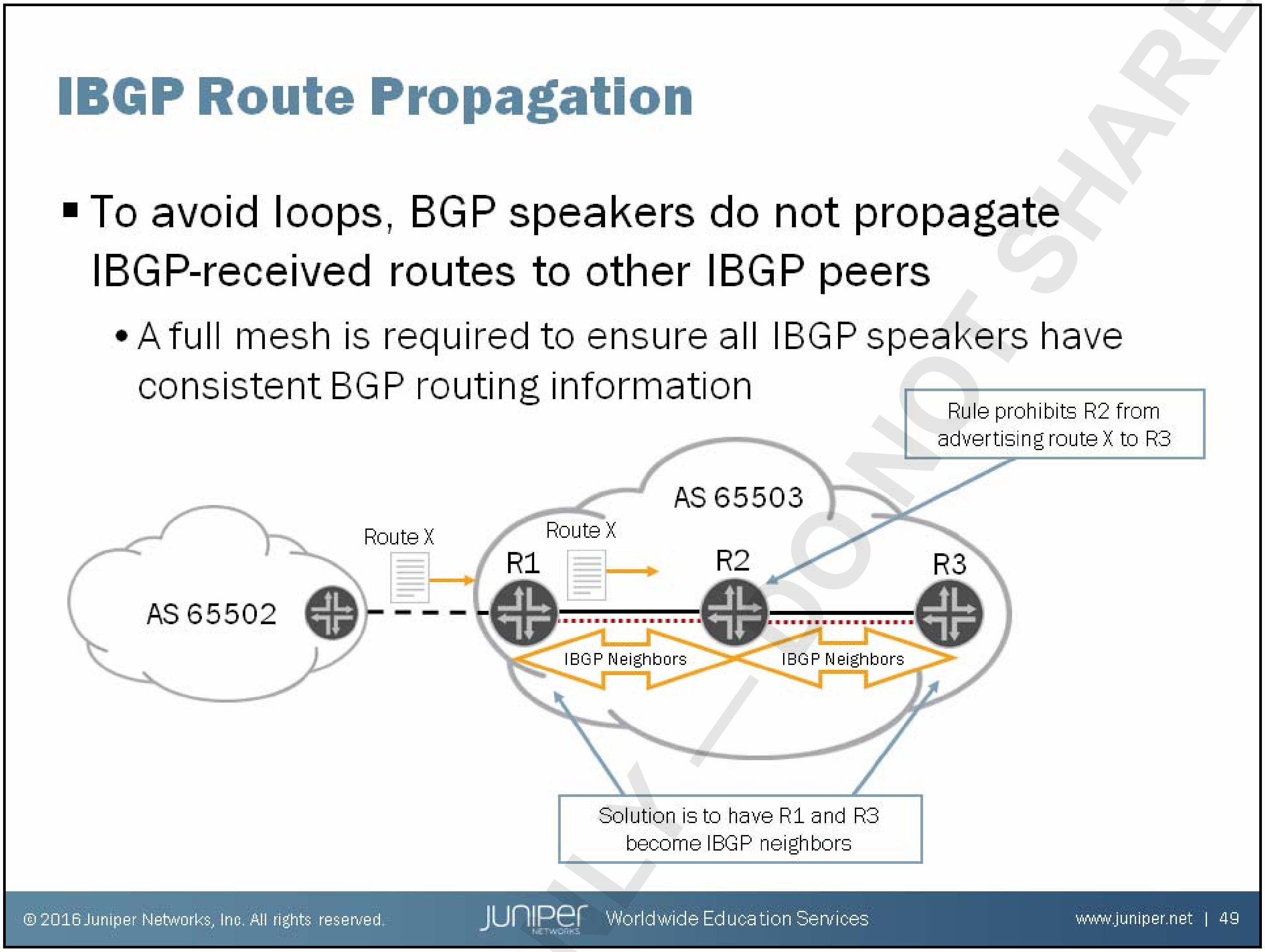
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 48

IBGP

Loopback peering maintains only one IBGP session between each internal peer. The IGP is used to maintain reachability between the loopback addresses regardless of the physical topology, allowing the IBGP sessions to stay up even when the physical topology changes. The physical topology is relevant in one respect: each router along the path between BGP speakers must have enough information to make consistent routing decisions about packet forwarding.

EBGP

EBGP sessions are simply BGP sessions between two routers in different ASs. When two EBGP peers have a single path between them, EBGP sessions are usually established over the shared subnet between two peers, using the IP addresses assigned to the interfaces on that subnet as the session endpoints. By establishing the EBGP session using the IP address assigned to the interfaces on the shared subnet, you gain many advantages. One of these advantages is that you prevent either AS from needing to maintain any routing information about the other AS (besides what it received through BGP). You also ensure that all traffic flows over this particular shared subnet.



IBGP Route Propagation

- To avoid loops, BGP speakers do not propagate IBGP-received routes to other IBGP peers
  - A full mesh is required to ensure all IBGP speakers have consistent BGP routing information

Rule prohibits R2 from advertising route X to R3

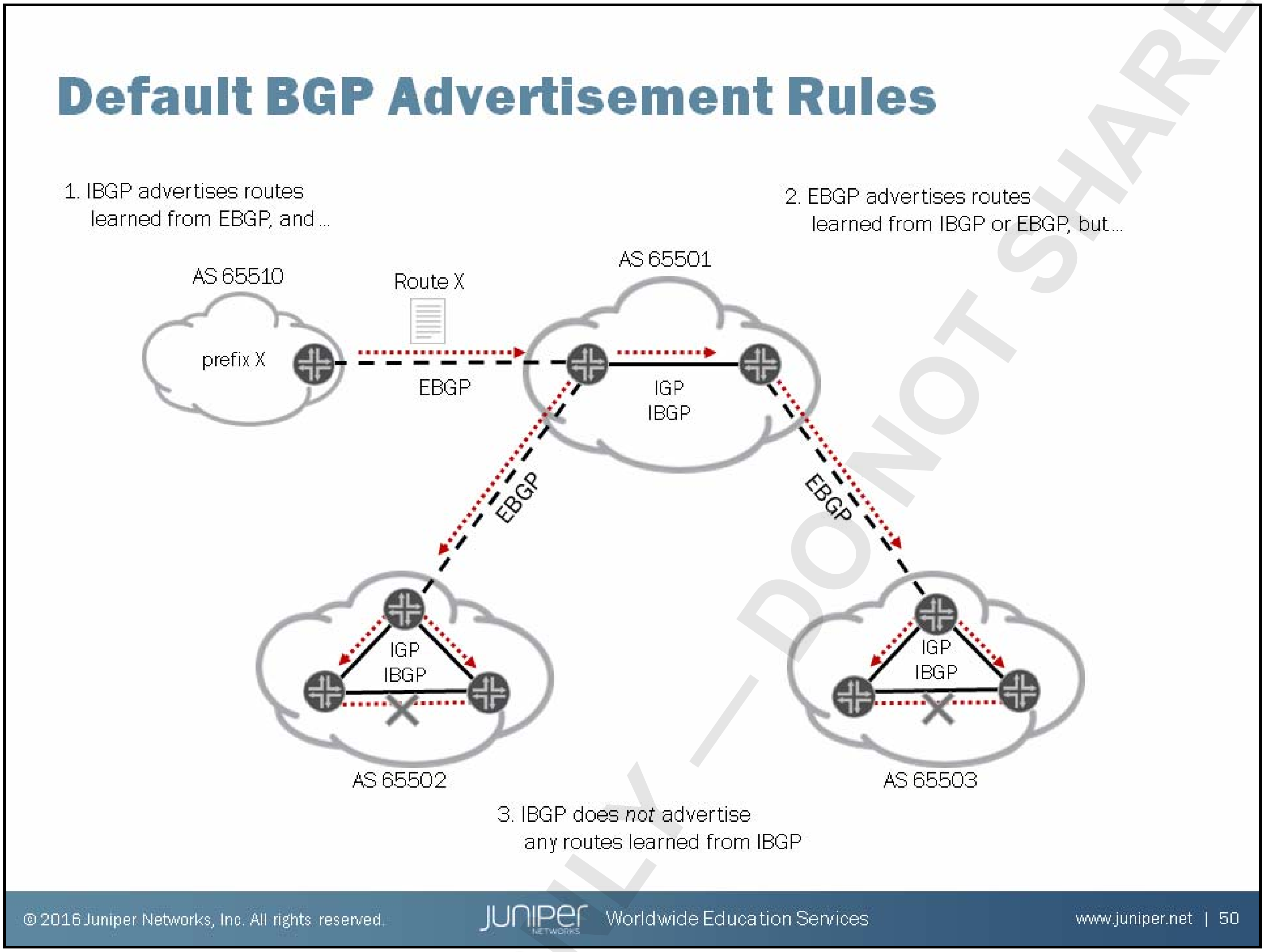
Solution is to have R1 and R3 become IBGP neighbors

Worldwide Education Services

www.juniper.net | 49

IBGP Route Propagation

IBGP speakers send routes to their IBGP peers that they received from EBGP peers and routes that they originated themselves. IBGP speakers never send routes to IBGP peers that they learned from other IBGP peers. For all IBGP speakers in an AS to have consistent routing information, there must be a full mesh of IBGP sessions between all BGP speakers. Without this full mesh, some BGP speakers might not receive all the required routing information. In the example on the slide, there is not a full mesh of IBGP sessions. R1 receives the announcement through an EBGP session. Because it is the best route it has for that prefix, it propagates the route to its IBGP peer R2. R2 also determines that route to be its best path for the prefix; however, it does not send the route to its IBGP peer R3. Because it received the route through IBGP, it cannot send the route to any IBGP peers. Therefore, R3 does not receive or install a route for the prefix advertised from AS 65502. This situation can be alleviated by adding an IBGP session between R1 and R3. (It is irrelevant whether the two routers are directly connected.) If IBGP routers re-advertised IBGP routes to other IBGP peers, a loop would form. Because the AS path is not updated by each router, but rather only when the associated prefix is advertised to an EBGP peer, the AS path cannot be used to detect loops for BGP routes advertised within an AS. For this reason, BGP enforces advertisement rules that require the full-mesh peering of IBGP routers to ensure consistent routing information on all IBGP routers within the AS. Route reflectors or confederations can eliminate the full-mesh requirement. These topics are discussed later in the chapter.



Default BGP Advertisement Rules

1. IBGP advertises routes learned from EBGP, and ...

2. EBGP advertises routes learned from IBGP or EBGP, but ...

3. IBGP does *not* advertise any routes learned from IBGP

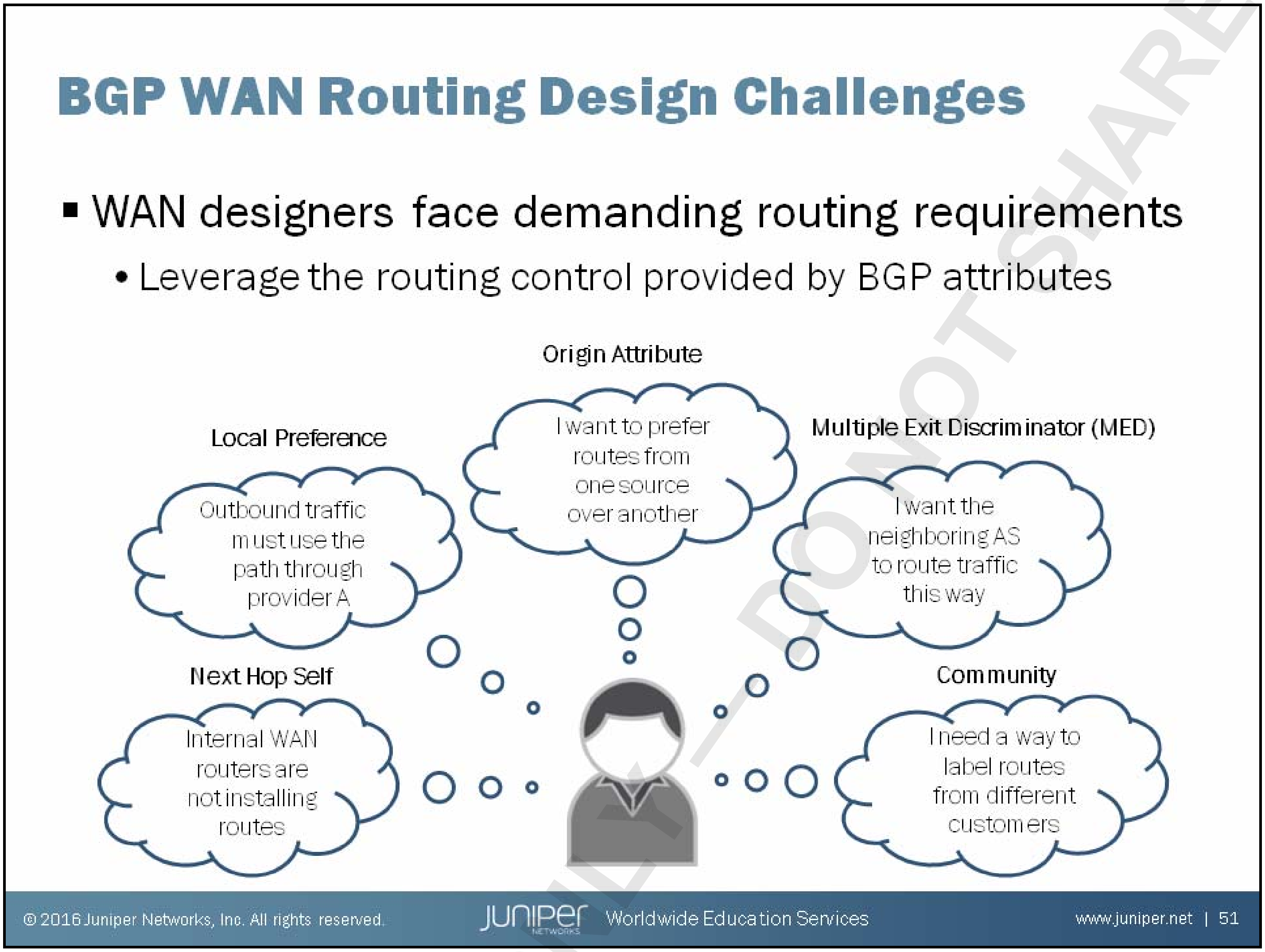
Juniper Networks Worldwide Education Services www.juniper.net | 50

Default BGP Advertisement Rules

By default, only active BGP routes are advertised. The slide illustrates the default BGP advertisement rules. The rules are as follows:

1. IBGP peers advertise routes received from EBGP peers to other IBGP peers.
2. EBGP peers advertise routes learned from IBGP or EBGP peers to other EBGP peers.
3. IBGP peers do not advertise routes received from IBGP peers to other IBGP peers.

The purpose of the advertisement rules is to prevent routing loops on a BGP network.



BGP WAN Routing Design Challenges

- WAN designers face demanding routing requirements
  - Leverage the routing control provided by BGP attributes

Origin Attribute

I want to prefer routes from one source over another

Local Preference

Outbound traffic must use the path through provider A

Multiple Exit Discriminator (MED)

I want the neighboring AS to route traffic this way

Next Hop Self

Internal WAN routers are not installing routes

Community

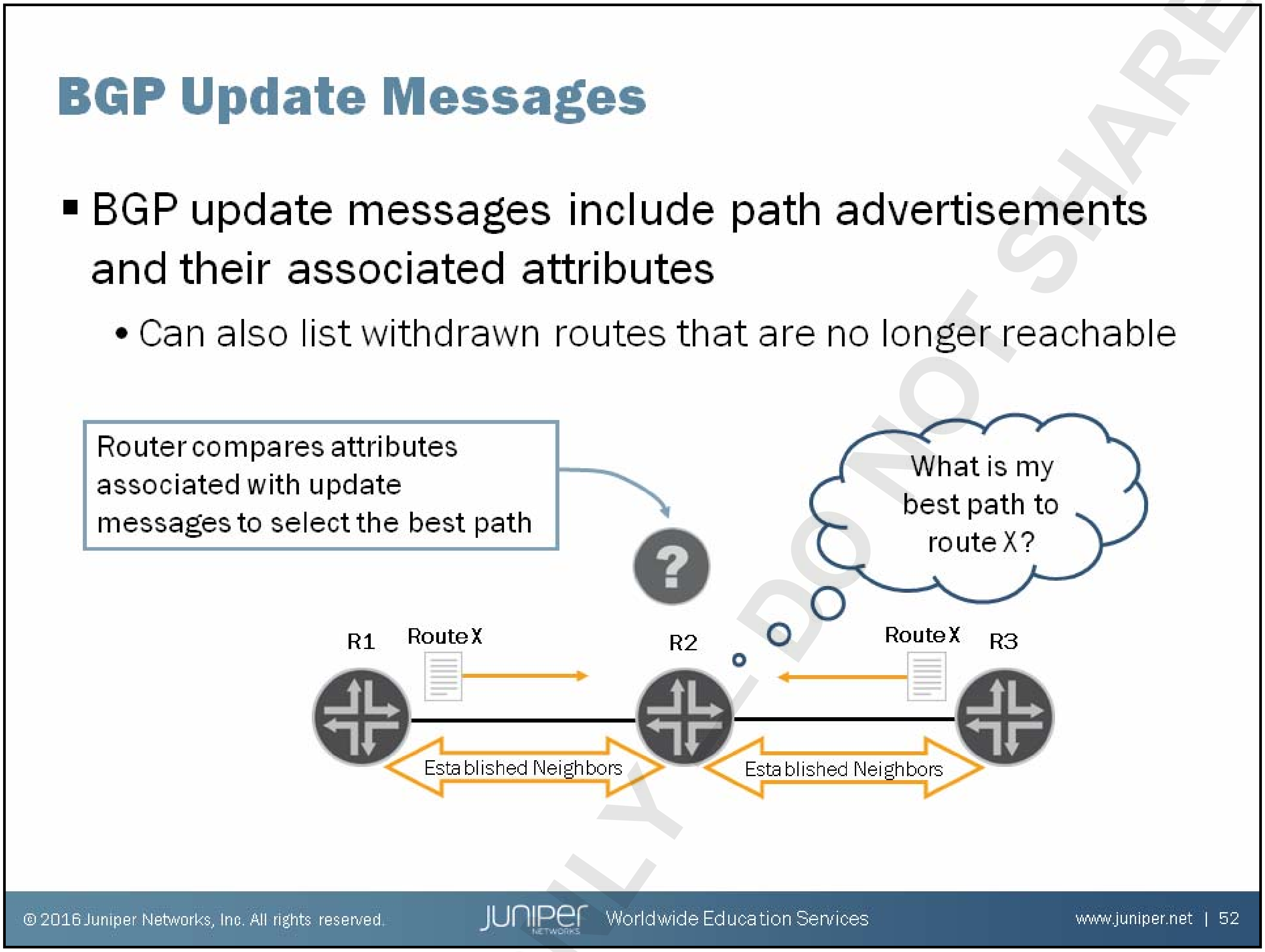
I need a way to label routes from different customers

Juniper Networks Worldwide Education Services

www.juniper.net | 51

BGP Routing Design Challenges

BGP attributes provide granular control of routing information. Attributes are analyzed when routes are received from neighbors and determine the active routes used to forward traffic and which routes to advertise to neighbors. Let’s analyze how BGP attributes are used.



BGP Update Messages

- BGP update messages include path advertisements and their associated attributes
  - Can also list withdrawn routes that are no longer reachable

Router compares attributes associated with update messages to select the best path

What is my best path to route X?

R1 Route X R2 Route X R3

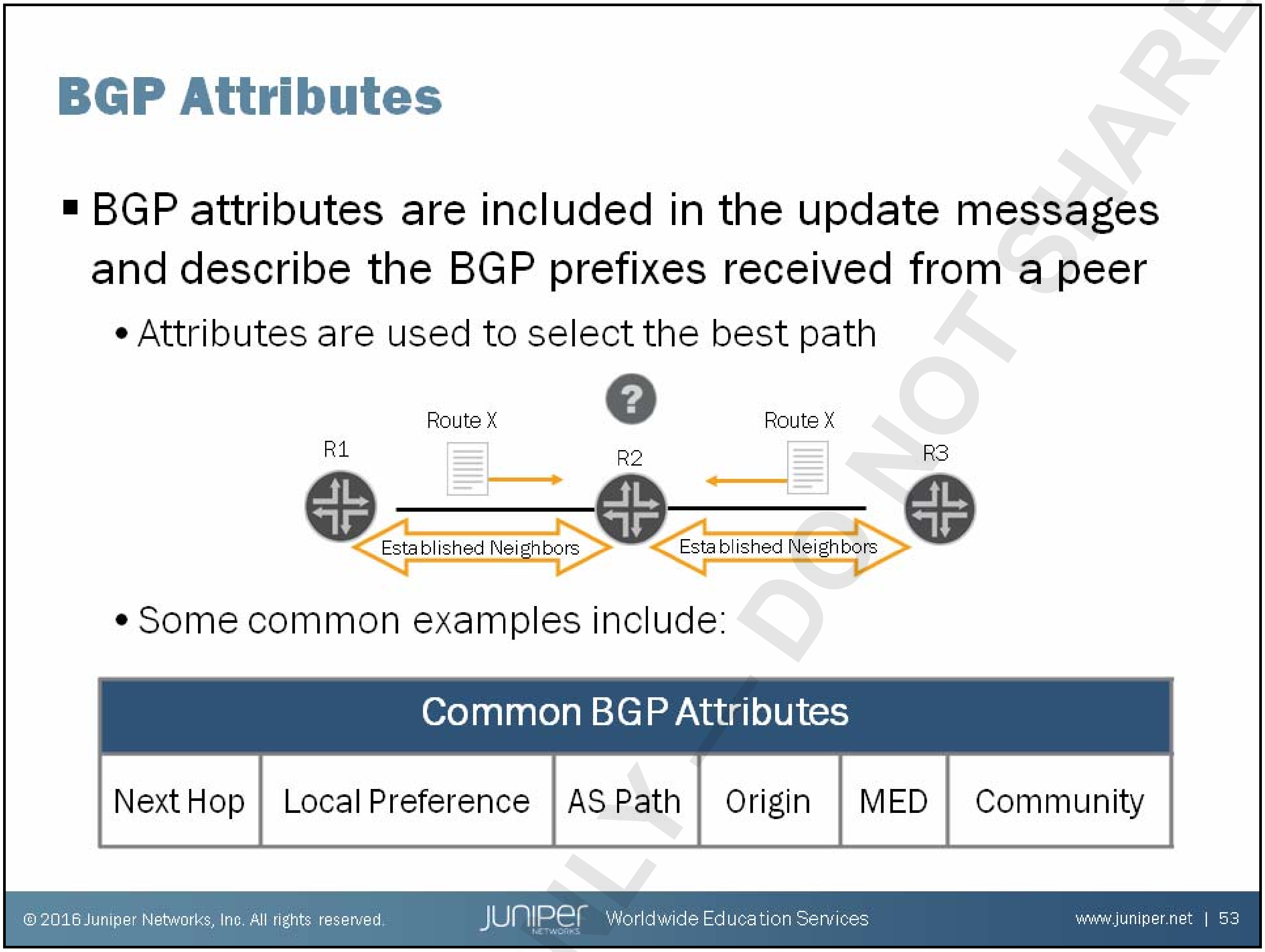
Established Neighbors

Established Neighbors

Juniper Networks Worldwide Education Services www.juniper.net | 52

BGP Update Messages

BGP update messages describe a single path and then list multiple prefixes that can be reached through this same path. BGP peers assume that this information is unchanged unless a subsequent update advertises a new path for a prefix or lists the prefix as unreachable. Updates can list any prefixes that are no longer reachable, regardless of the path associated with those prefixes. BGP peers use update messages to ensure that their neighbors have the most up-to-date information about BGP routes. BGP uses TCP to provide reliable communication, which ensures that BGP neighbors never miss an update. A system of keepalives also allows each BGP peer to ensure that its neighbor is still functioning properly. If a neighbor goes down, the BGP speaker deletes all routes learned from that peer and updates its other peers accordingly. BGP uses the information within the update messages, in particular the BGP attributes, to detect routing loops and determine the best path for a given destination prefix.



BGP Attributes

- BGP attributes are included in the update messages and describe the BGP prefixes received from a peer
  - Attributes are used to select the best path

?
Route X
R1
R2
Route X
R3
Established Neighbors
Established Neighbors

- Some common examples include:

Common BGP Attributes					
Next Hop	Local Preference	AS Path	Origin	MED	Community

Juniper Networks Worldwide Education Services

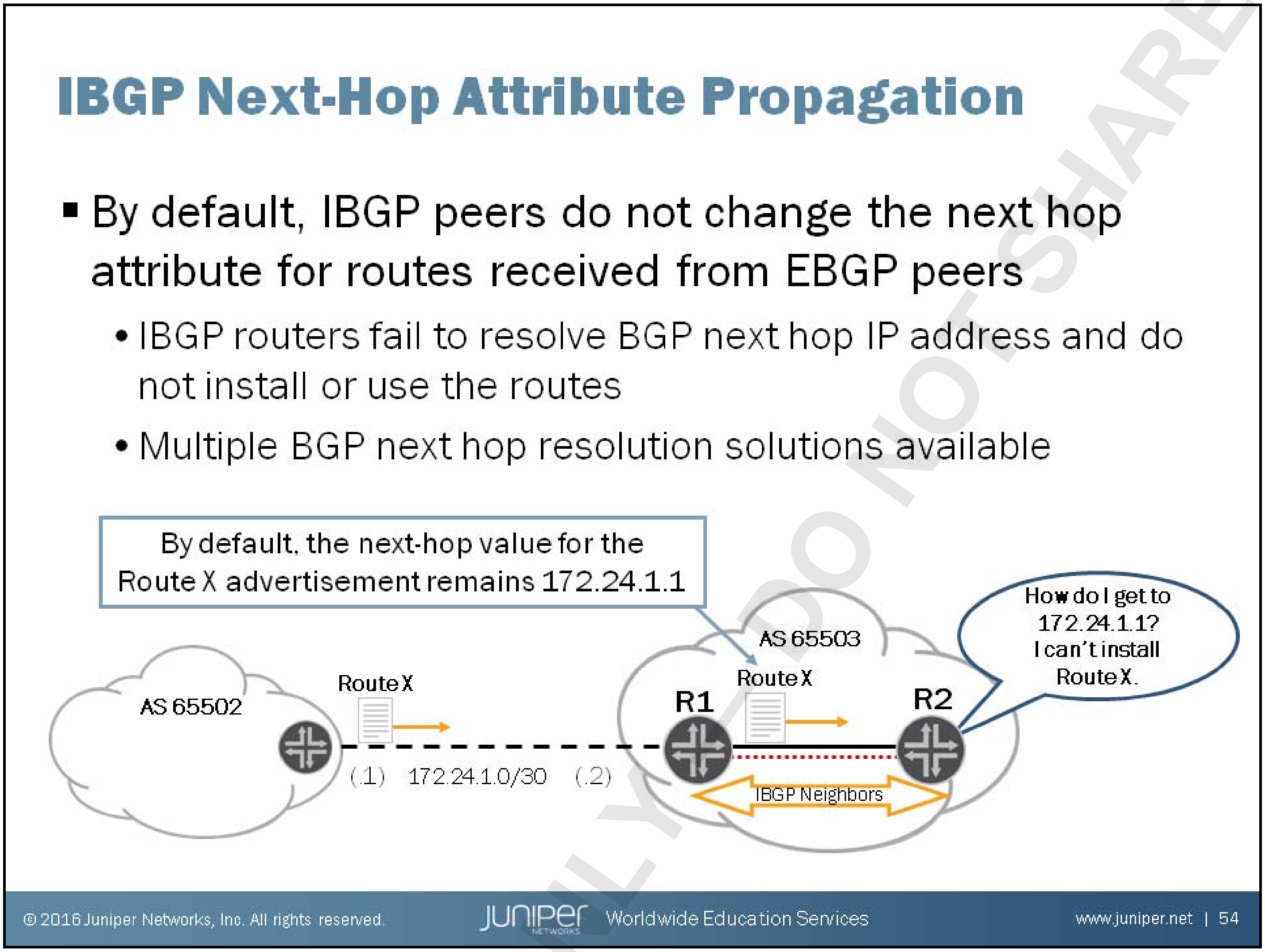
www.juniper.net | 53

BGP Attributes

The primary purpose of BGP is not to find the shortest path to a given destination; rather, its purpose is to find the best path. Each AS determines the best path to a prefix by taking into account its own outbound routing preferences, the inbound routing preferences of the route's originator (as updated by ASs along the path between the source and destination ASs), and some information that is collected about the path itself. All this information is contained in path attributes that describe the path to a prefix. The path attributes contain the information that BGP uses to implement the routing policies of source, destination, and transit autonomous systems.

The BGP attribute classes are described in RFC 1771 and include the well-known mandatory, well-known discretionary, optional transitive, and optional nontransitive classes.

- Well-known mandatory: Must be supported by all BGP implementations and must be included in every BGP update.
- Well-known discretionary: Must be supported in all BGP implementations, but do not have to be included in every BGP update.
- Optional transitive: Not required to be supported by all BGP implementations but, if they are, they should be passed along, unchanged, to other BGP peers.
- Optional nontransitive: Not required to be supported by all BGP implementations. If an optional nontransitive attribute is not recognized, it is ignored and not passed to other peers.



IBGP Next-Hop Attribute Propagation

- By default, IBGP peers do not change the next hop attribute for routes received from EBGP peers
  - IBGP routers fail to resolve BGP next hop IP address and do not install or use the routes
  - Multiple BGP next hop resolution solutions available

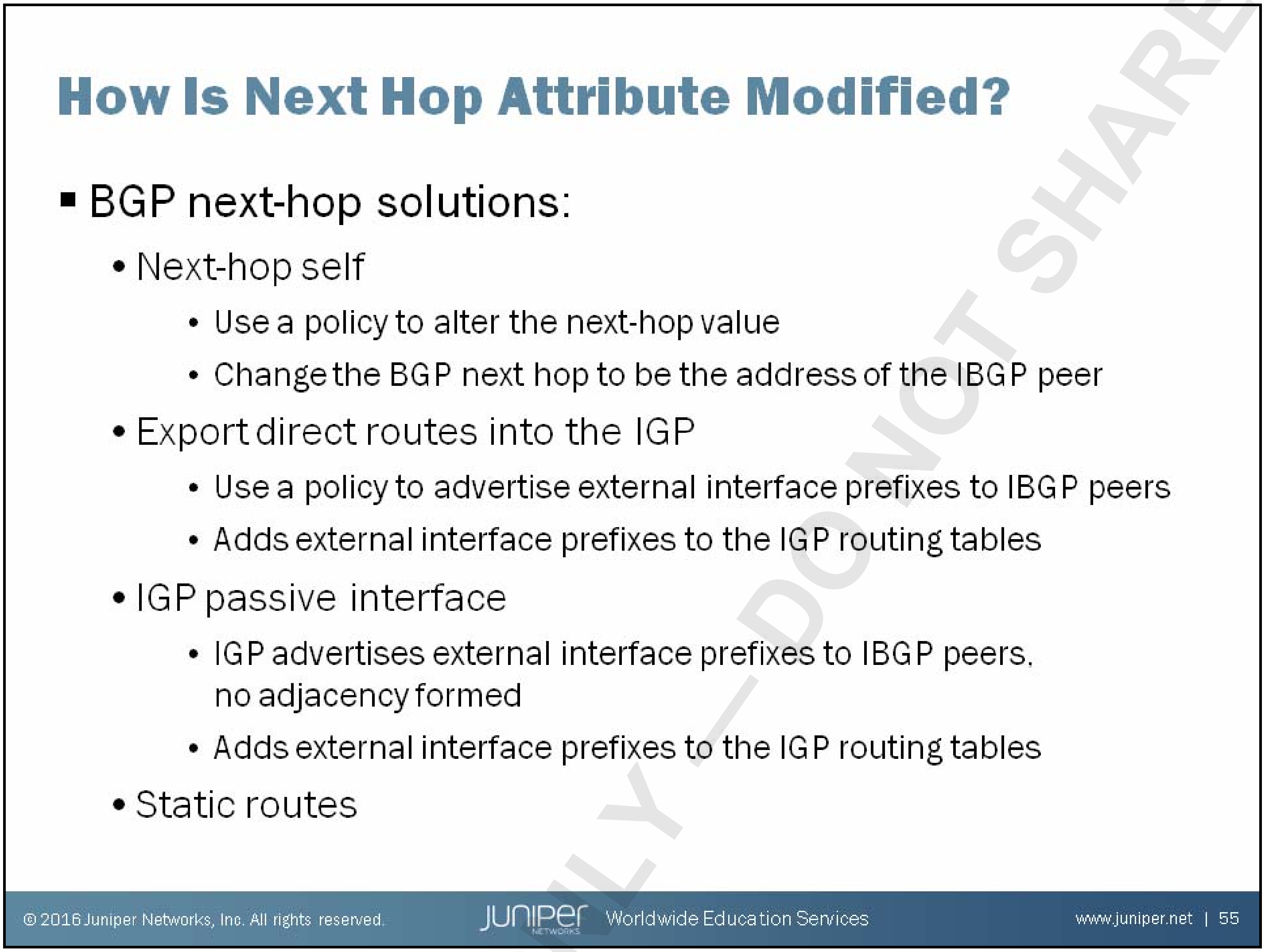
By default, the next-hop value for the Route X advertisement remains 172.24.1.1

How do I get to 172.24.1.1? I can't install Route X.

Juniper Networks Worldwide Education Services www.juniper.net | 54

IBGP Next-Hop Propagation

The BGP next-hop attribute is an IP address of a BGP peer. It is used to verify connectivity of a remote BGP peer. A BGP peer can be an immediately connected host or a remote host. Recursive lookups into routing tables accomplish the peer connectivity. The IP address specified in the next-hop field must be reachable by the local router before the route becomes active in the routing table. By default, the router that originally sourced the route into BGP places its peer address into the attribute field. The next-hop value is then typically changed when the route is transmitted across EBGP links. By default, the next-hop attribute attached to a route is unchanged as it passes through an AS. Because routers can use the BGP routes only if they already have a route to the next hop, you must either configure the routers to advertise external interfaces through the IGP, or configure the routers to change the next-hop attribute attached to BGP routes using policy. When EBGP speakers send routes to a peer, they set the next-hop attribute to the interface they share with that peer. In this example, R1 receives a route from its EBGP peer with the next-hop attribute set to 172.24.1.1. R1 sends this route to R2 without changing the next-hop attribute. Therefore, to use this route, R2 either must know how to reach 172.24.1.1 through the IGP or static routing, or R1 must send the routes with a different next hop.



How Is Next Hop Attribute Modified?

- BGP next-hop solutions:
  - Next-hop self
    - Use a policy to alter the next-hop value
    - Change the BGP next hop to be the address of the IBGP peer
  - Export direct routes into the IGP
    - Use a policy to advertise external interface prefixes to IBGP peers
    - Adds external interface prefixes to the IGP routing tables
  - IGP passive interface
    - IGP advertises external interface prefixes to IBGP peers, no adjacency formed
    - Adds external interface prefixes to the IGP routing tables
  - Static routes

Worldwide Education Services

www.juniper.net | 55

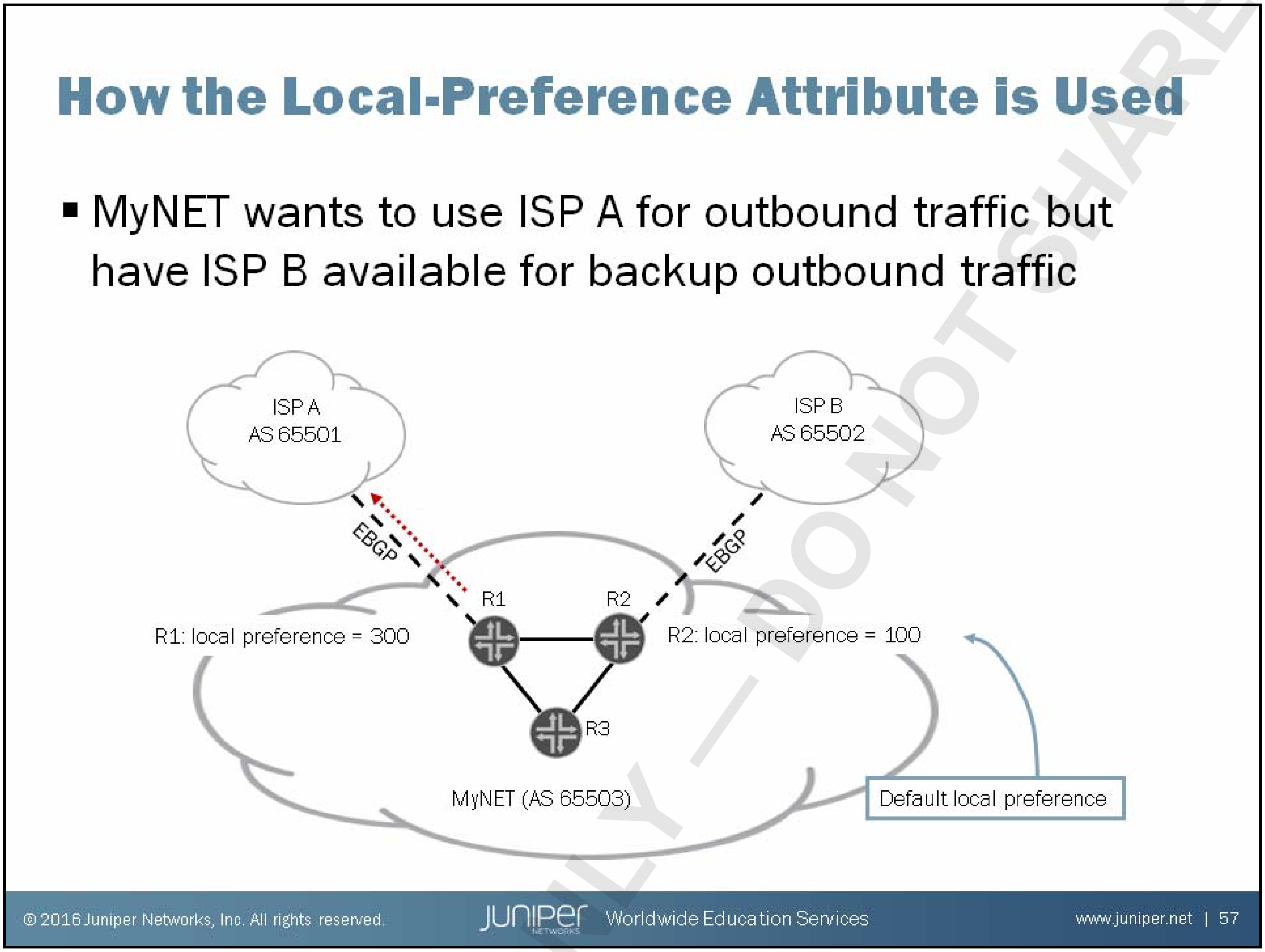
BGP Next-Hop Solutions

Numerous ways exist to solve this BGP next-hop reachability problem, and five examples are listed on the slide. Some of these examples do technically solve the reachability issue but are not best practices in a networking environment. The most commonly used (and recommended) solution is next-hop self. With this solution, when a BGP router advertises an EBGP-learned route to an IBGP peer, it alters the BGP next-hop attribute. The next-hop attribute's IP address of the remote EBGP peer is replaced with the IP address of the BGP router itself. Because the IBGP session was most likely established using the peer's loopback address, this new BGP next-hop value is reachable, and the advertised BGP route can be used. We create next-hop self by using a policy to match specific routes with an action of changing the next-hop attribute value. The Junos OS then applies this policy as an export policy to any IBGP peers. The next two options listed (export direct routes and IGP passive) are almost identical in their results. The difference between the two is in the approach that each takes to provide reachability. With export direct, the IGP operating in the AS with a routing policy advertises the address assigned to the point-to-point link between the two EBGP peers to all IBGP peers.

Continued on the next page.

BGP Next-Hop Solutions (contd.)

Export direct uses a Junos OS routing policy to retrieve the subnet information from the local routing table. Within inet.0, these networks are known as protocol direct. The policy matches these direct routes and accepts them. The Junos OS then applies this policy as an export policy to the local IGP. With IGP passive, the IGP is configured on the inter-AS link and advertises the interface addresses, but forms no adjacency (it is passive). Both methods inject the interface addresses into the local routing table for the IGP to use. An IGP passive interface allows the local IGP to advertise the subnet on a particular interface without forming an adjacency at the IGP level to the remote EBGP peer, which has the advantage of not using a policy, but it requires explicit configuration for each interface and subnet address that you want to advertise. The last two options listed on the slide (static routes and forming an IGP adjacency relationship with the remote EBGP router) have some severe disadvantages, but they both work. Static routes have an inherent scalability problem. You must configure each IBGP router in the network for a single static route for each remote EBGP peer. The more EBGP peers in the network, the more static routes required. The more IBGP peers in the AS, the more places that additions and changes must be made. Clearly, this option is not a real world option. The final option is to establish a full IGP adjacency between AS networks. Although reachability information can be provided by forming an IGP relationship with the remote EBGP peer, we do not recommend this practice because of the very trusting nature of the IGP protocols. Once this adjacency is formed, the protocol accepts any routing information the remote EBGP peer provides. This behavior is very dangerous because the remote AS might inject bad information into your network. In addition, this method potentially violates the entire idea of having autonomous systems in the first place.



How the Local-Preference Attribute is Used

- MyNET wants to use ISP A for outbound traffic but have ISP B available for backup outbound traffic

ISP A
AS 65501

ISP B
AS 65502

EBGP

EBGP

R1

R2

R3

R1: local preference = 300

R2: local preference = 100

Default local preference

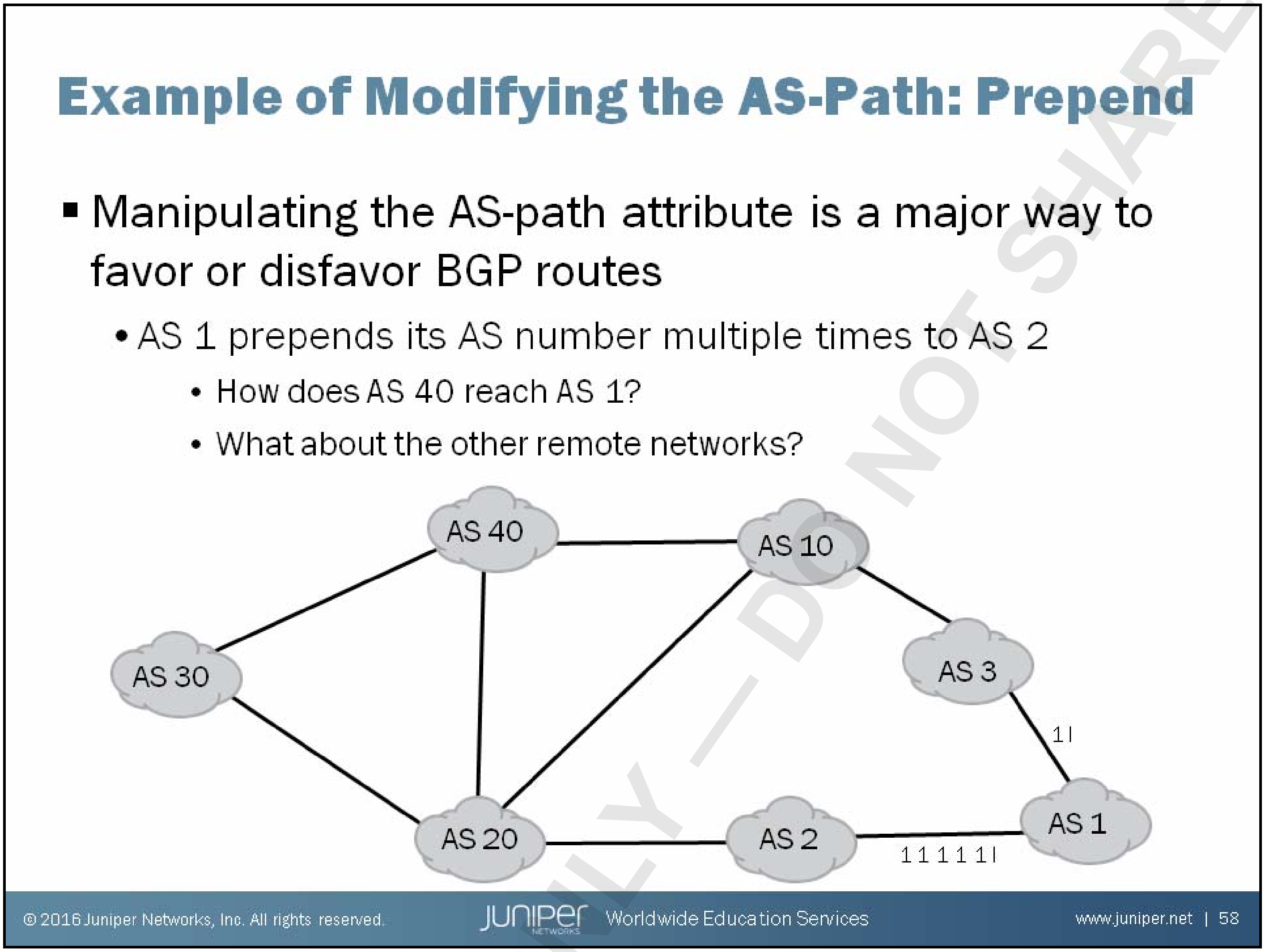
MyNET (AS 65503)

© 2016 Juniper Networks, Inc. All rights reserved. JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 57

Local-Preference Example

You can use the local-preference attribute to direct all outbound traffic through a specific peer. Before sending the traffic to the internal peers, the designated peer sets the local-preference value on all routes received. Then all the peers use those routes in their RIB-local tables. The local-preference attribute is a numeric value; higher values indicate a better metric. The Junos operating system allows you to set the local-preference value using BGP configuration or through routing policy. If you set a local-preference value through both configuration and routing policy, the system uses the value assigned through routing policy. BGP uses the local-preference attribute only within an AS. Local-preference values are not transmitted across EBGP links.

The slide shows an example of how the local-preference attribute is used within an AS. The network administrators in MyNET have decided that the routers in MyNET should use ISP A whenever possible and ISP B as its secondary path. This decision might be made for several reasons, including cost of service and performance factors. To use the ISP A path, the MyNET network administrators set the local-preference attribute for all routes received from ISP A to 300 while all routes received from ISP B use the default local-preference value of 100. We assume that ISP A and ISP B are advertising a similar set of routes. Because of the higher local-preference value associated with the routes received from ISP A, the MyNET routers choose the path through ISP A rather than the path through ISP B. In contrast to other BGP attribute values, higher local-preference values are preferred.



Example of Modifying the AS-Path: Prepend

- Manipulating the AS-path attribute is a major way to favor or disfavor BGP routes
  - AS 1 prepends its AS number multiple times to AS 2
    - How does AS 40 reach AS 1?
    - What about the other remote networks?

Manipulating the AS Path

The AS-path attribute describes the path of autonomous systems that the route has been through since it was sourced into BGP. When a BGP router receives routes in an update message, the first action is to examine the current AS path to see if the local AS number is in the path. If it is in the path, it indicates that the route has been through the AS already; accepting the route would cause a loop. Therefore, BGP drops the route. The Junos OS performs a verification to ensure the receiving router's AS number is not listed in the AS path. If the receiving router's AS number is listed in the AS path, the router does not advertise the route. By default, the AS-path value is changed as a route transitions between autonomous systems. The AS-path value is null until the associated route is advertised out of the originating AS. As the route leaves an AS, the BGP router adds the local AS number to the front of the path before sending it to a peer. Using routing policy, you can prepend your autonomous system number information to the AS-path attribute. By prepending your autonomous system number information to the AS-path multiple times, you can affect the routing decision made by routers in other autonomous systems and discourage those routes from using that path because of the longer AS-path. The AS-path attribute is mandatory; thus, it must always be present for all BGP routes.

If the AS-path attribute is changed before the route is re-advertised to other BGP routers, a route through the local AS can look less attractive to another AS. Note that making the route more attractive by shortening the AS path should not really be possible. However, you can lengthen the path to make the AS-path attribute another type of negative-bias path selection mechanism. In other words, unlengthened paths look more attractive for other AS networks to use than artificially lengthened AS paths. The only standard approach to alter the AS-path attribute is to add information to it by prepending. You can use a routing policy to extend (by prepending) AS-path information artificially onto an existing AS path. This type of a policy is often an attempt to influence traffic coming into an AS from another AS.

