BÀI 14. ĐỘ TIN CẬY VÀ UY TÍN

TS. HOÀNG SỸ TƯƠNG

BÀI #14

- Đánh giá độ tin cậy trong các hệ thống nối mạng
- Hệ thống danh tiếng mạng

NIỀM TIN VÀ DANH TIẾNG LÀ GÌ?

- **Niềm tin (Trust):**
  - Kỳ vọng chủ quan của một tác nhân nhận được kết quả tích cực từ một tác nhân khác trong một bối cảnh cụ thể
- **Danh tiếng (Reputation):**
  - Nhận thức toàn cầu về độ tin cậy của một đại lý trong một hệ thống
- Tại sao chúng ta quan tâm đến những vấn đề này?

ĐIỀU ĐÓ NGHĨA LÀ GÌ?

- Các tuyên bố về độ tin cậy do các thiết bị/người dùng khác đưa ra về danh tính, dịch vụ, sự kiện, v.v.
- Tin tưởng người khác để quản lý chính xác dữ liệu và dịch vụ
- Tin tưởng người khác sẽ cư xử như mong đợi/đã hứa
- Tin tưởng người khác công bằng / không tham lam
- Và hơn thế...

TIN TƯỞNG VÀO INTERNET

- Internet sử dụng mô hình tin cậy tập trung hoặc phân cấp dựa trên việc định dang các chứng chỉ
  - Cơ quan cấp chứng chỉ chứng thực danh tính và độ tin cậy của các cá nhân/nhóm bằng cách cấp khóa công khai được ký/chứng nhận
    - CA tuyên bố "X được định danh và đáng tin cậy"
    - X cung cấp chứng chỉ đã ký từ CA đến Y
    - Tín thác chuyển tiếp: CA→X, X→Y ==> CA→Y
  - Loại mô hình này cũng cung cấp một khái niệm về trách nhiệm giải trình

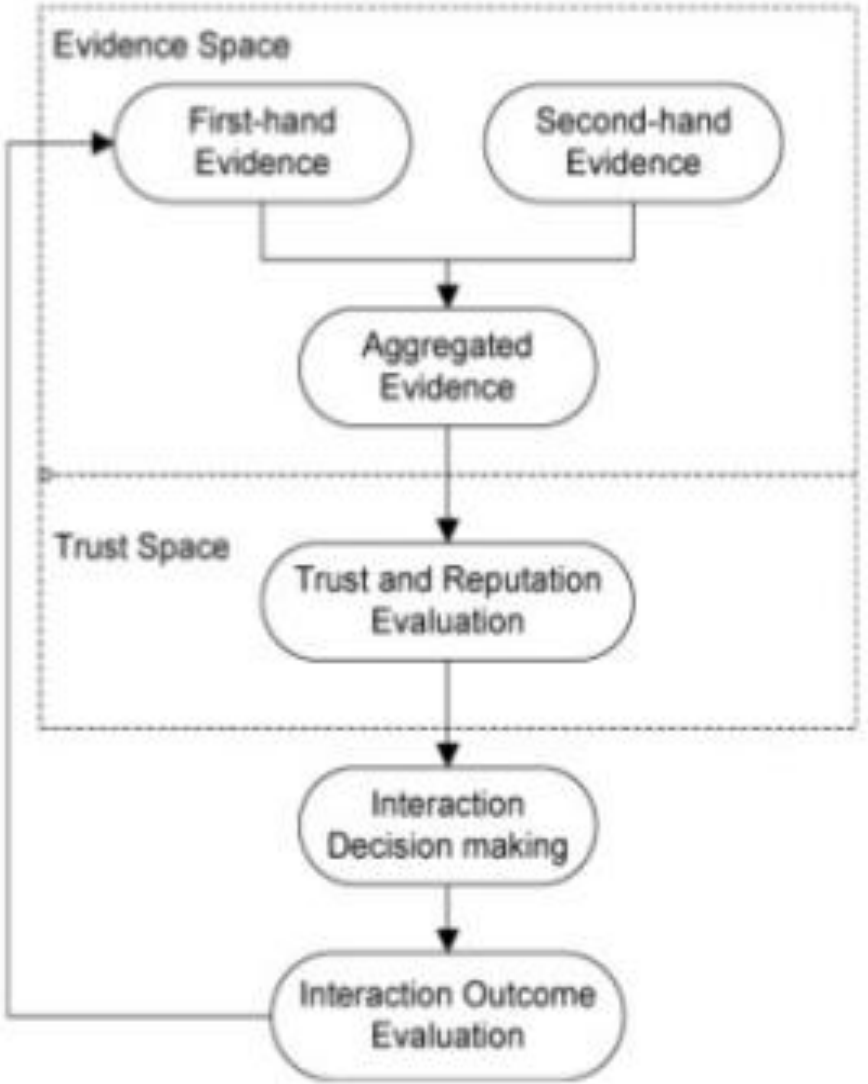
THỬ THÁCH ĐỘ TIN CẬY

- Trong MANET, thách thức lớn nhất là thiếu cơ quan quản lý tập trung nên không ai đóng vai trò là CA
  - *Làm thế nào để phân phối và xấp xỉ mô hình tin cậy CA?*
  - *Có mô hình nào khác hoạt động tốt/tốt hơn không?*
- Trong mạng hình lưới và WSN, độ trễ và đường dẫn đáng tin cậy là những thách thức lớn
  - *Làm cách nào để khởi động một đường dẫn an toàn/đáng tin cậy đến CA?*
- Trong DTN, độ trễ là một vấn đề lớn

