Aplicação de Criptografia Homomórfica na Mineração de Dados em Fluxos de Roteadores de Borda na Internet

Felipe M. F. Assis, Evandro L. C. Macedo,
Luís Felipe M. de Moraes

XXIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais
XVII Workshop de Trabalhos de Iniciação Científica e de Graduação (WTICG)

Oportunidades e Desafios

- **Grande quantidade de dados sendo gerada**
  - Mais de 95 zettabytes em 2022[1]
  - Oportunidade para Mineração de Dados

1 - Fonte: statista

Oportunidades e Desafios

- **Grande quantidade de dados sendo gerada**
  - Mais de 95 zettabytes em 2022[1]
  - Oportunidade para Mineração de Dados

- **Cresce a preocupação com privacidade**
  - LGPD
  - Não se pode fazer Mineração de Dados indiscriminadamente

1 - Fonte: statista

Solução? Criptografia

- **Criptografia homomórfica**
  - Permite operações no conjunto cifrado
  - Funções de Encriptação e Decriptação são homomorfismos
    - E(x)+E(y) = E(x+y)

Solução? Criptografia

- **Criptografia homomórfica**
  - Permite operações no conjunto cifrado
  - Funções de Encriptação e Decriptação são homomorfismos
    - $E(x)+E(y) = E(x+y)$

- **Criptografia de Limiar**
  - Chave privada distribuída por múltiplos participantes
    - Número mínimo necessário para decriptação!

Proposta

- **Mineração de Dados por meio da Criptografia**
  - Criação de Regras de Associação Distribuída
    - Conjunto de dados distribuído

Proposta

- **Mineração de Dados por meio da Criptografia**
  - Criação de Regras de Associação Distribuída
    - Conjunto de dados distribuído

- **Três métodos criados!**
  - Somente dois neste apresentação

Proposta

- **Mineração de Dados por meio da Criptografia**
  - Criação de Regras de Associação Distribuída
    - Conjunto de dados distribuído

- **Três métodos criados!**
  - Somente dois neste apresentação

- **Validado em fluxos de roteadores de borda**
  - Dados reais!

Regras de Associação

Quem viu, viu também:

Morango Orgânico 250g

R$15,99

Comprar

Banana Maçã Orgânica 1kg

R$10,99

R$10,99/Kg

Comprar

Exclusivo

Limão Taiti Orgânico Bio Vida 500g

R$8,99

Comprar

Melão Cantaloupe Orgânico 1kg

R$12,99

R$12,99/Kg

Comprar

Como são geradas?

Como são geradas?

As regras são do tipo $S_1 \Rightarrow S_2$

Como são geradas?

As regras são do tipo $S_1 \Rightarrow S_2$

$$Suporte_S = \frac{número\ de\ transações\ onde\ o\ conjunto\ S\ aparece}{número\ total\ de\ transações}$$

Como são geradas?

As regras são do tipo $S_1 \Rightarrow S_2$

$$Suporte_S = \frac{número\ de\ transações\ onde\ o\ conjunto\ S\ aparece}{número\ total\ de\ transações}$$

$$Confiança_{S_1 \Rightarrow S_2} = \frac{Suporte_{S_1 \cup S_2}}{Suporte_{S_1}} = \frac{n^o\ de\ transações\ onde\ S_1 \cup S_2\ aparece}{n^o\ de\ transações\ onde\ S_1\ aparece}$$

Suporte em um Ambiente Distribuído

$$Suporte_S = \frac{\text{número de transações onde o conjunto S aparece}}{\text{número total de transações}} = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} |DB_i|}$$

c_i : contagem de S do participante i

$|DB_i|$: tamanho da base de dados de i

s/100 : suporte mínimo escolhido

Suporte em um Ambiente Distribuído

$$Suporte_S = \frac{\text{número de transações onde o conjunto S aparece}}{\text{número total de transações}} = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} |DB_i|}$$

c_i : contagem de S do participante i

$|DB_i|$: tamanho da base de dados de i

s/100 : suporte mínimo escolhido

$$Suporte_S \geq s/100$$

$$\frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} |DB_i|} \geq s/100$$

$$100 \sum_{i=1}^{n} c_i \geq s \sum_{i=1}^{n} |DB_i|$$

$$\sum_{i=1}^{n} 100c_i - s|DB_i| \geq 0$$

Confiança em um Ambiente Distribuído

$$\text{Confiança}_{S_1 \Rightarrow S_2} = \frac{\text{nº de transações onde o conjunto } S_1 \cup S_2 \text{ aparece}}{\text{nº de transações onde o conjunto } S_1 \text{ aparece}} = \frac{\sum_{i=1}^{n} l_i}{\sum_{i=1}^{n} L_i}$$

