

Shareable Automation and Orchestration Workflows for scoring, sharing, and responding to Cyber Indicators of Compromise

July 2020

Acknowledgement:

This material is based upon work supported by the U.S. Department of Homeland Security / Cybersecurity & Infrastructure Security Agency under Grant Award Number DHS-19-CISA-128-SLT-001 State, Local, Tribal and Territorial Indicators of Compromise Automation Pilot.

Disclaimer:

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security / Cybersecurity & Infrastructure Security Agency.



Table of Contents

1.	<i>Introduction</i>	3
1.1	Scoring IOCs to enhance their value.....	3
1.2	What is a workflow?	5
1.3	Cyber Defense Use Cases addressed in this report	7
2.	<i>Shareable Workflows</i>	8
2.1	Shareable Workflows for generation of a scored IOC feed	8
2.2	Shareable Workflows for receiving IOCs from the feed	17
2.3	Shareable Workflow for processing IOCs from email submissions.....	21
2.4	Shareable Workflow for enrichment of threat data	22
2.5	Shareable Workflows for receipt and response to IP address IOCs.....	23
2.6	Shareable Workflows for receipt and response to Domain/URL IOCs.....	27
2.7	Shareable Workflows for receipt and response to File Hash IOCs.....	32
2.8	Shareable Workflows for receipt and response to Email Sender IOCs.....	36
3.	<i>Summary</i>	39

1. Introduction

The nature of the cybersecurity threat is consistently growing, and cyber adversaries move with speed and stealth, often utilizing automation to increase the scale of their attacks. To keep pace, all types of organizations need to be able to share information and respond to cyber risk in as close to real-time as possible.

Using a Department of Homeland Security Cybersecurity & Infrastructure Security Agency (DHS CISA) grant, the Johns Hopkins University Applied Physics Laboratory (JHU/APL) is conducting a joint pilot with the State, Local, Tribal and Territorial (SLTT) community to apply automation to enhance and speed the evaluation of cyber threat Indicators of Compromise (IOC) at the state and local government levels.

While the JHU/APL pilot effort has focused on collaboration with specific SLTT organizations, some of the artifacts from this work are applicable to any enterprise security effort. This document will provide some of those insights to the greater community of cyber defenders with hopes that it will be helpful.

In order to enhance the IOCs, JHU/APL developed a simple scoring rubric that provides clear indications of the impact of acting on an IOC. JHU/APL then developed automatable workflows that enable systems to act upon the IOCs without human intervention. These two steps will be described in the next two sections, followed by use cases and then explanations and examples of shareable workflows for those use cases.

1.1 Scoring IOCs to enhance their value

There are many sources of cyber threat intelligence available to network defenders today. However, these feeds often result in very little tactical utility for network defense because of poor data quality, and limited ability to rapidly screen information to identify the pertinent pieces of information and what to do with it.

A key component of JHU/APL's automation efforts is to curate a threat feed that network defenders can consume and act upon in an automated manner. One way to make threat intelligence more consumable is to provide a score that conveys context essential to decision making in a consistent and transparent manner. With that in mind, the scoring system defined in Table 1 was developed for IOCs. These IOCs can be extracted from Intrusion Detection/Prevention Systems (IDS/IPS), or other threat feeds. It is JHU/APL's vision that a scoring system like this can be applied to a circle of trust group sharing IOCs amongst themselves via a third party or within an organization's threat intelligence capability if one exists.

Table 1 Score Definitions

Score	Shared	What it means	Other notes/discussion points
0	No	Whitelisted or equivalent / High-Regret	These are IOCs that are associated with known good or member infrastructure. These items are considered High-Regret as blocking them automatically is assumed to have detrimental impact on an organization.
1	No	Undetermined	This is the default for any IOC that is not determined to be High- or Low-Regret.
2	Yes	Low-Regret	These are IOCs that demonstrate characteristics common to malicious activity/code. They have not been vetted by an analyst, but taking action on these IOCs is not expected to significantly impact an organization. This may include IOCs associated with non-malicious, but unnecessary or suspicious, activity/code (e.g., spyware).
3	Yes	Analyst vetted	These IOCs have been through some process and are determined to be suspicious or most likely malicious. They may also be vetted by external analysts or processes trusted to be equivalent. An analyst's input can override previous scores including those that were not previously shared.
4	Yes	Analyst validated	An internal or external analyst has determined with high confidence that the IOC is associated with malicious activity. An analyst's input can override previous scores including those that were not previously shared.

At the heart of this scoring methodology is the ability to determine if an IOC is high or low regret. This determination is being performed in a completely automated manner. A very conservative set of checks were implemented and are listed in

Table 2. More characteristics were considered, but only those that required information consistently available in an automated fashion using existing or free resources were

selected. Reputation services or other enrichment sources were not used for the JHU/APL effort, but organizations can easily add in these types of checks as appropriate.

If an organization chooses to utilize these reputation sources, it is recommended to verify that the number of potential IOCs from your sources will not exceed licensing restrictions once automation is in place.

Table 2 Regret Determination Checks

Regret Type	White List	Age	Signature	Reputation
High	The IOC is on a whitelist	Domains with a registered date >180 days	N/A	Domains on a Top 500 list
Low	N/A	Domains with a registered date <=30 days	Signatures that flag content as likely malicious with medium to high confidence	IOC is deemed malicious (e.g., on a block list) and not flagged as shared infrastructure.

The term signature is used loosely to mean any rule, signature, score, etc., that content is evaluated against to determine whether something is suspicious or potentially malicious. It is the value or condition identified by the source that can be compared to a set of rules to determine if the IOCs in the content should be considered high or low regret. In JHU/APL's effort, signature checks were only used to determine low regret, but values known to generate many false positives can be used to flag potential IOCs as high regret. For some sources, certain signatures are considered equivalent to an analyst vetting the condition or content. For example, an IOC with a certain severity score provided by another organization that is trusted could be considered analyst vetted. Analyst validated should be saved for only those IOCs whose association with malicious behavior has been confirmed.

1.2 What is a workflow?

Many organizations are looking to implement automation and orchestration via products such as Security Orchestration, Analysis, and Response (SOAR) platforms. However, a SOAR vendor's capabilities can be misaligned with an organization's policies and procedures. To help establish this connection, JHU/APL has identified three levels of abstraction for use by an organization:

- Playbooks (Process Oriented)
 - Represents a general security process at most basic level
 - Can be mapped to governance or regulatory requirements
 - Identifies Industry best practices for steps in the process

- Designed to be human readable
- Workflows (Technical Steps)
 - Implements an organizational playbook
 - Is repeatable and auditable
 - Can tailor the amount of automation depending on the needs and capabilities of the system and the desires of the organization
 - Is machine-to-machine sharable
- Local Instances of Workflows (Execution at System Level)
 - Is often thought of as a “run book” or “SOAR playbook”
 - Orchestrates and executes a workflow’s actions in a manner that:
 - Is consistent with local policies, procedures, thresholds, and decision process
 - Incorporates technologies, products, and assets deployed in the local environment
 - Responds to conditions or events that are occurring in the local environment

Figure 1 provides a comparison between these types of abstraction.

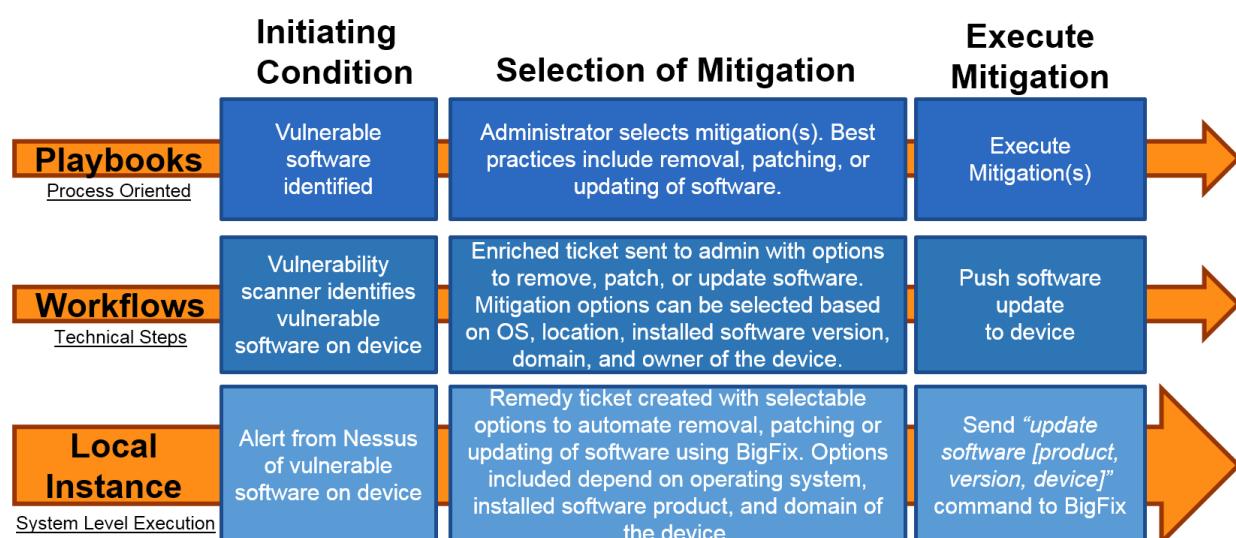


Figure 1 Levels of Automation and Orchestration Abstraction

This report provides examples of workflows to help multiple organizations understand how to implement security automation and orchestration if they have not done so before. These workflows are represented using Business Process Modeling Notation (BPMN). This is a standard for workflows that allows for representation of the process without requiring specific technologies. There are multiple free and non-free applications for editing and reading files in the BPMN format (e.g., Camunda Modeler, Flowable Modeler, etc.). While visual representations are provided in this report, JHU/APL will also make the XML based “.bpmn” files available for download as well. Figure 2

provides a style reference of the BPMN elements used in the workflows provided in this report.

Events		An Event is something that “happens” during the course of a Process.			
	Start Event	The Start Event indicates where a particular Process will start.		End Event	The End Event indicates where a Process will end.
	Intermediate Signal Catching Event	Intermediate Event catching a signal from a preceding event in another workflow.		Intermediate Signal Throwing Event	Intermediate Event throwing a signal to a subsequent event in another workflow.
Tasks		A Task is an atomic Activity within a Process flow. A Task is used when the work in the Process cannot be broken down to a finer level of detail.			
	Service Task	A Task that uses a service, which could be a Web service or an automated application.		User Task	A human performer Task with the assistance of a software application and is scheduled through a task list manager.
	Receive Task	Simple Task that waits for a Message to arrive from an external Participant.		Mail Task	Task where email is used to communicate a status indication or update.
Gateways		Gateways are used to control how Sequence Flows interact as they converge and diverge within a Process.			
	Exclusive Gateway	Only one of the paths can execute based on the decision logic. Often “yes/no” values.		Inclusive Gateway	A Default path executes and is indicated with a backslash through the Sequence Flow. Other paths are based on decision logic at the Gateway.
	Parallel Gateway	Multiple paths execute without priority order.		Sequence Flow	Creates a dependency where the performance of the 1st Task MUST be followed by the 2nd Task.

Figure 2 BPMN Style Reference

1.3 Cyber Defense Use Cases addressed in this report

While security automation and orchestration can be applied to a wide variety of use cases, JHU/APL based this report on the workflows conducted in our SLTT collaboration. These use cases include the following:

- Generation of a scored IOC feed
- Receiving IOCs from the feed
- Processing IOCs from email submissions
- Enrichment of threat intelligence data
- Receipt and response to IP address IOCs
- Receipt and response to Domain/URL IOCs
- Receipt and response to File Hash IOCs
- Receipt and response to Email Sender IOCs

For many of these use cases, JHU/APL is providing multiple versions of the workflows. This is to allow organizations with different business rules and risk profiles the ability to see alternate approaches that may better fit their organization.

2. Shareable Workflows

The JHU/APL team collaborated with multiple organizations to develop automation and orchestration workflows to support the use cases addressed in this report. These workflows were inherently tied to specific technologies for each partner environment. Due to this constraint, the orchestrator workflows by themselves are not immediately usable by other SLTT members and can require heavy modification if a pilot partner changes their security technology stack.

To address this issue, multiple examples of BPMN workflows are presented to showcase the different ways that one could approach the challenges within the use case. Some organizations may prefer a simpler approach and some may want more complex decision logic that is in accordance with their business practices and risk profiles. It is for this reason that multiple solutions exist and are presented in this report.

