DATA NEEDS AND FEEDS FOR FIGHTING DNS ABUSE

Raffaele Sommese - University of Twente

r.sommese@utwente.nl

IN THE PREVIOUS FALL

DarkDNS: Revisiting the Value of Rapid Zone Update

Raffaele Sommese
University of Twente
Enschede, The Netherlands
r.sommese@utwente.nl

Gautam Akiwate
Stanford University
Stanford, CA, USA
gakiwate@cs.stanford.edu

Antonia Affinito
University of Twente
Enschede, The Netherlands
a.affinito@utwente.nl

Moritz Müller
SIDN Labs
Arnhem, The Netherlands
University of Twente
Enschede, The Netherlands
moritz.muller@sidn.nl

Mattijs Jonker
University of Twente
Enschede, The Netherlands
m.jonker@utwente.nl

KC Claffy
CAIDA
San Diego, CA, USA
kc@caida.org

Abstract

Malicious actors exploit the DNS namespace to launch spam campaigns, phishing attacks, malware, and other harmful activities. Combating these threats requires visibility into domain existence, ownership and nameservice activity that the DNS protocol does not itself provide. To facilitate visibility and security-related study of the expanding gTLD namespace, ICANN introduced the Centralized Zone Data Service (CZDS) that shares daily zone file snapshots

1 Introduction

Malicious actors exploit (abuse) the DNS namespace to launch spam campaigns, phishing attacks, malware, and other harmful activities. In many dimensions the DNS ecosystem is more opaque than other aspects of Internet transport. Unlike BGP, DNS is a pull protocol, so learning internal dynamics requires an entry point *i.e.,* a domain name. Without knowing this entry point, any abuse that lies behind a domain remains opaque to everyone except the targets.

TRANSIENT DOMAINS?!

By leveraging CT Logs (stream) we demonstrated that:

- We detected 42% of newly registered domains before they appeared in the CZDS snapshot (Nov 2023 – Jan 2024).
- ~76K domains per day - - almost 1 domain per second.
- 1% of newly registered domains never appear in the next CZDS snapshot.
- Predominately malicious!
- With half of them died within their first 6 hours of life.

TRANSIENT DOMAINS?!

By leveraging CT Logs (stream) we demonstrated that:

- We detected 42% of newly registered domains before they appeared in the CZDS snapshot (Nov 2023 – Jan 2024).
- ~76K domains per day - - almost 1 domain per second.
- 1% of newly registered domains never appear in the next CZDS snapshot.
- Predominately malicious!
- With half of them died within their first 6 hours of life.

TRANSIENT DOMAINS?!

By leveraging CT Logs (stream) we demonstrated that:

- We detected 42% of newly registered domains before they appeared in the CZDS snapshot (Nov 2023 – Jan 2024).
- ~76K domains per day - - almost 1 domain per second.
- 1% of newly registered domains never appear in the next CZDS snapshot.
- Predominately malicious!
- With half of them died within their first 6 hours of life.

FROM RESEARCH TO OPERATION

- The paper focused on detecting transient domains in a post-mortem scenario.
- Being able to block them quicker can help against attackers who may leverage caching.
- How can we detect them as soon as they die?
  - Measuring newly registered domains* (every 10 min, for 48h. A/AAAA/NS (@TLDs)/MX).
  - Detect deletion: 3x NXDOMAIN at TLD Level.
  - Issuing 2 RDAP requests, one when the domain is first detected, one when the domain is marked as deleted.
  - Checking RDAP responses for status.

FROM RESEARCH TO OPERATION

- The paper focused on detecting transient domains in a post-mortem scenario.
- Being able to block them quicker can help against attackers who may leverage caching.
- How can we detect them as soon as they die?
  - Measuring newly registered domains* (every 10 min, for 48h. A/AAAA/NS (@TLDs)/MX).
  - Detect deletion: 3x NXDOMAIN at TLD Level.
  - Issuing 2 RDAP requests, one when the domain is first detected, one when the domain is marked as deleted.
  - Checking RDAP responses for status.