l_i : contagem de $S_1 U S_2$ do participante i

L_i : l_i : contagem de S_1 do participante i

c/100 : confiança mínima escolhida

Confiança em um Ambiente Distribuído

$$\text{Confiança}_{S_1 \Rightarrow S_2} = \frac{\text{nº de transações onde o conjunto } S_1 \cup S_2 \text{ aparece}}{\text{nº de transações onde o conjunto } S_1 \text{ aparece}} = \frac{\sum_{i=1}^{n} l_i}{\sum_{i=1}^{n} L_i}$$

l_i : contagem de $S_1 \cup S_2$ do participante i

L_i : l_i : contagem de S_1 do participante i

c/100 : confiança mínima escolhida

$$\text{Confiança}_{S_1 \Rightarrow S_2} \geq c/100$$

$$\frac{\sum_{i=1}^{n} l_i}{\sum_{i=1}^{n} L_i} \geq c/100$$

$$100 \sum_{i=1}^{n} l_i \geq c \sum_{i=1}^{n} L_i$$

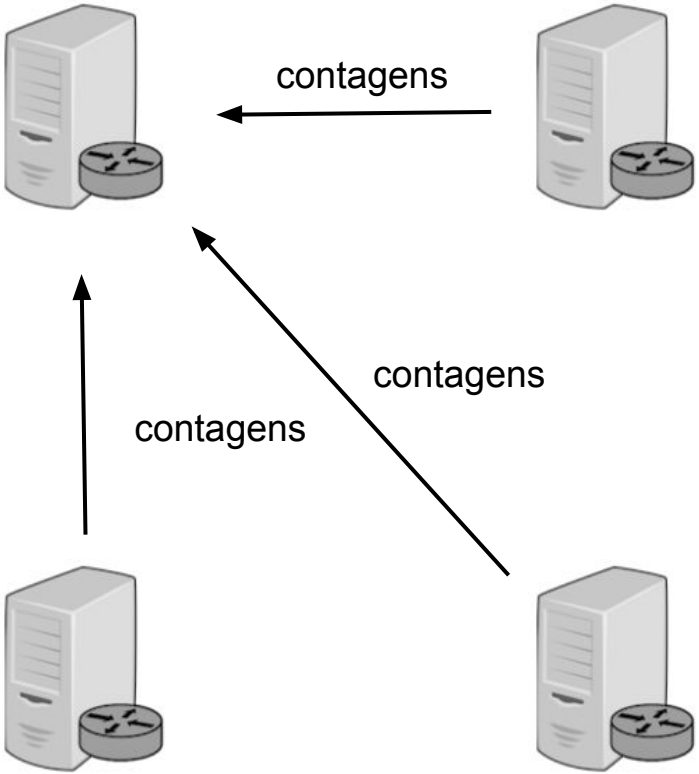
$$\sum_{i=1}^{n} 100 l_i - c L_i \geq 0$$

Métodos Propostos

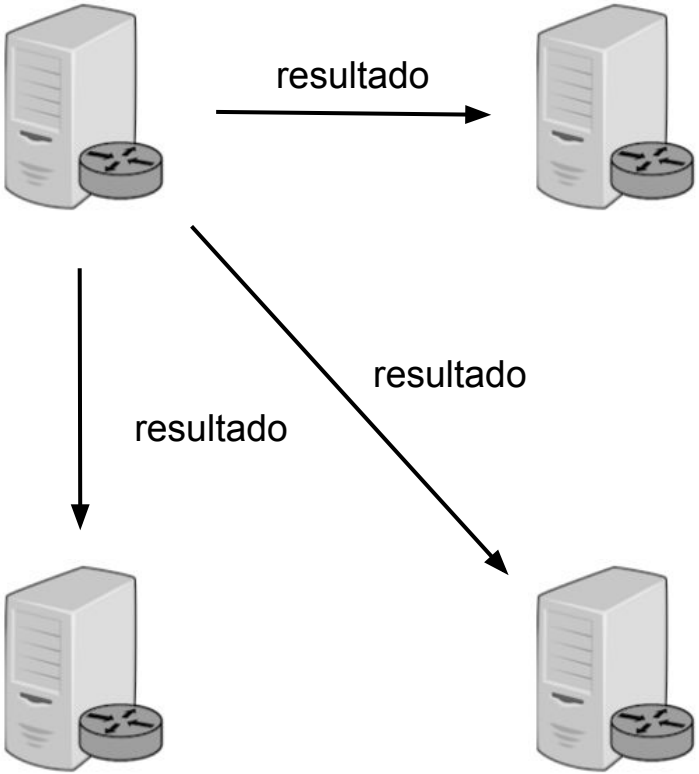
- **Usam OpenFHE**
  - Esquema Brakerski-Fan-Vercautere (BFV)

