

IxNetwork RoCEv2 Test Solution

Валидация фабрики ЦОД, обеспечивающей наилучшую производительность для рабочих нагрузок обучения ИИ

Введение

Обучение больших моделей ИИ привело к росту размеров кластеров и объёма рабочих нагрузок обучения. В таких системах множество вычислительных узлов — серверов с GPU — выполняют параллельные вычисления с коллективными операциями обмена данными между устройствами. Сеть, соединяющая эти устройства, должна обеспечивать высокую пропускную способность, низкую задержку и передачу трафика без потерь.

Сеть для обучения ИИ может иметь двух- или трёхуровневую архитектуру в зависимости от требуемого масштаба и выбранного проектного решения. Необходимо учитывать ECMP-хеширование, взаимоблокировки PFC и сквозную задержку передачи данных. Для валидации и сравнительной оценки производительности сетевой фабрики ИИ коммутационная фабрика должна задействовать RoCE Congestion Control и Priority Flow Control (PFC), чтобы оптимизировать управление буферами для рабочих нагрузок AI/ML.

Решение Keysight RoCEv2 Lossless Ethernet Test Solution включает высокоплотную экономичную тестовую платформу и тестовое приложение IxNetwork. Оно эмулирует соединения и потоки Queue-Pair (QP), формирует уведомления о перегрузке, выполняет динамическое управление скоростью на основе DCQCN, а также обеспечивает необходимую гибкость для тестирования пропускной способности, управления буферами и ECMP-хеширования с целью оптимизации производительности фабрики при передаче рабочих нагрузок обучения ИИ. Решение обеспечивает масштабируемую и экономичную проверку эффективности управления перегрузкой и сравнительную оценку производительности фабрики.

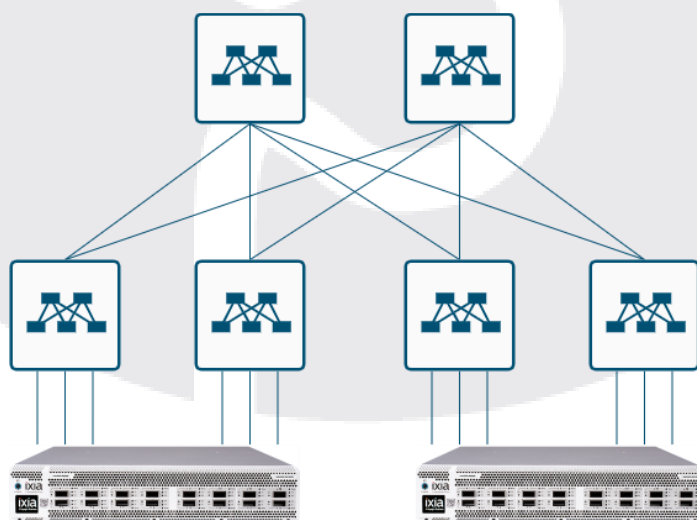


Рисунок 1. Keysight RoCEv2 Lossless Ethernet Test Solution

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

Ключевые функции

- Новый режим FPGA с поддержкой движка трафика потоков RoCEv2 Queue-Pairs (QP).
- Поддержка режимов скоростей RoCEv2 4×100GE NRZ и 8×100GE/4×200GE PAM4 (56G) на каждый Resource Group (RG).
- Эмуляция до 4K Q-Pairs, выполняющих RDMA WRITE с типом сервиса Reliable Connection (RC).
- Назначение кода DSCP для каждого Q-Pair.
- Автоматическая генерация номера Q-Pair или возможность задать номер Q-Pair пользователем.
- Передача данных с размером буфера до 252 MB при MTU 4K.
- Обнаружение сигнализации перегрузки ECN и генерация уведомления о перегрузке CNP.
- Выполнение алгоритма управления скоростью DCQCN для каждого Q-Pair с параметрами, настраиваемыми пользователем.
- Обработка кадров PFC Pause для приостановки и возобновления трафика.
- Поддержка шаблонов трафика in-cast (N:1), M:N и all-to-all с фиксированной или непрерывной передачей.
- Управление скоростью трафика в процентах от линейной скорости или через inter-batch-gap.
- Статистика RoCEv2 по каждому порту и по каждому QP.
- Поддержка MTU от 1500 байт до 14K байт.
- Комплексная поддержка TCL, Python/REST API для автоматизации.

