BÀI 12.
ẨN DANH VÀ QUYỀN RIÊNG TƯ

Bùi Trọng Tùng,
Viện Công nghệ thông tin và Truyền thông,
Đại học Bách khoa Hà Nội

Nội dung

- Quyền riêng tư
- Truyền tin ẩn danh
- Mạng ẩn danh Tor

1. QUYỀN RIÊNG TƯ

Bùi Trọng Tùng,
Viện Công nghệ thông tin và Truyền thông,
Đại học Bách khoa Hà Nội

Khái niệm

- Quyền riêng tư là quyền mỗi người có thể tự quyết và kiểm soát việc sử dụng các thông tin cá nhân của chính họ.
  - Thông tin liên lạc
  - Thông tin nhân thân
  - Thông tin sức khỏe
  - Các hoạt động thường ngày
  - ...
- Đọc thêm về nguyên tắc bảo vệ dữ liệu trong “Tuyên bố Montreux”

Luật An ninh mạng - 2015

1. Cá nhân tự bảo vệ thông tin cá nhân của mình và tuân thủ quy định của pháp luật về cung cấp thông tin cá nhân khi sử dụng dịch vụ trên mạng.
2. Cơ quan, tổ chức, cá nhân xử lý thông tin cá nhân có trách nhiệm bảo đảm an toàn thông tin mạng đối với thông tin do mình xử lý.
3. Tổ chức, cá nhân xử lý thông tin cá nhân phải xây dựng và công bố công khai biện pháp xử lý, bảo vệ thông tin cá nhân của tổ chức, cá nhân mình.
4. Việc bảo vệ thông tin cá nhân thực hiện theo quy định của Luật này và quy định khác của pháp luật có liên quan.
5. Việc xử lý thông tin cá nhân phục vụ mục đích bảo đảm quốc phòng, an ninh quốc gia, trật tự, an toàn xã hội hoặc không nhằm mục đích thương mại được thực hiện theo quy định khác của pháp luật có liên quan.

5

5

Luật bảo vệ dữ liệu cá nhân của EU

- Được xử lý một cách hợp pháp, công bằng và minh bạch
- Được thu thập cho các mục đích cụ thể, rõ ràng và hợp pháp và không được xử lý thêm theo cách không tương thích với các mục đích đó
- Chỉ sử dụng phù hợp và trong giới hạn cần thiết liên quan đến các mục đích mà chúng được xử lý
- Được lưu trữ chính xác và, khi cần thiết, được cập nhật; mọi bước hợp lý phải được thực hiện để đảm bảo rằng dữ liệu cá nhân không chính xác, liên quan đến các mục đích mà chúng được xử lý, được xóa hoặc khắc phục không chậm trễ

6

6

Luật bảo vệ dữ liệu cá nhân của EU

- Khi lưu trữ dưới dạng có thể xác định được danh tình chủ thể thì thời gian chỉ kéo dài ở giới hạn cần thiết cho mục đích xử lý
- Được xử lý một cách bí mật và toàn vẹn
- Các cá nhân, tổ chức có trách nhiệm giải trình về các hoạt động sử dụng dữ liệu cá nhân của người khác.

7

7

Google Privacy & Terms

We collect information about the apps, browsers and devices that you use to access Google services, which helps us provide features such as automatic product updates and dimming your screen if your battery runs low.

The information that we collect includes unique identifiers, browser type and settings, device type and settings, operating system, mobile network information including operator name and phone number and application version number. We also collect information about the interaction of your apps, browsers and devices with our services, including IP address, crash reports, system activity, and the date, time and referrer URL of your request.

8

8

Apple vs FBI

The New York Times

F.B.I. Asks Apple to Help Unlock Two iPhones

The request could reignite a fight bet
gian
tech

The New York Times

U.S. Says It Has Unlocked iPhone Without Apple

BBC Sign in Nev

NEWS

Home | Video | World | Asia | UK | Business | Tech | Science | Stories | Entertainment

Technology

Apple rejects order to unlock gunman's phone

9

9

Vi phạm của Cambridge Analytica

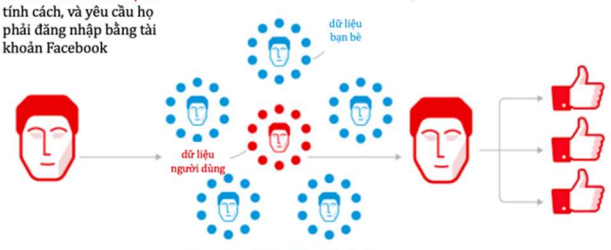
Cambridge Analytica làm thế nào đánh cắp được dữ liệu 50 triệu người dùng Facebook

1
Tầm 320,000 người bầu cử Mỹ ('người gieo hạt') được trả $2-$5 để làm bài kiểm tra chính trị về tính cách, và yêu cầu họ phải đăng nhập bằng tài khoản Facebook

2
Ứng dụng cũng sẽ thu thập dữ liệu như các nút like, thông tin cá nhân từ tài khoản Facebook của người làm bài kiểm tra...