- **Mesma entrada e saída**
  - Entrada: frequência de cada conjunto
  - Saída: Regras de Associação
  - Método Padrão como base

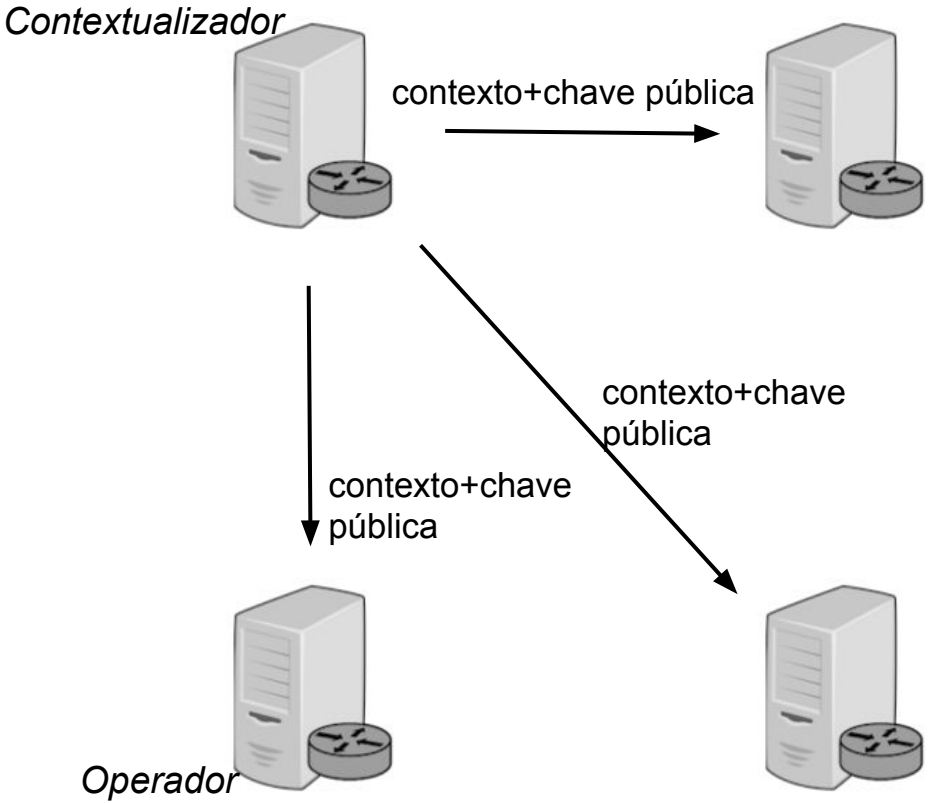