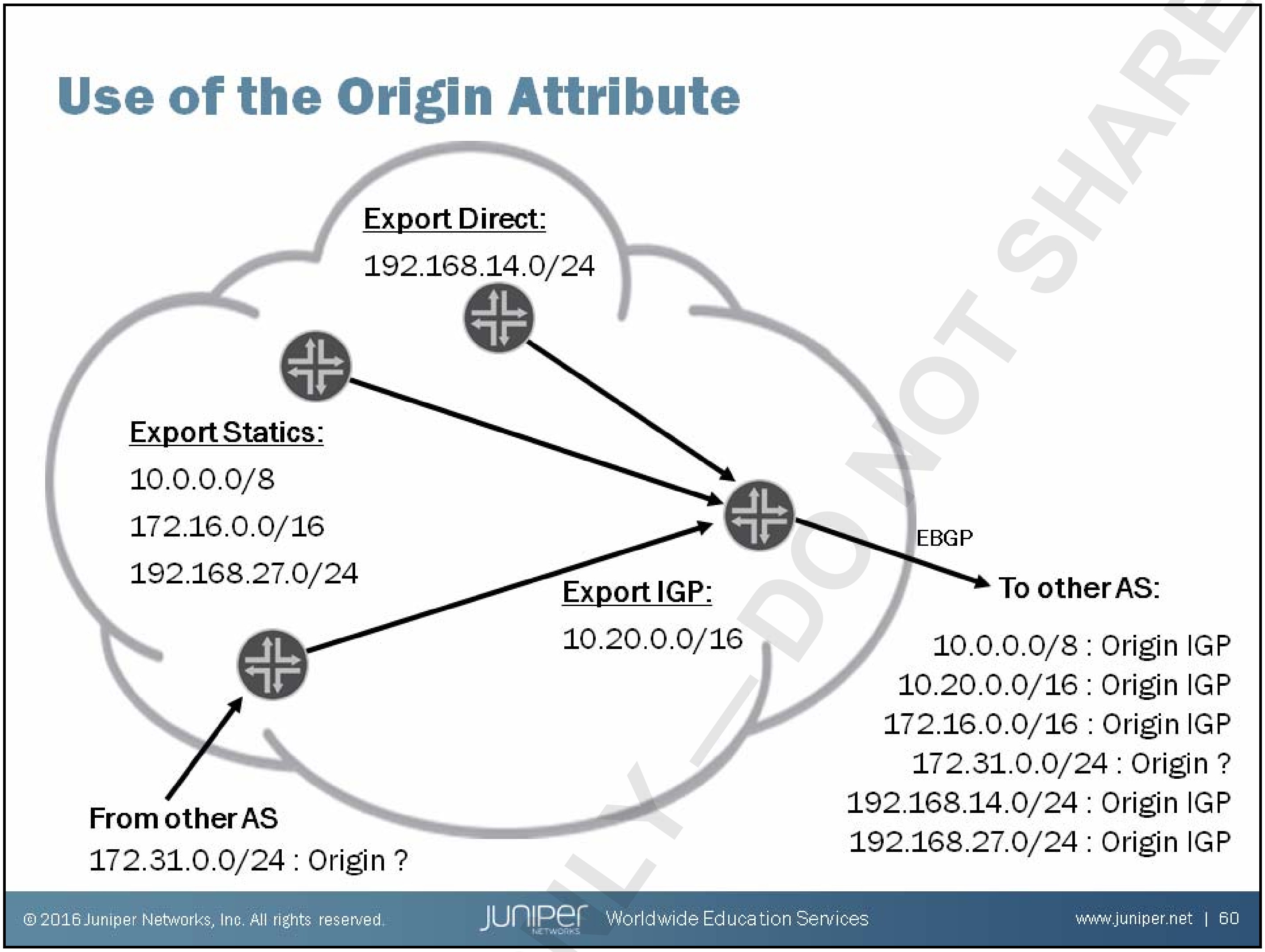
Continued on the next page.

Manipulating the AS Path (contd.)

On the slide, AS 1 announces its routes to both AS 2 and AS 3. Using a routing policy, AS 1 prepends its own AS number four times (shown as 1 1 1 1 1 on the slide) onto all route announcements to AS 2. This action causes the following:

- AS 2 will use AS 20 to forward packets to AS 1;
- AS 10 will use AS 3 to forward packets to AS 1;
- AS 20 will use AS 10 to forward packets to AS 1;
- AS 30 will use AS 40 to forward packets to AS 1; and
- AS 40 will use AS 10 to forward packets to AS 1.

Note that this behavior on the part of AS 2 (using AS 20 instead of sending directly to AS 1) is unexpected and would not occur without the routing policy. This behavior is extended to AS 20 as well, because AS 2 cannot shorten the AS path advertised by AS 1, even if AS 2 would like to shorten it.



Use of the Origin Attribute

Export Direct:
192.168.14.0/24

Export Statics:
10.0.0.0/8
172.16.0.0/16
192.168.27.0/24

Export IGP:
10.20.0.0/16

From other AS
172.31.0.0/24 : Origin ?

EBGP

To other AS:
10.0.0.0/8 : Origin IGP
10.20.0.0/16 : Origin IGP
172.16.0.0/16 : Origin IGP
172.31.0.0/24 : Origin ?
192.168.14.0/24 : Origin IGP
192.168.27.0/24 : Origin IGP

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 60

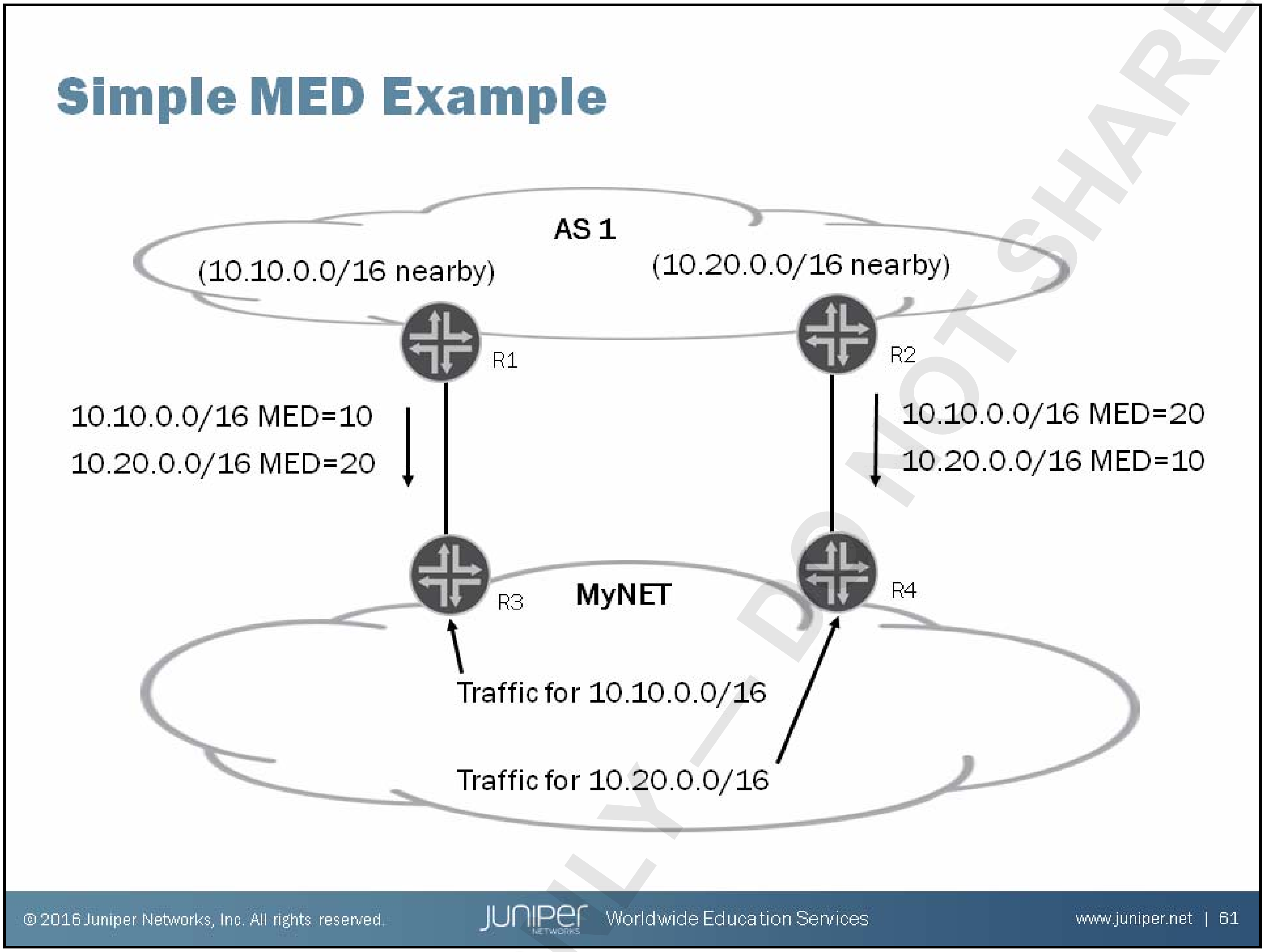
The Origin Attribute

The BGP router that injects a route into the BGP process is responsible for placing the origin attribute into the route. The origin attribute describes where the original router received the route. The following are the possible choices:

- IGP: BGP assigns an IGP route a value of 0. Examples of IGP routes include OSPF, IS-IS, static, and aggregate.
- EGP: BGP assigns EGP routes a value of 1. EGP routes are from the original EGP protocol, which was the predecessor to BGP.
- Incomplete: BGP assigns unknown routes a value of 2. Unknown routes are those that did not come from an IGP or from EGP.

The origin attribute is a mandatory attribute; it is transmitted across all BGP links. By default, the Junos OS assigns all routes injected into BGP an origin value of I for IGP. You can alter this default value using routing policy.

The slide shows the default BGP behavior within the Junos OS with regard to the origin code. Within the AS shown, the static routes of 10.0.0.0/8, 172.16.0.0/16, and 192.168.27.0/24 exist. An export policy placed these routes into BGP. A direct route of 192.168.14.0/24 exists that was exported into BGP. The route to 172.31.0.0/24 was learned from another AS altogether, and this route contains an origin attribute coded to 2, which indicates an unknown origin. Finally, an IGP-learned route of 10.20.0.0/16 exists in the network. The router does not know whether this route is an OSPF route or an IS-IS route, but the route had the appropriate export policy and, therefore, was placed into BGP. To the Juniper Networks router, it does not matter that these routes are advertised to another AS through EBGP; the BGP origin code is not altered as the routes are advertised to an EBGP peer. On the basis of the origin attribute alone, the 172.31.0/24 prefix appears less attractive to the remote AS.



Simple MED Example

AS 1

(10.10.0.0/16 nearby)

(10.20.0.0/16 nearby)

R1

R2

10.10.0.0/16 MED=10

10.20.0.0/16 MED=20

10.10.0.0/16 MED=20

10.20.0.0/16 MED=10

R3

MyNET

R4

Traffic for 10.10.0.0/16

Traffic for 10.20.0.0/16

The MED Attribute

By default, BGP uses the multiple exit discriminator (MED) value only when the BGP router's AS has two or more connections to the same upstream AS. You can determine the existence of multiple connections by examining the AS-path attribute to find that multiple routes from the same AS being advertised by multiple, different peers exist. An AS uses the MED value in an attempt to influence data traffic headed back toward the AS. The local AS sets the MED value differently on separate peers headed toward the same remote AS. The remote AS picks routes based on the lowest MED value it finds. The remote AS then uses that local peer to route traffic back to the local AS. BGP routes do not require the MED attribute. If it is missing, BGP assumes the MED value to be 0.

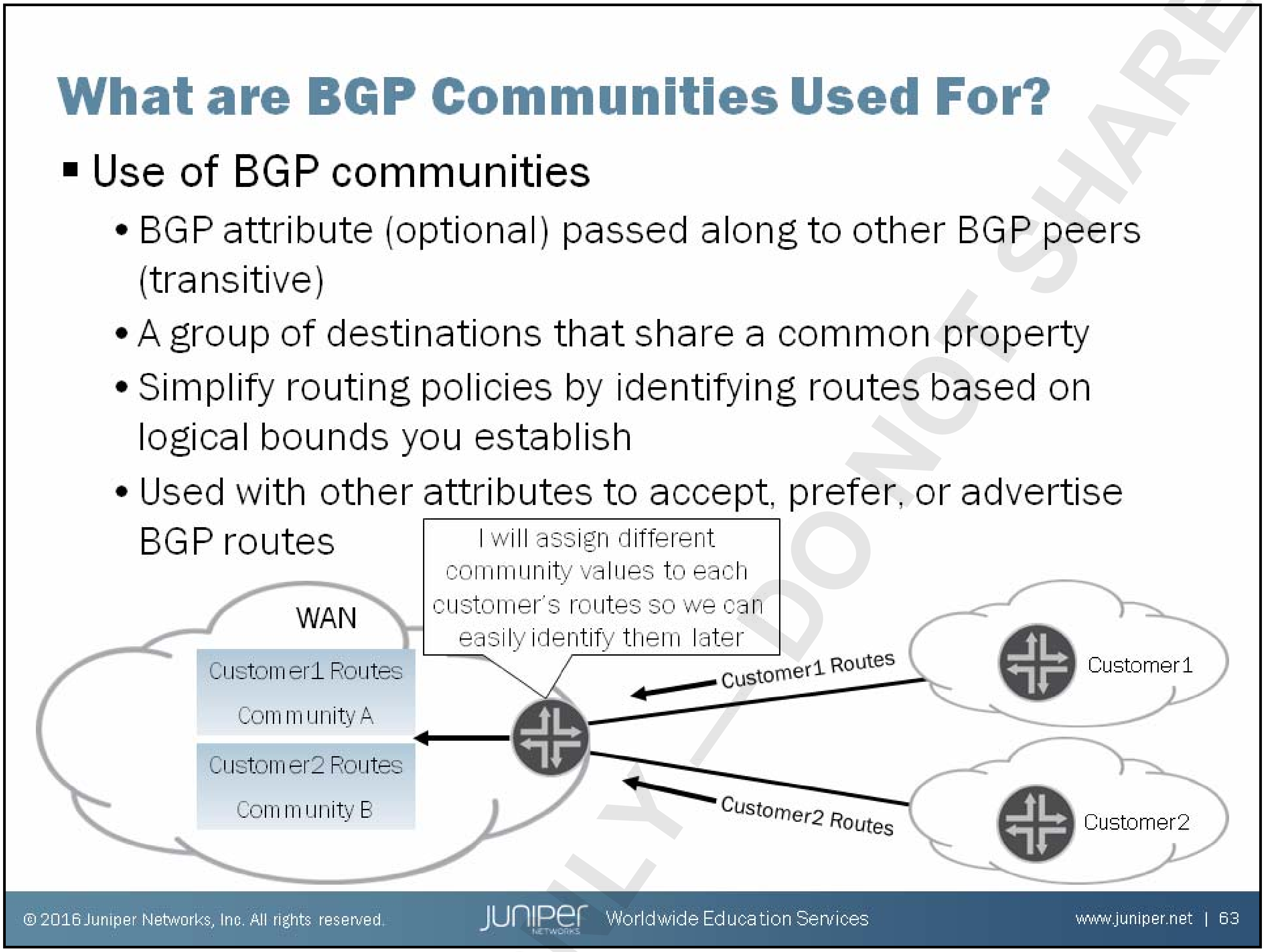
Simple MED Example

The slide shows a very basic example in the use of the BGP MED attribute to influence inter-AS traffic flows. AS 1 assigned its IP address spaces so that it can summarize its network into two major segments. Furthermore, AS 1 is relatively cleanly divided into networks near the left most router (10.10.0.0/16 networks are nearby) and networks near the right most router (10.20.0.0/16 networks are nearby). Perhaps the split is between Eastern and Western operations, but there are many other alternatives. AS 1 has two EBGP sessions to the customer Acme and advertises both the 10.10.0.0/16 and the 10.20.0.0/16 networks to Acme on each EBGP session, as shown on the slide. Naturally, AS 1 would like Acme to return traffic to the closest point in the network so that timely packet delivery and low latency can be achieved. Ordinarily, AS 1 would have no real way to convey this desire to Acme, and Acme would simply send traffic to AS 1 over whichever router Acme decides to use based on its own routing policies. However, the MED attribute offers a method for AS 1 to influence the routing policy on Acme for traffic sent to AS 1. To accomplish this closest point goal, AS 1 alters the MED values on the routes that it advertises to Acme with a BGP export routing policy.

Continued on the next page.

Simple MED Example (contd.)

Both of the networks in AS 1 are advertised across both links for redundancy. All things being equal, the routers in Acme still see multiple network paths to the routes in AS 1 as the AS 1 routes are passed along throughout Acme. The Acme routers, however, use the MED values of 10 and 20 (10 being preferred) to choose the BGP path to install in their local routing tables. Thus, within Acme, traffic to 10.10.0.0/16 networks flows to the left most router and out to AS 1, while traffic to 10.20.0.0/16 networks flows to the right-most router and out to AS 1. AS 1 influenced Acme, and, at the same time, achieved a primitive type of load balancing.



What are BGP Communities Used For?

- Use of BGP communities
  - BGP attribute (optional) passed along to other BGP peers (transitive)
  - A group of destinations that share a common property
  - Simplify routing policies by identifying routes based on logical bounds you establish
  - Used with other attributes to accept, prefer, or advertise BGP routes

I will assign different community values to each customer's routes so we can easily identify them later

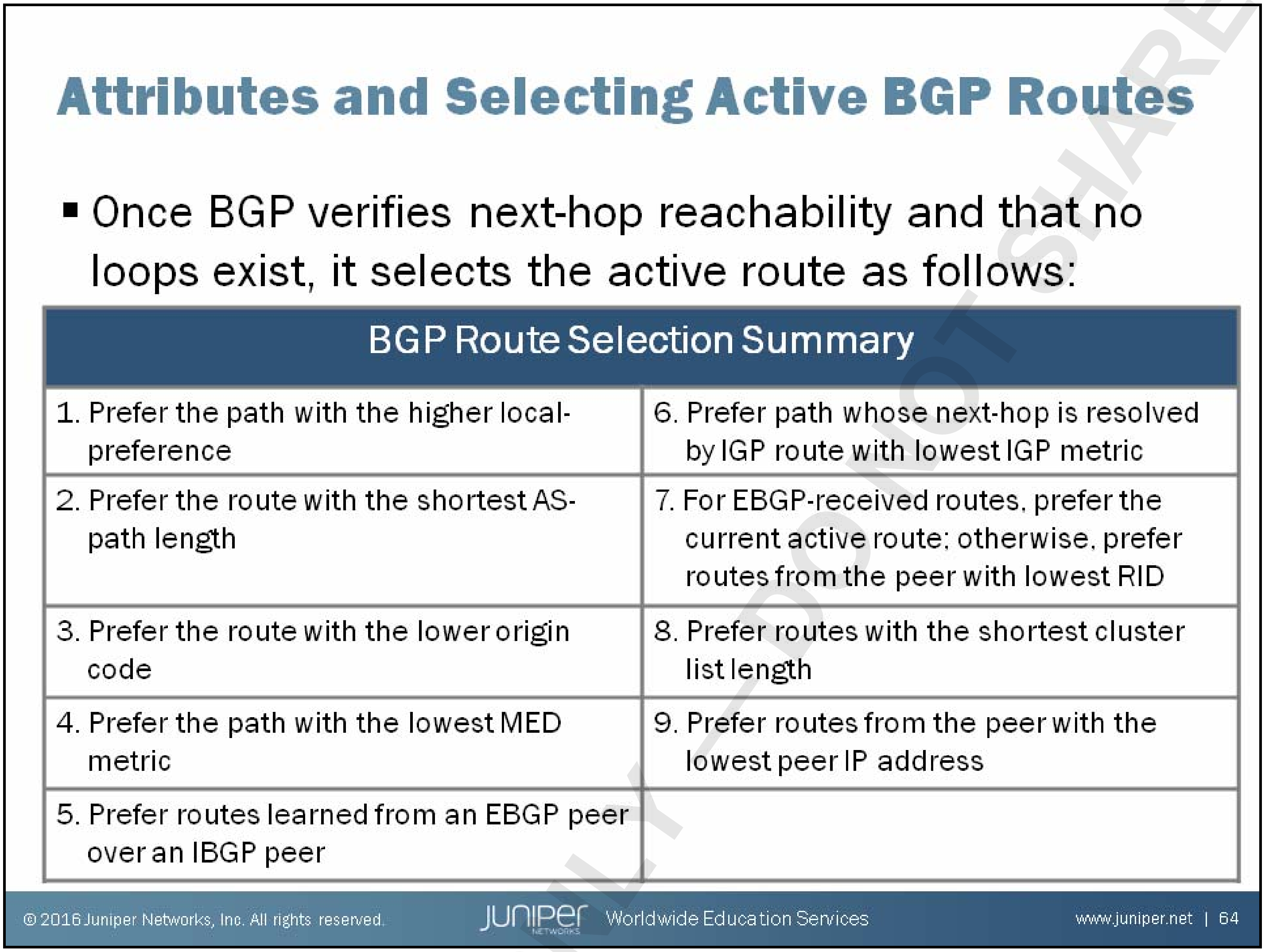
Juniper Networks Worldwide Education Services www.juniper.net | 63

The Community Attribute

A BGP community is an identifier that represents a group of destination prefixes that share a common property. Communities are used to tag specific routes that can be identified easily later for a variety of purposes. BGP includes community attribute information as a path attribute in BGP update messages. You define BGP communities and typically reference those communities in a routing policy. You can define routing policy matches based on the BGP communities. You can associate multiple communities with the same BGP route. BGP communities can be set, added, or deleted in a routing policy. The BGP community attribute is an optional attribute, so not all implementations of BGP must recognize and use communities. However, if BGP communities are associated with a BGP route, the BGP community attribute must be passed along to all other BGP peers, both within an AS and between AS networks (transitive).

The BGP community attribute is an administrative tag that you can use to associate routes together that share common properties. The BGP community attribute is not complex. The main role of the BGP community attribute is to make it easy to group routes that should be treated the same by a routing policy. BGP communities are very flexible: a BGP route can belong to many BGP communities. The attraction of BGP communities is that they can simplify routing policies. BGP communities identify routes based on the logical boundaries you establish and not what the AS number (too broad in most cases) or individual IP prefix (too granular in most cases) establishes.

You can use the community attribute within routing policies to accept or reject routes based on the values of the BGP community tags. In addition, you can use the community attribute values with other BGP attributes to implement routing policies that accept, prefer, or advertise BGP routes.



Attributes and Selecting Active BGP Routes

- Once BGP verifies next-hop reachability and that no loops exist, it selects the active route as follows:

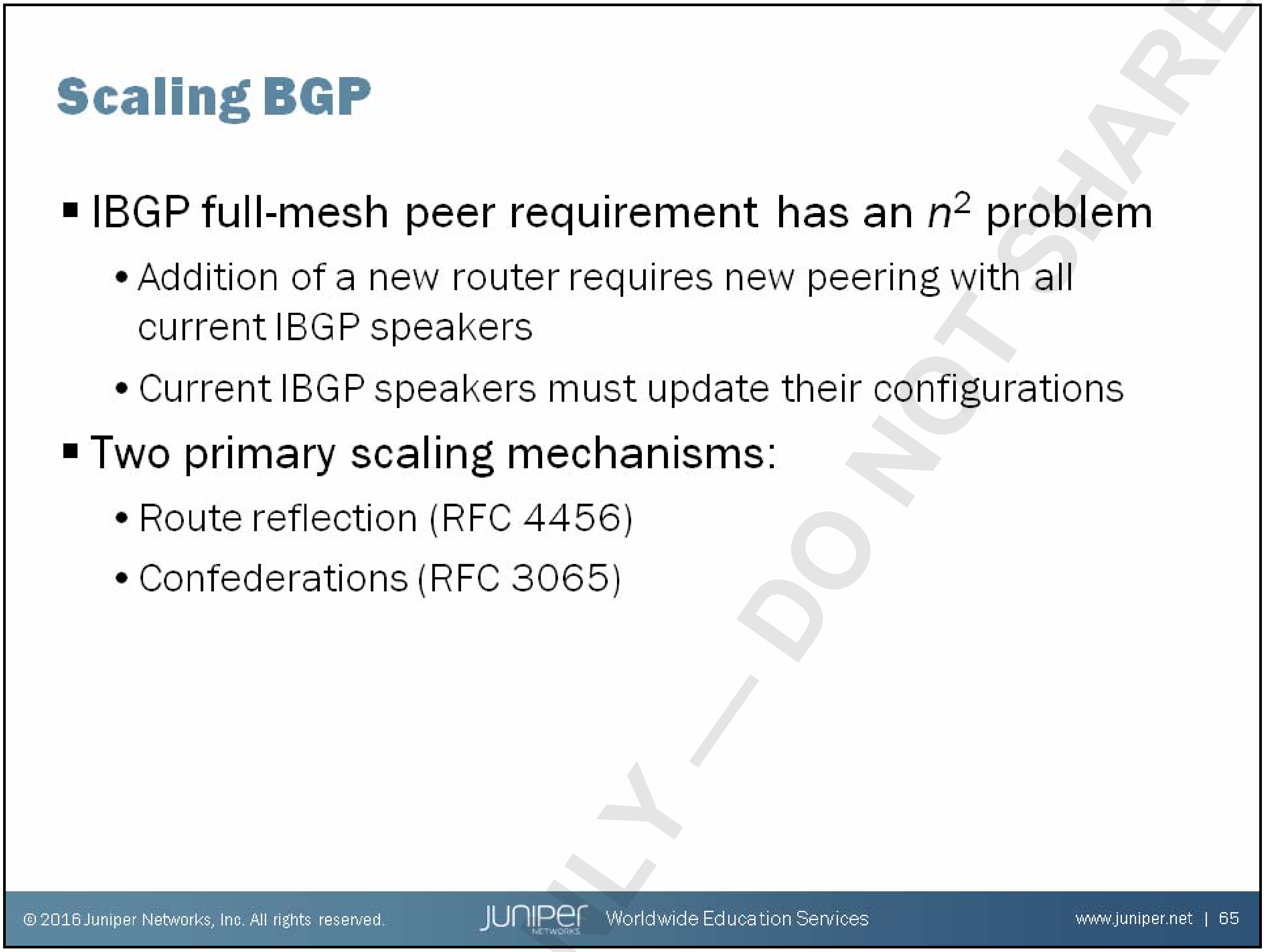
BGP Route Selection Summary	
1. Prefer the path with the higher local-preference	6. Prefer path whose next-hop is resolved by IGP route with lowest IGP metric
2. Prefer the route with the shortest AS-path length	7. For EBGP-received routes, prefer the current active route; otherwise, prefer routes from the peer with lowest RID
3. Prefer the route with the lower origin code	8. Prefer routes with the shortest cluster list length
4. Prefer the path with the lowest MED metric	9. Prefer routes from the peer with the lowest peer IP address
5. Prefer routes learned from an EBGP peer over an IBGP peer	

Worldwide Education Services

www.juniper.net | 64

Summary of BGP Active Route Selection

Before the router installs a BGP route, it must ensure that the BGP next-hop attribute is reachable and that no routing loops exist. If the BGP next hop cannot be resolved or if a loop is detected, the route is not evaluated through the BGP route selection process or installed in the route table. Before the Junos OS installs a BGP route in the routing table, the route preference is evaluated. Remember that the route preference can be changed through policy so the route preference can differ for the same prefix learned through different BGP paths. If the route preference for a BGP prefix learned through different BGP paths differs, the BGP route with the lower route preference is selected. Note that this evaluation occurs prior to the BGP selection process outlined on the slide. When a BGP route is installed in the routing table, it must go through a path selection process if multiple routes exist to the same destination prefix and the route preference is the same.



Scaling BGP

- IBGP full-mesh peer requirement has an n^2 problem
  - Addition of a new router requires new peering with all current IBGP speakers
  - Current IBGP speakers must update their configurations
- Two primary scaling mechanisms:
  - Route reflection (RFC 4456)
  - Confederations (RFC 3065)

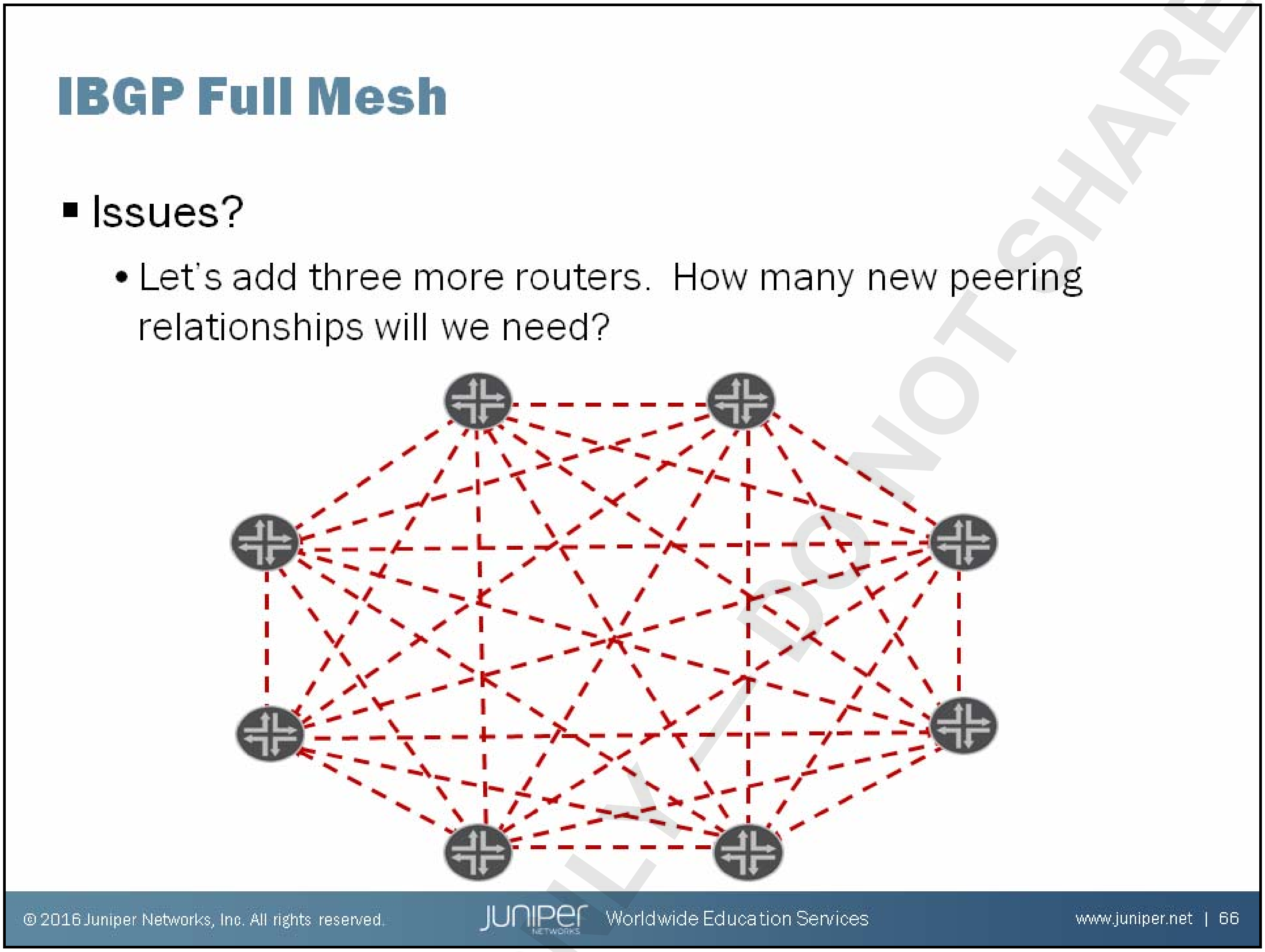
© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 65

IBGP Full Mesh

Unlike link-state routing protocols, IBGP does not flood routing updates. Instead, IBGP uses an explicit peering model that normally results in the exchange of routing information to peers that are connected in a full mesh. The need for a full-mesh IBGP topology stems from the fact that BGP uses the AS path attribute to provide loop detection, but IBGP speakers do not add the local AS number in the updates they send to other IBGP speakers. Lacking AS number based loop detection, IBGP speakers are normally precluded from re-advertising routes to other IBGP speakers when the route in question was learned from an IBGP speaker. This default IBGP behavior leads to the need for a full mesh of IBGP peering. Requiring that all IBGP peers within an AS be fully meshed has inherent scalability problems. For example, every time a new router is added to the AS, each existing IBGP router must have its configuration updated to include a peering statement for the router that has been added. This process can become quite an issue when there are 100, 200, or even 1000 routers in an AS. In fact, with only 100 routers in a full IBGP mesh, each router is required to maintain 99 IBGP peering sessions, with the network having to support a total of 4,950 IBGP sessions! Surely there has to be a better way.

Two Ways to Improve Scalability

The two primary ways to eliminate the need for a full BGP mesh are route reflection, as defined in RFC 4456, and BGP confederations, as defined in RFC 3065.



IBGP Full Mesh

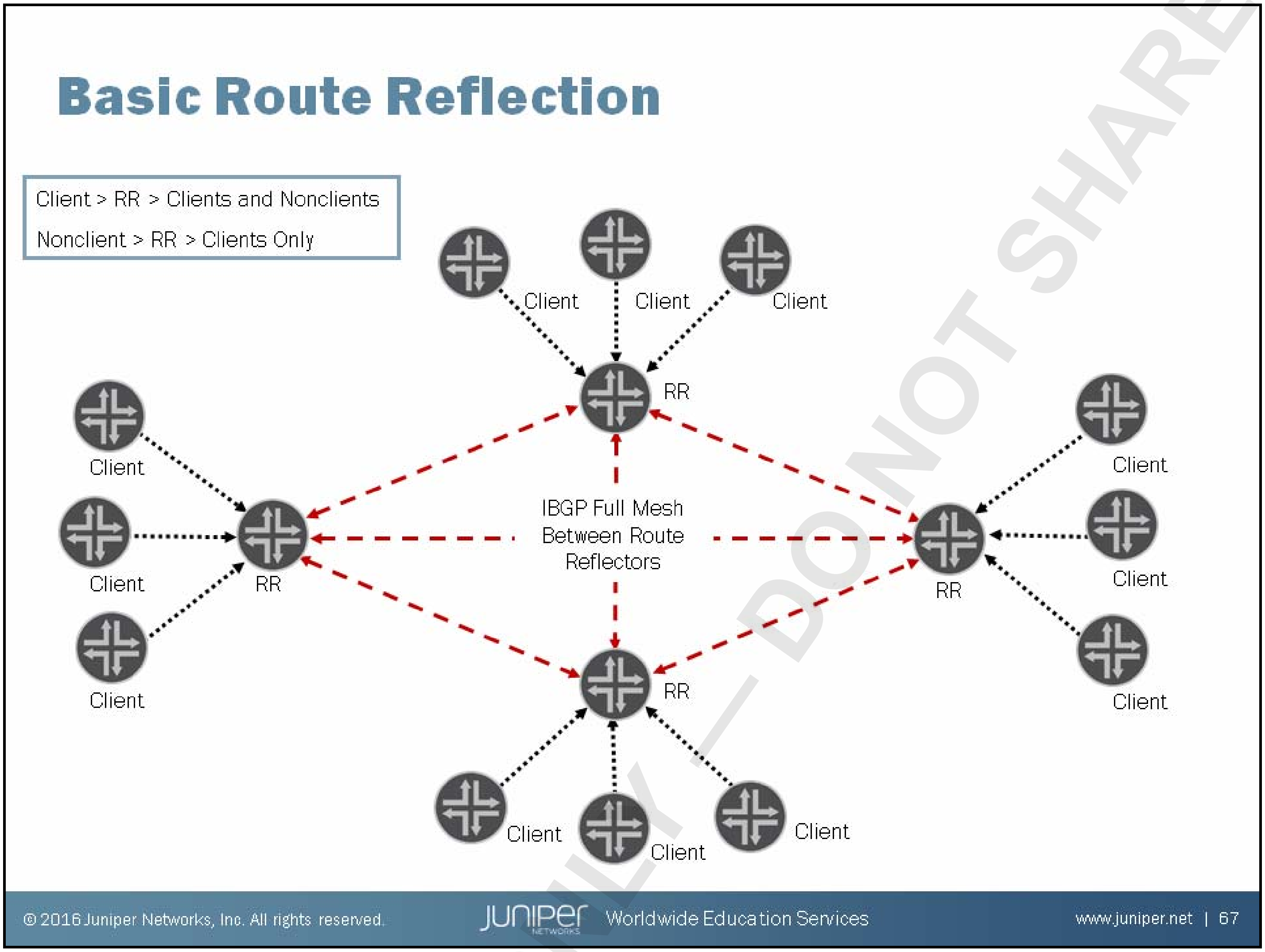
- Issues?
  - Let's add three more routers. How many new peering relationships will we need?

Juniper Networks Worldwide Education Services

www.juniper.net | 66

IBGP Full Mesh

This could turn into a scalability issue. This example demonstrates a small AS with each router maintaining seven IBGP peer relationships for a total of 56 BGP sessions throughout the AS. Add three more routers and the total number of peer relationships increases to 110. Let's look at solutions that alleviate this scalability issue.



Basic Route Reflection

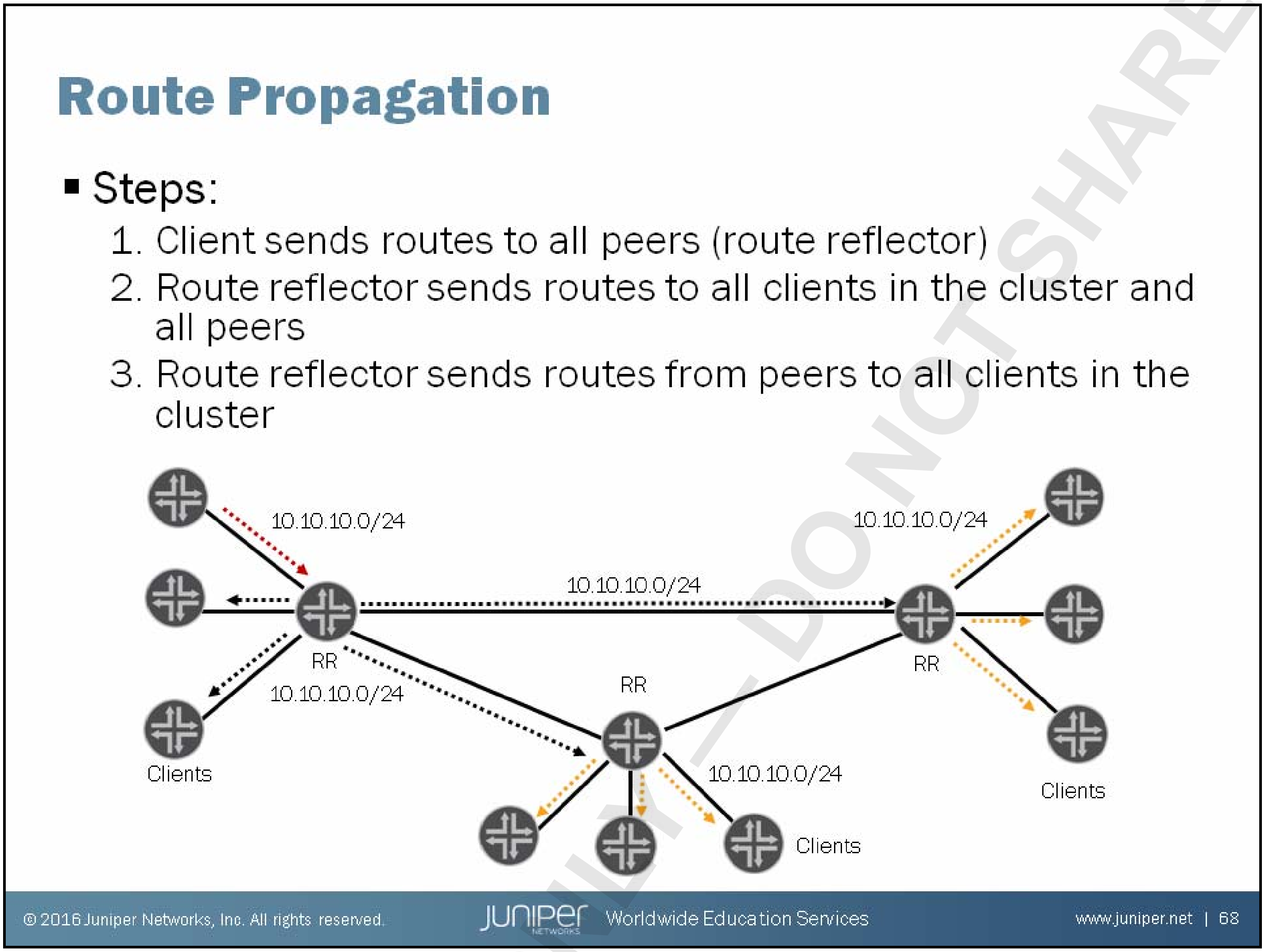
Client > RR > Clients and Nonclients

Nonclient > RR > Clients Only

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 67

Basic Route Reflection

The slide shows an AS network using a typical route reflection topology. BGP-speaking routers along the edge of the network all have a single peer configured in the form of the route reflector for the local cluster. The route reflectors are, in turn, fully meshed using standard IBGP peering procedures. The result is that all routes received by any BGP router will eventually be seen by all other BGP routers in the AS. It is a common best practice to have the logical route reflection topology follow the physical topology of the network.



Route Propagation

- Steps:
  1. Client sends routes to all peers (route reflector)
  2. Route reflector sends routes to all clients in the cluster and all peers
  3. Route reflector sends routes from peers to all clients in the cluster

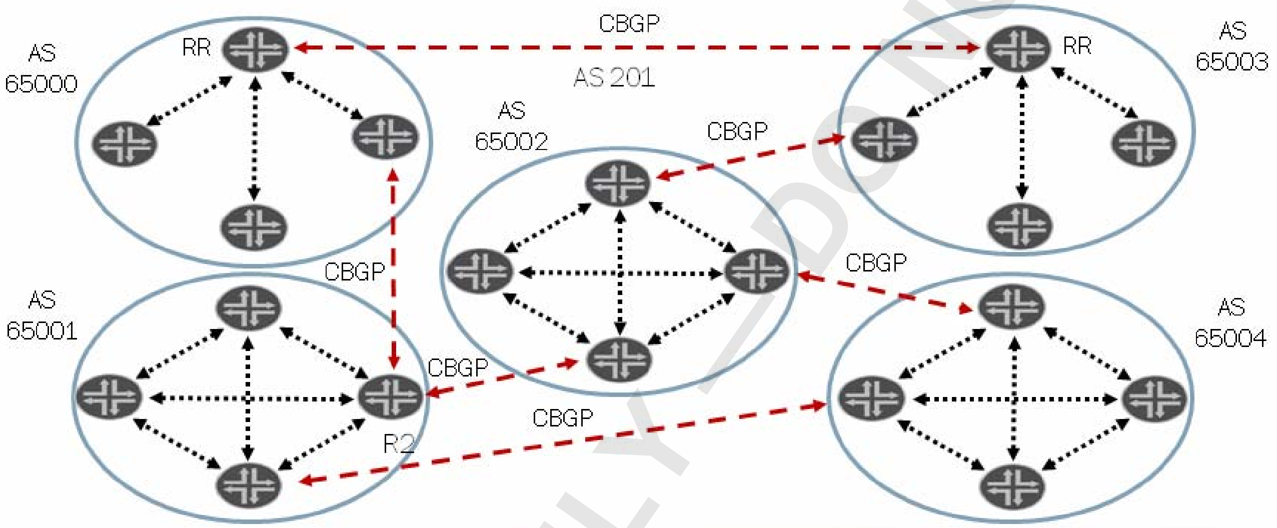
Worldwide Education Services www.juniper.net | 68

Route Propagation

The slide shows the flow of routing information in a route reflection network using a basic topology. We begin with a client in the left-most cluster advertising the 10.10.10.0/24 route to the cluster's route reflector. The slide details how the 10.10.10.0/24 route is re-advertised to all other clients in the cluster as well as to all non-client IBGP peers of the reflector. This process applies to all other routes the route reflector receives from a client in its cluster. This slide shows how the other route reflectors in the network readvertise all routes received from IBGP peers (the other reflectors in this example) to all of their cluster clients.

Confederations

- Breaks a global AS into multiple pieces (sub-AS)
  - Within each sub-AS:
    - Use private AS numbers
    - An IBGP full-mesh or route reflector topology is still required

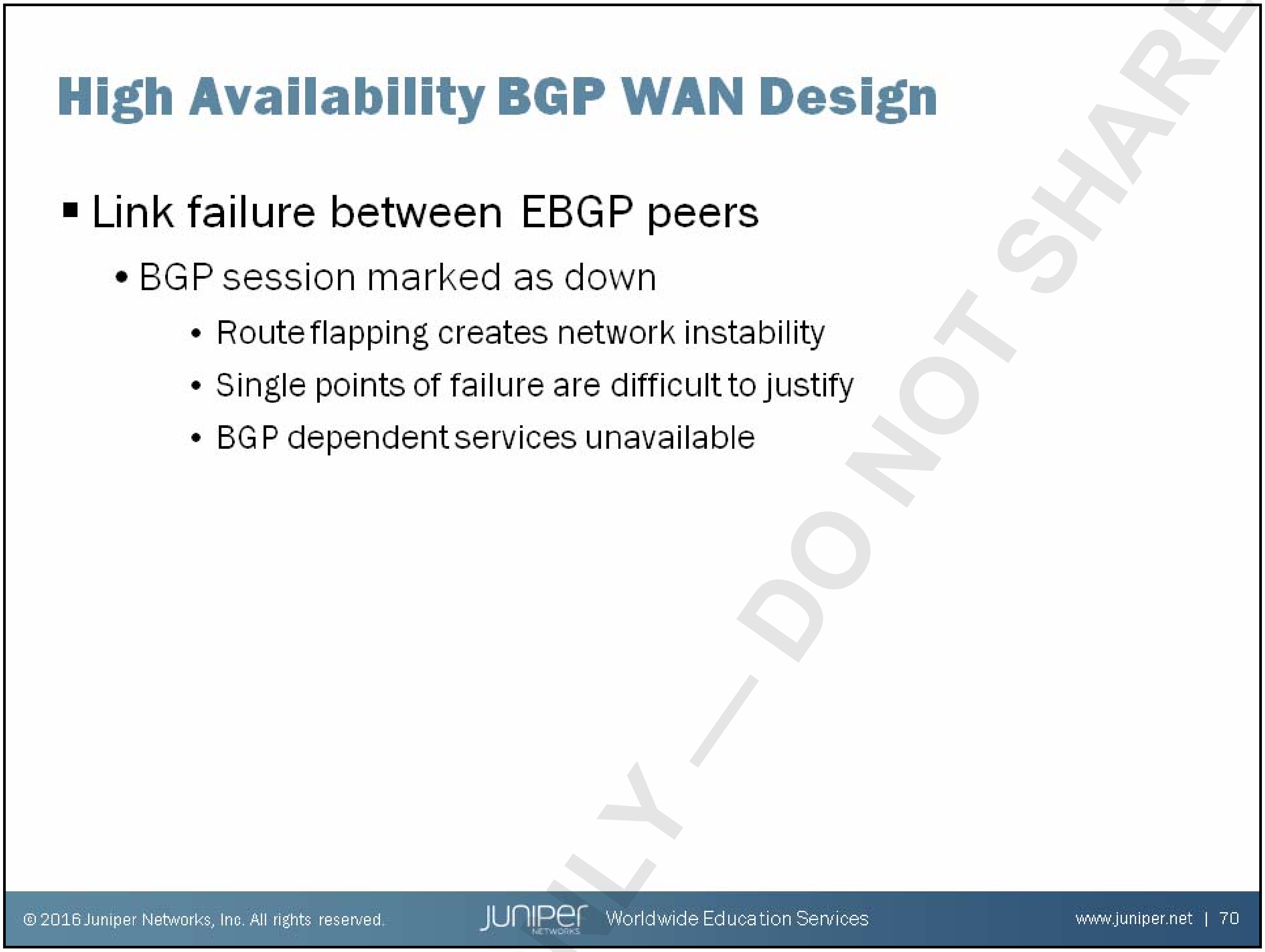


JUNIPer NETWORKS Worldwide Education Services www.juniper.net | 69

Break Up the Global AS

A confederation takes a global AS and breaks it into smaller subautonomous systems. These sub-AS networks are connected together to form the unified AS to which all other networks in the Internet peer. Within a Sub-AS The confederation sub-AS networks act just like a real AS; they require a full mesh of IBGP connectivity within themselves. Should the full mesh of the sub-AS grow too large, route reflection might be used within a sub-AS to scale the network. Each sub-AS must have a unique AS number defined, and most administrators use a private AS number from the 64512 to 65535 range. Between Each Sub-AS An EBGP-like connection that is often referred to as confederation BGP (CBGP) is used to interconnect the sub-AS networks. CBGP is a special type of EBGP; certain attributes, such as the BGP next hop, are handled differently across CBGP sessions. CBGP peers modify the AS path attribute to include the sub-AS numbers. This modification is performed to provide loop prevention within only the confederation network. All other BGP attributes, such as local preference and the BGP next hop, remain unchanged when sent across a CBGP link. The Internet views the confederation as a single autonomous system. The AS path received by other autonomous systems contains only the globally assigned AS number. The AS path contains only this number because all sub-AS numbers are removed from the AS path attribute as the route is advertised to EBGP peers. To operate a confederation network effectively, all BGP routers in the AS must know the globally unique AS number as well as all the configured sub-AS numbers.

At the edge of the confederation network, the routers remove all confederation-related AS numbers, both sequences and sets, from the AS path attribute of all routes prior to EBGP advertisement. This removal allows the details of the confederation network to be hidden to all other networks in the Internet.



