The Cyber WatchDog- DRAFT Use case

Author: Rich Wickersham

Enterprise Vulnerability WatchDog & RAG Applications:

Background and problem statement

In my twenty plus years in the Security industry, I have seen major technical and process driven improvements and efficiencies gained in many key areas across information security. One area that I **do not** believe has improved enough is the area of **vulnerability management**. The race to discover, assess, coordinate and continuously react to critical vulnerabilities (e.g. known exploitable or zero days) feels more like a hamster wheel of inefficiency than the well-oiled machine that it should be. The largest, most well-funded private and public organizations seem to struggle with this issue with results ranging from the scrambling of resources to address vulnerabilities to large avoidable incidents and news-worthy breaches. The need for a different approach to vulnerability management is paramount for organizations seeking to gain speed during a critical window, before an exploitable vulnerability provides an adversary a path to accomplishing their objective. With the emergence of Generative AI, I believe we have a viable path to solve this problem and build a new security capability.

Solution

Build a Privately hosted AI capability that will utilize retrieval-augmented generation, tuned large language models and machine learning to gain immediate insights from internal data sets that you control. External data sources will be used as a source of enrichment and verification that can reduce the time to identify, assess and respond to a zero day or a known exploitable vulnerability.

This approach and capability allows your security team to focus limited resources on other work that matters(e.g. Incident Response investigations) while the “AI-Cyberwatchdog” helps with the tactical work when a new threat or zero day emerges. This capability also provides your leadership with immediate answers, supportable metrics and the confidence in your AI driven response process that allows you to build toward a future SOC which is driven by data science, AI and efficient automation.

Use Case 1: Enterprise Vulnerability Watchdog:

The use case requires pulling data and summarizing data from at least three External Sources and correlation with at least one internal source. The first action that triggers the workflow is a retrieval and summarization of an RSS feed from CISA Known Exploitable Vulnerabilities(KEV). This was the most logical thread to use to represent a starting point that many of us have experienced firsthand. A Threat Intelligence Platform(TIP) or other source could also be used to trigger the workflow if the POC is expanded. On a daily basis the CyberWatchDog will listen, pull or search for the notification of a new zero day or known exploitable vulnerability and validate whether the vulnerability is applicable to your environment. Relevant mentions in other sources(e.g. Breach Forums) that align to these Known Exploitable Vulnerabilities in your environment will be also be pulled in. We will tailor this methodology to pull in data from CISA(or a TIP) at a specific interval, add our customer controlled CMDB data, Shodan, Pastebin, X, Mastodon, and Cyber Crime forum data, etc. An analyst will be able to have further interaction with the data via a web based UI and chatbot(Open-webUI) to pivot in whatever direction they need to during the analysis or an investigation.

The Vulnerability Watch Dog will generate summary data that is marked as new for a new threat report, or provide tailored updates to existing threat reports based on new information(gained through our tuned model that utilized RAG to pull in more relevant data into the vector DB). The process to execute this is as follows:

1. Pull from CISA.gov RSS a list of KEV (note this is a RAG App)
   https://www.cisa.gov/sites/default/files/feeds/known_exploited_vulnerabilities.json

2. Generate a summary of the new exploitable vulnerability(see examples below) or update to an existing vulnerability
   - 2.1. Prompts tuned and created to test summary data
   - 2.2. Summary data will include relevant terminology that can be used later(product name, version number, vuln, ports, protocols, etc)

3. Inventory: API pull from CMDB(I would suggest that we write a connector for Service Now) but we will connect to a databased with Synthetic data for now. CMDB tells us to moderate degree of confidence(I never trust an enterprise inventory as a single source of truth) whether the vulnerable product is in our environment, the software and firmware versions and the patch level.
   - 3.1. Claud AI will be used to generate the synthetic data until CMDB connector is built but the connector will need to be built for ServiceNow out of the box.
   - 3.2. Our LLM agent will include this in the output report with a confidence rating based on the assumed accuracy of the CMDB(in large organizations the CMDB should not be treated as a source of truth)

4. We will perform a RAG operation against OSINT data source (Shodan.io) because who can really trust a CMDB without verifying?
   - 4.1. Shodan will use several potential variables to narrow down a search to a specific client/target. These may include specific IP ranges, names, etc
     - 4.1.1. It may also be useful to capture the entire inventory of internet facing vulnerabilities in a region, sector or globally for further analysis. If a third party that has access to our customer environment also has the vulnerability, we will want that as part of our Threat Intel product.