3
Kết quả bài kiểm tra nhân cách được ghép cặp với dữ liệu Facebook của họ, chẳng hạn các nút like nhằm tìm ra các mẫu tâm lý

4
Thuật toán kết hợp dữ liệu với các nguồn khác như hồ sơ người bỏ phiếu nhằm tạo ra một bồ hồ sơ tối ưu (hồ sơ ban đầu gồm 2 triệu người trong 11 bang chính*) với hàng trăm điểm dữ liệu cho một người



... cũng như dữ liệu bạn bè của họ, thu được hơn 50 triệu dữ liệu người dùng Facebook

Những cá nhân này sau đó sẽ thành mục tiêu của các quảng cáo đã được cá nhân hoá dựa trên dữ liệu tính cách của họ

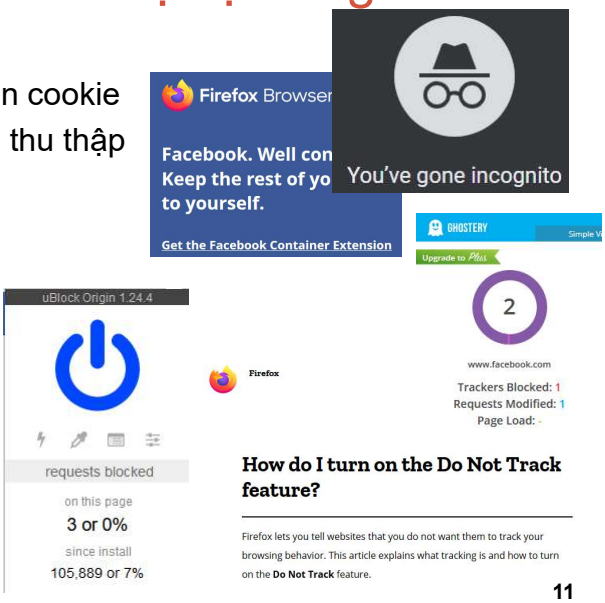
Nguồn: Guardian graphic. *Arkansas, Colorado, Florida, Iowa, Louisiana, Nevada, New Hampshire, North Carolina, Oregon, South Carolina, West Virginia

10

10

Các công cụ bảo vệ sự riêng tư

- Tiện ích ngăn chặn cookie
- Tiện ích cảnh báo thu thập
- Mã hóa
- Tiện ích dọn dẹp
- Ẩn danh
- Công cụ vật lý



11

11

CASE STUDY: BẢO VỆ THÔNG TIN CÁ NHÂN TRONG CSDL

Bùi Trọng Tùng,
Viện Công nghệ thông tin và Truyền thông,
Đại học Bách khoa Hà Nội

12

12

Bảo vệ thông tin cá nhân trong CSDL

- CSDL có thể chứa thông tin riêng tư:
  - CSDL y tế
  - CSDL giáo dục
  - CSDL nhân viên
  - …
- Một người dùng/nhóm có quyền truy cập một số loại dữ liệu nào đó, nhưng không phải tất cả, ví dụ
  - Có quyền thực hiện truy vấn tính toán thống kê nhưng không được truy vấn tới từng giá trị cụ thể
  - Thực hiện truy vấn để khai phá dữ liệu (data mining)

13

13

Vấn đề cần giải quyết

- Có thể suy luận thông tin không có quyền truy cập:
  - Kết hợp các thông tin được phép truy cập
  - Kết hợp với các kiến thức khác
- Mô hình giải quyết:
  - Truy vấn phi tương tác
    - ✓ Dữ liệu đã được xử lý trước, giấu danh tính cá nhân, chuẩn hóa
    - ✓ Người dùng được truy vấn tới mọi dữ liệu
  - Truy vấn tương tác: kết quả truy vấn được lọc

14

14

Ví dụ

- Không được phép biết lương của nhân viên cụ thể

ID	Name	Years of service	Salary
001	Alice	12	65,000$
003			000$
004	David	30	80,000$

SELECT … ne = 'Alice'

Request denied!