CHI TIẾT VỀ CÁCH LẬP MÔ HÌNH VÀ ĐO LƯỜNG SỰ TIN CẬY

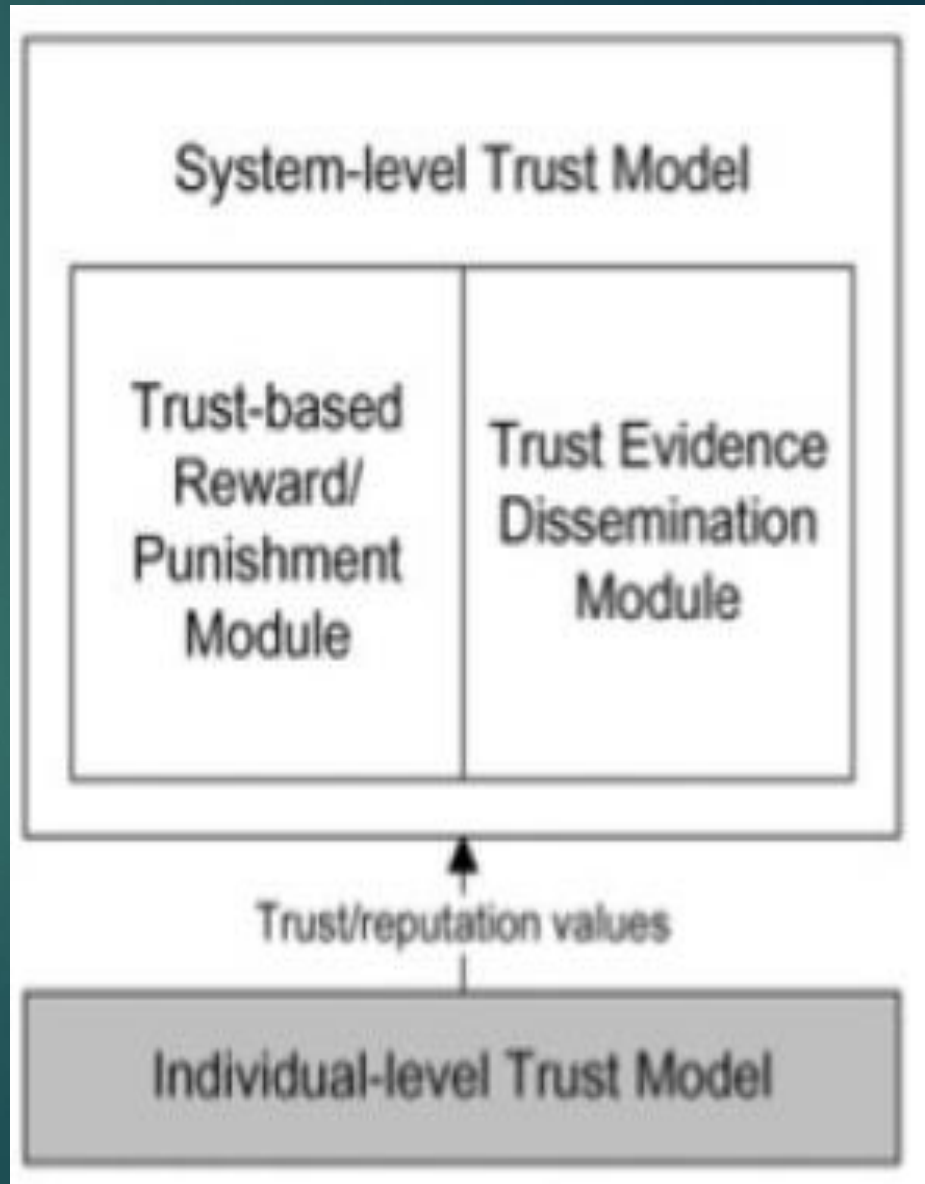
MÔ HÌNH ĐỘ TIN CẬY CÁ NHÂN

- Mỗi đại lý đánh giá sự tin tưởng của mình đối với một đại lý khác
  - Kết hợp quan sát trực tiếp và gián tiếp
  - Bao gồm hành vi trong quá khứ
- Niềm tin là một quan điểm
  - Nó có thể được thể hiện/chia sẻ, sửa đổi, thay đổi, v.v.



DANH TIẾNG, HAY CÒN ĐƯỢC GỌI LÀ ĐỘ TIN CẬY HỆ THỐNG

- Danh tiếng là một quan điểm được chia sẻ trên toàn cầu về sự tin tưởng vào một đại lý
  - Sắp xếp tổng hợp các giá trị tin cậy riêng lẻ
  - Cho phép hành động nhất quán với các đại lý không hợp tác



CÁC VẤN ĐỀ VỀ LÒNG TIN

- Làm thế nào để khởi tạo/bootstrap tin cậy?
  - *Ví dụ: Tôi đang đánh giá X, nhưng tôi chưa gặp họ bao giờ (và không ai trong số những người liên hệ của tôi đã gặp họ trước đây)*
- Làm thế nào để cân nhắc các sự kiện trong quá khứ và hiện tại?
  - *Ví dụ: X bất hợp tác 2 tuần trước nhưng từ đó trở nên tốt đẹp*
- Làm thế nào để cân nhắc các quan sát trực tiếp và gián tiếp?
  - *Ví dụ: X hợp tác với hàng xóm của tôi (được cho là) nhưng không hợp tác với tôi*

CÁC VẤN ĐỀ VỀ LÒNG TIN

- Làm cách nào để ánh xạ các sự kiện tới thước đo độ tin cậy?
  - *Ví dụ: X hợp tác 9 lần, từ chối 1 lần*
- Làm thế nào để nắm bắt được động sự biến động của niềm tin?
  - *Ví dụ: [X hợp tác 9x và từ chối 1x]*
- so với

  [X hợp tác 4x, từ chối 1x, hợp tác 5x]
- Làm thế nào để sử dụng các thước đo niềm tin sau khi được đánh giá?

LÒNG TIN – ĐỘ TIN CẬY

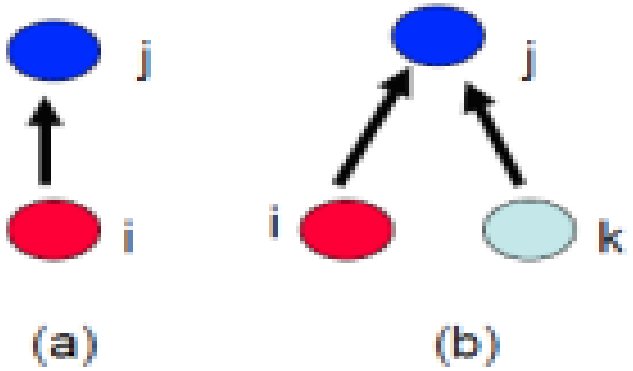
- Hầu hết các kỹ thuật đánh giá niềm tin đều sử dụng một số khái niệm chung
  - Niềm tin khó xây dựng nhưng dễ đánh mất
  - Tầm quan trọng của các sự kiện trong quá khứ giảm dần theo thời gian
  - Niềm tin phải mạnh mẽ đối với các sự kiện "tự nhiên"
    - *Ví dụ: có thể kết hợp sự không chắc chắn hoặc tự tin*
  - Bản thân cơ chế tin cậy phải mạnh mẽ đối với các hành vi sai trái