4.2. The IP address range, company name(DNS) or the following query can be used, see query for vuln:<CVE ID> as a general starting point.

4.3. Our LLM agent will include a summary in the output report in a discoverability section

5. RAG operation against OSINT data sources or APIs for Pastebin, X, Mastodon, Google and add more.

5.1. A separate step will be integrated for all sources that are relevant to the vulnerability use case.   These use cases will require search strings/seed files to include the vulnerability name, POC code name, Company name(s), IP's, products.

5.2. Pastebin is the easiest to add and we have already created code for X, assuming the API costs don't exceed the ROI of the use case. Other code will be developed as time allows.

5.3. Our LLM agent will include this mentions summary in the output report

6. RAG and Python scraping of Cyber-Crime forums and possibly Telegram channels with keywords(from step 1) against 2-3 known cybercrime forums(Breach, RAMP, Exploit, PwnedForum, XSS) for mention of vulnerability key words, company name(e.g. Citrix, Ivanti), threats, a PoC or relevant data.   NOTE: This requires commercial proxy infrastructure that will need to be provided by the client(e.g. a CTI vendor acting on behalf of the client) or built out as part of the service.

6.1. Seed/Keyword files based on step #1 and preexisting target/customer information (seed files) will be used to tune the prompts and the output

6.2. We can potentially deliver mentions of exploits that are not yet in KEV database for further analysis. If threat actors are discussing a zero day, an exploit, TTPs we want to know in advance. This is outside of the core use case but still would provide value.

6.3. We can allow variations on the queries, so the analysts can ask questions about new vulnerability/TTP discussions that are relevant to the target company, specific threats that are actively discussed in these forums regarding the customer. Its just a matter of knowing the proper keywords(e.g. customer name, cve, etc) to construct the prompt.

6.4. Our LLM Agent will deliver a summary report and this data will be stored in low cost storage for further analysis down the line.

7. Generate repeatable deliverables and allow for interaction(chat web-ui) with the data set we have generated:
   - 7.1. Human Analysts runs queries in query engine: Is a working PoC available for Ivanti Pulse VPN? How many threat actors are talking about this exploit/what is approximate interest level on a scale of 1-10 (LOW<5 Med 10<HIGH).
     - 7.1.1. Note, that the interest level will require tuning and parameters and must be a score that is based on mentions or other weighted criteria)?
   - 7.2. Output should reference links to the code or requests if data is copied and locally stored. Additional questions should be expected once the basic information is retrieved.
   - 7.3. Note that some prompts will be saved and indexed for reuse. A large number of analysts will want to build off the work that others have done.
     - 7.3.1. The private AI system(Ollama Web UE) captures all prompt data so we can run Generative AI against the prompt data to provide a list of commonly used/high value prompts over a time period for this purpose.

8. Deliver summary reports/ PowerBI/Tablau Products, Dashboards for Recommend Mitigations, remediations and actions
   - 8.1. Separate deliverables are provided for each consumer group

---REPORT SAMPLE-------------------

Begin tailored Summary REPORTS(raw text version from LLM):

The threat landscape for your corporation is elevated due to two known exploitable vulnerabilities that we are currently investigating. Two critical CVE's for ScreenWise and Ivanti VPN were published on (list date here). Our analysis reveals forum posts and other artifacts from several threat sources showing interest(mentions) in exploiting critical vulnerabilities in **Ivanti Pulse VPN**, and **ScreenConnect.**

- **IVANTI: We have determined through verification against our internal CMDB and External Shodan Validation that the Ivanti Pulse VPN devices in our environment are external facing and running a vulnerable version of the code.**
- **ScreenConnectOur initial analysis of Screen-Connect indicates that it is only accessible internally.**

NOTE: This effort intentionally focuses on zero-day exploits/Known-Exploitable Vulns, and the availability of proof-of-concept (PoC) code, or mentions of data that is relevant to the target on the "dark web".

