Bộ môn An toàn Thông tin – Khoa MMT&TT
Trường Đại học Công nghệ Thông tin (UIT)

BÁO CÁO THỰC HÀNH

Môn học: Pháp chứng kỹ thuật số

Kỳ báo cáo: Seminar-Cyber Threat Intelligence (CTI) and OSINT 1

GVHD: Phan Thế Duy

Team: NBG

- **THÔNG TIN CHUNG:**
  (Liệt kê tất cả các thành viên trong nhóm)
  Lớp: NT334.M21.ANTN

STT	Họ và tên	MSSV	Email
1	Lâm Thanh Ngân	19521884	19521884@gm.uit.edu.vn
2	Lê Hồng Bằng	19520396	19520396@gm.uit.edu.vn
3	Nguyễn Ngọc Quỳnh Giang	19520500	19520500@gm.uit.edu.vn

Table of Contents

BÁO CÁO CHI TIẾT

1. Ngữ cảnh

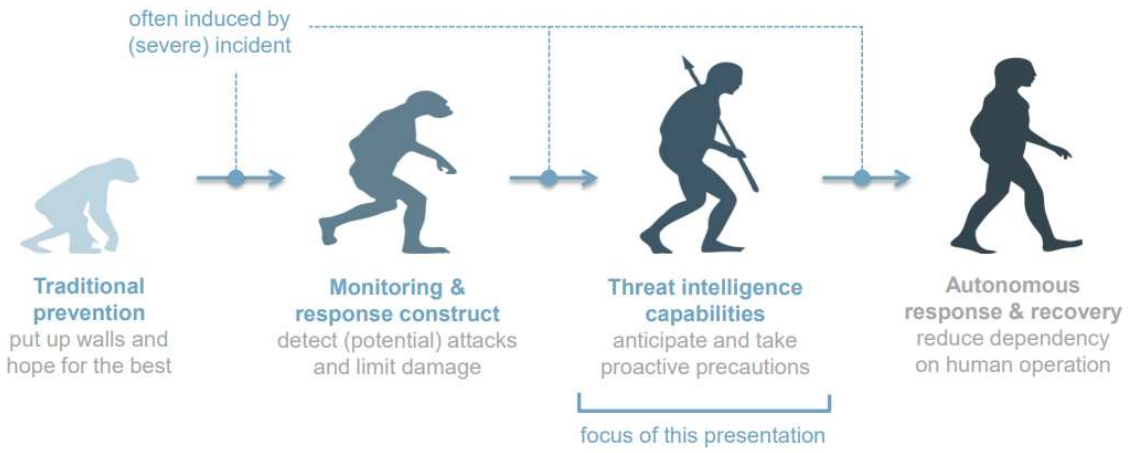


Figure 1: Sự phát triển của việc chống lại các cuộc tấn công xâm nhập

Những tôi phạm mạng không có kỹ thuật cao cũng có thể khởi chạy các cuộc tấn công bằng cách sử dụng các bộ công cụ có thể tải xuống trực tuyến miễn phí hoặc với chi phí rất nhỏ. Kết hợp điều đó với các tài nguyên trực tuyến như YouTube và có hàng ngàn tùy chọn tấn công và hàng triệu người dùng sẵn sàng. Theo kịp tất cả những điều này là một công việc không hề dễ dàng, dẫn đến việc mở rộng một lĩnh vực mới của an ninh mạng được gọi là cyber threat intelligence.

2. Cyber threat intelligent and OSINT

Cyber threat intelligence – Thông tin mối đe dọa an ninh mạng (CTI) là một lĩnh vực của an ninh mạng tập trung vào việc thu thập và phân tích thông tin về các cuộc tấn công hiện tại và tiềm năng đe dọa sự an toàn của một tổ chức hoặc tài sản của tổ chức. Lợi ích của thông tin mối đe dọa là đó là một biện pháp bảo mật chủ động, ngăn chặn vi phạm dữ liệu và tiết kiệm cho bạn chi phí tài chính để dọn dẹp sau khi xảy ra sự cố. Mục đích của nó là cung cấp cho các công ty một sự hiểu biết sâu sắc về các mối đe dọa gây rủi ro lớn nhất cho cơ sở hạ tầng của họ và cho biết những gì họ có thể làm để bảo vệ doanh nghiệp.

Làm thế nào để sử dụng cyber threat intelligence?