Comunicação no Método Padrão



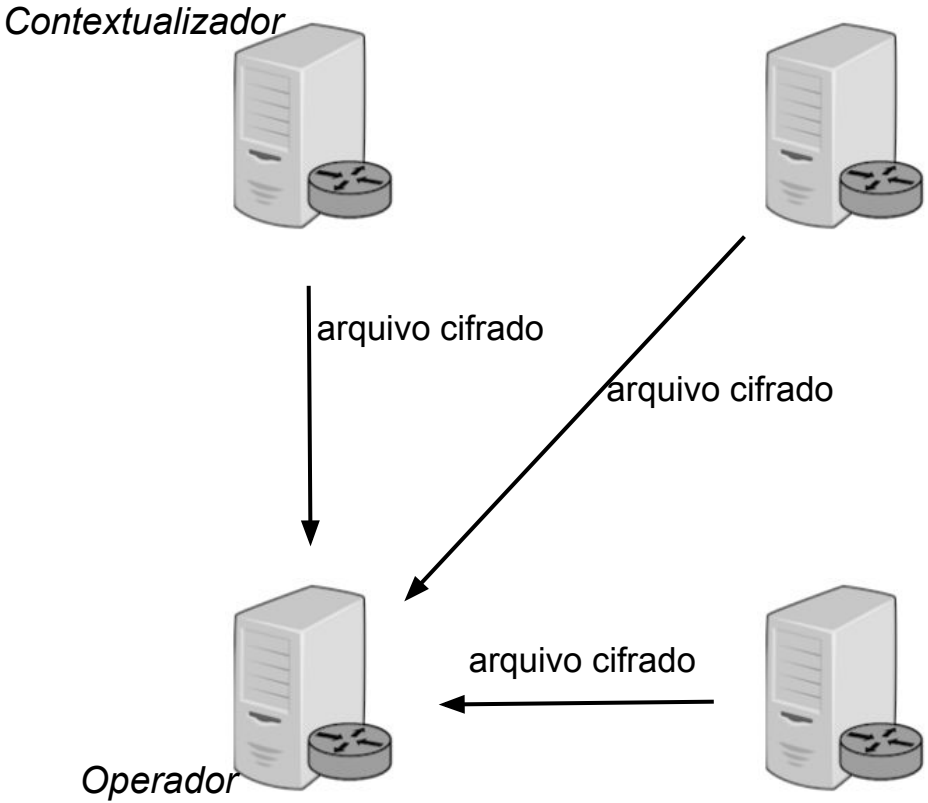
Comunicação no Método Padrão



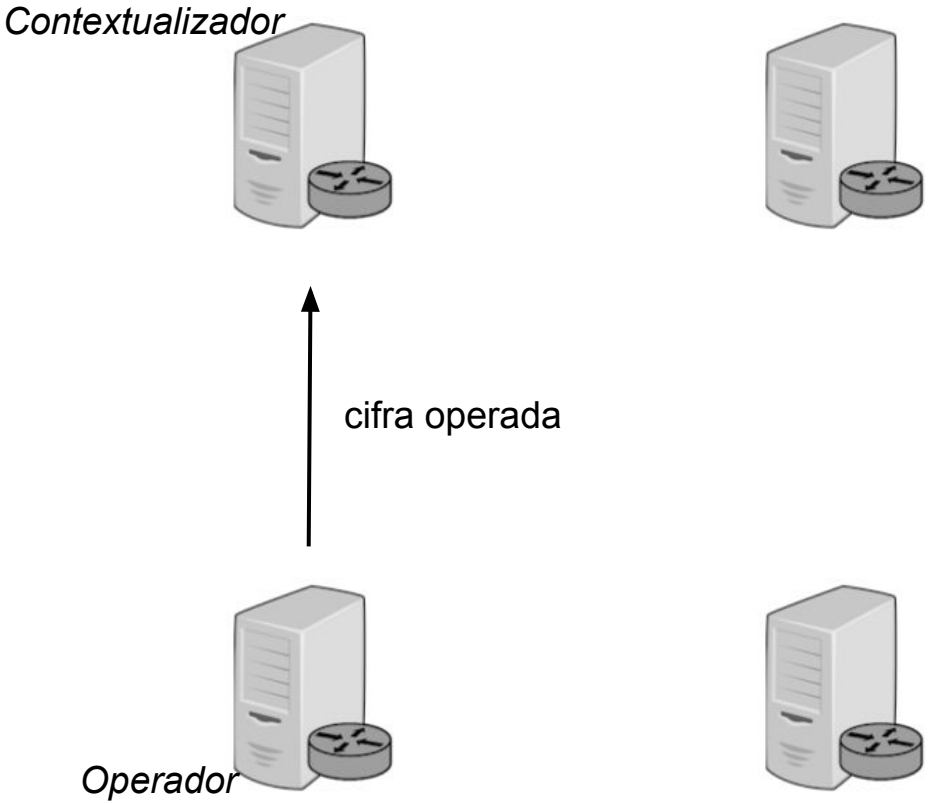
Método 1: Criptografia com Soma e Produto



