

mix_protocols number of requests

Rafilx

2022-06-12

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Loading required package: gridExtra

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##   combine

## Loading required package: viridisLite

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

R Markdown

Analisar a porcentagem de equisições por protocolo, dividindo por períodos. Essa análise não envolve os payloads, apenas os quantitativos de requisições.

Resultados esperados:

- gráficos de linhas e de barras mostrando a evolução

```
db <- dbConnect(RSQLite::SQLite(), dbname="../db/database-2022-05-11/mix_protocol.sqlite")

data_unfetch <-dbSendQuery(db, "
  SELECT *, CAST(CAST(year AS text) || CAST(period AS text) as integer) as year_period
  FROM (
    SELECT *, strftime(\"%Y\", tempo_inicio) as year, ((strftime(\"%m\", tempo_final) - 1) / 3) + 1
    FROM MIX_PROTOCOL
  )
")
data <- fetch(data_unfetch)

dbDisconnect(db)
```

```
## Warning in connection_release(conn@ptr): There are 1 result in use. The
## connection will be released when they are closed
```

- Agrupamento realizado por período (trimestre) e “attack_protocol” é o protocolo utilizado no ataque ["chargen", "cldap", "coap", "dns", "memcached", "ntp", "qotd", "ssdp", "steam_games", "outros"]
- Somando a quantidade de requisições utilizadas por cada protocolo e período

```
data['tempo_final_cast'] = as.POSIXct(data[['tempo_final']], format = "%Y-%m-%d %H:%M:%S")
data['tempo_inicio_cast'] = as.POSIXct(data[['tempo_inicio']], format = "%Y-%m-%d %H:%M:%S")
```

```
data_grouped_period_protocol = data %>%
  mutate(year_period_int = year_period,
         year_period = as.factor(year_period),
         attack_protocol = as.factor(attack_protocol)) %>%
  group_by(year_period, attack_protocol) %>%
  summarise(sum_requests_per_attack = sum(requests_per_attack),
            number_of_attacks = n(),
            tempo_inicio=min(tempo_inicio_cast),
            tempo_final=max(tempo_final_cast))
```

```
## 'summarise()' has grouped output by 'year_period'. You can override using the
## '.groups' argument.
```

```
data_grouped_period_protocol_percentage = data_grouped_period_protocol %>%
  ungroup() %>%
  group_by(year_period) %>%
  summarise(attack_protocol = attack_protocol,
            number_of_attacks = number_of_attacks,
            tempo_inicio = tempo_inicio,
            tempo_final = tempo_final,
            sum_period_number_of_attacks = sum(number_of_attacks),
            sum_period_requests_per_attack = sum(sum_requests_per_attack),
            sum_requests_per_attack = sum_requests_per_attack) %>%
  mutate(number_of_attacks_percentage = (number_of_attacks / sum_period_number_of_attacks) * 100,
         number_of_requests_percentage = (sum_requests_per_attack / sum_period_requests_per_attack) * 100)
```

```
## 'summarise()' has grouped output by 'year_period'. You can override using the
## '.groups' argument.
```

```

minimum_percentage_as_others = 5
decimals_digits = 1

data_grouped_period_protocol_others_percentage = data_grouped_period_protocol_percentage %>%
  mutate(
    attack_protocol = case_when(
      number_of_requests_percentage < minimum_percentage_as_others ~ "OUTROS",
      TRUE ~ as.character(attack_protocol)
    )
  ) %>%
  group_by(year_period, attack_protocol) %>%
  summarise(number_of_requests_percentage = sum(number_of_requests_percentage))

## 'summarise()' has grouped output by 'year_period'. You can override using the
## '.groups' argument.