15

15

Ví dụ

ID	Name	Years of service	Salary
004	David	30	80,000$

SELECT name FROM employees

SELECT salary FROM employees

Alice ↔ 65,000$

David ↔ 80,000$

Giải pháp?

16

16

Giải pháp

- Thông thường không có giải pháp nào là toàn vẹn
  - Luôn phải đánh đổi sự riêng tư/an toàn bảo mật với tính tiện dụng
- Giải pháp 1: Chia nhỏ CSDL thành các CSDL khác nhau
- Giải pháp 2: Phát hiện ra cách thức suy luận khi truy vấn
  - Lưu trữ tạm thời kết quả của tất cả truy vấn từ một người dùng và tìm kiếm các suy luận trái phép trước khi gửi kết quả
  - Hạn chế: Không ngăn cản được sự câu kết giữa những người dùng khác nhau
  - → Lưu kết quả của tất cả truy vấn của mọi người dùng

17

17

Chia nhỏ CSDL

- DB1 lưu trữ (Name, Years of Service), có thể truy cập bởi mọi người dùng
- DB2 lưu trữ (UID, Salary), có thể truy cập bởi mọi người dùng
- DB3 lưu trữ (UID, name) chỉ có thể truy cập bởi người dùng quản trị

18

18

CSDL thống kê

- Xây dựng CSDL chỉ cung cấp dữ liệu là các giá trị thống kê(giá trị trung bình, độ lệch chuẩn)
  - Phương pháp 1: CSDL thuần thống kê chỉ lưu trữ dữ liệu thống kê
  - Phương pháp 2: CSDL lưu trữ mọi dữ liệu nhưng chỉ trả lời truy vấn thống kê
- Với phương pháp thứ 2, cần thiết phải phát hiện và ngăn chặn các suy luận từ kết quả có được
  - Nguyên tắc chung: giới hạn các dữ liệu thống kê có chứa thông tin cá nhân.

19

19

Biến đổi CSDL thô thành CSDL thống kê

- Phương pháp 2: Giới hạn truy vấn được trả lời
  - Liệt kê các hàm, mệnh đề truy vấn được phép
  - Có thể thiết lập các ngưỡng đối với các hàm thống kê
  - Ví dụ: chỉ cho phép sử dụng hàm tính tổng với tối thiểu 3 bản ghi

Name	Genre	Years of service	Salary
Alice	F	12	65 000$
Charlie	M	20	70 000$
David	M	20	70 000$

SELECT SUM(salary) FROM employees WHERE Genre ='F'

Request denied!

20

20

Giới hạn truy vấn

Name	Genre	Years of service	Salary
Alice	F	12	65.000$

SELECT SUM(salary) FROM employees;

SELECT SUM(salary) FROM employees WHERE Genre ='M'

Lương của Alice

- Vấn đề:
  - Có thể suy luận từ kết quả của các truy vấn khác nhau
  - Có thể suy luận từ kết quả của cùng 1 câu truy vấn nếu CSDL có thay đổi. Ví dụ, có thể biết được lương của nhân viên mới
- Giải pháp: Dựa trên lịch sử truy vấn để xác định cho phép hay không cho phép thực hiện truy vấn tiếp theo

21

21

2. ẨN DANH

Bùi Trọng Tùng,
Viện Công nghệ thông tin và Truyền thông,
Đại học Bách khoa Hà Nội

22

22

Ẩn danh là gì?

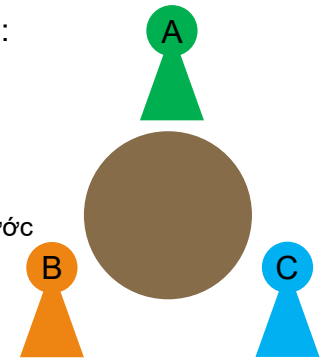
- Ẩn danh(Anonymity): che giấu danh tính của chủ thể
- Truyền thông ẩn danh (Anonymous Communication):
  - Ẩn danh người gửi: Không xác định được ai là người đã gửi thông tin trong một tập những người có khả năng
  - Ẩn danh người nhận
  - Ẩn danh người gửi-người nhận: không xác định được cặp giao tiếp trong các cặp có thể
- Mức độ ẩn danh được đánh giá qua lực lượng tập người gửi/người nhận:
  - Tập càng lớn, mức độ ẩn danh càng cao