- CVE-2024-22024- CRITICAL PRIORITY

- Ivanti CVE-2024-21887: A command injection vulnerability with a CVSS score of 9.1
- Ivanti CVE-2023-46805: An authentication bypass vulnerability with a CVSS score of 8.2

- Connectwise CVE-2024-1709: An authentication bypass that uses an alternate path or channel. This vulnerability has a CVSS score of 10, which is rated as critical.
- ConnectWise CVE-2024-1708: A path traversal issue with a CVSS score of 8.4, which is rated as high

Ivanti CVE 2024-21887 URL: Here

This CVE is an external entity injection (XXE) vulnerability that effects Ivanti Connect Secure and Policy Secure devices. A threat actor could take control of the impacted devices to gain a foothold in the environment if they are not patched immediately. Our CMDB indicates that these devices are in our inventory and running an exploitable version of the software. Our External research indicates that our devices are discoverable on the internet(see Shodan reference link query). Our Dark web forum search indicated that we have **33 mentions of POCs**, **22 mentions of exploitation** by groups including Ransomware Initial Access Brokers indicating that active exploit may be ongoing. We have no mention of our company or subsidiaries at this time.

Ivanti CVE 2024-22024 Actions:

Our AI driven recommendations are to patch these Ivanti devices immediately and to begin investigative threat hunting until we have proven that the Ivanti devices were not compromised. Our threat hunting will include the TTPs that have been associated with the threat actors that are actively exploiting these devices.

Screen Connect CVE-2024-1709:

These vulnerabilities affect Screen Connect versions 23.9.7 and earlier. Cybercriminals have chained the two vulnerabilities, first using the authentication bypass CVE-2024-1709 and then moving through the system with the path traversal CVE-2024-1708.

ConnectWise states that **on-premise** Screen Connect, partners active with maintenance are recommended to upgrade to the most current release of 23.9.8 or later.

Screen Connect Actions:

Our AI driven recommendation is to patch these to the latest patch level immediately and validate that inbound firewall rules block connections from untrusted networks to TCP/8040-8041. Additional threat hunting will be required if any unauthorized internal use was detected prior to applying the patch.

--------------------------------END---

9. This report will be created and built or supplemented with aggregated data from prior runs(if the report is an update), all data will be stored in a low-cost storage(SNOW, ADLS, S3, etc) or directly in the vector DB and we will run our private, opensource language models against it.

   9.1. We will refine/tune with system/user prompts to deliver the virtual analyst results related to a KEV that are relevant to our customer environment.

   9.2. The primary goal that we achieved was to identify the vulnerability, validate the exposure in our environment and to determine/add other data points that are contextually relevant to our queries or aligned to recent critical CVE's or technologies that we have in our client environment like Ivanti, Screenwise or mentions from OSINT and dark web sources. This has enabled our analysts to run a private or public LLM to learn from the data going forward.

Note: A public LLM can also be used via a hybrid deployment if the infrastructure(UI and proxy and DSPM) is built for the customer to consume OpenAI to generate summaries without exposing any of our data or research

1 Short Analysis and Summary of the entire use case(used for the demo):

9.3. Picked up a known exploitable vuln from CISA KEV or our via our TIP from another source and delivered summary

9.4. Retrieval Augmented generation for our CVE data against our CMDB added to summary

9.5. Retrieval Augmented generation to validate external availability of vulnerable resources via Shodan and added to summary

9.6. Retrieval Augmented generation for the top dark web forums for useful matching cybercrime data(Exploit, RAMP, Breach, PwnedForum, XSS) and added to summary as mentions

9.7. Retrieval Augmented generation of multiple OSINT data sources that we have scraped from Shodan, Pastebin, X, Mastadon ,Google or other sources added to summary

9.8. Recommendations on immediate actions that should be taken delivered as summary reports with links to data sources(some of these data sources must be sandboxed)

