Ways to Anonymize IP Addresses

The too-short version

Paul Hoffman

DNS-OARC 28 Lightning talks
09 March 2018

What

- You have IP addresses you want to anonymize
- They are both IPv4 and IPv6
- You think this is easy
- It *is* easy, but there are a lot of choices
- This lightning talk **is not enough** to understand the pros and cons

The yet-to-be-published RSSAC document says

1. Cryptographically mix each new address seen with AES-128 and a random value, then truncate the output of IPv4 addresses to 32 bits
2. Cryptographically mix each new address with the Cryptopan method (mix bit-by-bit with a random value)
3. Cryptographically encrypt each new IPv4 address seen with ipcrypt and a random value, and encrypt each new IPv6 address with AES-128

- All of these require a large secret random number per "dataset"
  - If others discover the secret, you lose anonymization

Mix-and-truncate for IPv4, mix for IPv6

```
define mixing_function(input_address):
    set output_holder to aes-128(msg=input_address, key=dataset_random)
    if length(input_address) is 4 bytes:
        set truncated_value to leftmost 4 bytes of output_holder
        return truncated_value
    else:
        return output_holder
```

Cryptopan

```
define mixing_function(input_address):
    set output_holder to length of input_address (32 or 128 bits)
     for each bit_position in length(input_address):
        set this_message to input_address[0:bit_position] | dataset_random
        set this_bit to highest bit returned from aes-128(msg=this_message, key=0)
        set output_holder[bit_position] to this_bit
    return output_holder
```

ipcrypt

```
define mixing_function(input_address):
  if length(input_address) is 4 bytes:
    return ipcrypt(msg=input_address, key=dataset_random)
  else:
    return aes-128(msg=input_address, key=dataset_random)
```

- It turns out that this might expose the key much faster than expected

Advantages and Disadvantages

- They are different for people anonymizing and people consuming the data
- Costs of someone de-anonymizing are hard to predict
- Cost of collisions in the addresses are hard to predict

Engage with ICANN

Thank You and Questions

Visit us at **icann.org**
Email: email

 @icann

 facebook.com/icannorg

 youtube.com/icannnews

 flickr.com/icann

 linkedin/company/icann

 slideshare/icannpresentations

 soundcloud/icann