Come bloccare i set di caratteri basati sul tipo di contenuto

Sommario

Introduzione

In questo documento viene descritto come scrivere e configurare un filtro per rilevare ed eseguire azioni sui set di caratteri basati sul tipo di contenuto in Cisco Email Security Appliance (ESA). Il documento seguente può essere usato per rilevare i caratteri basati su lingue straniere rilevati nei messaggi di posta indesiderata.

Premesse

Gli amministratori ESA possono ricevere un afflusso di messaggi di posta che contengono lingue straniere basate su caratteri che non sono messaggi legittimi per la loro società o il loro dominio. Un modo per risolvere il problema con l'ESA, abbiamo tre opzioni:

3. Scrivere un filtro utilizzando la condizione Message Language. Questa opzione è una nuova funzionalità di AsyncOS Email Security 10.0.0-203 e versioni successive.

Come bloccare i set di caratteri basati sul tipo di contenuto

Scrivi un filtro per rilevare il tipo di contenuto

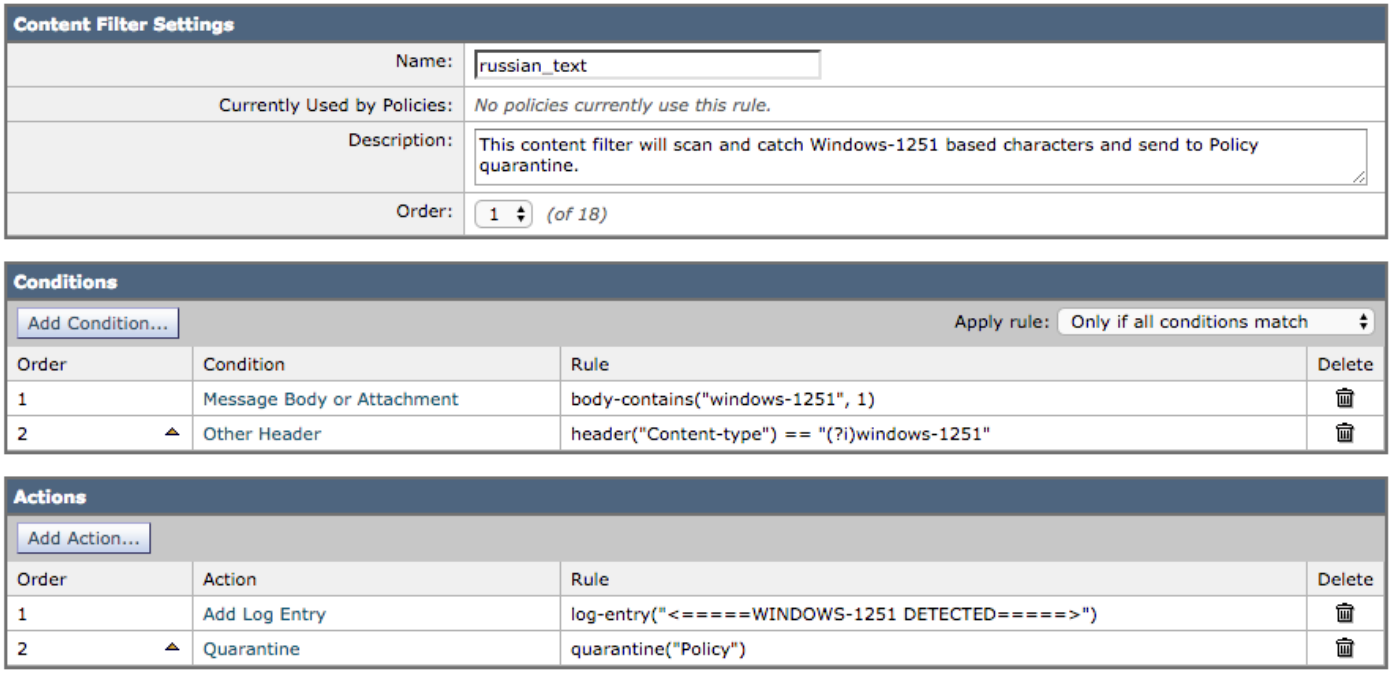
La prima opzione consente all'amministratore di scrivere e configurare un filtro e di associarlo a un criterio di posta, in base alle esigenze.

