WHITE PAPER

ПОДКЛЮЧЕНИЕ РЕСУРСОВ К СЕРВЕРНОЙ ОС: UDEV DEVICE MANAGER

YADRO TATLIN

О WHITE PAPERS

Данный документ описывает архитектурные принципы и подходы, а также рекомендации по их применению при построении корпоративных решений на основе продуктов производства компании YADRO.

Документ предназначен для широкой аудитории, но также может включать материалы, требующие специальной профессиональной подготовки, ориентированные на системных архитекторов, администраторов, сервисных инженеров и других IT-специалистов, участвующих в планировании, проектировании и внедрении комплексных решений.

! Представленная информация основана на результатах исследований, а также экспертизе специалистов YADRO. Документ носит не предписывающий, а информационный характер, целью которого является информирование получателей документа о предмете произведенного исследования и о его результатах. При этом конкретные результаты и эффективность конечных решений зависят от множества факторов, которые могут различаться в зависимости от особенностей инфраструктуры и условий эксплуатации. Ни при каких обстоятельствах YADRO не несет ответственность за последствия применения получателем документа рекомендаций, инструкций, подходов и решений, указанных в документе.

Запрещается копировать, воспроизводить, изменять, публиковать, передавать, распространять, продавать информацию (часть информации), содержащуюся в документе, или создавать производные работы на основе содержания документа без предварительного письменного разрешения YADRO.

Любая информация или утверждения относительно возможностей, емкости, производительности или других характеристик продукции, указанной в документе, предоставляются исключительно на условиях «как есть» и приведена исключительно в целях описания произведенного исследования. YADRO оставляет за собой право вносить изменения/обновлять информация в документе без уведомления.

«YADRO» является зарегистрированным товарным знаком ООО «КНС ГРУПП». Другие названия продуктов и компаний, упомянутые здесь, могут быть товарными знаками или торговыми наименованиями их соответствующих владельцев.

Актуальные версии документации публикуются на информационном ресурсе YADRO по адресу service.yadro.com.

ОГЛАВЛЕНИЕ

1. ОБ ЭКСПЕРТАХ 4
2. ВВЕДЕНИЕ 5
3. СОЗДАНИЕ ПРАВИЛ UDEV ДЛЯ НАЗНАЧЕНИЯ ИМЕН СЕТЕВЫХ КАРТ (NIC) 6
4. ИСПОЛЬЗОВАНИЕ UDEV ДЛЯ НАСТРОЙКИ ПАРАМЕТРОВ БЛОЧНЫХ УСТРОЙСТВ 8
5. ПРАКТИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ UDEV DEVICE MANAGER 11

ОБ ЭКСПЕРТАХ

Дмитрий Боровской — специалист с более чем 30-летним опытом работы в сфере ИТ, включая работу с системами хранения данных различных вендоров (EMC, Dell, Hitachi) и отечественных производителей (YADRO). В YADRO занимает должность ведущего специалиста отдела Мультивендорной Сервисной Экспертизы. Специализируется на администрировании Linux, конфигурировании и поддержке систем хранения, а также интеграции продуктов TATLIN в инфраструктуру заказчиков. Эксперт в области сетей хранения и передачи данных (SAN, IPSAN).

ВВЕДЕНИЕ

`Udev` — это подсистема в Linux, отвечающая за динамическое управление устройствами. `Udev` является частью системы управления устройствами (device management) и выполняет следующие функции:

- Динамическое создание и удаление устройств в `/dev`.
- Автоматическое создание файлов устройств (`/dev/sdX`, `/dev/ttyUSB0`, `/dev/input/mouse0` и т. д.) при их подключении.
- Назначение постоянных имен устройствам.
- Запуск скриптов и выполнение команд при появлении или изменении состояния устройств.
- Обеспечение взаимодействия ядра с пространством пользователя через события `udev`.