10. The Cyber WatchDog will ultimately provide this to a client so they can create recurring reports for investigations they are interested in. Reports will be generated that serve as a starting point. In the future state the clients can use prompts to interact with the data to **deliver the data they need when they need it** through the web ui**.**

Value proposition:

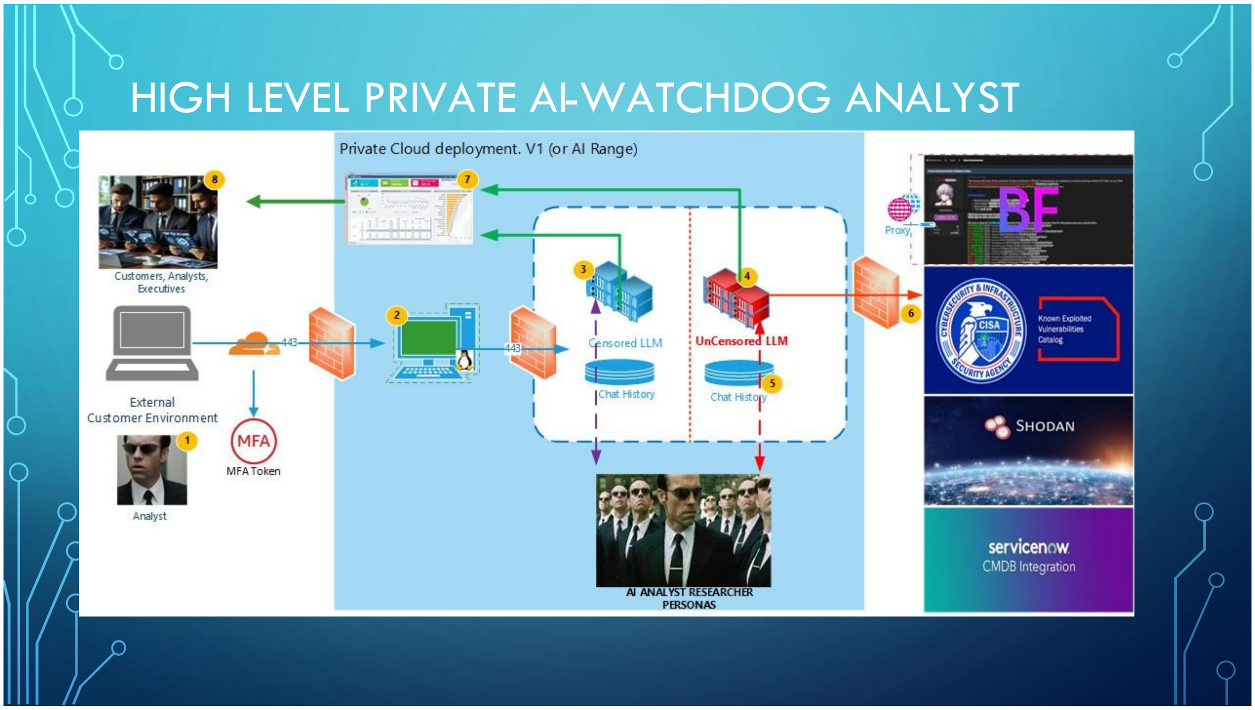
We will make analysts more productive by taking the routine work that happens in a fire drill fashion when a zero day is released and generating organized summary reporting for further action. The WatchDog will generate actionable intelligence via a virtual analyst(s) for a targeted audience(CISO, Director, Analyst). We can use multiple AI agents or another AI capability to hand from one virtual Analyst to another and to create products for an audience member(e.g. Tier III CTI Analyst, or Director, or C level executive)

Analysis:

Every capable adversary(nation state, ransomware) on the planet is using Private AI to scan and exploit your environment. If you are not building this capability today you are behind and its time to catch up!