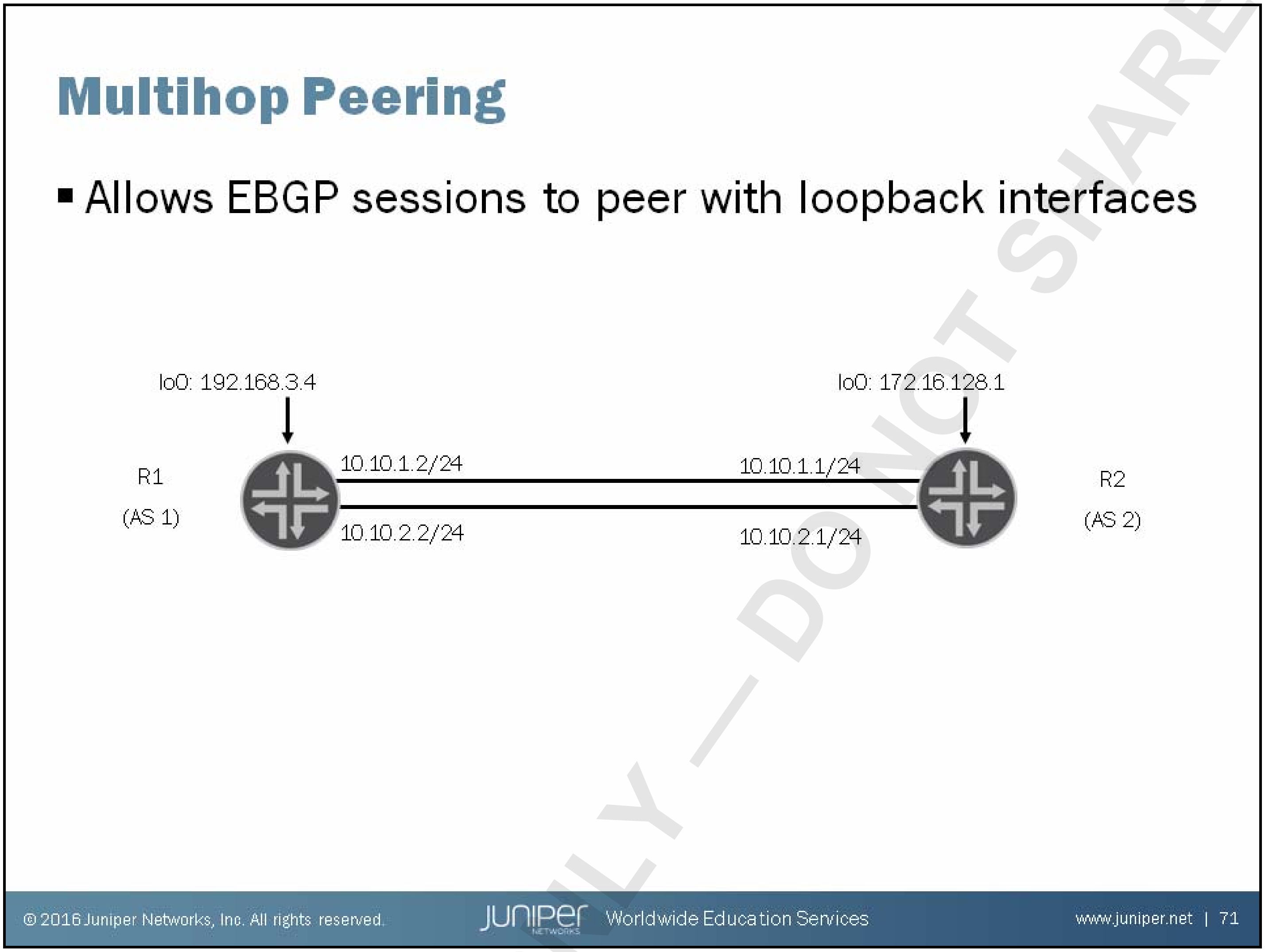
High Availability BGP WAN Design

- Link failure between EBGP peers
  - BGP session marked as down
    - Route flapping creates network instability
    - Single points of failure are difficult to justify
    - BGP dependent services unavailable

JUNIPer NETWORKS
Worldwide Education Services
www.juniper.net | 70

BGP High Availability

Link failure consequences for BGP include BGP session failures that cause routing to be re-calculated. BGP is configured in most WANs to carry additional signaling information for customer VPN traffic. The next few slides discuss some of BGP's high availability features.



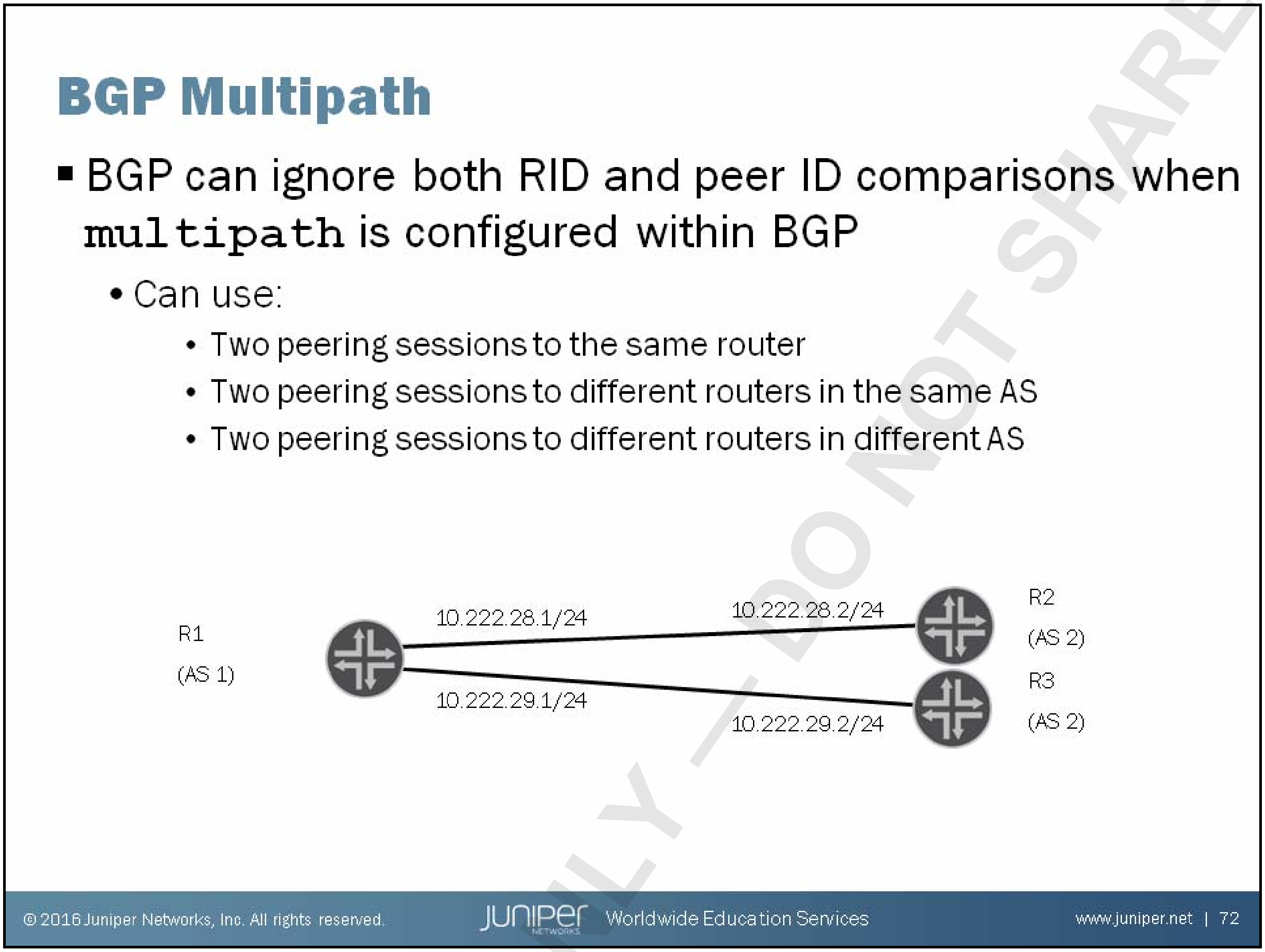
Multihop Peering

- Allows EBGP sessions to peer with loopback interfaces

Worldwide Education Services

BGP Multihop Peering

The default for an EBGP connection is to peer over a single physical hop using the physical interface address of the peer. In some cases, altering this default, one-hop, physical peering EBGP behavior is advantageous. One such case is when multiple physical links connect two routers that are to be EBGP peers. In this case, if one of the point-to-point links fails, reachability on the alternate link still exists. You must take extra configuration steps to accomplish a single BGP peering session across these multiple physical links. First, each router must establish the peering session with the loopback address of the remote router. In addition, you can also specify a time-to-live (TTL) value in the BGP packets to control how far they propagate. On the slide, we use a TTL value of 1 to ensure that the session cannot be established across any other backdoor links in the network. Third, each router must have IP routing capability to the remote router’s loopback address.

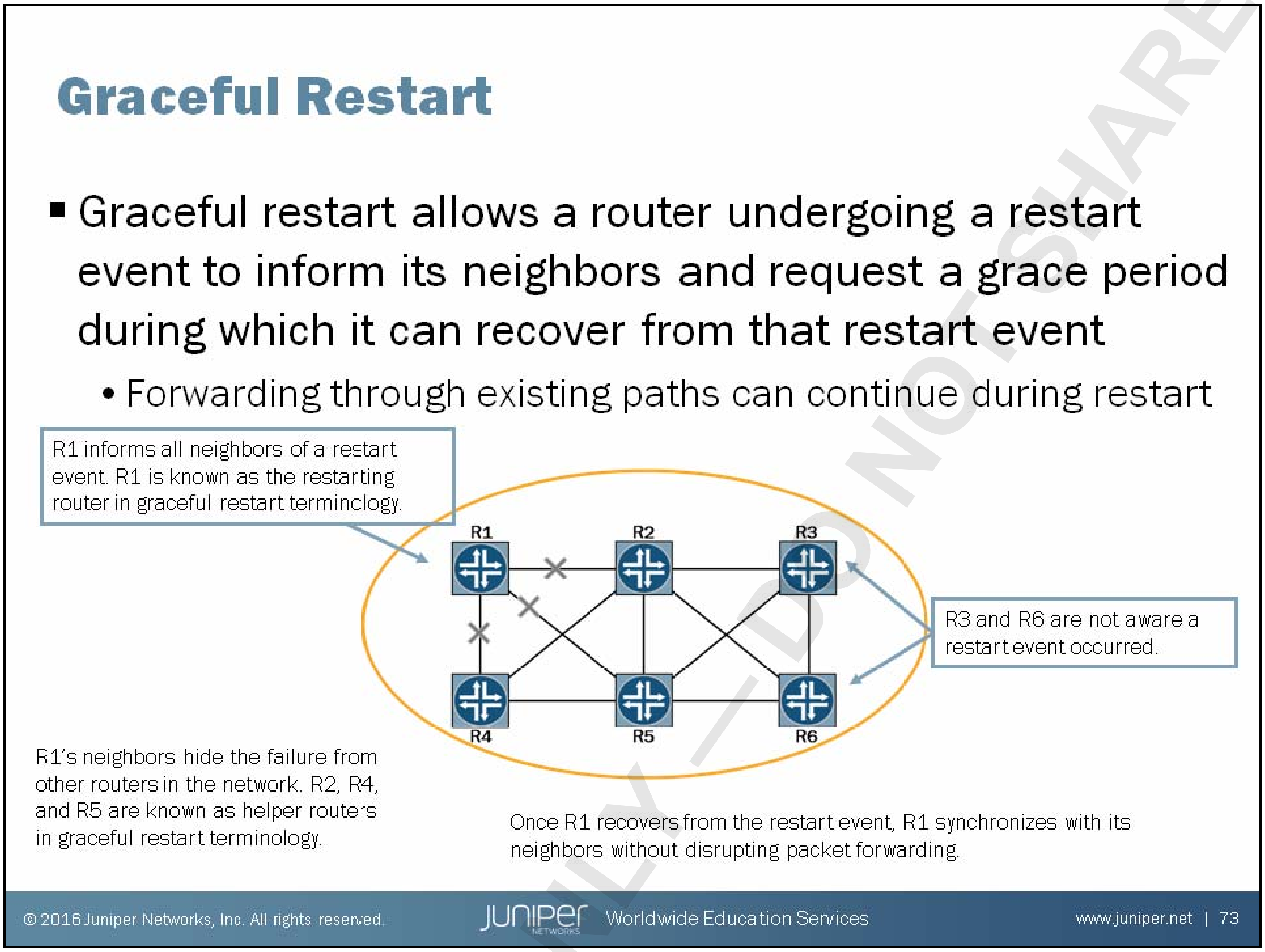


BGP Multipath

- BGP can ignore both RID and peer ID comparisons when `multipath` is configured within BGP
  - Can use:
    - Two peering sessions to the same router
    - Two peering sessions to different routers in the same AS
    - Two peering sessions to different routers in different AS

BGP Multipath

When you configure multipath on a BGP router, the active route selection algorithm ignores both the RID and the peer ID selection criteria. Should multiple copies of a route reach those portions of the route selection process, BGP installs all copies into the local routing table. Each version is listed in the table with only one of them marked as active. This active route is the version of the route that would have been selected by the algorithm had the multipath option not been configured. However, the next-hop values for the non-active routes are also listed as valid next hops for the active route. The multipath command allows multiple copies of a route from the same remote router. It also allows multiple copies of a route from two different routers in the same AS (either a local or remote AS) or two different routers in different AS. The entire concept centers around resiliency and redundancy. The effect of the multipath command on the routes from AS 2 is that the next hop for the routes from R3 (10.222.29.2) are now added to the version of the route from R2.The overall benefit of this system is the total amount of traffic sent from AS 1 to AS 2 can now be load-balanced over the two inter-AS links.



Graceful Restart

- Graceful restart allows a router undergoing a restart event to inform its neighbors and request a grace period during which it can recover from that restart event
  - Forwarding through existing paths can continue during restart

R1 informs all neighbors of a restart event. R1 is known as the restarting router in graceful restart terminology.

R3 and R6 are not aware a restart event occurred.

R1's neighbors hide the failure from other routers in the network. R2, R4, and R5 are known as helper routers in graceful restart terminology.

Once R1 recovers from the restart event, R1 synchronizes with its neighbors without disrupting packet forwarding.

Juniper NETWORKS Worldwide Education Services

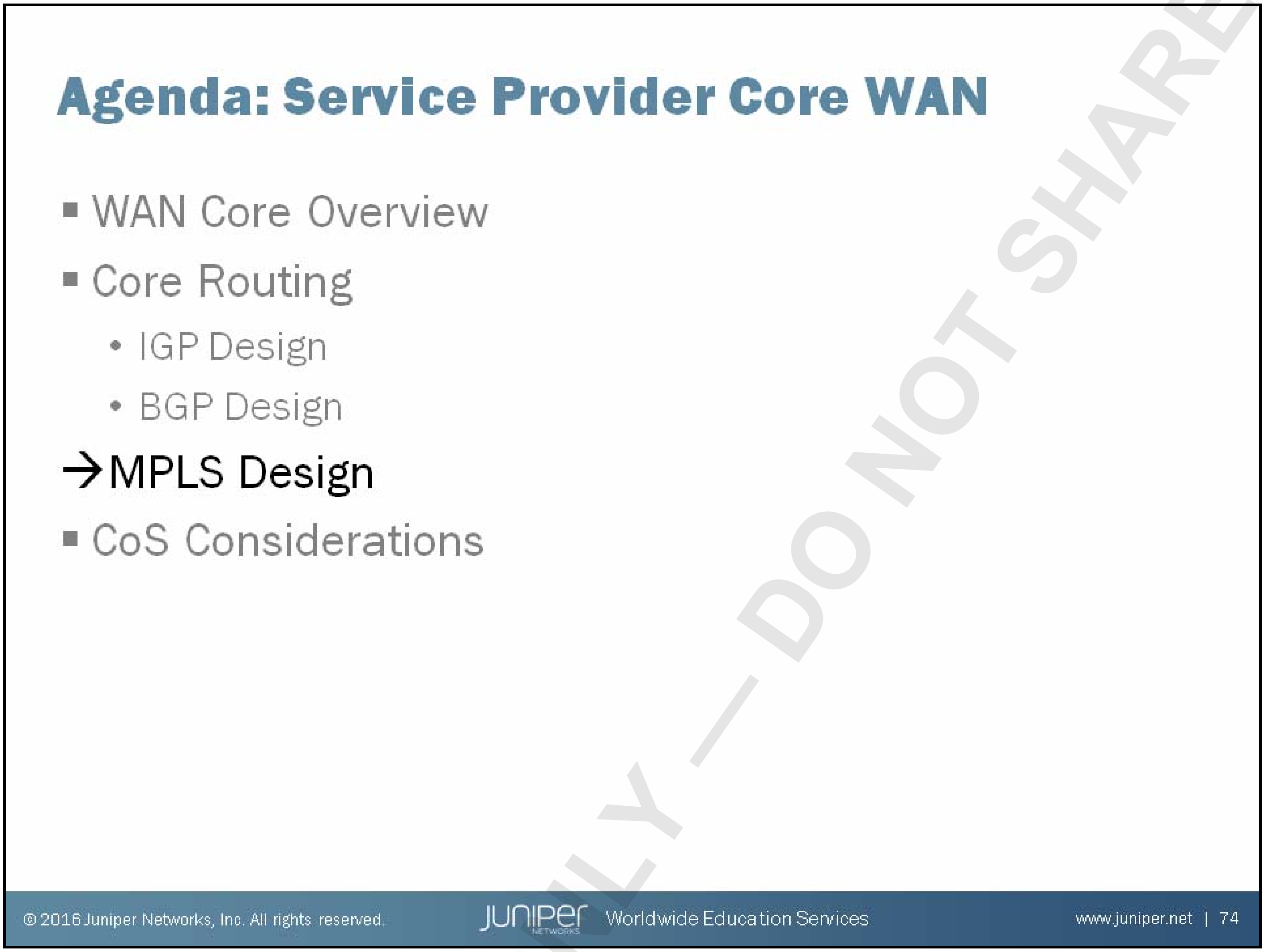
www.juniper.net | 73

Graceful Restart

What If R1's routing process restarts? When the routing protocol process (rpd) restarts, all configured protocols are affected, which means the established network topology and communication paths are also affected. In the scenario presented on the slide, all of R1's BGP neighbors must recalculate any path that traverses R1 because of the topology change. Typically, when rpd restarts, the effect on the network is temporary. In other words, once rpd restarts and the affected protocols re-establish their respective adjacencies or peering sessions, the topology and data paths return to their original state, thus causing multiple, network wide disruptions. These temporary disruptions can have a significant impact on a user's experience, especially considering today's modern networks, which include voice and video communications.

Graceful restart allows a router undergoing a restart event, including a restart of rpd, to inform its adjacent neighbors and peers of its condition. The restarting router requests a grace period from the neighbor or peer, which can then cooperate with the restarting router. When a restart event occurs and graceful restart is enabled, the restarting router can still forward traffic during the restart period, and convergence in the network is not disrupted. The neighbors or peers of the restarting router, also known as helper routers, hide the restart event from other devices not directly connected to the restarting router. In other words, the restart is not visible to the rest of the network, and the restarting router is not removed from the network topology. The graceful restart request occurs only if the following conditions are met:

- The network topology is stable;
- The neighbor or peer cooperates;
- The restarting router is not already cooperating with another restart already in progress;
- The grace period does not expire.



Agenda: Service Provider Core WAN

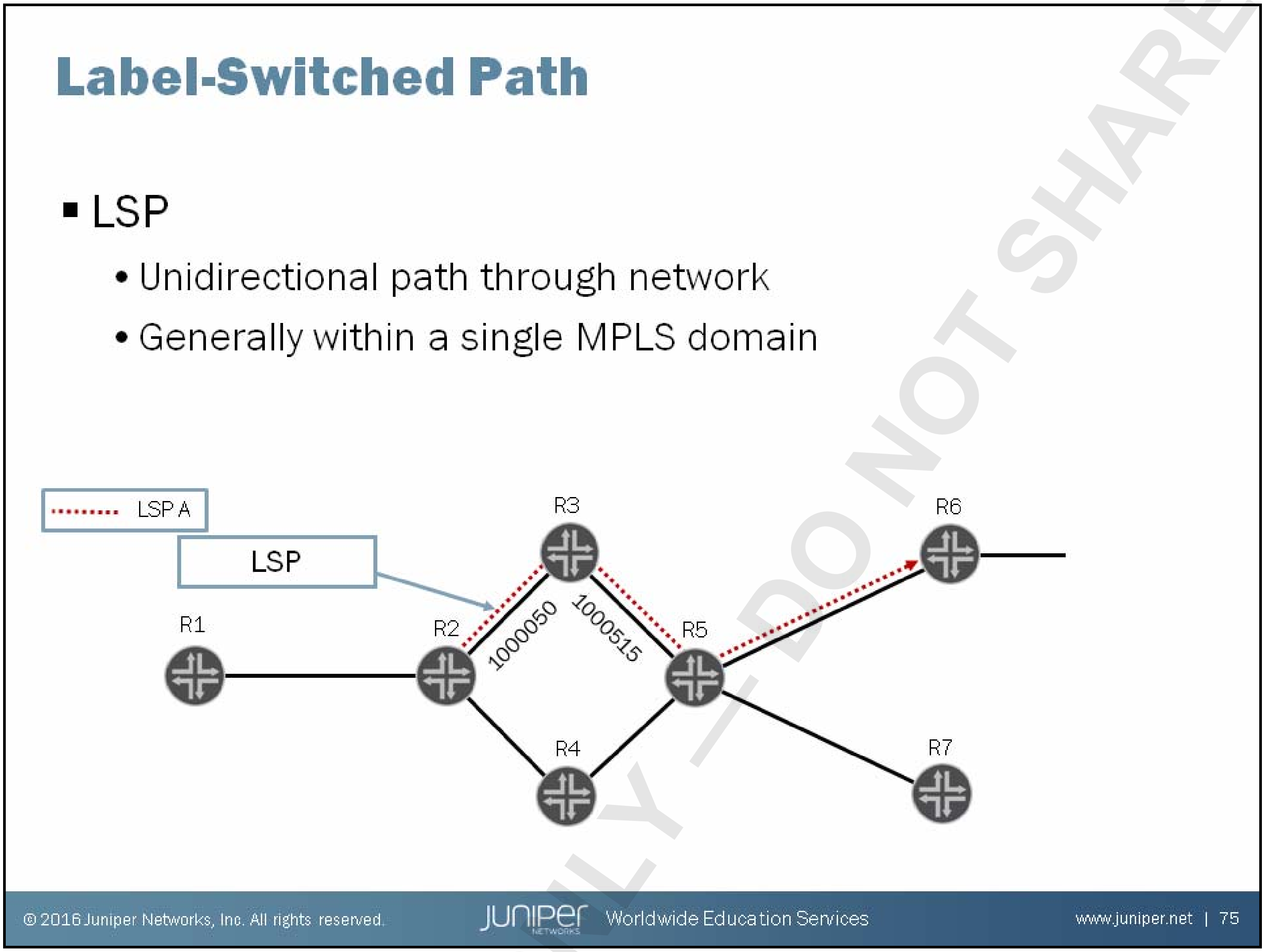
- WAN Core Overview
- Core Routing
  - IGP Design
  - BGP Design
- →MPLS Design
- CoS Considerations

Juniper Networks Worldwide Education Services

www.juniper.net | 74

MPLS Design

The slide highlights the topic we discuss next.

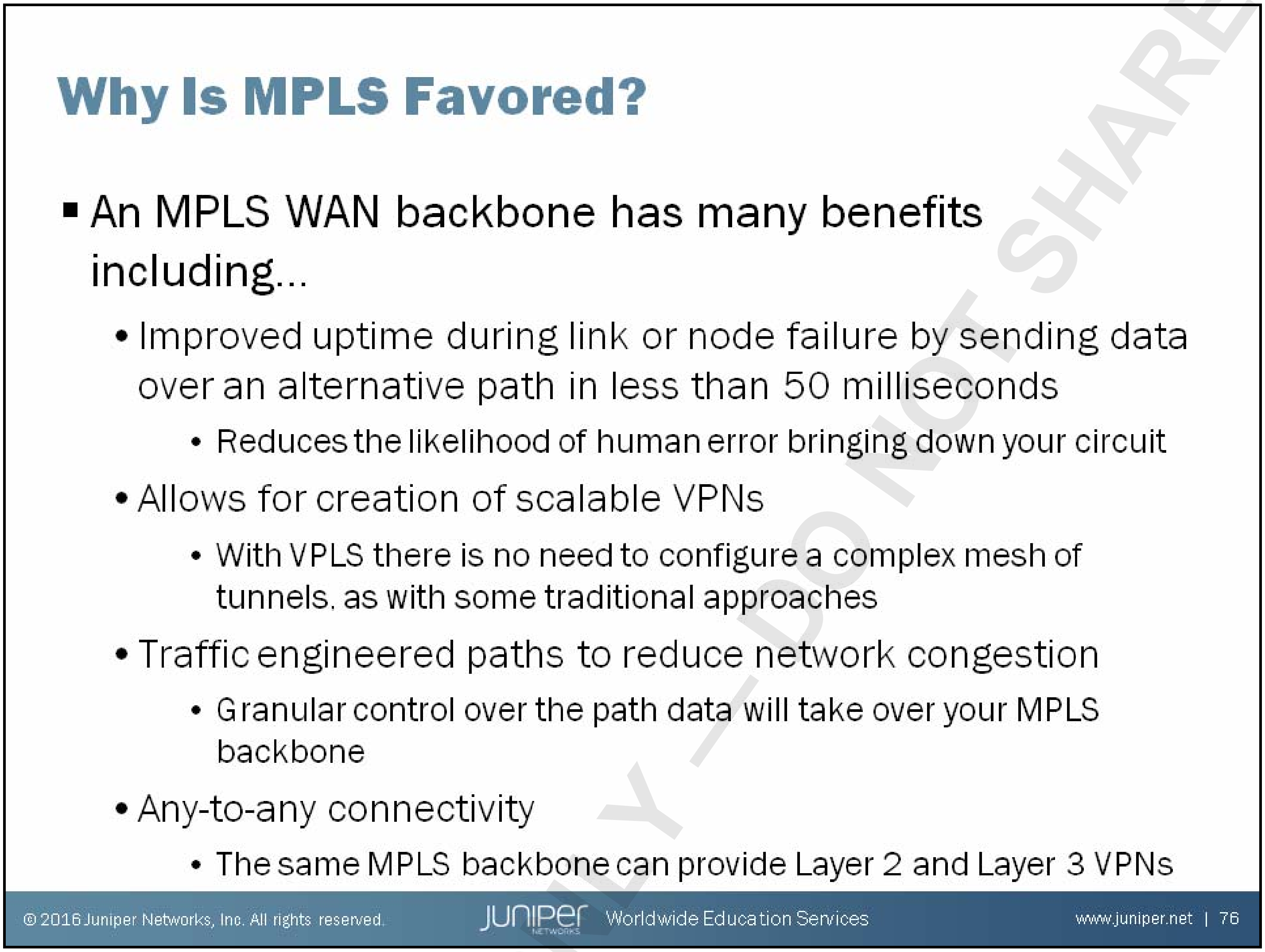


Label-Switched Path

- LSP
  - Unidirectional path through network
  - Generally within a single MPLS domain

Label-Switched Path

An LSP is a one-way (unidirectional) flow of traffic, carrying packets from beginning to end. Packets must enter the LSP at the beginning (ingress) of the path, and can only exit the LSP at the end (egress). Packets cannot be injected into an LSP at an intermediate hop. Generally, an LSP remains within a single MPLS domain. That is, the entrance and exit of the LSP, and all routers in between, are ultimately in control of the same administrative authority. This ensures that MPLS LSP traffic engineering is not done haphazardly or at cross purposes but is implemented in a coordinated fashion.



Why Is MPLS Favored?

- An MPLS WAN backbone has many benefits including...
  - Improved uptime during link or node failure by sending data over an alternative path in less than 50 milliseconds
    - Reduces the likelihood of human error bringing down your circuit
  - Allows for creation of scalable VPNs
    - With VPLS there is no need to configure a complex mesh of tunnels, as with some traditional approaches
  - Traffic engineered paths to reduce network congestion
    - Granular control over the path data will take over your MPLS backbone
  - Any-to-any connectivity
    - The same MPLS backbone can provide Layer 2 and Layer 3 VPNs

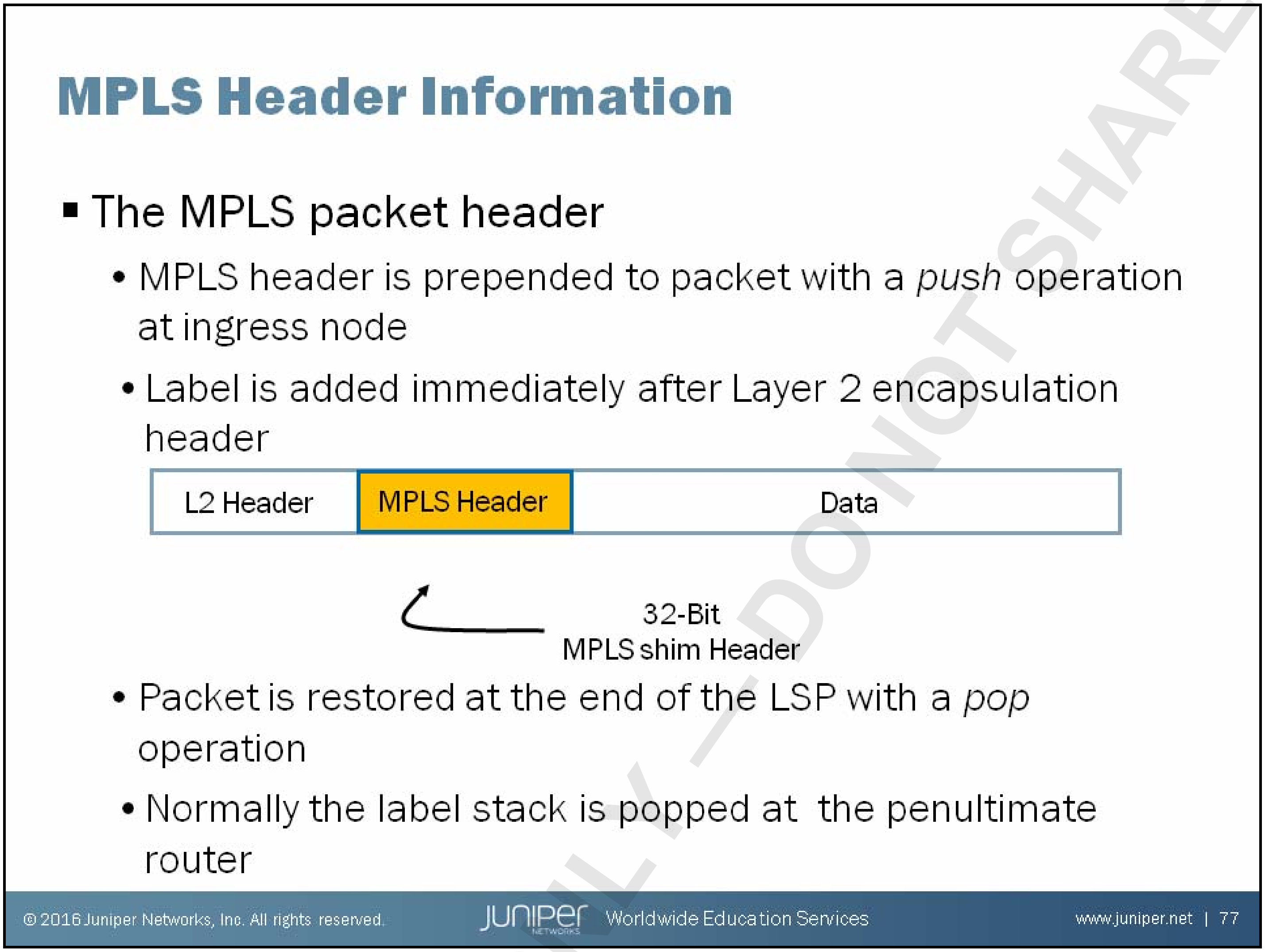
Juniper Worldwide Education Services

www.juniper.net | 76

MPLS Advantages

There are several advantages to using an MPLS network:

- Fast failover between MPLS nodes: Fast reroute and Node/Link protection are two features of an MPLS network that allow for 50ms or better recovery time in the event of a link failure or node failure along the path of an MPLS label switched path (LSP).
- Scalable VPNs: VPLS, EVPN, L3 MPLS VPNs are technologies that use MPLS to transport frames. These same technologies allow for the interconnection of many sites (potentially hundreds) without the need for the manual setup of a full mesh of tunnels between those sites. In most cases, adding a new site only requires administrator to configure the devices at the new site. The remote sites do not need to be touched.
- Traffic engineering: MPLS allows for the administrator to decide the path takes over the MPLS network. You no longer have to take the same path calculated by the IGP (i.e., all data takes the same path between sites). You can literally direct different traffic types to take different paths over the MPLS network.
- Any-to-any connectivity: An MPLS backbone will allow you the flexibility to provide any type of MPLS-based Layer 2, Layer 3, or both combinations. An MPLS backbone is a network that can generally support most types of MPLS or IP-based connectivity at the same time.



MPLS Header Information

- The MPLS packet header
  - MPLS header is prepended to packet with a *push* operation at ingress node
  - Label is added immediately after Layer 2 encapsulation header

L2 Header	MPLS Header	Data

32-Bit
MPLS shim Header

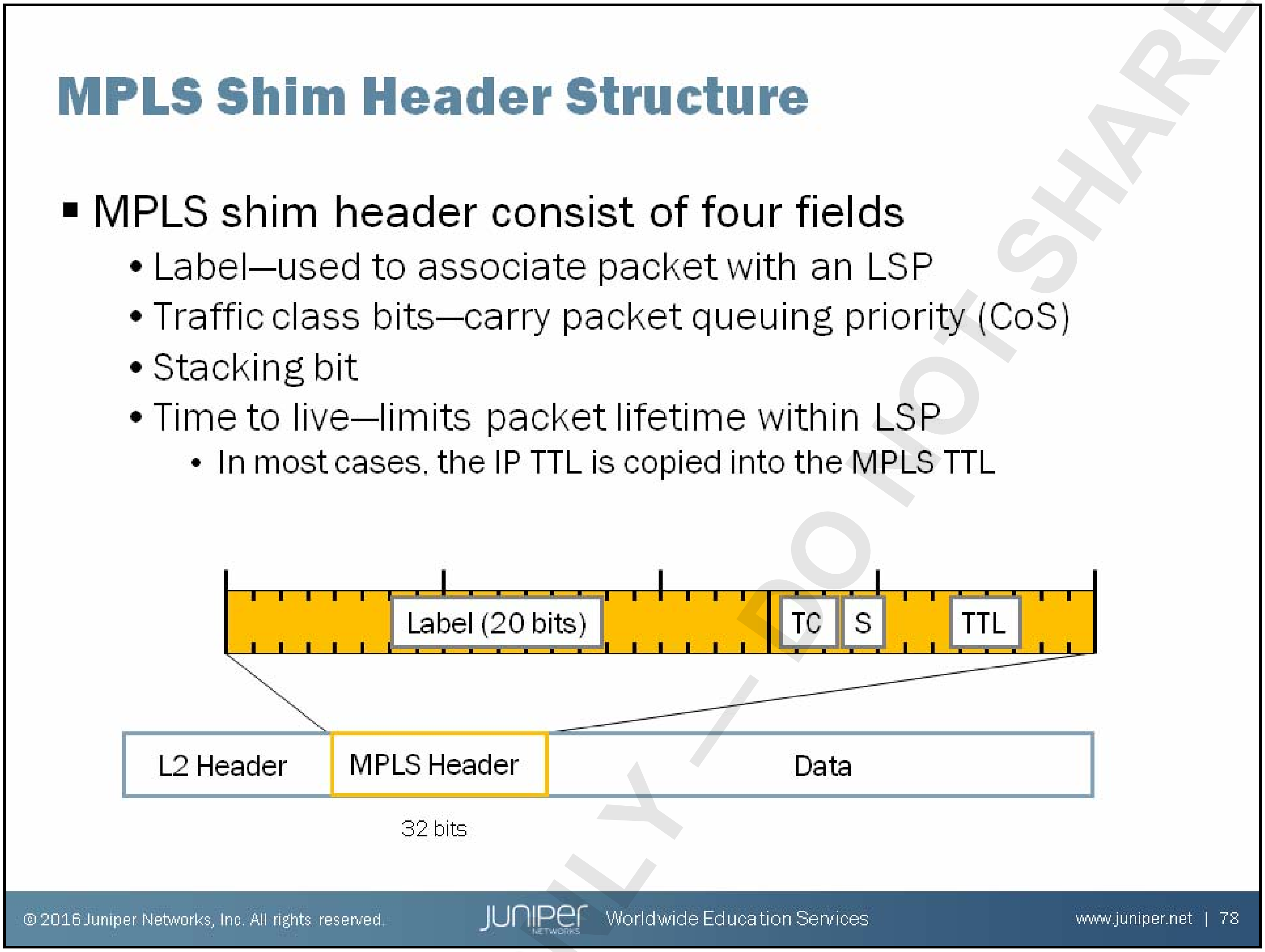
  - Packet is restored at the end of the LSP with a *pop* operation
  - Normally the label stack is popped at the penultimate router

MPLS Packet Header

MPLS is responsible for directing a flow of IP packets along a predetermined path across a network. This path is the LSP, which is similar to an ATM virtual circuit in that it is unidirectional. That is, the traffic flows in one direction from the ingress router to an egress router. Duplex traffic requires two LSPs—that is, one path to carry traffic in each direction. An LSP is created by the concatenation of one or more label-switched hops that direct packets between LSRs to transit the MPLS domain. When an IP packet enters a label-switched path, the ingress router examines the packet and assigns it a label based on its destination, placing a 32-bit (4-byte) label in front of the packet's header immediately after the Layer 2 encapsulation. The label transforms the packet from one that is forwarded based on IP addressing to one that is forwarded based on the fixed-length label.

The slide shows an example of a labeled IP packet. Note that MPLS can be used to label non-IP traffic, such as in the case of a Layer 2 VPN. MPLS labels can be assigned per interface or per router. The Junos OS currently assigns MPLS label values on a per-router basis. Thus, a label value of 10234 can only be assigned once by a given Juniper Networks router. Multicast and IPv6 labels are assigned independently of unicast packet labels. The Junos OS currently does not support labeled multicast or IPv6, except in the context of a Layer 2 or Layer 3 VPN.

At egress the IP packet is restored when the MPLS label is removed as part of a pop operation. The now unlabeled packet is routed based on a longest-match IP address lookup. In most cases, the penultimate (or second to last) router pops the label stack in penultimate hop popping. In some cases, a labeled packet is delivered to the ultimate router—the egress label-switching router (LSR)—when the stack is popped, and the packet is forwarded using conventional IP routing.



MPLS Shim Header Structure

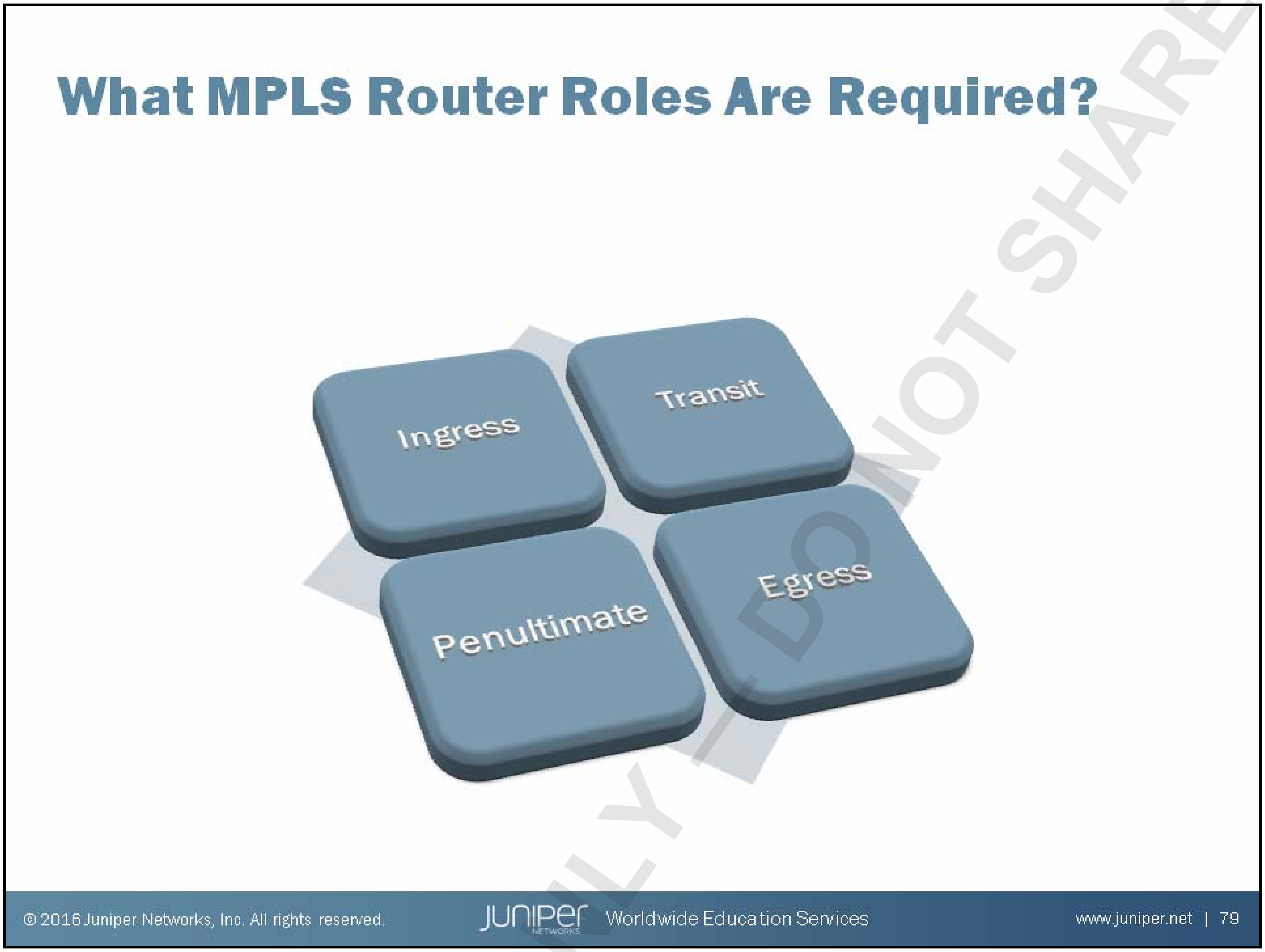
- MPLS shim header consist of four fields
  - Label—used to associate packet with an LSP
  - Traffic class bits—carry packet queuing priority (CoS)
  - Stacking bit
  - Time to live—limits packet lifetime within LSP
    - In most cases, the IP TTL is copied into the MPLS TTL

Worldwide Education Services www.juniper.net | 78

The MPLS Header (Label) Structure

The 32-bit MPLS header consists of the following four fields:

- 20-bit label: Identifies the packet to a particular LSP. This value changes as the packet flows on the LSP from LSR to LSR.
- Traffic class bits (TC): Indicates queuing priority through the network. This field was initially just the CoS field, but lack of standard definitions and use led to the current designation of this field as experimental. In other words, this field was always intended for CoS, but which type of CoS is still experimental. At each hop along the way, the CoS value determines which packets receive preferential treatment within the tunnel.
- Bottom of stack bit: Indicates whether this MPLS packet has more than one label associated with it. The MPLS implementation in the Junos OS supports unlimited label stack depths for transit LSR operations. At ingress up to three labels can be pushed onto a packet. The bottom of the stack of MPLS labels is indicated by a 1 bit in this field; a setting of 1 tells the LSR that after popping the label stack an unlabeled packet will remain.
- Time to live (TTL): Contains a limit on the number of router hops this MPLS packet can travel through the network. It is decremented at each hop, and if the TTL value drops below 1, the packet is discarded. The default behavior is to copy the value of the IP packet into this field at the ingress router.



What MPLS Router Roles Are Required?

Ingress

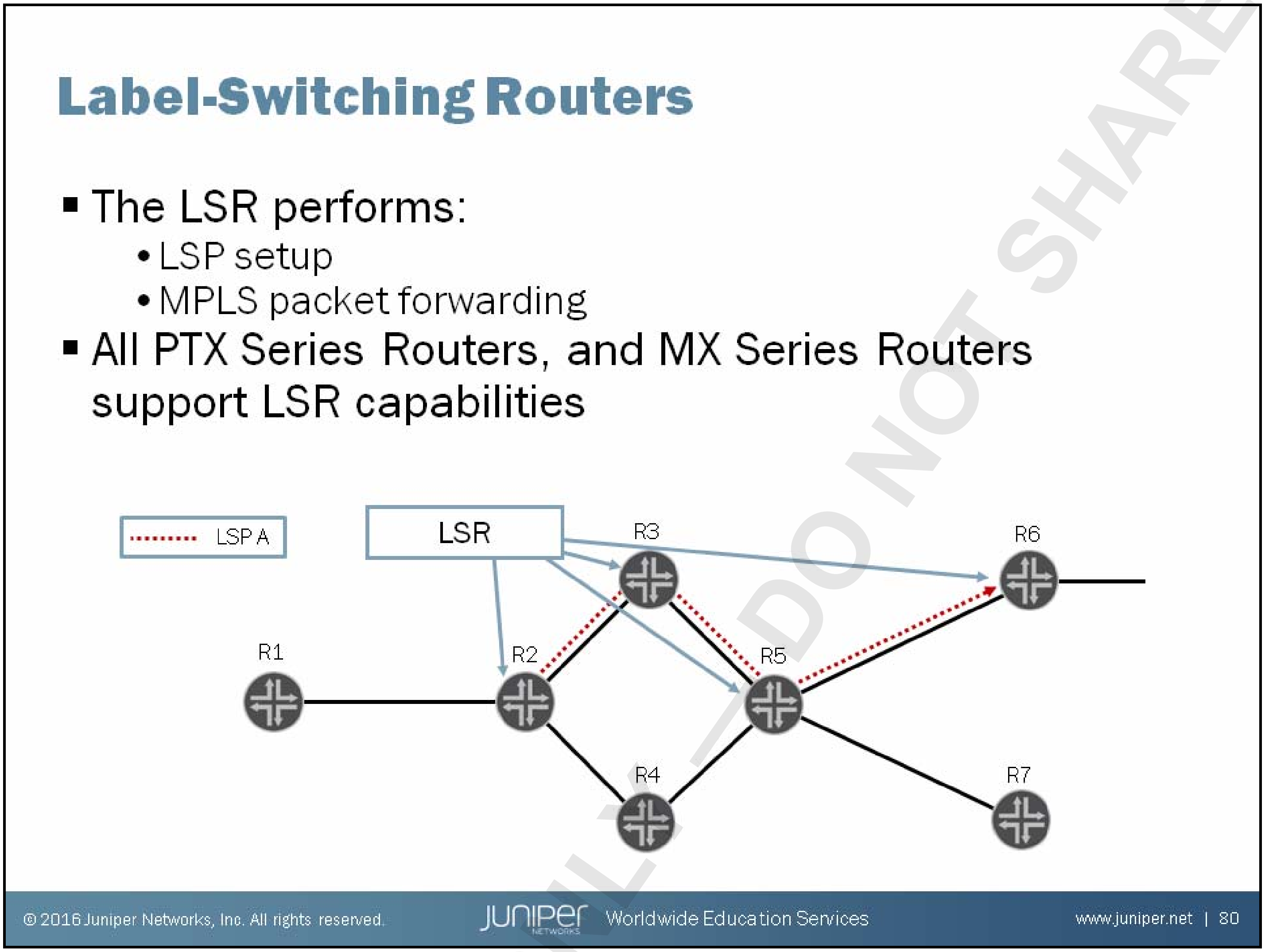
Transit

Penultimate

Egress

MPLS Router Roles

Routers play different roles in a MPLS environment. Each router has a specific unction to perform to complete the establishment of the LSP and forward traffic. The next few slides examine these different roles.