FROM RESEARCH TO OPERATION

- The paper focused on detecting transient domains in a post-mortem scenario.
- Being able to block them quicker can help against attackers who may leverage caching.
- How can we detect them as soon as they die?
  - Measuring newly registered domains* (every 10 min, for 48h. A/AAAA/NS (@TLDs)/MX).
  - Detect deletion: 3x NXDOMAIN at TLD Level.
  - Issuing 2 RDAP requests, one when the domain is first detected, one when the domain is marked as deleted.
  - Checking RDAP responses for status.

FROM RESEARCH TO OPERATION

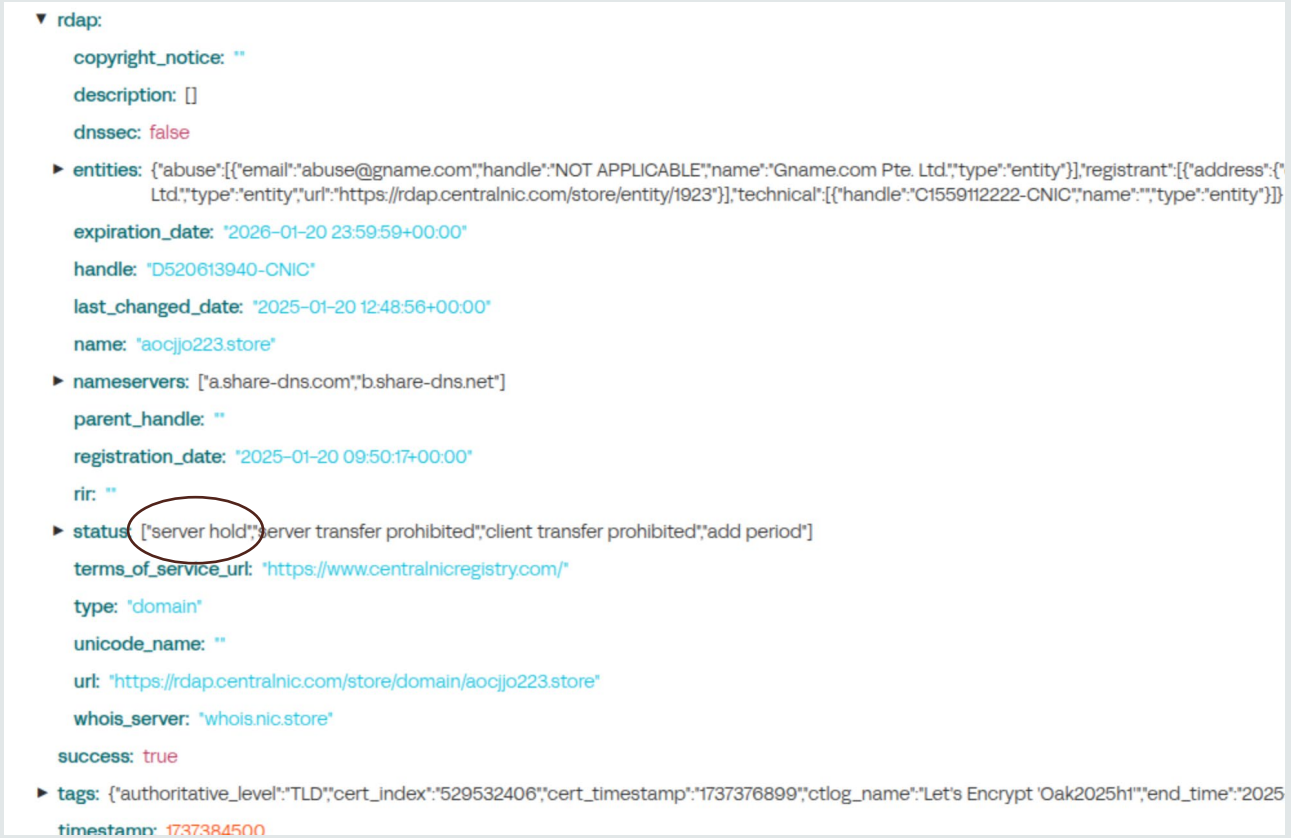
- The paper focused on detecting transient domains in a post-mortem scenario.
- Being able to block them quicker can help against attackers who may leverage caching.
- How can we detect them as soon as they die?
  - Measuring newly registered domains* (every 10 min, for 48h. A/AAAA/NS (@TLDs)/MX).
  - Detect deletion: 3x NXDOMAIN at TLD Level.
  - Issuing 2 RDAP requests, one when the domain is first detected, one when the domain is marked as deleted.
  - Checking RDAP responses for status.