2.1 Shareable Workflows for generation of a scored IOC feed

The development of a threat feed has been a critical aspect of this pilot. In this section, we will present the general process required to create, score, and disseminate IOCs via automation and orchestration. Figure 3 shows the high-level end to end scoring and dissemination process being implemented. This process involves:

1. Polling multiple information sources
2. Extracting potential IOCs
3. Determining if the IOC needs to be scored
4. Performing regret determination
5. Assigning associated scores
6. Generating the STIX message
7. Posting the STIX message to the TAXII server for dissemination

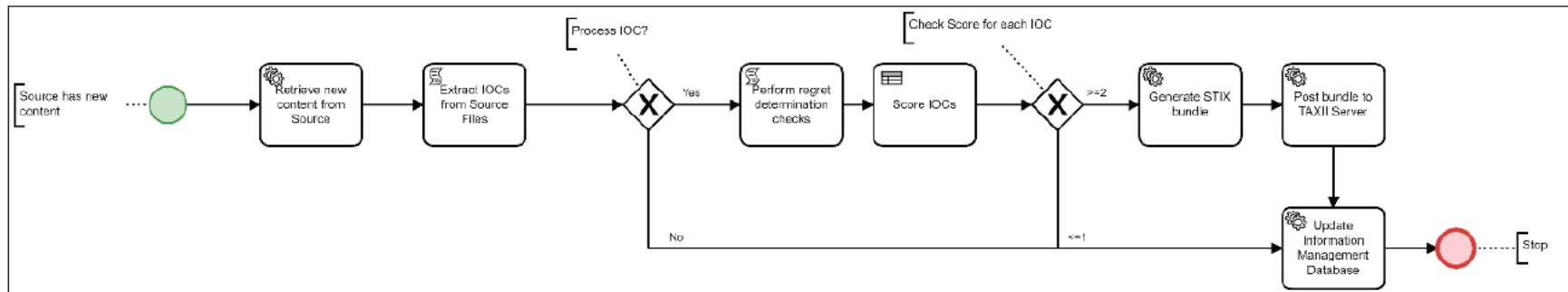


Figure 3 High-Level Pilot Threat Feed Workflow

Figure 4 shows the Parsing workflow in more detail, which implements steps 1-3. The workflow can be initiated in one of two ways: based on elapsed time and polling a source or a trigger is provided by a source to indicate new content. A previously scored IOC will go through the regret determination process again if it: a) comes from a new source or b) it has been 7 days since it was last scored. The different sources have different levels of accuracy and provide different context when identifying IOCs, so when a new source has findings for a previously seen observable, the score may be different once the new information is processed by the Regret Determination workflow.

The reason to reevaluate after 7 days is to prevent consumer organizations from aging off IOCs that are still in use. Most IOCs are associated with malware Command and Control (C2) infrastructure and have a very short half-life. Many products and operations will implement a process to undo response actions (e.g., remove firewall block) for IOCs that are 7-30 days old. For that reason, IOCs seen after 7 days from previous scoring will be either reevaluated or resent with a new valid until value in the STIX message.

IOCs that have been previously scored as a 3 or 4 will not be rescored, only resent. If the previous score was less than 3, the IOC will go through the regret determination and scoring process again. An organization can determine if they want to add in more complex logic for reevaluating IOCs that were previously considered analyst vetted or validated.

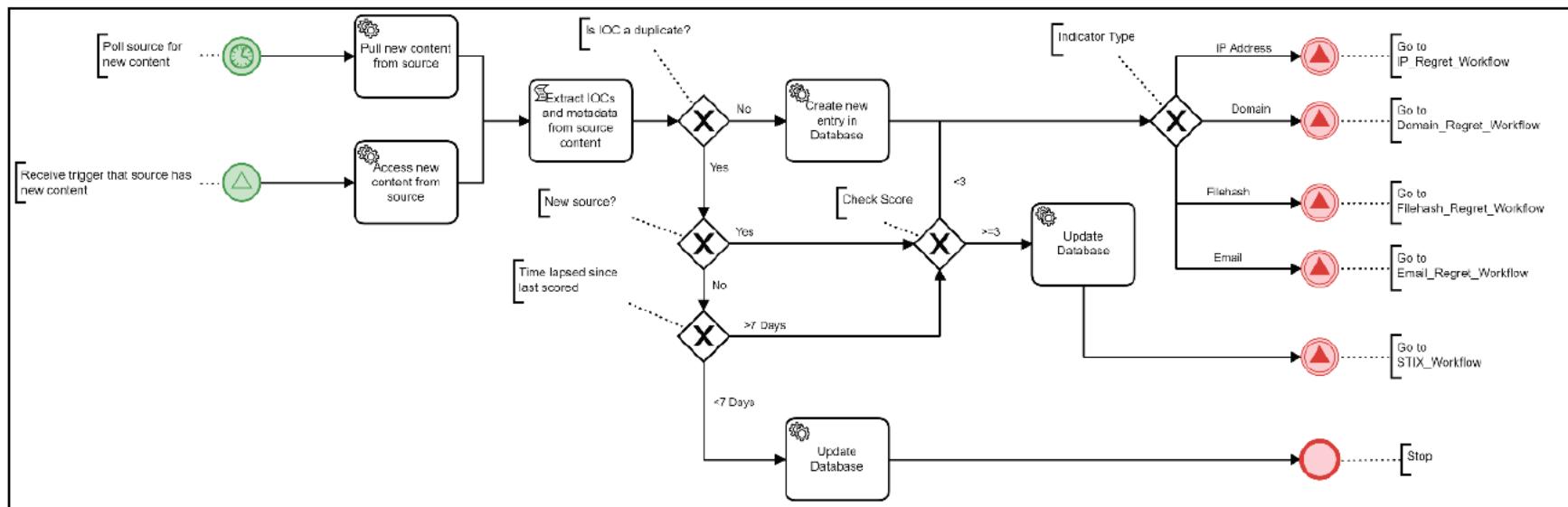


Figure 4 IOC Parsing Workflow

Figure 5 documents the Regret Determination workflow for an IP Address. The first step is to check the IP Address against a whitelist. Analysts have different levels of confidence in the signature accuracy for difference sources. Therefore, for high confidence sources, the signature itself can be used to determine if an IP Address is low-regret. For all other sources, the signature checks are used to filter the IP Addresses that should be checked against the block list. If an IP Address is on the block list, another check is performed to determine how many domains are associated with that IP. Only if the number of domains is ≤ 1 is the IP Address then marked as low regret.

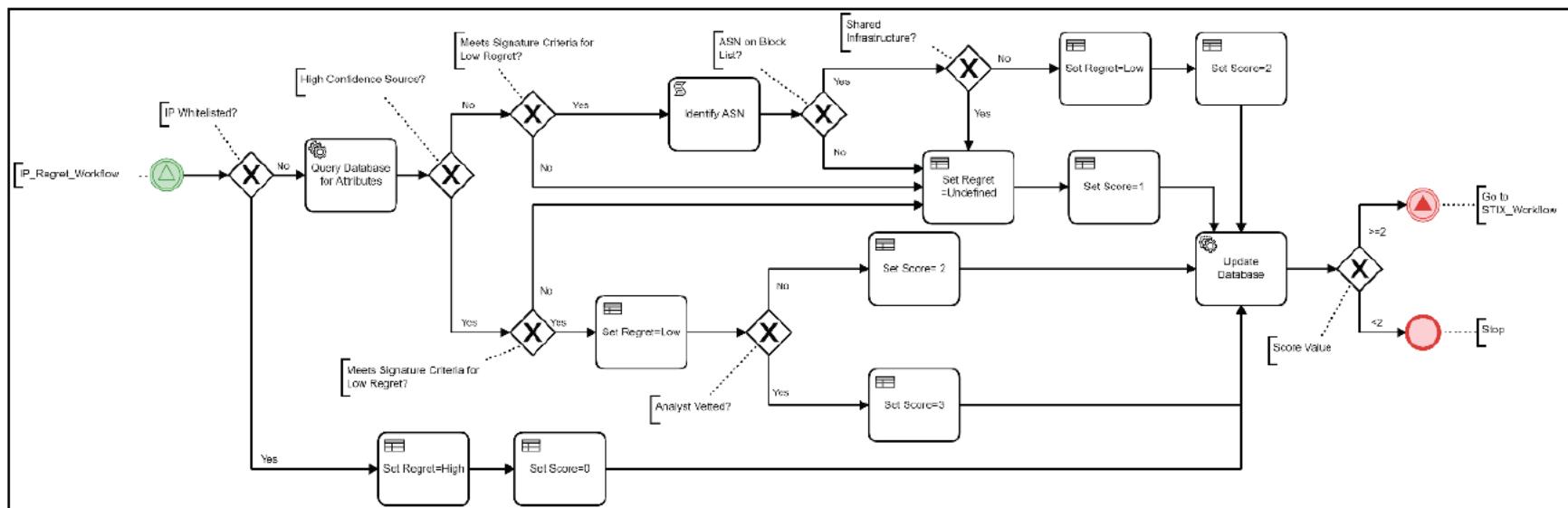


Figure 5 IP Address Regret Determination Workflow

Figure 6 documents the Regret Determination workflow for a domain. Any domain from any source that was registered less than 30 days ago is determined to be low regret.

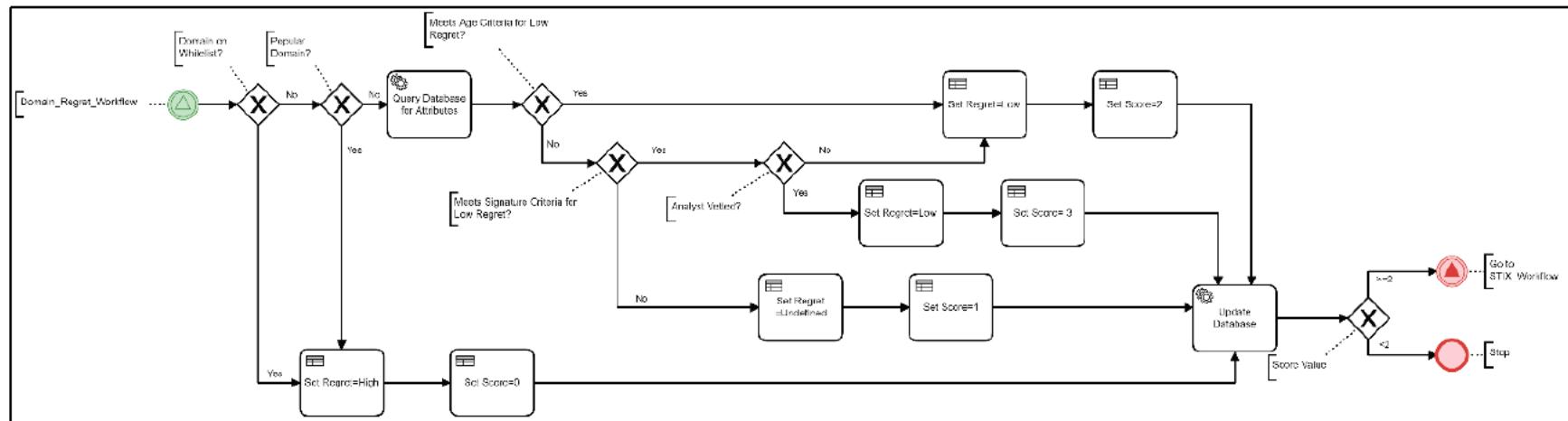


Figure 6 Domain Regret Determination Workflow

Figure 7 and Figure 8 document the Regret Determination workflow for a file hash or email respectively.