> **Nota**: scrivere e configurare questo filtro come filtro messaggi potrebbe essere dispendioso in termini di risorse per poter analizzare il corpo dei messaggi di posta elettronica per i set di caratteri.

> **Nota:** La configurazione di questo filtro come filtro dei contenuti è consigliata, in quanto i filtri

dei contenuti si verificano dopo la scansione della posta indesiderata. Tuttavia, può essere scritto e configurato come filtro messaggi, se necessario.

Nell'esempio seguente viene preso in considerazione un messaggio di posta contenente caratteri basati su russo (cirillico) tramite il set di caratteri basato su Windows-1251. Scritto come filtro contenuti:



Content Filter Settings

Name:	russian_text
Currently Used by Policies:	*No policies currently use this rule.*
Description:	This content filter will scan and catch Windows-1251 based characters and send to Policy quarantine.
Order:	1 (of 18)

Conditions

Add Condition... Apply rule: Only if all conditions match

Order	Condition	Rule	Delete
1	Message Body or Attachment	body-contains("windows-1251", 1)	
2	Other Header	header("Content-type") == "(?i)windows-1251"	

Actions

Add Action...

Order	Action	Rule	Delete
1	Add Log Entry	log-entry("<=====WINDOWS-1251 DETECTED=====>")	
2	Quarantine	quarantine("Policy")	

Il messaggio di prova utilizzato conterrà quanto segue nel corpo del messaggio:

```
Russian uses , , , , o, , , , ,  as vowels. You could create a message filter set to "Matches
any of the following" that test whether "Body" "contains" "", "Body" "contains" "" and so forth
until you covered all of the vowels. Ssince English also uses "a" , "e" , "o", and "y" letters
don't test for them. The reason for "Matches any of the following" is to logically OR them - you
want the action to take place if any of those letters are found.
```

Con il filtro dei contenuti configurato come sopra, i log di posta verrebbero registrati in modo simile al seguente:

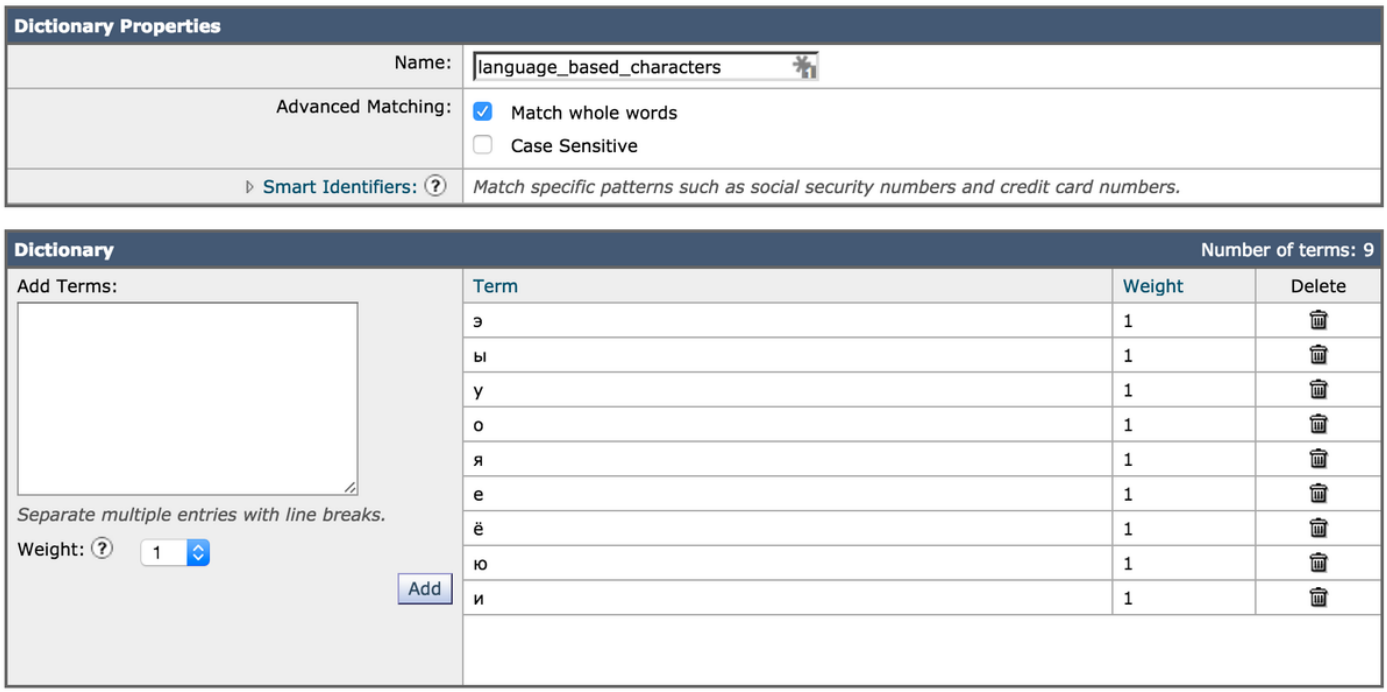
```
Thu Sep 10 14:50:09 2015 Info: Start MID 164993 ICID 266729
Thu Sep 10 14:50:09 2015 Info: MID 164993 ICID 266729 From: <end_user@test.com>
Thu Sep 10 14:50:09 2015 Info: MID 164993 ICID 266729 RID 0 To: <recpient@my_co.com>
Thu Sep 10 14:50:09 2015 Info: MID 164993 using engine: SPF Verdict Cache using cached verdict
Thu Sep 10 14:50:09 2015 Info: MID 164993 Message-ID '<7A961F85-A5F1-413F-87CB-
C31D2E5605EC@my_co.com>'
Thu Sep 10 14:50:09 2015 Info: MID 164993 Subject 'russian test'
Thu Sep 10 14:50:09 2015 Info: MID 164993 ready 2302 bytes from <end_user@test.com>
Thu Sep 10 14:50:09 2015 Info: MID 164993 matched all recipients for per-recipient policy
DEFAULT in the inbound table
Thu Sep 10 14:50:09 2015 Info: MID 164993 AMP file reputation verdict : CLEAN
Thu Sep 10 14:50:09 2015 Info: MID 164993 using engine: GRAYMAIL negative
Thu Sep 10 14:50:09 2015 Info: MID 164993 Custom Log Entry: <====== WINDOWS-1251 DETECTED
======>
Thu Sep 10 14:50:09 2015 Info: MID 164993 quarantined to "Policy" (content filter:russian_text)
Thu Sep 10 14:50:09 2015 Info: Message finished MID 164993 done
```

Èpossibile utilizzare altre lingue e set di caratteri. Per ulteriori informazioni, vedere la sezione Riferimenti.

Scrivere un filtro per fare riferimento a un dizionario basato su caratteri

La seconda opzione consiste nell'aggiungere l'elenco dei set di caratteri a un file di testo del dizionario e fare riferimento a tale elenco nel filtro.

Esempio di aggiunta di caratteri al dizionario:



Dictionary Properties

Name:	language_based_characters
Advanced Matching:	☑ Match whole words ☐ Case Sensitive
▷ Smart Identifiers:	*Match specific patterns such as social security numbers and credit card numbers.*

Dictionary — Number of terms: 9

Add Terms:

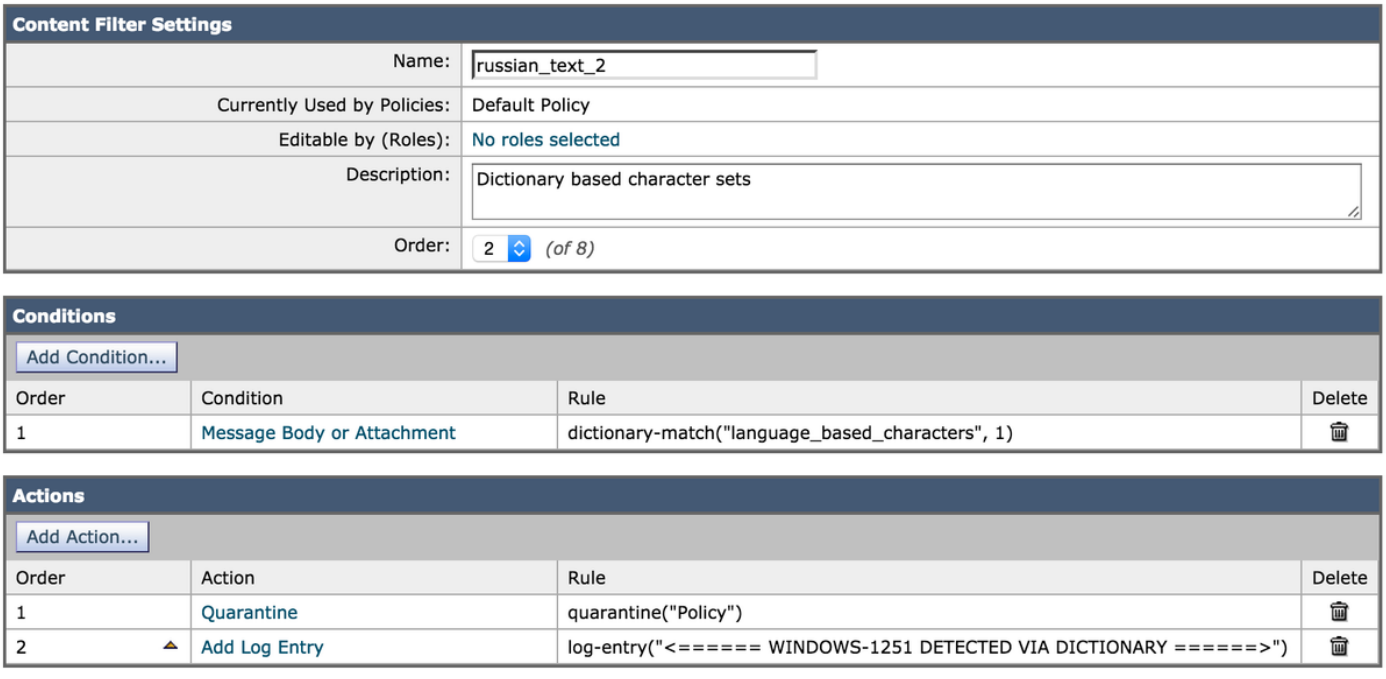
Separate multiple entries with line breaks.

Weight: 1

Add

Term	Weight	Delete
э	1	
ы	1	
у	1	
о	1	
я	1	
е	1	
ё	1	
ю	1	
и	1	

I caratteri vengono ora assegnati al dizionario e viene fatto riferimento al dizionario stesso nelle condizioni del filtro:



Content Filter Settings

Name:	russian_text_2
Currently Used by Policies:	Default Policy
Editable by (Roles):	No roles selected
Description:	Dictionary based character sets
Order:	2 *(of 8)*

Conditions

Add Condition...

Order	Condition	Rule	Delete
1	Message Body or Attachment	dictionary-match("language_based_characters", 1)	

Actions

Add Action...

Order	Action	Rule	Delete
1	Quarantine	quarantine("Policy")	
2	Add Log Entry	log-entry("<====== WINDOWS-1251 DETECTED VIA DICTIONARY ======>")	

Utilizzando lo stesso messaggio di prova di cui sopra, contiene quanto segue nel corpo del messaggio:

```
Russian uses , , , , о, , , , ,  as vowels. You could create a message filter set to "Matches
```

any of the following" that test whether "Body" "contains" "", "Body" "contains" "" and so forth until you covered all of the vowels. Ssince English also uses "a" , "e" , "o", and "y" letters don't test for them. The reason for "Matches any of the following" is to logically OR them - you want the action to take place if any of those letters are found.