. . . AND

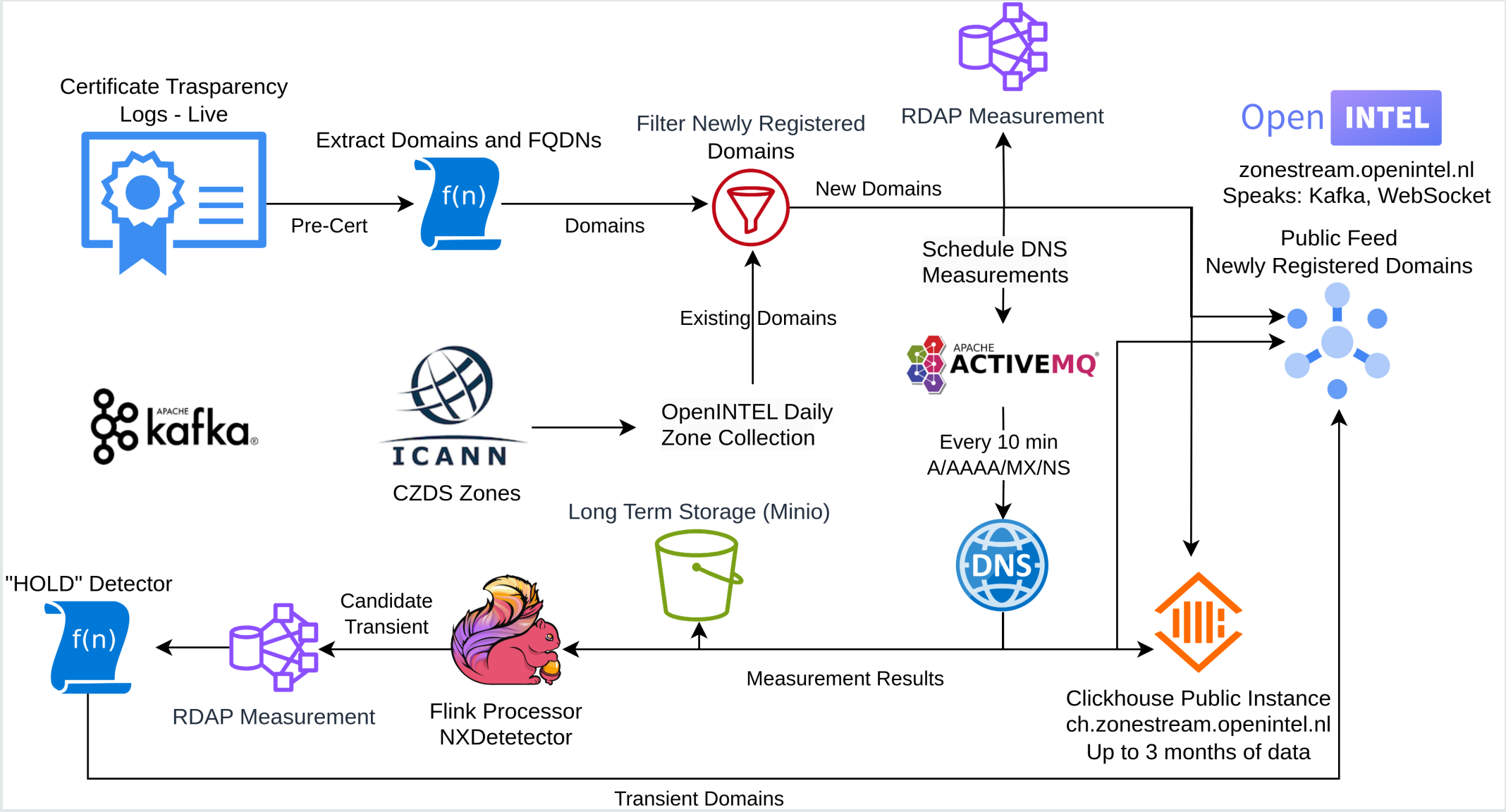
It seems someone (the registry) is holding those transient domains back.