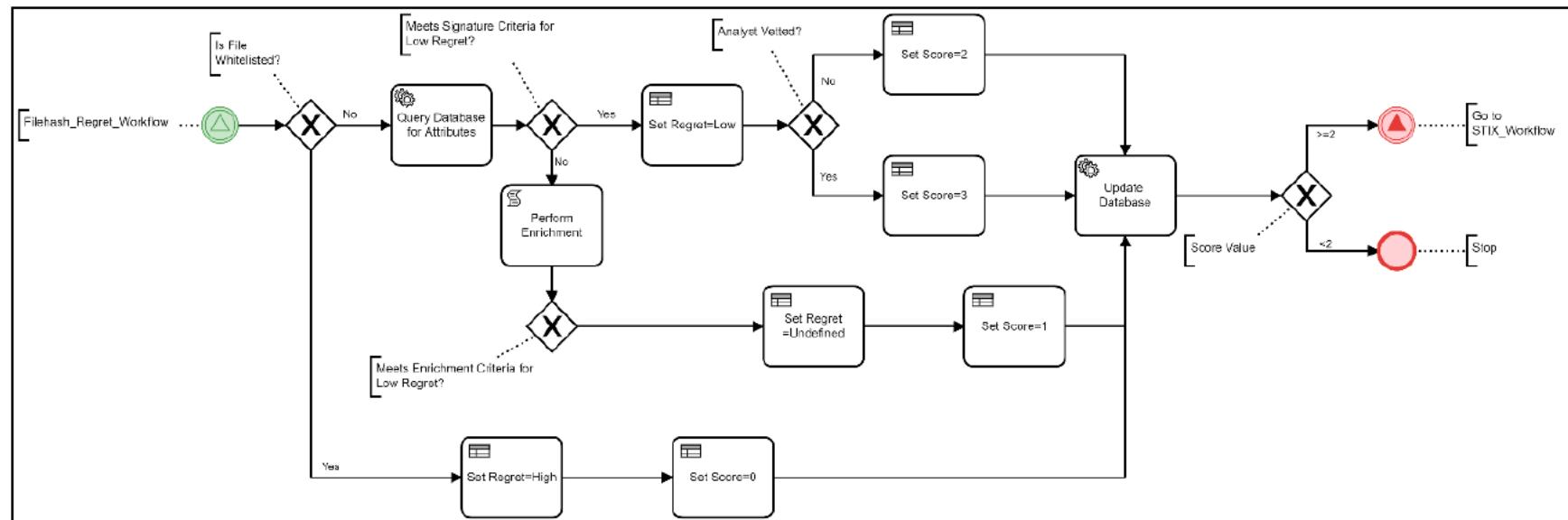


Figure 7 File Hash Regret Determination Workflow

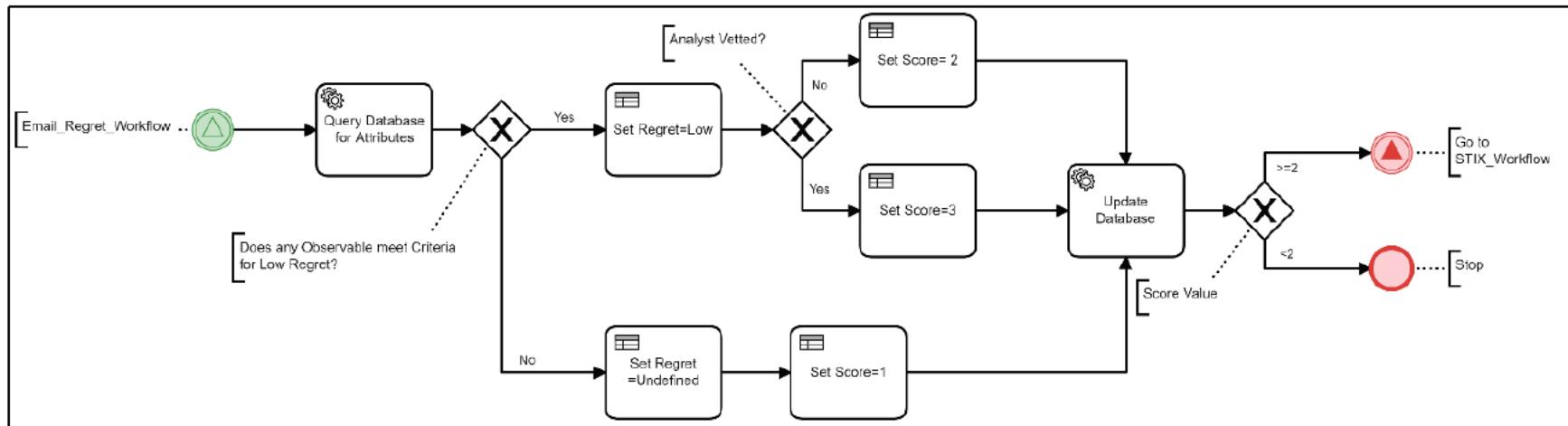


Figure 8 Email Regret Determination Workflow

Figure 9 documents the STIX Generation and Sharing workflow, which includes steps 6 and 7 highlighted in the discussion of the high-level process.

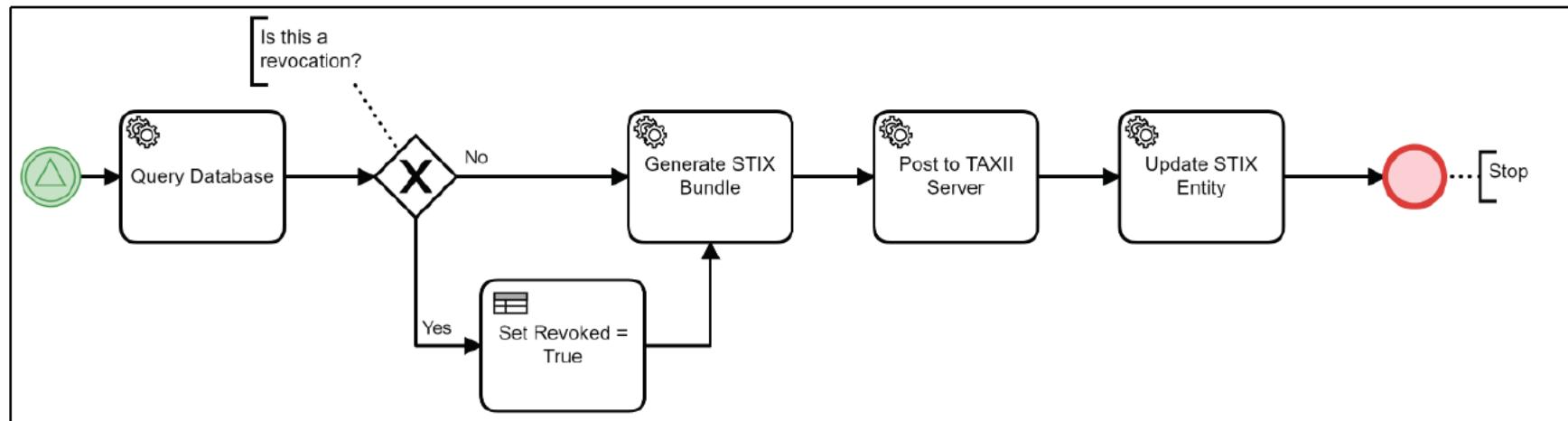


Figure 9 STIX Generation and Sharing Workflow

Figure 10 represents the two processes that occur after an analyst has completed evaluation of any alert, message, or other content that contained an IOC. The first process is triggered when the conditions are met that an IOC is to be considered analyst vetted. IOCs determined to be low regret and previously shared as a 2 are rescored as a 3 and resent.

The second process is triggered when an IOC meets the conditions to be considered analyst validated. These IOCs are rescored as a 4 and resent.

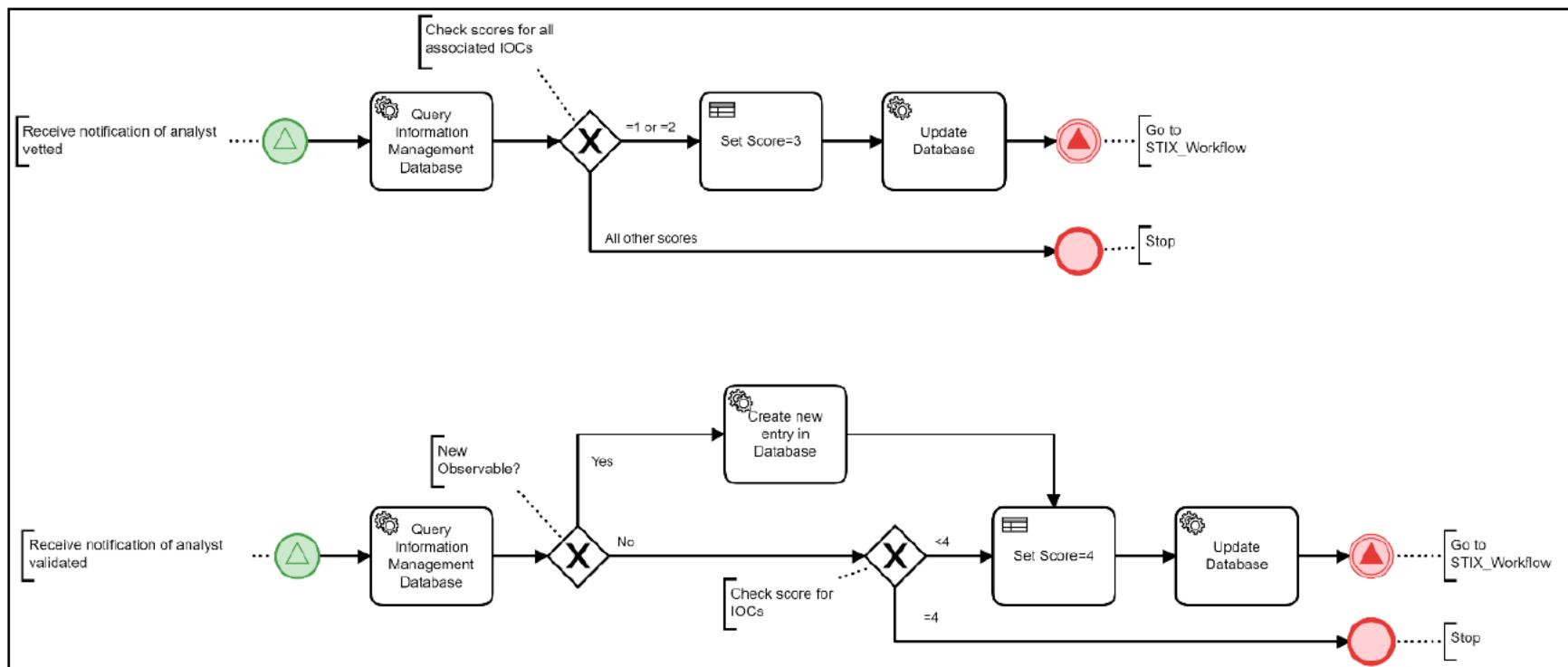


Figure 10 Post Analysis Score Refinement Workflow

Figure 11 is a Revocation workflow. This workflow is meant to correct scores for IOCs that are inaccurately marked as low-regret. This can occur because the content that the IOC is associated with has been deemed to be a false-positive or because a decision has been made to revoke the IOC.

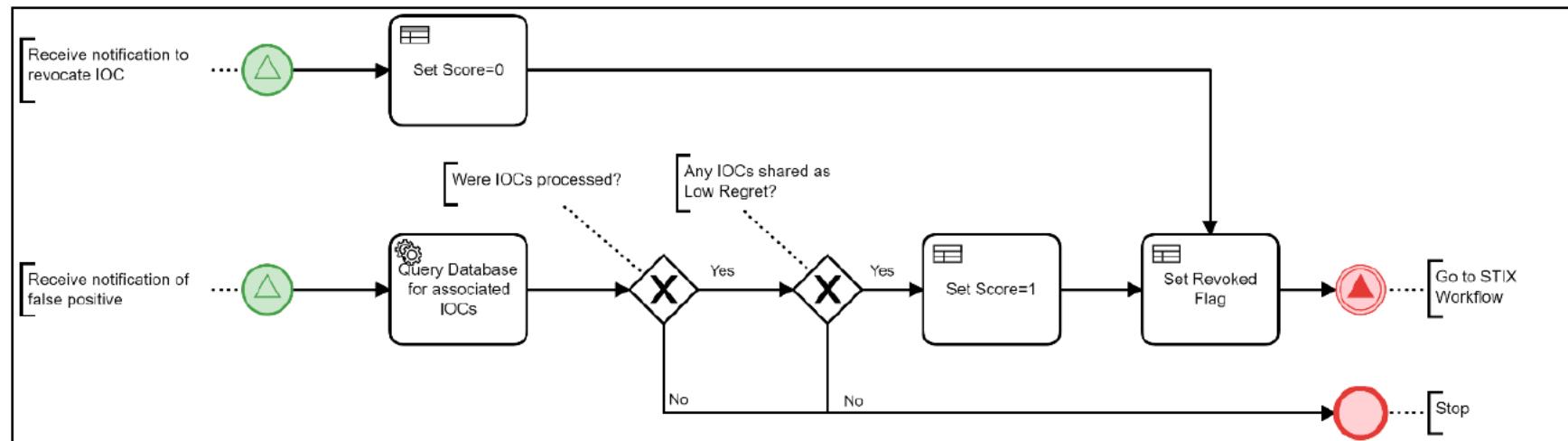


Figure 11 Revocation Workflow

2.2 Shareable Workflows for receiving IOCs from the feed

Retrieving threat intelligence such as IOCs from an intelligence feed is a common task conducted by many SOCs throughout the SLTT community and others. It is very often a manual task consisting of analysts copying and pasting hundreds to thousands of unique IOCs daily from emails, attachments, and websites. In this section, several options are presented to showcase how one could apply automation and orchestration to address this task.

Figure 12 illustrates one technique for retrieving threat intelligence. Automation extracts the IOC from a feed, updates local records as to whether or not that specific piece of information has been seen before and presents that data to a human for a decision. This is an example of a case that uses automation for only the lowest risk, most mundane and repeatable task but still allows the human to have maximum control of every case.