23

23

Bài toán nhà mật mã học ăn tối

- Có 3 nhà mật mã học ăn tối cùng nhau:
  - Một người trong số họ muốn tiết lộ thông tin nhưng không muốn lộ danh tính
  - Giả sử, bản tin là 1 bit, cách thực hiện?
- Trao đổi khóa:
  - Mỗi người trao đổi bí mật 1 khóa có kích thước 1 bit với người bên cạnh
  - Mỗi người sẽ có 2 khóa k_{left} và k_{right}
- Công bố thông tin:
  - Nếu có thông tin m, công bố: $m \oplus k_{left} \oplus kright$
  - Nếu không, công bố: $k_{left} \oplus kright$



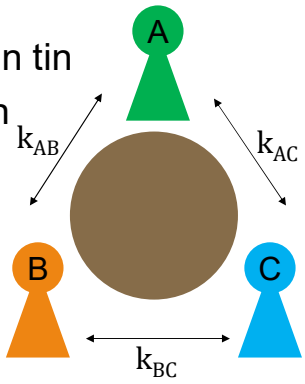
24

24

Bài toán nhà mật mã học ăn tối

- Nhận thông tin: XOR tất cả các bản tin
- Giải thích: Giả sử A tiết lộ thông tin
  - A: $m_A = m \oplus k_{AC} \oplus kAB$
  - B: $m_B = k_{AB} \oplus kBC$
  - C: $m_C = k_{BC} \oplus kAC$
  - Kết quả:

$m_A \oplus m_B \oplus m_C =$
$(m \oplus k_{AC} \oplus kAB) \oplus$
$(k_{AB} \oplus kBC) \oplus$
$(k_{BC} \oplus kAC) = m$



25

25

Kết quả thực hiện giao thức

- Tất cả đều biết:
  - Khóa của họ trao đổi với người bên cạnh
  - Nội dung thông tin
- Không ai biết giá trị bit còn lại
- Không ai biết người đã công bố thông tin
- Ví dụ

26

26

Bài toán nhà mật mã học ăn tối

- Chứng minh tính đúng đắn: David Chaum, *The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability*
- Ưu điểm:
  - Giao thức đơn giản
  - Các bên không cần tương tác theo cặp sau khi chia sẻ khóa
  - Rất khó để gian lận: tất cả những người còn lại hiệp sức mới có thể biết ai là người công bố thông tin → số người càng lớn, độ an toàn của giao thức càng cao
- Hạn chế:
  - Tình trạng đụng độ: giao thức không hoạt động nếu có >1 người cùng công bố
  - Bất kỳ ai trong nhóm cũng có thể phá hoại giao thức
  - Khóa không thể dùng lại

27

27

3. MẠNG TOR

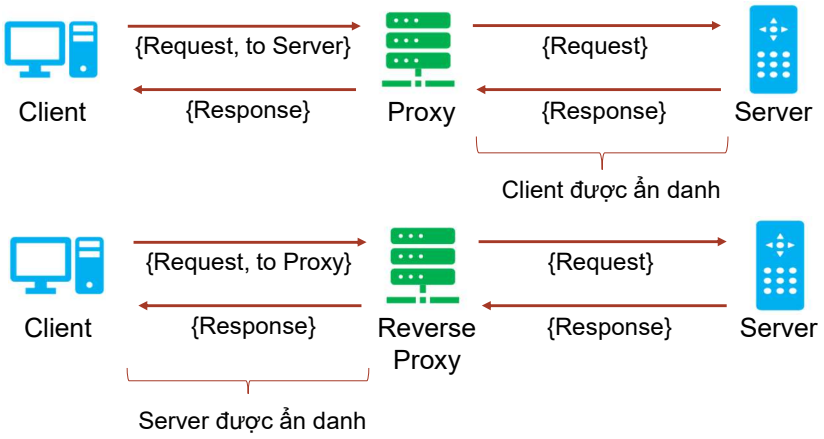
Bùi Trọng Tùng,
Viện Công nghệ thông tin và Truyền thông,
Đại học Bách khoa Hà Nội