Основная задача демона `udevd` — динамическое создание специальных файлов устройств (special device files) в каталоге `/dev` и динамическое применение дополнительных настроек. Параметры создаваемых файлов устройств и дополнительные настройки задаются путем создания правил для `udev`. Файлы с пользовательскими настройками обычно хранятся в каталоге `/etc/udev/rules.d` и имеют имя в формате `*.rules`. Демон, получив от ядра системы событие, связанное с добавлением, удалением или изменением устройств, выполняет действия, заданные в правилах. Данный механизм обеспечивает удобное и централизованное управление периферией.

Кроме специальных файлов устройств, `udev` может создавать символические ссылки на файлы устройств, которые могут задавать «говорящие» имена устройств.

Каждое событие, получаемое от ядра демоном `udev`, содержит основную информацию об устройстве, такую как его имя, подсистему, которой оно принадлежит, тип устройства, старший и младший номер специального файла устройств (major and minor number), а также тип события. На основе этой информации, а также данных из каталога `/sys`, можно задавать фильтры и обработчики событий, которые будут выполнять определенные действия, связанные с устройствами. Например, можно изменять владельца, права доступа для специального файла устройства, изменять имя устройства, вызывать внешние обработчики событий (например, скрипты) и т.д.

ВЕРСИИ ИСПОЛЬЗУЕМОГО ПО	
TATLIN.UNIFIED Gen1/Gen2	2.7–3.1.x
Linux Debian based	10–12
Linux RedHat based	7–9
FabricOS	8.1.x–9.2.x

СОЗДАНИЕ ПРАВИЛ UDEV ДЛЯ НАЗНАЧЕНИЯ ИМЕН СЕТЕВЫХ КАРТ (NIC)

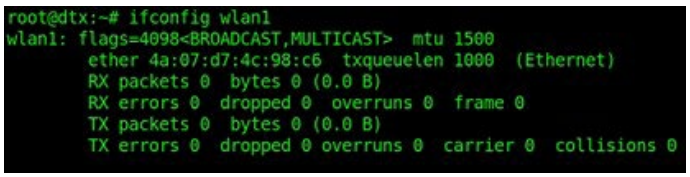
Присвоение «говорящего» имени устройству на примере USB WiFi адаптера D-Link DWA-126:

> **!** Инструкция актуальна для любой сетевой карты (NIC), независимо от способа подключения.

> **!** Приведенные в документе команды выполнены в профиле суперпользователя `root`. Также возможно их выполнение через команду `sudo` (Substitute User DO) и вводом пароля суперпользователя.

1. Определите параметры или атрибуты данного устройства, на которые можно создать фильтр.

   Пример для устройства с именем wlan1

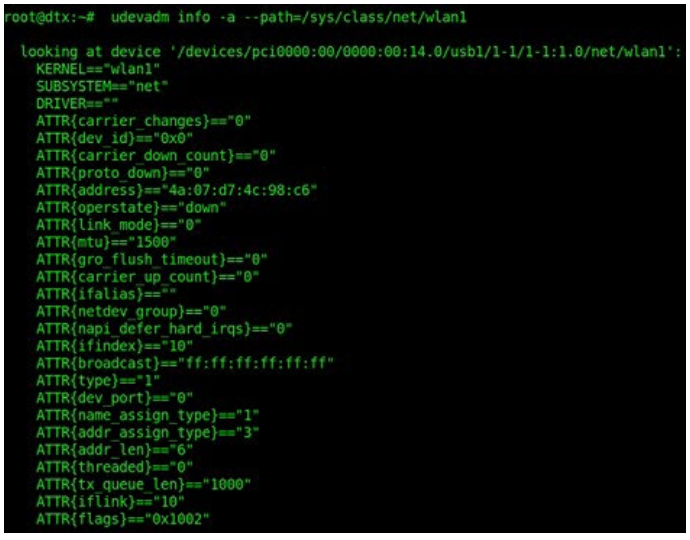


2. Определите, какие характеристики данного устройства можно использовать для создания фильтра:

```
udevadm info -a --path=/sys/class/net/wlan1
```

   Вывод команды представлен в виде списка в формате: `<VARIABLE_NAME>==<VALUE>`. В качестве имени переменной также может выступать массив с различными атрибутами устройства (в примере ниже – `ATTR{}`).