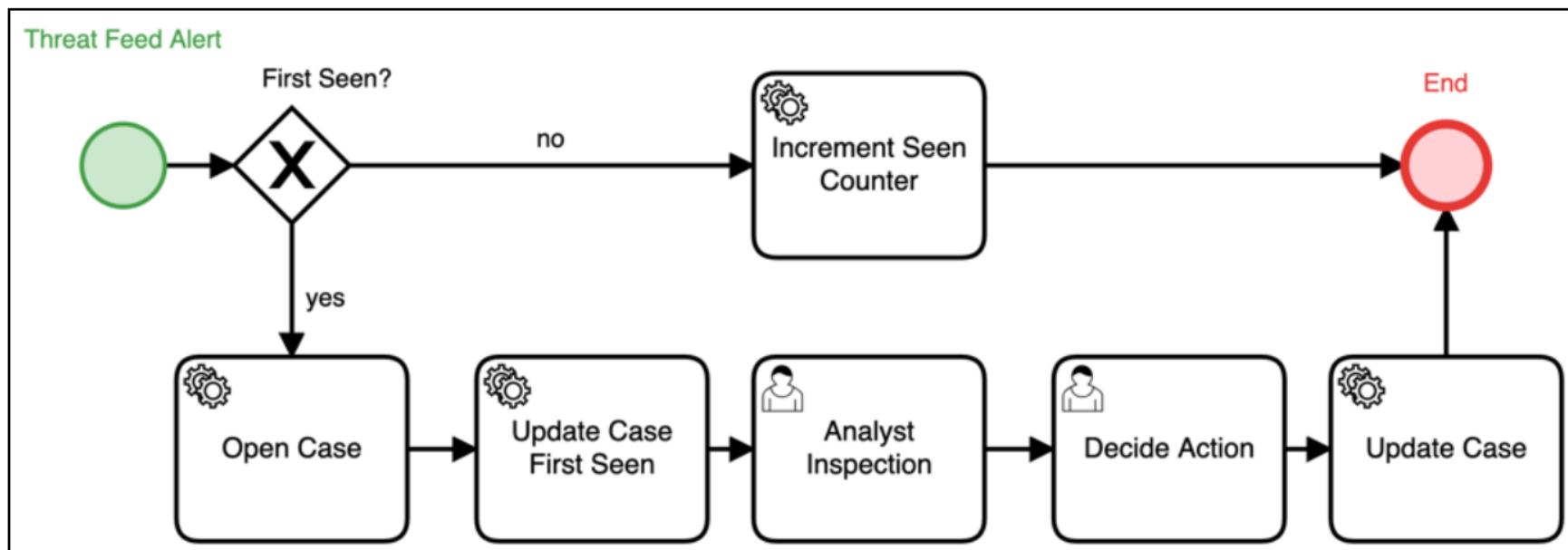


Figure 12 Threat Feed Ingestion example 1

Figure 13 represents a case for extracting threat intelligence in an environment where automation and orchestration is more heavily relied upon. In this workflow, the SOAR platform extracts and parses the STIX objects from the feed and invokes additional analysis and response workflows based on the IOC type. In this scenario, the human analyst is only involved when the workflow has errors. The modular design demonstrated in this approach allows for SOCs to have more trust in this automation as it is narrow in scope and easy to verify proper handling.

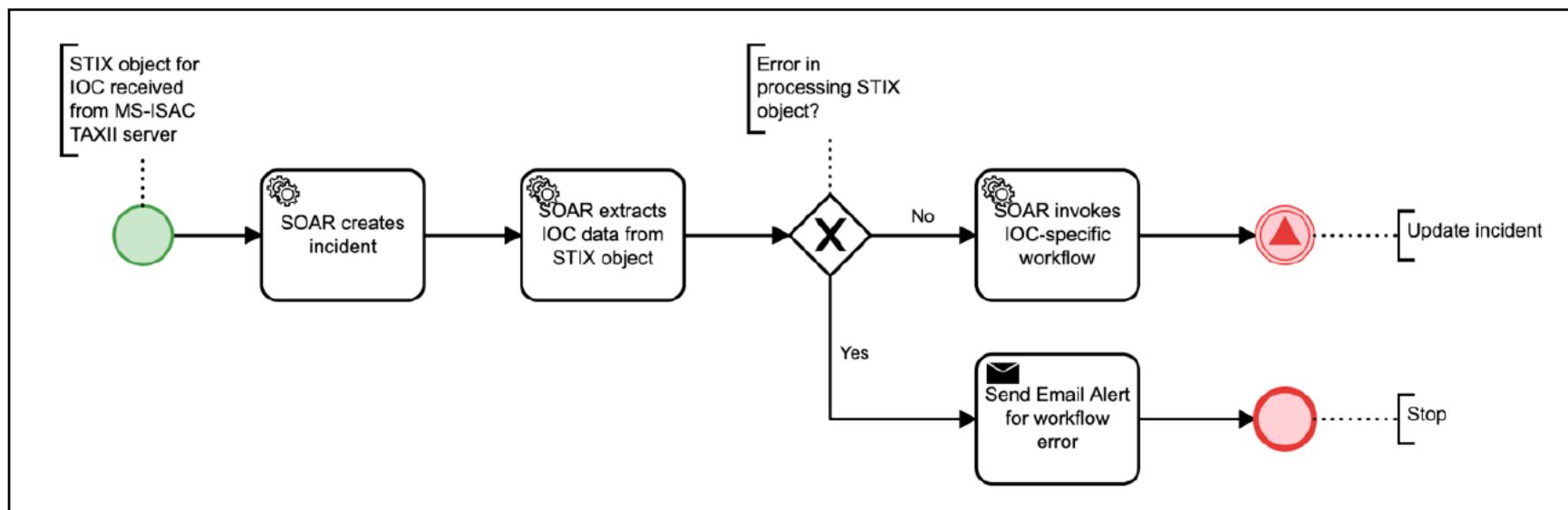


Figure 13 Threat Feed Ingestion example 2

Figure 14 provides an example of how an organization can use additional automation to augment the IOC extraction process. This workflow is conducting additional tasks to augment the organization's logging requirements and potentially use other existing processes and procedures already found in the operational environment.

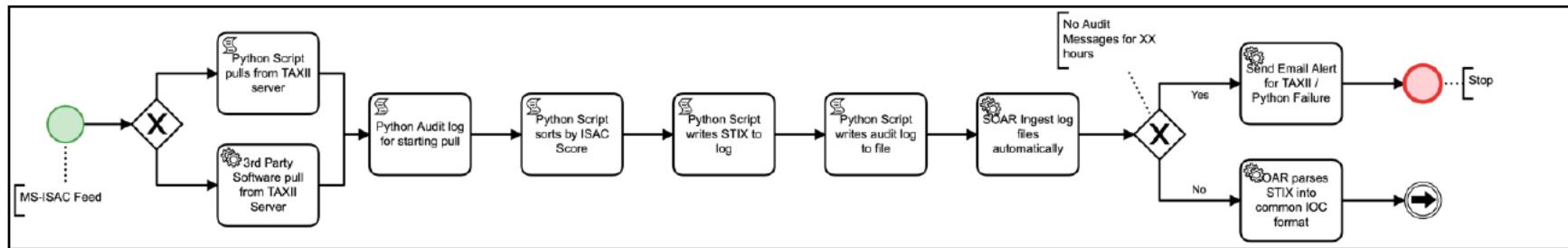


Figure 14 Threat Feed Ingestion example 3

2.3 Shareable Workflow for processing IOCs from email submissions

While STIX/TAXII provide machine readable infrastructure to pass cyber threat intelligence quickly and at scale, many organizations still rely upon email as a primary means for receipt of threat intelligence. Figure 15 provides a workflow for extracting the IOCs from an email, and forwarding the relevant data to analysts or additional workflows as deemed appropriate by the SOC's policies and procedures.

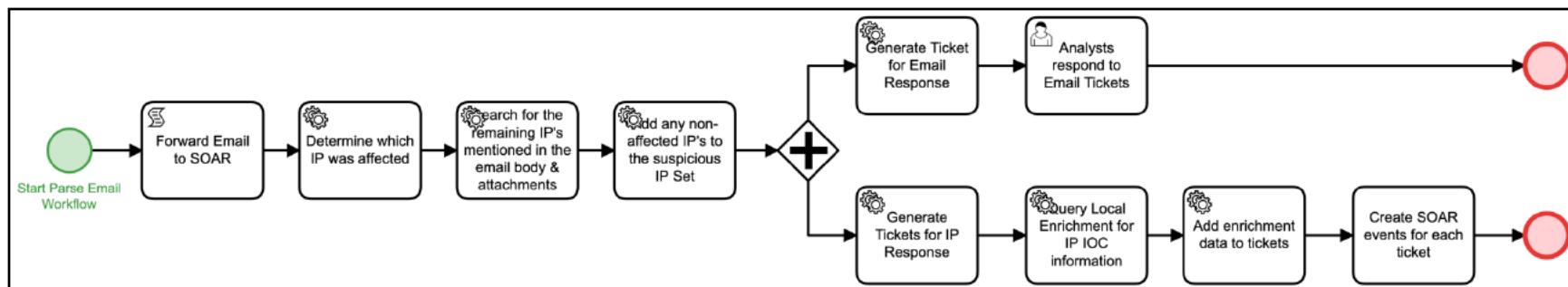


Figure 15 Processing IOCs from Email Submissions

2.4 Shareable Workflow for enrichment of threat data

Once IOCs have been received, a very common task for many SOC analysts is to conduct both external and local enrichment (e.g. VirusTotal scores, local prevalence, WHOIS, domain age, etc.) to better inform which actions should be taken. This is an ideal case for the use of automation and one such process is documented in Figure 16. In this example, the automation conducts many of the lookups that would have to be done manually so that the human analyst no longer has to do those repetitive tasks and can begin the case with an enriched ticket providing not just the IOC but the relevant information needed to make a decision and act on the intelligence.

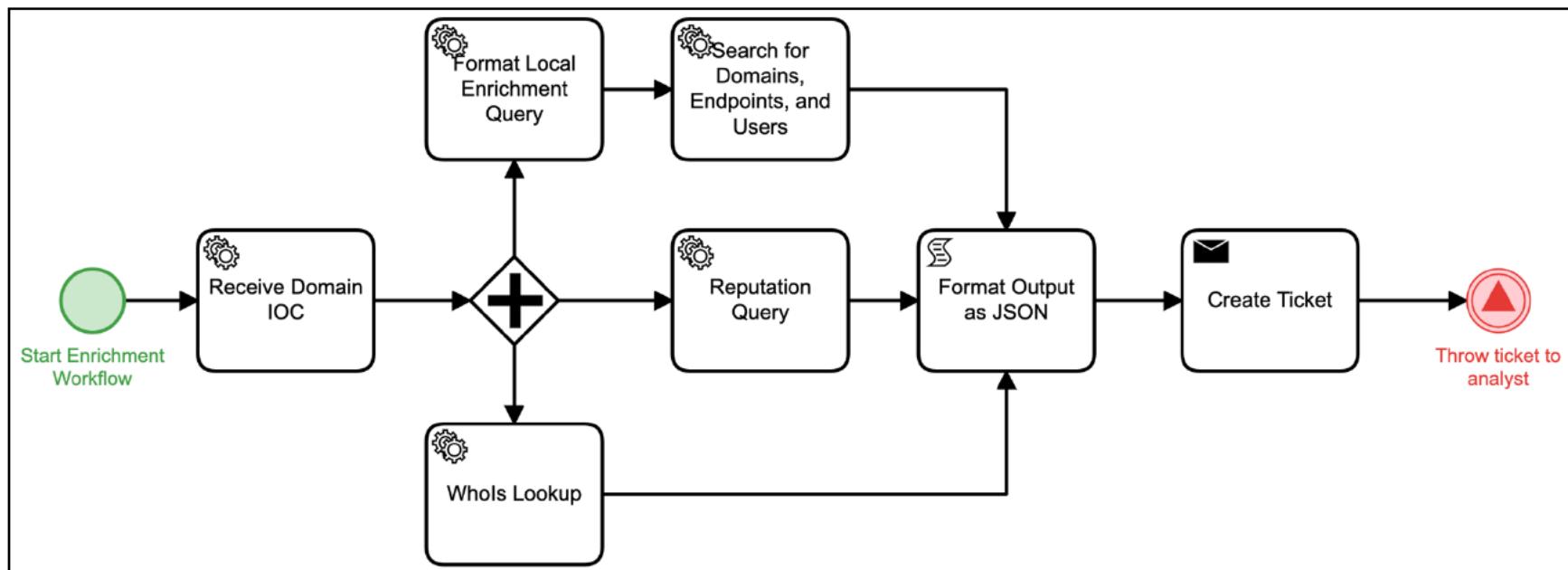


Figure 16 Threat Intel Enrichment

2.5 Shareable Workflows for receipt and response to IP address IOCs

IP addresses are one of the most common IOCs received via threat feeds. Due to the ephemeral nature of IP addresses (e.g. due to DHCP an IP address can represent a malicious server one day and a legitimate one the next), the decision on whether or not to block an IP address and how long to block it can be difficult. To assist with this case in the pilot, multiple approaches are presented.

Figure 17 provides an example of using a “low-regret” strategy for processing IP IOCs. After the SOAR platform verifies the latest IOC reputation, it searches for prevalence in the local network. If the IOC is malicious and no systems have connected to it, it will be automatically blocked as there is low risk that any legitimate process will be impacted. For all other cases, the human analyst will be provided the summarized data so a decision can be quickly made.

