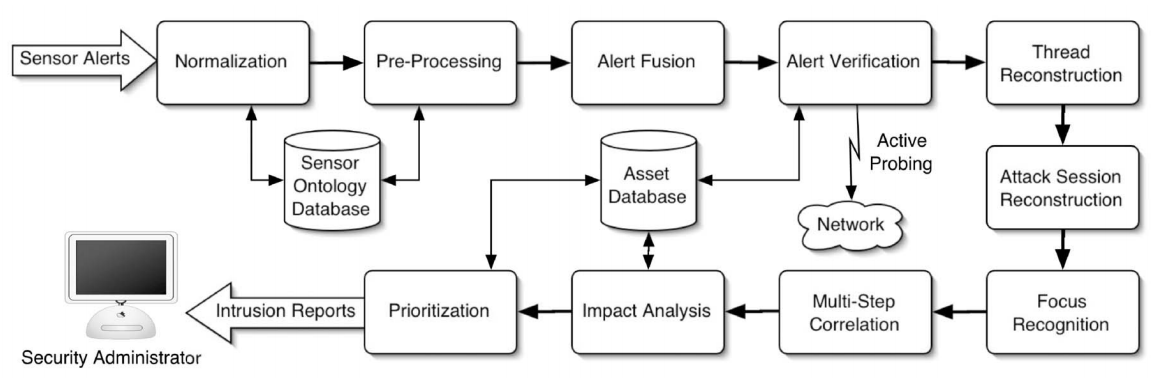
# A comprehensive approach to intrusion detection alert correlation



Hình vẽ thể hiện các bước trong mô hình correlation của tác giả. Tuy mô hình thể hiện ở dạng pipeline nhưng các module có thể thực hiện song song hoặc thứ tự có thể b5i thay đổi.

Mô hình thử nghiệm với nhiều loại data set khác nhau và có nêu ưu và nhược điểm của từng loại data st nên có thể dùng để tham khảo: darpa,defcon…

Nội dung của từng module:

1. Normalization
   * Chuyển tất cả các loại alert sang dạng IDMEF
   * Các alert name có thể lấy từ CVE làm chuẩn
   * Các alert attr được chuyển sang 1 dạng thống nhất
   * Dĩ nhiên nếu chỉ sử dụng 1 IDS thì có thể ko cần normalize
2. Preprocessing
   * Mục đích:cũng như preprocessing trong DM thì nó sẽ thêm vào các giá trị mà alert bị thiếu, cũng như chuẩn hóa giữa các alert với nhau
     1. Ví dụ: như lấy time nào trong IDMEF vì có 3 attr Time trong IDMEF; thêm src, dest của alert…
   * Tác giả cho rằng các sensor có thể thiếu 1 số attr và preprocessing ảnh hưởng rất lớn đến các module sau
3. Fusion
   * Mục đích: nếu có nhiều sensor thì alert fusion sẽ gọp các alert của cùng 1 event (tác giả ghi là attack occurrence) mà các sensor khác nhau tạo ra.
   * Cách thực hiện: hầu như các module của các giả đều dựa trên sliding window time và matching của các alert attr
     1. Các alert đc lưu trong 1 queue sắp xếp theo time. Các alert trong queue này nếu nằm trong 1 window. Window này sẽ đc di chuyển. Alert nào ra khỏi window sẽ đc gửi tới module kế. Alert nào ở trong window sẽ đc xử lí để gọp lại.
     2. Các alert đc merge lại bằng cách so trùng attr và tạo alert mới.
   * Tác giả cho rằng việc fusion chỉ có ý nghĩa khi có redundancy do việc sử dụng nhiều sensor(ví dụ sensor cho cả network và vùng DMZ)
4. Verification
   * Mục đích: giảm thiểu false positive bằng cách kiểm ra lại các alert với các thông tin khác như network topo, know vuln… Thông qua việc này sẽ phân biệt giữa các cuộc tấn công nào là success và fail.
   * Phương pháp: có 2 cách
     1. Passive:sử dụng thông tin biết trc về host hay network topo và các service
        1. Pros: ko ảnh hưởng tới network band,
        2. Cons: phải cập nhật KB thường xuyên, thông tin lưu trong KB có thể ko đủ để verify 1 attack(ví dụ như offset ko đúng trong lỗi buffer)
     2. Active: kiểm tra thông tin trên host về các track mà cuộc tấn cong để lại.
        1. Pros: thông tin này là mới và live
        2. Cons: ảnh hưởng network band, các alert tạo ra bởi active có thể bị nhầm, phải có local access và quyền trên host
5. Thread reconstruction
   * Mục đích: gọp các alert có cùng source và destination (tức 1 attacker tấn công 1 target)
   * Phương pháp: tương tự fusion nhưng gọp các alert có cùng source và dest. Alert có thể cùng 1 sensor.
6. Attack session reconstruction( có thể học tập)
   * Mục đích: kết hợp host based IDS và NIDS.
   * Phương pháp: tạo KB để map giữa port và service(process) chạy trên port đó. Khi đó 1 alert tạo ra chứa port co thể map với process chạy trên port đó.
     1. Vẫn sử dụng slide window và kết hợp timeout
7. Attack focus recognition:
   * Mục đích: xác định one2many và many2one attack
   * Phương pháp: slide window và timeout. Gôm các alert có cùng src cho one2many, tương tự cho many2one.
8. Multistep correlation
   * Mục đích: tạo 1 alert ở mức tổng quát hơn bằng cách nối các alert tạo ra bởi attk đó.
   * Tác giả cho rằng có 2 loại multistep attck: recon-breakin-escalate và island-hopping
   * Phương pháp: sử dụng STATL để mô hình. Slide window và timeout để hiện thực.