28

28

Làm thể nào để truyền tin ẩn danh

- Sử dụng dịch vụ proxy



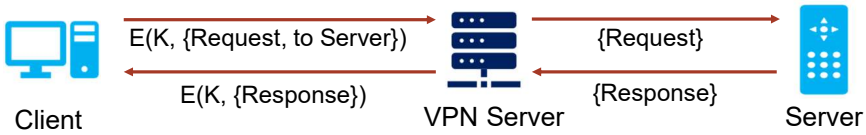
Tuy nhiên, proxy vẫn biết được danh tính 2 bên

29

29

Làm thể nào để truyền tin ẩn danh

- Sử dụng VPN(Virtual Private Network)



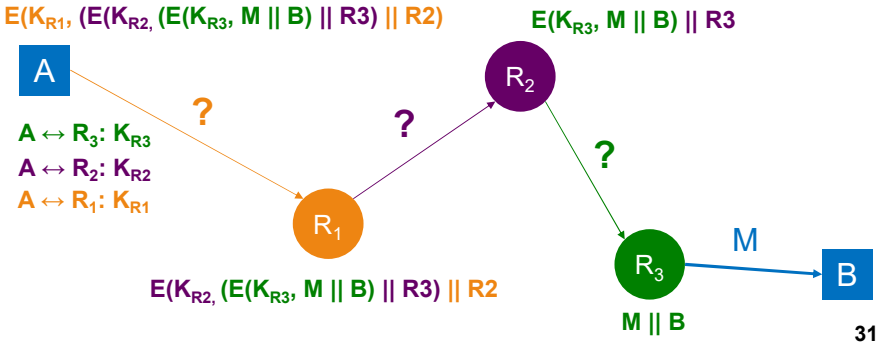
- Ẩn danh được bên gửi và bên nhận
- Tuy nhiên, VPN Server vẫn biết được danh tính 2 bên
- Cách thực hiện tốt hơn?

30

30

Định tuyến củ hành – Onion Routing

- Ý tưởng: Sử dụng một số lượng các nút tùy ý trung gian để chuyển dữ liệu
- Nguyên tắc: Không nút nào được biết đồng thời danh tính của cả 2 bên
- Thảo luận: Thiết lập như thế nào?



31

31

Tấn công vào Onion Routing

- Khai thác thông tin từ nhân viên điều hành các nút
- Chiếm quyền điều khiển một số lượng đủ lớn các nút
- Tấn công kênh bên: phân tích thời gian bên gửi phát dữ liệu và bên nhận thu được dữ liệu

32

32

Tor Project

- Tor network: cung cấp kết nối ẩn danh, chuyển tiếp dữ liệu thông qua hệ thống Onion Router
- Tor browser: phát triển từ trình duyệt Mozzila Firefox với các tính năng bảo vệ quyền riêng tư(privacy), sử dụng Tor network để kết nối tới Web server
- Tor Onion Service: cung cấp cơ chế cho các dịch vụ chỉ có thể truy cập thông qua Tor network
  - Ví dụ: Dark Web
- Tor bridge: cơ chế đóng gói dữ liệu truyền tới Tor network để tránh kiểm duyệt

33

33

Tor network

- Bao gồm hàng nghìn Onion Router, gọi là Tor node
  - Mỗi node có cặp khóa bất đối xứng
  - Giao tiếp với các node khác thông qua kết nối TLS (Transport Layer Security)
- Vanilla Tor: thiết lập một kênh truyền (Tor circuit) giữa nguồn và đích:
  - Bao gồm một số lượng ngẫu nhiên các Tor node
- Dữ liệu trên kênh truyền được bên gửi mã hóa theo nhiều lớp:
  - Số lớp bằng với số chặng
  - Qua mỗi một chặng sẽ giải mã 1 lớp