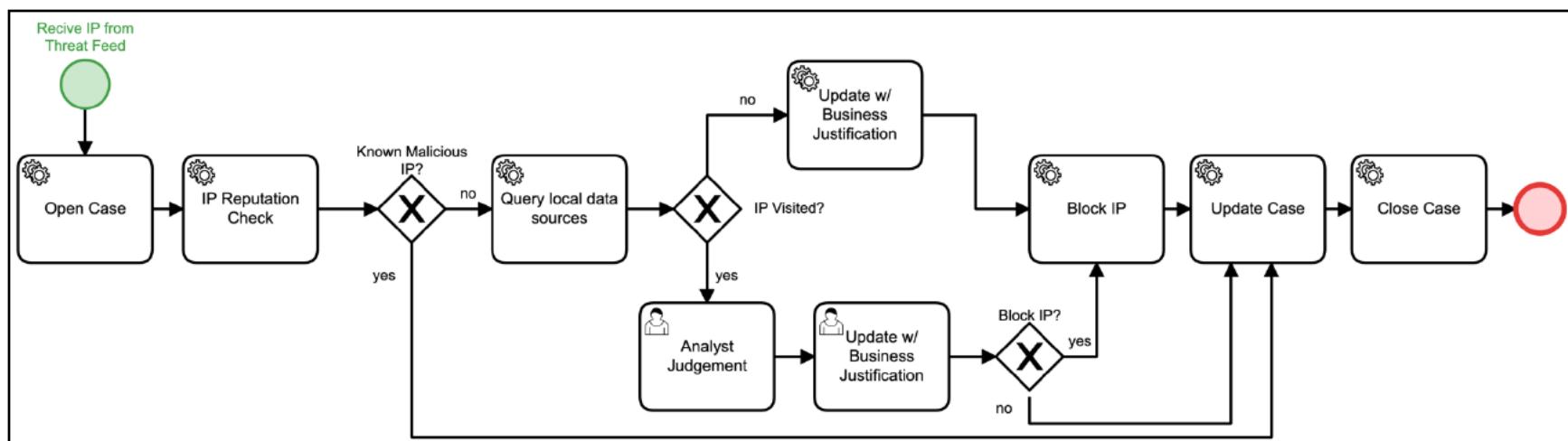


Figure 17 Response to IP IOC example 1

Other organizations may have very detailed vetting and processing logic for their threat intelligence. Often this is done to support a multiple stage triage process so that limited resources (analyst time, paid-for enrichment license limitations, etc.) can be optimized for their utility. Figure 18 provides an example of how an orchestrated workflow can be enhanced to support this process for the vetting and decision points with respect to IP IOCs.

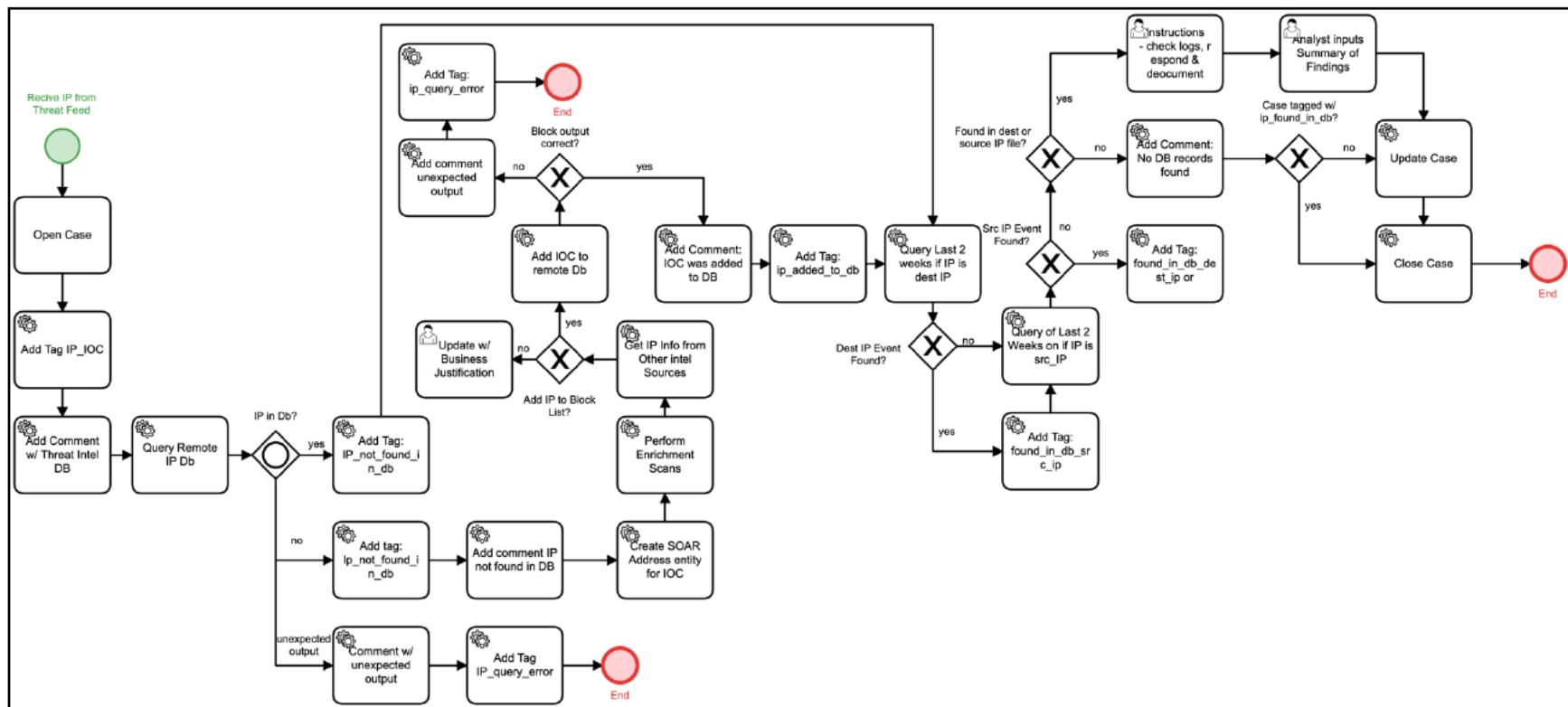


Figure 18 Response to IP IOC example 2

Whereas some organizations may keep their local enrichment in a common “data lake”, others may need to query their tools directly to learn necessary information for making decisions on IP IOCs. Figure 19 provides an example where a SOAR platform may interact with multiple tools not just to block the IOC but to gain the necessary information to support both manual and automated decisions.

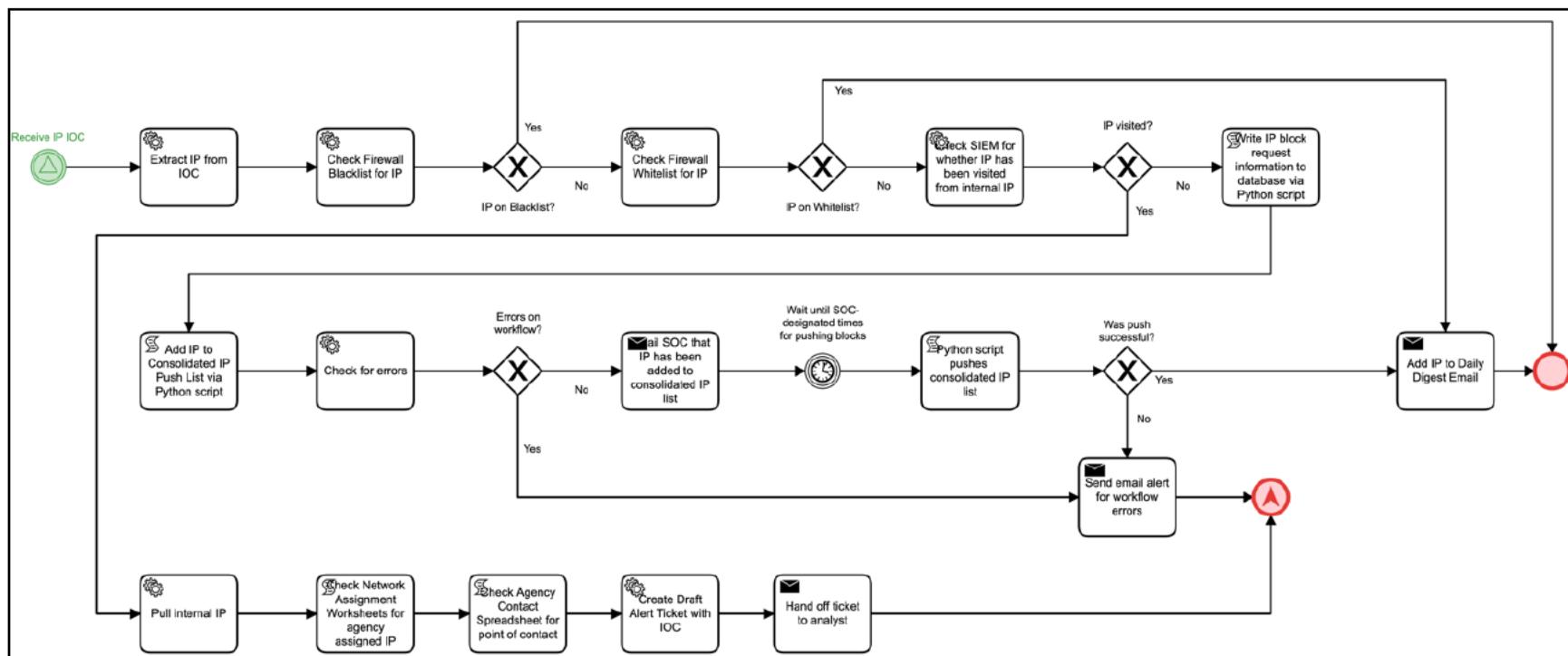


Figure 19 Response to IP IOC example 3

Another case worth noting is with respect to local enrichment. The type of asset that is potentially impacted by a cyber threat can require different courses of action for response either due to the nature of the asset or the jobs assigned to it. Figure 20 provides an example of how those types of decisions can be applied to the response of a malicious IP address.

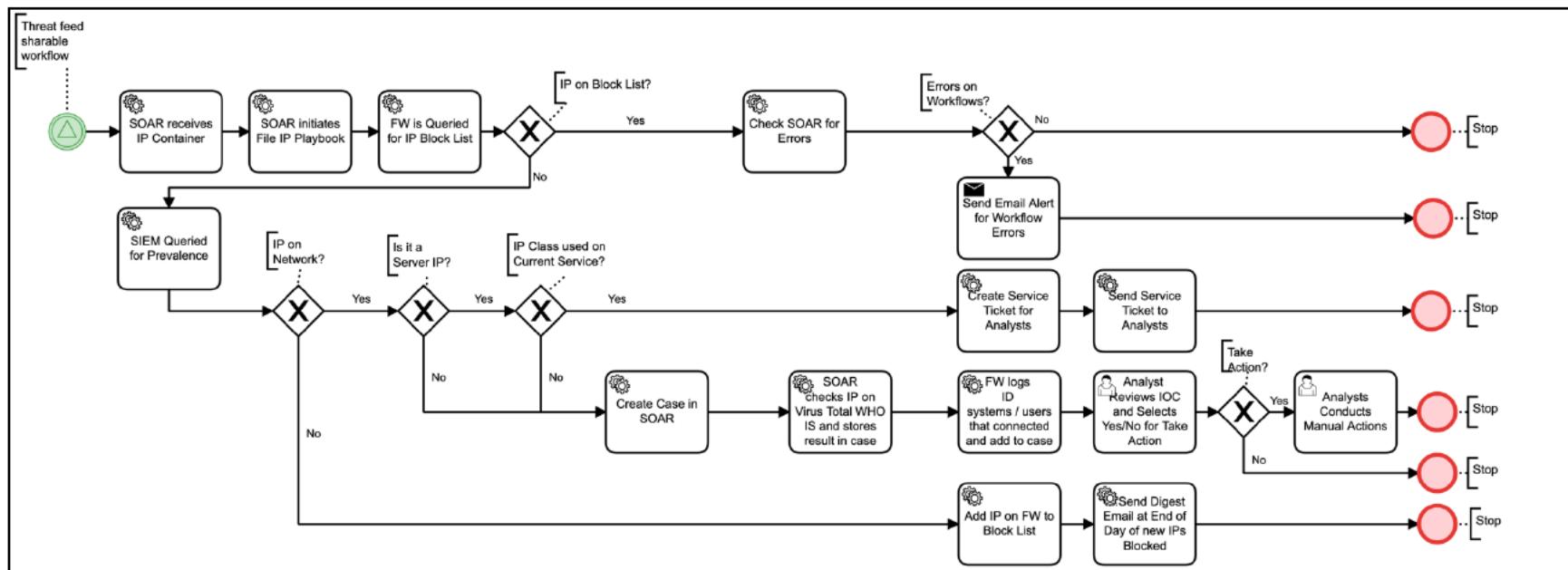


Figure 20 Response to IP IOC example 4

2.6 Shareable Workflows for receipt and response to Domain/URL IOCs

The response to domain IOCs is very similar to that of an IP IOC with respect to the types of technology used for enrichment and blocking. However, web domains tend to remain as viable candidates for blocking for significantly longer time periods (with exceptions for watering hole attacks and the rare case that an expired malicious domain is later purchased for legitimate use). These workflows are provided to show multiple ways to apply automation and orchestration to respond to these IOCs.

Figure 21 provides an example of how to apply an organization's policies to block malicious domains while taking a "low-regret" approach toward automatically blocking domains that appear to have no impact to operations. For cases that require additional review, a human is brought into the process so that policy and operations can be maintained.