Likely abuse!

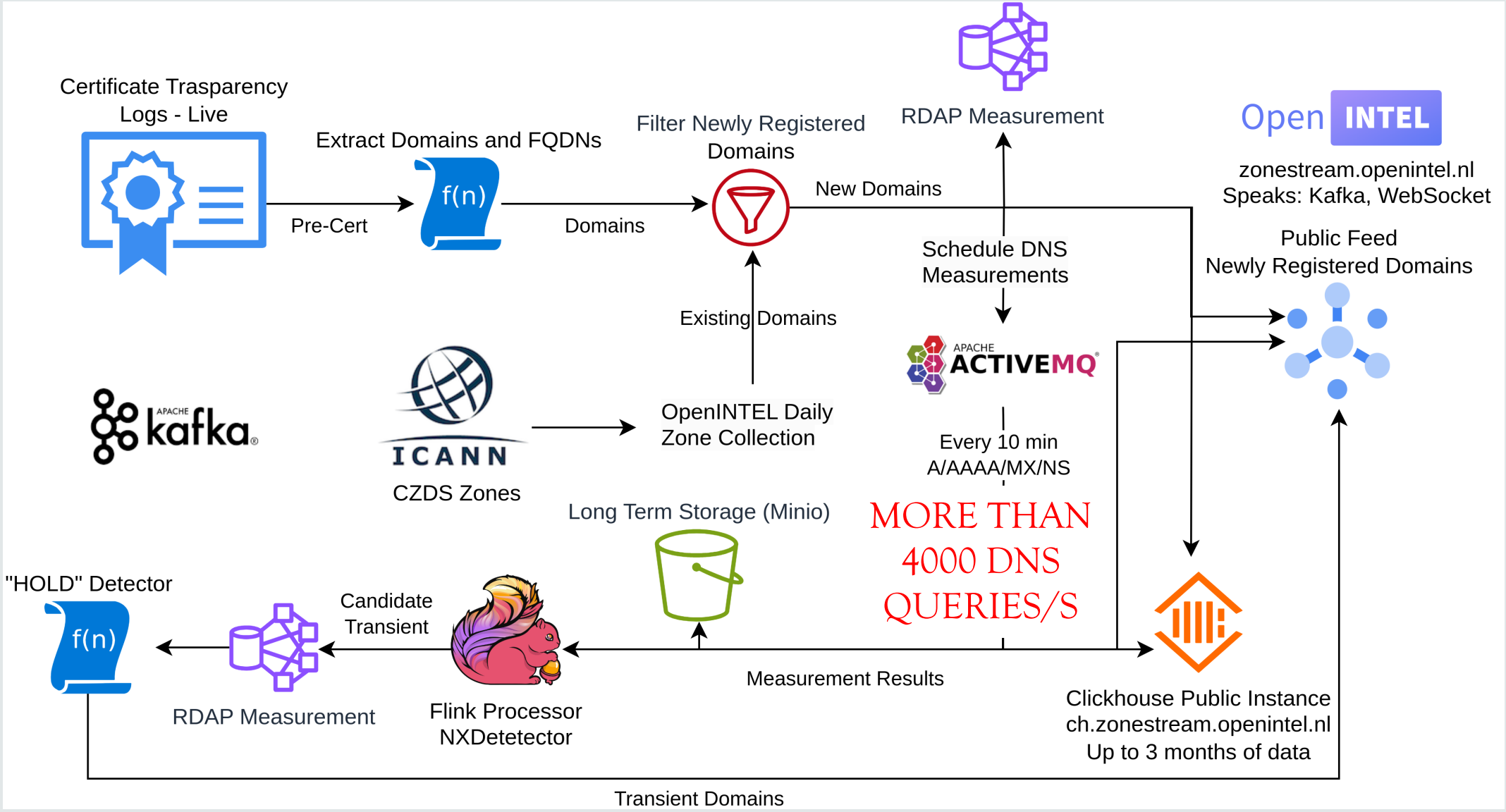
(Sometimes is client hold, registrar acting)



