BAWMAN

Randomized MAC Addresses

System Impacts and Implications

Baw Chng
info@bawman.com

IEEE

First presented at an IEEE event on 2021 October 19

Biography

BAW CHNG @ BAWMAN LLC

- Proud Senior Member of IEEE. Founder and President of BAWMAN LLC
- Consult in the Wireless Networking and Telecommunications industry
- **Services offered**:
  - **Strategic Consulting** – business strategy, product positioning, technology roadmap development, RFI/RFP
  - **Standards Consulting** – standards development, standards & trade group representation
  - **Technology Consulting** – network system architecture, technology development, evaluation and trial
- **Clientele**: communications service providers, technology suppliers, equipment vendors, start-ups, established enterprises, investors; domestic and international
- **Expertise**: networking systems, wireless communications, standards and protocols
  - cellular, Wi-Fi, IoT, LPWAN, optical, NFV, SDN, cloud, IEEE, IETF, 3GPP, WBA, CableLabs, Broadband Forum
- **Contact**: www.bawman.com ; info@bawman.com

Outline

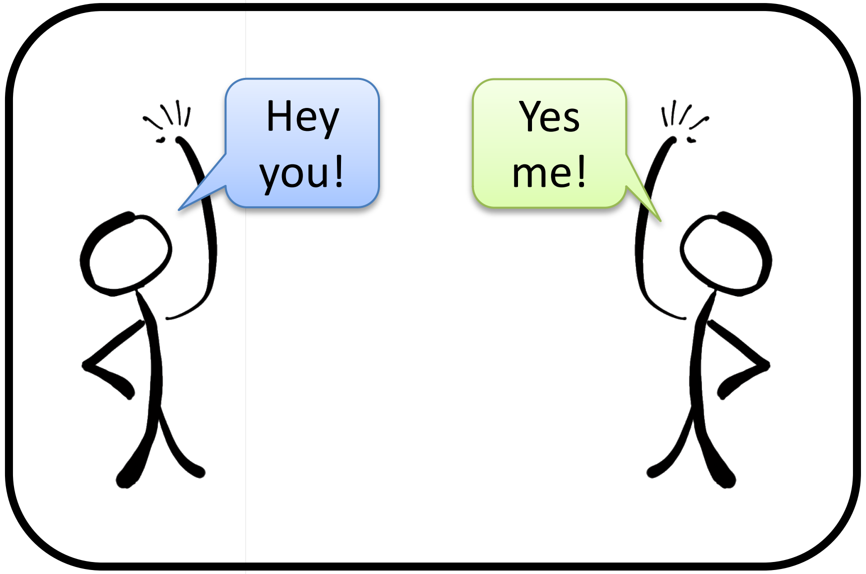
- Biography
- Introductory Background Information
  - High level concepts, what are MAC addresses
  - Motivation: Why randomize MAC addresses? Why now?
- How are MAC addresses randomized?
- Issues from MAC address randomization
  - Analysis
  - Mitigations
- Summary