Основные особенности

- Высокоплотная и экономичная тестовая платформа 400GE с запасом на будущее.
- Аппаратная генерация уведомлений о перегрузке ECN/CNP и управление скоростью DCQCN для каждого Q-Pair.
- Нагрузочное тестирование сетевой фабрики ИИ реалистичным трафиком RoCEv2, эмулирующим рабочую нагрузку ИИ.
- Комплексная статистика для поиска и устранения неисправностей и сокращения времени вывода продукта на рынок.

Высокоплотная и экономичная тестовая платформа

[AresONE-S 400GE QSFP-DD](#) с фиксированным шасси на 16 портов — самая высокоплотная в отрасли тестовая платформа 400GE. Она поддерживает 16 × 400GE / 32 × 200GE / 64 × 100GE PAM4 56G, а также скорость 100GE NRZ и является подходящей платформой для валидации AI Fabric.

RoCEv2 FPGA представляет собой выбираемый режим, включающий движок потоков RoCEv2 для каждого Resource Group (RG). Каждый RG поддерживает 2 × 400GE, 4 × 200GE и 8 × 100GE PAM4, а также 4 × 100GE NRZ. Он обнаруживает ECN-CE, формирует уведомления о перегрузке CNP и выполняет управление скоростью DCQCN.

Тестовая платформа поддерживает многопользовательский режим — до 8 пользователей на систему. Она одновременно поддерживает обычное тестирование control plane и data plane L23 и

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

тестирование RoCEv2 в разных RG, что повышает эффективность использования и снижает совокупную стоимость владения.



Рисунок 2. Keysight RoCEv2 Lossless Ethernet Test Solution — [AresONE-S 400GE QSFP-DD](#)

Эмуляция конечных точек RDMA

IxNetwork эмулирует конечные точки RDMA, устанавливающие соединения Q-Pair, выполняет операции RDMA WRITE в режиме Reliable Connection (RC), генерирует шаблоны трафика in-cast (N:1), M:N и all-to-all с фиксированной или непрерывной передачей и предоставляет статистику RoCEv2 для каждого Q-Pair.

