UNCLASSIFIED

Secure and confidential rule matching

February 2020

Communications Security Establishment
Centre de la sécurité des télécommunications

Canada

Outline

- Context
- Long term vision
- The challenge
- Questions

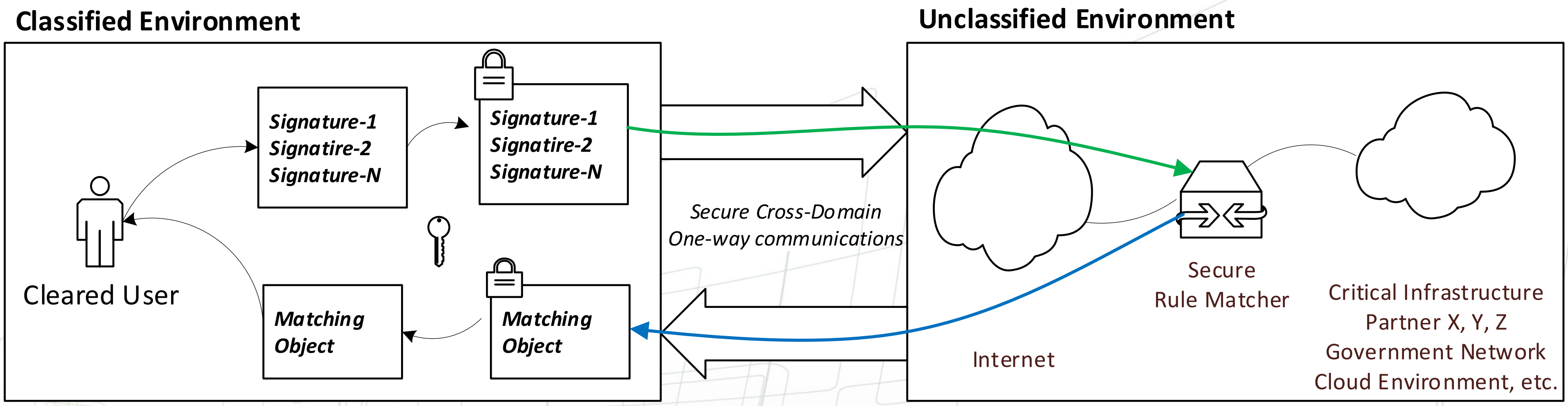
Context

- The security and intelligence (S&I) community have access to sensitive cyber threat information that is not always sharable publically
- A portion of this information can be encoded with enough precision to identify the threat actors by detecting their cyber modus-operandi in observing their presence in network traffic and system telemetry
- This sensitive information will be classified (at least for a period of time)
- We would like to allow the provisioning of classified cyber security signatures in appliances that could be deployed in unclassified networks such as Government or national critical infrastructure networks.

The long term vision

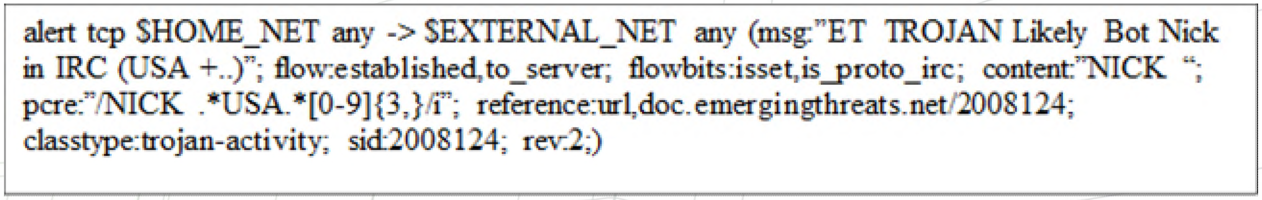
To be able to evaluate packet matching rules in unsecure environments without revealing the signatures or what network traffic is being matched by those signatures

Envisioned system (1)



Envisioned system (2)