Bạn có thể có được thông tin về mối đe dọa bằng cách thuê một nhà cung cấp dịch vụ CTI, người sẽ làm việc với nhóm bảo mật hoặc IT của bạn một cách thường xuyên. Họ sẽ giải thích không chỉ các mối đe dọa mà còn cách phòng ngừa. Khi nhóm của bạn có

thông tin đó, bạn có thể thực hiện các điều chỉnh để đảm bảo doanh nghiệp của bạn sẽ không trở thành nạn nhân của bất kỳ mối đe dọa nào.

Có thể lợi ích lớn nhất của CTI là nó cung cấp cho bạn khả năng phòng thủ chủ động, đảm bảo bạn có thể tự bảo vệ mình trước khi bạn phải chịu bất kỳ chi phí nào. Nó cũng có thể giúp bạn tìm ra nếu bạn đã bị vi phạm bằng cách sử dụng các chỉ số thỏa hiệp (IOC) để xác định xem hệ thống của bạn có bị nhiễm phần mềm độc hại hay không. Một phần mềm độc hại càng lâu không bị phát hiện trên một hệ thống, nó sẽ càng đánh cắp nhiều thông tin và càng tốn nhiều chi phí trong thời gian dài.

Một ví dụ phổ biến về điều này là một loại phần mềm độc hại được gọi là phần mềm gián điệp, có thể được cài đặt trên thiết bị máy tính mà bạn không biết để có được dữ liệu sử dụng internet và thông tin nhạy cảm khác. Trong môi trường kinh doanh, đây có thể là thông tin thẻ tín dụng, thông tin cá nhân của khách hàng và nhân viên, v.v.

Phần mềm độc hại đắt nhất trong lịch sử, Mydoom, gây ra thiệt hại ước tính 38,7 tỷ đô la và là vi rút lây lan nhanh nhất từ trước đến nay. Một số sự cố có thể đã được ngăn chặn nếu các công ty biết nó lan truyền như thế nào, chủ yếu qua email sử dụng tám dòng chủ đề chính. Ngay cả dịch vụ CTI cơ bản nhất cũng sẽ bắt và ngăn chặn virus. Thông tin mối đe dọa an ninh mạng tốt sẽ cung cấp cho bạn các IOC giúp bạn phát hiện phần mềm độc hại như thế này trước khi nó tiêu tốn của bạn một khoản tiền vô lý.

Threat Data quan trọng như thế nào?

Đối với môi trường mạng hiện đại có rất nhiều thực thể vô cùng phức tạp nên để bảo vệ tốt cho môi trường mạng đòi hỏi người bảo vệ hệ thống mạng cần hiểu biết chi tiết cách thức vận hành của các thành phần trong hệ thống mạng. Hiểu rõ cách thức hoạt động của môi trường mạng giúp bạn ứng phó với các mối đe dọa nhanh hơn khi mà các phương thức tấn công mạng ngày càng tinh vi.

Phần lớn các tổ chức không đủ khả năng tài chính để đầu tư vào công nghệ phát hiện hoặc chống các phương thức tấn công ngày càng phát triển. Hơn nữa đầu tư công nghệ thường không hiệu quả bằng việc đầu tư vào các nhà phân tích mối đe dọa giỏi, những người có thể thu thập và hiểu nhanh dữ liệu về các mối đe dọa (Threat data) mà tổ chức phải đối mặt.

Dữ liệu về mối đe dọa (Threat data) khi được cung cấp trong bối cảnh thích hợp sẽ đưa ra dữ liệu hưu ích về mối đe dọa hoặc thông tin về các tác nhân độc hại và hành vi của chúng, thông tin này cho phép những người bảo vệ hệ thống hiểu rõ hơn về môi trường hoạt động của họ. Một số sản phẩm có thể được sử dụng nhằm cung cấp cho những người ra bảo vệ hệ thống mạng một bức tranh rõ ràng về những gì đang thực sự xảy ra trên mạng, giúp họ đưa ra quyết định tự tin hơn và chính xác, tăng chi phí và thời gian cho những kẻ tấn công, cải thiện thời gian ứng phó của người vận hành hạ tầng trong trường hợp xảy ra sự cố, giảm thời gian phục hồi đối với tổ chức.

Như vậy để có được dữ liệu về mối đe dọa chúng ta cần có giải pháp tình báo về mối đe dọa (threat intelligence) là yếu tố cần thiết trong kế hoạch ứng phó an ninh thông tin trong môi trường mạng hiện đại.