GETTING TO THIS....



GETTING TO THIS....



SOME NICE BY-PRODUCTS AS LIVE STREAM (AND CLICKHOUSE STORED DATA)

- newly_registered_domain – List of newly registered domains detected from CT Logs.
- confirmed_newly_registered_domain – Same of above, RDAP-Confirmed.
- newly_registered_fqdn – List of FDQNs of newly registered domains detected from CT Logs.
- newly_registered_domains_measurements – A/AAAA/NS/MX (every 10 min for 48h).
- newly_registered_fqdn_measurements – A/AAAA (every 10 min for 48h).
- certstream – Certificate Transparencty Logs stream (based on https://certstream.calidog.io/).
- certstream_domains – Stream of domain names learned from CT Logs.
- newly_issued_certificates_measurements - A/AAAA/NS for every domain name learned from CT Logs.

Publicly available as best-effort and for non-for-profit usage only.
How to access: https://kafka.zonestream.openintel.nl/

SOME NICE BY-PRODUCTS AS LIVE STREAM

- newly_registered_domain – List of newly registered domains detected from CT Logs.
- confirmed_newly_registered_domain – Same of above, RDAP-Confirmed.
- newly_registered_fqdn – List of FDQNs of newly registered domains detected from CT Logs.
- newly_registered_domains_measurements – A/AAAA/NS/MX (every 10 min for 48h).
- newly_registered_fqdn_measurements – A/AAAA (every 10 min for 48h).
- certstream – Certificate Transparencty Logs stream (based on https://certstream.calidog.io/).
- certstream_domains – Stream of domain names learned from CT Logs.
- newly_issued_certificates_measurements - A/AAAA/NS for every domain name learned from CT Logs.

Publicly available as best-effort and for non-for-profit usage only.
How to access: https://kafka.zonestream.openintel.nl/

SOME NICE BY-PRODUCTS AS LIVE STREAM

- newly_registered_domain – List of newly registered domains detected from CT Logs.
- confirmed_newly_registered_domain – Same of above, RDAP-Confirmed.
- newly_registered_fqdn – List of FDQNs of newly registered domains detected from CT Logs.
- newly_registered_domains_measurements – A/AAAA/NS/MX (every 10 min for 48h).
- newly_registered_fqdn_measurements – A/AAAA (every 10 min for 48h).
- certstream – Certificate Transparencty Logs stream (based on https://certstream.calidog.io/).
- certstream_domains – Stream of domain names learned from CT Logs.
- newly_issued_certificates_measurements - A/AAAA/NS for every domain name learned from CT Logs.

Publicly available as best-effort and for non-for-profit usage only.
How to access: https://kafka.zonestream.openintel.nl/

SOME NICE BY-PRODUCTS AS LIVE STREAM

- newly_registered_domain – List of newly registered domains detected from CT Logs.
- confirmed_newly_registered_domain – Same of above, RDAP-Confirmed.
- newly_registered_fqdn – List of FDQNs of newly registered domains detected from CT Logs.
- newly_registered_domains_measurements – A/AAAA/NS/MX (every 10 min for 48h).
- newly_registered_fqdn_measurements – A/AAAA (every 10 min for 48h).
- certstream – Certificate Transparencty Logs stream (based on https://certstream.calidog.io/).
- certstream_domains – Stream of domain names learned from CT Logs.
- newly_issued_certificates_measurements - A/AAAA/NS for every domain name learned from CT Logs.