```

```

filter_in_period_2020 = c(20201, 20202, 20203, 20204)
filter_in_period_2018_2019 = c(20181, 20182, 20183, 20184, 20191, 20192, 20193, 20194)
filter_in_period_2021_2022 = c(20211, 20212, 20213, 20214, 20221, 20222, 20223)

```

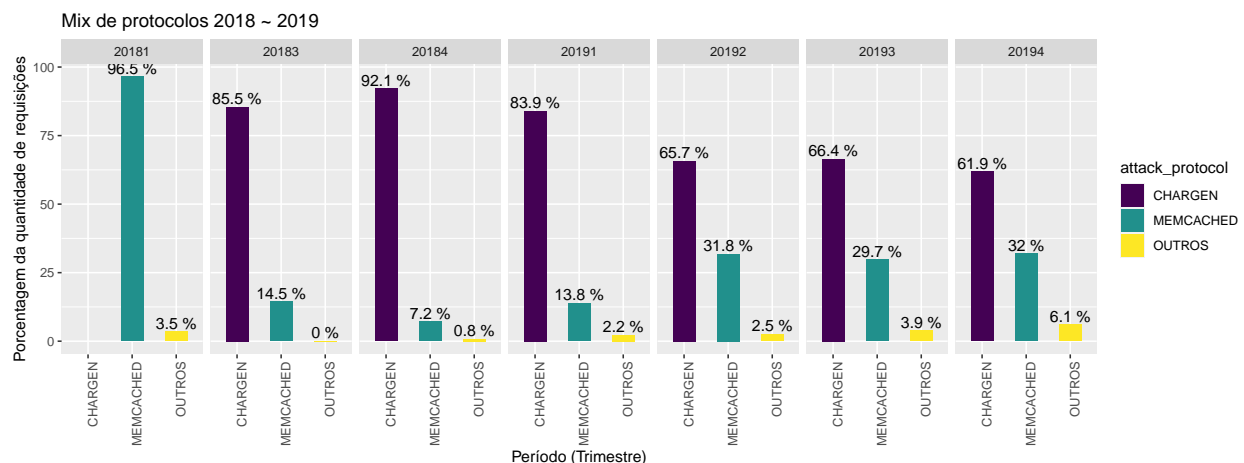
Porcentagem menores que 5 foram agrupadas como “OUTROS”

- Gráfico de barras 2018 e 2019

```

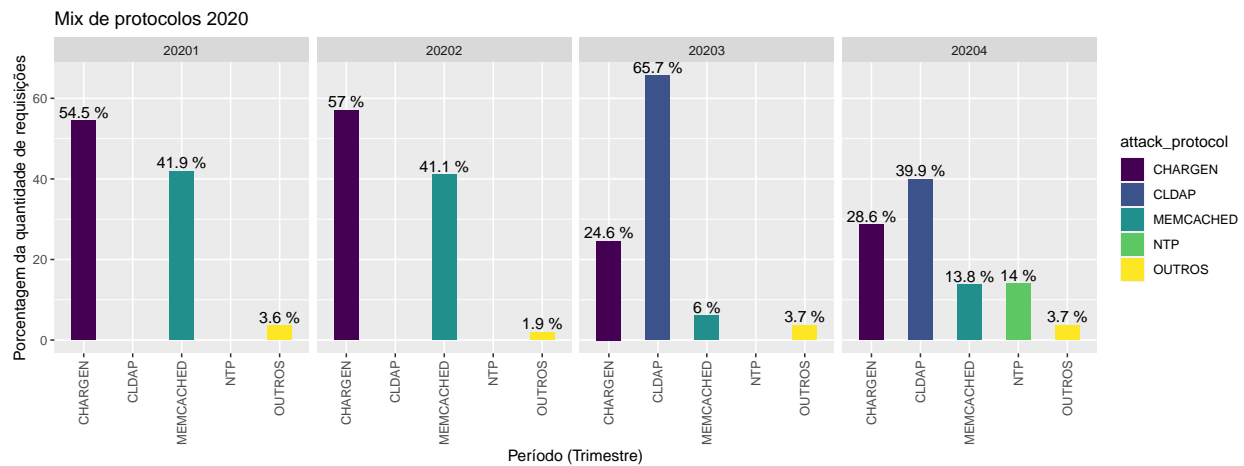
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2018_2019) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
  geom_bar(stat="identity", width = 0.5, position="dodge") +
  geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -1)) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
  facet_grid(~year_period) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2018 ~ 2019")