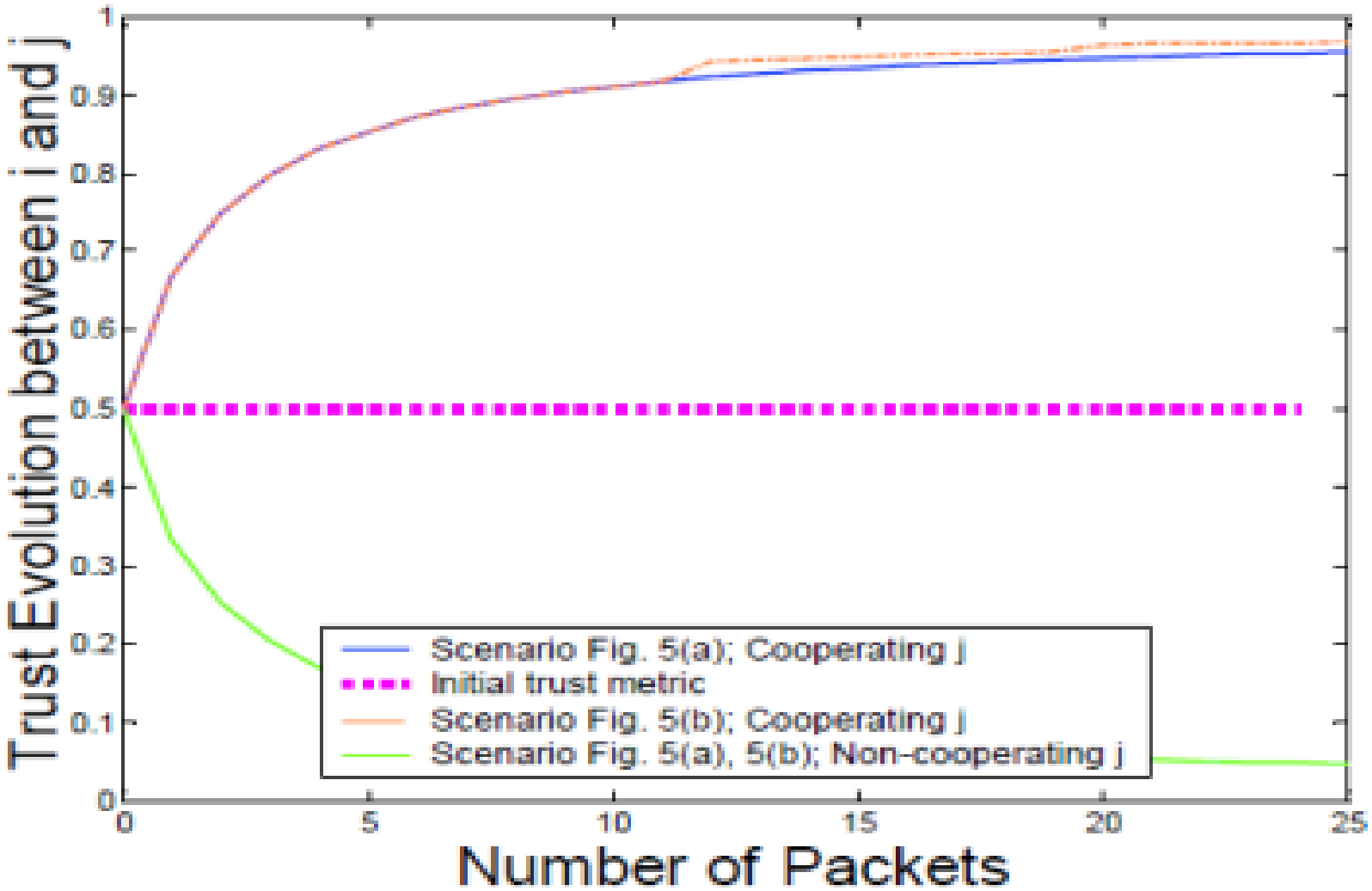
MA TRẬN ĐỘNG VỀ ĐỘ TIN CẬY

- Cách tiếp cận khác nhau sử dụng các chính sách đánh giá khác nhau, chẳng hạn như:
  - Đối với mỗi hành động tích cực/tiêu cực, cộng/trừ một hằng số vào/từ giá trị tin cậy
  - Đối với mỗi hành động tích cực, thêm một hằng số; đối với mỗi hành động tiêu cực, nhân với một phần không đổi
  - Đối với mỗi hành động tích cực, thêm một hằng số; đối với mỗi hành động tiêu cực, hãy giảm ranh giới xuống dưới (0 hoặc -1)

VÍ DỤ

- Từ [Ganeriwal & Srivastava, 2004]





CÒN CÁC CUỘC TẤN CÔNG VÀO CHÍNH HỆ THỐNG NIỀM TIN/DANH TIẾNG THÌ SAO?

TẤN CÔNG LÒNG TIN/DANH TIẾNG

- Mô hình tấn công:
  - Kẻ tấn công là người bên trong hệ thống, có thể hợp tác/tuân thủ hoặc lựa chọn hành vi sai trái
  - Được thúc đẩy bởi mục đích ích kỷ/không công bằng hoặc ác ý
  - Có thể làm việc một mình hoặc thông đồng với người khác
- Nói chung, mục đích tấn công nhằm đạt được một trong ba mục tiêu:
  - Tăng giá trị niềm tin giả(bản thân hoặc bạn bè)
  - Làm giảm giá trị niềm tin (mục tiêu tấn công)
  - Từ chối dịch vụ

TỰ THÚC ĐẨY TẤN CÔNG

- **Mục tiêu:** đạt được sự tin tưởng cao hơn giữa các láng giềng và/hoặc danh tiếng trong hệ thống
- **Phương tiện:** tạo phản hồi tích cực hoặc sửa đổi các giá trị danh tiếng trong quá trình truyền, có thể gây thiệt hại cho người khác
- **Giả định:** (i) hệ thống danh tiếng dựa trên phản hồi tích cực, (ii) cơ chế này có thể khai thác được

TẤN CÔNG TẨY TRẮNG - WHITEWASHING

- **Mục tiêu:** nhanh chóng sửa chữa giá trị niềm tin/danh tiếng sau khi hành động ích kỷ/ác ý được thực hiện
- **Phương tiện:** sau khi lạm dụng, nhập lại hoặc khai thác hệ thống để đặt lại các giá trị tin cậy về trạng thái mặc định hoặc trước đó, có thể kết hợp với các cuộc tấn công khác
- **Giả định:** (i) hệ thống danh tiếng có thể cần phản các phản hồi tiêu cực, (ii) cơ chế này có thể khai thác được