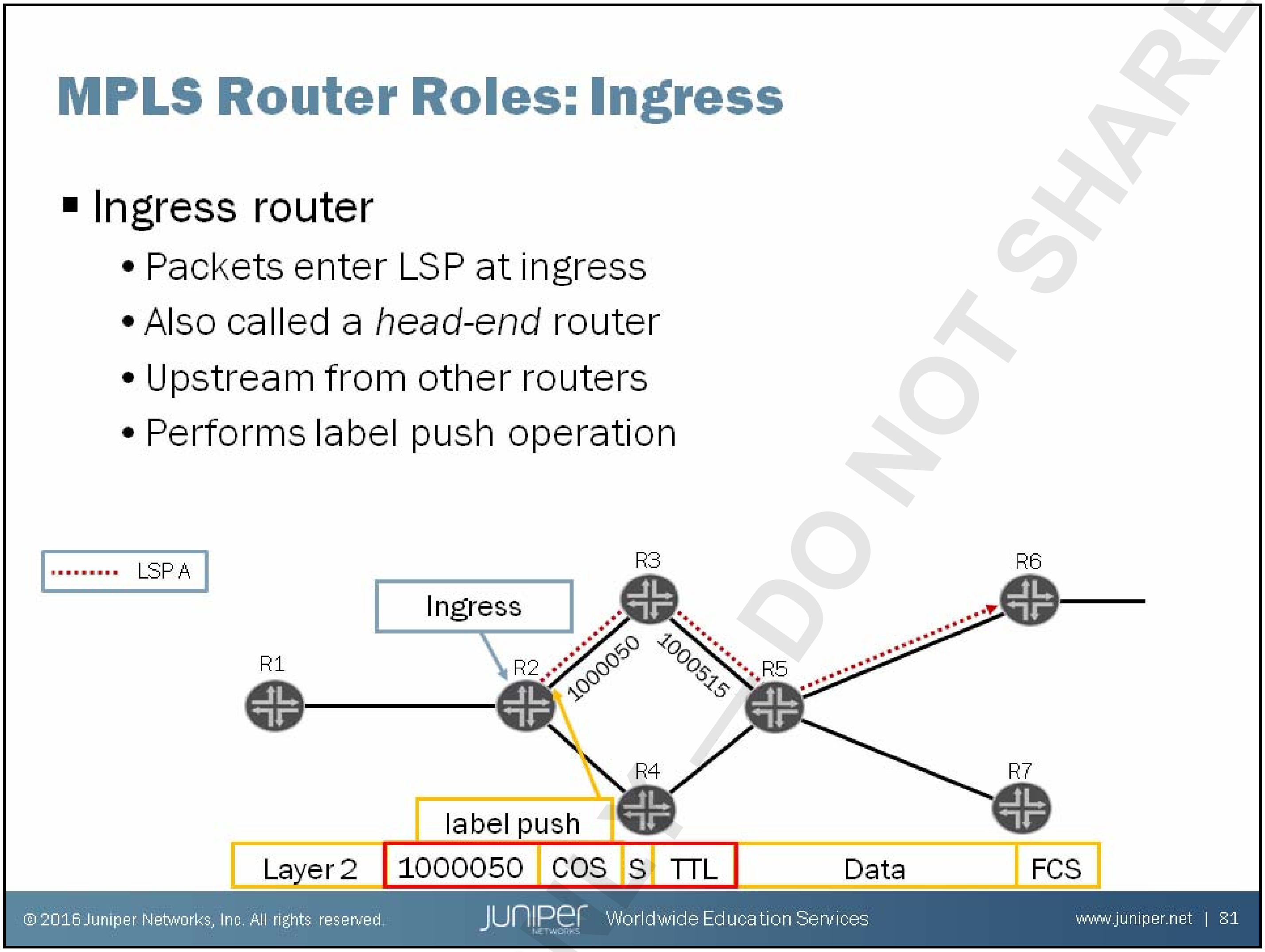
Label-Switching Routers

- The LSR performs:
  - LSP setup
  - MPLS packet forwarding
- All PTX Series Routers, and MX Series Routers support LSR capabilities

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 80

Label-Switching Routers

An LSR understands and forwards MPLS packets, which flow on, and are part of, an LSP. In addition, an LSR participates in constructing LSPs for the portion of each LSP entering and leaving the LSR. For a particular destination, an LSR can be at the start of an LSP, the end of an LSP, or in the middle of an LSP. An individual router can perform one, two, or all of these roles as required for various LSPs. However, a single router cannot be both entrance and exit points for any individual LSP. This course uses the terms LSR and router interchangeably because all Junos OS routers are capable of being an LSR.



MPLS Router Roles: Ingress

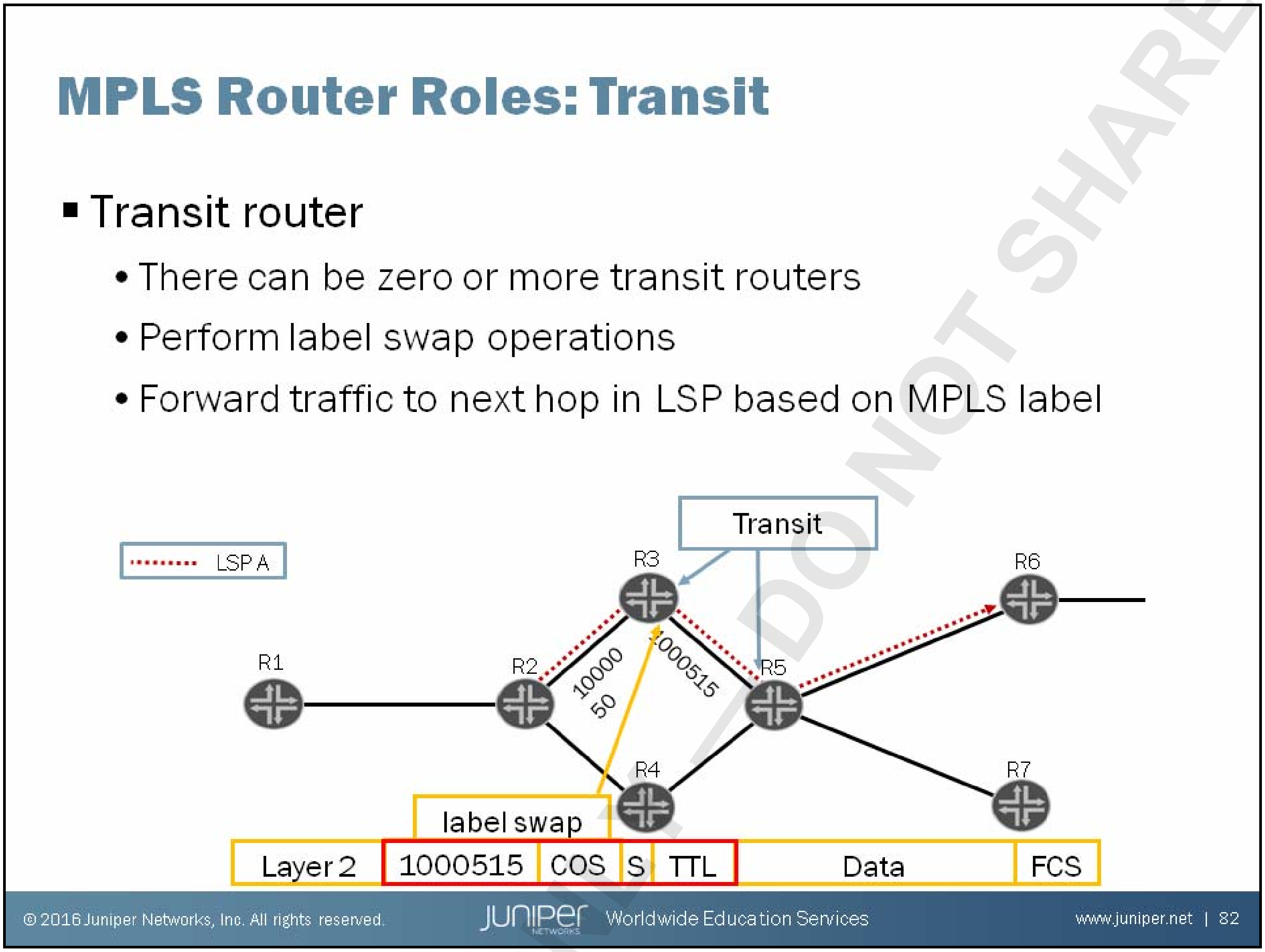
- Ingress router
  - Packets enter LSP at ingress
  - Also called a *head-end* router
  - Upstream from other routers
  - Performs label push operation

Worldwide Education Services

www.juniper.net | 81

The Functions of the Ingress Router

Each router in an MPLS path performs a specific function and has a well-defined role based on whether the packet enters, transits, or leaves the router. At the beginning of the tunnel, the ingress router encapsulates an IP packet that will use this LSP to R6 by adding the 32-bit MPLS shim header and the appropriate data link layer encapsulation before sending it to the first router in the path. Only one ingress router in a path can exist, and it is always at the beginning of the path. All packets using this LSP enter the LSP at the ingress router. In some MPLS documents, this router is called the head-end router, or the label edge router (LER) for the LSP. In this course, we call it simply the ingress router for this LSP. An ingress router always performs a push function, whereby an MPLS label is added to the label stack. By definition, the ingress router is upstream from all other routers on the LSP. In our example we see the packet structure. We can identify that the label number is 1000050 and the ingress router action is to push this shim header in between the Layer 2 Frame and the IP header.



MPLS Router Roles: Transit

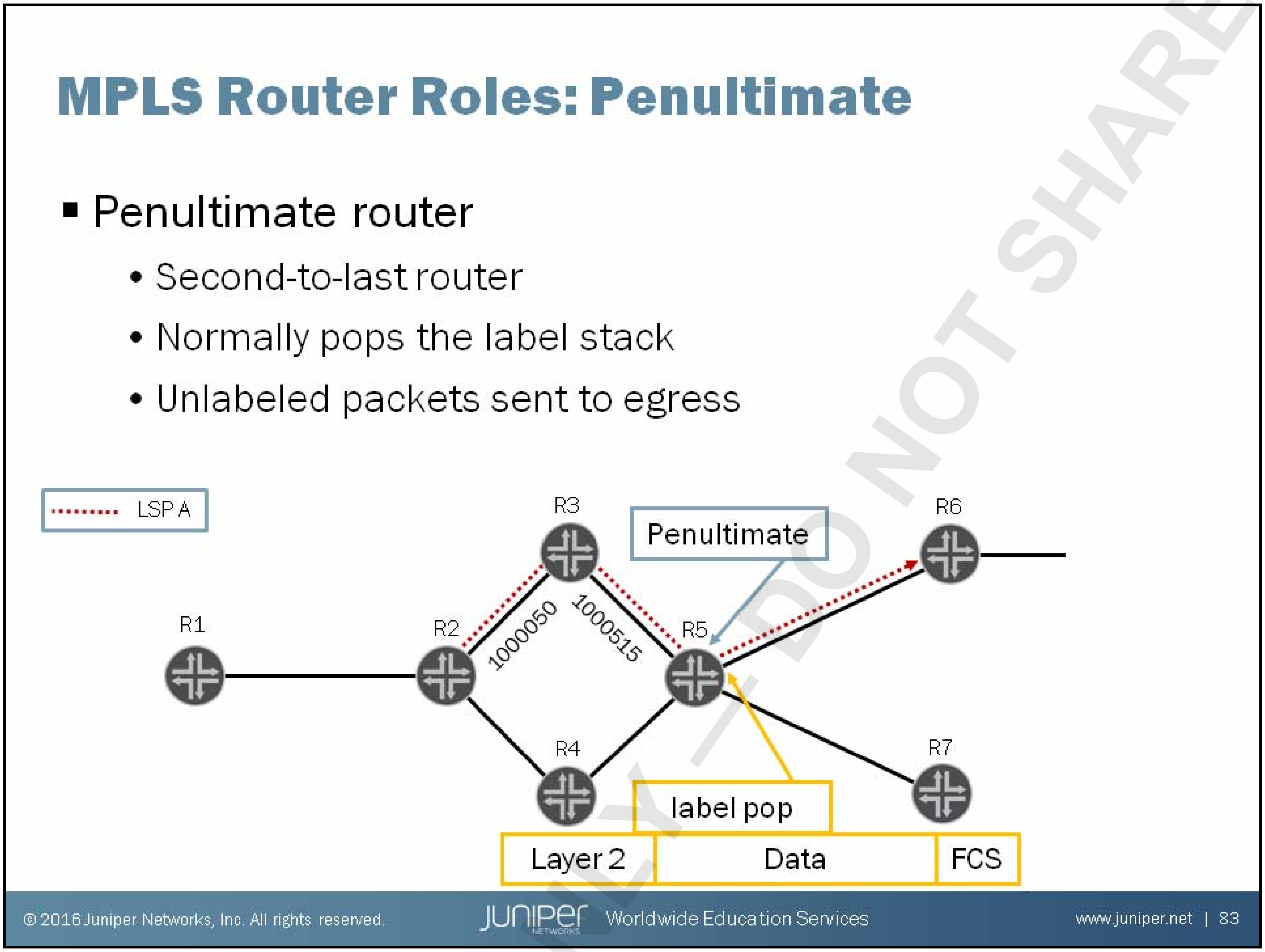
- Transit router
  - There can be zero or more transit routers
  - Perform label swap operations
  - Forward traffic to next hop in LSP based on MPLS label

JUNIPER NETWORKS Worldwide Education Services

www.juniper.net | 82

The Functions of the Transit Router

An LSP might have one or more transit routers along the path from ingress router to egress router. A transit router forwards a received MPLS packet to the next hop in the MPLS path. Zero or more transit routers in a path can exist. In a fully meshed collection of routers forming an MPLS domain, because each ingress router is connected directly to an exit point by definition, every LSP does not need a transit router to reach the exit point (although transit routers might still be configured, based on traffic engineering needs). MPLS processing at each transit point is a simple swap of one MPLS label for another. In contrast to longest-match routing lookups, the incoming label value itself can be used as an index to a direct lookup table for MPLS forwarding, but this is strictly an MPLS protocol implementation decision. The MPLS protocol enforces a maximum limit of 253 transit routers in a single path because of the 8 bit TTL field. In our example we know that the packet was sent to us with the label value of 1000050 as the previous slide indicated. Since this is a transit router we swap out the incoming label value with the outgoing label value for the next section of the LSP. We now see that the label has a value of 1000515.



MPLS Router Roles: Penultimate

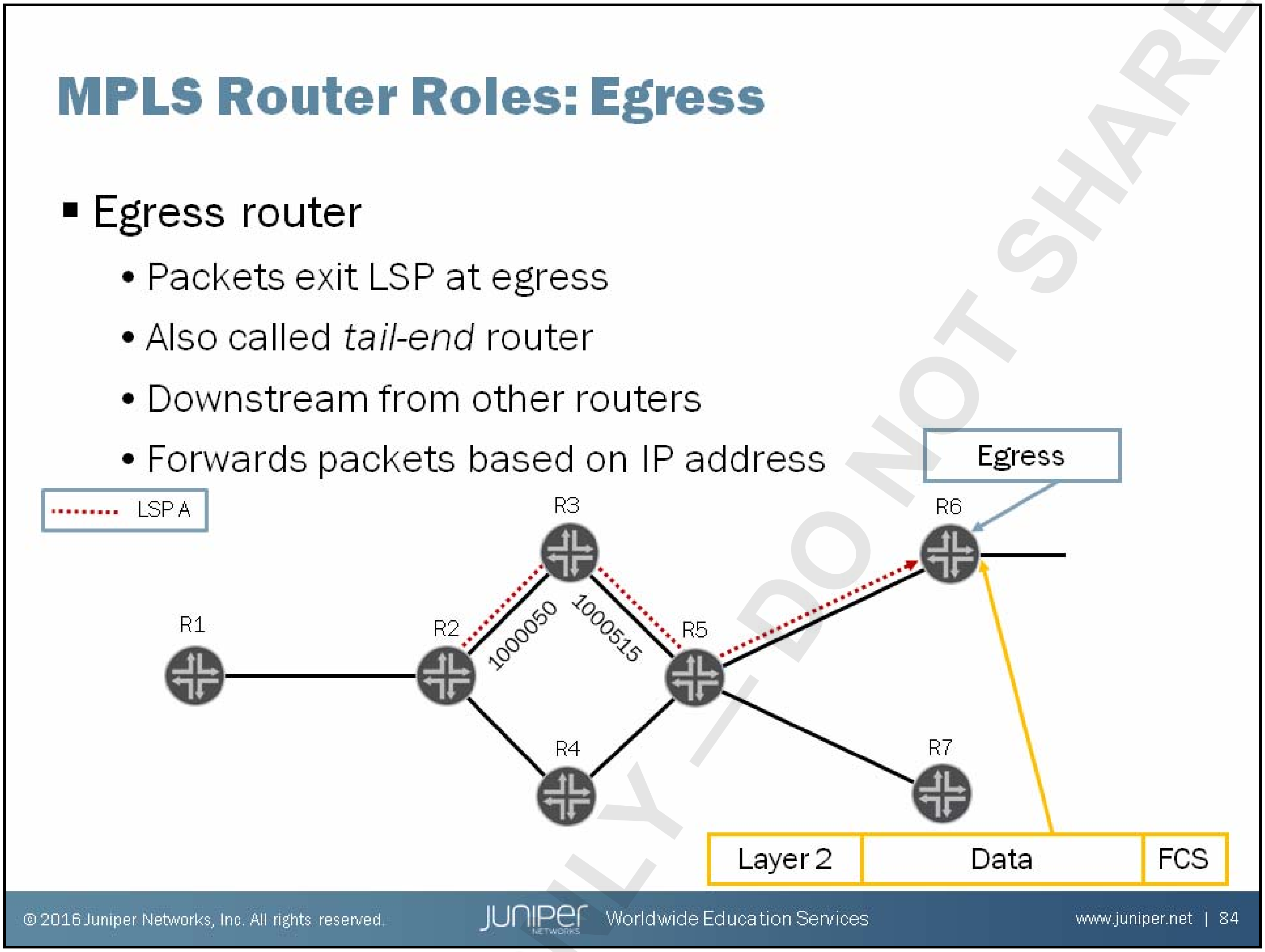
- Penultimate router
  - Second-to-last router
  - Normally pops the label stack
  - Unlabeled packets sent to egress

Worldwide Education Services

www.juniper.net | 83

The Function of the Penultimate Router

The second-to-last router in the LSP often is referred to as the penultimate hop—a term that simply means second to the last. In most cases the penultimate router performs a label pop instead of a label swap operation. This action results in the egress router receiving an unlabeled packet that then is subjected to a normal longest-match lookup. Penultimate-hop popping (PHP) facilitates label stacking and can improve performance on some platforms because it eliminates the need for two lookup operations on the egress router. Juniper Networks routers perform equally well with, or without, PHP. Label stacking makes use of multiple MPLS labels to construct tunnels within tunnels. In these cases, having the penultimate node pop the label associated with the outer tunnel ensures that downstream nodes will be unaware of the outer tunnel's existence. PHP behavior is controlled by the egress node by virtue of the label value that it assigned to the penultimate node during the establishment of the LSP. In our example you can see that the MPLS header has been popped and the router is sending the packet on to the egress router without the MPLS information.



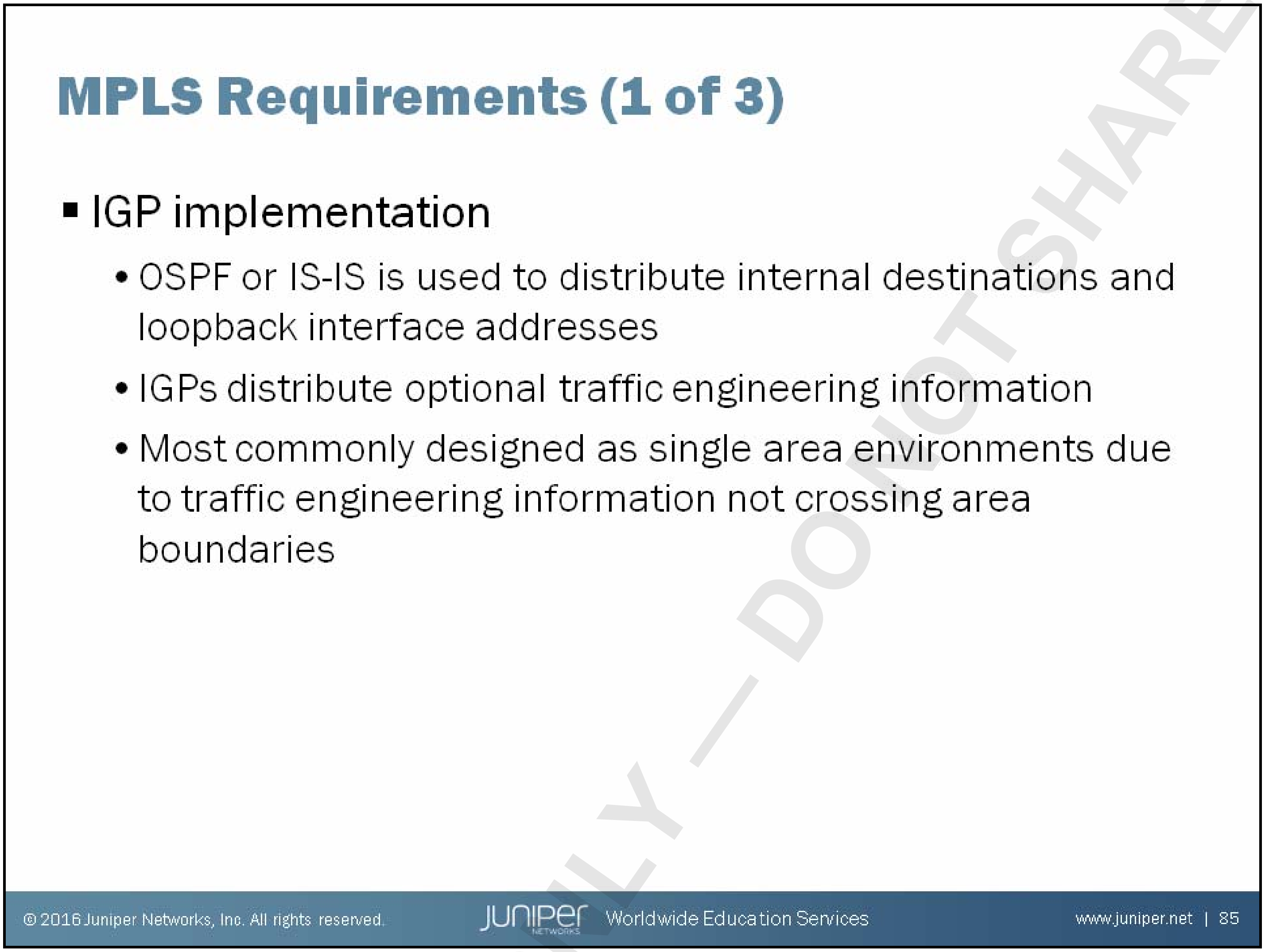
MPLS Router Roles: Egress

- Egress router
  - Packets exit LSP at egress
  - Also called *tail-end* router
  - Downstream from other routers
  - Forwards packets based on IP address

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 84

The Functions of the Egress Router

The final type of router defined in MPLS is the egress router. Packets exit the LSP at the egress router and revert to normal, IGP-based, next-hop routing outside the MPLS domain. At the end of an LSP, the egress router routes the packet based on the native information and forwards the packet toward its final destination using the normal IP forwarding table. Only one egress router can exist in a path. In many cases, the use of PHP eliminates the need for MPLS processing at the egress node. The egress router is sometimes called the tail-end router, or LER. We do not use these terms in this course. By definition, the egress router is located downstream from every other router on the LSP.



MPLS Requirements (1 of 3)

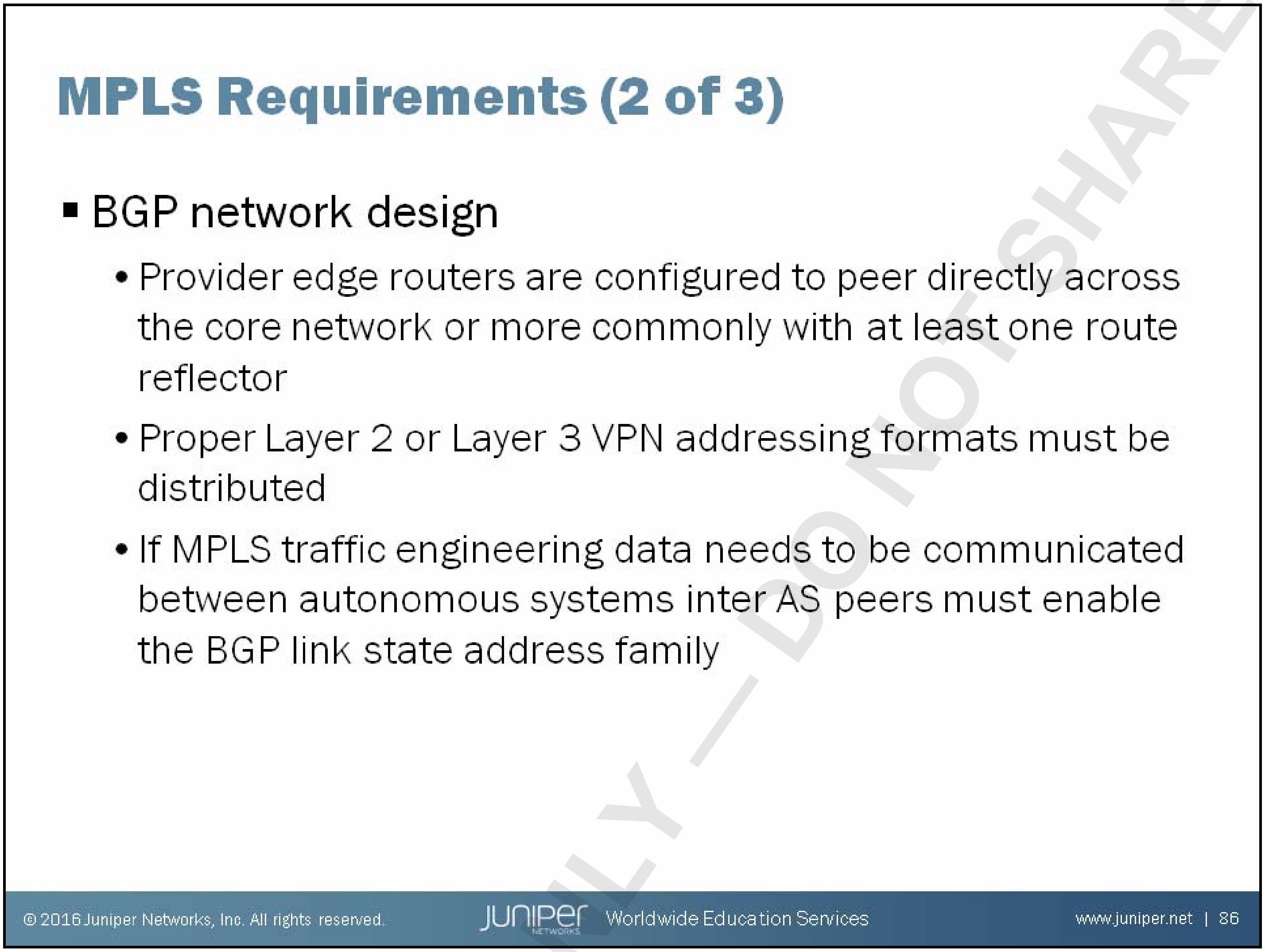
- IGP implementation
  - OSPF or IS-IS is used to distribute internal destinations and loopback interface addresses
  - IGPs distribute optional traffic engineering information
  - Most commonly designed as single area environments due to traffic engineering information not crossing area boundaries

Worldwide Education Services

www.juniper.net | 85

IGP Implementation

The foundation of an MPLS implementation is the IGP. IGPs distribute internal network addressing, router loopback IP addresses, and optionally traffic engineering information. Without traffic engineering information the LSPs are limited to being established along the IGP metric shortest path. If traffic engineering information is used to assist in the routing of LSPs, the IGP is normally configured as single area environment because traffic engineering information is not distributed across area boundaries. LSPs spanning multiple areas lose the ability to route edge to edge using the granularity traffic engineering information provides.



MPLS Requirements (2 of 3)

- BGP network design
  - Provider edge routers are configured to peer directly across the core network or more commonly with at least one route reflector
  - Proper Layer 2 or Layer 3 VPN addressing formats must be distributed
  - If MPLS traffic engineering data needs to be communicated between autonomous systems inter AS peers must enable the BGP link state address family

Juniper Worldwide Education Services

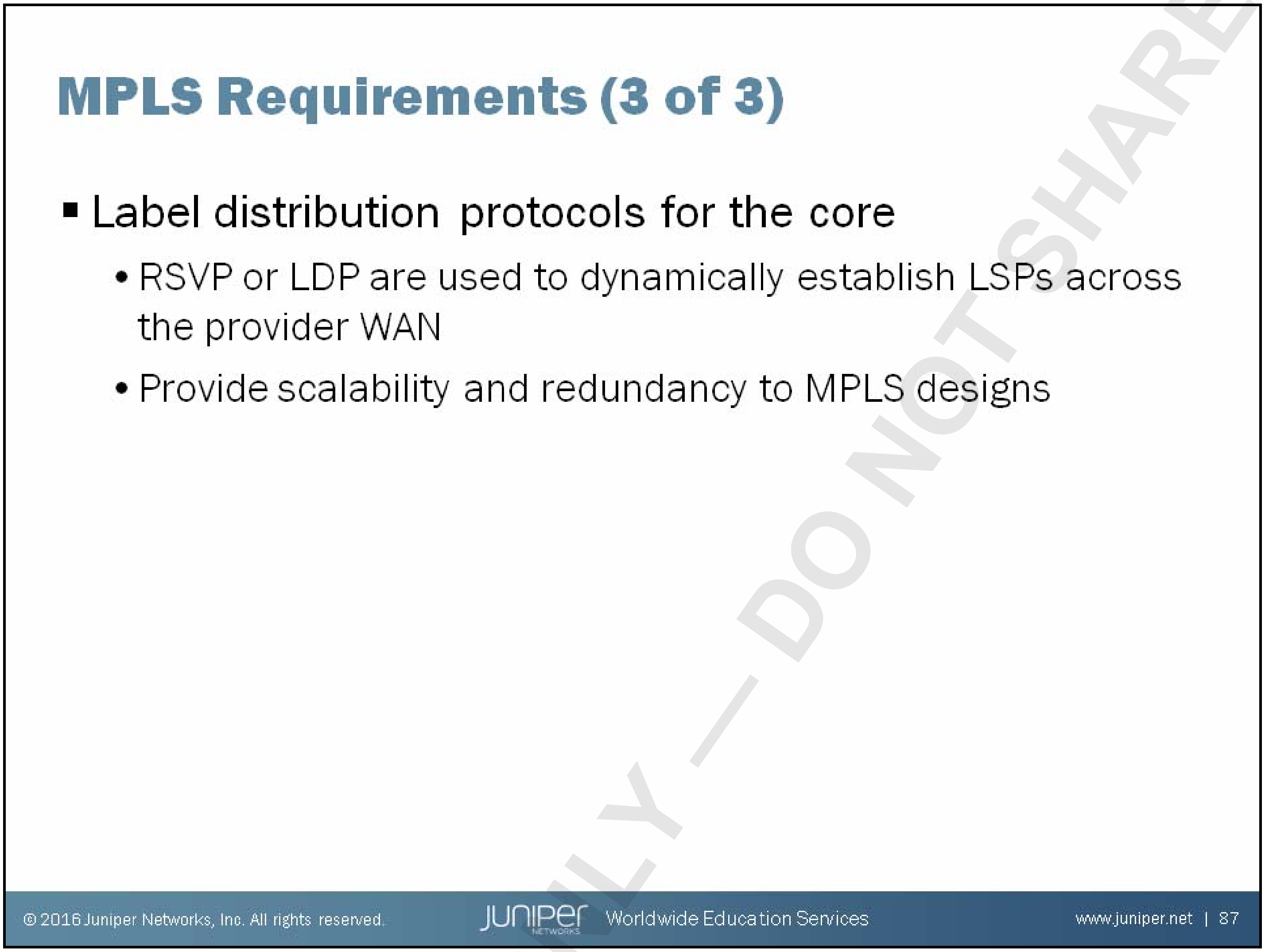
www.juniper.net | 86

BGP Network Design

In a MPLS core network edge routers are required to establish BGP neighbor relationships for the purpose of exchanging customer addressing information. Customer addressing information is stored in separate routing instances on the edge routers. Increasing the number of edge routers increases the number of required peering relationships so for scalability reasons route reflection is normally included in the core BGP design.

Different Layer 2 and Layer 3 VPN services require different BGP addressing information. BGP routers can exchange addressing information in many different formats typically called address families. Depending on the connectivity service offerings provided to WAN customers, BGP neighbors will be required to communicate properly formatted addressing information.

MPLS provides granular control of the path selection process by using link characteristics stored in a traffic engineering database during the path computation process. Traffic engineering information is distributed by IGPs but is not allowed to cross IGP area boundaries. BGP can use the link state address family to communicate the contents of a traffic engineering to peers in different IGP areas.



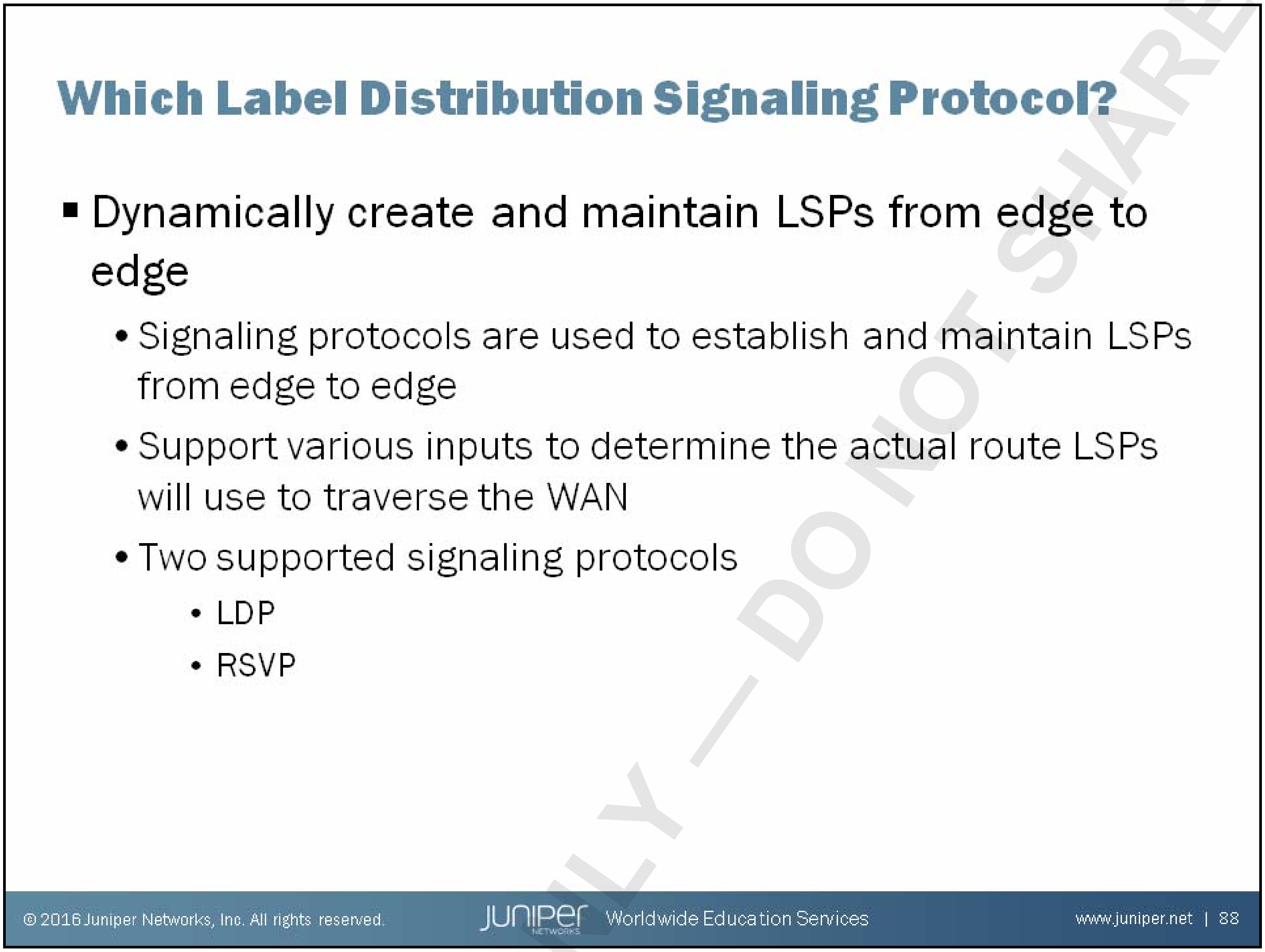
MPLS Requirements (3 of 3)

- Label distribution protocols for the core
  - RSVP or LDP are used to dynamically establish LSPs across the provider WAN
  - Provide scalability and redundancy to MPLS designs

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 87

Label Distribution Protocols

Label distribution protocols create and maintain the label-to-forwarding equivalence class (FEC) bindings along an LSP from the MPLS ingress LSR to the MPLS egress LSP. A label distribution protocol is a set of procedures by which one LSR informs a peer LSR of the meaning of the labels used to forward traffic between them. MPLS uses this information to create the forwarding tables in each LSR. Label distribution protocols are often referred to as signaling protocols. However, label distribution is a more accurate description of their function and is preferred in this course. Label distribution protocols create and maintain an LSP dynamically with little or no user intervention. Once the label distribution protocols are configured for the signaling of an LSP, the egress router of an LSP will send label (and other) information in the upstream direction towards the ingress router based on the configured options. The Junos OS supports two different label distribution protocols: RSVP and LDP. To reduce the chance of confusion this course will use the acronym LDP when referring to the particular protocol. RSVP is a generic label distribution protocol that was adapted for use in MPLS. LDP on the other hand was developed specifically to be used with MPLS.



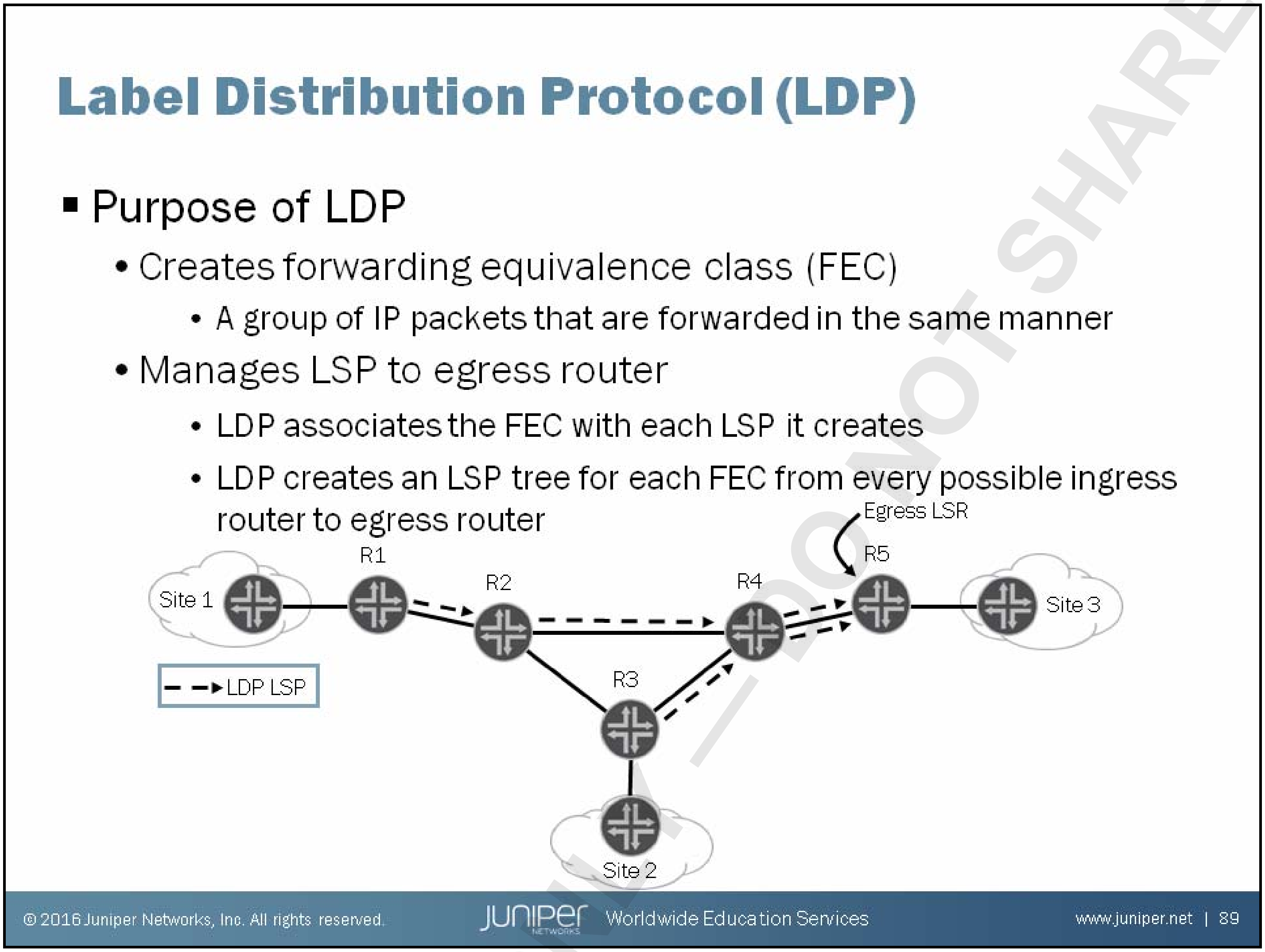
Which Label Distribution Signaling Protocol?

- Dynamically create and maintain LSPs from edge to edge
  - Signaling protocols are used to establish and maintain LSPs from edge to edge
  - Support various inputs to determine the actual route LSPs will use to traverse the WAN
  - Two supported signaling protocols
    - LDP
    - RSVP

Juniper NETWORKS
Worldwide Education Services
www.juniper.net | 88

Label Distribution Signaling Protocols

MPLS LSPs are dynamically signaled and maintained between ingress and egress edge routers. Two protocols are available for use for the signaling process and each protocol has different features and capabilities. In the next few slides we examine the LDP and RSVP signaling protocols.



Label Distribution Protocol (LDP)

- Purpose of LDP
  - Creates forwarding equivalence class (FEC)
    - A group of IP packets that are forwarded in the same manner
  - Manages LSP to egress router
    - LDP associates the FEC with each LSP it creates
    - LDP creates an LSP tree for each FEC from every possible ingress router to egress router

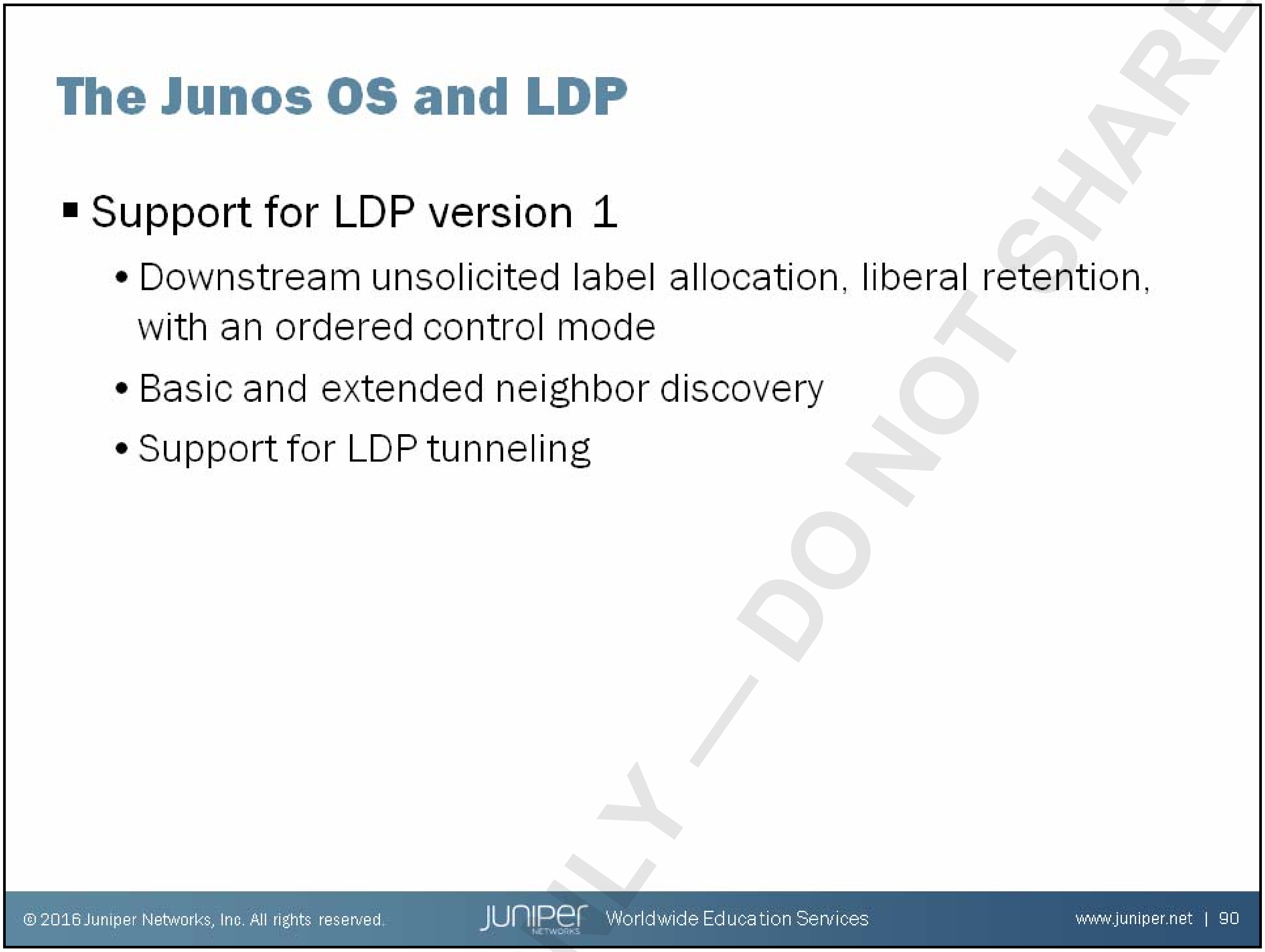
Egress LSR
R1
R2
R3
R4
R5
Site 1
Site 2
Site 3
LDP LSP

JUNIPER NETWORKS Worldwide Education Services

www.juniper.net | 89

Purpose of LDP

LDP associates a set of destinations (prefixes) with each LSP. This set of destinations is called the FEC. These destinations share a common LSP path and egress router, as well as a common unicast routing path. LDP maps groups of prefixes to an egress router at the end of an LSP. LDP manages the LSP to the egress router for each FEC. LDP is not related to RSVP or traffic engineering concepts from previous lectures. LDP maps the FECs (prefixes) to label values. The LSP forwarding paths look like a unicast forwarding path, in that MPLS traffic for the ultimate destination is forwarded along the unicast forwarding tree. LDP allows multiple prefixes to share the same label mapping. No constraints are allowed when signaling the LSPs. The LSPs must follow the IGP best path. LDP merges together traffic from different tunnels, which results in fewer total tunnels than would be required with RSVP. LDP will create a LSP tree for each FEC from every possible ingress point in the LDP network to the egress point. Each LDP speaking router will advertise the addresses reachable via a MPLS label into the LDP domain. The label information is exchanged in a hop by hop fashion so every LSR in the domain will become an ingress router to all other routers in the network. This process creates a full mesh LDP environment. The slide displays what LSPs will be generated for the FEC egressing at R5.



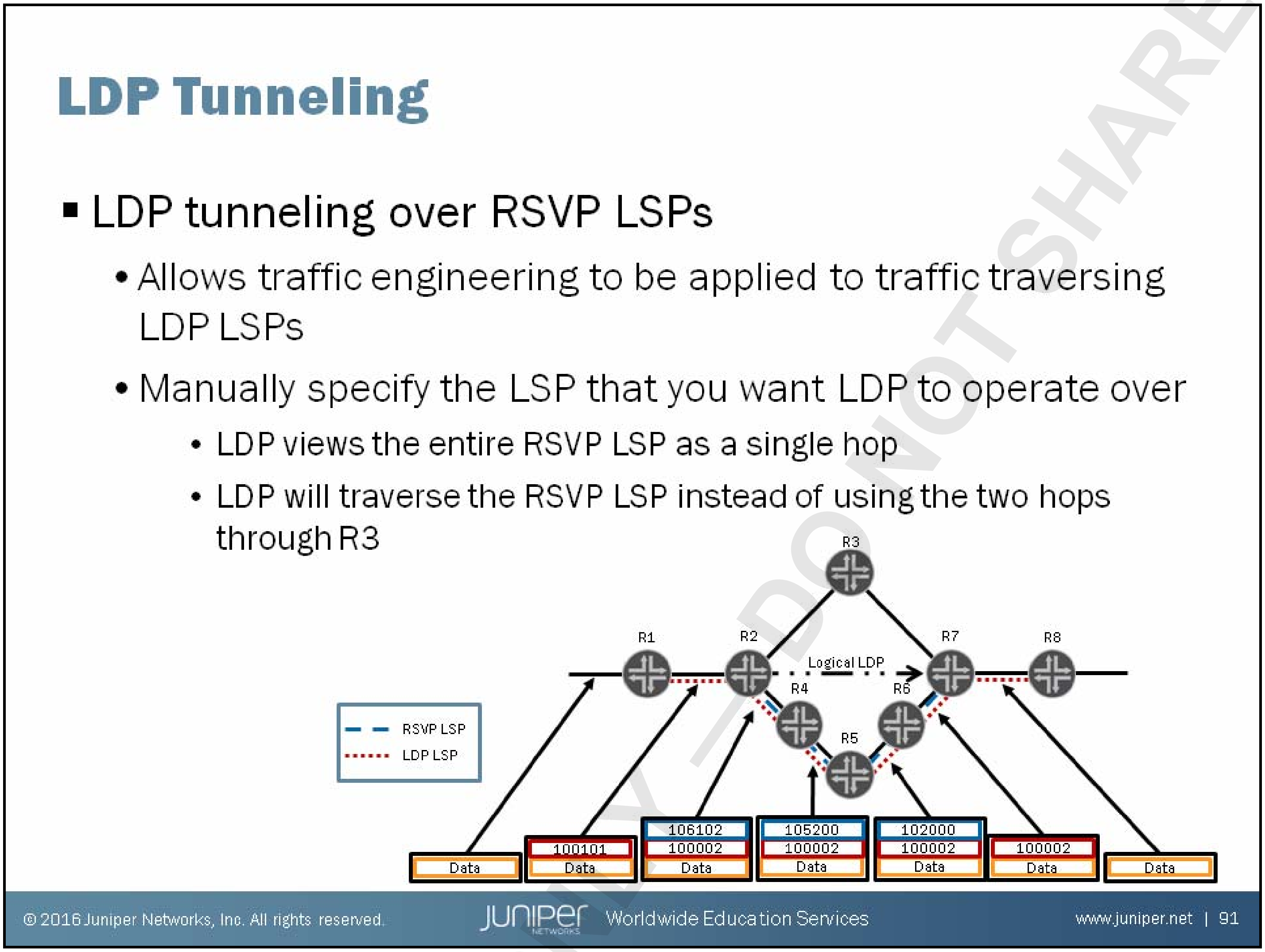
The Junos OS and LDP

- Support for LDP version 1
  - Downstream unsolicited label allocation, liberal retention, with an ordered control mode
  - Basic and extended neighbor discovery
  - Support for LDP tunneling

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 90

Junos OS LDP Implementation

The Junos OS implementation of LDP supports LDP Version 1. Constraint-Based Routed Label Distribution Protocol (CR-LDP) is not supported. The Junos OS implementation of LDP supports the “ordered downstream unsolicited with liberal label retention” mode defined in RFC 3036. This means that each LDP peer will store all label bindings received (liberal retention), that each downstream peer will advertise all FECs for which it is prepared to receive labeled traffic (downstream unsolicited), and that FECs are only advertised when the router is the traffic’s egress point, or it has received a label mapping for the traffic’s next hop (ordered). With the Junos OS using the minimum LDP configuration, LSRs will form LSPs to the /32 router ID of all LDP capable routers that are reachable. Basic neighbor discovery forms an LDP session with a directly connected neighbor because the hello messages have a destination address of 224.0.0.2; messages sent to these addresses are not routed. Extended discovery allows peers to establish LDP sessions through an RSVP-signaled LSP, thus allowing some level of traffic engineering for LDP traffic. You explicitly configure the destination address of the hello messages when using extended discovery; because a routable IP address is specified, the LDP peer can be reached via IP routing and no longer needs to be directly connected.