2 Appendix:

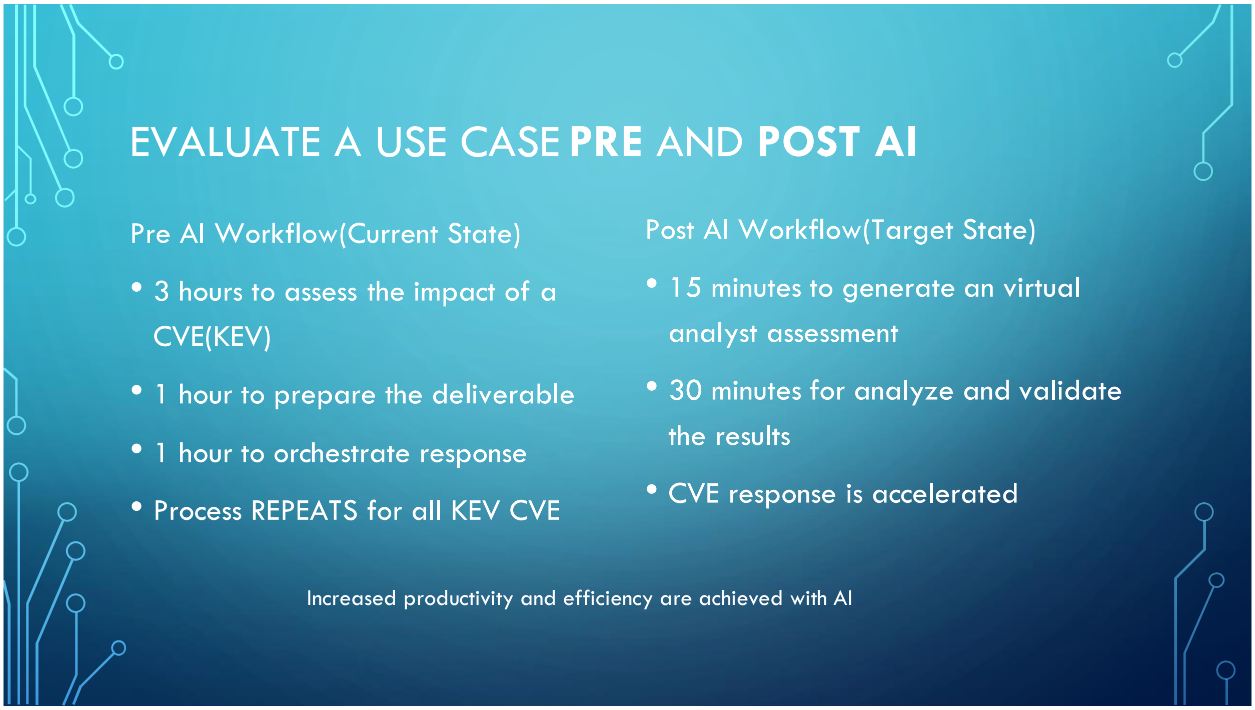
Slides from planning deck:



HIGH LEVEL PRIVATE AI-WATCHDOG ANALYST

Private Cloud deployment. V1 (or AI Range)
8
Customers, Analysts, Executives
7
External Customer Environment
443
2
443
3
Censored LLM
Chat History
4
UnCensored LLM
Chat History
5
1
Analyst
MFA
MFA Token
AI ANALYST RESEARCHER PERSONAS
Proxy
BF
6
CISA
Known Exploited Vulnerabilities Catalog
SHODAN
servicenow
CMDB Integration

Outcome: Pre and Post AI



EVALUATE A USE CASE **PRE** AND **POST AI**

Pre AI Workflow(Current State)

- 3 hours to assess the impact of a CVE(KEV)
- 1 hour to prepare the deliverable
- 1 hour to orchestrate response
- Process REPEATS for all KEV CVE

Post AI Workflow(Target State)

- 15 minutes to generate an virtual analyst assessment
- 30 minutes for analyze and validate the results
- CVE response is accelerated

Increased productivity and efficiency are achieved with AI

3 Appendix

See Chart of current zero-day process vs updated process with the help of AI

Process	**Current state**	**Target State**
Identify	TIP/CISA drives manual kick off	RAG based automation
Identify and analyze	CMDB	RAG Automation
External reconnaissance	Human driven	RAG Shodan CVE search
Summary and Action Plan	Running multiple private models to gauge quality of prompts	Multi-models, prompt based interrogation of the data set.

Code: added current project code to GitHub