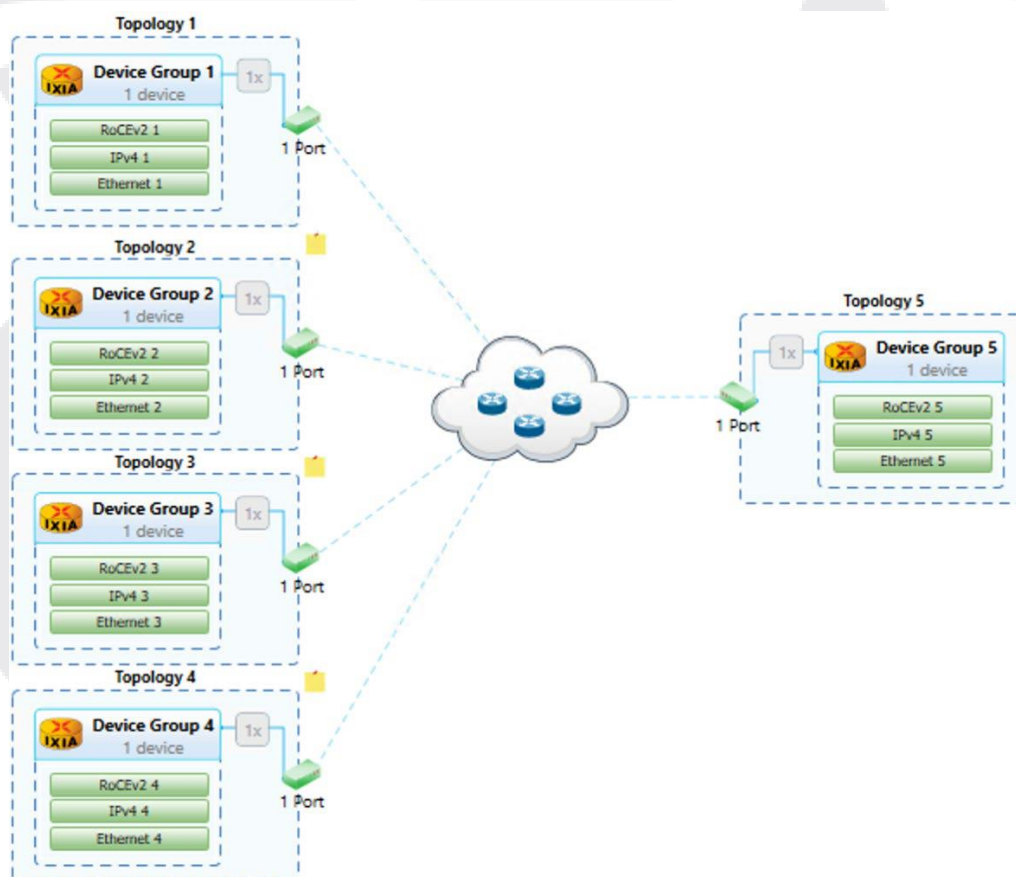


Рисунок 3. Эмуляция конечных точек IxNetwork RoCEv2 — In-cast 4:1

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

В конфигурации Q-Pair номер Q-Pair генерируется автоматически либо задаётся пользователем. DSCP можно сопоставлять на уровне каждого QP; поддерживается размер буфера до 256 MB при MTU 4K.

Device#	Local IP	Remote IP	Auto QP Number	Custom QP	Custom QP Number	DSCP	UDP Source Port	Execute Commands	Buffer Size	Buffer Size Unit
1 (x 171)					Inc:2, 1	0	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1	31.1.1.10	31.9.1.10	2		2	0	49152	RDMA WRITE	1	MB
# 1.2	31.1.1.10	31.9.1.10	3		3	0	49153	RDMA WRITE	1	MB
# 1.3	31.1.1.10	31.9.1.10	4		4	0	49154	RDMA WRITE	1	MB
# 1.4	31.1.1.10	31.9.1.10	5		5	0	49155	RDMA WRITE	1	MB
# 1.5	31.1.1.10	31.9.1.10	6		6	0	49156	RDMA WRITE	1	MB
# 1.6	31.1.1.10	31.9.1.10	7		7	0	49157	RDMA WRITE	1	MB
# 1.7	31.1.1.10	31.9.1.10	8		8	0	49158	RDMA WRITE	1	MB
# 1.8	31.1.1.10	31.9.1.10	9		9	0	49159	RDMA WRITE	1	MB
# 1.9	31.1.1.10	31.9.1.10	10		10	0	49160	RDMA WRITE	1	MB
# 1.10	31.1.1.10	31.9.1.10	11		11	0	49161	RDMA WRITE	1	MB

Рисунок 4. Конфигурация Q-Pair IxNetwork RoCEv2

Генерация потоков Q-Pair одним щелчком предоставляет подробные сведения о потоках QP. Пользователь может управлять скоростью трафика либо в процентах от линейной скорости, либо с помощью inter batch period.

Статистика RoCEv2 по каждому QP содержит число операций RDMA WRITE с успешными и неуспешными операциями, количество пакетов и задержку, а также счётчики ECN/CNP/ACK/NAK для проверки перегрузки и поиска причин отказов.