TẤN CÔNG VU KHỐNG - SLANDERING ATTACK

- **Mục tiêu:** làm giảm giá trị niềm tin/danh tiếng của (những) tác nhân khác một cách sai trái
- **Phương tiện:** như tên cho thấy, lan truyền ý kiến sai lệch của người khác, thường thông qua phản hồi tiêu cực có thể rất tai hại
- **Các giả định:** (i) hệ thống danh tiếng cần phản hồi tiêu cực, (ii) cơ chế có thể khai thác được, (iii) có thể yêu cầu thông đồng

TẤN CÔNG PHỐI HỢP - ORCHESTRATED ATTACKS

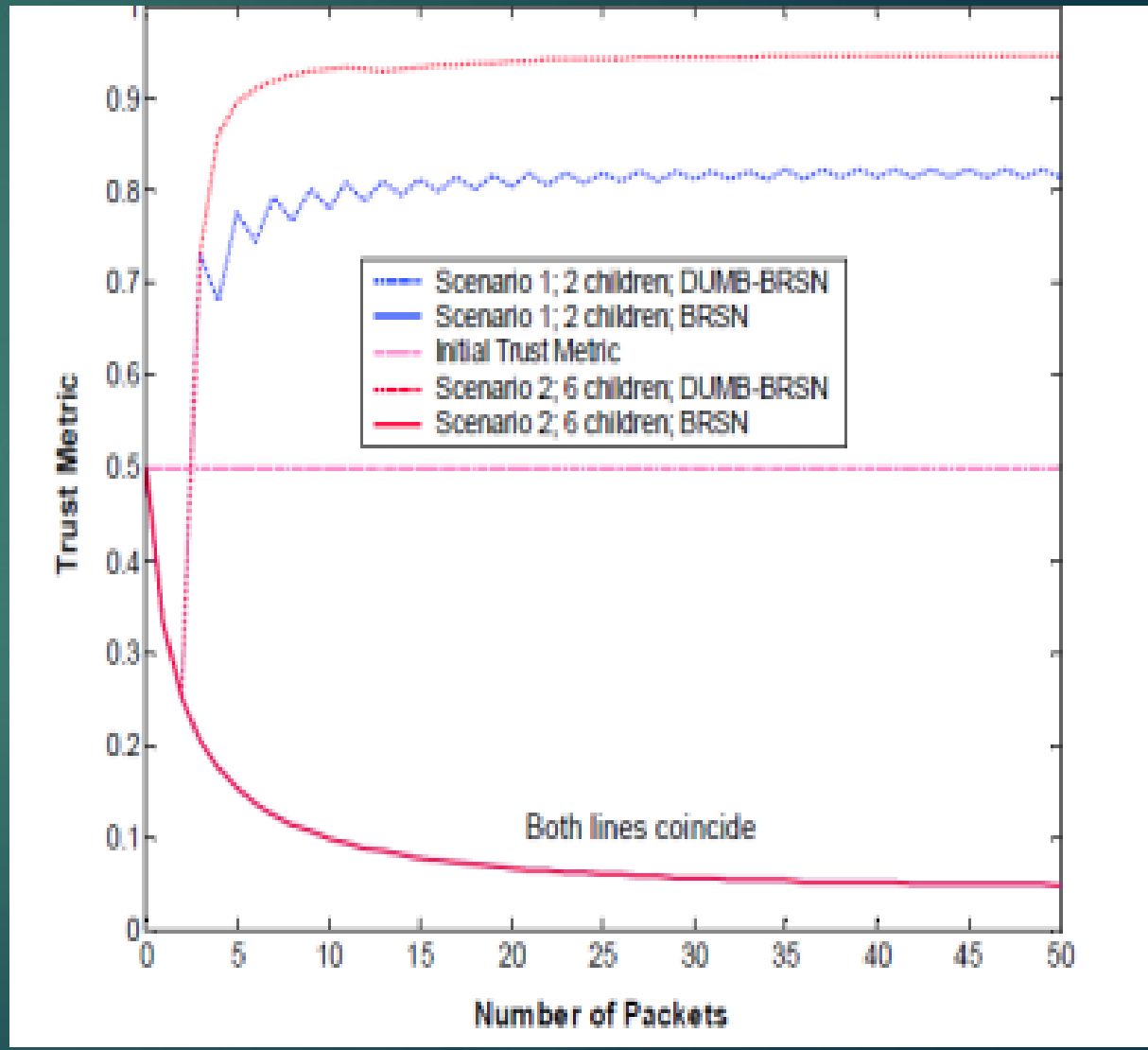
- **Mục tiêu:** nhiều kẻ tấn công thông đồng để ép hệ thống vào một trạng thái mong muốn cụ thể
- **Phương tiện:** kết hợp quảng cáo, tẩy trắng và vu khống khi cần thiết cho một mục tiêu cụ thể
  - *Ví dụ: tấn công dao động - chia đội, 1/2 vu khống 1/2 thăng cấp, thỉnh thoảng chuyển đổi*
- **Các giả định:** (i) thông đồng, (ii) bất kỳ giả định nào cần thiết cho các cuộc tấn công thành phần

TẤN CÔNG DOS

- **Mục tiêu:** ngăn chặn việc tính toán và phổ biến các giá trị tin cậy/danh tiếng, từ chối mọi giao thức/ứng dụng được hỗ trợ
- **Phương tiện:** làm quá tải hệ thống hoặc chặn tin nhắn theo một cách nào đó (thường thông qua một số hình thức tấn công DoS hiện có)
  - ***Ví dụ:*** *cập nhật ngập lụt danh tiếng để không ai có thể xử lý tất cả các vấn đề; kẹt/thả tin nhắn cập nhật danh tiếng*
- **Giả định:** (i) đủ tài nguyên, (ii) thông đồng như trong DDoS