34

34

Thiết lập khóa cho kênh truyền

Sử dụng sơ đồ trao đổi khóa Diffie-Hellman

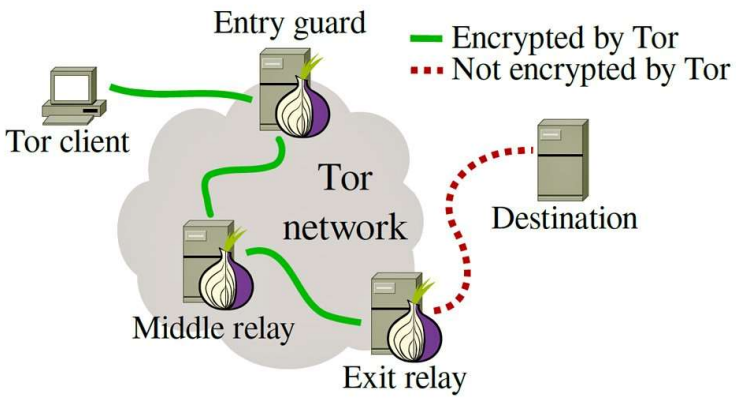
1) Bên gửi thiết lập khóa với nút thứ nhất (OR1)
   - Nút đầu tiên được gọi là nút canh giữ (guard node)
2) Bên gửi báo cho OR1 mở rộng kênh tới nút thứ 2 (OR2)
   - Thiết lập khóa với OR2 (bí mật với OR1)
   - OR2 không biết danh tính của bên gửi
3) Bên gửi báo cho OR2 mở rộng kênh tới nút thứ 3 (OR3)
   - Thiết lập khóa với OR3
   - OR3 không biết danh tính bên gửi
   - OR1 không biết kênh được mở rộng tới OR3
4) Tới nút cuối cùng gọi là nút ra (exit node, exit relay), kênh được hoàn thành

35

35

Tor network

- Kênh phải có tối thiểu có 3 nút
  - Tại sao?
- Encryption stack: $E(K_{OR1},E(K_{OR2}(E(K_{OR3}, M \,||\, B)||OR3)||OR2)$



36

36

Tor Onion Service

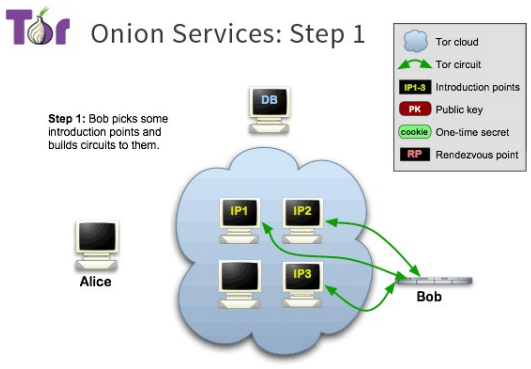
- Với Vanilla Tor và Tor Browser, danh tính client đã được che dấu
- Làm cách nào để cung cấp một dịch vụ mà máy chủ được ẩn danh?
- Tor Onion Service cho phép cung cấp 1 dịch vụ trên Tor network mà không để lộ tên miền và địa chỉ IP của máy chủ
  - Tên miền dạng *.onion
  - Khóa công khai của dịch vụ được băm để tạo ID cho dịch vụ
  - Ví dụ, chợ đen AlphaBay: http://pwoah7foa6au2pul.onion

37

37

Thiết lập Tor Onion Service

- B1: Chọn một số Tor node ngẫu nhiên để thiết lập kênh. Những nút này được gọi là điểm giới thiệu dịch vụ (introduction point)

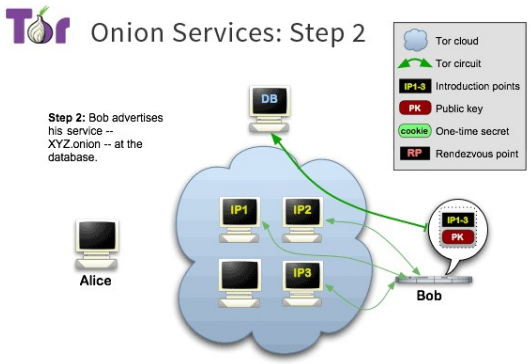


38

38

Thiết lập Tor Onion Service

- B2: Tạo bản mô tả dịch vụ bao gồm khóa công khai và danh sách các điểm giới thiệu dịch vụ. Bản mô tả được ký bằng khóa riêng. Lưu bản mô tả bằng CSDL với ID của dịch vụ là mã băm(16 byte) của khóa công khai.