```



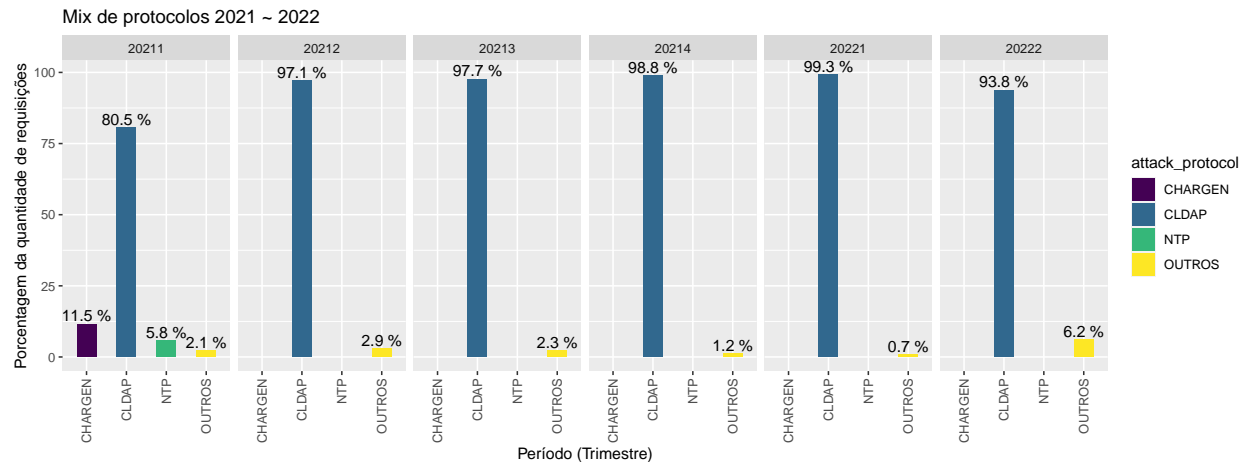
- Gráfico de barras 2020

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2020) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -0.5)) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de requisições") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2020")
```



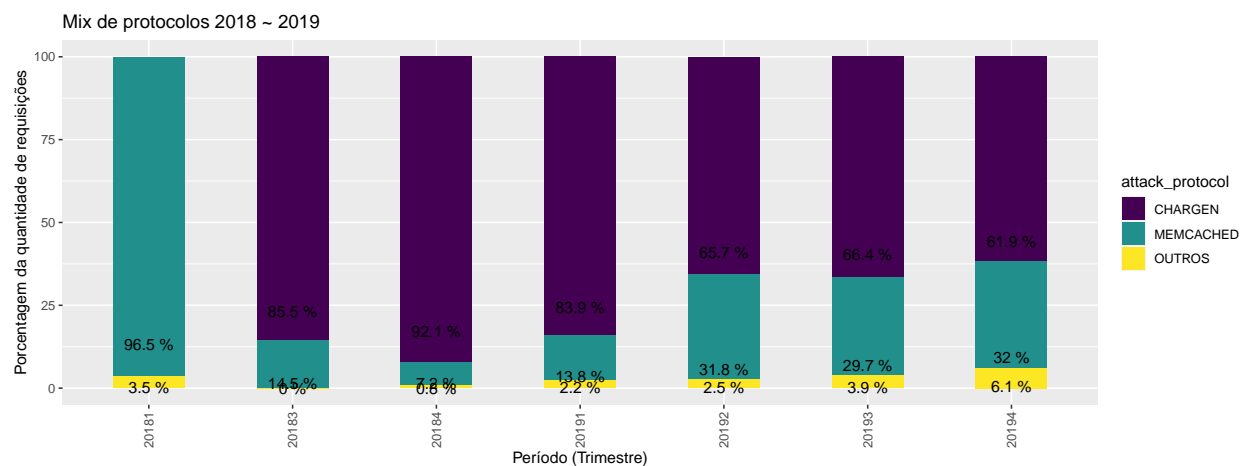
- Gráfico de barras 2021 ~ 2022

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2021_2022) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5, position="dodge") +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -0.5)) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    facet_grid(~year_period) +
    ylab("Porcentagem da quantidade de requisições") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2021 ~ 2022")
```