RoCEv2 Flow Group

RoCEv2 Port Configurations

RoCEv2 Traffic Flow Groups

☒ Enable RoCEv2 Traffic

Enabled	Flow Group Name	Tx Port	Rx Port	Destination QP	Packets	Frame Size (Byte)	Source IP	Destination IP	Source MAC	Destination MAC	Udp Source Port	Burst Mode	Burst Count
☒	RoCEv2 Flow Group 0001	Ethernet - 008	Ethernet - 009	856	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49152	Continuous	
☒	RoCEv2 Flow Group 0025	Ethernet - 008	Ethernet - 009	857	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49153	Continuous	
☒	RoCEv2 Flow Group 0026	Ethernet - 008	Ethernet - 009	858	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49154	Continuous	
☒	RoCEv2 Flow Group 0027	Ethernet - 008	Ethernet - 009	859	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49155	Continuous	
☒	RoCEv2 Flow Group 0028	Ethernet - 008	Ethernet - 009	860	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49156	Continuous	
☒	RoCEv2 Flow Group 0029	Ethernet - 008	Ethernet - 009	861	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49157	Continuous	
☒	RoCEv2 Flow Group 0030	Ethernet - 008	Ethernet - 009	862	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49158	Continuous	
☒	RoCEv2 Flow Group 0031	Ethernet - 008	Ethernet - 009	863	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49159	Continuous	
☒	RoCEv2 Flow Group 0032	Ethernet - 008	Ethernet - 009	864	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49160	Continuous	
☒	RoCEv2 Flow Group 0033	Ethernet - 008	Ethernet - 009	865	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49161	Continuous	
☒	RoCEv2 Flow Group 0034	Ethernet - 008	Ethernet - 009	866	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49162	Continuous	
☒	RoCEv2 Flow Group 0035	Ethernet - 008	Ethernet - 009	867	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49163	Continuous	
☒	RoCEv2 Flow Group 0036	Ethernet - 008	Ethernet - 009	868	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49164	Continuous	
☒	RoCEv2 Flow Group 0037	Ethernet - 008	Ethernet - 009	869	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49165	Continuous	
☒	RoCEv2 Flow Group 0038	Ethernet - 008	Ethernet - 009	870	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49166	Continuous	
☒	RoCEv2 Flow Group 0039	Ethernet - 008	Ethernet - 009	871	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49167	Continuous	

Flow groups

Select Views...

Port Statistics

RoCEv2 Per Port

RoCEv2 Flow Statistics

Tx Port	Rx Port	Traffic Item	Dest QP	Data Frames Tx	Data Frames Rx	Frames Delta	WRITE Tx	WRITE Complete Rx	WRITE Fail	Avg Latency (ns)	Min Latency (ns)	Max Latency (ns)	ECN-CR Rx	CNP Tx	CNP Rx	ACK Tx	ACK Rx
1 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	2	2,404,620	2,409,000	4,380	3,294	3,300	6	20,864	692	98,020	205,486	183,411	182,870	3,306	3,306
2 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	3	2,404,241	2,409,000	4,759	3,293	3,300	7	19,902	687	101,335	185,464	185,126	184,158	3,306	3,306
3 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	4	2,404,268	2,409,000	4,732	3,293	3,300	7	19,894	690	99,085	186,425	186,126	185,087	3,305	3,305
4 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	5	2,404,140	2,409,000	4,860	3,293	3,300	7	19,903	690	42,437	185,158	184,857	183,855	3,306	3,306
5 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	6	2,404,127	2,408,958	4,831	3,293	3,299	6	19,864	690	103,375	184,838	184,585	183,593	3,306	3,306
6 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	7	2,404,291	2,409,000	4,709	3,293	3,300	7	19,880	680	102,822	185,322	185,030	184,020	3,306	3,306
7 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	8	2,404,338	2,409,000	4,662	3,293	3,300	7	19,926	690	75,307	186,634	186,342	185,338	3,306	3,306
8 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	9	2,404,301	2,409,000	4,699	3,293	3,300	7	19,868	690	98,140	185,633	185,330	184,344	3,305	3,305
9 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	10	2,404,275	2,409,000	4,725	3,293	3,300	7	19,892	692	103,015	186,136	185,865	184,836	3,305	3,305
10 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	11	2,404,208	2,408,775	4,567	3,293	3,299	6	19,885	690	102,952	185,157	184,833	183,906	3,305	3,305
11 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	12	2,404,361	2,409,000	4,639	3,293	3,300	7	19,896	682	98,772	186,220	185,965	184,938	3,305	3,305
12 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	13	2,404,165	2,409,000	4,835	3,293	3,300	7	19,854	690	103,377	184,202	183,842	182,900	3,305	3,305
13 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	14	2,404,278	2,409,000	4,722	3,293	3,300	7	19,869	677	103,015	185,381	184,986	183,956	3,305	3,305
14 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	15	2,404,213	2,409,000	4,787	3,293	3,300	7	19,874	685	101,335	185,928	185,562	184,557	3,305	3,305
15 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	16	2,404,169	2,409,000	4,831	3,293	3,300	7	19,855	692	97,882	184,642	184,331	183,292	3,305	3,305
16 Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	17	2,404,144	2,408,876	4,732	3,293	3,299	6	19,848	686	103,015	185,381	184,986	183,956	3,305	3,305
SUM=2,462,0...				SUM=2,468,8...	SUM=4,930...	SUM=2,468,8...	SUM=3,293...	SUM=3,381,531	SUM=9,161	Max=21,245	Max=692	Max=106,485	SUM=18...	SUM=18...	SUM=18...	SUM=3,3...	SUM=3,3...