Nền tảng của threat intelligence.

Tình báo truyền thống (threat intelligence) liên quan đến việc thu thập và xử lý thông tin chuyển về từ nước ngoài của các nhà hoạt động tính báo. Thường được tiến hành dưới sự điều hành của chính phủ, các hoạt động tình báo được thực hiện nhằm phục vụ các mục tiêu chính sách đối ngoại xa hơn là hỗ trợ an ninh quốc gia. Một khía cạnh quan trọng khác là bảo vệ các hoạt động tình báo, những người và tổ chức liên quan và các hoạt động tình báo chống lại việc tiết lộ trái phép. Thông tin tình báo được chia thành nhiều loại và các khu vực mà nó được thu thập, như thể hiện trong bảng bên dưới. Về mặt chức năng, chính phủ có thể sắp xếp một cá nhân hoặc một số cơ quan hoạt đông lĩnh vực tình báo.

Kỹ Thuật Mô Tả

SIGINT Signal Intelligence là thông tin được thu thập thông qua nghe lén trò chuyện dựa trên thiết bị điện tử

HUMINT Human intelligence là nguồn thông tin được chuyển từ những người hoạt động trong lĩnh vực tình báo

OSINT Open source intelligence là thông tin được phân tích và thu thập từ các thông tin được công bố như công thông tin điện tử……

MASINT Measurement and signature intelligence là dữ liệu tình báo không phải là hình ảnh và SIGINT

GEOINT Geospatial intelligence là phân tích hình ảnh và dữ liệu không gian địa lý liên quan đến các hoạt động, liên quan đến an ninh trên trái đất

All SOURE Là thông tin tình báo được gửi từ tất cả các nguồn.

Ví Dụ: Hoạt Động Tình Báo Truyền Thống

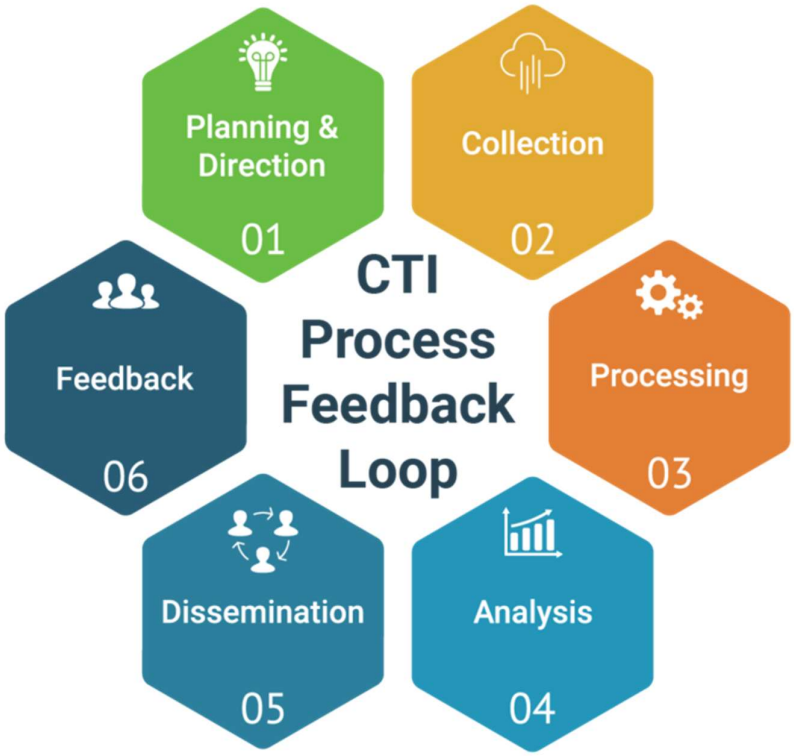
Nguồn thông tin tình báo mối đe dọa.

Các đội tình báo về mối đe dọa bên ngoài chính phủ thường không có đủ công cụ hoặc thẩm quyền tình báo theo yêu cầu cho những người trong chính phủ. Công ty năng lượng và nhà các nhà cung cấp dịch vụ internet (ISP) họ không triển khai hoạt động tình báo dựa trên HUMINT hay là SIGINT, nhưng họ sẽ tập trung vào việc sử dụng bất kỳ nguồn lực công cộng, thương mại hoặc nội bộ nào có sẵn. May mắn thay, có một số nguồn miễn phí và trả phí để giúp các nhóm đáp ứng các yêu cầu tình báo của họ, bao gồm các nhà cung cấp thương mại thông tin tình báo về mối đe dọa, các đối tác trong ngành và các tổ chức chính phủ. Và tất nhiên, cũng có sự giám sát rộng rãi trên mạng xã hội và tin tức để có dữ liệu liên quan.