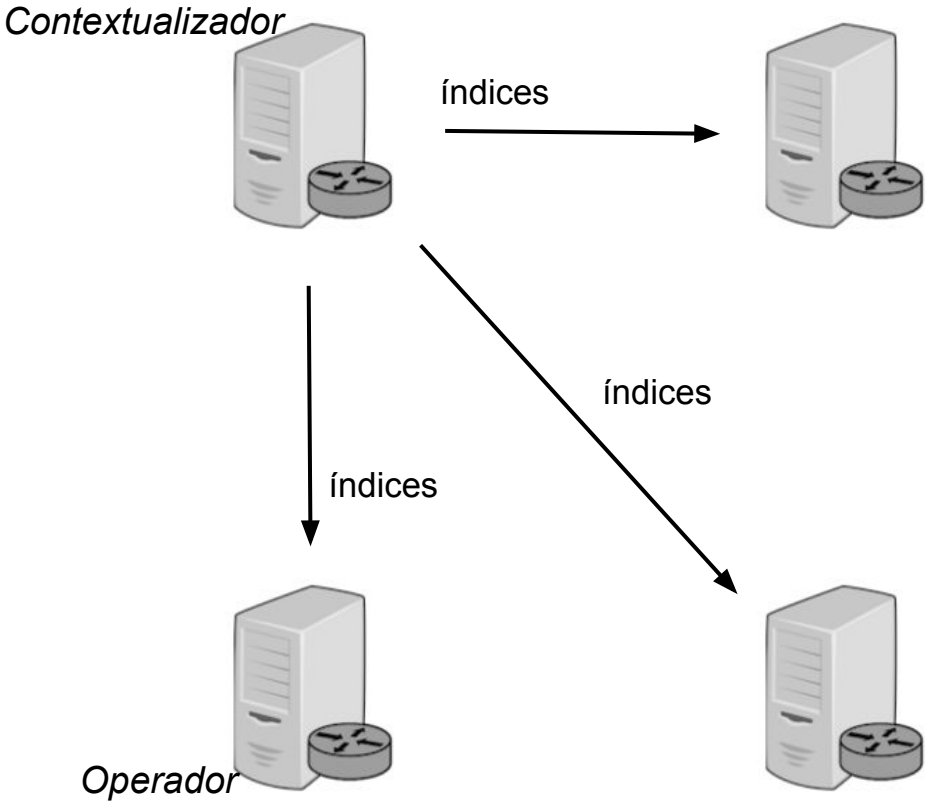
Comunicação no Método 1



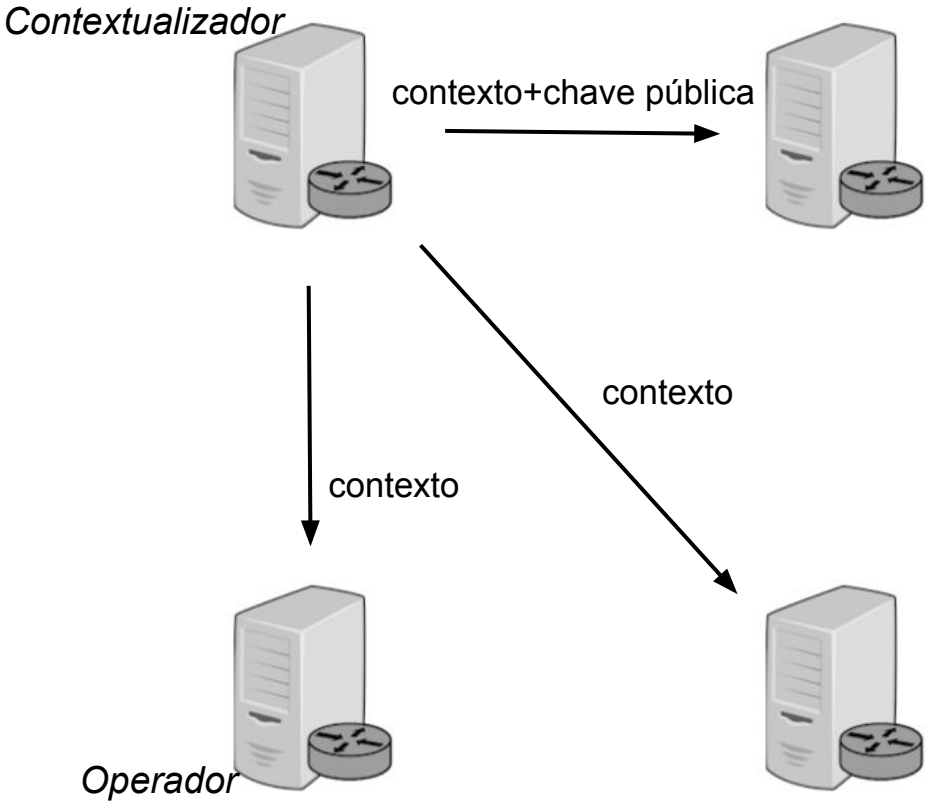
Comunicação no Método 1