Технические характеристики

ISIS segment routing

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

Аппаратная платформа	• AresONE-S 400GE QSFP-DD, модель фиксированного шасси на 16 портов • AresONE-S 400GE QSFP-DD, модель фиксированного шасси на 8 портов • AresONE-M 800GE QSFP-DD800, модель фиксированного шасси на 8 портов • AresONE-M 800GE QSFP-DD800, модель фиксированного шасси на 4 порта
Скорости Ethernet	• 100GE NRZ • 100GE PAM4 56G • 200GE PAM4 56G
Конфигурация Q-Pairs	• Локальный и удалённый IP • Auto QP Number или Custom QP Number • Сопоставление DSCP • Execute Command: RDMA WRITE • Размер буфера и единица измерения • Соединение: Connect Request, Connect Reply, ReadyToUse
Управление перегрузкой	• Обнаружение ECN-CE • Генерация CNP и приоритет DSCP • Параметры управления скоростью DCQCN • CNP Delay Timer
Конфигурация потоков трафика	• Q-Pair Mesh: In-cast (N:1), All-to-all, Partial mesh (M:N) • Burst mode: Fixed, Continuous • Rate: Target % Line Rate, Inter batch Period • DCQCN Setting
Статистика	• Количество пакетов и задержка пакетов • RDMA WRITE Count: Complete or Fail • ECN Rx, CNP Tx/Rx, ACK Tx/Rx, NAK Tx/Rx • Sequency error

Информация для заказа

Партномер	Описание
Партномера RoCEv2	
905-1092	Опция Keysight RoCEv2 Lossless Ethernet Enablement, ЗАВОДСКАЯ УСТАНОВКА, для моделей фиксированных шасси AresONE-S 400GE и AresONE-M 800GE (905-1092).

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.

905-1093	Опция Keysight RoCEv2 Lossless Ethernet Enablement, ПОЛЕВОЕ ОБНОВЛЕНИЕ, для моделей фиксированных шасси AresONE-S и AresONE-M (905-1093).
930-2208	Тестовый пакет Keysight IxNetwork RoCEv2 Lossless Ethernet Test Package для моделей фиксированных шасси AresONE-S 400GE и AresONE-M 800GE (930-2208).
Комплекты RoCEv2	
947-4071	Комплект Keysight RoCEv2 Lossless Ethernet Test Bundle для модели фиксированного шасси AresONE-S 400GE QSFP-DD на 16 портов (947-4071).
947-4072	Комплект Keysight RoCEv2 Lossless Ethernet Test Bundle для модели фиксированного шасси AresONE-S 400GE QSFP-DD на 8 портов (947-4072).
947-4073	Комплект Keysight RoCEv2 Lossless Ethernet Test Bundle для модели фиксированного шасси AresONE-M 800GE QSFP-DD800 на 8 портов (947-4073).
947-4074	Комплект Keysight RoCEv2 Lossless Ethernet Test Bundle для модели фиксированного шасси AresONE-M 800GE QSFP-DD800 на 4 порта (947-4074).

Тел: +7 (495) 252-00-96

Email: info-site@akmetron.ru

Адрес: Москва, ул. Рабочая, д.93 стр.2, под. 2.