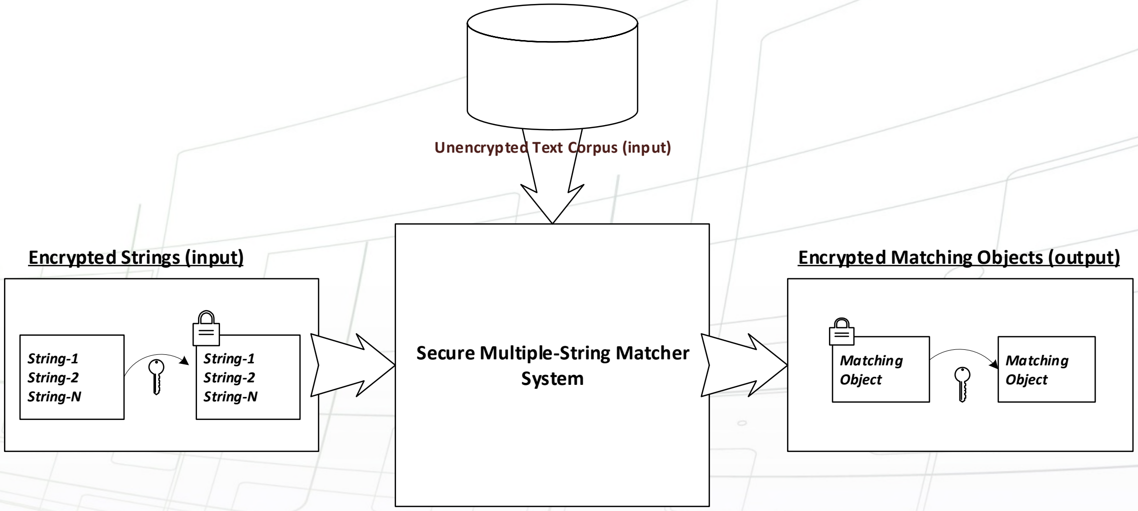
- Signatures are analogous to signature matching algorithms of popular open source intrusion detection systems (IDS), like Suricata or SNORT (https://suricata-ids.org, https://www.snort.org)



```
alert tcp $HOME_NET any -> $EXTERNAL_NET  any (msg:"ET  TROJAN Likely  Bot Nick
in IRC (USA +..)"; flow:established,to_server;  flowbits:isset,is_proto_irc;  content:"NICK  ";
pcre:"/NICK  .*USA.*[0-9]{3,}/i";  reference:url,doc.emergingthreats.net/2008124;
classtype:trojan-activity;  sid:2008124;  rev:2;)
```

The challenge

- The challenge is a simplified version of the ultimate envisioned system where signatures are simple strings and network traffic is a text corpus


Unencrypted Text Corpus (input)
Encrypted Strings (input)
String-1
String-2
String-N
String-1
String-2
String-N
Secure Multiple-String Matcher System
Encrypted Matching Objects (output)
Matching Object
Matching Object

Essential Outcomes

- Match a collection of simple rules on a corpus of unencrypted text
- Rules are simple character strings
- Keep rules confidential (encrypted) during matching process
- Keep it impossible to deduce the rules by analysing the execution of the instructions of the matching system at run time
- Keep matching objects confidential (encrypted)
- Signatures and corresponding matching objects are protected with a key only available to individuals with appropriate security clearance
- Rules are matched without errors, exactly as the system would run without encryption
- The solution fits in a reduced form factor equivalent to 4 unit spaces of a standard data centre rack

Additional Outcomes

- Scale to support a higher number of signatures (target is 20 000)
- Allow for more complex rule specification language (better than simple strings). E.g.:
  - Allow wild cards
  - Multi-criteria Boolean rules
  - Regular Expressions
- Match signatures on unencrypted packetized data (as opposed to unencrypted text corpus)
- Match 20 000 signatures at a rate of 1 Gbps of packetized network traffic
- Aim at an algorithm complexity of O(size-of-text + number_of_matching_occurences_in_corpus)

Anticipated Q & A

- Q: What security level do we require?
  - A: For confidentiality of information up to TOP SECRET, we require security equivalent to AES (See FIPS Pub 197) with 256-bit keys.”
- Q : What information should be returned when there is a match?
  - A: We need to know what signature matched where (offset in text, or packet IP 5-tuple)
- Q: What output data rate is acceptable?
  - A: We need to attempt to minimize the output data rate (i.e. the size of the matching objects). Batching can certainly be considered as a valid option when applying the secure rule matching process.
- Q: Will all the signatures be classified or just a subset?
  - A: All the signatures will be classified
- Q: What is the assumed input record size? (i.e. How many records per second in that Gb/s stream, or what’s the max length of a string being compared?)
  - A: In the case of packetized network traffic, it can be assumed that packets have a maximum size of 1500 bytes.

Questions?