Randomized MAC Addresses
Introductory Background Information

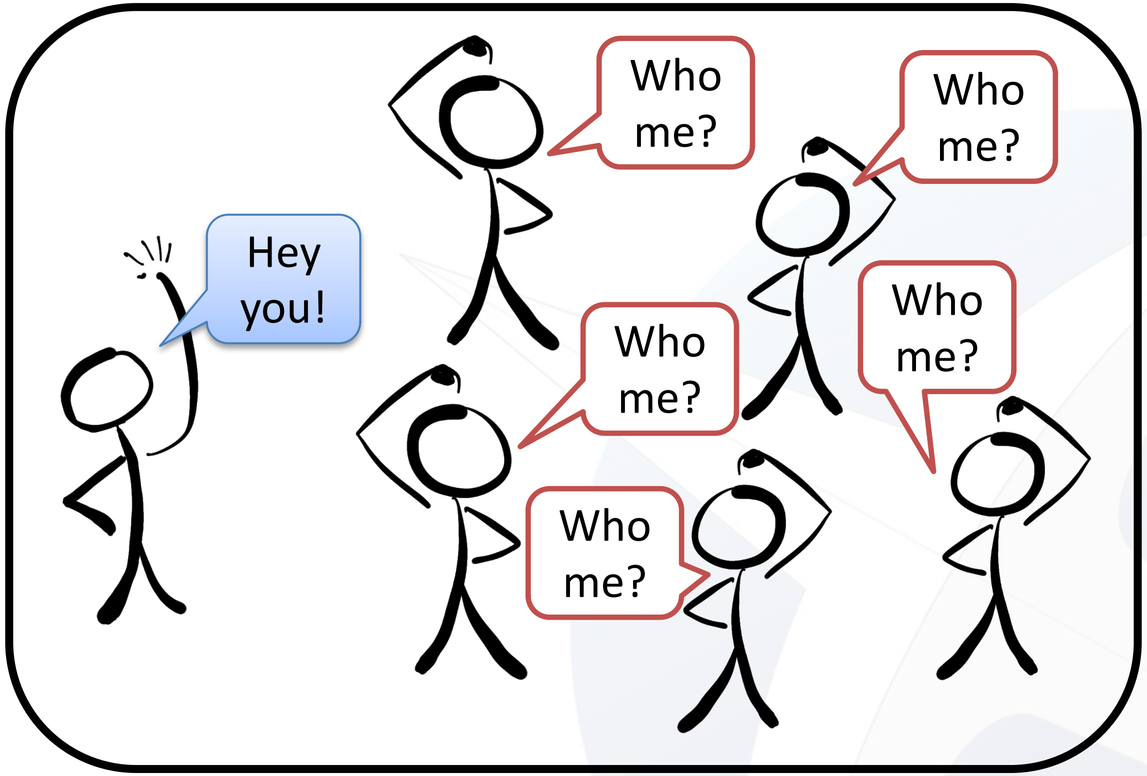
Concepts. What are MAC addresses? Why randomize them now?

Before we get technical … (1)

- When and why do we need “addresses” or “identifiers”?



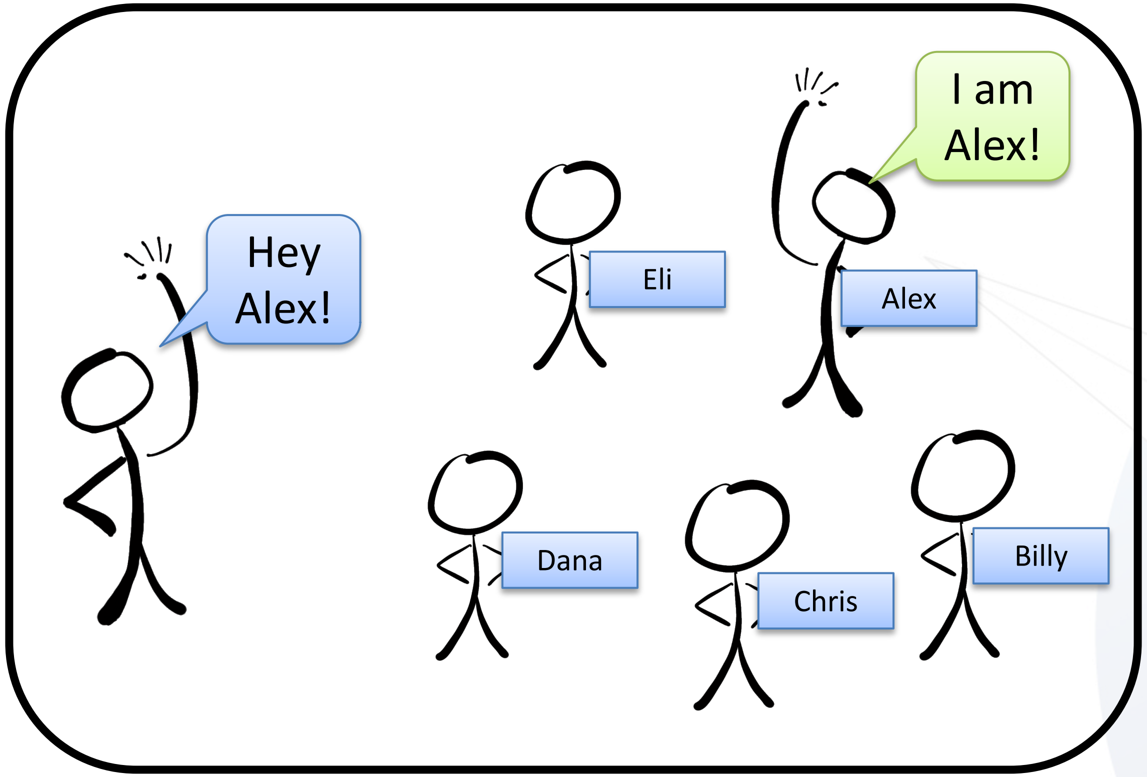
Not needed with two or fewer parties.
(E.g., “point to point” communications.)