- Gráfico de barras empilhadas 2018 ~ 2019

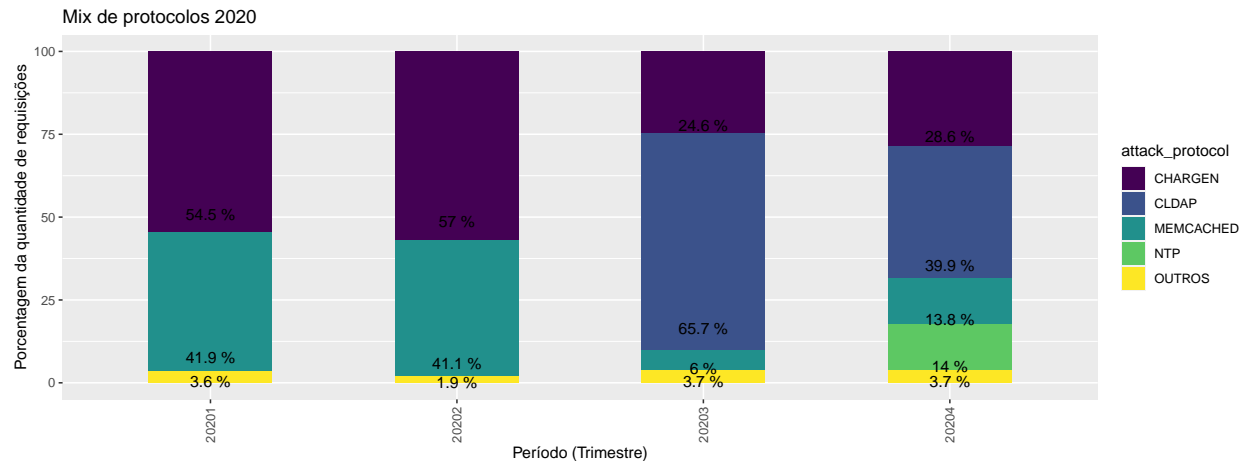
```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2018_2019) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")), position = "top", size = 10) +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de requisições") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2018 ~ 2019")
```



- Gráfico de barras empilhadas 2020

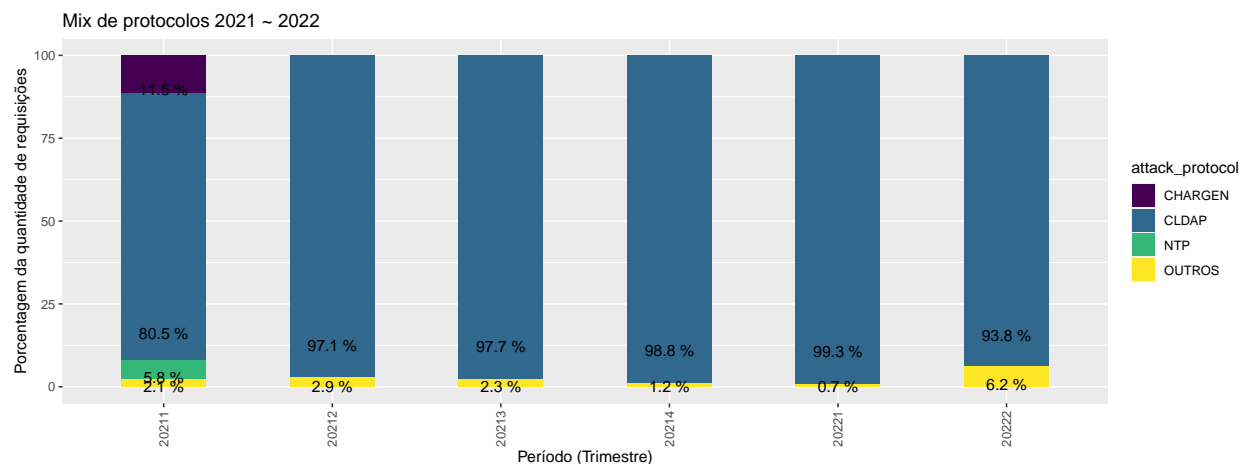
```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2020) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")), position = "top", size = 10) +
    scale_fill_viridis(discrete=TRUE) +
```

```
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
ylab("Porcentagem da quantidade de requisições") +
xlab("Período (Trimestre)") +
ggtitle("Mix de protocolos 2020")
```