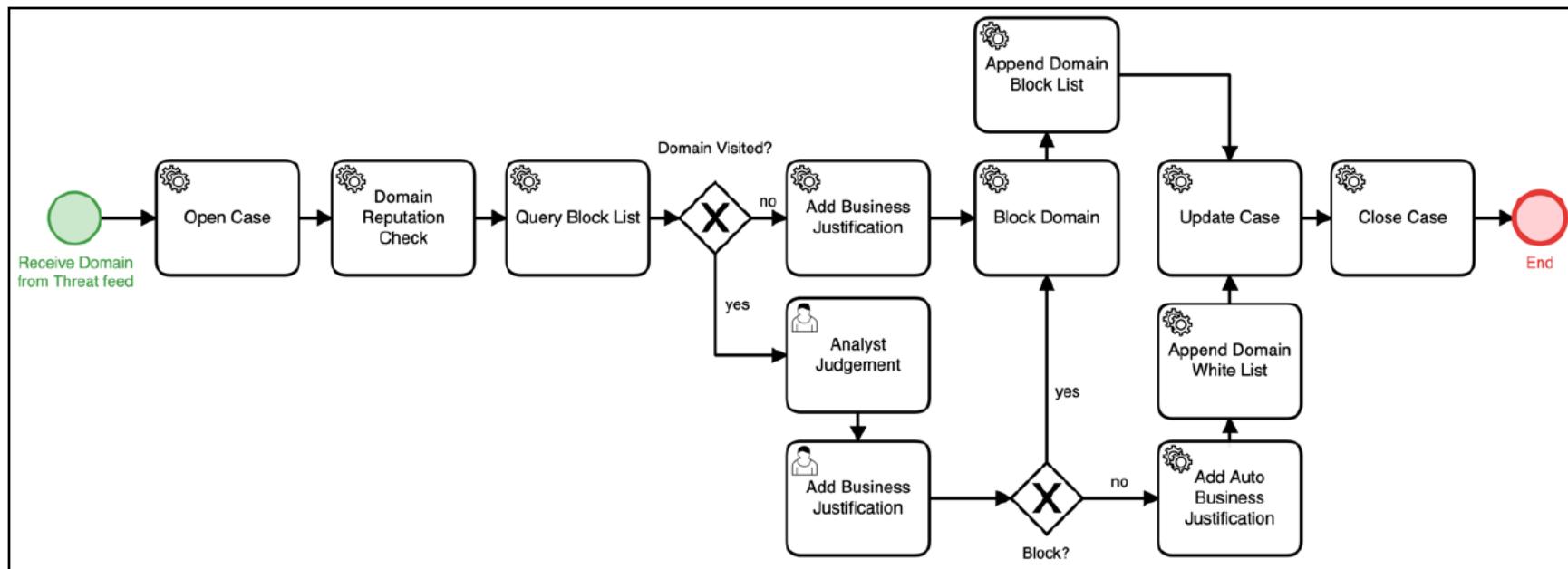


Figure 21 Response to Domain IOC example 1

Some organizations may have additional teams, policies and resources that require the information derived from the SOC analyst. Figure 22 provides an example of an orchestration workflow that allows information to be stored in multiple locations to support these other tasks.

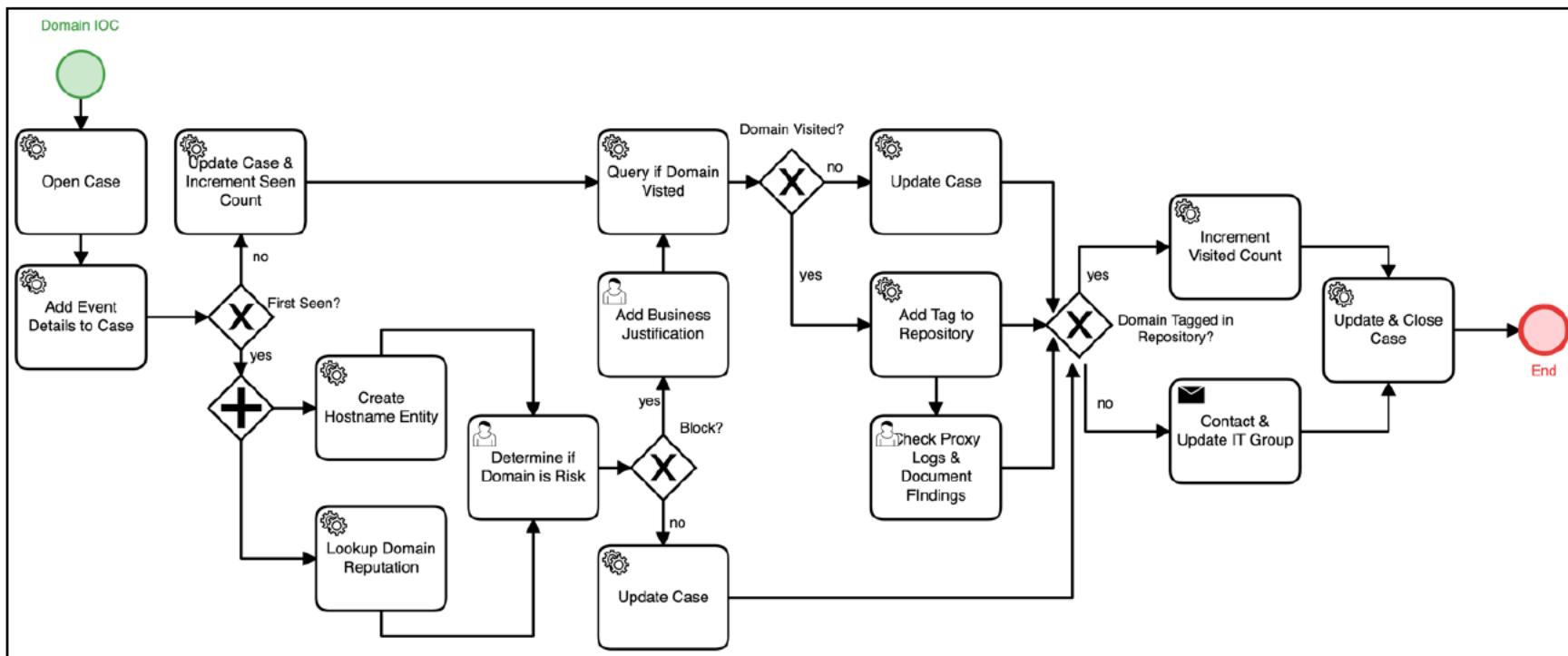


Figure 22 Response to Domain IOC example 2

In other environments, the resources available to evaluate a domain may be limited. For these situations, an organization may wish to optimize the number of IOCs evaluated and responded to via automation as early as possible in the process. Figure 23 and Figure 24 provide such examples.

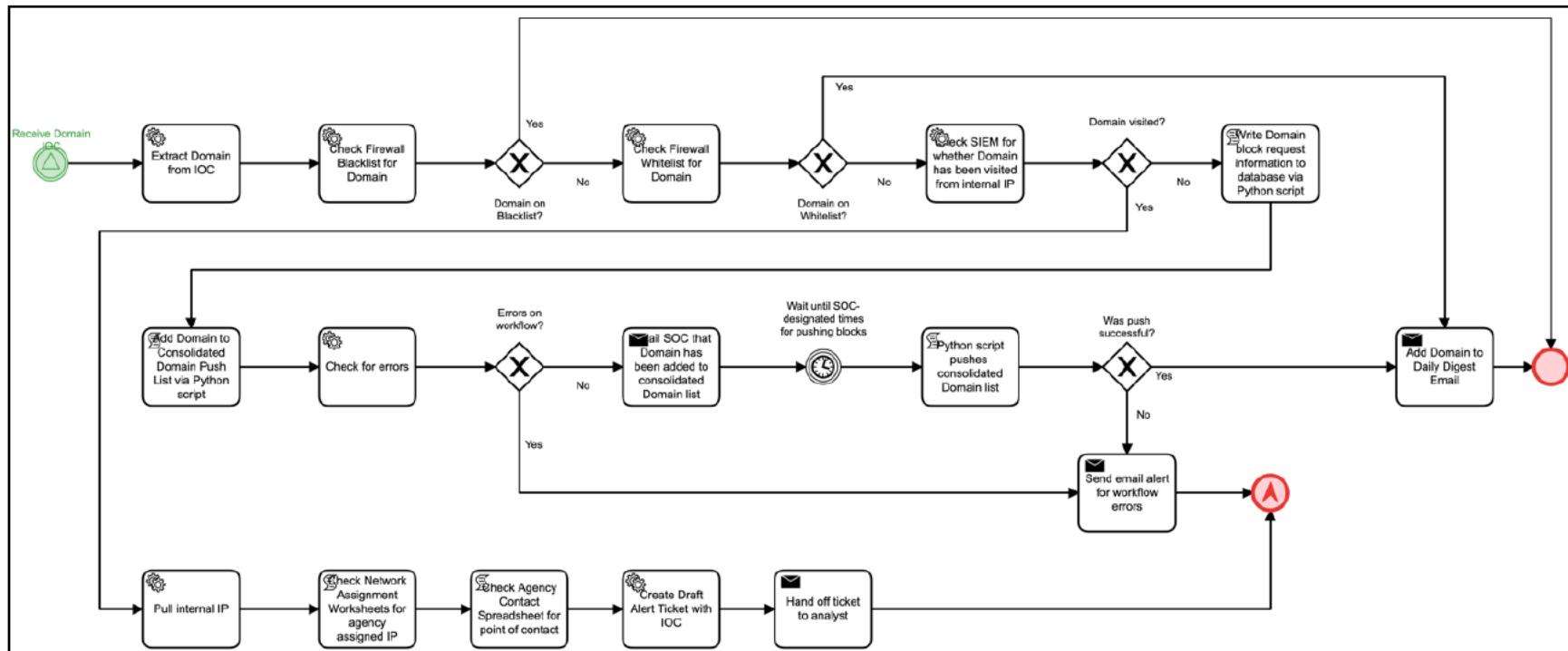


Figure 23 Response to Domain IOC example 3

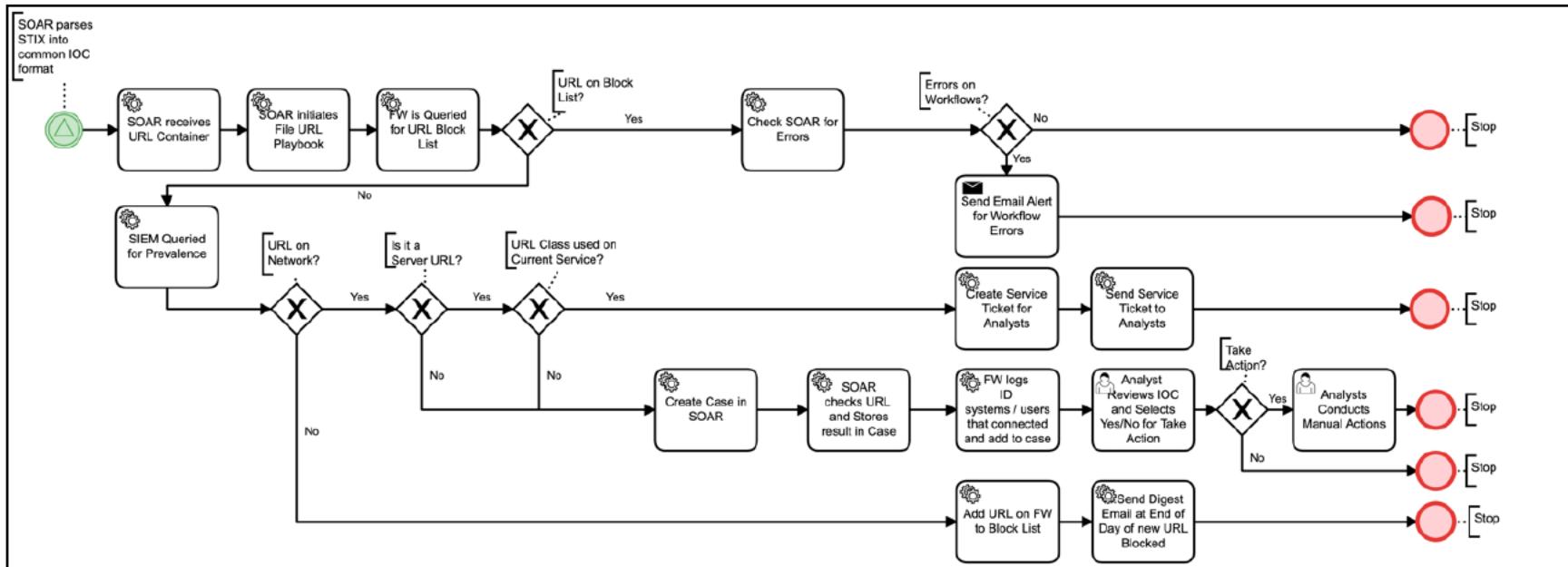


Figure 24 Response to Domain IOC example 4

There are other organizations that may not be willing to provide a high level of trust in automation due to their own risk tolerance and policy. Figure 25 provides an example of how automation may be used to remove the most mundane and repetitive tasks while keeping a human analyst in the loop for all decisions with respect to domain IOCs.