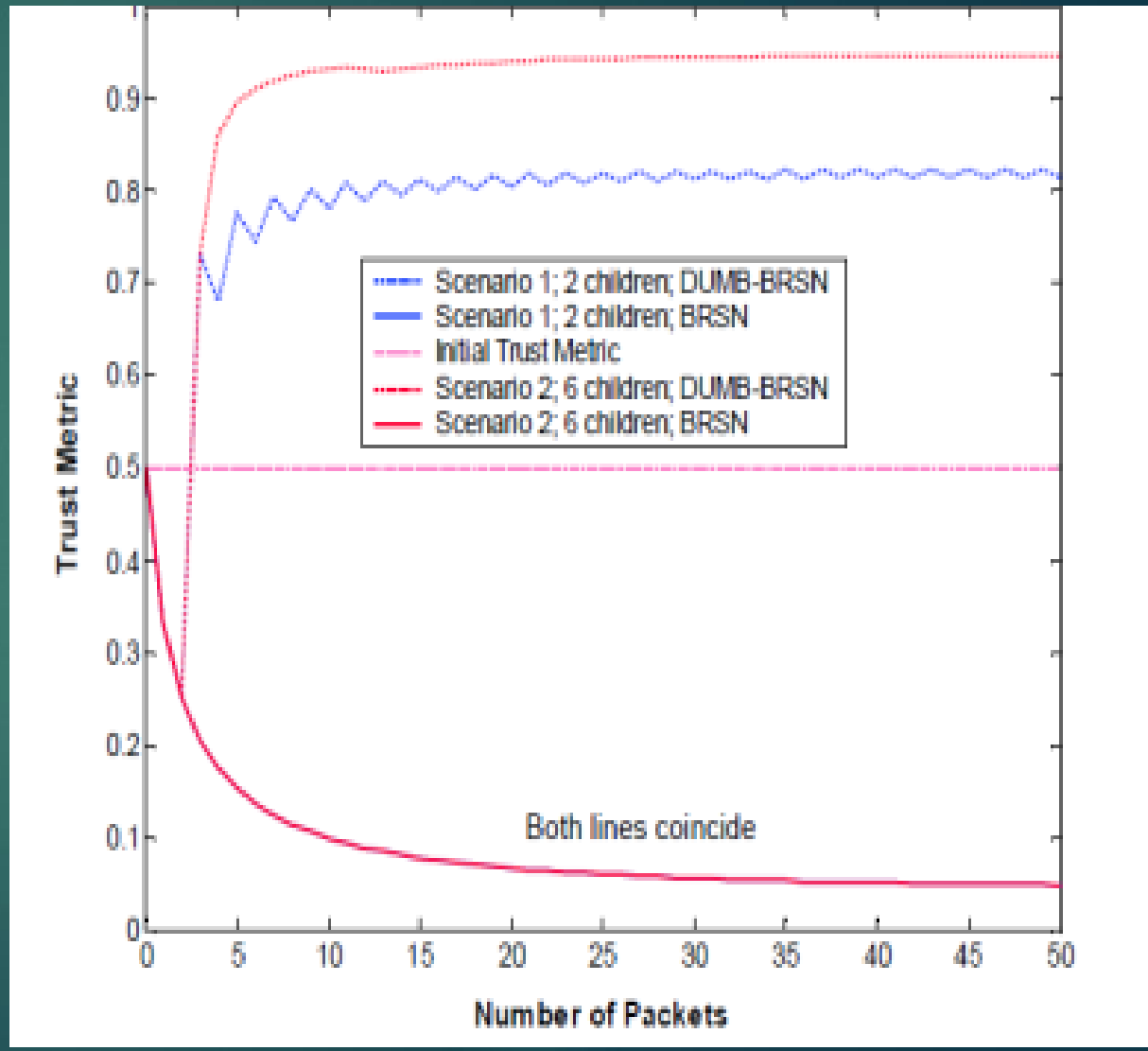
TẤN CÔNG VÀO DANH TIẾNG

- Danh tiếng tiêu cực không công bằng hoặc mã độc
  - "Tấn công sử dụng Những lời nói xấu mang tính công kích"
  - Khả năng phòng thủ: ứng lượng các phản hồi tiêu cực



TẤN CÔNG VÀO DANH TIẾNG

- Danh tiếng tích cực ác ý không công bằng hoặc mã độc
  - "Các cuộc tấn công nhồi nhét lá phiếu"
  - Khả năng phòng thủ: phát hiện sau đó thiên vị kết quả trong tương lai



LÀM THẾ NÀO ĐỂ CHỐNG LẠI TẤN CÔNG VÀO NIỀM TIN/DANH TIẾNG?

PHÒNG THỦ TRƯỚC CÁC CUỘC TẤN CÔNG CỦA SYBIL

- **Vấn đề:** nhiều cuộc tấn công ở trên là do hành vi giống như Sybil (nhiều định danh trên mỗi nút)
  - Cho phép kẻ tấn công trình bày nhiều ý kiến
- Khả năng phòng thủ:
  - Quản lý danh tính tập trung hoặc phân tán, có khả năng liên kết ID với thiết bị, địa chỉ hoặc thông số tĩnh khác
  - ID cũng có thể dựa trên "web tin cậy" xã hội

GIẢM THIỂU TIN ĐỒN THẤT THIỆT

- **Vấn đề:** kẻ tấn công có thể bịa đặt tin đồn thất thiệt để thay đổi tính toán danh tiếng
- Phòng thủ # 1 (giảm nhẹ):
  - *Ràng buộc các báo cáo bằng cách sử dụng bảo vệ bằng mật mã như chữ ký số (đối với trách nhiệm giải trình)*
- Phòng thủ #2 (giảm thiểu lây lan):
  - *Lọc ra các báo cáo không khớp với các báo cáo khác bằng cách sử dụng biểu quyết hoặc tính nhất quán với các quan sát trực tiếp*
  - *Không chuyển tiếp bất kỳ báo cáo nào không nhất quán*

GIẢM NHẸ LẠM DỤNG NGẮN HẠN

- **Vấn đề:** kẻ tấn công có thể xử lý sai trong một thời gian tương đối ngắn và sau đó chơi đẹp (hoặc đặt lại bằng ID mới) để khôi phục danh tiếng
- Khả năng phòng vệ:
  - Các diễn viên mới bắt đầu với danh tiếng thấp và cần xây dựng trước khi nhận dịch vụ
  - Thực hiện các hình phạt nghiêm khắc đối với hành vi sai trái với tốc độ cải thiện chậm

GIẢM THIỂU CÁC CUỘC TẤN CÔNG DOS

- **Vấn đề:** Các cuộc tấn công DoS nhằm phổ biến và cập nhật độ tin cậy có thể ngăn cản việc xây dựng danh tiếng
- Khả năng phòng vệ:
  - Phân phối nhiệm vụ phổ biến/cập nhật trên nhiều tác nhân để tạo sự đa dạng
  - Sử dụng các chiến lược giảm thiểu DoS và độ tin cậy mạng chung
    - ACK/NACK, định tuyến đa đường, cơ chế tin đồn, mã sửa lỗi, v.v.

HỆ THỐNG NIỀM TIN VÀ DANH TIẾNG HOẠT ĐỘNG NHƯ THẾ NÀO?