Nguồn thông tin tình báo mở.

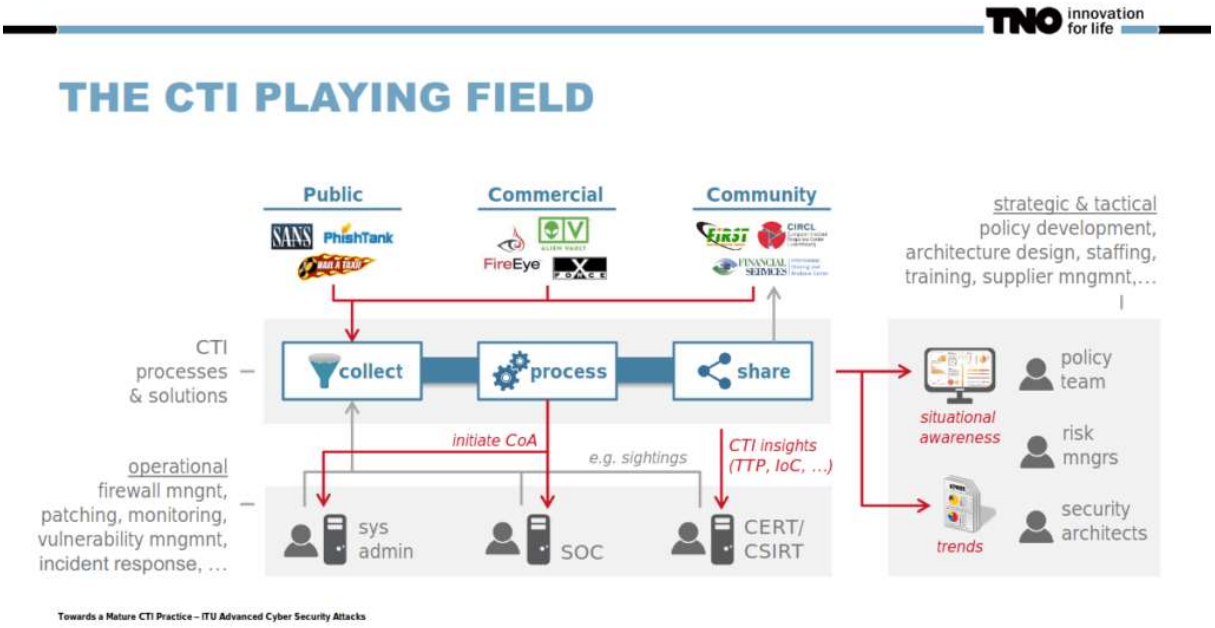
Có nhiều cách để lấy dữ liệu về mối đe dọa (Threat data) miễn phí liên quan đến hoạt động của tình báo, có thể là độc hại hoặc không độc hại. Một cách phổ biến để thu thập dữ liệu mối đe dọa mà không cần tương tác với các bên là thu thập thông tin tình báo nguồn mở (Open source intelligence – OSINT), thông tin miễn phí được thu thập theo những cách hợp pháp từ các nguồn công cộng như các trang tin tức, thư viện và công cụ tìm kiếm.

Cyber Threat Intelligence Lifecycle



- Direction: lập kế hoạch và định hướng, đề cập đến một câu hỏi hoặc 1 số câu hỏi cần trả lời, cần xác định càng cụ thể càng tốt.
- Collection: quá trình thu thập tất cả các dữ liệu cần thiết để trả lời câu hỏi được đặt ra. Collection được thu thập từ nhiều nguồn (open source, internal source, private source, commercial source, etc.).
    a. URLhauls: nguồn chia sẻ các URL độc hại đang được sử dụng để phân phối các phần mềm độc hại.
- Processing: gồm việc định hình dữ liệu thành một dạng có thể sử dụng (tiện dụng). bao gồm việc chuyển đổi định dạng dữ liệu, dịch ngôn ngữ, giải mã dữ liệu… đòi hỏi kỹ thuật cao và một số hệ thống mã nguồn mở hoặc thương mại => cần dùng tool
- Analysis: bước quan trọng nhất, chủ yếu là con người (các nhà phân tích). Đây là nơi các dữ liệu được được vào các context (bối cảnh) nhất định. Các nhà phân tích sẽ phân tích tất cả các dữ liệu được thu thập và xử lí trước đó => suy nghĩ để viết ra các sự kiện và suy nghĩ kỹ càng về mối đe dọa và các câu hỏi trước đó.
- Dissemination: các câu trả lời sẽ được cung cấp cho các bên liên quan thích hợp. chú ý: những nhóm người khác nhau sẽ cần kết quả ở định dạng khác nhau. Vd: CISO sẽ không xem dữ liệu thô mà chỉ nhận dưới dạng PDF
- Feedback: là bước các bên liên quan sẽ liên lạc lại với nhà phân tích.