Con il filtro dei contenuti configurato come sopra utilizzando la condizione di corrispondenza del dizionario, i log di posta verranno registrati in modo simile al seguente:

```
Thu Sep 10 15:26:08 2015 Info: Start MID 164995 ICID 266737
Thu Sep 10 15:26:08 2015 Info: MID 164995 ICID 266737 From: <end_user@test.com>
Thu Sep 10 15:26:08 2015 Info: MID 164995 ICID 266737 RID 0 To: <recpient@my_co.com>
Thu Sep 10 15:26:08 2015 Info: MID 164995 using engine: SPF Verdict Cache using cached verdict
Thu Sep 10 15:26:08 2015 Info: SPF Verdict Cache cache status: hits = 6, misses = 4, expires =
1, adds = 4, seconds saved = 0.50, total seconds = 0.85
Thu Sep 10 15:26:08 2015 Info: MID 164995 Message-ID '<BCC88307-EB91-476E-8732-
334E9EE84EC8@my_co.com>'
Thu Sep 10 15:26:08 2015 Info: MID 164995 Subject 'russian test 3'
Thu Sep 10 15:26:08 2015 Info: MID 164995 ready 2316 bytes from <end_user@test.com>
Thu Sep 10 15:26:08 2015 Info: MID 164995 matched all recipients for per-recipient policy
DEFAULT in the inbound table
Thu Sep 10 15:26:08 2015 Info: MID 164995 AMP file reputation verdict : CLEAN
Thu Sep 10 15:26:08 2015 Info: MID 164995 using engine: GRAYMAIL negative
Thu Sep 10 15:26:08 2015 Info: MID 164995 Custom Log Entry: <====== WINDOWS-1251 DETECTED VIA
DICTIONARY ======>
Thu Sep 10 15:26:08 2015 Info: MID 164995 quarantined to "Policy" (content
filter:russian_text_2)
Thu Sep 10 15:26:08 2015 Info: Message finished MID 164995 done
```

Scrivere un filtro contenuti utilizzando la condizione "Message Language"

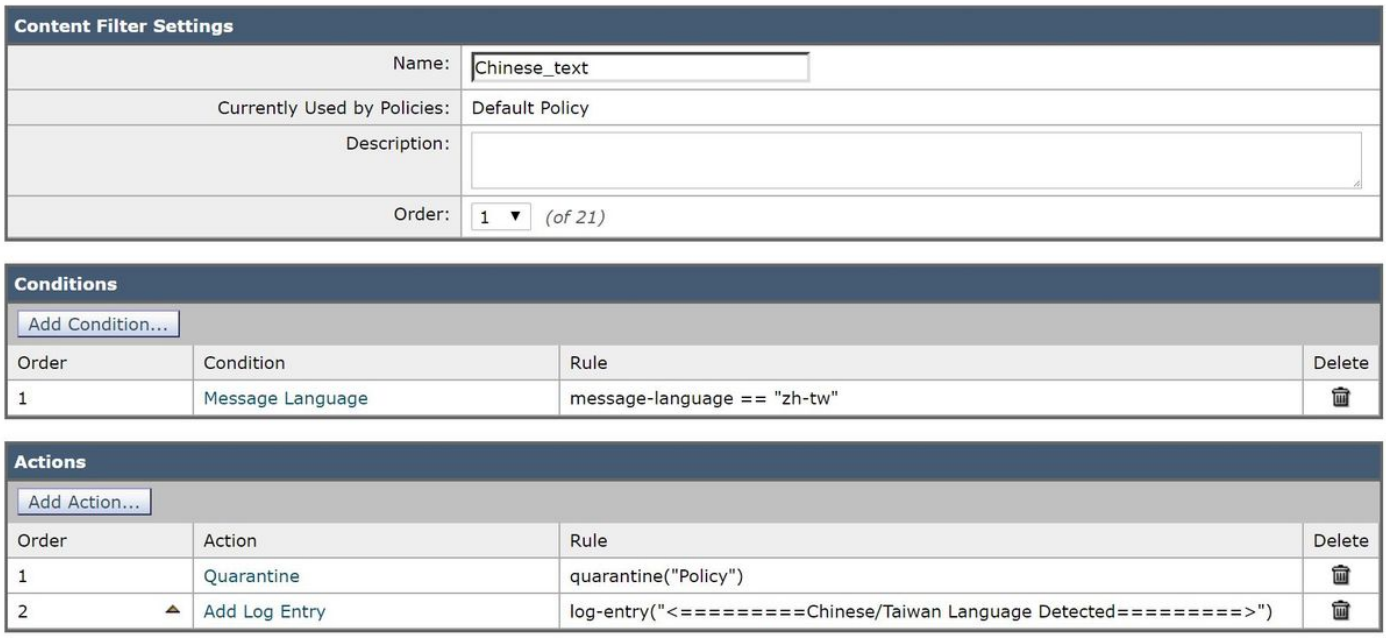
La terza opzione consiste nell'utilizzare la condizione "lingua del messaggio". L'ESA utilizza il motore di rilevamento della lingua incorporato per rilevare la lingua in un messaggio. L'accessorio estrae l'oggetto e il corpo del messaggio e lo passa al motore di rilevamento della lingua.