LDP Tunneling

- LDP tunneling over RSVP LSPs
  - Allows traffic engineering to be applied to traffic traversing LDP LSPs
  - Manually specify the LSP that you want LDP to operate over
    - LDP views the entire RSVP LSP as a single hop
    - LDP will traverse the RSVP LSP instead of using the two hops through R3

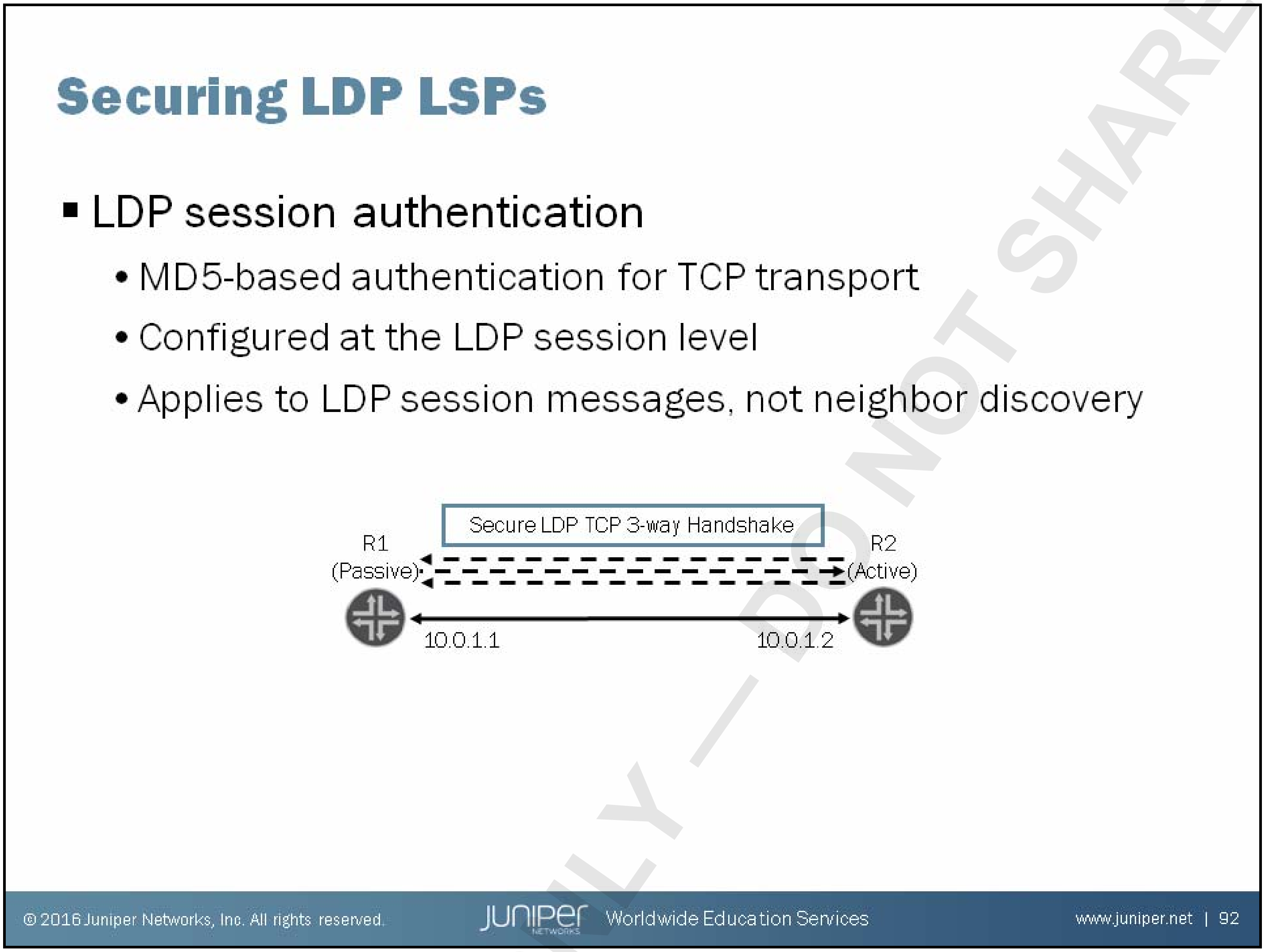
JUNIPer NETWORKS Worldwide Education Services

www.juniper.net | 91

LDP Tunneling

You can tunnel LDP LSPs over RSVP-signaled LSPs using label stacking. Note that in Junos you must enable LDP on the lo0.0 interface to support extended neighbor discovery needed for this application. Additionally, you must configure the LSPs over which you want LDP to operate.

This slide shows that LDP-over-RSVP tunneling results in LDP traffic being forwarded through the RSVP tunnel, which itself takes a traffic engineered path. By default, LDP always follows the IGP's shortest path, which in this case, would be the 3-hop path at the top of the slide. LDP views the RSVP LSP as a single hop, therefore the RSVP path becomes the more preferred path even though the LSP actually traverses 5 hops. The label assignment is also shown on this slide. When the traffic enters into the LDP LSP it pushes a label value of 100101. When received on R2 it accepts the packet based on the assigned incoming label. R2 will lookup the route and identify that the route will be sent over the RSVP LSP. R2 pushes on the LDP label value of 100002 and then stacks an outer label value of 106102 for the RSVP label. When the packet is received at R4 it accepts the packet based on the RSVP label and swaps the outer label with the label assigned to the outgoing interface (105200) and forwards the traffic to the next LSR. R5 received the packet based on the incoming RSVP label. R5 swaps the RSVP label value with the next label (102000) and forwards it on to R6. R6 will also process the packet based on the RSVP label. Since R6 is the penultimate LSP for the RSVP LSP it will pop the RSVP label and forward to the next LSR with the LDP label and R7 will accept and forward the pact based on the LDP label.



Securing LDP LSPs

- LDP session authentication
  - MD5-based authentication for TCP transport
  - Configured at the LDP session level
  - Applies to LDP session messages, not neighbor discovery

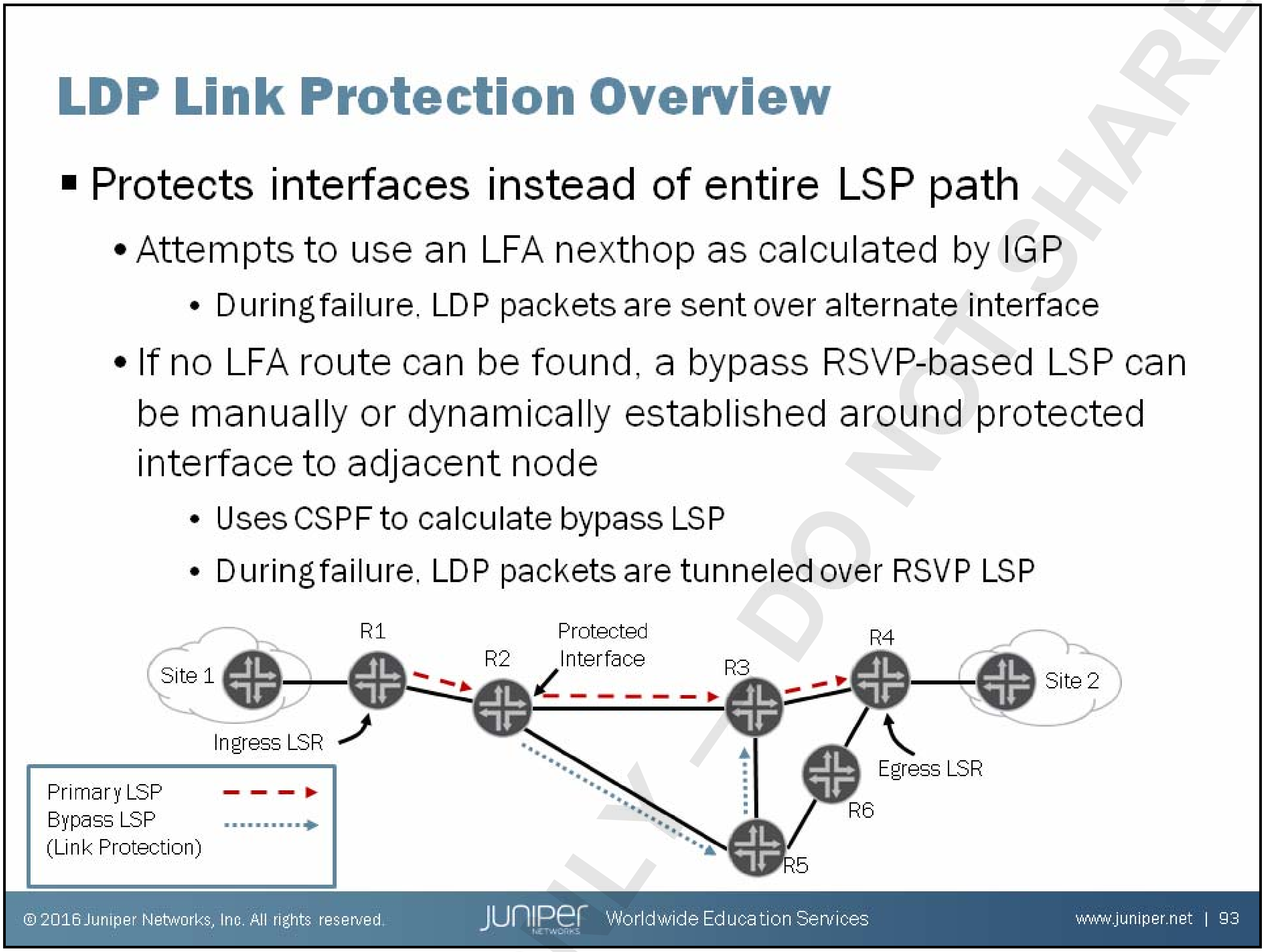
Secure LDP TCP 3-way Handshake
R1
(Passive)
R2
(Active)
10.0.1.1
10.0.1.2

Worldwide Education Services

www.juniper.net | 92

Securing LDP

If desired you can configure MD5-based authentication for the TCP transport protocol that supports LDP sessions. LDP session authentication does not apply to the UDP-based neighbor discovery mechanism. Thus, mismatched LDP authentication settings permit LDP neighbor discovery and adjacency formation, but the LDP session will not establish without compatible authentication values.



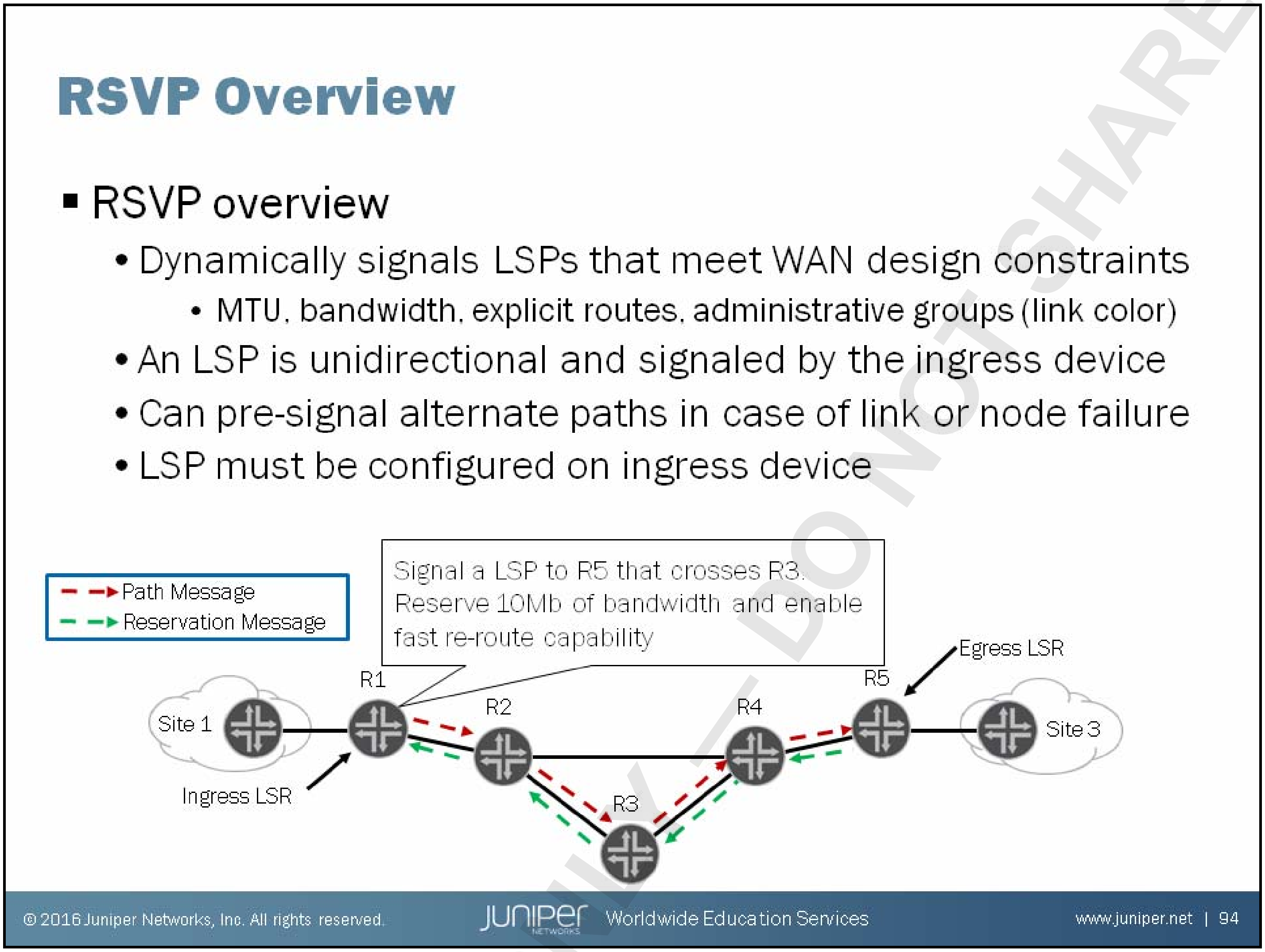
LDP Link Protection Overview

- Protects interfaces instead of entire LSP path
  - Attempts to use an LFA nexthop as calculated by IGP
    - During failure, LDP packets are sent over alternate interface
  - If no LFA route can be found, a bypass RSVP-based LSP can be manually or dynamically established around protected interface to adjacent node
    - Uses CSPF to calculate bypass LSP
    - During failure, LDP packets are tunneled over RSVP LSP

JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 93

LDP Link Protection Overview

Label Distribution Protocol (LDP) link protection is similar to the RSVP link protection in that it protects the failure of an individual interfaces instead of the entire path. As you know, LDP relies on the IGP to determine the label and interface that should be used as the next hop of an LSP. The same is true for link protection feature of LDP. IS-IS and OSPF each have a feature called link protection. When a link is configured under the IGP for link protection, the IGP attempts to find a loop-free alternate (LFA) next hop for all destinations that have a primary next hop of the protected interface. The calculated LFA next hop is then installed in the routing table as if it was an equal cost next hop to the destination. If there is a link failure on the protected interface, the LFA next hop is already available in the routing table for forwarding. LDP will also use the LFA next hop and associated label during a failure. It is possible that the IGP may be unable to calculate a loop free path to a downstream router on an LDP LSP. In that case, a bypass RSVP-signaled LSP can be either manually or dynamically created between the two LDP neighbors (LDP tunneling is used over the RSVP LSP).



RSVP Overview

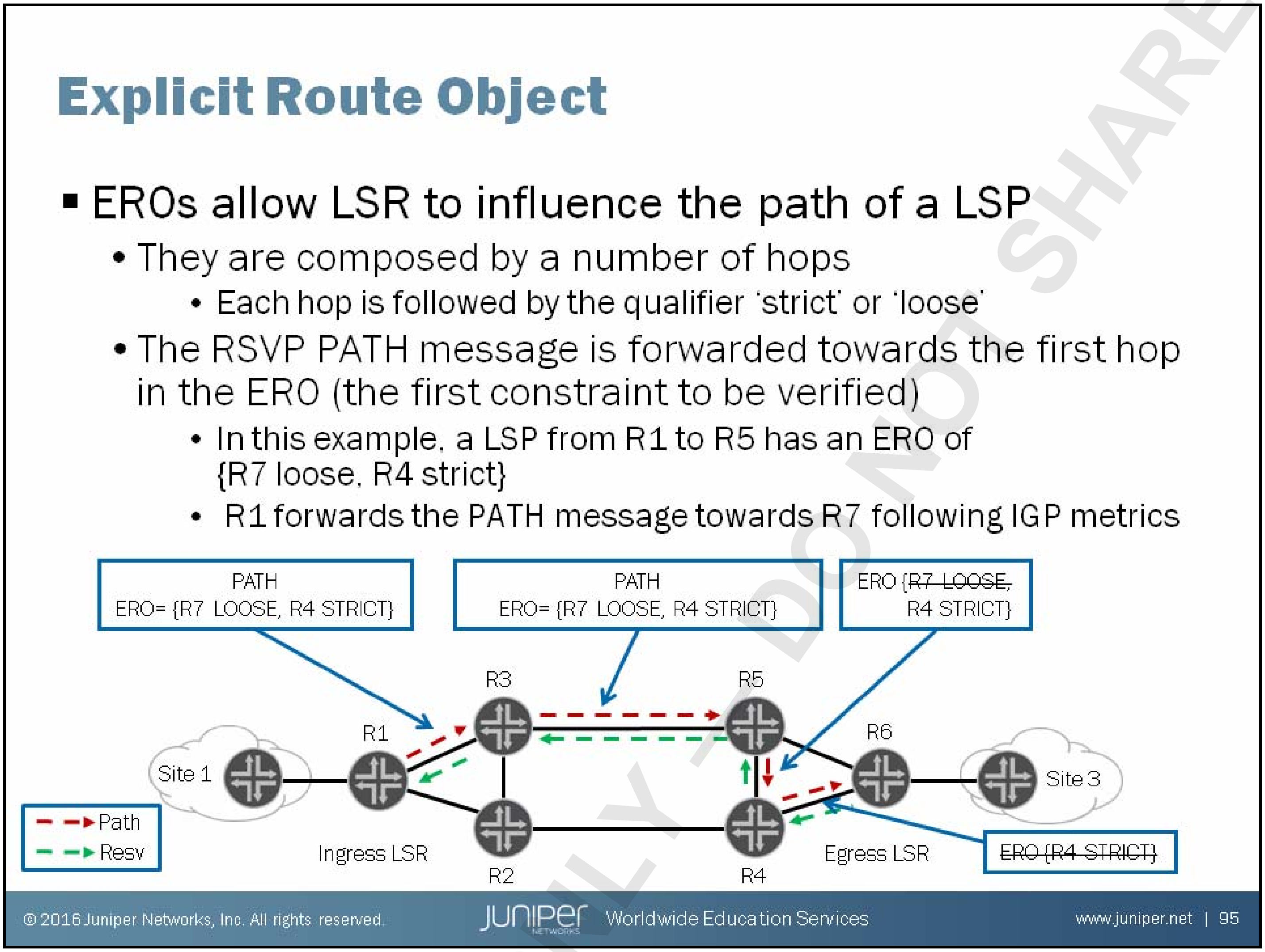
- RSVP overview
  - Dynamically signals LSPs that meet WAN design constraints
    - MTU, bandwidth, explicit routes, administrative groups (link color)
  - An LSP is unidirectional and signaled by the ingress device
  - Can pre-signal alternate paths in case of link or node failure
  - LSP must be configured on ingress device

Signal a LSP to R5 that crosses R3. Reserve 10Mb of bandwidth and enable fast re-route capability

Worldwide Education Services www.juniper.net | 94

RSVP Overview

RSVP is a generic signaling protocol designed originally to be used by applications to request and reserve specific quality-of-service (QoS) requirements across an internetwork. Resources are reserved hop by hop across the internetwork; each router receives the resource reservation request, establishes and maintains the necessary state for the data flow (if the requested resources are available), and forwards the resource reservation request to the next router along the path. As this behavior implies, RSVP is an internetwork control protocol, similar to Internet Control Message Protocol (ICMP), Internet Group Management Protocol (IGMP), and routing protocols. RSVP does not transport application data, nor is it a routing protocol. It is simply a label distribution protocol. RSVP uses unicast and multicast IGP routing protocols to discover paths through the internetwork by consulting existing routing tables. The Junos OS uses RSVP as the label distribution protocol for traffic engineered LSPs. RSVP Data Flows RSVP requests resources for unidirectional data flows. Each reservation is made for a data flow from a specific sender to a specific receiver. While RSVP messages are exchanged between the sender and receiver, the resulting path itself is unidirectional. Although the application data flow is from the sender to the receiver, the reservation itself is initiated by the receiver. The sender notifies the receiver of a pending flow and characterizes the flow, and the receiver is responsible for requesting the resources. This design choice was made to accommodate heterogeneous receiver requirements and for multicast flows in which multiple receivers join and leave a multicast group.



Explicit Route Object

- EROs allow LSR to influence the path of a LSP
  - They are composed by a number of hops
    - Each hop is followed by the qualifier 'strict' or 'loose'
  - The RSVP PATH message is forwarded towards the first hop in the ERO (the first constraint to be verified)
    - In this example, a LSP from R1 to R5 has an ERO of {R7 loose, R4 strict}
    - R1 forwards the PATH message towards R7 following IGP metrics

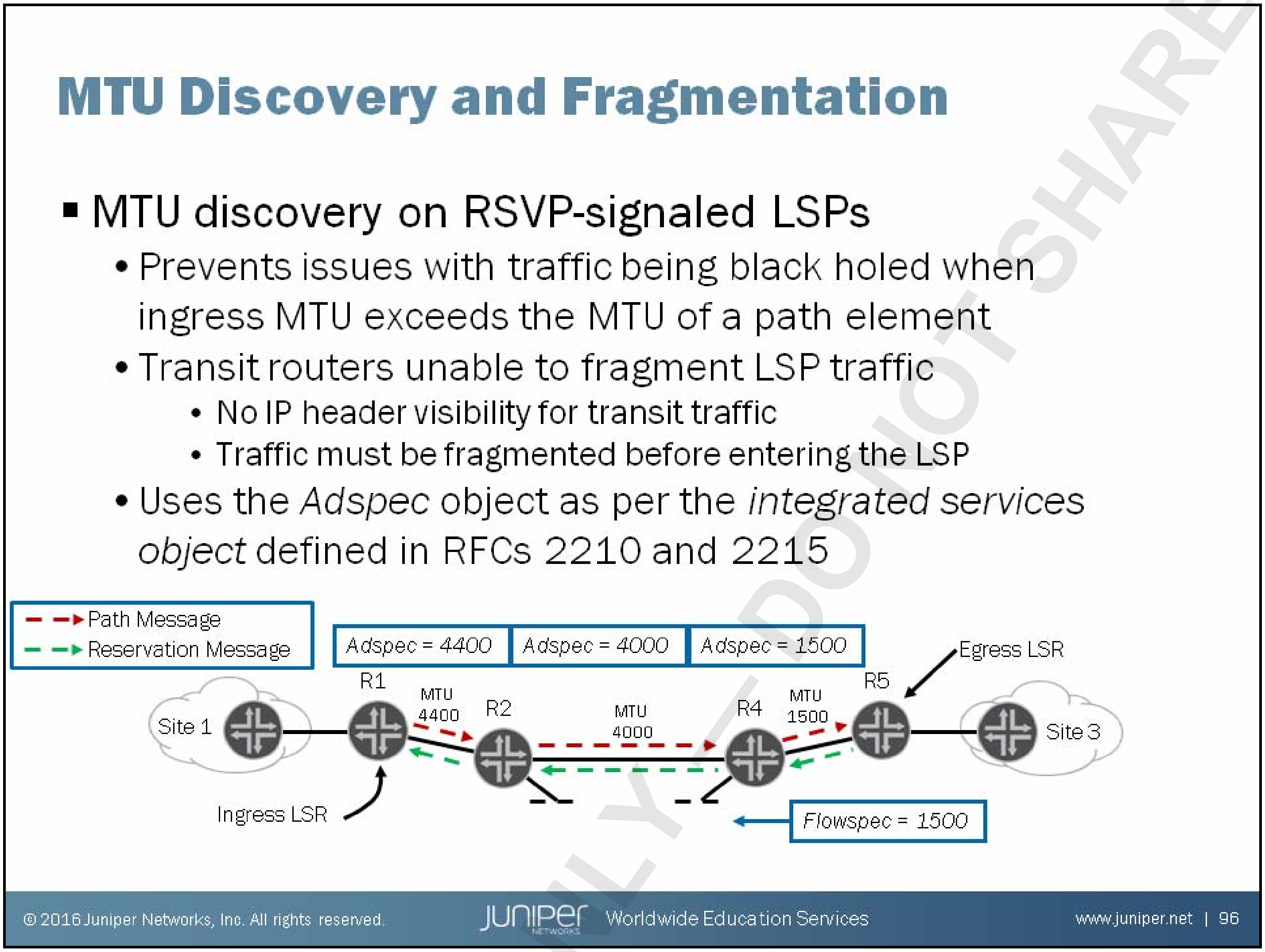
PATH
ERO= {R7 LOOSE, R4 STRICT}
PATH
ERO= {R7 LOOSE, R4 STRICT}
ERO {~~R7 LOOSE,~~ R4 STRICT}
R3
R5
R1
R6
Site 1
Site 3
Path
Resv
Ingress LSR
Egress LSR
~~ERO {R4 STRICT}~~
R2
R4

Worldwide Education Services

www.juniper.net | 95

EROs Allow LSR to Influence the Path of a LSP

By including the EXPLICIT_ROUTE Object (ERO) in the PATH message, the ingress LSR can control the path through which a LSP is established, independently from the IGP best path. This is by far the most important feature of RSVP as a label distribution protocol, as it allows it to be used for traffic engineering. You can think about an EROs as a list of hops that the LSP should be traversing, or a list of path constraints that the LSP needs to verify. Each hop within an ERO has either a qualifier of strict (the default) or loose. The use of EROs changes during the RSVP LSP setup process. To understand how, we can follow a simple example of an ERO composed of two constraints, a loose hop followed by a strict one.



MTU Discovery and Fragmentation

- MTU discovery on RSVP-signaled LSPs
  - Prevents issues with traffic being black holed when ingress MTU exceeds the MTU of a path element
  - Transit routers unable to fragment LSP traffic
    - No IP header visibility for transit traffic
    - Traffic must be fragmented before entering the LSP
  - Uses the *Adspec* object as per the *integrated services object* defined in RFCs 2210 and 2215

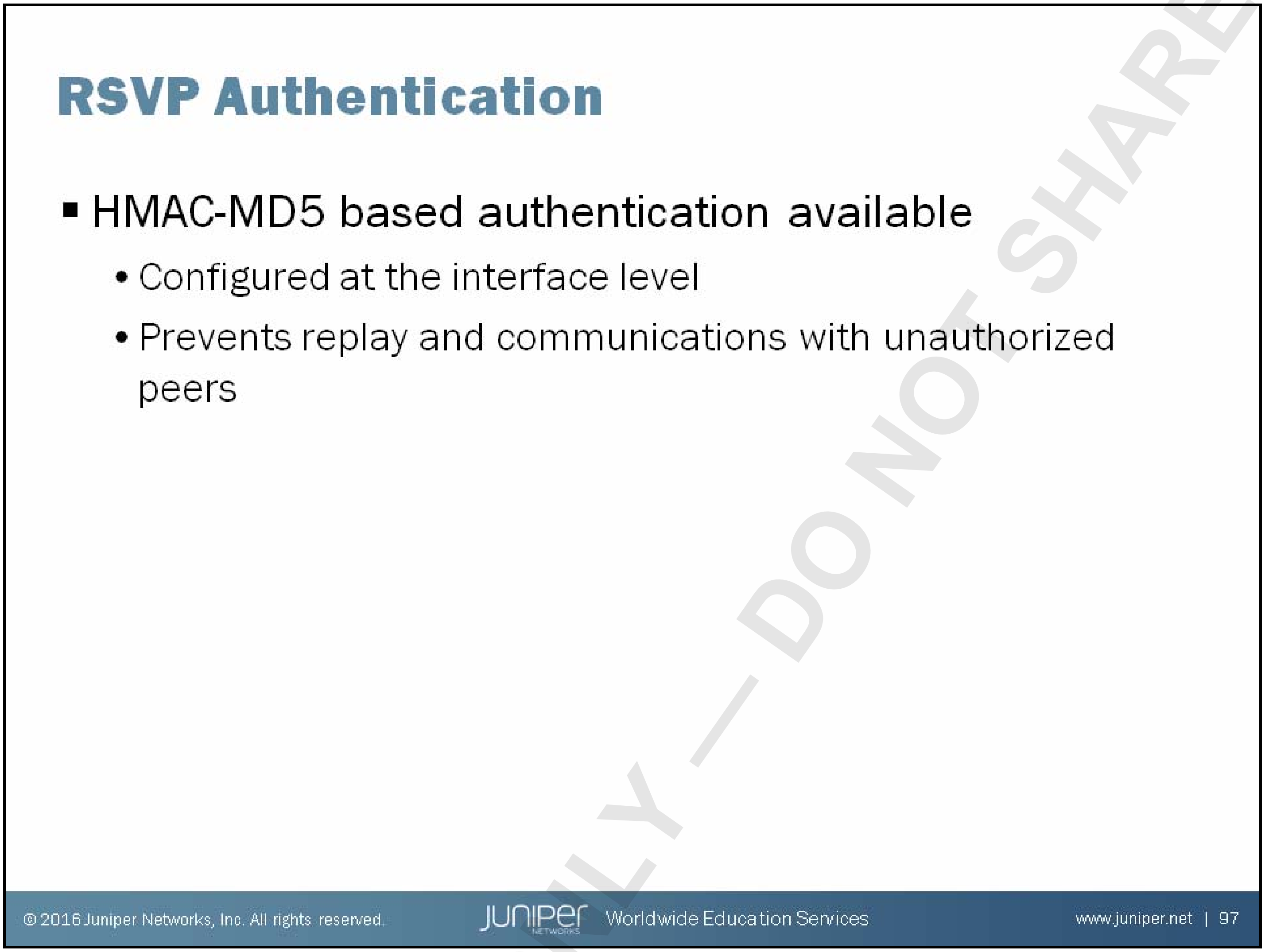
Worldwide Education Services

www.juniper.net | 96

MTU Discovery for RSVP-Signaled LSPs

The Junos OS supports maximum transmission unit (MTU) discovery when using RSVP signaling. The discovery mechanism is performed according to the integrated services object as defined in RFCs 2210 and 2215. This feature helps to prevent the black hole condition that is normally associated with mismatched MTUs along an the elements that make up an LSP.

In operation, the ingress LSR sets the M value in the TSPEC to 9192 and codes the egress interface's IP MTU in the ADSPEC object in the path message. At each hop transit LSRs update the MTU value in the ADSPEC object with the minimum of the incoming value and egress interface MTU. When the path message is received by the egress LSR the smaller of the two values coded in the TSPEC and ADSPEC objects is signaled back to the ingress router using the Flowspec object in the Resv message. This behavior is shown on the slide where the 1500-byte MTU is correctly reported to the egress router in the ADSPEC object.



RSVP Authentication

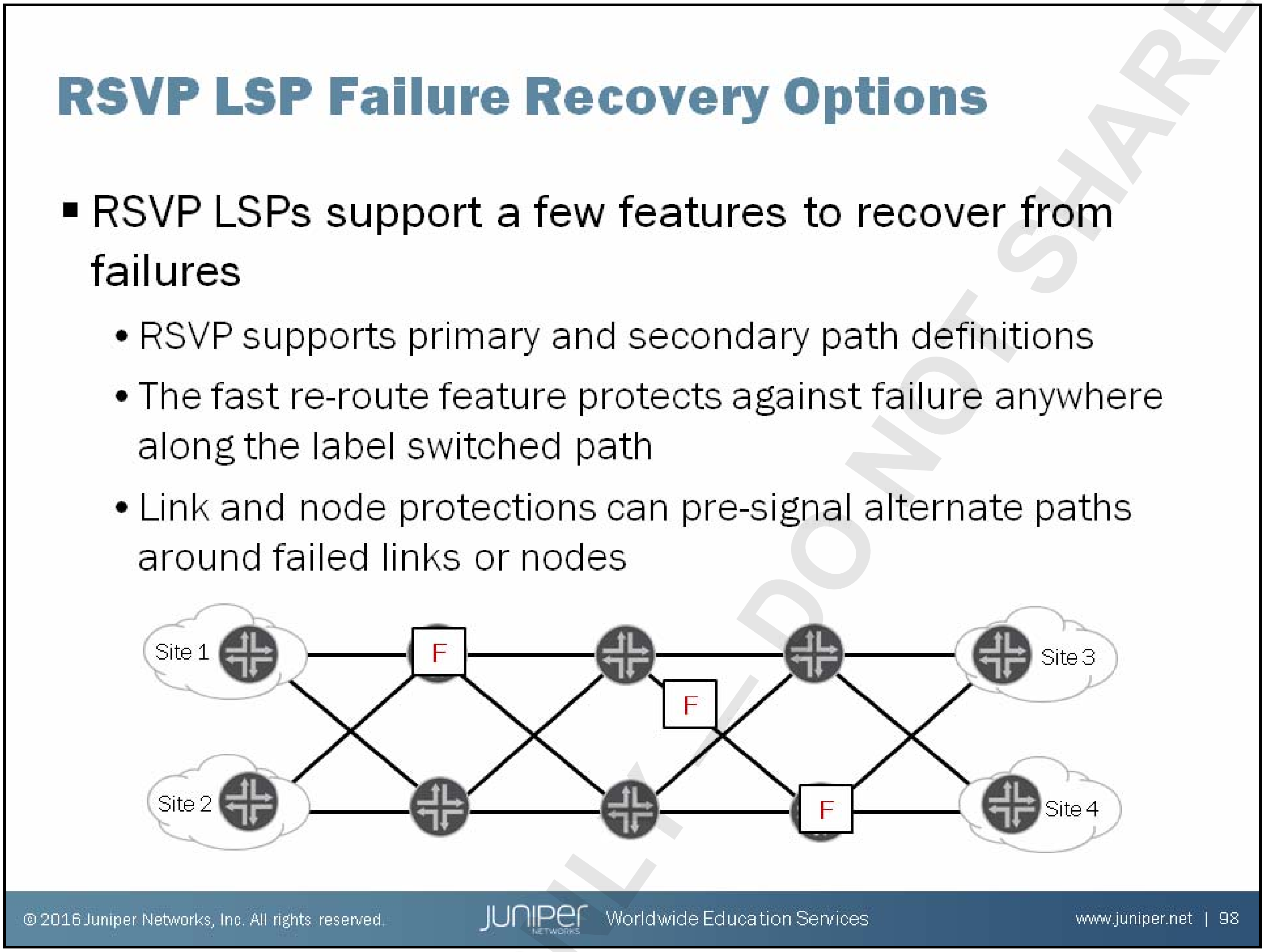
- HMAC-MD5 based authentication available
  - Configured at the interface level
  - Prevents replay and communications with unauthorized peers

Worldwide Education Services

www.juniper.net | 97

Authenticate RSVP Exchanges

When necessary, designers can configure Hashed Message Authentication Code (HMAC)-Message Digest 5 (MD5) authentication for RSVP exchanges based on the procedures defined in RFC 2747. Once configured, all RSVP messages are authenticated using a message digest based on a shared secrete key. Sequence numbers are added to all messages to prevent replay attacks.



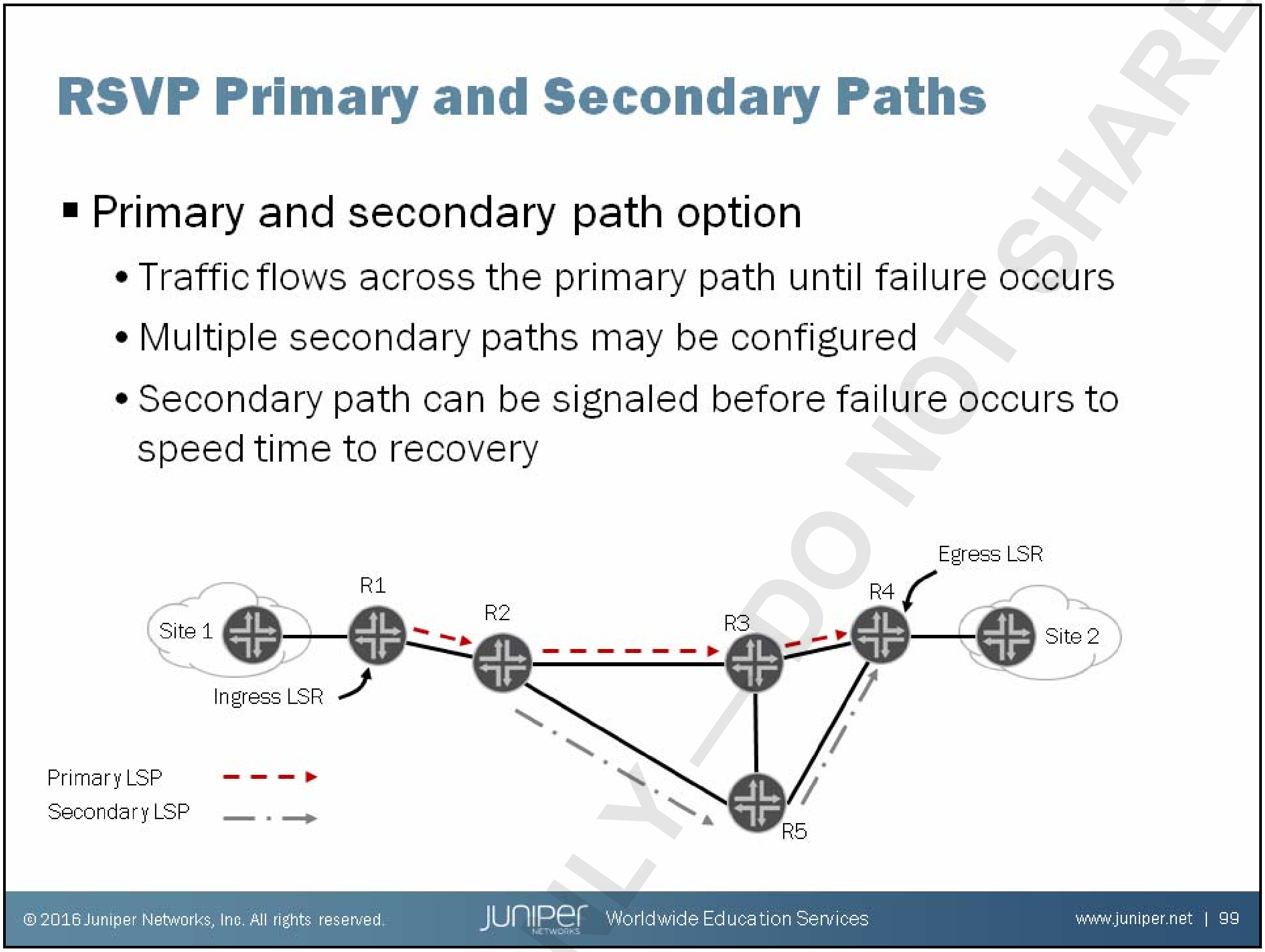
RSVP LSP Failure Recovery Options

- RSVP LSPs support a few features to recover from failures
  - RSVP supports primary and secondary path definitions
  - The fast re-route feature protects against failure anywhere along the label switched path
  - Link and node protections can pre-signal alternate paths around failed links or nodes

Worldwide Education Services www.juniper.net | 98

LSP Failure Recover Options

RSVP allows LSPs to signal alternate paths that can route traffic around link and node failures. The next few slides examine some RSVP options to help overcome network failures.



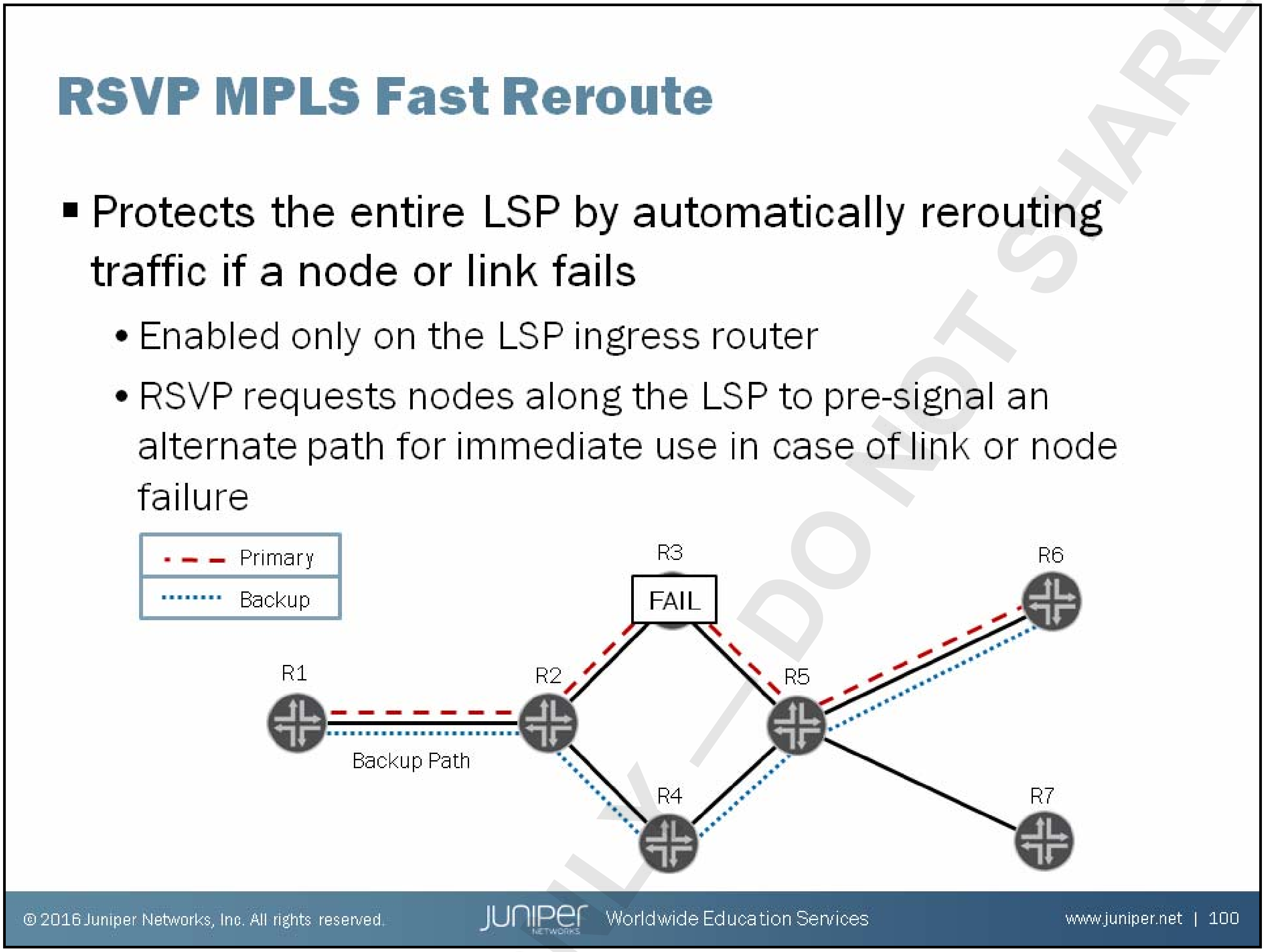
RSVP Primary and Secondary Paths

- Primary and secondary path option
  - Traffic flows across the primary path until failure occurs
  - Multiple secondary paths may be configured
  - Secondary path can be signaled before failure occurs to speed time to recovery

Juniper Networks Worldwide Education Services www.juniper.net | 99

Primary and Secondary Paths

Primary paths are optional. Only one primary path is permitted per LSP definition. Within the primary physical path you can specify parameters, like explicit route objects, bandwidth or priority, that affect only the primary physical path. Like primary paths, secondary paths are also optional. By default, a secondary path becomes active when a primary, or another secondary, physical path fails. The Junos operating system does not require that a primary and secondary path share the same parameters. You can decide to configure your primary paths with stringent resource requirements while your secondary paths are far more lax in their demands. Such asymmetric settings helps to ensure that your secondary paths can be established during periods of diminished resources. You can specify standby for a secondary path. This command causes the router to signal the secondary path, even though the secondary path is not currently needed, that is, the primary path has not yet failed. Note that standby secondaries result in routers having to maintain additional state in the form of the pre-established standby LSPs.



RSVP MPLS Fast Reroute

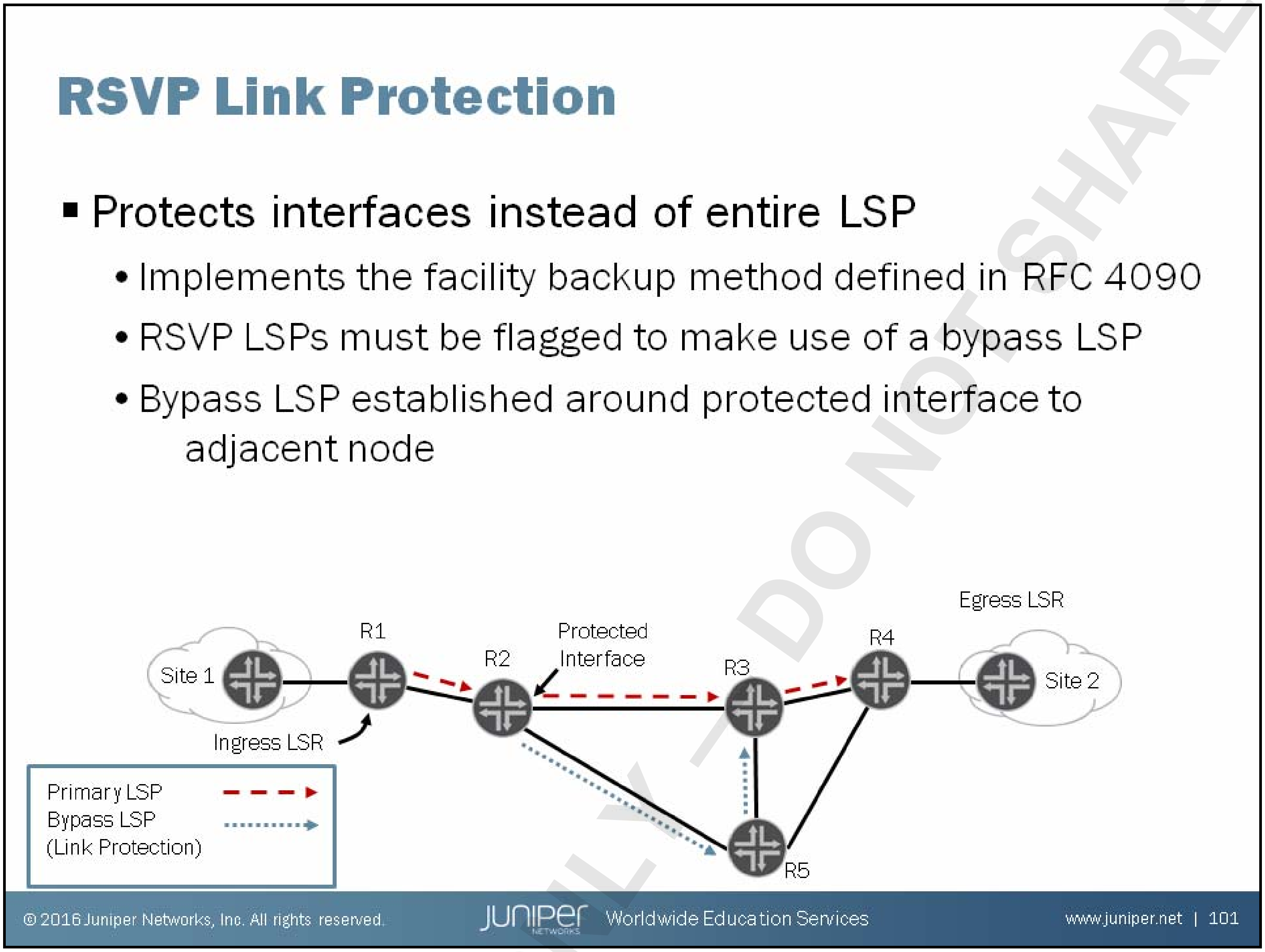
- Protects the entire LSP by automatically rerouting traffic if a node or link fails
  - Enabled only on the LSP ingress router
  - RSVP requests nodes along the LSP to pre-signal an alternate path for immediate use in case of link or node failure

Worldwide Education Services

www.juniper.net | 100

RSVP MPLS Fast Reroute

Fast reroute provides a way for intermediate LSRs to immediately start forwarding traffic over an alternate route while simultaneously alerting the ingress LSR to the presence of downstream link or node failures You configure fast reroute to minimize the effects of a LSP failure. Fast reroute enables a router upstream of the failure to quickly route around the failure while the primary path is torn down and resignaled. The router that detects the primary path failure signals the outage to the ingress router. Fast reroute serves as an interim connectivity mechanism during the establishment of a new primary path. Once the new primary path is signaled, the fast reroute detours associated with the original paths are torn down; fast reroute is a short-term solution. When fast reroute is enabled, the ingress router adds an object to the RSVP Path messages requesting that downstream routers establish reroute detours. These downstream routers then originate detour Path messages to detour the LSP around that LSR's downstream link and node. When an active physical path fails and a detour is available, the upstream router sends a PathErr message to the ingress router. This message triggers new CSPF computations and a switchover to an alternate path if available. If a fast reroute detour is not available, the downstream node sends a ResvTear message and begins withdrawing the MPLS labels, which brings down the LSP. A fast reroute path might stay up indefinitely if an alternative primary path is not found.