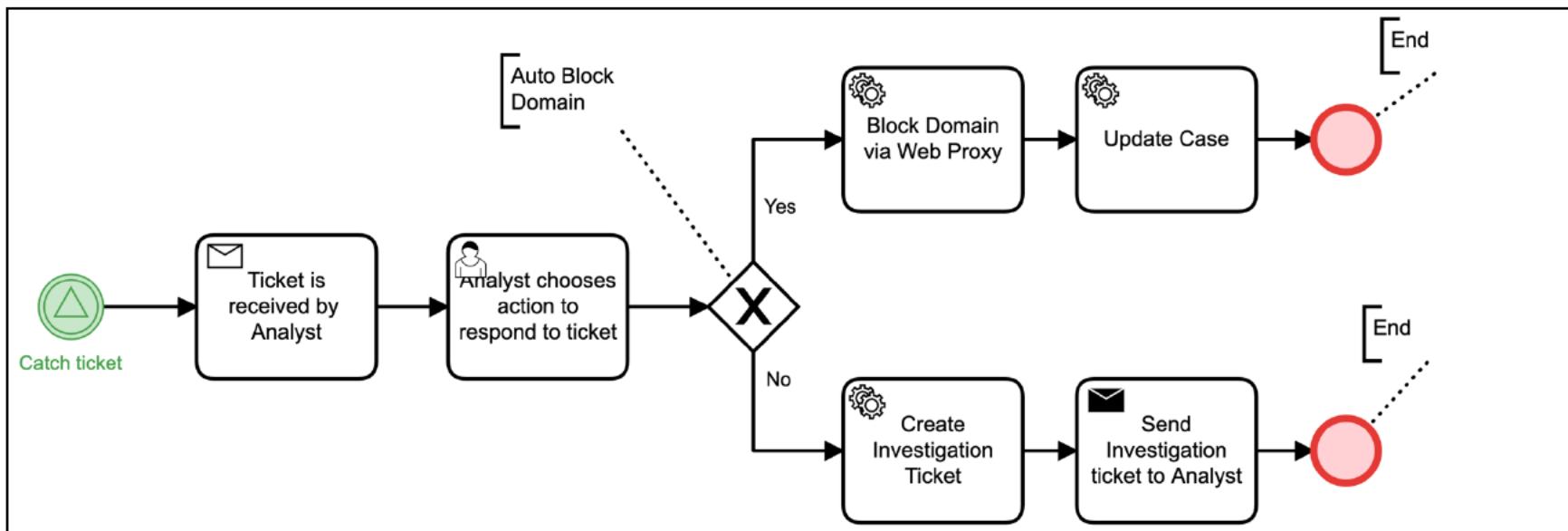


Figure 25 Response to Domain IOC example 5

2.7 Shareable Workflows for receipt and response to File Hash IOCs

Response to file hash IOCs have unique features due to the permanence of the IOC. Since a file hash is fairly permanent, if it is attributed to malware, it will always be attributed to malware. This allows for more permanent action as well. These workflows provide examples of how organizations can apply automation and orchestration against these threats.

Figure 26 provides an example of how to apply an organization's policies to block malicious files while taking a "low-regret" approach toward automatically blocking files that appear to have no impact to operations. For cases that require additional review, a human is brought into the process so that policy and operations can be maintained.

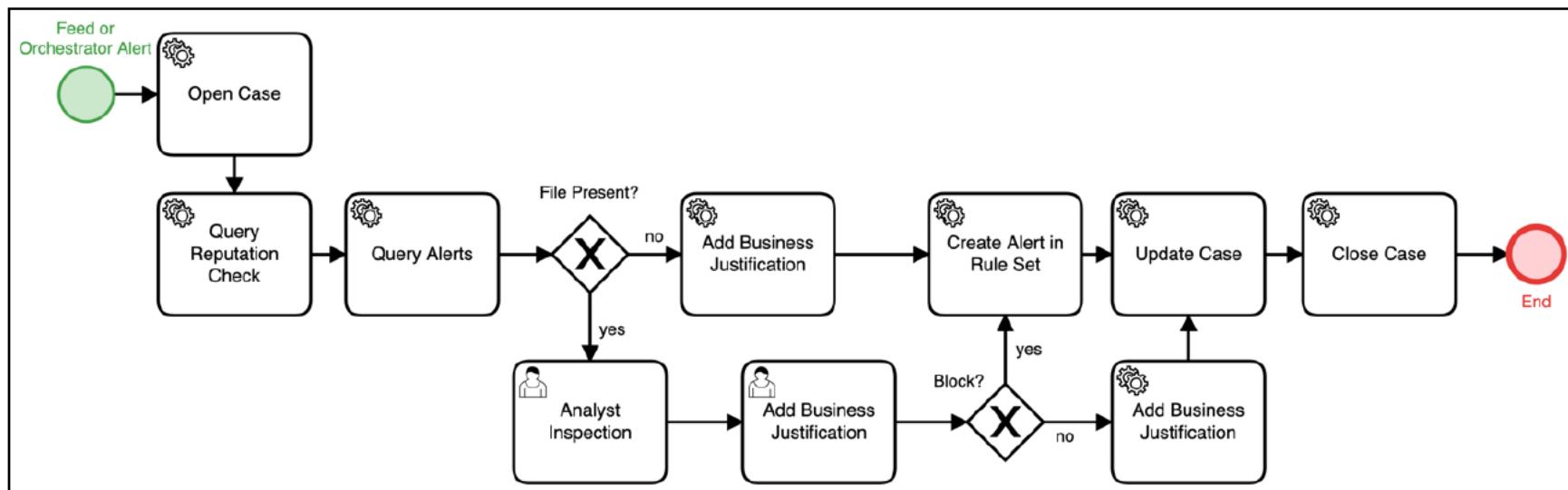


Figure 26 Response to File Hash IOC example 1

Some organizations may have additional “rules of thumb” applied for assessing potential risk. When these policies are fairly static, they can easily be adopted by automation. Figure 27 provides an example of using automation for this purpose, while preserving the key decision points for the human analyst, mainly due to the permanence of the action for banning files from the network.

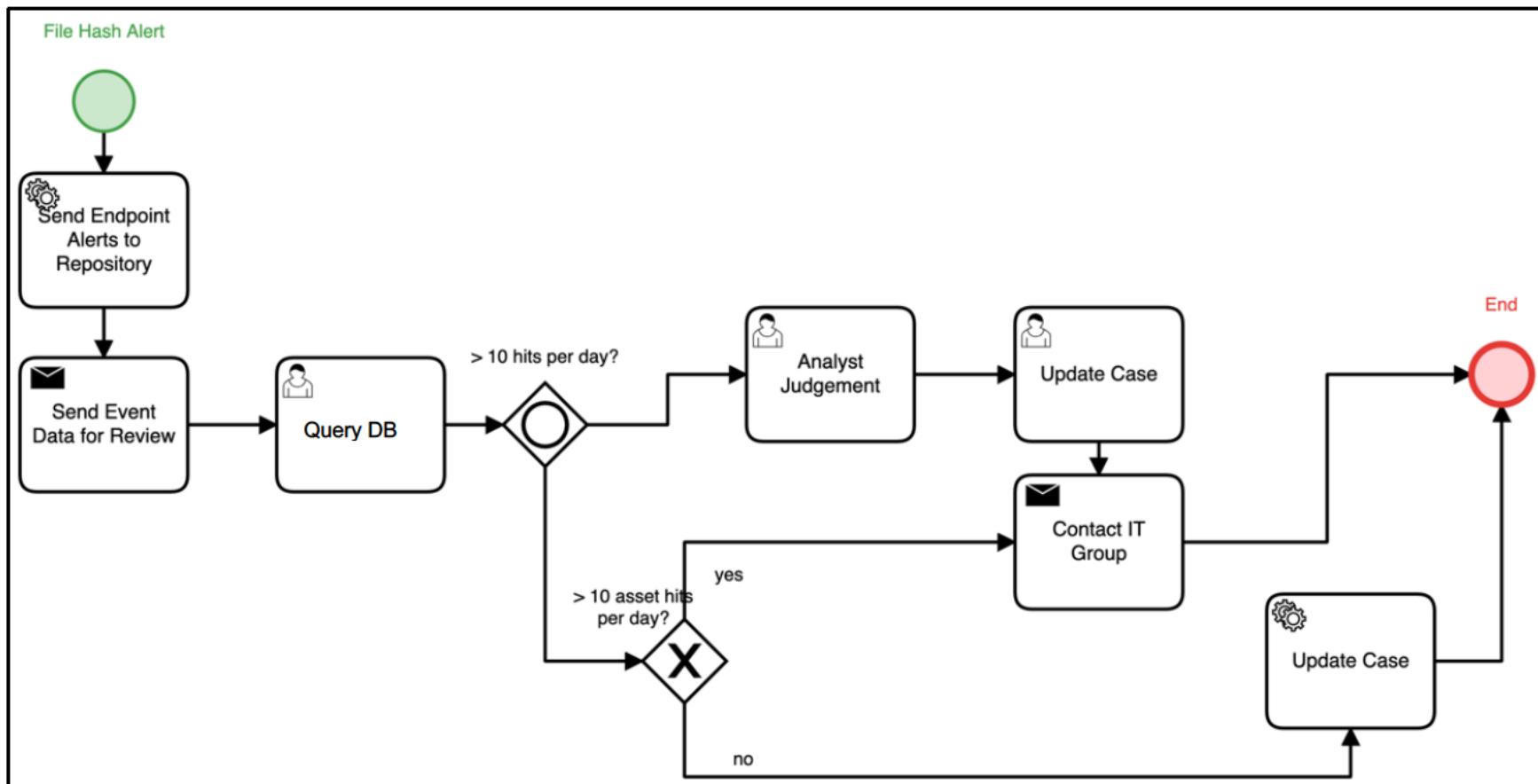


Figure 27 Response to File Hash IOC example 2

Whereas some organizations may keep their local enrichment in a common “data lake”, others may need to query their tools directly to learn necessary information for making decisions on file hash IOCs. Figure 28 provides an example where a SOAR platform may interact with multiple tools not just to block the IOC but to gain the necessary information to support both manual and automated decisions.

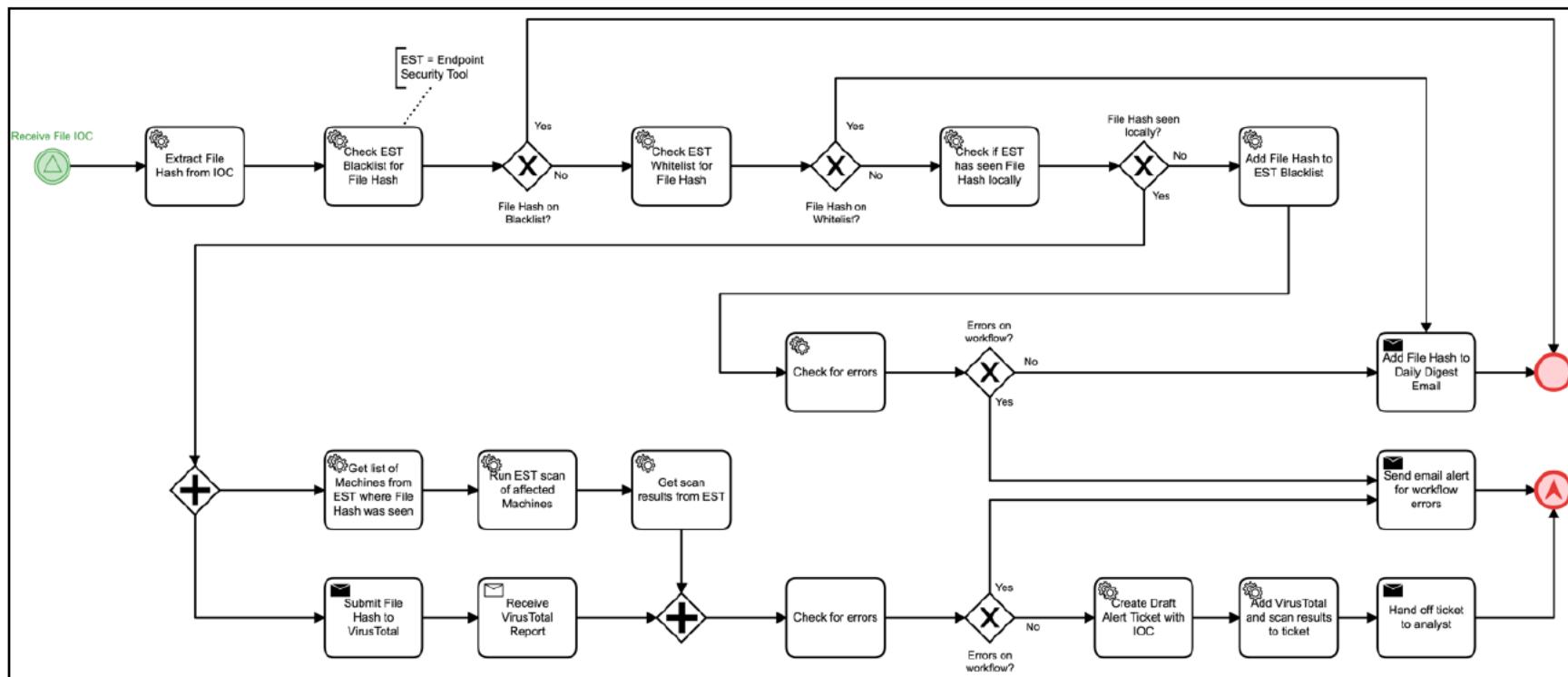


Figure 28 Response to File Hash IOC example 3

The type of asset that is potentially impacted by a cyber threat can require different courses of action for response either due to the nature of the asset or the jobs assigned to it. Figure 29 provides an example of how those types of decisions can be applied to the response of a malicious file hash.