   Пример



Свяжитесь с нами:
support@yadro.com
www.yadro.com
Москва, ул. Рочдельская, 15, стр.13
+7 800 777-06-11

3. Сделайте привязку нового имени NIC к MAC-адресу. Для этого используйте атрибуты: `SUBSYSTEM=="net"`, `ATTR{address}=="4a:07:d7:4c:98:c6"`.
4. Создайте правило `udev`. Для этого в файле `/etc/udev/rules.d/70-persistent-net.rules` добавьте строки :

```
--- cut ---
ACTION=="add", \
      SUBSYSTEM=="net", \
      ATTR{address}=="fc:75:16:67:08:7f", \
      NAME="dlink-wifi0"
--- cut ---
```

Правило можно записать в одну строку, но для удобства можно разделить на несколько, поставив в конце очередной строки символ "\". В первой строке указывается тип события, на которое срабатывает правило `add` (добавление нового устройства). Вторая и третья строка задают фильтр для данного правила (указанные условия обрабатываются через логическое И). Четвертая строка — выполняемое действие. В данном случае атрибуту `NAME` присваивается значение "`dlink-wifi0`". Обратите внимание, что в фильтре стоит два знака равно (==), а в присвоении значения атрибуту – только один. Кроме того, присваиваемое значение и значение в фильтре должно быть в кавычках.

5. «Уведомите» `udevd` о том, что настройки изменились и нужно перечитать файлы конфигурации:

```
udevadm control --reload
```

После этого, при каждом новом подключении или после перезагрузки, данная сетевая карта будет иметь имя `dlink-wifi0`:

```
root@dtx:~# ifconfig -a
dlink-wifi0: flags=4098<BROADCAST,MULTICAST>  mtu 1500
        ether 76:88:43:0d:40:8f  txqueuelen 1000  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 0  bytes 0 (0.0 B)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

Привязка имени устройства к MAC-адресу выполнена.

В некоторых случаях необходимо, чтобы сетевая карта (NIC) в определенном слоте PCI имела определенное имя независимо от MAC-адреса. Переименование NIC на примере устройства с первоначальным именем `ens33`:

1. Дополнительно установите MTU 9000. В этом случае привязка выполняется по адресу на PCI-шине.
2. Проверьте атрибуты данного устройства (секцию `SUBSYSTEMS=="pci"`):

```
[root@dtx]:/# udevadm info -a --path=/sys/class/net/ens33

  {SKIP}

  looking at parent device '/devices/pci0000:00/0000:00:11.0/0000:02:01.0':
    KERNELS=="0000:02:01.0"
    SUBSYSTEMS=="pci"
    DRIVERS=="e1000"
    ATTRS{d3cold_allowed}=="1"
    ATTRS{ari_enabled}=="0"
    ATTRS{msi_bus}=="1"
    ATTRS{consistent_dma_mask_bits}=="32"
    ATTRS{numa_node}=="-1"
    ATTRS{subsystem_device}=="0x0750"
    ATTRS{local_cpulist}=="0-3"
    ATTRS{broken_parity_status}=="0"
    ATTRS{label}=="Ethernet0"
    ATTRS{class}=="0x020000"
    ATTRS{reset_method}=="pm"
    ATTRS{local_cpus}=="00000000,00000000,00000000,0000000f"
    ATTRS{subsystem_vendor}=="0x15ad"
    ATTRS{device}=="0x100f"
    ATTRS{revision}=="0x01"
    ATTRS{enable}=="1"
    ATTRS{irq}=="19"
    ATTRS{vendor}=="0x8086"
    ATTRS{power_state}=="D0"
    ATTRS{acpi_index}=="16777736"
    ATTRS{dma_mask_bits}=="32"
    ATTRS{driver_override}=="(null)"
```

3. В качестве фильтра используйте адрес устройства на PCI-шине: `KERNELS=="0000:02:01.0"`. Если привязку необходимо сделать для конкретной модели сетевой карты (NIC), используйте дополнительно атрибуты `ATTRS{vendor}` и `ATTRS{device}` (фактически, это `device PCI ID`). В данном случае `0x8086:0x100f` — это Intel 82545EM Gigabit Ethernet Controller.
4. Создайте правило `udev`:

```
--- cut ---
ACTION=="add", \
      SUBSYSTEMS=="pci", \
      KERNELS=="0000:02:01.0", \
      ATTRS{vendor}=="0x8086", \
      ATTRS{device}=="0x100f", \
      NAME="LC0", \
      ATTR{mtu}="9000"
--- cut ---
```

5. Перезагрузите демон `udev`.
   После перезагрузки:

```
dim@dtx:~# ifconfig
LC0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 9000
        inet 192.168.1.22  netmask 255.255.0.0  broadcast 192.168.255.255
        inet6 fe80::a471:85b8:924b:722f  prefixlen 64  scopeid 0x20<link>
        ether 00:0c:29:2d:56:66  txqueuelen 1000  (Ethernet)
        RX packets 10526  bytes 14837546 (14.8 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 2302  bytes 219181 (219.1 KB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

ИСПОЛЬЗОВАНИЕ UDEV ДЛЯ НАСТРОЙКИ ПАРАМЕТРОВ БЛОЧНЫХ УСТРОЙСТВ

Обратитесь к рекомендациям по настройке LUN, передаваемых из массива TATLIN.UNIFIED при подключении по FC или iSCSI. Согласно документации по настройке СХД требуется:

1. Установить для `SCSI command timeout` значение 120 секунд.
2. Дополнительно установить планировщик ввода-вывода — `none` или `noop` (в зависимости от версии ядра системы).
3. Установить флаг, указывающий, что LUN, передаваемый с СХД, не является механическим диском (rotational device).
   В этом случае файл `/etc/udev/rules.d/70-tatlin.rules` с правилами `udev` будет иметь следующий вид:

```
--- cut ---
# For TATLIN LUNs:
# SCSI command timeout: 120 seconds, scheduler: none and it's not rotational disk

ACTION=="add|change", \
      SUBSYSTEM=="block", \
      ENV{ID_VENDOR}=="YADRO", ENV{ID_MODEL}=="TATLIN", \
      ATTR{device/timeout}="120", \
      ATTR{queue/scheduler}="none", \
      ATTR{queue/rotational}="0"
--- cut ---
```

ПАРАМЕТРЫ	
`ACTION=="add|change"`	Правило срабатывает при добавлении нового LUN или изменении существующего.
`SUBSYSTEM=="block"`	Используется подсистема блочных устройств.
`ENV{ID_VENDOR}=="YADRO"` `ENV{ID_MODEL}=="TATLIN"`	Правило срабатывает только для LUN, имеющих соответствующие атрибуты.
`ATTR{device/timeout}="120`	Установить для SCSI command timeout значение 120 секунд.
`ATTR{queue/scheduler}="none"`	Установить планировщик ввода-вывода.
`ATTR{queue/rotational}="0"`	Установить флаг, указывающий, что устройство не является механическим диском.

Чтобы посмотреть список всех атрибутов для блочного устройства:

```
udevadm info -a --path=/sys/class/block/{DEVICE}
```

`{DEVICE}` — имя блочного устройства (`sda`, `dm-0` и т.п.)

Дополнительно, для многопутевых устройств (multipathing devices) можно поменять планировщик ввода-вывода и установить флаг, указывающий, что LUN, передаваемый с СХД, не является механическим диском (rotational device). Для этого:

1. Создайте файл `/etc/udev/rules.d/10-dm.rules`:

```
--- cut ---
# Set scheduler and rotational flag for /dev/dm-*ACTION=="add|change", \
      KERNEL=="dm-*", \
      ATTR{queue/rotational}="0", \
      ATTR{queue/scheduler}="none"
--- cut ---
```

2. «Уведомите» демон `udevd` об изменении настроек:

```
udevadm control --reload
```

3. Сгенерируйте событие "`change`", чтобы применились новые настройки:

```
udevadm trigger --verbose --action=change --attr-match="subsystem=block"
```