MẠNG DỰA TRÊN NIỀM TIN

- Các nút mạng có thể được chọn lọc về các quyết định liên lạc và kết nối mạng bằng cách sử dụng các chính sách dựa trên niềm tin
  - Một nút có thể quyết định chỉ giao tiếp cục bộ với các nút mà nó tin cậy trên ngưỡng τ
  - Một nút có thể xây dựng/chọn đường dẫn định tuyến bằng cách sử dụng chỉ số độ tin cậy/danh tiếng của đường dẫn tổng hợp

WATCHDOG & PATHRATER

- Giám sát watchdog chuyển tiếp bằng cách nghe lén các lần truyền tiếp theo

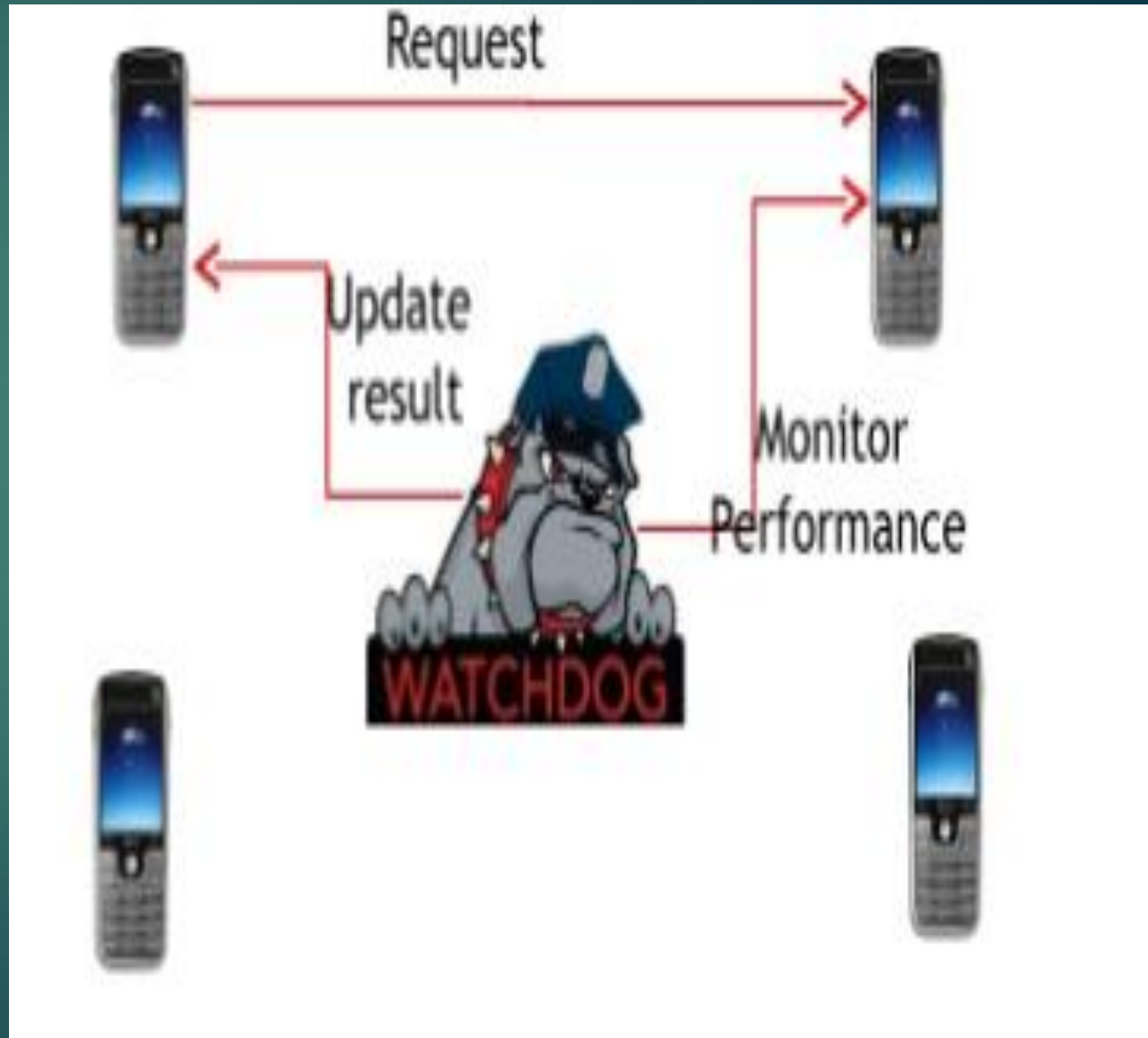
- Nếu $A \rightarrow B$ and $B \rightarrow C$, thì A có thể nghe và phân tích hành vi chuyển tiếp của B
- Pathrater sử dụng số liệu thống kê được quan sát để chọn đường dẫn nào đáng tin cậy nhất
  - Có thể hữu ích trong việc lựa chọn tuyến đường cho định tuyến nguồn, ví dụ: DSR

CÁC VẤN ĐỀ VỚI WATCHDOG

- Không có gì đáng ngạc nhiên, watchdog+pathrate không cải thiện độ tin cậy đối với hành vi sai trái hoặc lỗi mạng
  - Va chạm, làm mờ, liên kết không đối xứng và nhiều sự kiện khác được coi là hành vi sai trái
- Không cải thiện độ tin cậy trước nhiều kiểu tấn công
  - Các cuộc tấn công vu khống/đóng khung được thực hiện bởi watchdog ảnh hưởng đến tỷ lệ đường dẫn tổng hợp

LÕI – CORE

- CORE kết hợp các giá trị tin cậy trực tiếp và gián tiếp để tạo ra một tập hợp các chức năng khác nhau
  - **Ví dụ:** chuyển tiếp, khám phá tuyến đường, quản lý mạng, quản lý vị trí, v.v.
  - Xây dựng trên cơ chế Watchdog sử dụng mô hình người yêu cầu-nhà cung cấp



YÊU CẦU CORE

- Khi người yêu cầu đưa ra yêu cầu dịch vụ:
  - watchdog giám sát yêu cầu và trả lời
  - Nhà cung cấp chỉ chấp nhận yêu cầu nếu giá trị uy tín của người yêu cầu đủ cao
  - watchdog có thể cập nhật nhà cung cấp giá trị danh tiếng nếu nó thay đổi, ví dụ: nếu người yêu cầu là nhà cung cấp DoS-ing
- CORE ngăn chặn một số cuộc tấn công bằng cách chỉ cho phép lan truyền các báo cáo tích cực; báo cáo tiêu cực chỉ đi 1 hop