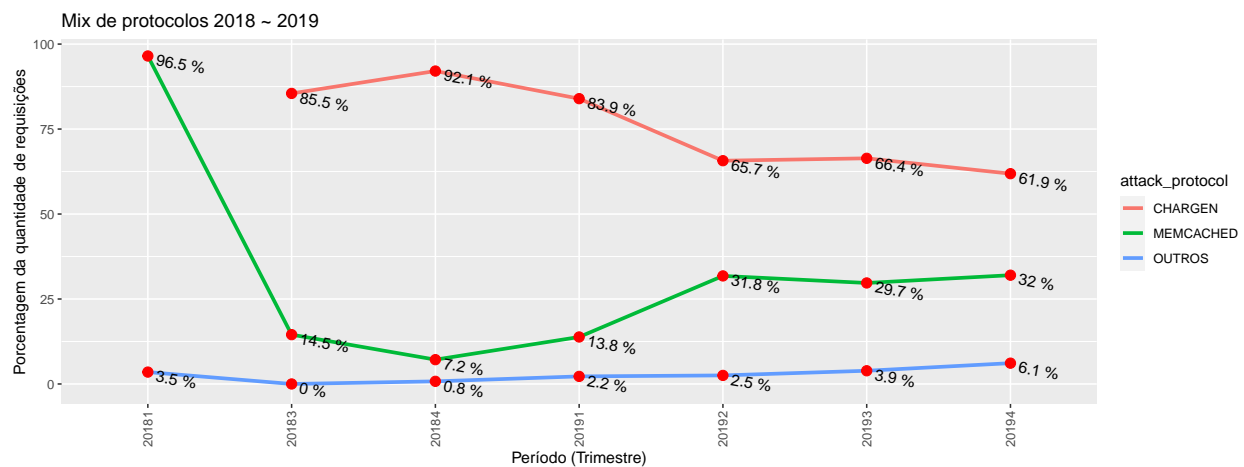
- Gráfico de barras empilhadas 2021 ~ 2022

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2021_2022) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, fill=attack_protocol)) +
    geom_bar(stat="identity", width = 0.5) +
    geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")), position = "top") +
    scale_fill_viridis(discrete=TRUE) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
    ylab("Porcentagem da quantidade de requisições") +
    xlab("Período (Trimestre)") +
    ggtitle("Mix de protocolos 2021 ~ 2022")
```



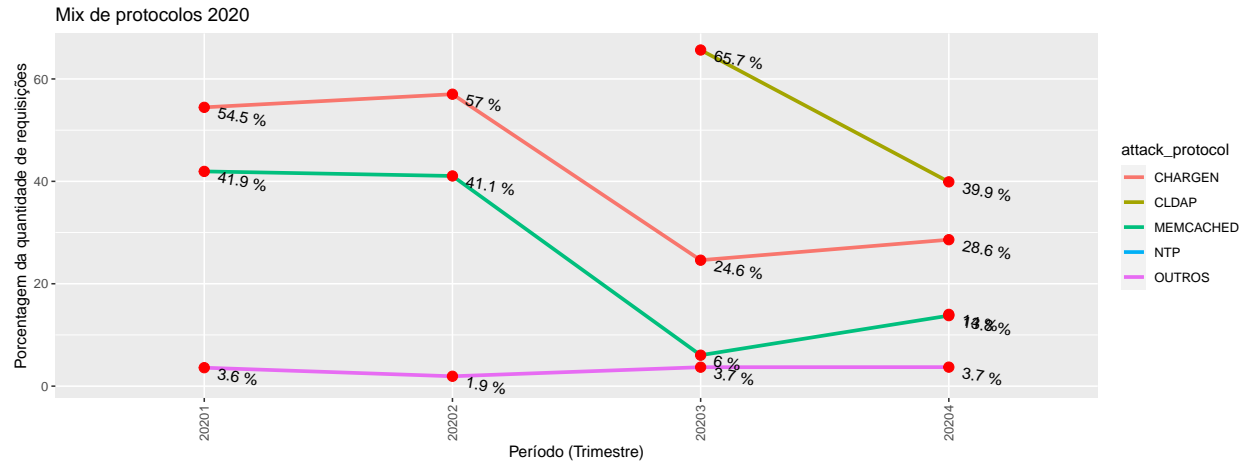
- Gráfico de linhas 2018 ~ 2019

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2018_2019) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, group=attack_protocol)) +
  geom_line(size=1.2, aes(color=attack_protocol)) +
  geom_point(color="red", size=3, aes(color=attack_protocol)) +
  geom_text(
    aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")),
    hjust = -0.03, nudge_x = 0.05, nudge_y = -1, angle = -10,
  ) +
  scale_fill_viridis(discrete=TRUE) +
  theme(
    axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
  ) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2018 ~ 2019")
```