Minh họa mối quan hệ giữa CTI và các yếu tố, thành phần bảo mật khác của hệ thống



- Operational threat intelligence
  - Là kiến thức về các cuộc tấn công mạng, events, các chiến dịch đang diễn ra. Nó cung cấp cho các nhóm ứng phó sự cố thông tin chi tiết chuyên biệt giúp họ hiểu bản chất, mục đích và thời gian cụ thể khi chúng đang diễn ra. Và thường source của nó là từ machine.
  - Đôi khi được gọi là technical threat intelligent. Vì nó thường bao gồm các thông tin kỹ thuật (techni) về các cuộc tấn công (vector tấn công nào đang được sử dụng, lỗ hổng nào đang bị khai thác và command nào hay control domain nào đang được sử dụng) => thường hữu ích cho những nhân viên trực tiếp tham gia vào việc bảo vệ một tổ chức (administrator, security staff)
  - Nguồn cung cấp: threat data feed (nguồn cung cấp dữ liệu về mối đe dọa), thường tập trung vào một loại cụ thể như: malware hash, suspicious domain. Data feed này như mình nói ở trên, cung cấp thông tin đầu vào nhưng bản thân nó không phải là threat intelligent.
  - Lợi ích: (loại này được sử dụng làm gì?) nhằm cải tiến hệ thống bảo mật hiện có, đồng thời tang tốc độ phản ứng với các sự cố của hệ thống vì nó sẽ trả lời đuoccác câu câu hỏi khẩn cấp dành riêng cho tổ chức của bạn - chẳng hạn như "Lỗ hổng nghiêm trọng này, đang được khai thác trong ngành của tôi, có trong hệ thống của tôi không?".
  -
- Strategic Threat Intelligence
  - cung cấp một cái nhìn tổng thể về toàn cảnh mối đe dọa của một tổ chức.

- đòi hỏi yếu tố con người vì cần thời gian và suy nghĩ để đánh giá và thử nghiệm các chiến thuật, kỹ thuật và quy trình mới của đối thủ chống lại các biện pháp kiểm soát an ninh hiện có
- Các phần của quá trình này có thể được tự động hóa, nhưng phần lớn cần phải có bộ não của con người
- cần cung cấp cái nhìn sâu sắc về những rủi ro liên quan đến các hành động nhất định, các mô hình rộng rãi trong các chiến thuật và mục tiêu của tác nhân đe dọa, các sự kiện và xu hướng địa chính trị cũng như các chủ đề tương tự.
- common source:
  - Tài liệu chính sách từ các quốc gia / tổ chức phi chính phủ
  - Tin tức từ các phương tiện truyền thông địa phương, quốc gia; các bài báo trong các ấn phẩm của ngành và chủ đề cụ thể / các chuyên gia
  - White paper, research paper

⇨ Các tổ chúc cần một đội ngủ phân tích cho chuyên môn cả trong lẫn ngoài ngành an ninh mạng – đặc biệt là sự hiểu biết sâu sắc về kinh doanh & chính trị xã hội