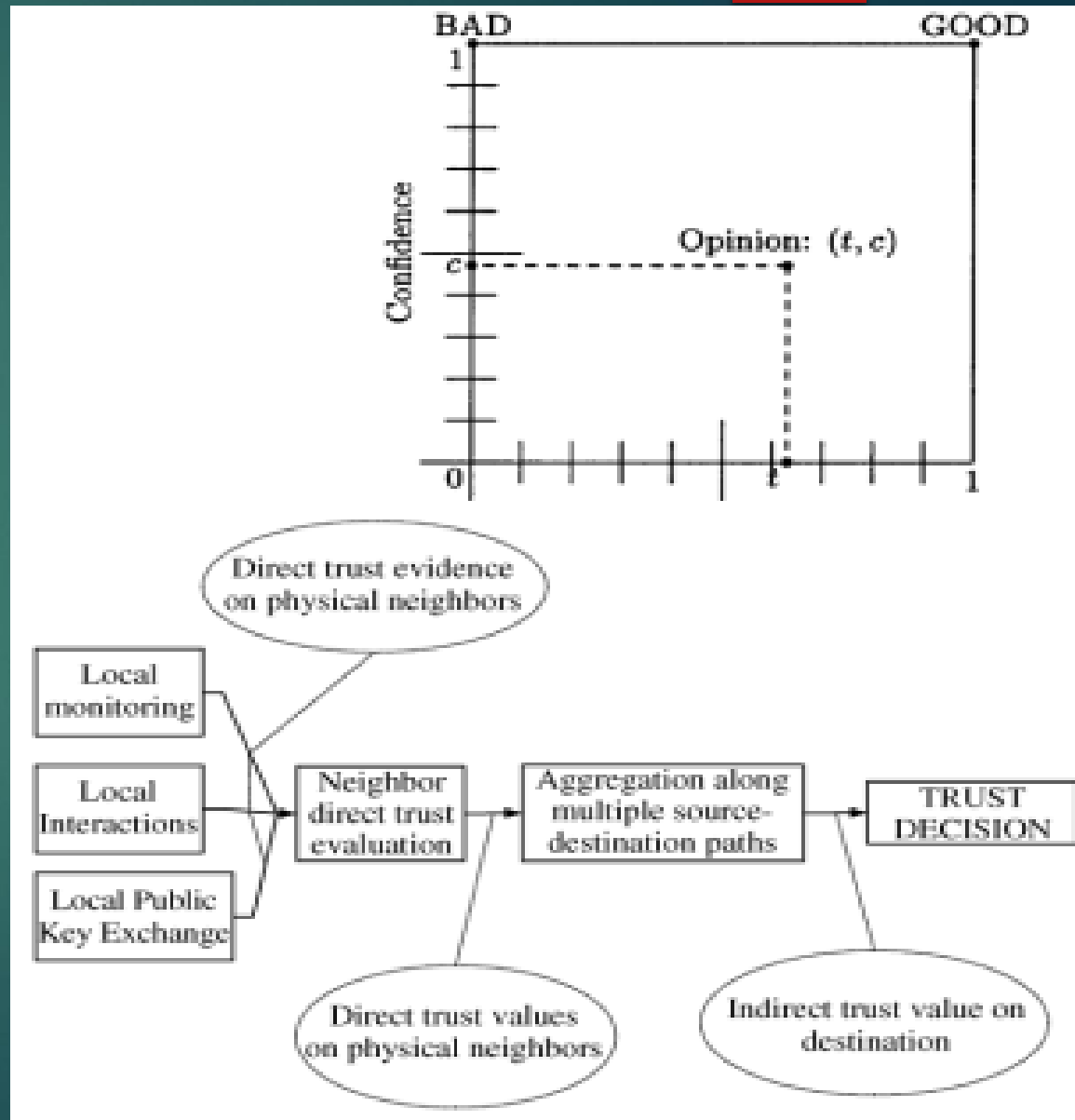
HẠN CHẾ CORE

- CORE chấp nhận tất cả các giới hạn của watchdog
- Không thể mở rộng vì cần có watchdog ở mọi khu vực lân cận

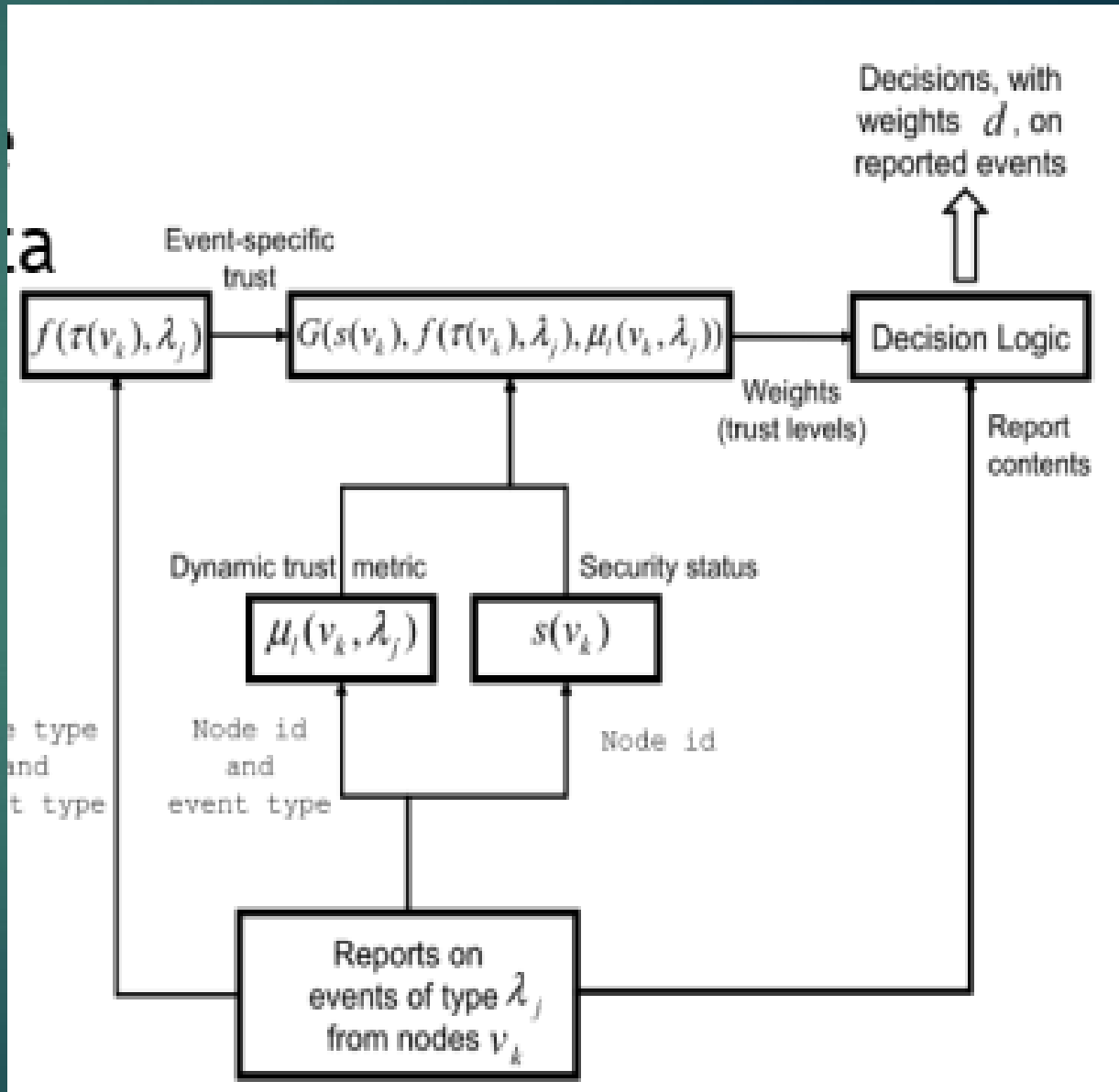
Ngoài ra, watchdog cần thông tin toàn cầu (hoặc ít nhất là E2E)