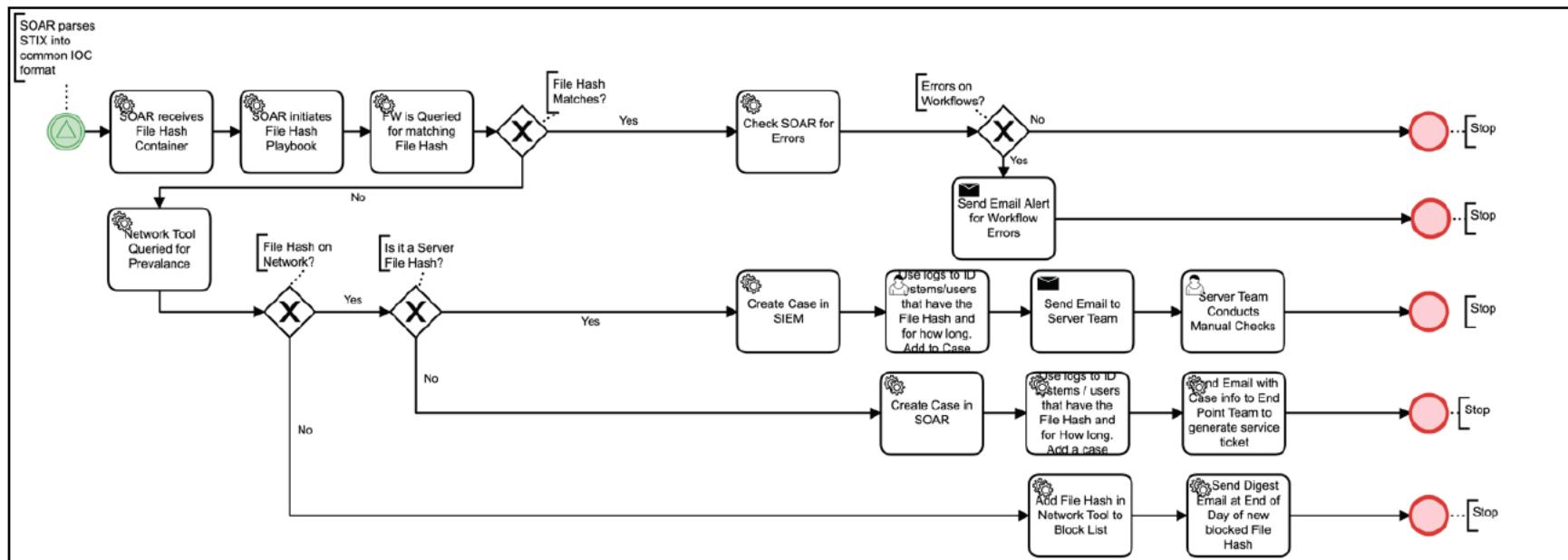


Figure 29 Response to File Hash IOC example 4

2.8 Shareable Workflows for receipt and response to Email Sender IOCs

IOCs for malicious email senders tend to have a very direct impact on operations. When the sender is sending malicious content, it may be unintentional. Additionally, banning a sender from emailing users within the enterprise may or may not have significant impact to operations. The following examples provide a guide for how different organizations within the SLTT community have chosen to address this challenge.

Figure 30 and Figure 31 provide examples of how to apply an organization's policies to block malicious email senders while taking a "low-regret" approach toward automatically blocking senders that appear to have no impact to operations. For cases that require additional review, a human is brought into the process so that policy and operations can be maintained. The difference in organizational policy directly impacts the changes in decision logic and complexity between these examples.

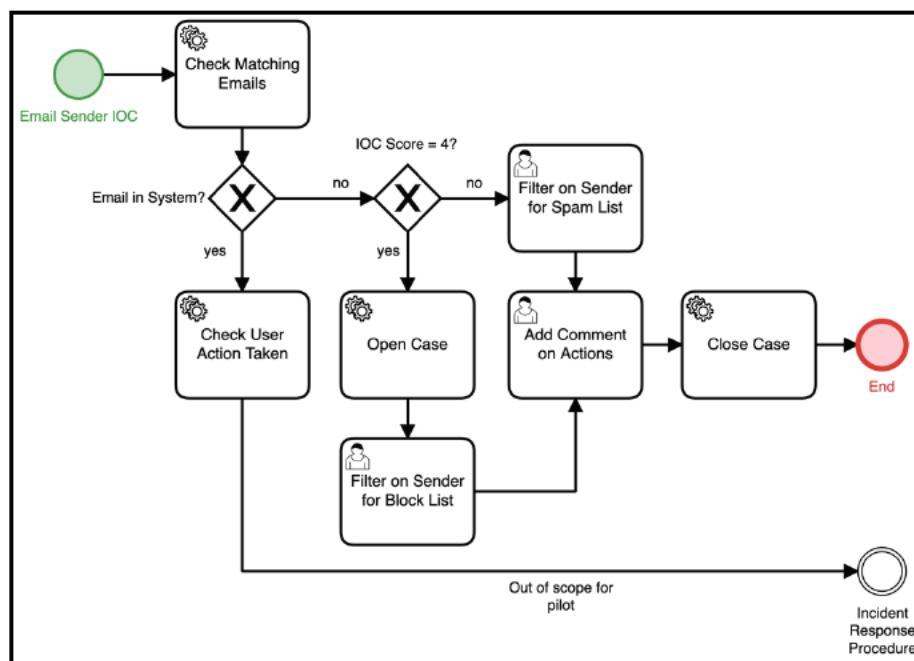


Figure 30 Response to Email IOC example 1

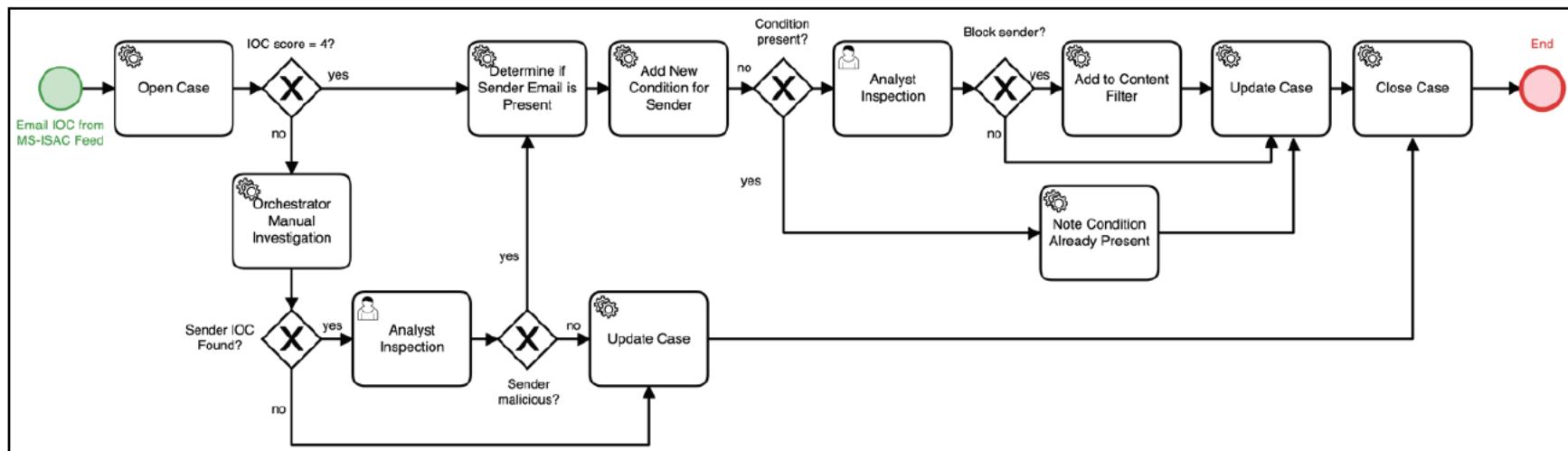


Figure 31 Response to Email IOC example 2

Sometimes the resources available to evaluate an email sender may be limited. For these situations, an organization may wish to optimize the number of IOCs evaluated and responded to via automation as early as possible in the process. Figure 32 provides such an example.

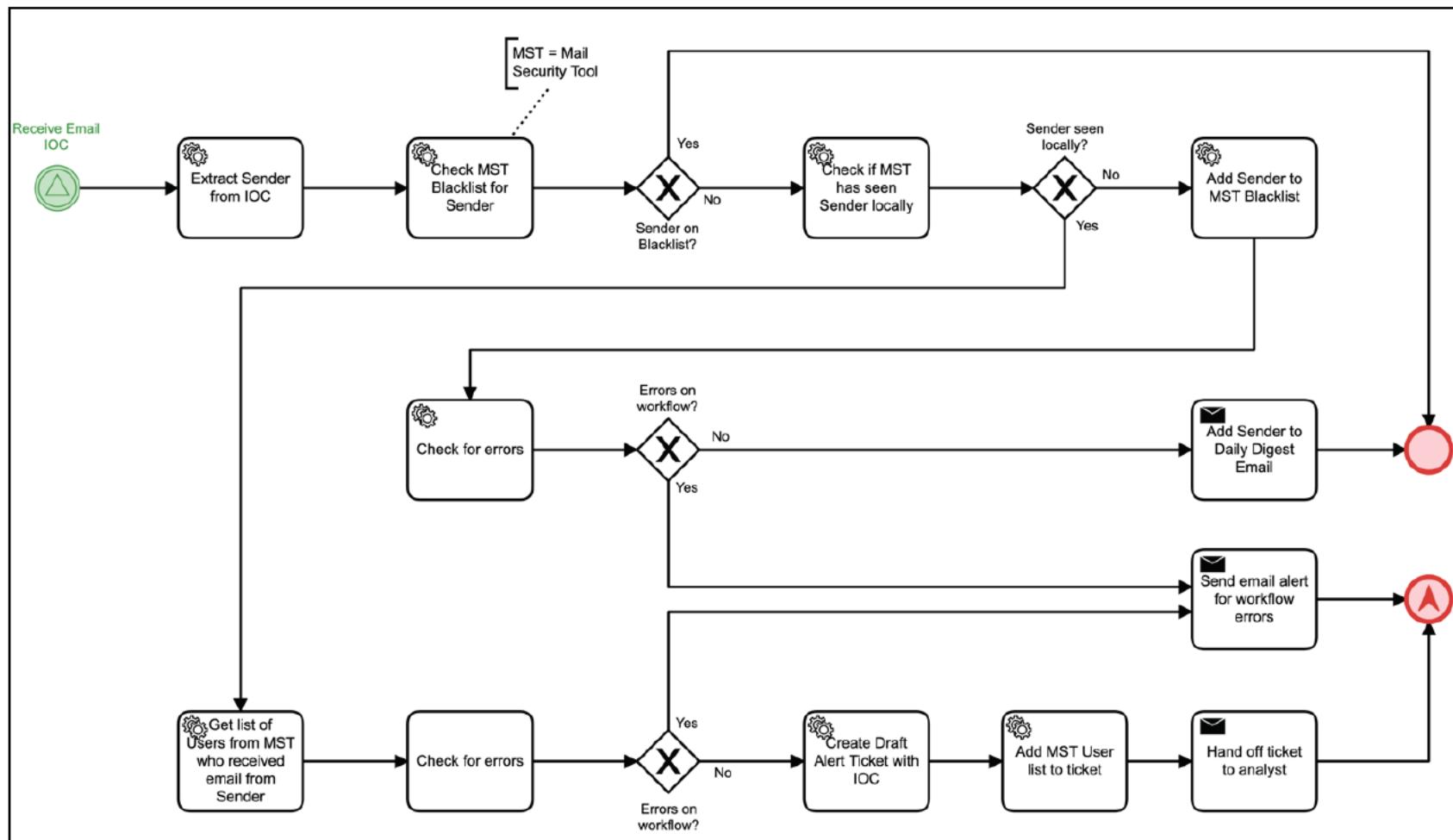


Figure 32 Response to Email IOC example 3

3. Summary

Shareable workflows for security automation and orchestration allow us to provide the community with simple guidelines for the design of their own SOAR local instances / runbooks. It is JHU/APL's intent that this document provides a starting point for organizations to initiate their efforts in designing and employing their own workflows.

For more information regarding orchestration, playbooks, and workflows, JHU/APL recommends guidance found from the Integrated Adaptive Cyber Defense (IACD) framework (<https://iacdautomate.org>) . IACD provides a large amount of information for free on the topics of orchestration and cyber threat information sharing. More detail on the topic of playbooks and workflows can be found at the following page on the IACD website:

<https://www.iacdautomate.org/intro-to-playbooks-and-workflows>