Limitations of CTI: Nhìn chung, CTI có thể xem như là một đại diện của các mối đe dọa mạng trên internet, nơi mà thông tin dựa trên giao tiếp thông qua TCP/IP quy mô lớn xảy ra trên web. Mặt khác, hầu hết các gói mạng được tạo ra trong cơ sở hạ tầng mạng IOT đều được truyền đến mạng nội bộ hoặc giao tiếp qua các giao thức chuyên dụng liên mạng. Trong trường hợp CTI dựa trên thông tin định danh là IP, nó cũng không ảnh hưởng đến CTI trong mạng nội bộ được hình thành bởi DHCP. Trên internet, các mối đe dọa mạng tồn tại là do các mã độc hại. Trong IOT, việc truyền các file hiếm khi xảy ra. Các mối đe dọa mạng có thể dẫn đến truyền dữ liệu bất thường hoặc các cuộc tấn công DDOS thông qua CTI giả mạo được tạo ra bởi các nguồn cung cấp dữ liệu được thu thập trên internet, không thể được áp dụng trực tiếp cho mạng IOT. Và tuyệt nhiên cũng khó đảm bảo được độ tin cậy của các chỉ số rủi ro của nguồn dữ liệu CTI được một bên thứ ba cung cấp. Ví dụ như với một tài nguyên nào đó trên mạng, nếu một số nguồn cung cấp dữ liệu CTI đưa ra kết luận là nó nguy hiểm, trong khi một số khác thì không và cho rằng nó vô hại. Điều này thể hiện tính thiếu nhất quán của các nguồn cung cấp dữ liệu. Lúc đó, rất khó để ta đưa ra quyết dịnh để quyết định dữ liệu nào đáng tin hơn. Vấn đề này luôn xảy ra vì dữ liệu của CTI không được chia sẻ hết cho nhau. Một số bên thứ ba cũng có cách nhận các report từ con người và xem đó như một cách thu thập nguồn dữ liệu. Điều này có thể bị khai thác để làm mất uy tín của người dùng vô tội nào đó (bằng một report giả), từ đó làm hỏng tính khả dụng của hệ thống và dịch vụ. Những điều trên cho thấy việc xác định độ tin cậy của dữ liệu đối với thông tin CTI là rất quan trọng.

3. Tính tin cậy của CTI feed và CTI data

Dữ liệu CTI nên được xác định độ tin cậy của chúng. Sử dụng các OSINT có thể gây ra các vấn đề về chất lượng của dữ liệu. Bằng cách gây nhầm lẫn cho người tiêu dùng CTI, việc quá tải cũng có thể làm giảm độ chính xác của CTI.

Mô hình đề xuất:

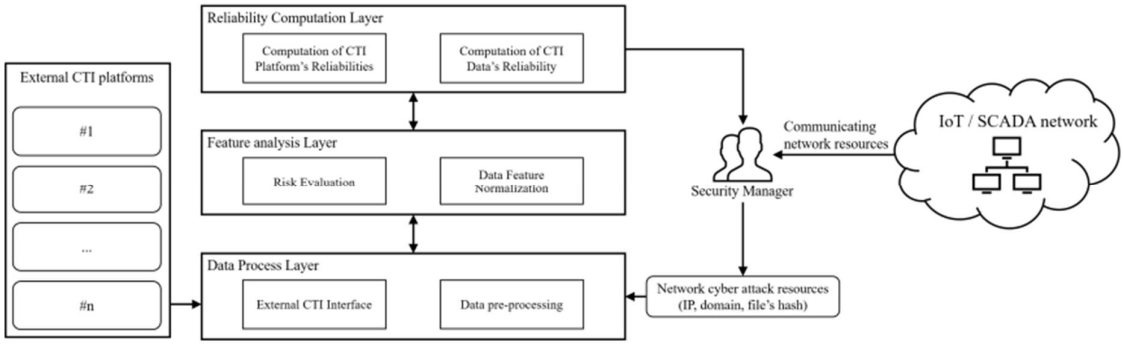


Fig. 1. Layer structure of proposal model

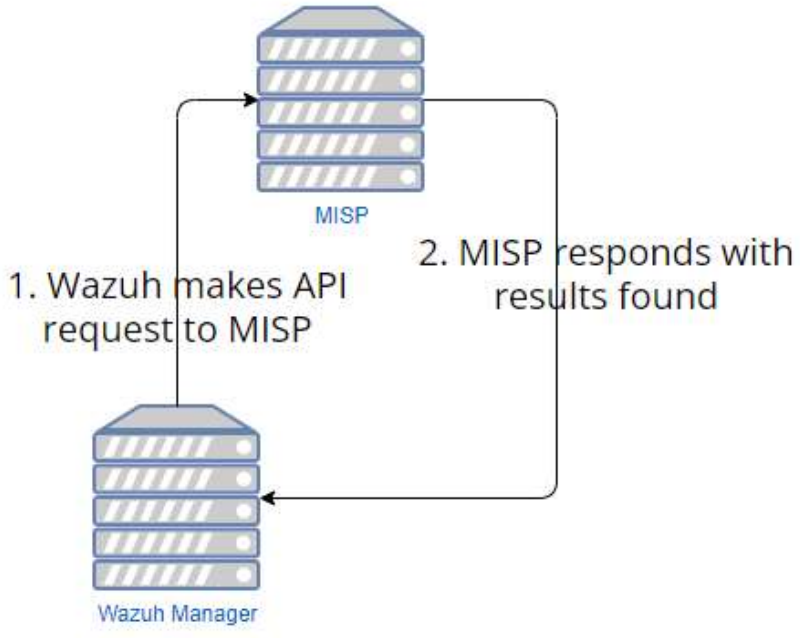
- Collecting CTI data from each CTI feeds (tạm dịch: thu thập dữ liệu CTI từ từng nguồn cung cấp CTI)
  - Mô hình đề xuất trong bài này sẽ thu thập dữ liệu CTI từ các nguồn dữ liệu CTI bằng cách sử dụng các tài nguyên mạng được cho là sử dụng để tấn công mạng: IP, domain theo thứ tự nhận biết, hàm băm từ báo cáo bảo mật và danh sách tệp vô hại.
- CTI data feature analysis and normalization (tạm dịch: phân tích và chuẩn hóa)
  - Mỗi nguồn dữ liệu CTI cung cấp dữ liệu dưới nhiều hình thức khác nhau. Nhóm tác giả chọn ra 10 đặc điểm đặc trưng của dữ liệu để làm đại diện và chỉ ra các mối đe dọa mạng. Từ thông tin báo cáo về các mối đe dọa và báo cáo bảo mật được quan sát, các đặc điểm này có thể giả định thông tin về kẻ tấn công, phạm vi của chúng, hoạt động, khuôn mẫu và danh tính như DNS, các sub-domain, anti-virus scan,… (miêu tả rõ trong chương 4)
- Getting reliabilities of CTI feeds using weight distances (tạm dịch: xác định độ tin cậy của nguồn CTI bằng cách xác định sự khác nhau giữa các dữ liệu trong nguồn CTI)
  - Với dữ liệu CTI đã được thu thập, tiếp theo mô hình này sẽ xác định tính hợp lệ của nguồn dữ liệu. Bằng cách so sánh giữa các dữ liệu CTI thu thập được từ các nguồn cung cấp dữ liệu, mô hình sẽ ước tính độ tin cậy của nguồn cung cấp dữ liệu. Độ tin cậy này trở thành cơ sở để xác định độ tin cậy của chính dữ liệu CTI (như đã nói độ tin cậy có hai phần: CTI feed & CTI data itself).
  - *weight distance ở đây hiểu là khoảng cách của trọng số, sự khác nhau của các dữ liệu thu thập được, khoảng cách càng lớn đồng nghĩa với sự khác nhau càng lớn.
- Getting reliability of CTI data based on the reliability of feed (tạm dịch: xác định độ tin cậy của dữ liệu CTI dựa trên độ tin cậy của nguồn cung cấp dữ liệu)
  - Dựa trên độ tin cậy của nguồn cung cấp dữ liệu ở trên, mô hình này sẽ tính toán độ tin cậy của dữ liệu. Gía trị sau khi tính được độ tin cậy này chứng tỏ được tính 'có thể tin cậy' của dữ liệu, đồng nghĩa với khả năng một mối đe dọa nào đó có thể xảy ra với tài nguyên hoặc dữ liệu đang quan sát.