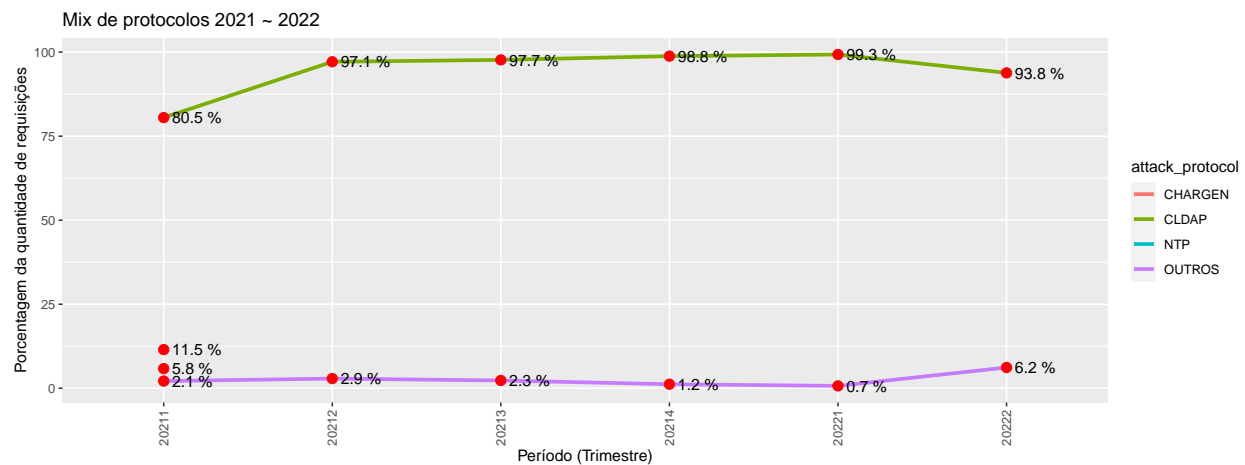
- Gráfico de linhas 2020

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2020) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, group=attack_protocol)) +
  geom_line(size=1.2, aes(color=attack_protocol)) +
  geom_point(color="red", size=3, aes(color=attack_protocol)) +
  geom_text(
    aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")),
    hjust = -0.03, nudge_x = 0.05, nudge_y = -1, angle = -10,
  ) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2020")
```



- Gráfico de linhas 2021 ~ 2022

```
data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2021_2022) %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, group=attack_protocol)) +
  geom_line(size=1.2, aes(color=attack_protocol)) +
  geom_point(color="red", size=3, aes(color=attack_protocol)) +
  geom_text(
    aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")),
    hjust = 0, nudge_x = 0.05,
  ) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2021 ~ 2022")
```



Todos os períodos no mesmo gráfico

- Barras


```

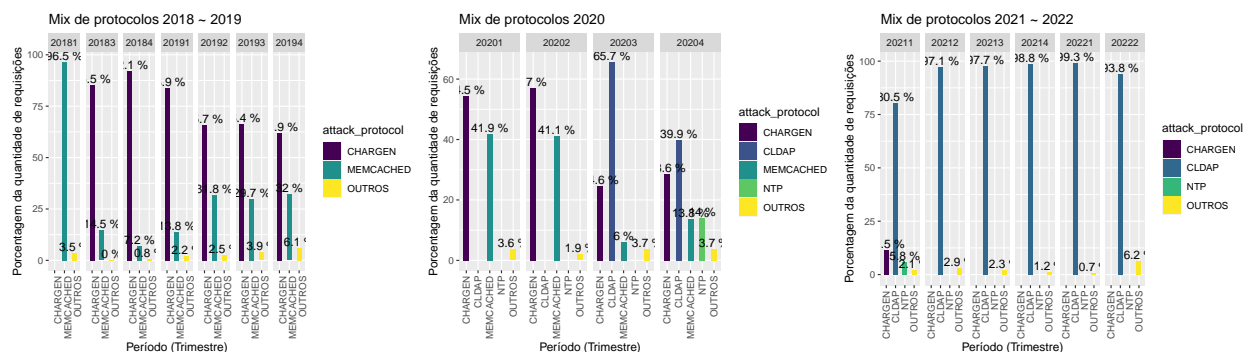
barras_2018_2019 = data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2018_2019) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
  geom_bar(stat="identity", width = 0.5, position="dodge") +
  geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -1)) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 1, hjust=1)) +
  facet_grid(~year_period) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2018 ~ 2019")