Publicly available as best-effort and for non-for-profit usage only.
How to access: https://kafka.zonestream.openintel.nl/


ISN'T THERE A
BETTER WAY?

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

DNS: FROM PULL TO PUSH

- DNS is a *pull protocol!?* Right?
- (Rapid) Zone Updates? -> EPP
- A signal for any update in the zone, enabling a CT Logs-like system to detect changes.
- DNS-Transparency -> https://www.internetfire.org/projects/dns-transparency
- Why?
  - Detecting early removal of domain names.
    - Evicting from cache of public and private resolvers.
    - Revoking certificates of expired domains.
    - Build better ML models to detect future threats.
  - Identifying hijacking
  - ...add your use case that requires you monitoring continuously the DNS

WILL THIS BE ENOUGH TO MITIGATE ABUSE?

WILL THIS BE ENOUGH TO MITIGATE ABUSE?

DEFINITELY NOT!

WE MAY NEED TO GO BEYOND DNS

BEYOND DNS: WHERE? WHO? MONEY?

- Where?
  - Exposing the registrar information in a more public manner (e.g., in a DNS Transparency Log) will enable researchers and security firms to detect and identify bad registrars' behaviors (… someone said bad CAs identified by CT Logs?)
- Who?
  - GDPR complicate the matter veiling RDAP registrant information's behind privacy. But we need the ability to correlate behaviors of actors across different system, not of knowing their identity!
- How to follow the money?
  - Registrant may still lie about their identity, the (global, unique, anonymous) identifier should enable independent third-party, vetted security researcher, law enforcement to track the money flow (e.g., a hash linked to the payment method).

BEYOND DNS: WHERE? WHO? MONEY?

- Where?
  - Exposing the registrar information in a more public manner (e.g., in a DNS Transparency Log) will enable researchers and security firms to detect and identify bad registrars' behaviors (… someone said bad CAs identified by CT Logs?)
- Who?
  - GDPR complicate the matter veiling RDAP registrant information's behind privacy. But we need the ability to correlate behaviors of actors across different system, not of knowing their identity!
- How to follow the money?
  - Registrant may still lie about their identity, the (global, unique, anonymous) identifier should enable independent third-party, vetted security researcher, law enforcement to track the money flow (e.g., a hash linked to the payment method).

BEYOND DNS: WHERE? WHO? MONEY?

- Where?
  - Exposing the registrar information in a more public manner (e.g., in a DNS Transparency Log) will enable researchers and security firms to detect and identify bad registrars' behaviors (… someone said bad CAs identified by CT Logs?)
- Who?
  - GDPR complicate the matter veiling RDAP registrant information's behind privacy. But we need the ability to correlate behaviors of actors across different system, not of knowing their identity!
- How to follow the money?
  - Registrant may still lie about their identity, the (global, unique, anonymous) identifier should enable independent third-party, vetted security researcher, law enforcement to track the money flow (e.g., a hash linked to the payment method).

WILL THIS SOLVE THE PROBLEM OF ABUSE?

- Probably not!
- Malicious actor may still evade detection by using fake identities.
- But we will make their life harder (and ours easier)!
- We need a zone of trust where those information can be shared and design a mechanism to enable anonymization of those identifiers without violating the privacy of the users.

WILL THIS SOLVE THE PROBLEM OF ABUSE?

- Probably not!
- Malicious actor may still evade detection by using fake identities.
- But we will make their life harder (and ours easier)!
- We need a zone of trust where those information can be shared and design a mechanism to enable anonymization of those identifiers without violating the privacy of the users.

- Or we can keep playing the cat and the mouse game!

QUESTIONS?

Raffaele Sommese
r.sommese@utwente.nl

https://zonestream.openintel.nl
https://kafka.zonestream.openintel.nl/

Open INTEL

DACS

UNIVERSITY OF TWENTE.