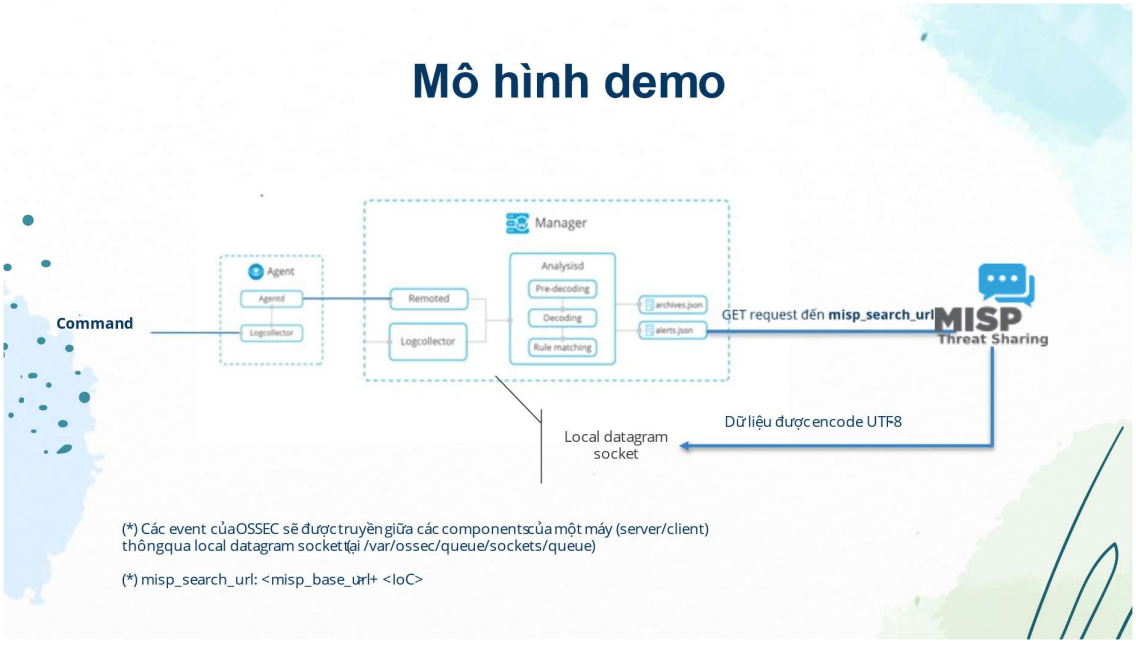
⇨ Trong mô hình này, nhóm tác giả sẽ cung cấp thông tin độ tin cậy của dữ liệu CTI về loại tài nguyên mạng cụ thể: IP, domain, file hash. Các tài nguyên này được xem như các chỉ số tiêu biểu, đặc trưng và cơ bản nhất của một cuộc tấn công mạng. Mô hình sẽ thu thập CTI từ các nguồn cung cấp dữ liệu bằng cách sử dụng các tài nguyên mạng này. Sau khi thu thập, dữ liệu CTI sẽ được chuẩn hóa vì mỗi dữ liệu CTI được cung cấp từ nguồn cung cấp có đặc tính và ngữ cảnh khác nhau. Trong quá trình này, để phản ánh sự khác biệt về format của dữ liệu, sự khác nhau (distance) giữa mối dữ liệu sẽ được gom lại. Gía trị trung bình của các distance này (distance giữa các data) trở thành distance giữa các nguồn cung cấp dữ liệu. Sau khi phân tính các đặc điểm, độ tin cậy của nguồn dữ liệu CTI có thể được tính toán dựa trên xác thực chéo của từng dữ liệu trong nguồn cung cấp. Độ tin cậy của nguồn cung cấp được cập nhật theo thời gian thực…

4. Phần Demo

Video demo tính hợp MISP và WAZUH

- Wazuh là một mã nguồn mở dùng việc bảo vệ an ninh.
- Được xây dựng từ các thành phần: OSSEC HIDS, OpenSCAP và Elastic Stack
  - Wazuh agent: Giám sát, thu thập dữ liệu hệ thống và ứng dụng. Dữ liệu được chuyển tới manager thông qua một kênh được mã hóa và xác thực (bước đăng ký).
  - Wazuh manager: Phụ trách việc phân tích dữ liệu nhận từ agent, tạo các ngưỡng cảnh báo khi 1 event ánh xạ với rule.
- MISP: Là một nền tảng mã nguồn mở cho phép thu thập, lưu trữ, chia sẻ các mối đe dọa mạng.
- Mô hình demo:





- Tích hợp MISP – Wazuh:
  - Cấu hình Sysmon rules cho agent
  - File ossec.conf được cấu hình chạy file custom-misp.py khi gặp các event Sysmon
  - Cấu hình file script custom-misp.py trong thư mục /var/ossec/integration
  - Thực hiện gửi GET request misp_base_url kèm theo IoC, xác thực bằng API key của MISP
  - Cấu hình custom rule để sinh ra aleart ở mức high security nếu MISP phản hồi.

https://youtu.be/oBhmUSS3Atk

Link drive:
https://drive.google.com/drive/folders/1fASxOEwBNf5CyDYemk_Z4vMkb-SQf5pq?usp=sharing

HẾT