barras_2020 = data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2020) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
  geom_bar(stat="identity", width = 0.5, position="dodge") +
  geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -1)) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  facet_grid(~year_period) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2020")

barras_2021_2022 = data_grouped_period_protocol_others_percentage %>%
  filter(year_period %in% filter_in_period_2021_2022) %>%
  ggplot( aes(x=attack_protocol, y=number_of_requests_percentage, fill=attack_protocol)) +
  geom_bar(stat="identity", width = 0.5, position="dodge") +
  geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%"), vjust = -1)) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  facet_grid(~year_period) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2021 ~ 2022")

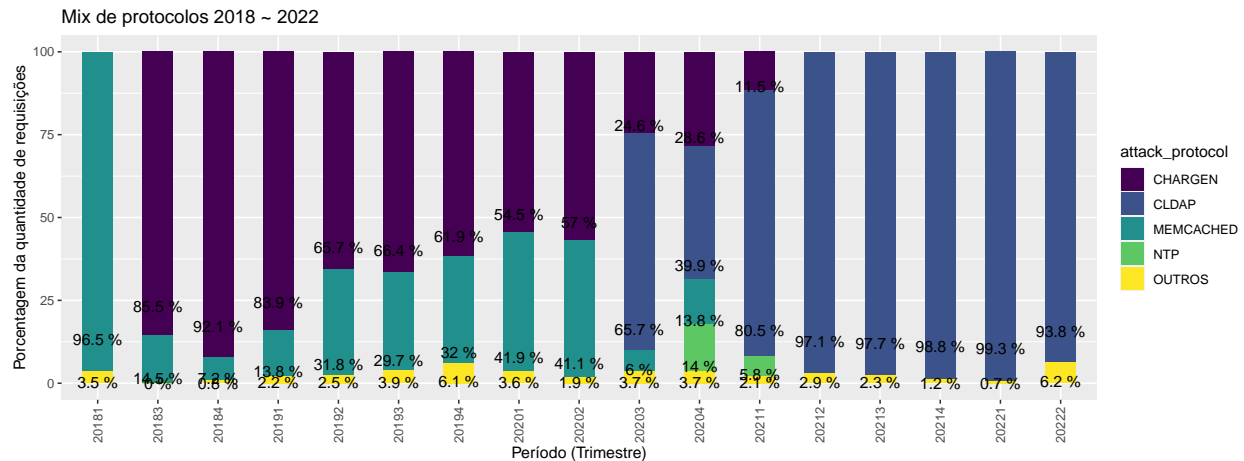
grid.arrange(barras_2018_2019, barras_2020, barras_2021_2022, ncol=3)

```



- Barras empilhadas

```
data_grouped_period_protocol_others_percentage %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, fill=attack_protocol)) +
  geom_bar(stat="identity", width = 0.5) +
  geom_text(aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")), position = "top",
    scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2018 ~ 2022")
```



- Gráfico de linhas

```
data_grouped_period_protocol_others_percentage %>%
  ggplot( aes(x=year_period, y=number_of_requests_percentage, group=attack_protocol)) +
  geom_line(size=1.2, aes(color=attack_protocol)) +
  geom_point(color="red", size=3, aes(color=attack_protocol)) +
  geom_text(
    aes(label = paste(round(number_of_requests_percentage, decimals_digits), "%")),
    hjust = 0, nudge_x = 0.05,
  ) +
  scale_fill_viridis(discrete=TRUE) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  ylab("Porcentagem da quantidade de requisições") +
  xlab("Período (Trimestre)") +
  ggtitle("Mix de protocolos 2018 ~ 2022")
```