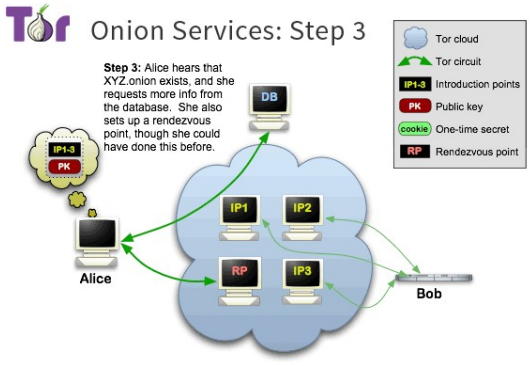


39

39

Kết nối Tor Onion Service

- B3: Client thiết lập kênh ẩn danh và yêu cầu 1 nút ngẫu nhiên là điểm hẹn(rendezvous point)

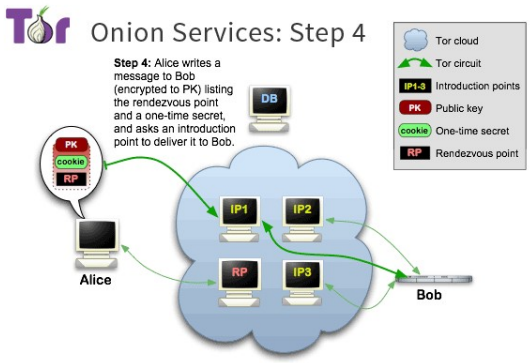


40

40

Kết nối Tor Onion Service

- B4: Client gửi yêu cầu kết nối tới một trong các điểm giới thiệu dịch vụ. Yêu cầu này được mã hóa bởi khóa công khai của dịch vụ, nội dung bao gồm địa chỉ điểm hẹn, giá trị bí mật cookie

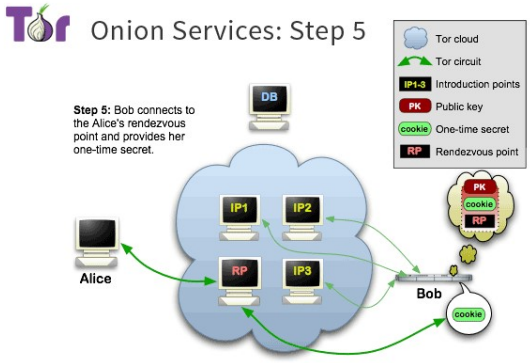


41

41

Kết nối Tor Onion Service

- B5: Máy chủ dịch vụ kết nối tới điểm hẹn và gửi đi giá trị ngẫu nhiên

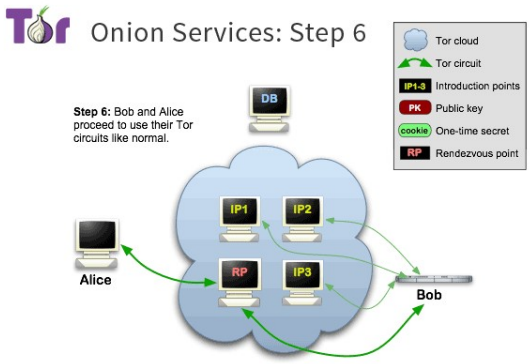


42

42

Kết nối Tor Onion Service

- B6: Nút điểm hẹn gửi thông báo tới client đã thiết lập được kết nối tới dịch vụ.



43

43

Các vấn đề của Tor Onion Service

- Máy chủ phải sử dụng nút canh giữ (guard node) tin cậy trong một thời gian dài để thiết lập kênh truyền
- Điểm giới thiệu dịch vụ và điểm hẹn phải là các nút khác nhau
- Tốc độ rất chậm:
  - Số chặng mỗi kết nối: 6+

44

44

Tor có thực sự giúp ẩn danh?

- Các cơ chế ẩn danh chỉ hiệu quả khi nó được sử dụng trong một đám đông:
  - Bạn không khác biệt với những người khác
- Mặc định, rất dễ để chỉ ra rằng "Ai đang truy cập vào Tor network"
- Ví dụ:

5

45

Tor brigde

- Truyền tin ẩn danh cần phải thực hiện 2 yêu cầu:
  - Ẩn danh tính các bên
  - Chống lại các cơ chế kiểm duyệt
- Vanilla Tor chỉ đáp ứng tốt yêu cầu đầu tiên
- Tor bridge cung cấp cơ chế để đóng gói dữ liệu của mạng Tor bằng các giao thức khác:
  - obs3
  - Meek

46

46

Truy nguyên Tor Onion Service

- Thu thập thông tin từ dịch vụ
  - Công cụ: onion scan
- Tấn công dịch vụ
- Truy nguyên client truy cập
  - FBI sử dụng các kỹ thuật Network Investigative Technique khai thác lỗ hổng của Tor Browser
- Đưa các nút kiểm soát vào mạng Tor
  - Hiện nay Tor project rất thận trọng cấp phép cho một nút mới

47