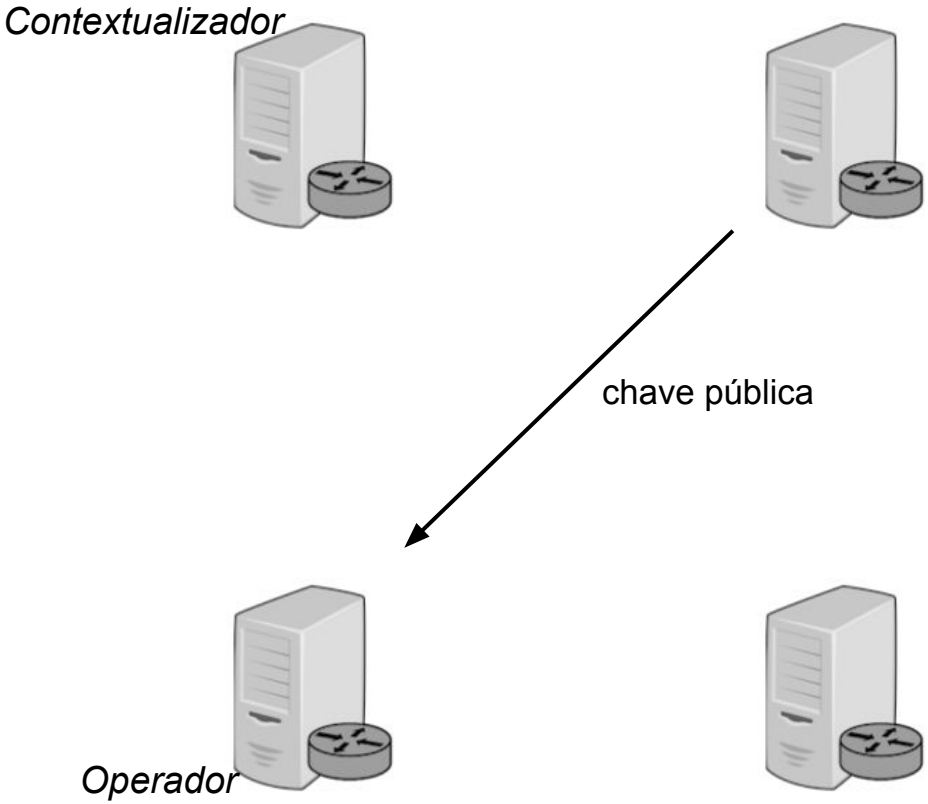
Método 1: Criptografia com Soma e Produto



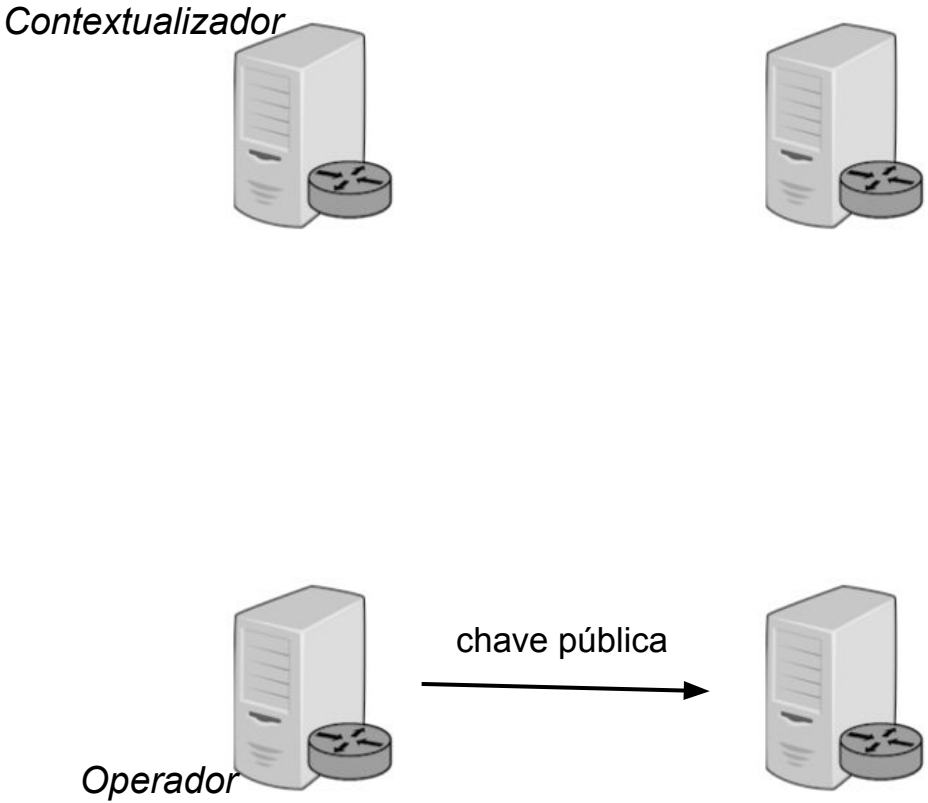
Método 2: Criptografia de Limiar com Soma