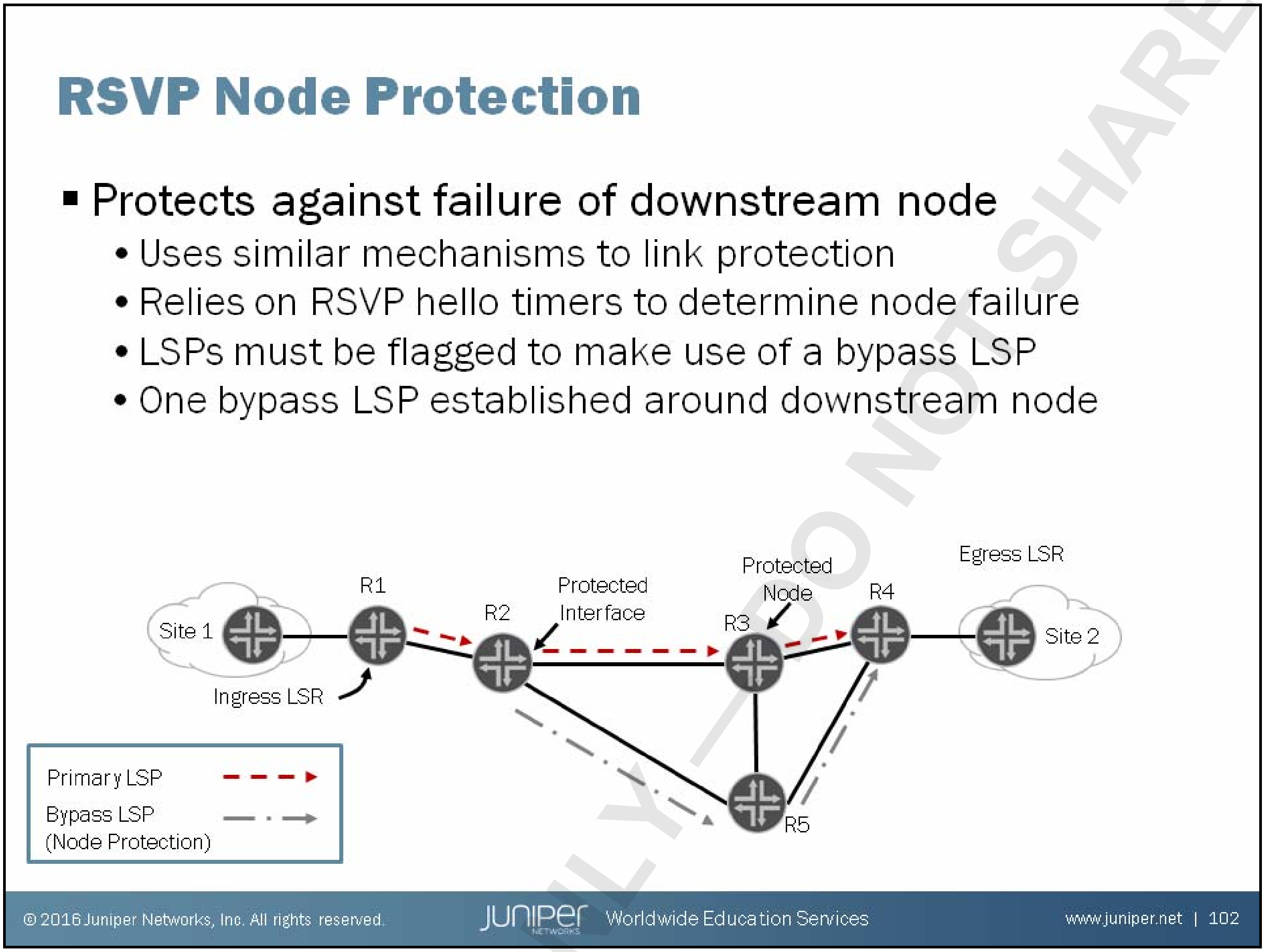


RSVP Link Protection

- Protects interfaces instead of entire LSP
  - Implements the facility backup method defined in RFC 4090
  - RSVP LSPs must be flagged to make use of a bypass LSP
  - Bypass LSP established around protected interface to adjacent node

Protect Interfaces

Link protection is the Junos OS nomenclature for the facility backup feature defined in RFC 4090. The link protection feature is interface based, rather than LSP based. The slide shows how the R2 node is protecting its interface and link to R3 through a bypass LSP that is calculated using CSPF and the node’s TED. While fast reroute attempts to protect the entire path of a given LSP, you can apply link protection on a per-interface basis as needed. LSPs must be tagged for them to make use of a bypass LSP, and you can provide an ERO list to influence the CSPF-based routing of the bypass LSP.



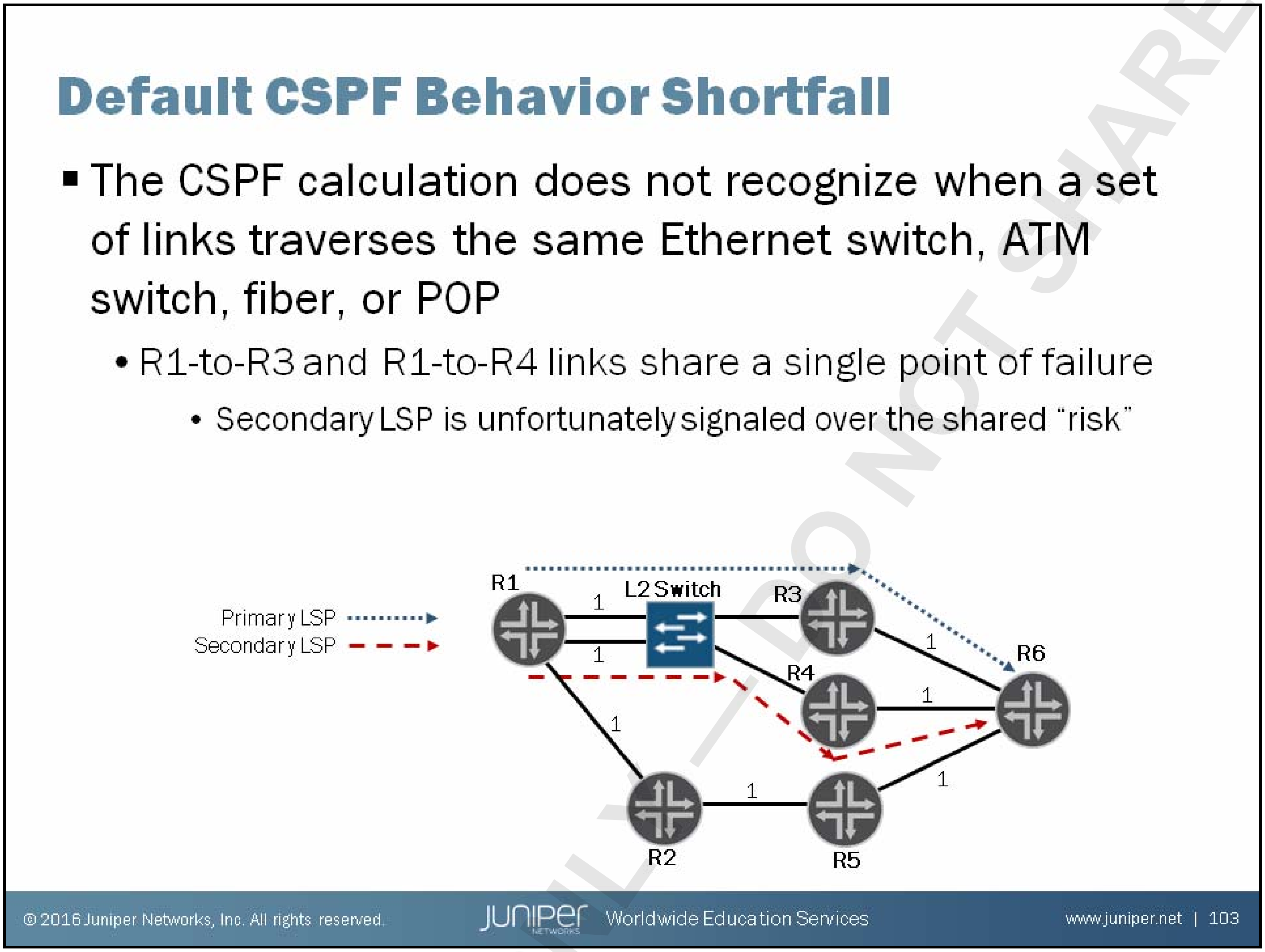
RSVP Node Protection

- Protects against failure of downstream node
  - Uses similar mechanisms to link protection
  - Relies on RSVP hello timers to determine node failure
  - LSPs must be flagged to make use of a bypass LSP
  - One bypass LSP established around downstream node

JUNIPer NETWORKS Worldwide Education Services www.juniper.net | 102

Node Protection

Node protection is the Junos OS nomenclature for the facility backup feature defined in RFC 4090. Node protection uses the same messaging as link protection. The slide shows that R2 is protecting against the complete failure of R3 through a bypass LSP that is calculated using CSPF. LSPs must be tagged for node-link protection to make use of the bypass LSPs, and you can provide an ERO list to influence the CSPF-based routing of the bypass LSP.



Default CSPF Behavior Shortfall

- The CSPF calculation does not recognize when a set of links traverses the same Ethernet switch, ATM switch, fiber, or POP
  - R1-to-R3 and R1-to-R4 links share a single point of failure
    - Secondary LSP is unfortunately signaled over the shared "risk"

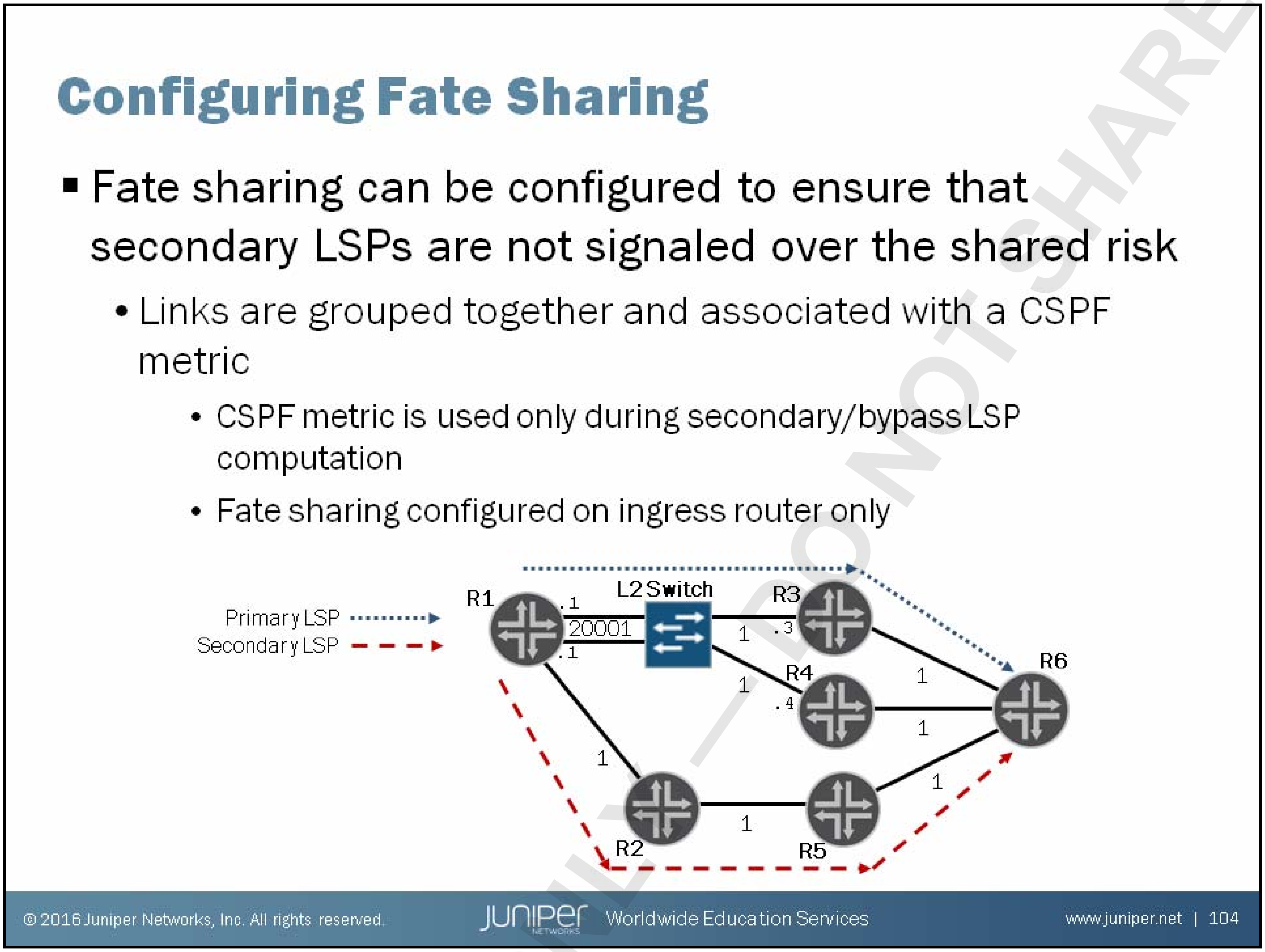
© 2016 Juniper Networks, Inc. All rights reserved.

Worldwide Education Services

www.juniper.net | 103

CSPF Behavior Shortfall

The default fate sharing behavior might cause the Junos OS to not make a very good choice for the path of a secondary LSP. By default, the fate sharing behavior is not able to recognize when multiple links are traversing the same fiber (different wavelength), ATM switch, Ethernet switch, or point of presence (POP). The CSPF algorithm is simply attempting to find the best sequence of IP hops to reach the egress router. It does not take into consideration the underlying layer 1 or layer 2 topology. R1 will chose the R1-R4-R6 path for the secondary because of the lower IGP cost to the egress router. Unfortunately, both primary and secondary LSPs are now traversing the same single point of failure, the Ethernet Switch.



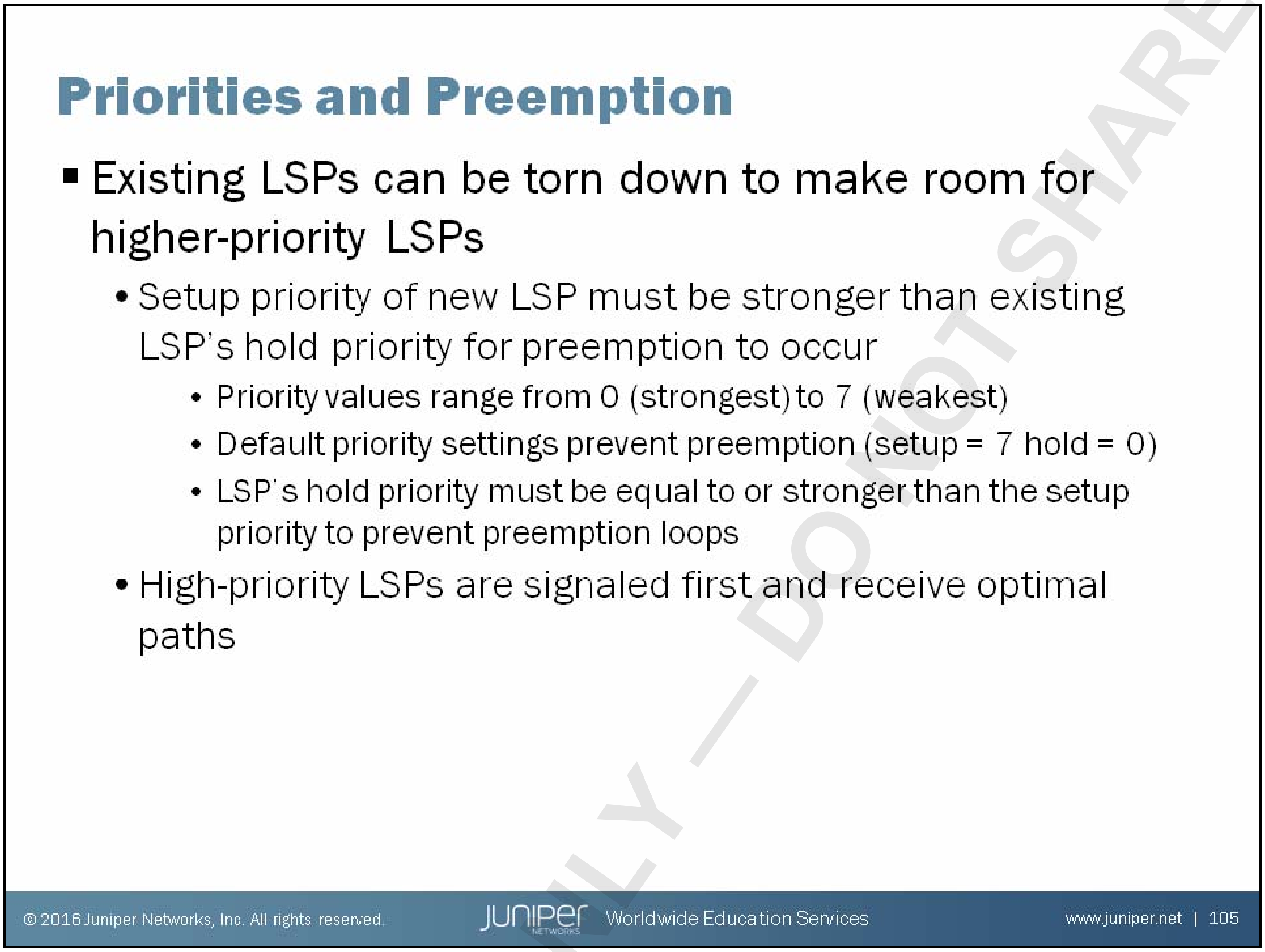
Configuring Fate Sharing

- Fate sharing can be configured to ensure that secondary LSPs are not signaled over the shared risk
  - Links are grouped together and associated with a CSPF metric
    - CSPF metric is used only during secondary/bypass LSP computation
    - Fate sharing configured on ingress router only

JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 104

Configuring Fate Sharing

The slide shows how you can group a set of links or nodes together and associate a CSPF metric with them. The fate sharing configuration is used only by the ingress router during the CSPF calculation of a secondary LSP. In the example on the slide, R1 determines that the primary LSP traverses one of the links or nodes in the ethernet-switch fate sharing group. Prior to running the CSPF algorithm to determine the path of the secondary LSP, R1 adds the cost (20000 in the example) to those links. Although the links in the fate sharing group are still available to be used, they are not chosen for the path of the secondary LSP because the cost to get to R6 along the R1-R2-R5-R6 path is only 3 (compared to 20002 along the R1-R4-R6 path).



Priorities and Preemption

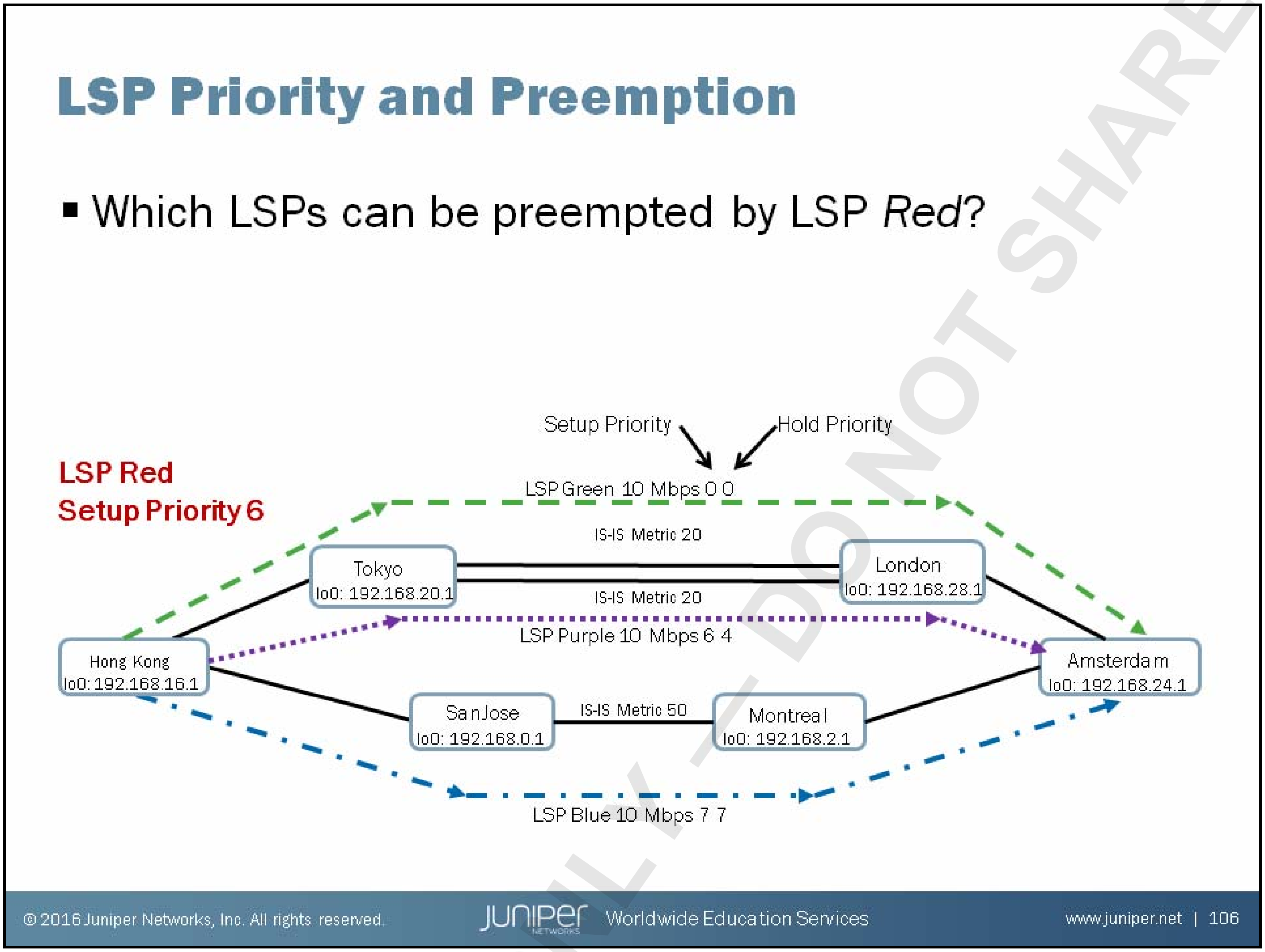
- Existing LSPs can be torn down to make room for higher-priority LSPs
  - Setup priority of new LSP must be stronger than existing LSP's hold priority for preemption to occur
    - Priority values range from 0 (strongest) to 7 (weakest)
    - Default priority settings prevent preemption (setup = 7 hold = 0)
    - LSP's hold priority must be equal to or stronger than the setup priority to prevent preemption loops
  - High-priority LSPs are signaled first and receive optimal paths

Worldwide Education Services

www.juniper.net | 105

LSP Priorities and Preemption

RSVP-signaled LSPs support setup and hold priorities. These priorities work together to determine the relative priority of a new LSP that must be established versus the hold priority of existing LSPs. When insufficient network resources exist to accommodate all LSPs simultaneously, an LSP with a strong setup priority preempts—or causes the teardown—of an existing LSP with a weaker hold priority. At software startup, LSPs are signaled in order from strongest to weakest setup priority; this behavior ensures that high-priority LSPs are established first and are afforded optimal paths. LSP setup and hold priorities range in value from 0 (the strongest) to 7 (the weakest). The default settings disable preemption by assigning all LSPs the weakest setup priority (7) and the strongest hold priority (0).

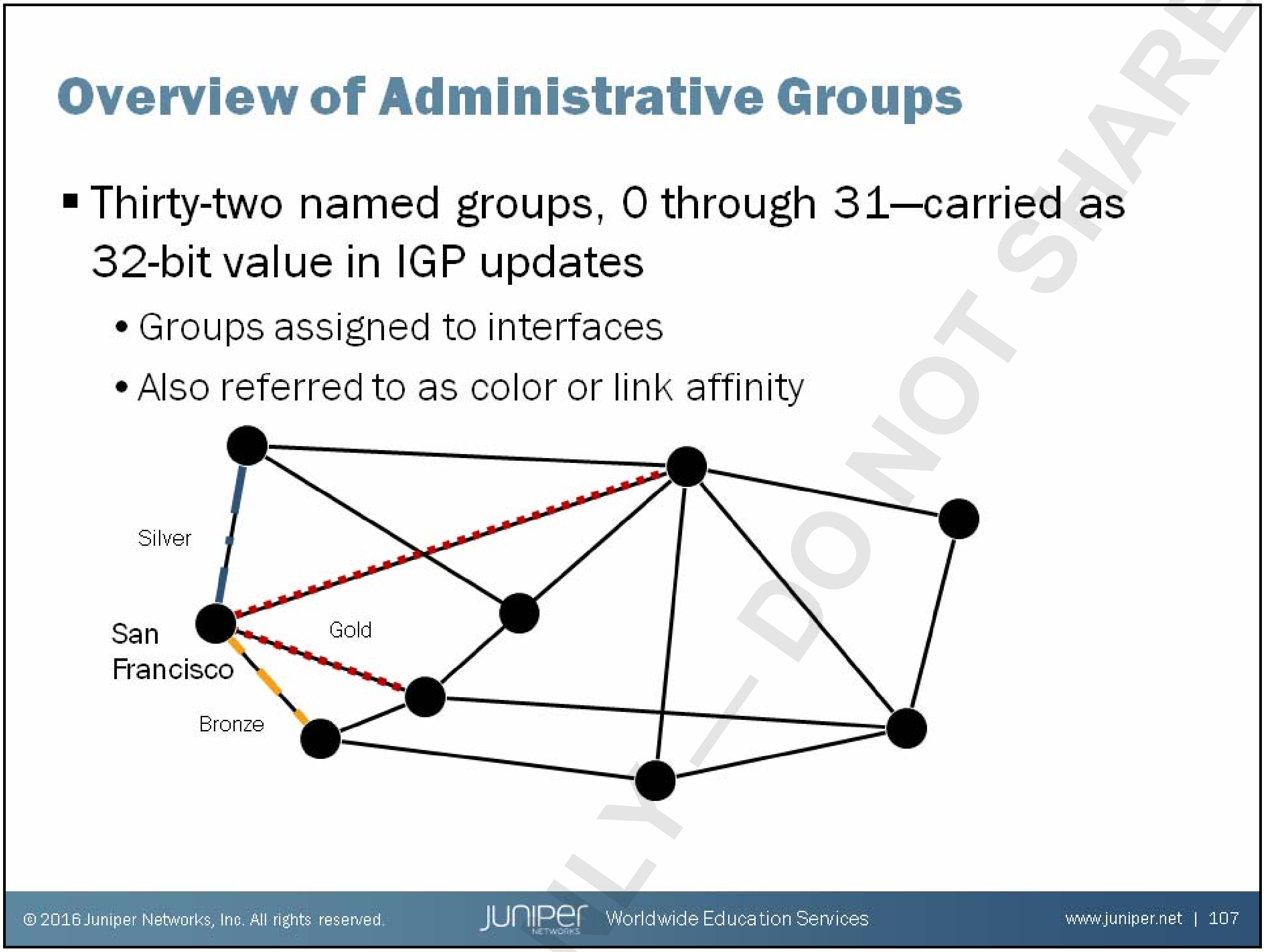


LSP Priority and Preemption

- Which LSPs can be preempted by LSP *Red*?

LSP Priority and Preemption

The setup priority for LSP Red is 6. The hold priority (the second number) for LSPs Green and Purple are both less than 6, which gives these LSPs a stronger hold priority and will prevent their preemption. In contrast, LSP Blue has a hold a priority of 7, which is weaker than LSP Red's setup priority. Thus, LSP Red can only preempt LSP Blue. Note the IS-IS metric has no effect on LSP preemption.



Overview of Administrative Groups

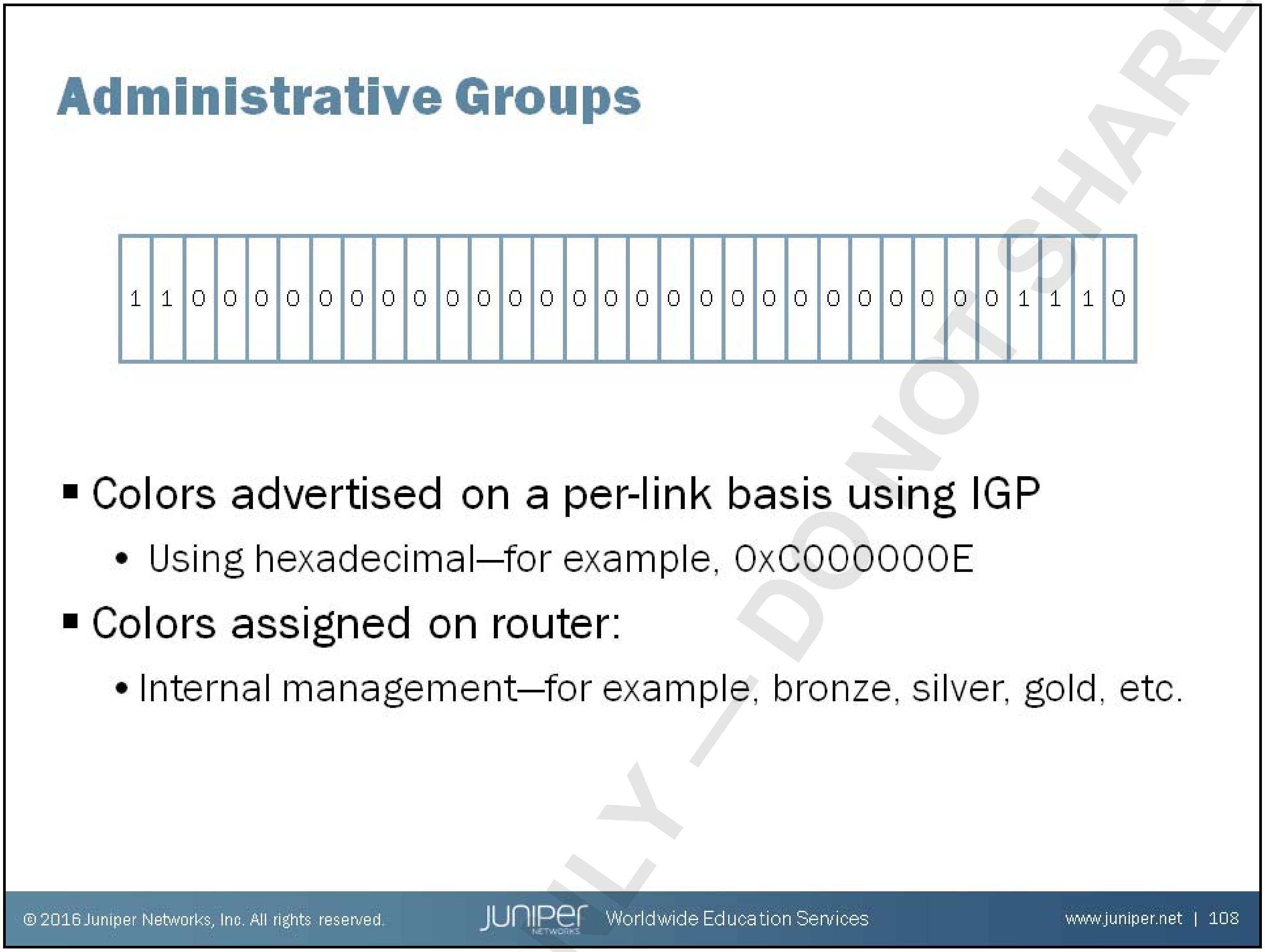
- Thirty-two named groups, 0 through 31—carried as 32-bit value in IGP updates
  - Groups assigned to interfaces
  - Also referred to as color or link affinity

Worldwide Education Services

www.juniper.net | 107

Administrative Group Overview

Administrative groups allow you to constrain the routing of an LSP to the set of links that meet the prescribed administrative groupings. Each interface can support 32 different administrative groups. The administrative groups associated with each interface is communicated through the extended IGP for storage in the TED. When the ingress router performs a CSPF computation, it includes or excludes links based on their associated colors, as specified in the LSP's definition. The net result is that the routing of the LSP will be controlled by its need to avoid, or make use of, links with the specified colors. If you use administrative groups, you must configure them identically on all routers participating in an MPLS domain. Great confusion results when a pair of routers do not agree on the color associated with mutually attached link. You can assign more than one administrative group to each physical link, or you might opt to leave one or more links uncolored by not assigning any administrative group values.



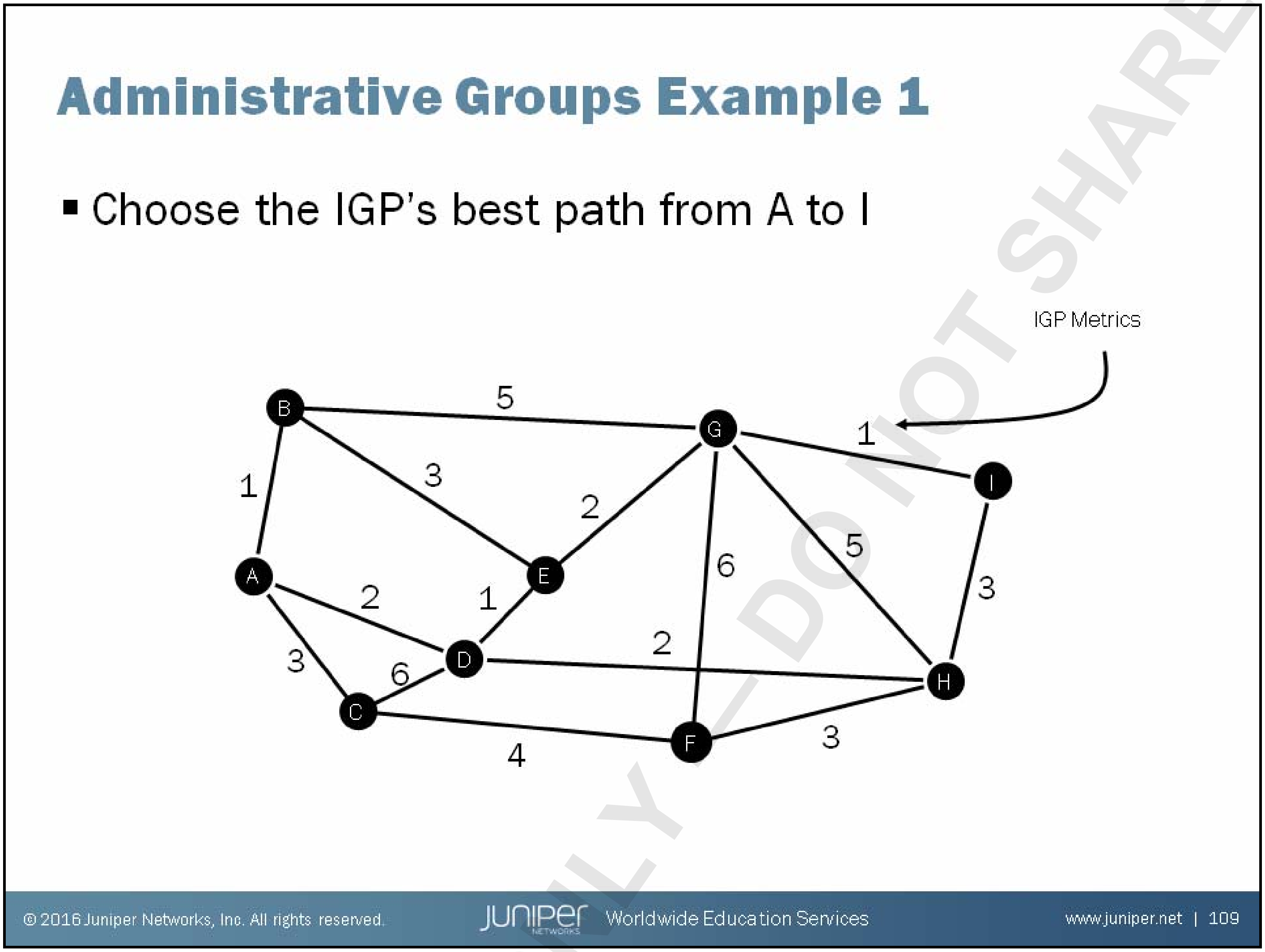
Administrative Groups

1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0

- Colors advertised on a per-link basis using IGP
  - Using hexadecimal—for example, 0xC000000E
- Colors assigned on router:
  - Internal management—for example, bronze, silver, gold, etc.

Administrative Group IGP Advertisements

A traffic engineering aware IGP communicates the administrative group of each interface as a 32-bit (4 bytes) bit mask. Each of the bit values in 32-bit sequence represents a different administrative group. Color Assignments Each bit value is correlated through configuration to a human-friendly name within Junos OS. This capability helps to simplify router management, as the name silver often means more to the typical human than the hexadecimal value of 0x02, for example. These names are often assigned as colors, but they do not have to be a color; they can be any descriptive term you want. Each link can have one or more bits enabled, and can therefore be associated with one or more colors simultaneously.



Administrative Groups Example 1

- Choose the IGP's best path from A to I

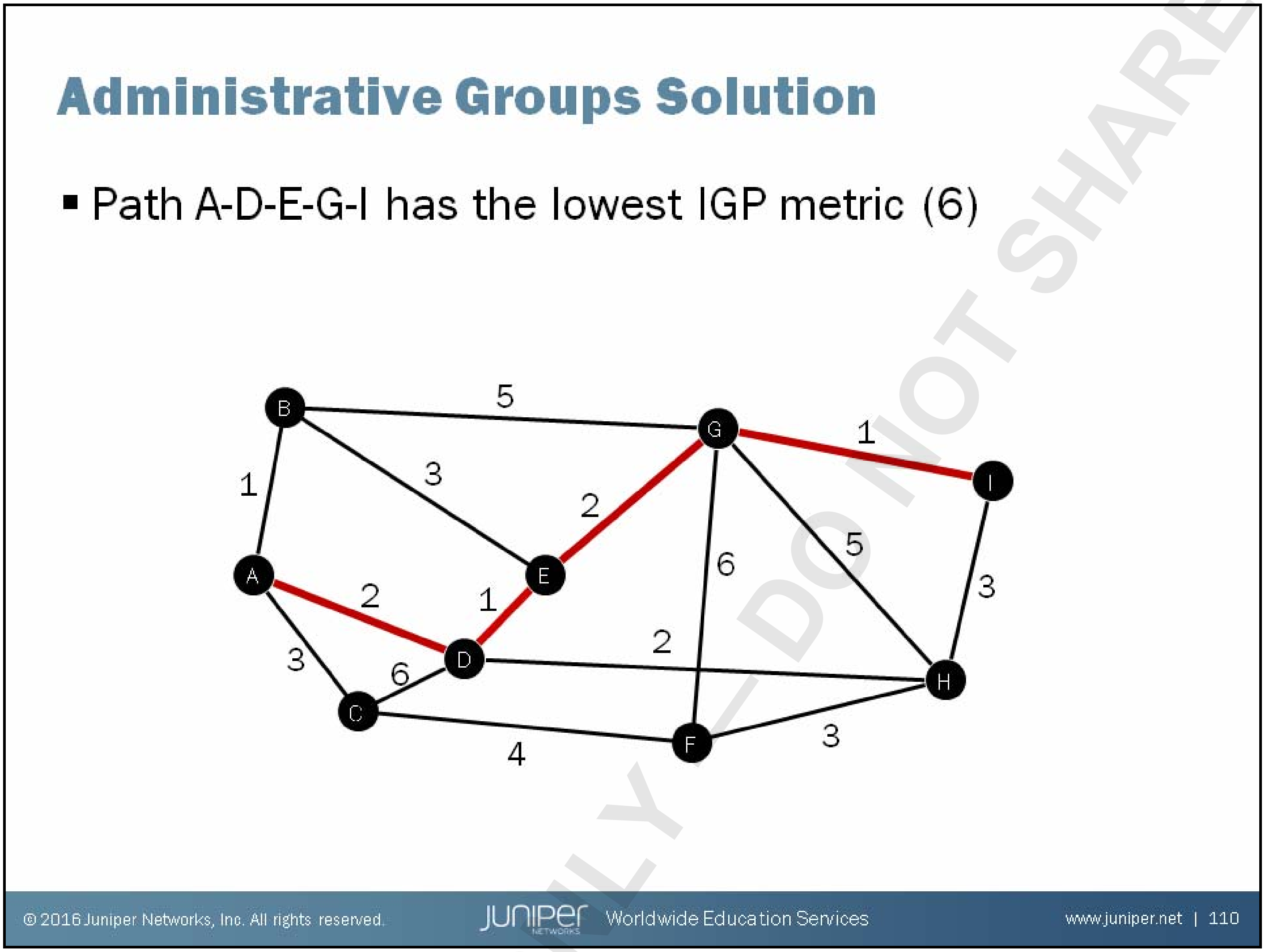
IGP Metrics

Worldwide Education Services

www.juniper.net | 109

Administrative Groups Example 1

In this initial example, you must determine the shortest path from A to I according to the perspective of the IGP. Each link displays the associated IGP metric value. It should not take you long to determine that the IGP's shortest path from A to I is path A-D-E-G-I, with a total cost of 6. This calculation reflects normal IGP processing and therefore, the default routing of an RSVP-signaled LSP. You can influence LSP routing with the inclusion of administrative constraints, as is demonstrated in subsequent pages in this section.



Administrative Groups Solution

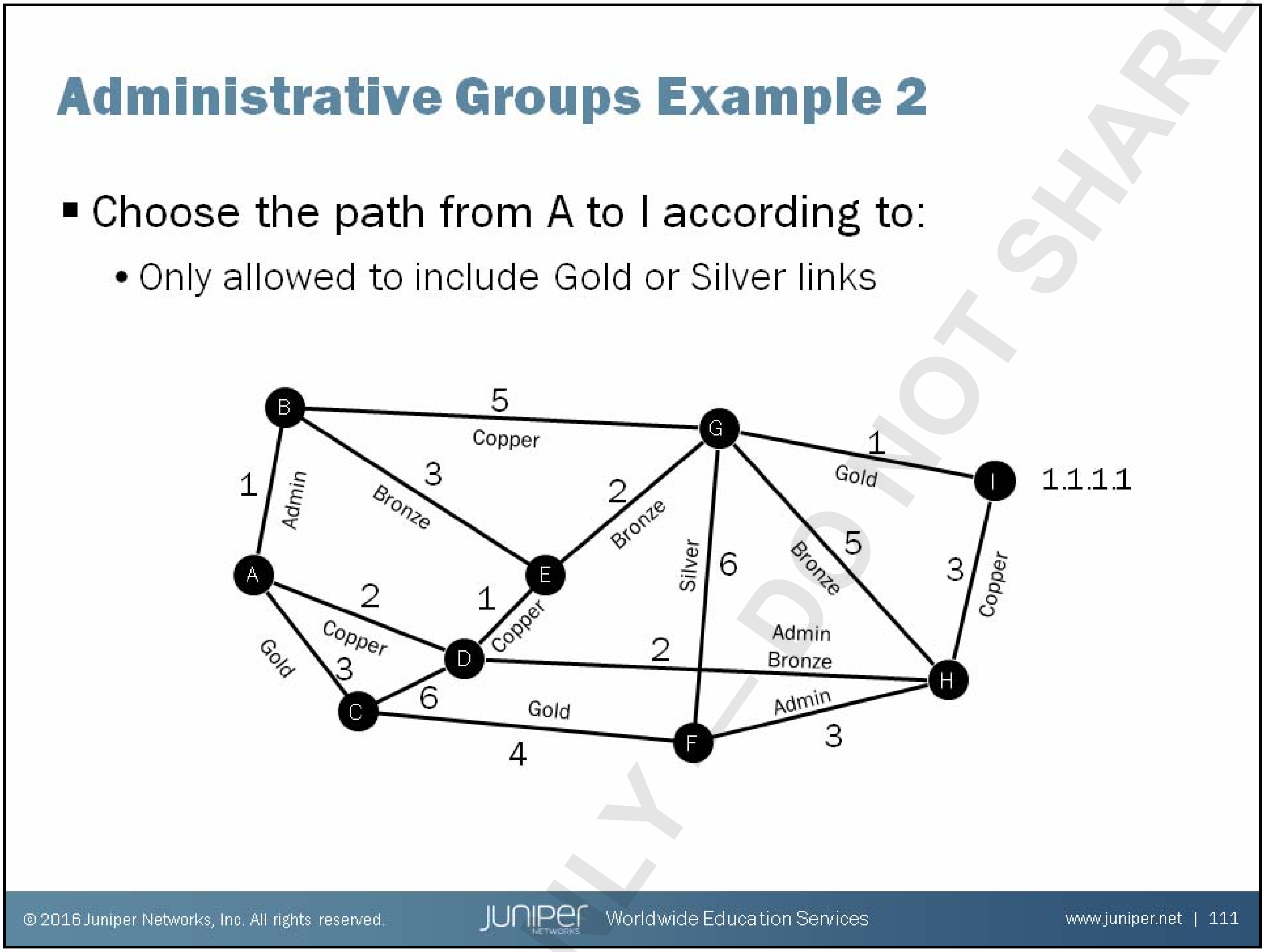
- Path A-D-E-G-I has the lowest IGP metric (6)

© 2016 Juniper Networks, Inc. All rights reserved.

www.juniper.net | 110

Administrative Groups Solution

This slide displays the solution to the question asked on the previous slide. In this case, the IGP's shortest path has a metric of 6 and consists of the path A, D, E, G, and I.



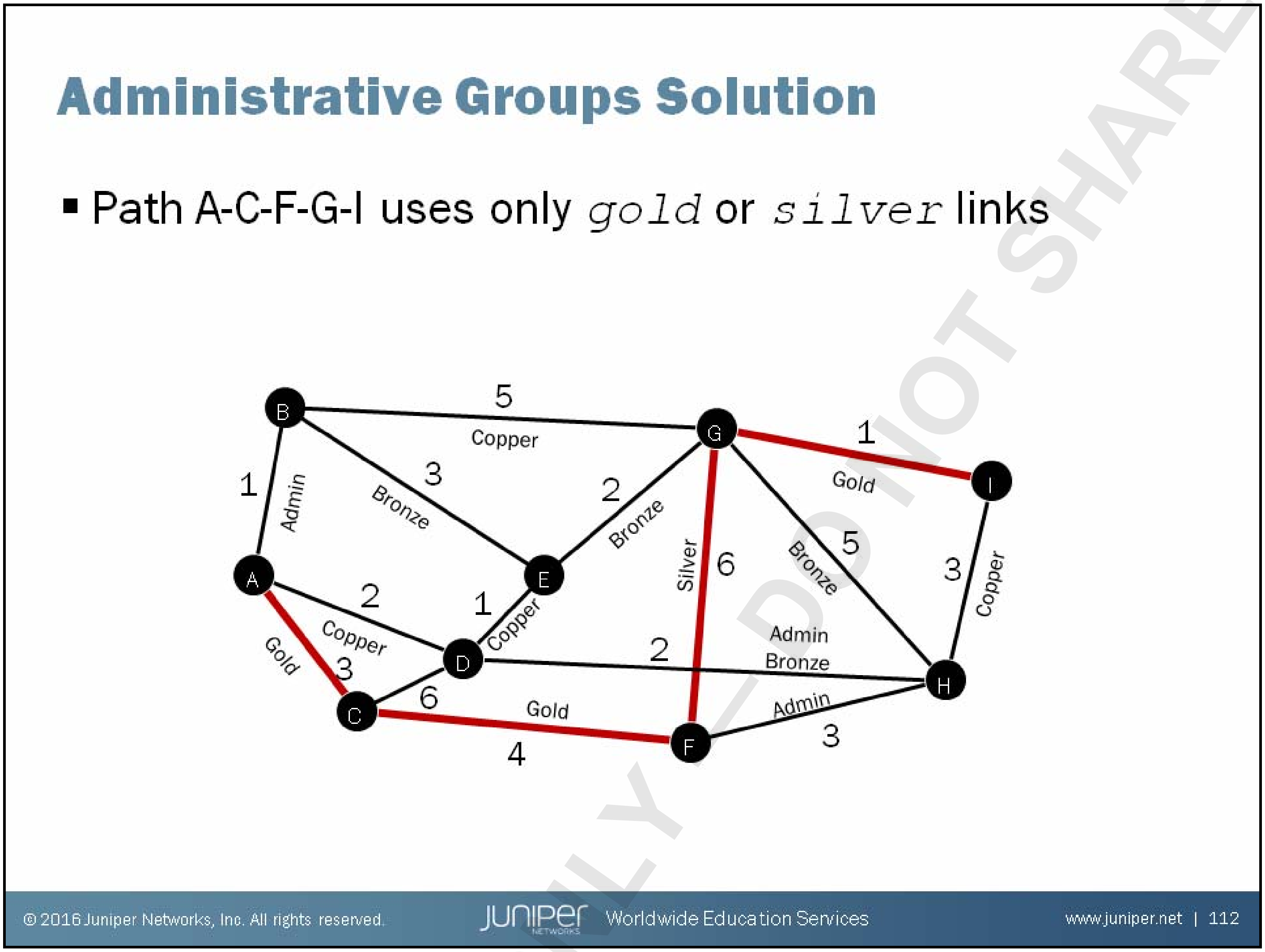
Administrative Groups Example 2

- Choose the path from A to I according to:
  - Only allowed to include Gold or Silver links

Worldwide Education Services www.juniper.net | 111

Administrative Groups Example 2

The LSP definition in this example requires that the link include either the color gold or the color silver. The CSPF algorithm begins by pruning the following links because they do not include the required colors: A-B, A-D, C-D, B-E, B-G, D-E, E-G, D-H, F-H, G-H, or H-I. The links that do comply with the constraints are A-C, C-F, F-G, and G-I. A shortest path is computed from the links that remain, which in this case yields only one viable path. The only path available, given these constraints, is shown on the next slide.