4. Проверьте, что настройки применились:

```
lsblk -t
```

Пример вывода

```
NAME     ALIGNMENT MIN-IO  OPT-IO PHY-SEC LOG-SEC ROTA SCHED       RQ-SIZE   RA WSAME
sda              0    512       0     512     512    1 mq-deadline     254  128    0B
├─sda1           0    512       0     512     512    1 mq-deadline     254  128    0B
├─sda2           0    512       0     512     512    1 mq-deadline     254  128    0B
└─sda3           0    512       0     512     512    1 mq-deadline     254  128    0B
sdb              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpatha         0   8192 4194304     512     512    0 none           2048 8192   32M
sdc              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpatha         0   8192 4194304     512     512    0 none           2048 8192   32M
sdd              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpatha         0   8192 4194304     512     512    0 none           2048 8192   32M
sde              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpatha         0   8192 4194304     512     512    0 none           2048 8192   32M
sdh              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpathd         0   8192 4194304     512     512    0 none           2048 8192   32M
sdk              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpathd         0   8192 4194304     512     512    0 none           2048 8192   32M
sdn              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpathd         0   8192 4194304     512     512    0 none           2048 8192   32M
sr0              0    512       0     512     512    1 mq-deadline      64  128    0B
sdq              0   8192 4194304     512     512    1 none            113 8192   32M
└─mpathd         0   8192 4194304     512     512    0 none           2048 8192   32M
```

ИЗ ПРИМЕРА ВЫВОДА	
NAME	Имя блочного устройства.
ALIGNMENT	Выравнивание раздела.
MIN-IO	Минимальный IOsize в байтах.
OPT-IO	Оптимальный IOsize в байтах.
PHY-SEC	Физический размер блока.
LOG-SEC	Логический размер блока.
ROTA	Значение флага `rotational device`. Дополнительная настройка, используемая некоторыми планировщиками ввода-вывода, например, `mq-deadline`.
SHED	Используемый планировщик ввода-вывода.
RQ-SIZE	Глубина очереди ввода-вывода для данного блочного устройства.
RA	Размер предвыборки для устройства (read-ahead of the device).
WSAME	Размер в байтах, поддерживаемый SCSI командой `WRITE SAME (10, 16 or 32 byte)`. Подробнее см. `man(8)sg_write_same`.

В старых версиях ядра планировщик ввода-вывода может иметь название `noop`. Чтобы посмотреть список доступных планировщиков для данного устройства:

```
cat /sys/class/block/{DEVICE}/queue/scheduler
```

Пример:

```
root@dtx:~# cat /sys/class/block/sdm/queue/scheduler
[mq-deadline] none

root@dtx:~# cat /sys/class/block/dm-0/queue/scheduler
[none] mq-deadline
```

Используемый планировщик выведен в квадратных скобках.

При необходимости можно поменять планировщик вручную без изменения конфигурации udev, но такие изменения действуют только до перезагрузки. Чтобы изменить планировщик:

```
echo {SchedulerName} > /sys/class/block/{DEVICE}/queue/scheduler
```

{SchedulerName} — название планировщика.

Пример:

```
root@ubuntu-x64:~# echo none > /sys/class/block/sdm/queue/scheduler
root@ubuntu-x64:~# cat /sys/class/block/sdm/queue/scheduler
[none] mq-deadline
```

Подробнее о планировщиках ввода-вывода см. RedHat KB 109223: «How to use the Noop or None IO Schedulers» и Раздел 8.1.1. I/O Schedulers в «Performance Tuning Guide».

ПРАКТИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ UDEV DEVICE MANAGER

Использование подсистемы `udev` позволяет существенно упростить администрирование ОС Linux, обеспечить предсказуемую реакцию ОС на добавление новых устройств, обеспечить прогнозируемое конфигурирование новых устройств.
Также подсистема `udev` широко используется при написании системных скриптов управления и мониторинга, позволяя выполнять фильтрацию событий в привязке к конкретному физическом у устройству.

При конфигурировании подсистемы `udev` рекомендуется использовать для наименования устройств «говорящие» имена, позволяющие идентифицировать назначение устройства. При работе с СХД TATLIN необходимо создавать правило, устанавливающее для параметра `SCSI command timeout` значение 120 секунд. Данное правило является критичным для обеспечения устойчивости системы при обновлении ПО контроллеров хранения и/или проведения ремонтных работ на контроллерах хранения СХД.