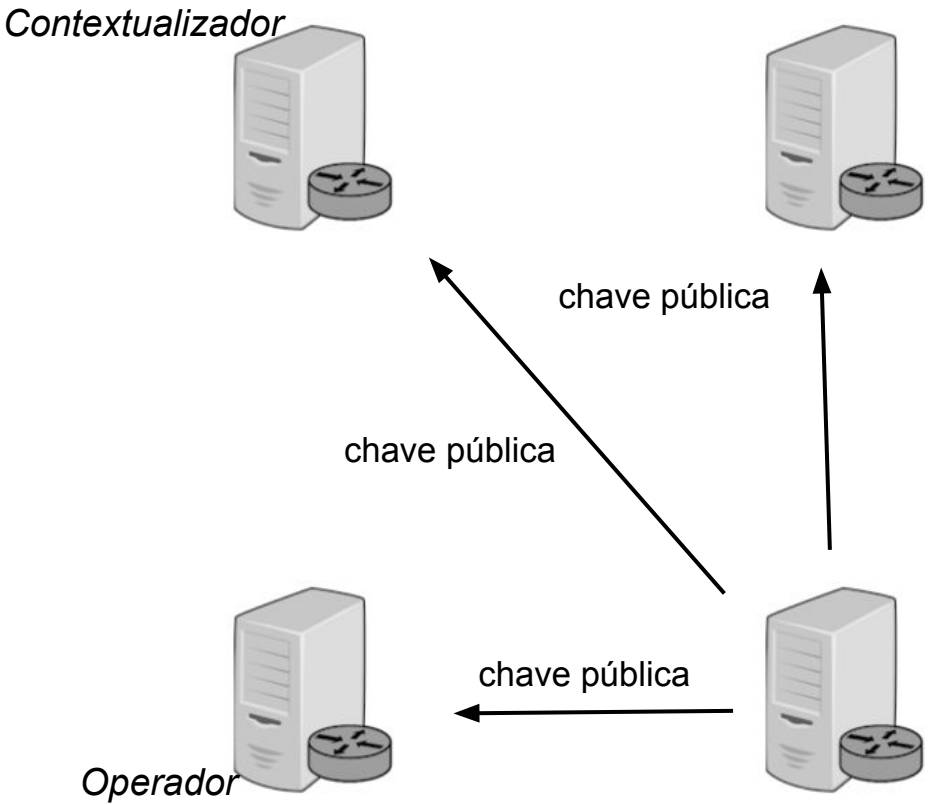
Comunicação no Método 2



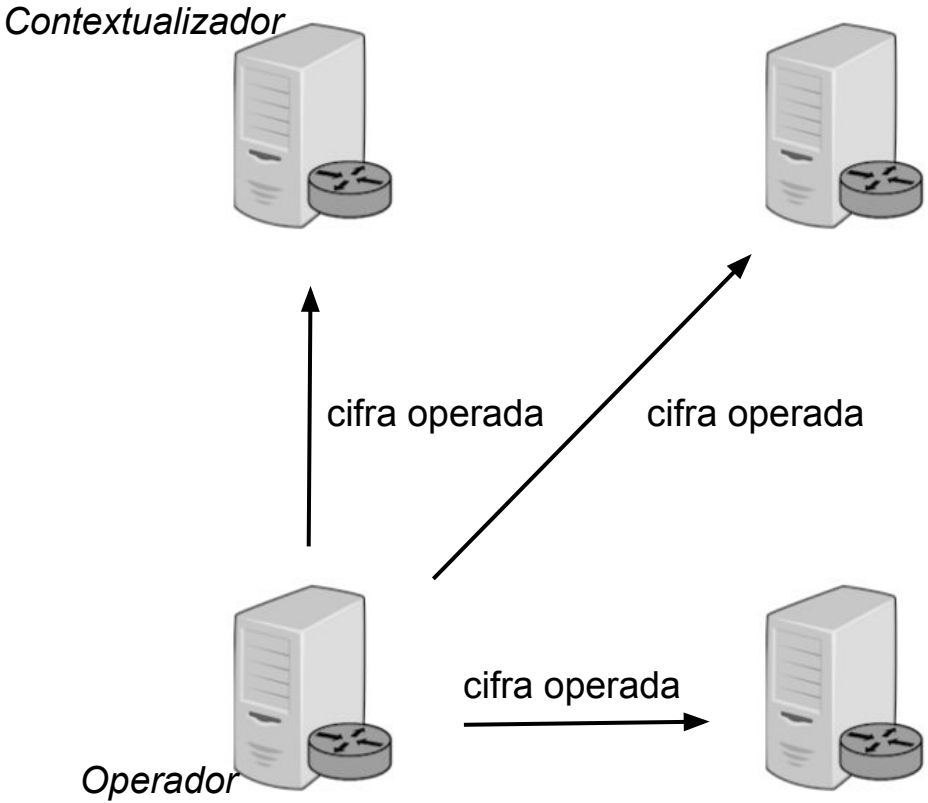
Comunicação no Método 2



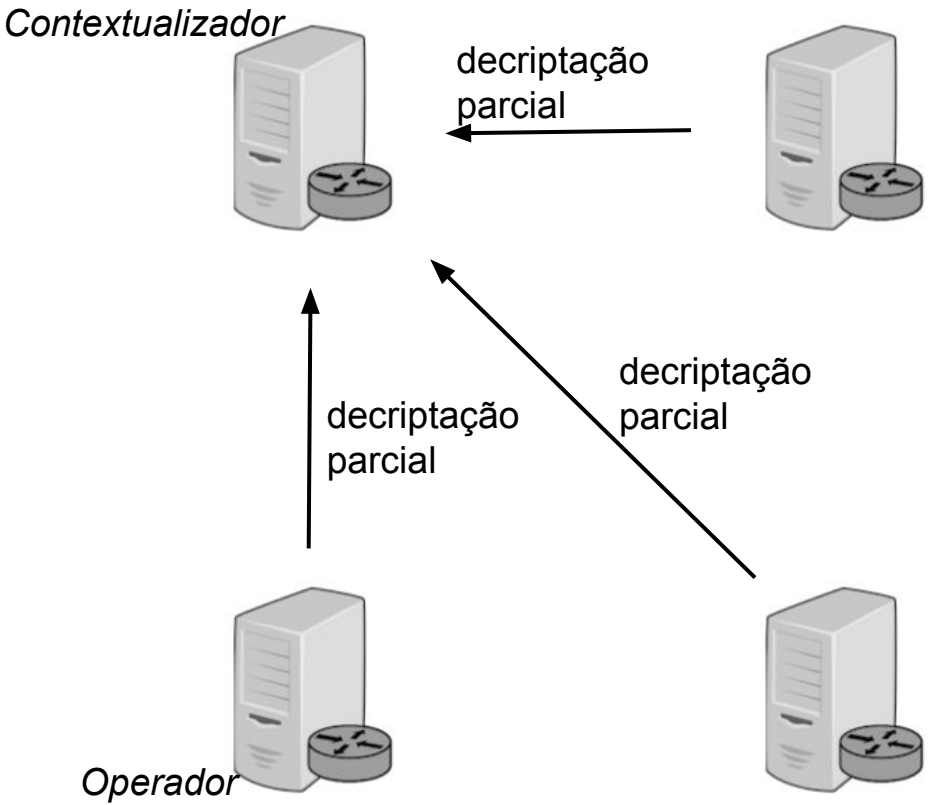
Comunicação no Método 2



Comunicação no Método 2



Comunicação no Método 2



Dados Utilizados

- **Testes utilizando fluxos de roteadores de borda da Rede-Rio**
  - Mais de 20GB, em arquivos de 5 minutos

- **Campos tratados e fluxos distribuídos em 4 conjuntos**
  - Agrupar para Regras de Associação
  - Distribuir para simular ambiente distribuído

- **Resultados**
  - Nesta versão: apenas uma baseline
  - Versão completa: comparação completa

Conclusão e Trabalhos Futuros

- **Métodos construídos com sucesso**
  - Formam a mesma saída que a padrão
  - Proveem privacidade

- **Muito mais na versão final**
  - Mais um método, combinando os anteriores
  - Maior análise de segurança
  - Análise prática completa

Universidade Federal do Rio de Janeiro Politécnica UFRJ RAVEL FAPERJ RedeRio FAPERJ Engenharia de Sistemas e Computação PESC/Coppe

Obrigado!

assis@ravel.ufrj.br
evandro@ravel.ufrj.br
moraes@ravel.ufrj.br