Needed with three or more parties.
(E.g., “multiple access” communications.)

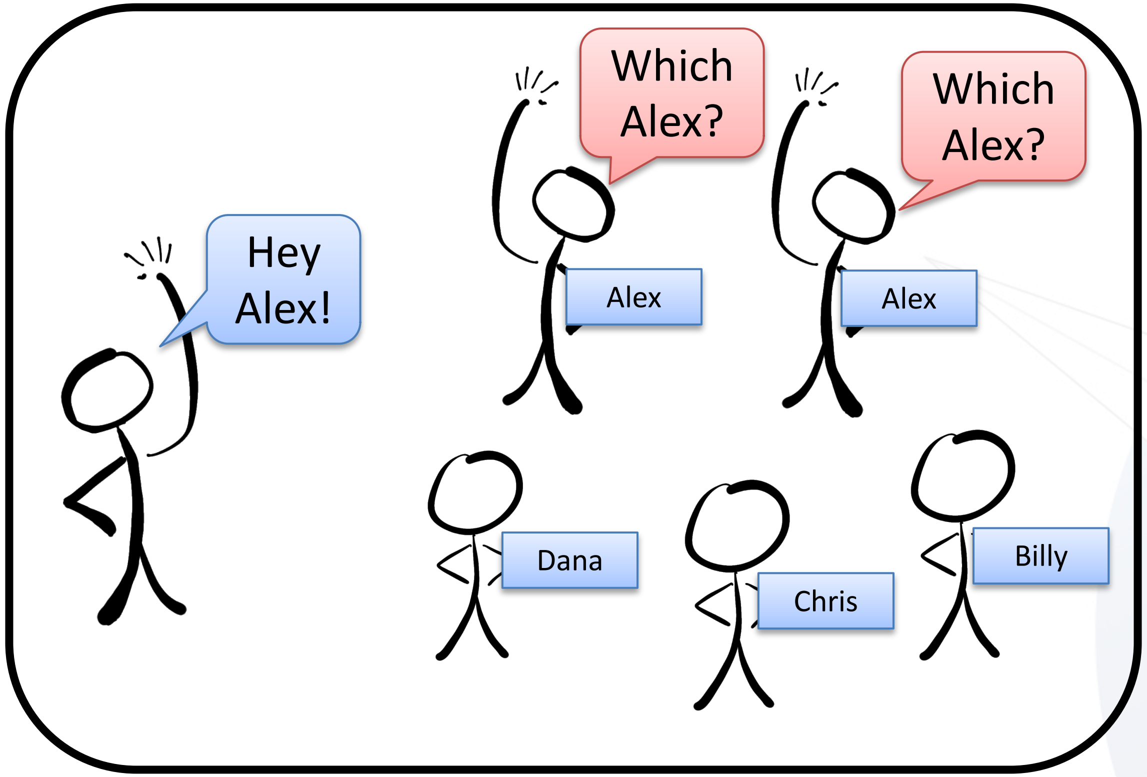
Before we get technical … (2)