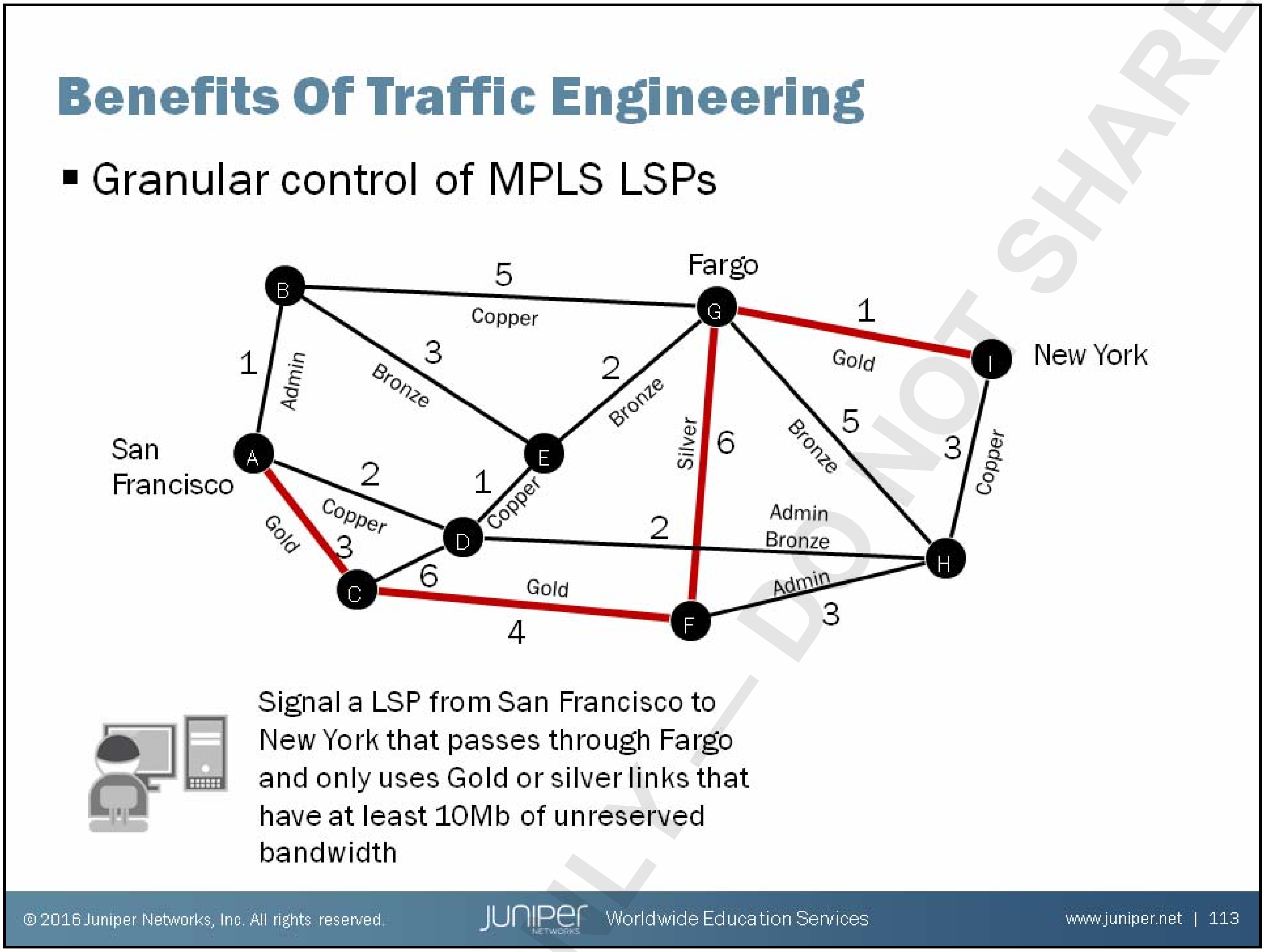
Administrative Groups Solution

- Path A-C-F-G-I uses only `gold` or `silver` links

© 2016 Juniper Networks, Inc. All rights reserved. JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 112

Administrative Groups Solution

This slide displays the solution to the question asked on the previous slide. In this case, the only path meeting the provided include-any constraints consists of the path A, C, F, G, and I.



Benefits Of Traffic Engineering

- Granular control of MPLS LSPs

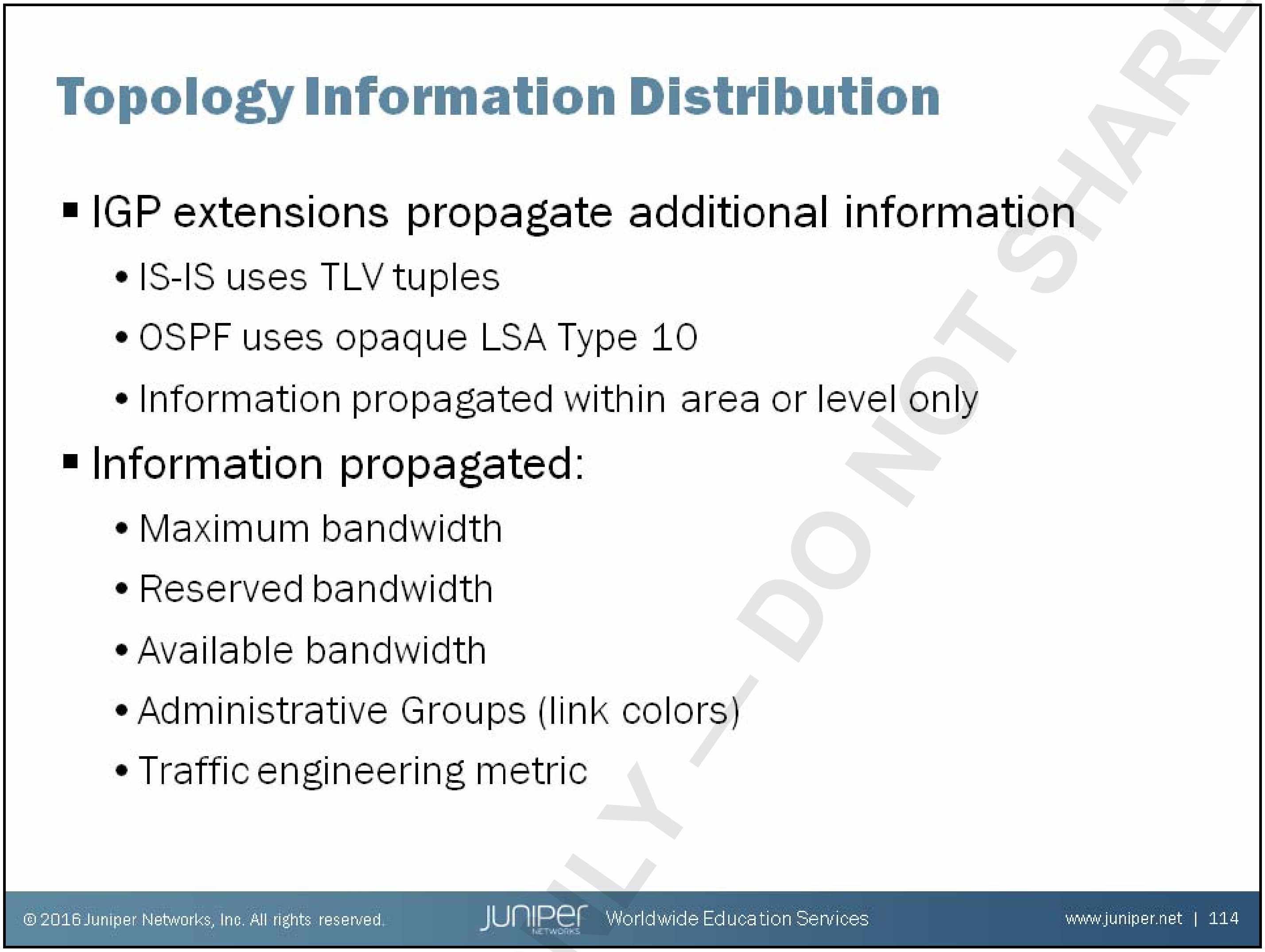
Signal a LSP from San Francisco to New York that passes through Fargo and only uses Gold or silver links that have at least 10Mb of unreserved bandwidth

JUNIPer NETWORKS Worldwide Education Services

www.juniper.net | 113

Traffic Engineering

The slide shows an example of additional routing constraints that can be specified by MPLS designers. Instead of simply routing customer traffic to Fargo based on the IGP metric shortest path, designers can route customer traffic in a way that meets customer demands for bandwidth, redundancy and service levels.



Topology Information Distribution

- IGP extensions propagate additional information
  - IS-IS uses TLV tuples
  - OSPF uses opaque LSA Type 10
  - Information propagated within area or level only
- Information propagated:
  - Maximum bandwidth
  - Reserved bandwidth
  - Available bandwidth
  - Administrative Groups (link colors)
  - Traffic engineering metric

Worldwide Education Services

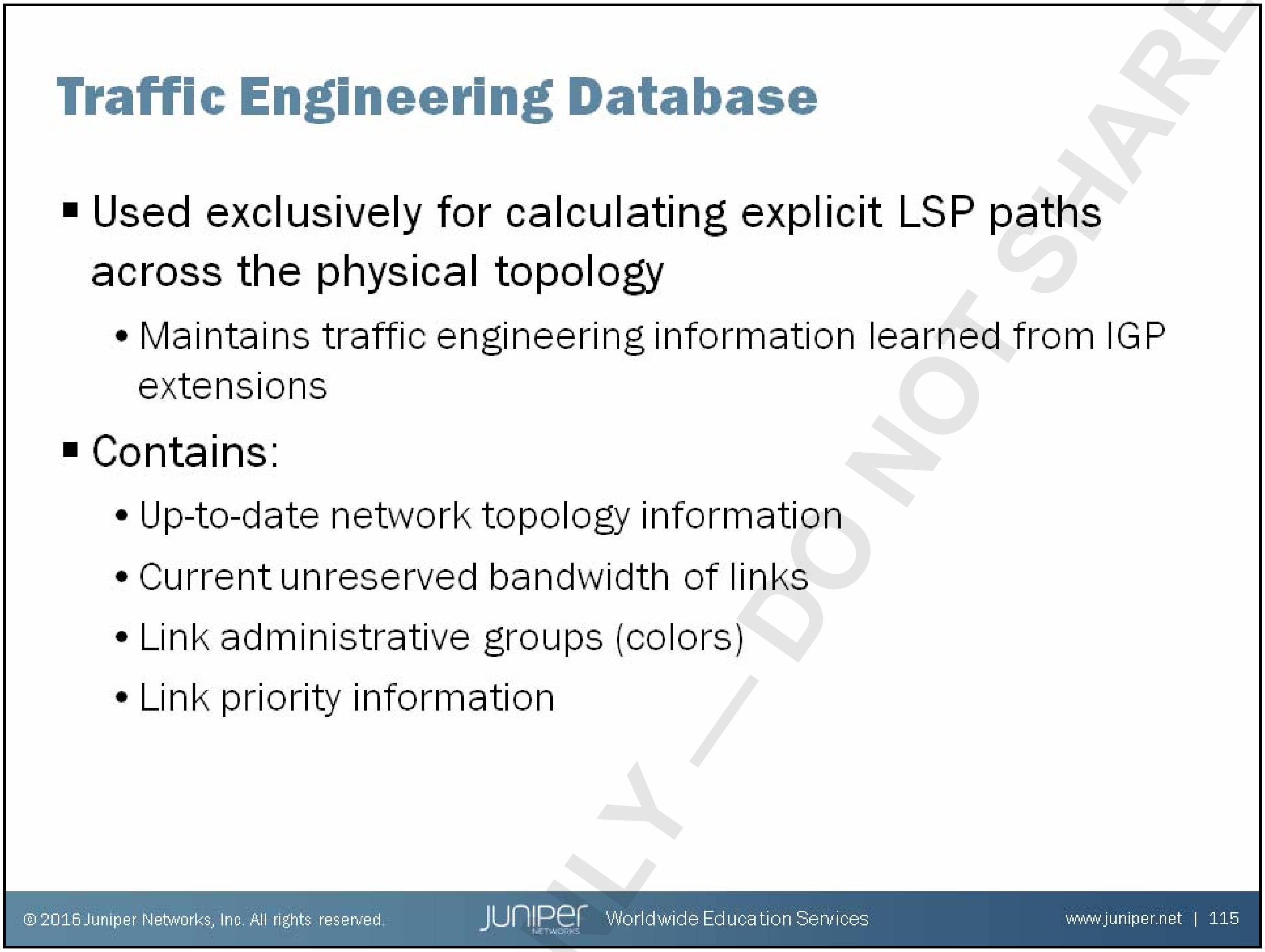
www.juniper.net | 114

IGP Extensions

Both OSPF and IS-IS can propagate additional information through some form of extension. IS-IS carries different parameters in type/length/value (TLV) tuples, which are propagated within a level; these TLVs do not propagate between levels. OSPF, on the other hand, uses Type 10 opaque LSAs to carry traffic engineering extensions. Type 10 LSAs have an area flooding scope, meaning that the information is propagated within a given area only; OSPF traffic engineering extensions do not cross area border routers (ABRs). The MPLS Traffic Engineering Information carried by these IGP extensions is defined in RFCs 3630 and 4203 for OSPF, and RFCs 3784 and 4205 for IS-IS.

Information Propagated

Maximum, available, and reserved bandwidth as well as link administrative group and traffic engineering specific metrics are communicated by IGPs. This data is stored on each IGP node in a traffic engineering database.



Traffic Engineering Database

- Used exclusively for calculating explicit LSP paths across the physical topology
  - Maintains traffic engineering information learned from IGP extensions
- Contains:
  - Up-to-date network topology information
  - Current unreserved bandwidth of links
  - Link administrative groups (colors)
  - Link priority information

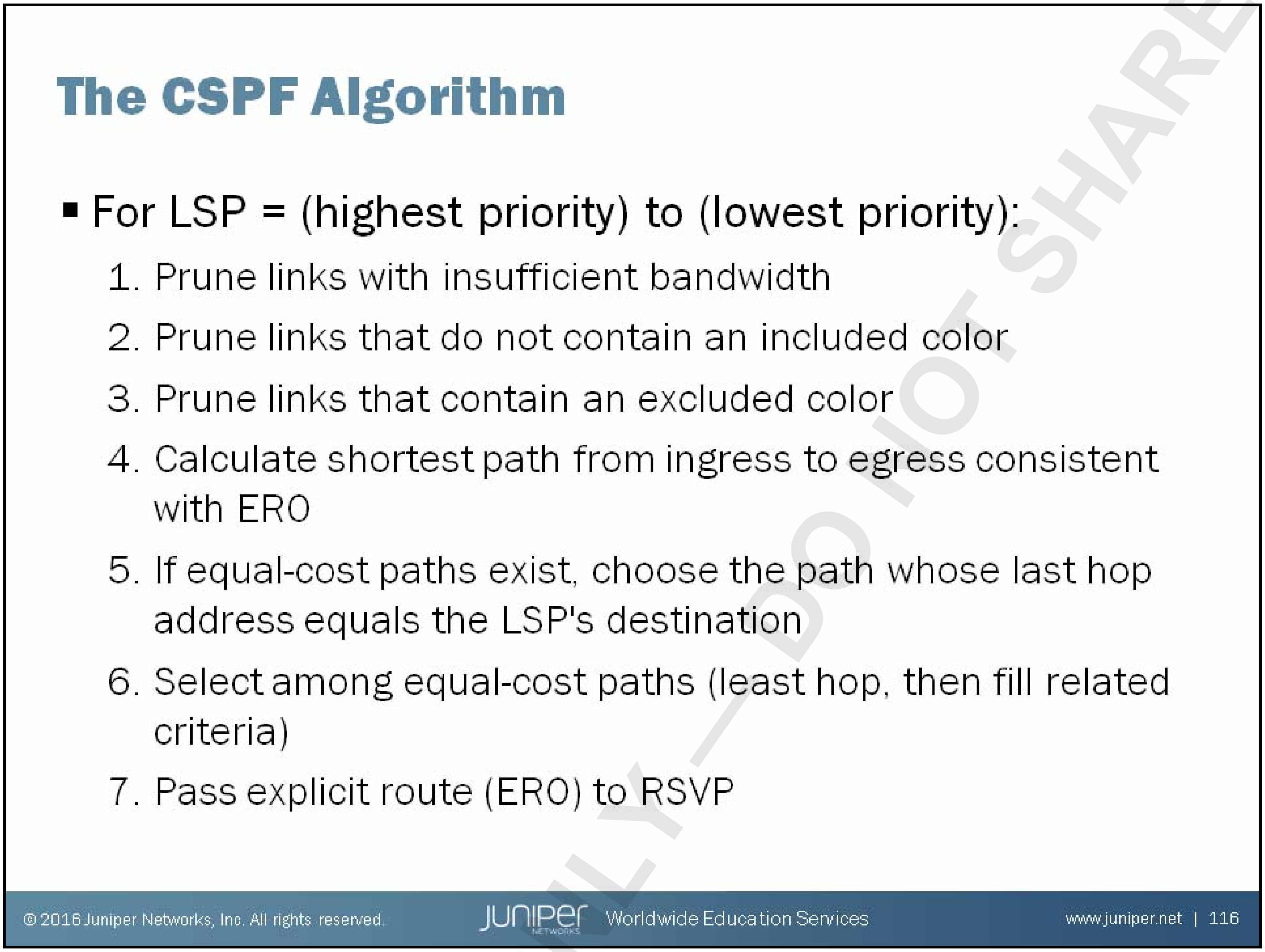
© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 115

Used Exclusively for LSP Path Computation

Each router maintains network link attributes and topology information in its TED. The TED is used exclusively for calculating explicit paths for the placement of LSPs across the physical topology. Because the TED does not know about existing LSPs, the TED does not allow a CSPF LSP to form over an LSP (because a non-CSPF LSP consults the routing table on a hop-by-hop basis to forward the RSVP messages, a non-CSPF LSP might try to form over an existing LSP) if features like forwarding adjacencies or traffic engineering shortcuts are enabled.

TED Contents

CSPF uses the TED to calculate explicit paths across the physical topology. It is similar to IGP link-state database (LSDB) and relies on extensions to the IGP, but it is stored independently of the IGP database. Traffic engineering requires detailed knowledge about the network topology as well as dynamic information about network loading. The information distribution component is implemented by defining relatively simple extensions to the IGPs so that link attributes are included as part of each router's link-state advertisement (LSA). The standard flooding algorithm used by the link-state IGPs ensures that link attributes are distributed to all routers in the routing domain. Some of the traffic engineering extensions to be added to the IGP link-state advertisement include maximum link bandwidth, maximum reserved link bandwidth, current bandwidth reservation levels, and link coloring.



The CSPF Algorithm

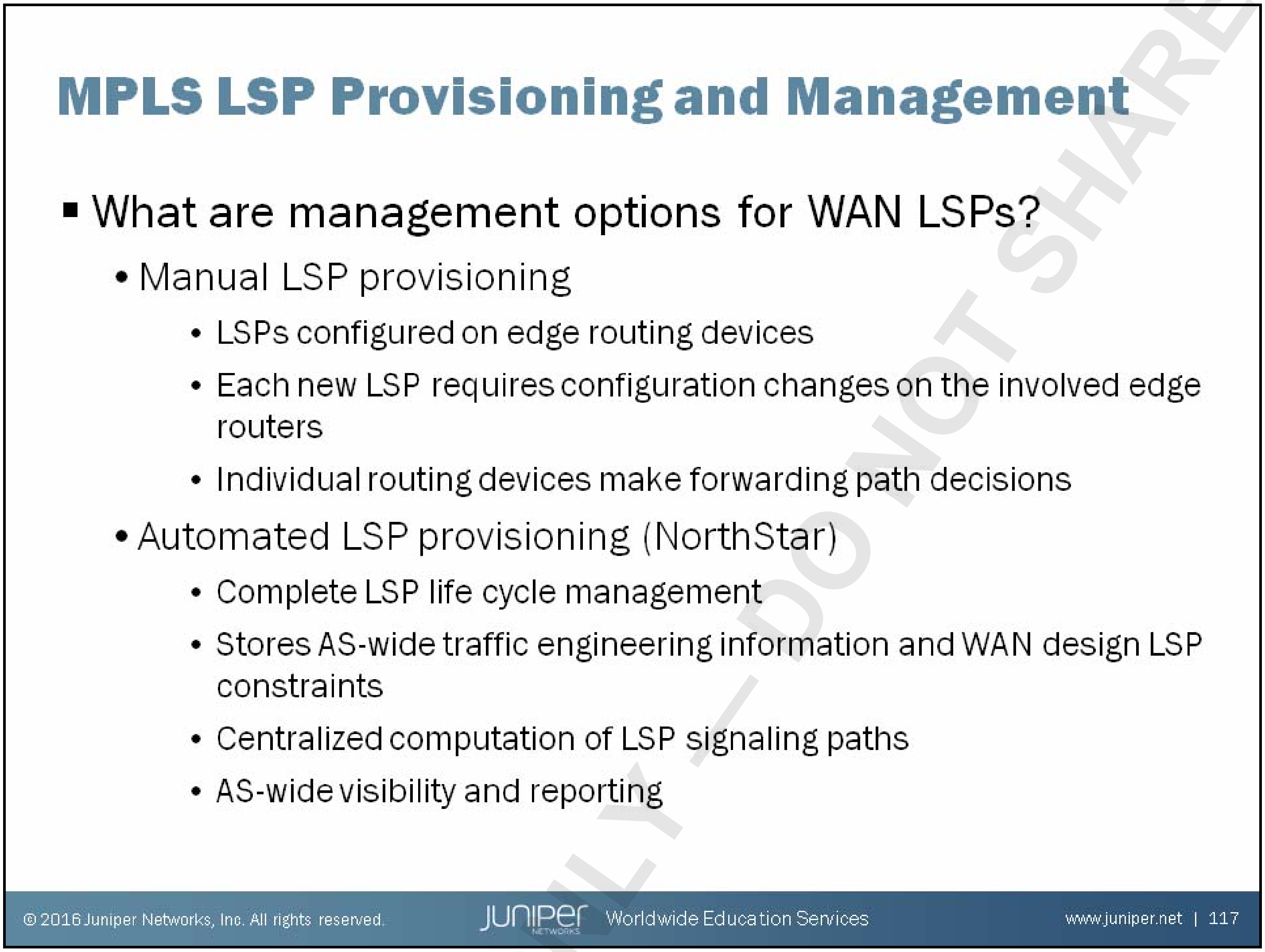
- For LSP = (highest priority) to (lowest priority):
  1. Prune links with insufficient bandwidth
  2. Prune links that do not contain an included color
  3. Prune links that contain an excluded color
  4. Calculate shortest path from ingress to egress consistent with ERO
  5. If equal-cost paths exist, choose the path whose last hop address equals the LSP's destination
  6. Select among equal-cost paths (least hop, then fill related criteria)
  7. Pass explicit route (ERO) to RSVP

JUNIPER NETWORKS Worldwide Education Services www.juniper.net | 116

How CSPF Selects a Path

The CSPF algorithm computes the path of LSPs one at a time, beginning with the highest-priority LSP (the one with the numerically lowest setup priority value). We cover LSP priority settings in the next chapter. Among LSPs of equal priority, CSPF begins with those that have the highest bandwidth requirement. For each such LSP, the following sequence is executed:

1. Prune the topology database (TED) of all the links that are not full duplex and do not have sufficient reservable bandwidth.
2. If the LSP configuration contains an include-any statement, prune all links that do not have at least one of the included colors assigned, including those links with no color assigned. If the LSP configuration contains an include-all statement, prune all links that do not have all of the included colors assigned.
3. If the LSP configuration contains an exclude statement, prune all links that contain excluded colors; links with no color are not pruned.
4. Find the shortest path towards the LSP's egress router, taking into account ERO constraints. For example, if the path must pass through Router A, two separate SPFs are computed, one from the ingress router to Router A, the other from Router A to the egress router.
5. If several paths have equal cost, choose the one whose last hop address is the same as the LSP's destination.
6. If several equal-cost paths remain, select the one with the fewest number of hops. If equal-cost paths still remain, apply the CSPF load-balancing rule configured on the LSP (least fill, most fill, or random).
7. When a path is chosen, pass the complete ERO list to RSVP for signaling.



MPLS LSP Provisioning and Management

- What are management options for WAN LSPs?
  - Manual LSP provisioning
    - LSPs configured on edge routing devices
    - Each new LSP requires configuration changes on the involved edge routers
    - Individual routing devices make forwarding path decisions
  - Automated LSP provisioning (NorthStar)
    - Complete LSP life cycle management
    - Stores AS-wide traffic engineering information and WAN design LSP constraints
    - Centralized computation of LSP signaling paths
    - AS-wide visibility and reporting

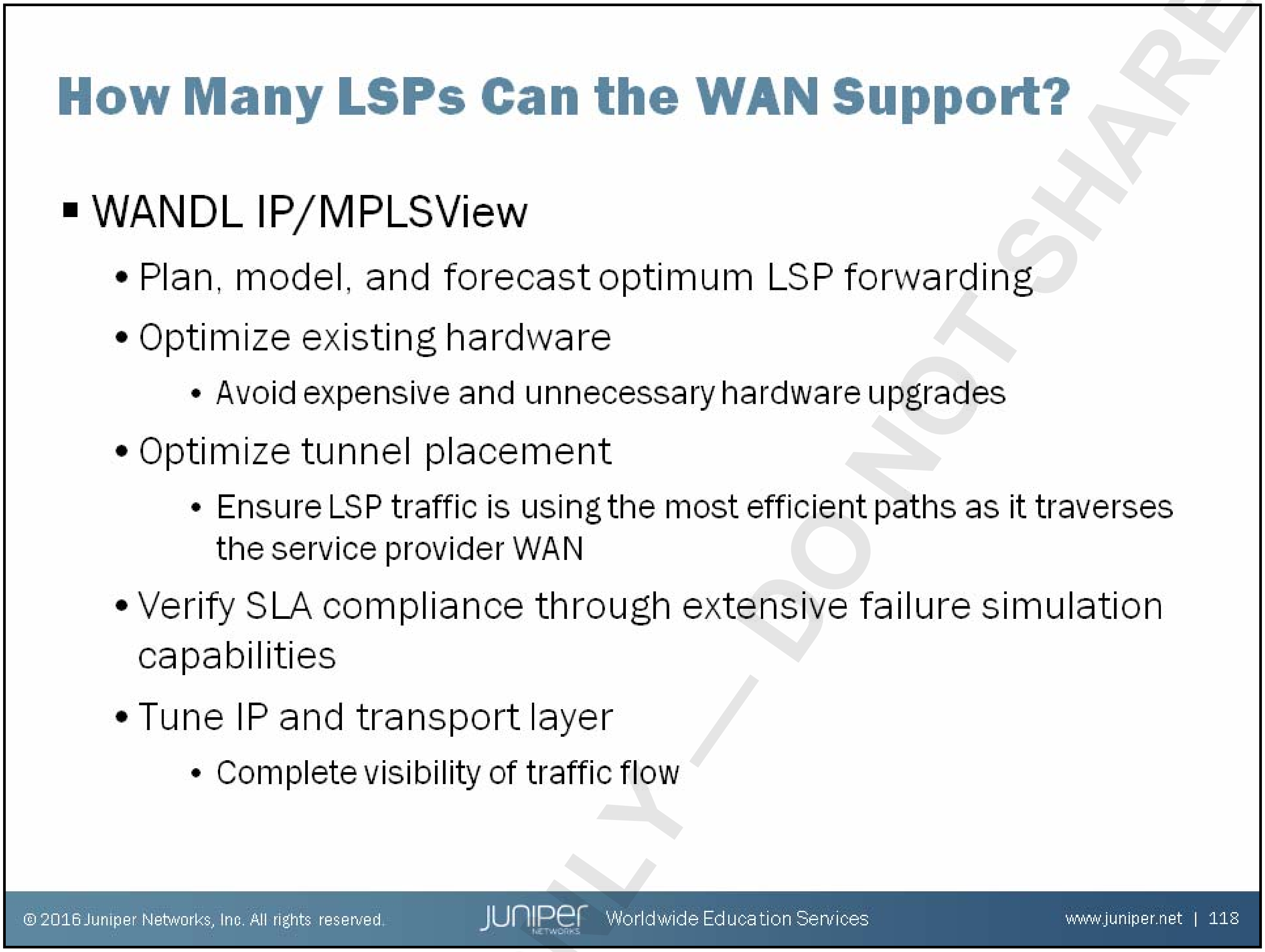
Worldwide Education Services

www.juniper.net | 117

MPLS LSP Provisioning and Management

RSVP signaled LSPs must be provisioned on the LSP ingress router and signaled across the WAN to the egress LSR. LSPs can be manually provisioned on the ingress router. The ingress router uses its local routing tables, traffic engineering database, and takes into account any administrator defined routing constraints to perform the CSPF algorithm and determine the LSP signaling path.

Software Defined Networking (SDN) can automate the provisioning process, eliminating the need to manually provision and manage LSPs. SDN removes inconsistency and adds scalability to the LSP design process. A centralized SDN controller stores a copy of the WAN traffic engineering database and all LSP definitions and constraints. The SDN controller communicates dynamically with the ingress routers information about any new LSPs causing the ingress router to begin the LSP signaling process. NorthStar is Juniper Network’s implementation of a centralized SDN controller that can manage the complete LSP life cycle. NorthStar provides granular visibility into, and control over, IP/MPLS flows in large service provider and enterprise networks. Using NorthStar Controller, operators can optimize their network infrastructure through proactive monitoring and planning, and dynamically create explicit routing paths using a global view that’s based on user-defined constraints.



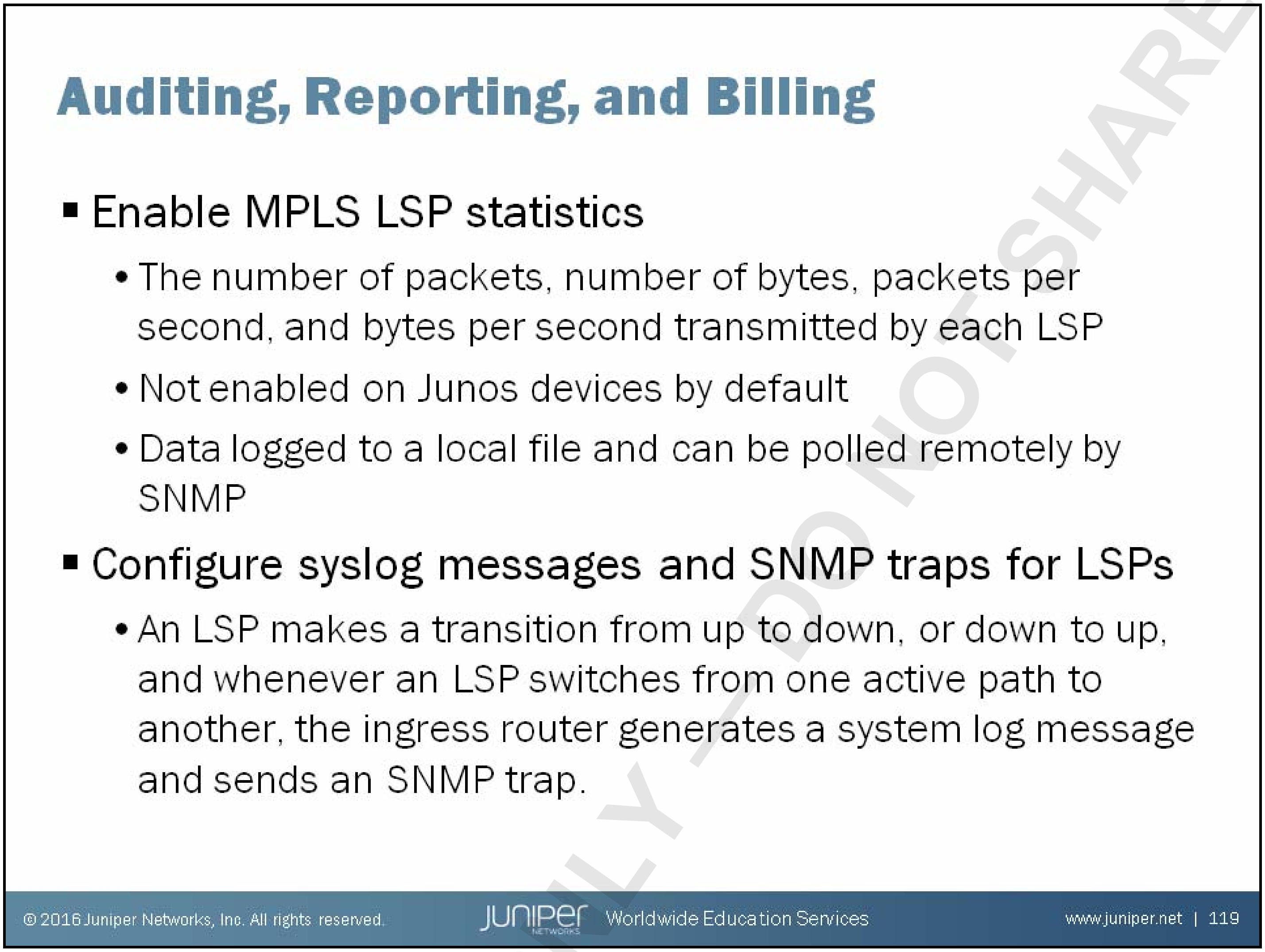
How Many LSPs Can the WAN Support?

- WANDL IP/MPLSView
  - Plan, model, and forecast optimum LSP forwarding
  - Optimize existing hardware
    - Avoid expensive and unnecessary hardware upgrades
  - Optimize tunnel placement
    - Ensure LSP traffic is using the most efficient paths as it traverses the service provider WAN
  - Verify SLA compliance through extensive failure simulation capabilities
  - Tune IP and transport layer
    - Complete visibility of traffic flow

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 118

WANDL IP/MPLSView

WANDL IP/MPLSView is a Juniper Network's product that builds a model of the WAN network and uses this model to perform traffic load analysis and capacity planning. IP/MPLSView provides insights as to why traffic flows or tunnels fail to route and helps identify which trunks become congested under certain failures or what-if scenarios. With IP/MPLSView, users can study the impact of extensive node, link, site, card, and Shared Risk Link Group (SRLG) multilayer failure scenarios. They can also analyze the way traffic is rerouted and the effect on network links (e.g., worst-case trunk utilization), and even perform exhaustive single, double, and triple failure tests. IP/MPLSView automatically determines where to add links to satisfy traffic for resiliency against any failure scenario. IP/MPLSView allows WAN designers to validate day-to-day network changes, or model and simulate network migration, network expansion, or the merging of multiple networks. This provides the ability to analyze the impact of these changes in a safe, virtual environment, and also experiment with changing parameters, protocols, topology, and so on. IP/MPLSView helps designers build effective designs that result in lower hardware and maintenance costs.



Auditing, Reporting, and Billing

- Enable MPLS LSP statistics
  - The number of packets, number of bytes, packets per second, and bytes per second transmitted by each LSP
  - Not enabled on Junos devices by default
  - Data logged to a local file and can be polled remotely by SNMP
- Configure syslog messages and SNMP traps for LSPs
  - An LSP makes a transition from up to down, or down to up, and whenever an LSP switches from one active path to another, the ingress router generates a system log message and sends an SNMP trap.

Juniper Worldwide Education Services

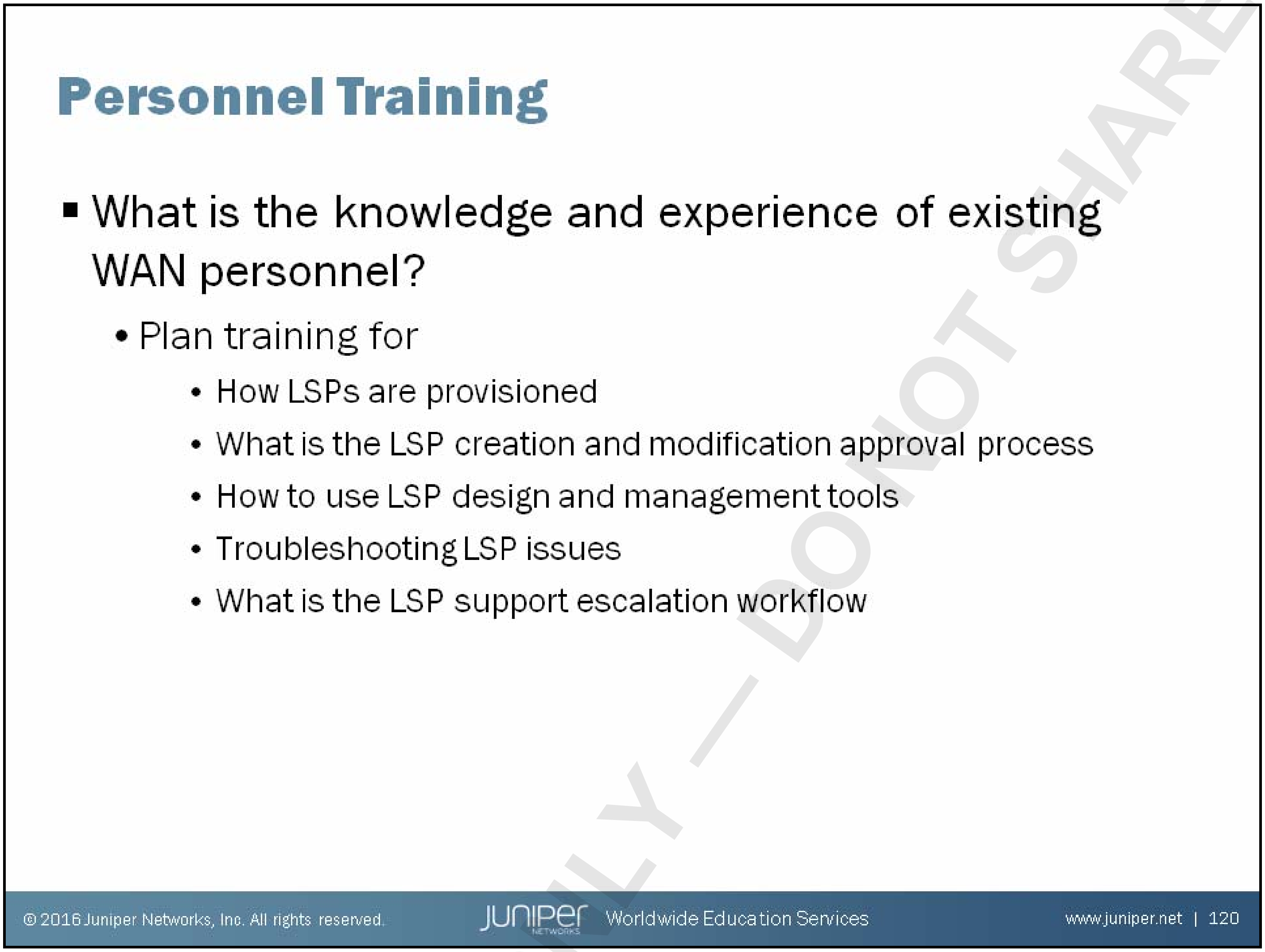
www.juniper.net | 119

Enable MPLS LSP Statistics

Junos devices can generate MPLS statistical information about the LSPs it manages. Information such as the number of packets, number of bytes, packets per second, and bytes per second transmitted can be stored in files on the router and polled by SNMP. MPLS statistics are not generated by default but can be enabled through configuration.

Configure Syslog and SNMP

SNMP traps can be generated when LSP state changes occur. Ingress routers use SNMP traps to notify WAN administrators of up/down status changes and any path changes that occur. Syslog messages also give WAN administrators information about the status of LSP tunnels.



Personnel Training

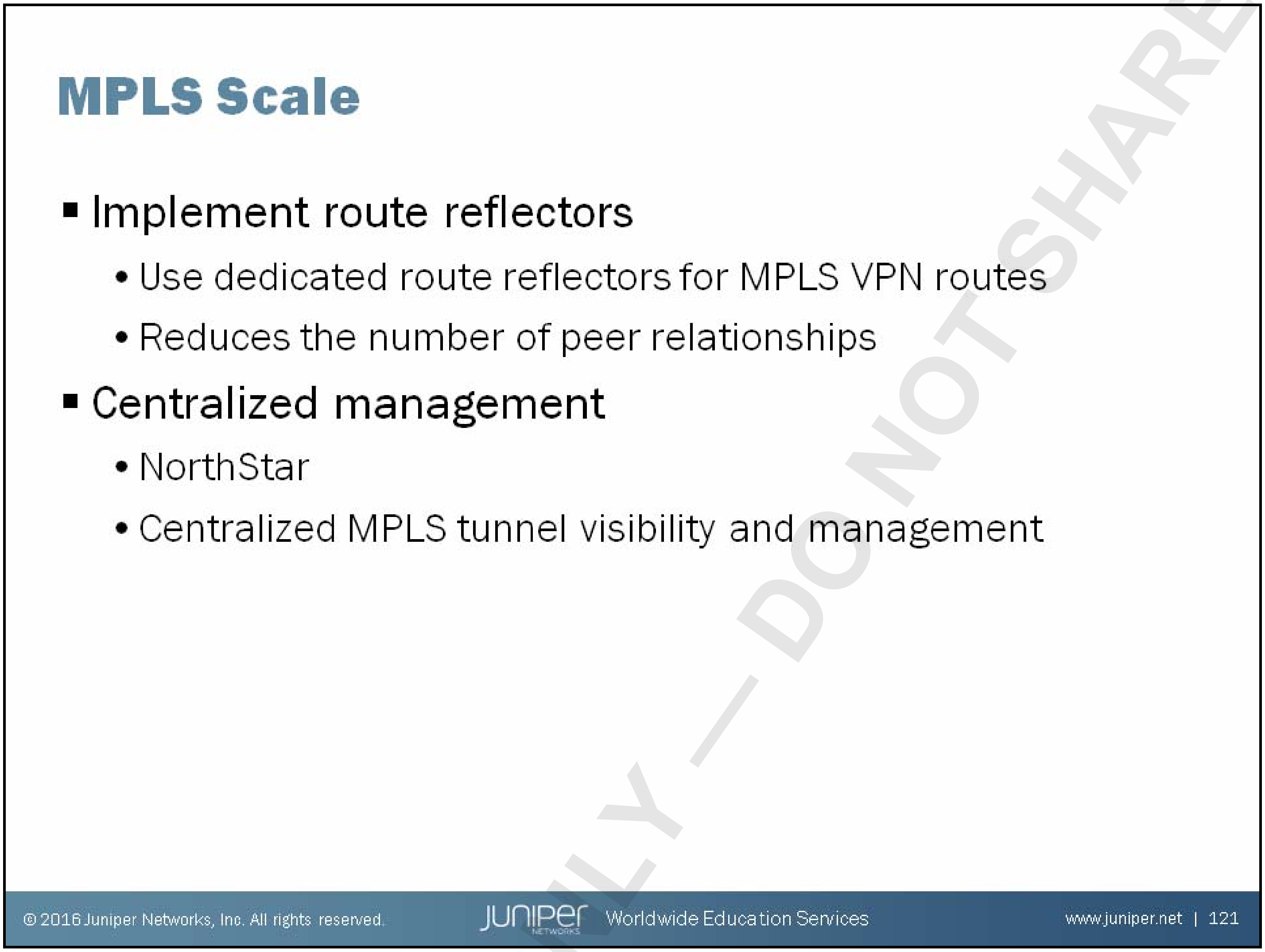
- What is the knowledge and experience of existing WAN personnel?
  - Plan training for
    - How LSPs are provisioned
    - What is the LSP creation and modification approval process
    - How to use LSP design and management tools
    - Troubleshooting LSP issues
    - What is the LSP support escalation workflow

JUNIPER NETWORKS Worldwide Education Services

www.juniper.net | 120

Personnel Up Training

If MPLS is a new service offering for your WAN the skill set of the current network management team should be evaluated. New processes and responsibilities will be required for a successful implementation of MPLS. How will these new LSPs be provisioned? What is the approval process for new tunnels and who is responsible for support? What are common troubleshooting processes necessary to reduce time to recovery? These are some of the topics day to day WAN administrators and support staff will need to be trained on. Determining how many people should be trained is a key decision.



MPLS Scale

- Implement route reflectors
  - Use dedicated route reflectors for MPLS VPN routes
  - Reduces the number of peer relationships
- Centralized management
  - NorthStar
  - Centralized MPLS tunnel visibility and management

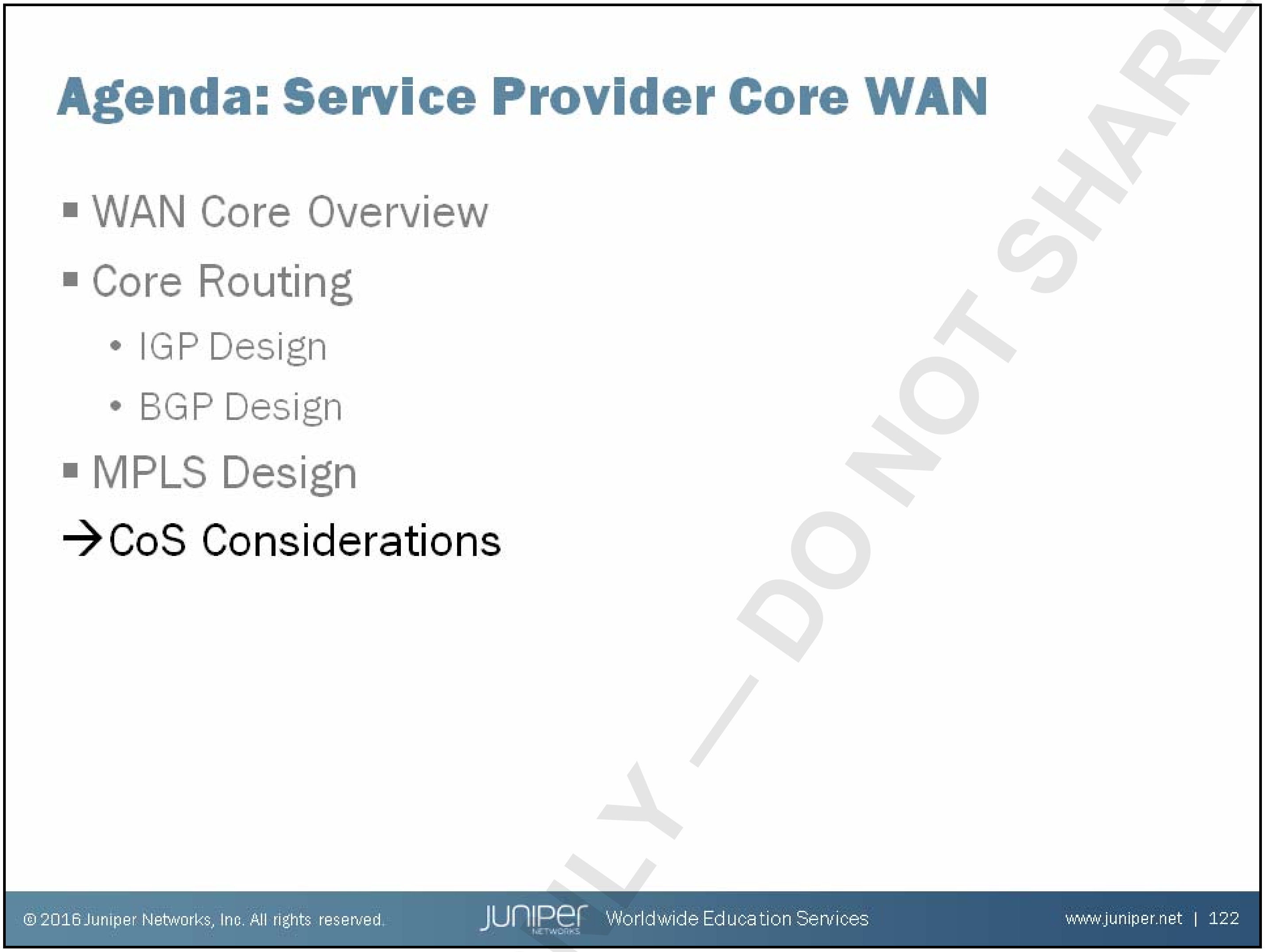
Worldwide Education Services
www.juniper.net | 121

MPLS Scale

Large BGP implementations take advantage of the scalability provided by route reflection. Requiring edge routers to create a full mesh of adjacencies to other edge routers reduces the scale of the MPLS design.