- Tính cơ động có thể phá vỡ CORE

- Bộ định tuyến báo cáo độ tin cậy của mỗi bước nhảy tiếp theo, bao gồm cả độ tin cậy được ước lượng.
- Nguồn tính toán tổng hợp trên từng đường dẫn để thông báo lựa chọn đường dẫn

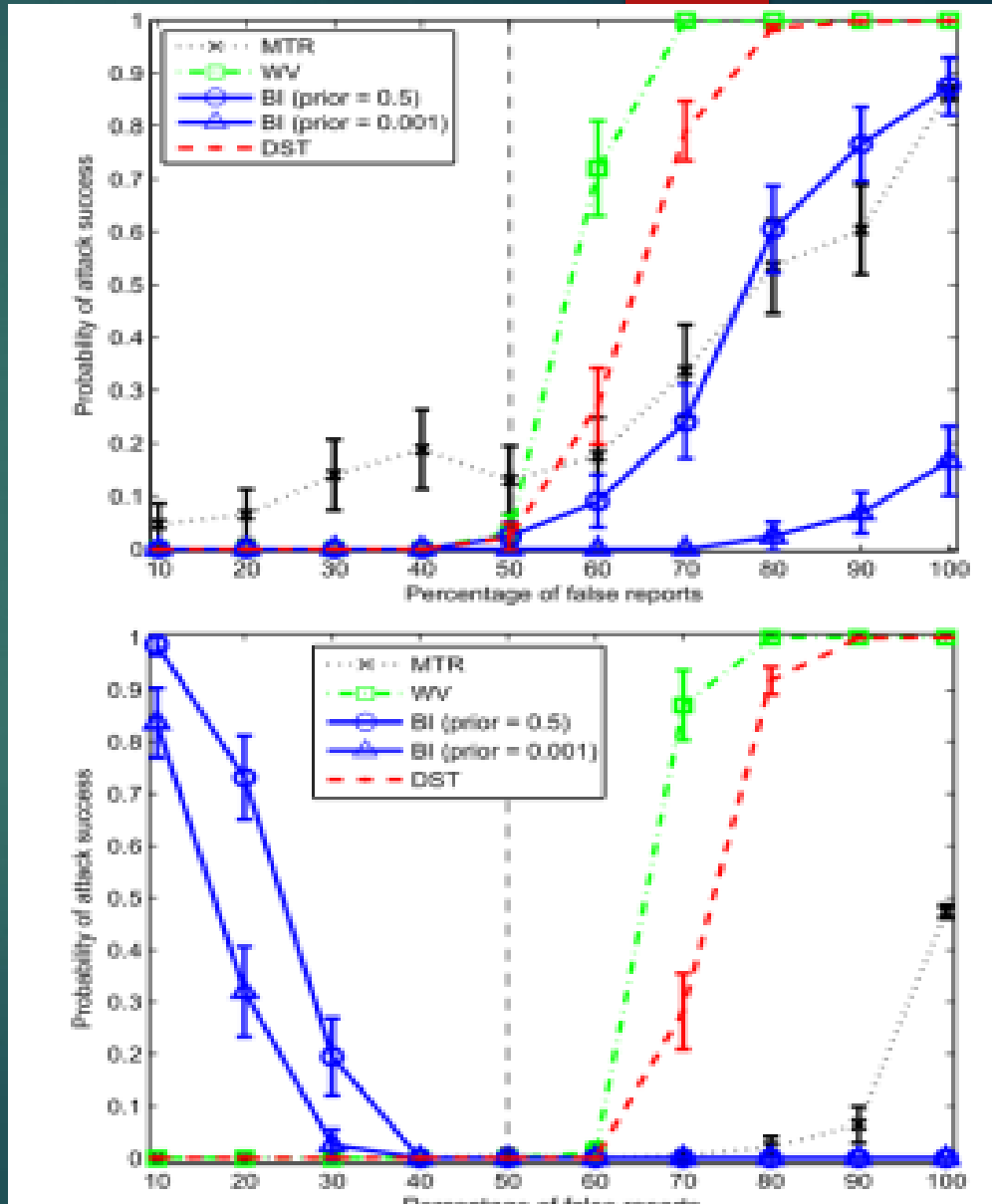


- Trong các môi trường động như VANET, độ tin cậy có thể được tính toán trên dữ liệu thay tác nhân
  - Bao gồm các yếu tố động như địa điểm và thời gian
  - Tất cả các yếu tố liên quan có thể được tính trọng số riêng lẻ để đưa ra giá trị tin cậy trong từng phần dữ liệu



ĐÁNH GIÁ ĐỘ TIN CẬY DỮ LIỆU

- Bài báo cung cấp một khung để đánh giá mức độ tin cậy bằng cách sử dụng một số phương pháp thống kê nhau
  - Các kỹ thuật khác nhau cung cấp khả năng phục hồi chống lại các một số dạng tấn công.



CÂU HỎI MỞ

- Làm cách nào để chọn đúng loại danh tiếng động cho một hệ thống/nhiệm vụ/dữ liệu nhất định?
- Làm thế nào để phát hiện các sự kiện làm tăng và giảm danh tiếng?
- Làm thế nào để giảm thiểu ảnh hưởng của lỗi được phát hiện?
- Danh tiếng có hiệu quả trước các cuộc tấn công nổi tiếng (vu khống, dối trá, v.v.) không?

BÀI 14:

AN NINH TẦNG VẬN TẢI KHÔNG DÂY;