Il motore di rilevamento della lingua determina la probabilità di ciascuna lingua nel testo estratto e la trasmette all'accessorio. L'accessorio considera la lingua con la probabilità più alta come lingua del messaggio. L'accessorio considera la lingua del messaggio come "indeterminata" in uno dei seguenti scenari:

- Se la lingua rilevata non è supportata dall'ESA
- Se l'accessorio non è in grado di rilevare la lingua del messaggio
- Se le dimensioni totali del testo estratto inviato al motore di rilevamento della lingua sono inferiori a 50 byte.

  Nota: Questa opzione è una nuova funzionalità di AsyncOS Email Security 10.0.0-203 e versioni successive.

Nell'esempio seguente verrà preso in considerazione un messaggio di posta contenente un set di caratteri basato su cinese/Taiwan. Scritto come filtro contenuti:



Con il filtro dei contenuti configurato come sopra, i log di posta verrebbero registrati in modo simile al seguente:

```
Tue Feb 28 06:53:18 2017 Info: Start MID 481 ICID 27
Tue Feb 28 06:53:18 2017 Info: MID 481 ICID 27 From: <end_user@test.com>
Tue Feb 28 06:53:18 2017 Info: MID 481 ICID 27 RID 0 To: <recipient@my_co.com>
Tue Feb 28 06:53:18 2017 Info: MID 481 Subject 'Chinese text test'
Tue Feb 28 06:53:18 2017 Info: MID 481 ready 1047 bytes from <end_user@test.com>
Tue Feb 28 06:53:18 2017 Info: MID 481 matched all recipients for per-recipient policy DEFAULT
in the inbound table
Tue Feb 28 06:53:18 2017 Info: MID 481 interim verdict using engine: CASE spam negative
Tue Feb 28 06:53:18 2017 Info: MID 481 using engine: CASE spam negative
Tue Feb 28 06:53:18 2017 Info: MID 481 interim AV verdict using Sophos CLEAN
Tue Feb 28 06:53:18 2017 Info: MID 481 antivirus negative
Tue Feb 28 06:53:18 2017 Info: MID 481 using engine: GRAYMAIL negative
Tue Feb 28 06:53:18 2017 Info: MID 481 Message language: 'Chinese/Taiwan'
Tue Feb 28 06:53:18 2017 Info: MID 481 Custom Log Entry: <==========Chinese/Taiwan Language
Detected==========>
Tue Feb 28 06:53:18 2017 Info: MID 481 Outbreak Filters: verdict negative
Tue Feb 28 06:53:18 2017 Info: MID 481 quarantined to "Policy" (content filter:Chinese_text)
Tue Feb 28 06:53:18 2017 Info: Message finished MID 481 done
```

Riferimenti

- Microsoft fornisce i nomi dei set di caratteri (*nome .NET*) nella loro Identificatori tabella codici a cui è possibile fare riferimento durante la scrittura e la configurazione dei filtri.

  Nota: le tabelle codici ANSI possono variare a seconda del computer oppure possono essere modificate per un singolo computer, con conseguente danneggiamento dei dati. Per ottenere risultati coerenti, le applicazioni devono utilizzare Unicode, ad esempio UTF-8 o UTF-16, anziché una tabella codici specifica.

- Mozillazina In vengono fornite informazioni dettagliate sul tipo di contenuto: nell'intestazione, nelle lettere straniere, nelle parole straniere e altro ancora Posta indesiderata in lingua straniera

Informazioni correlate

- Attacchi di phishing avanzati omoglifi
- Guide per l'utente finale di Cisco Email Security Appliance
- Documentazione e supporto tecnico – Cisco Systems