Centralized Management

Large MPLS WAN environments benefit from the scale, consistency, and visibility centralized management provides. Automation is the present and future of network design. Juniper's NorthStar Controller stores AS wide traffic engineering information and WAN design LSP constraints providing centralized computation of LSP signaling paths.



Agenda: Service Provider Core WAN

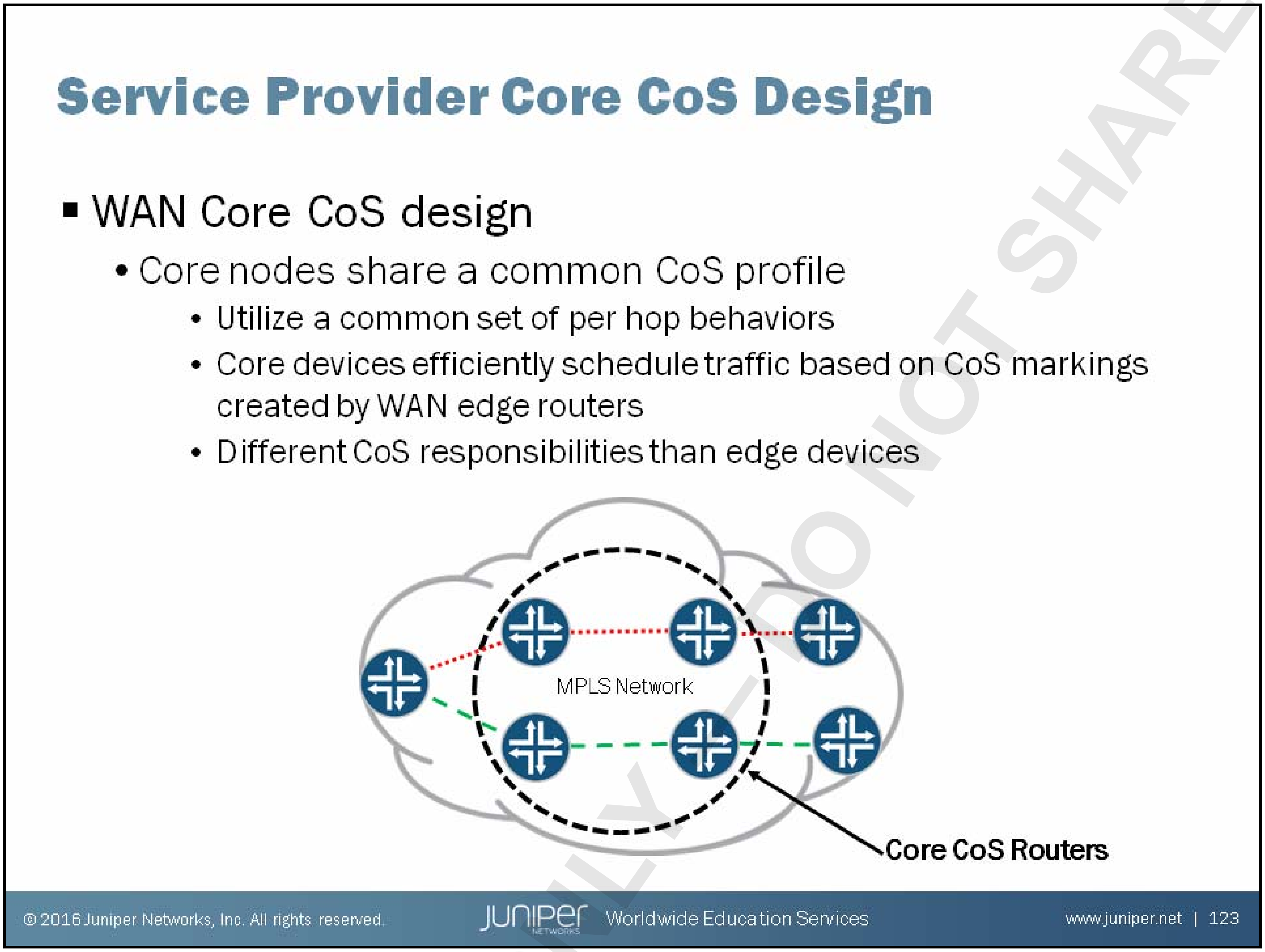
- WAN Core Overview
- Core Routing
  - IGP Design
  - BGP Design
- MPLS Design

→CoS Considerations

© 2016 Juniper Networks, Inc. All rights reserved. Juniper Networks Worldwide Education Services www.juniper.net | 122

CoS Considerations

The slide highlights the topic we discuss next.



Service Provider Core CoS Design

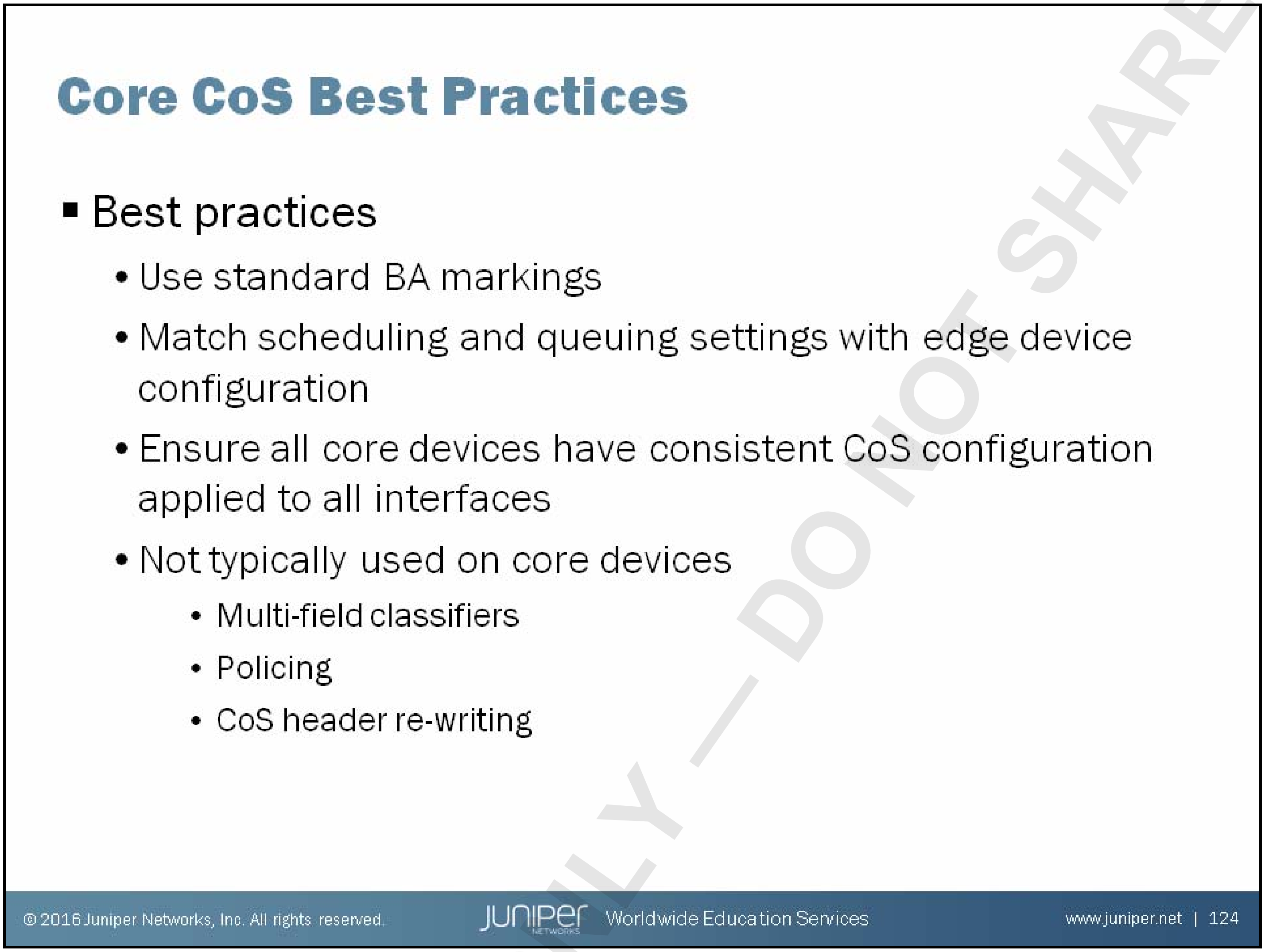
- WAN Core CoS design
  - Core nodes share a common CoS profile
    - Utilize a common set of per hop behaviors
    - Core devices efficiently schedule traffic based on CoS markings created by WAN edge routers
    - Different CoS responsibilities than edge devices

MPLS Network

Core CoS Routers

WAN Core CoS Design

Core networking devices have different CoS responsibilities than edge devices. Traffic is multi-field classified and marked at the network edge and forwarded to the network core. Core routers use the CoS markings in traffic headers to perform behavior aggregate classification and queue and schedule the data. Core routers should be configured to use identical CoS parameters for queuing and scheduling traffic to guarantee consistent service across the WAN. Different hardware models and vendors may be used for edge and core networking devices. Verify CoS feature compatibility between different models and vendors before beginning the CoS design.



Core CoS Best Practices

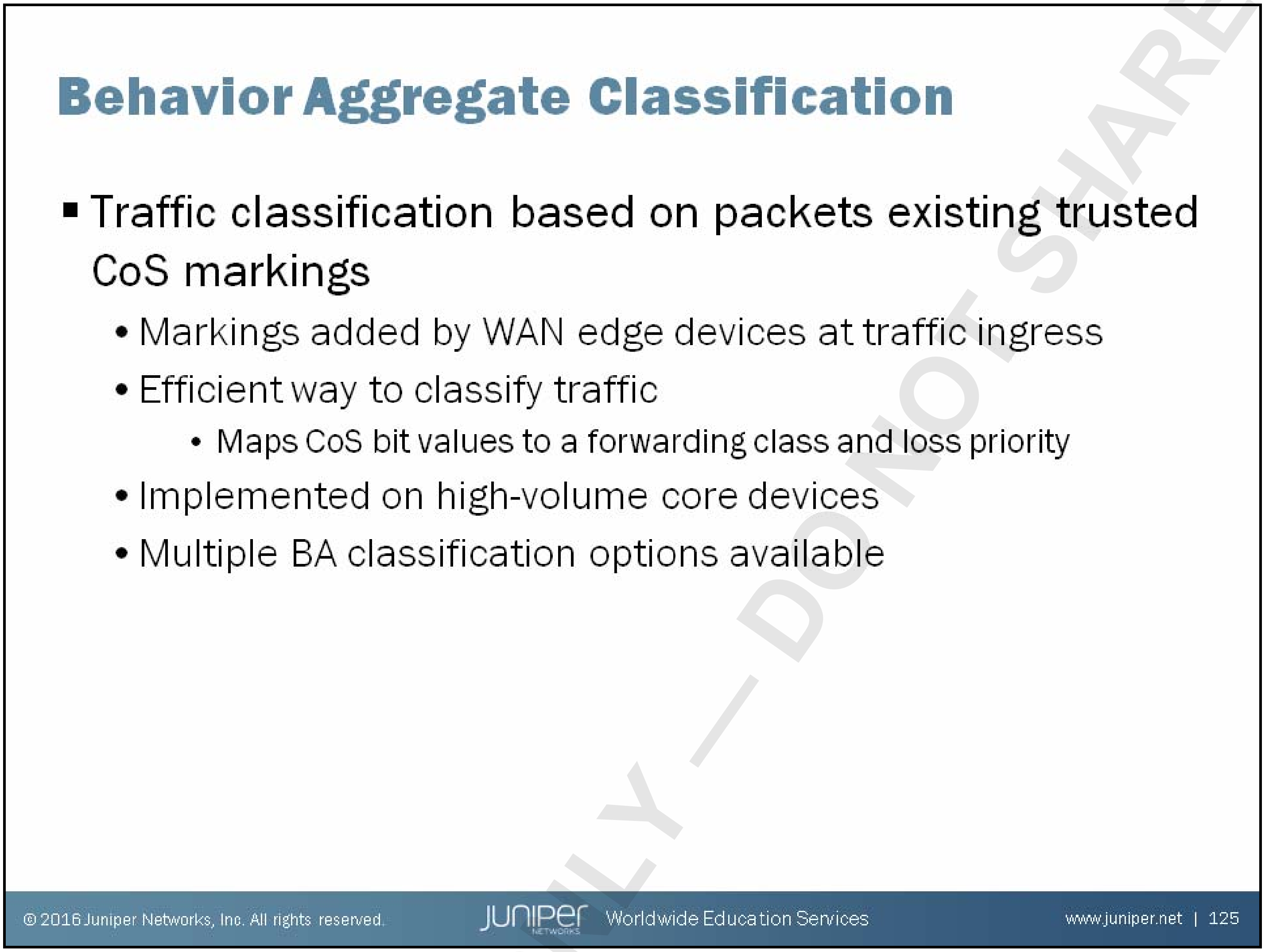
- Best practices
  - Use standard BA markings
  - Match scheduling and queuing settings with edge device configuration
  - Ensure all core devices have consistent CoS configuration applied to all interfaces
  - Not typically used on core devices
    - Multi-field classifiers
    - Policing
    - CoS header re-writing

Worldwide Education Services

www.juniper.net | 124

Core CoS Best Practices

Several RFCs define standard traffic header locations and marking values that should be supported by vendors of CoS capable networking equipment. A full mesh of LSPs are generally configured across the WAN core and traffic can be forwarded across the core using different paths depending on network conditions. Compliance assessment is necessary to verify that CoS configuration is consistent from edge to edge across the WAN core. Verify that all core networking devices are using behavior aggregate classification and are configured with the same scheduling and queuing values defined on WAN edge routers. WAN core routers typically do not use multi-field classifiers to classify traffic or perform traffic policing. These functions are performed by WAN edge routing devices as traffic enters the service provider network.



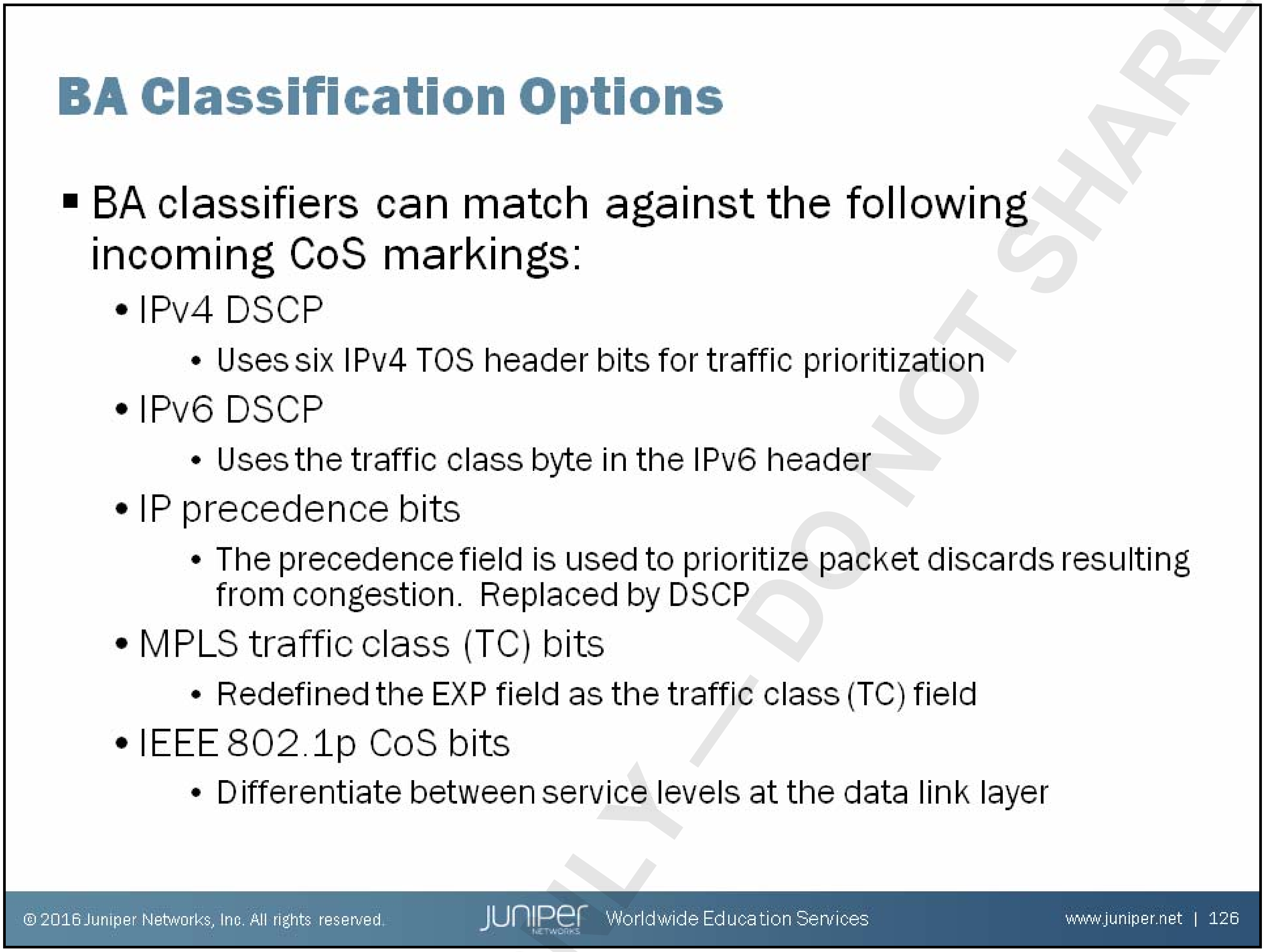
Behavior Aggregate Classification

- Traffic classification based on packets existing trusted CoS markings
  - Markings added by WAN edge devices at traffic ingress
  - Efficient way to classify traffic
    - Maps CoS bit values to a forwarding class and loss priority
  - Implemented on high-volume core devices
  - Multiple BA classification options available

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 125

Behavior Aggregate Classification—CoS Marking-Based Method

When traffic arrives from a neighboring node with CoS markings present, the router can perform behavior aggregate (BA) classification. BA classification can be applied per logical interface or virtual LAN (VLAN), and it provides a simple way to directly map a marked packet to a forwarding class and packet loss priority (PLP) value. Forwarding classes are associated with queues on the outbound interface and the queues scheduling properties are evaluated to determine when the packet is forwarded. BA classification is more efficient than multifield (MF) classification, because it requires less packet analysis. The efficiency benefit makes BA classification a good choice for devices with high traffic volumes, such as routers in a network core. BA classification is based entirely on existing CoS markings. It treats all traffic with a given CoS value in the same way.



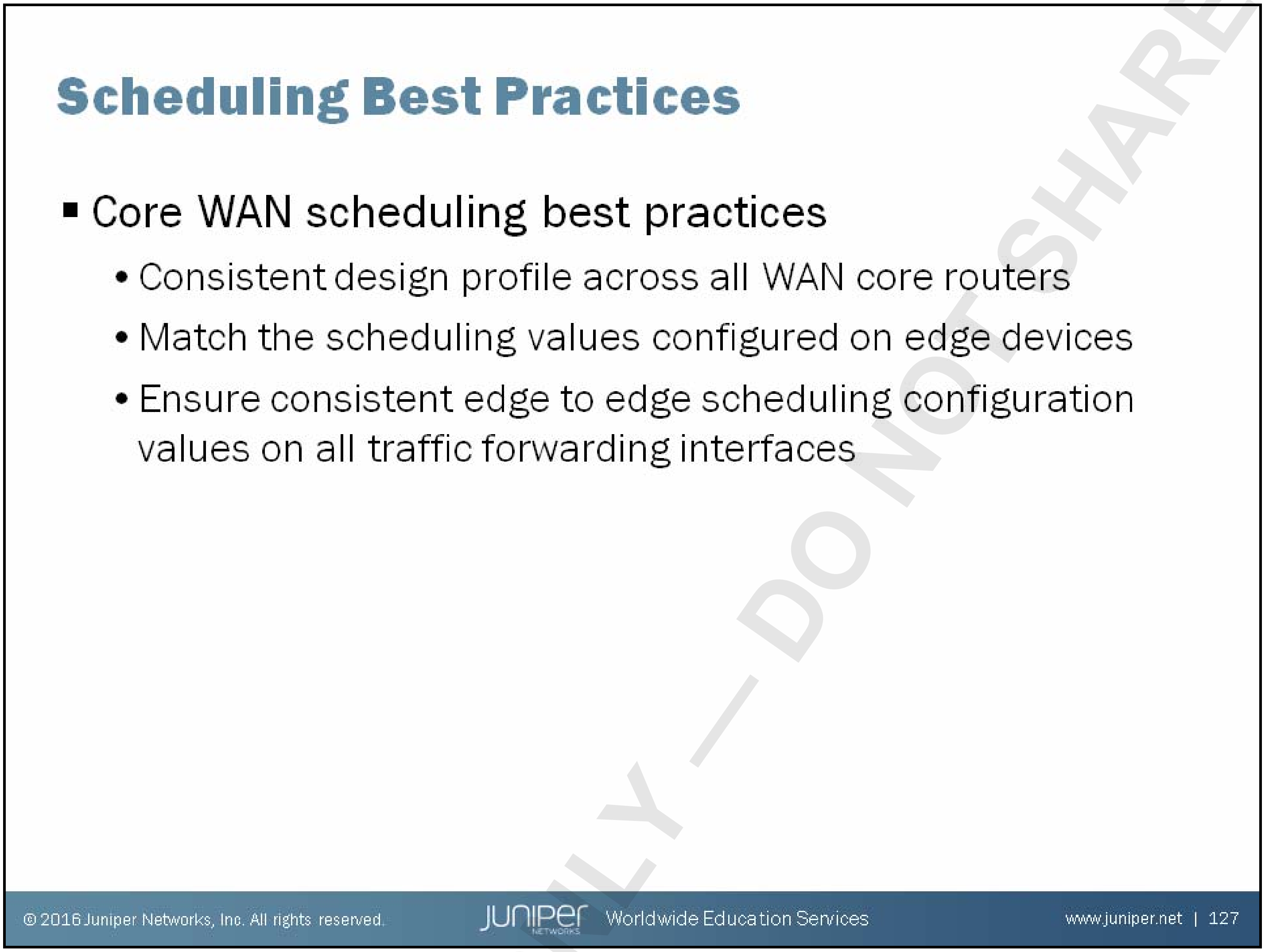
BA Classification Options

- BA classifiers can match against the following incoming CoS markings:
  - IPv4 DSCP
    - Uses six IPv4 TOS header bits for traffic prioritization
  - IPv6 DSCP
    - Uses the traffic class byte in the IPv6 header
  - IP precedence bits
    - The precedence field is used to prioritize packet discards resulting from congestion. Replaced by DSCP
  - MPLS traffic class (TC) bits
    - Redefined the EXP field as the traffic class (TC) field
  - IEEE 802.1p CoS bits
    - Differentiate between service levels at the data link layer

BA Classification Options

Different options are available for WAN designers interested in behavior aggregate classification. Several RFCs have been defined that allow CoS traffic markings to be placed in the headers of many different protocols and traffic types. Choose the appropriate methods to match your WAN traffic types. IP ToS/DiffServ Differentiated Services Code Point (DSCP) occupies the type of service (ToS) byte, using the first six bits for traffic prioritization. The DSCP field supersedes the ToS field, but provides backward compatibility. DSCP is implemented similarly in IPv4 and IPv6.

- IP Precedence Bits: IP precedence is primarily used to prevent discards of network control packets.
- MPLS Traffic Class: The MPLS header has three bits that were originally defined as experimental (EXP). However, once MPLS came into common use, the field was designated for CoS purposes. In February 2009, RFC 5462 redefined the EXP field as the traffic class (TC) field, in support of the standardization of the usage of these bits across the industry. The MPLS EXP field occupies three bits, providing eight classes of service.
- IEEE802.1p: Three bits at the beginning of the Tag Control Information (TCI) field of the 802.1p header. These priority bits can be used to differentiate between service levels at the data link or media access control (MAC) layer. The 802.1p standard provides eight levels of priority that are defined in ways similar to those of the IP ToS field.



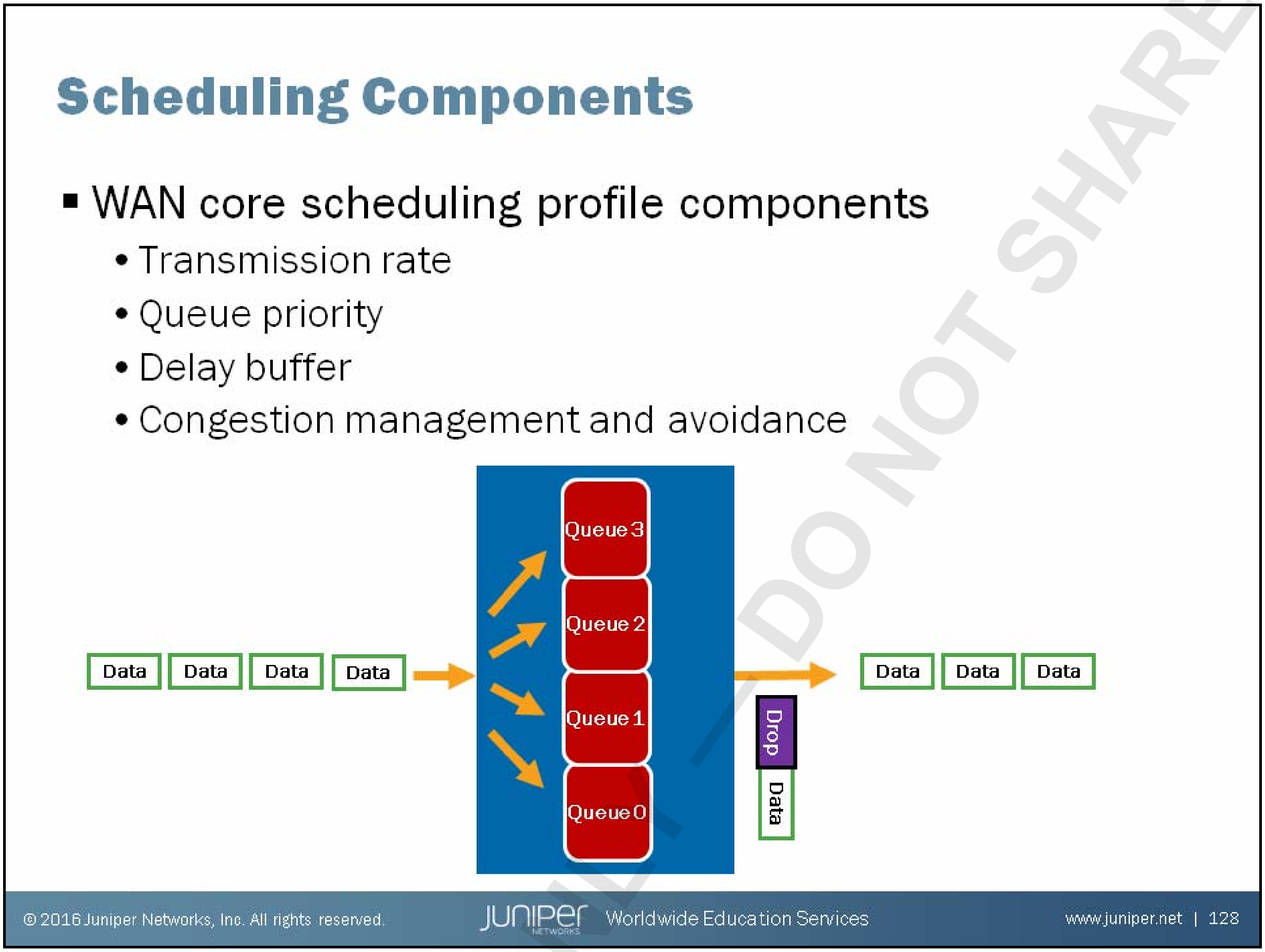
Scheduling Best Practices

- Core WAN scheduling best practices
  - Consistent design profile across all WAN core routers
  - Match the scheduling values configured on edge devices
  - Ensure consistent edge to edge scheduling configuration values on all traffic forwarding interfaces

© 2016 Juniper Networks, Inc. All rights reserved. Worldwide Education Services www.juniper.net | 127

Scheduling Best Practices

Consistency is key for high performance core CoS designs. Edge routers add CoS markings at traffic ingress and forward them towards core routing devices. The core routing devices CoS configuration for queuing and scheduling should match the settings configured on the edge routing devices. Verify all core routers are configured to treat CoS traffic identically for consistent WAN traffic forwarding.

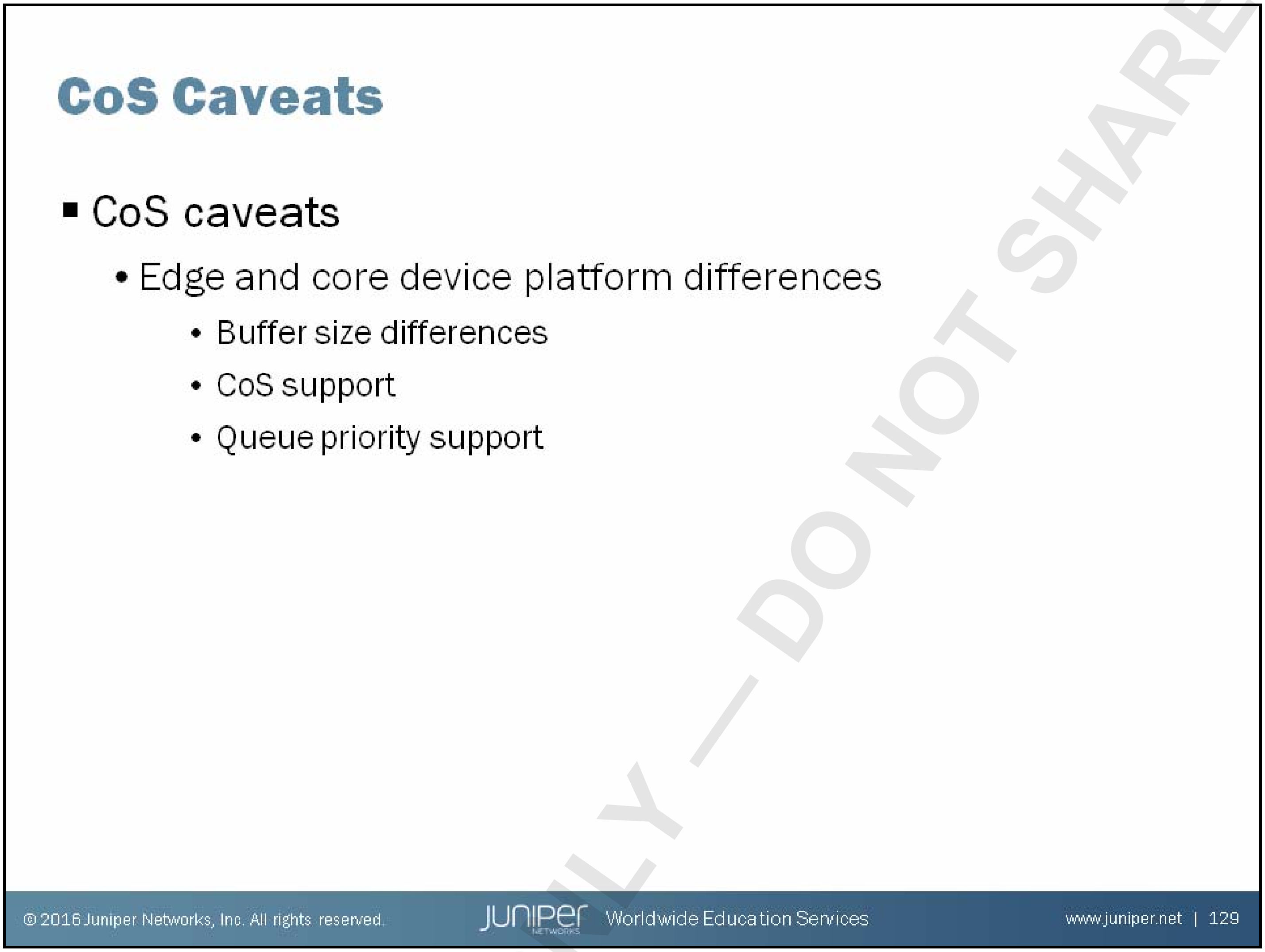


Scheduling Components

- WAN core scheduling profile components
  - Transmission rate
  - Queue priority
  - Delay buffer
  - Congestion management and avoidance

WAN Core Scheduling Components

Transmission rate, queue priority, delay buffer size, and congestion management are all components of scheduling outbound traffic. To guarantee consistent treatment of traffic from WAN edge to WAN edge these parameters core routers should be configured with the same values defined on the WAN edge devices. Audit changes to any CoS related configuration.



CoS Caveats

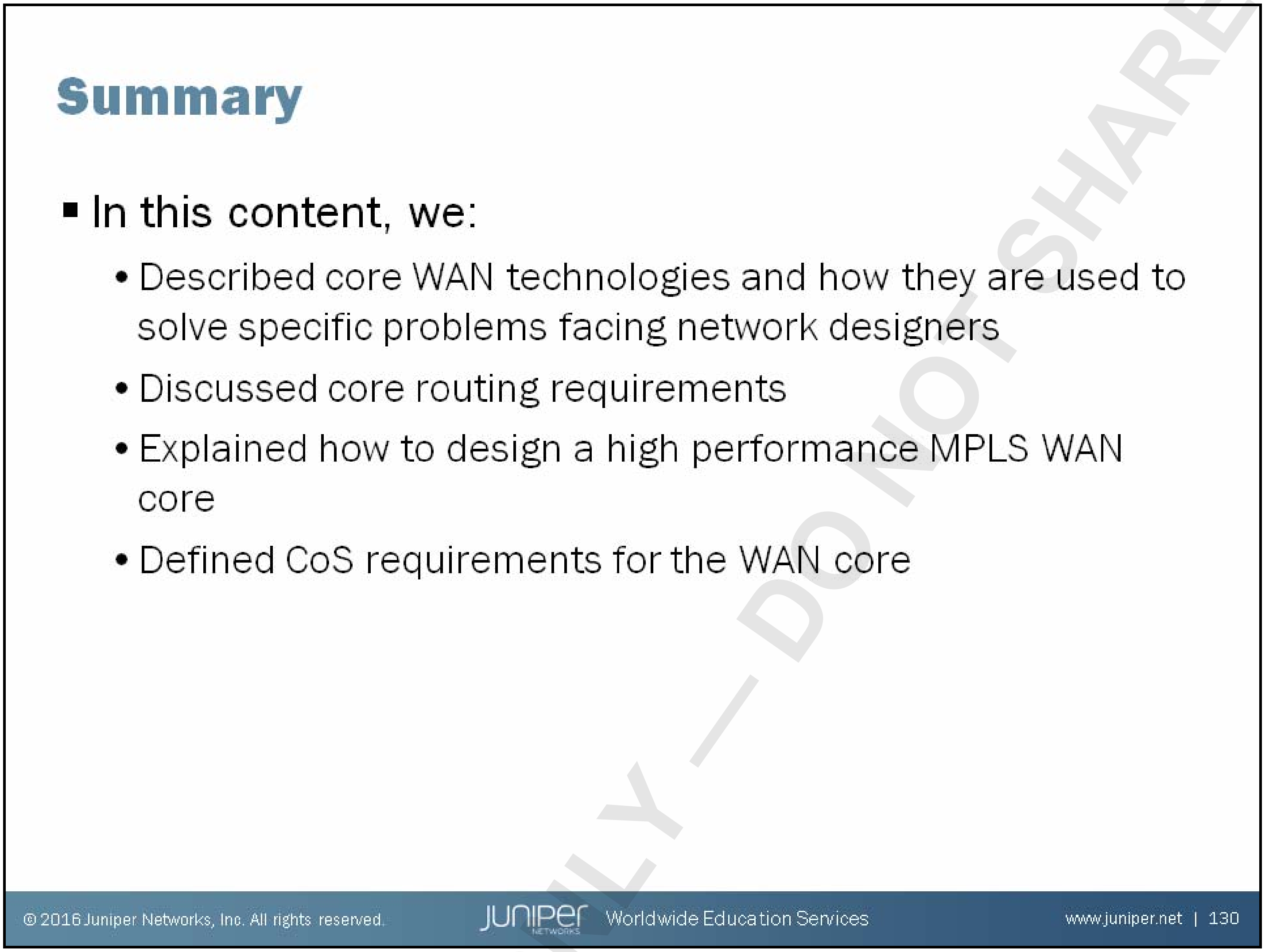
- CoS caveats
  - Edge and core device platform differences
    - Buffer size differences
    - CoS support
    - Queue priority support

Worldwide Education Services

www.juniper.net | 129

CoS Caveats

Platform differences that exist between devices that make up a large WAN environment present challenges to CoS designers. WAN edge devices initiate the CoS process at network ingress and are designed to provide different features and services than core WAN devices. Even routing products from the same vendor can implement different memory and processor architectures that affect the amount of device resources available for CoS. Consult WAN networking vendors to verify support for CoS capabilities.



Summary

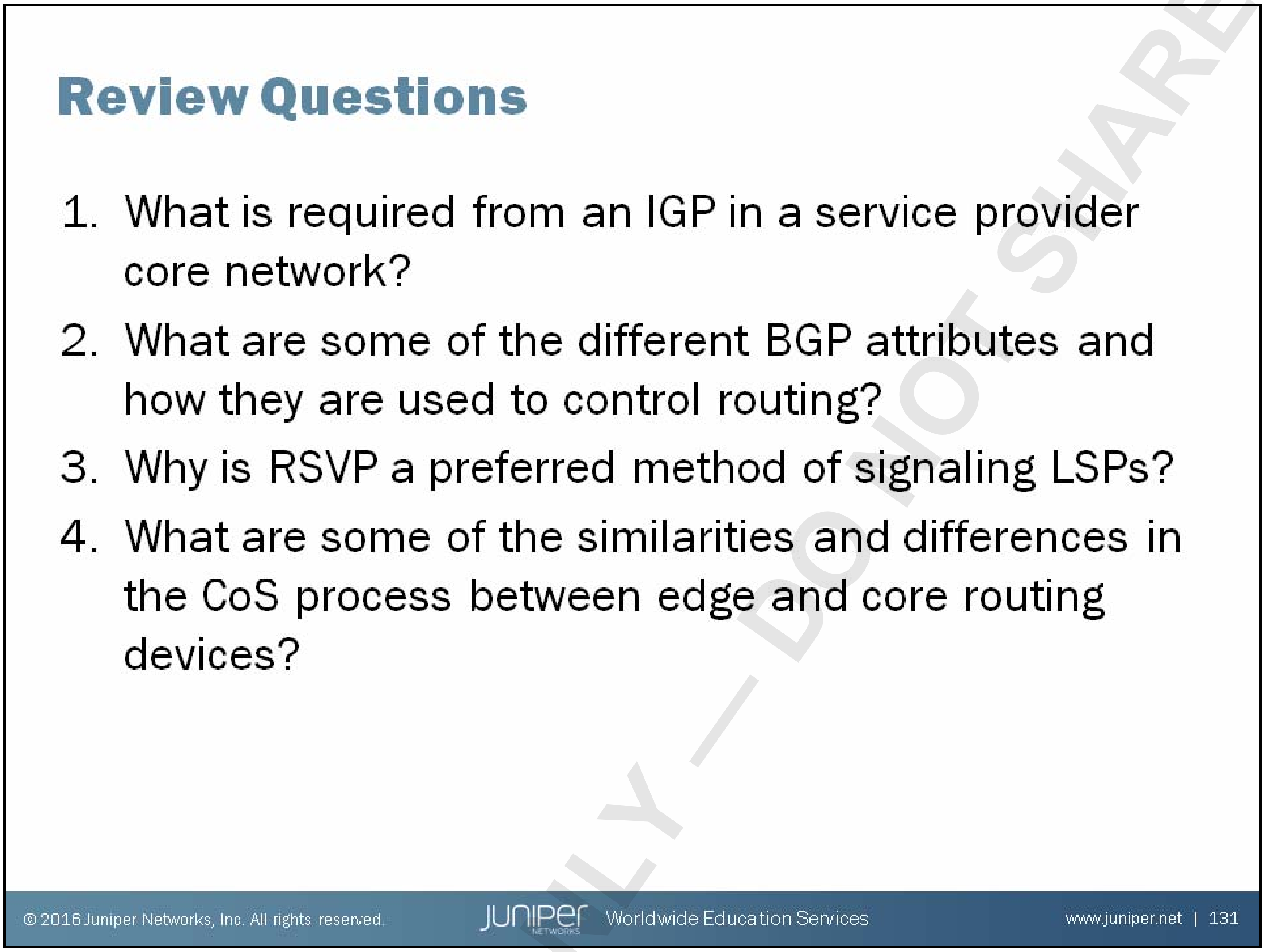
- In this content, we:
  - Described core WAN technologies and how they are used to solve specific problems facing network designers
  - Discussed core routing requirements
  - Explained how to design a high performance MPLS WAN core
  - Defined CoS requirements for the WAN core

Worldwide Education Services

www.juniper.net | 130

We Discussed:

- Core WAN technologies and how they are used to solve specific problems facing network designers;
- Core routing requirements;
- How to design a high performance MPLS WAN core; and
- CoS requirements for the WAN core.



Review Questions

1. What is required from an IGP in a service provider core network?
2. What are some of the different BGP attributes and how they are used to control routing?
3. Why is RSVP a preferred method of signaling LSPs?
4. What are some of the similarities and differences in the CoS process between edge and core routing devices?

Worldwide Education Services

www.juniper.net | 131

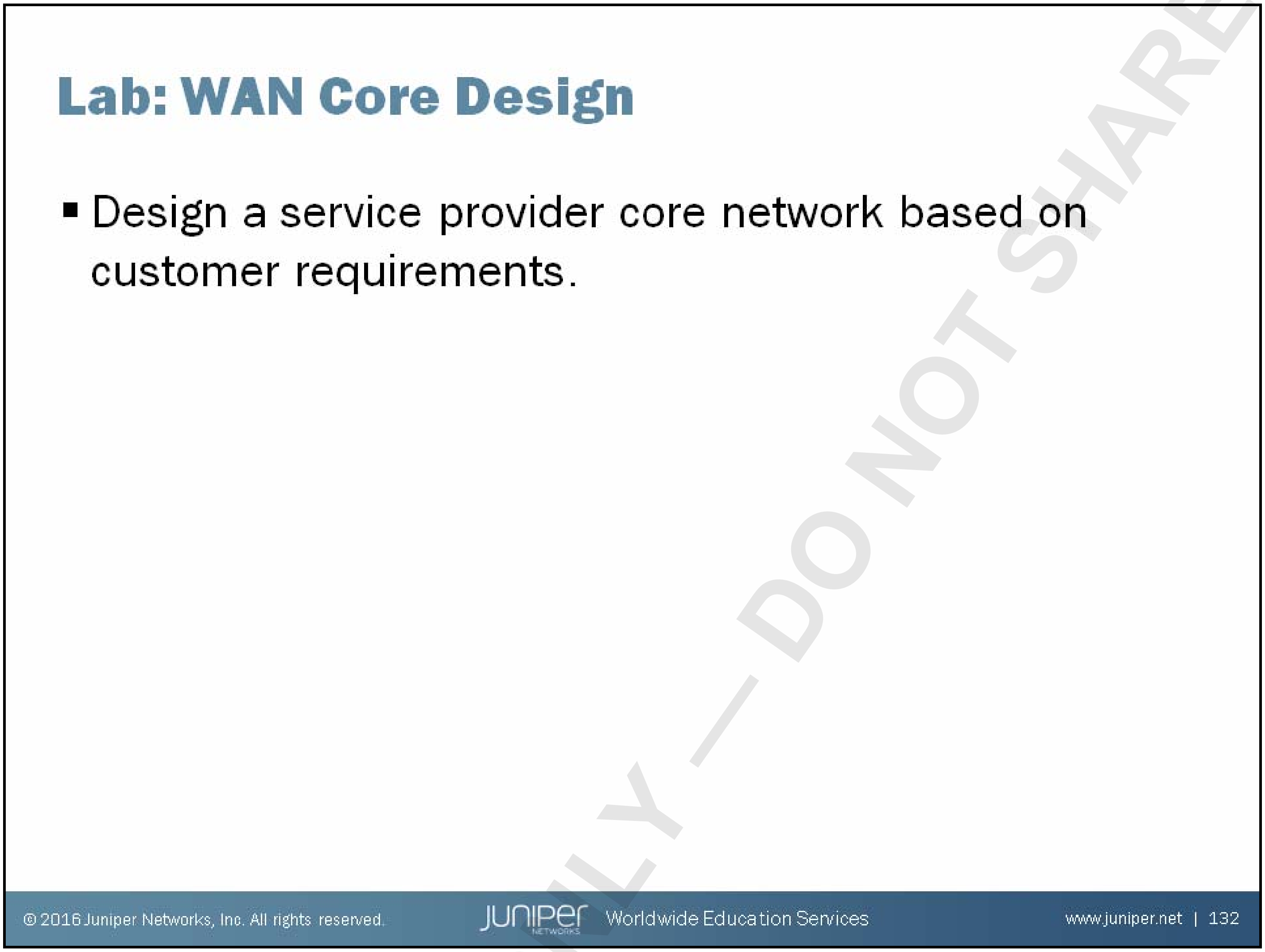
Review Questions

1.

2.

3.

4.



Lab: WAN Core Design

- Design a service provider core network based on customer requirements.

Worldwide Education Services

Lab: WAN Core Design

The slide provides the objective for this lab.

Answers to Review Questions

1.

Discover neighbors and establish adjacencies. Communicate internal network destinations including WAN router loopback interface addresses. Distribute traffic engineering data.

2.

Local preference, AS path, MED, and Origin attributes can be manipulated to specify preferred paths.

3.

RSVP can signal LSPs that use a path that differs from the IGP metric shortest path

4.

Both core and edge devices perform CoS. Edge devices classify traffic as arrives from the customer network and add CoS markings to the appropriate headers. Traffic is then forwarded into the core. Core devices analyze the CoS markings to classify the data.

Acronym List

ABR	area border router
AE	aggregated Ethernet
AF	assured forwarding
AS	autonomous system
BA	behavior aggregate
BE	best-effort
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
BGP4	BGP version 4
CBF	class-of-service-based forwarding
CBGP	confederation BGP
CBS	committed burst size
CCC	circuit cross connect
CIR	committed information rate
CLI	command-line interface
CLNP	Connectionless Network Protocol
CoS	class of service
CR-LDP	Constraint-Based Routed Label Distribution Protocol
CS	class selector
DEI	drop eligible indicator
DLCI	data-link connection identifier
DS	DiffServ
DSCP	DiffServ code point
EBS	excess burst size
EF	expedited-forwarding
EGP	exterior gateway protocol
ERO	EXPLICIT_ROUTE Object
ES	end system
ES-IS	End System–to–Intermediate System
EXP	experimental
FEC	forwarding equivalence class
FPC	Flexible PIC Concentrator
GRES	graceful Routing Engine switchover
GUI	graphical user interface
HA	high availability
HMAC	Hashed Message Authentication Code
IAB	Internet Advisory Board
IANA	Internet Assigned Numbers Authority
IBGP	internal BGP
ICMP	Internet Control Message Protocol
IGMP	Internet Group Membership Protocol
IGP	interior gateway protocol
IQ	Intelligent Queuing
IRB	integrated routing and bridging
IS	Intermediate System
IS-IS	Intermediate System-to-Intermediate System
ISO	International Organization for Standardization
JNCP	Juniper Networks Certification Program
L2CP	Layer 2 Control Protocol
l2cpd	Layer 2 Control Protocol process
LAG	link aggregation group
LDP	Label Distribution Protocol
LER	label edge router
LFA	loop-free alternate
LSA	link-state advertisement

LSDB	link-state database
LSP	label switched path
LSR	label-switching router
MAC	media access control
MD5	Message Digest 5
MED	multiple exit discriminator
MF	multifield
MIC	modular interface card
MPC	modular port concentrator
MSTP	Multiple Spanning Tree Protocol
MTBF	mean time between failure
MTTR	mean time to repair
MTU	maximum transmission unit
NLRI	network layer reachability information
NSB	nonstop bridging
NSR	nonstop active routing
OS	operating system
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
PBS	peak burst size
PDU	protocol data unit
PFE	Packet Forwarding Engine
PHB	Per-hop behavior
PHP	penultimate-hop popping
PIM	Protocol Independent Multicast
PIR	peak information rate
PLP	packet loss priority
POP	point of presence
QoS	quality-of-service
RED	random early detection
RIP	Routing Information Protocol
RIPng	RIP next generation
rpd	routing protocol process
RSTP	Rapid Spanning Tree Protocol
RSVP	Resource Reservation Protocol
SDN	software-defined network
SLA	service level agreement
SPF	shortest-path first
SRLG	Shared Risk Link Group
srTCM	Single-rate tricolor marking
STP	Spanning Tree Protocol
TC	traffic class
TCC	translational cross-connect
TCI	Tag Control Information
TED	traffic engineering database
TLV	type/length/value
ToS	type of service
trTCM	Two-rate tricolor marking
TTL	time to live
VLAN	virtual LAN
VPN	virtual private network
WAN	wide area network
WRED	weighted random early detection
WRR	weighted round-robin