- “Addresses” or “Identifiers” as means to disambiguate



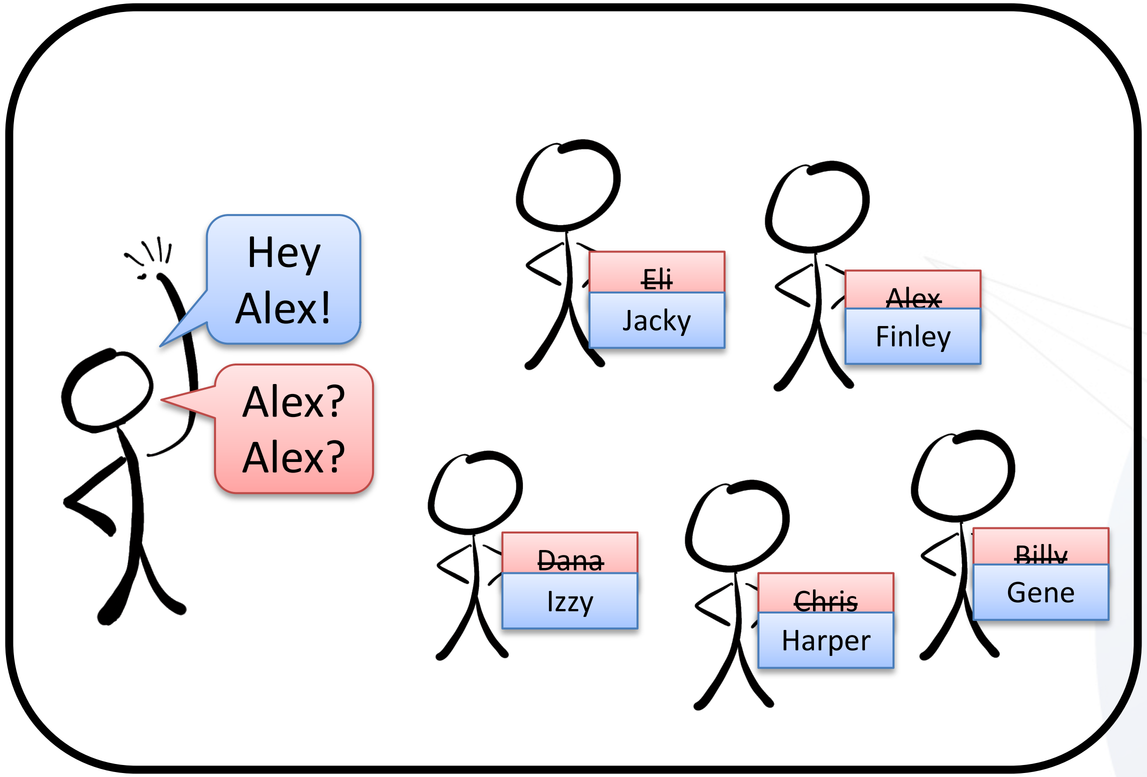
Before we get technical … (3)

- "Uniqueness" matters



Before we get technical … (4)

- Some notion of “permanence” or “continuity” matters



What are MAC Addresses?

- "MAC" = "media access control"
- Used to address/identify network interfaces
- Defined in IEEE 802 (standard)
- Focus on the EUI-48 (48 bit) form of MAC address
- EUI-48 is used by:
  - Ethernet (IEEE 802.3)
  - Wi-Fi (IEEE 802.11)
  - Bluetooth (started out as 802.15.1, now Bluetooth SIG)
- Address space allocation managed by IEEE

MAC Address Structure

- EUI-48 typical notation:
  - Examples:
    - **`F2:D3:CF:88:9A:07`**
    - **`F2-D3-CF-88-9A-07`**
    - **`F2D3CF889A07`**

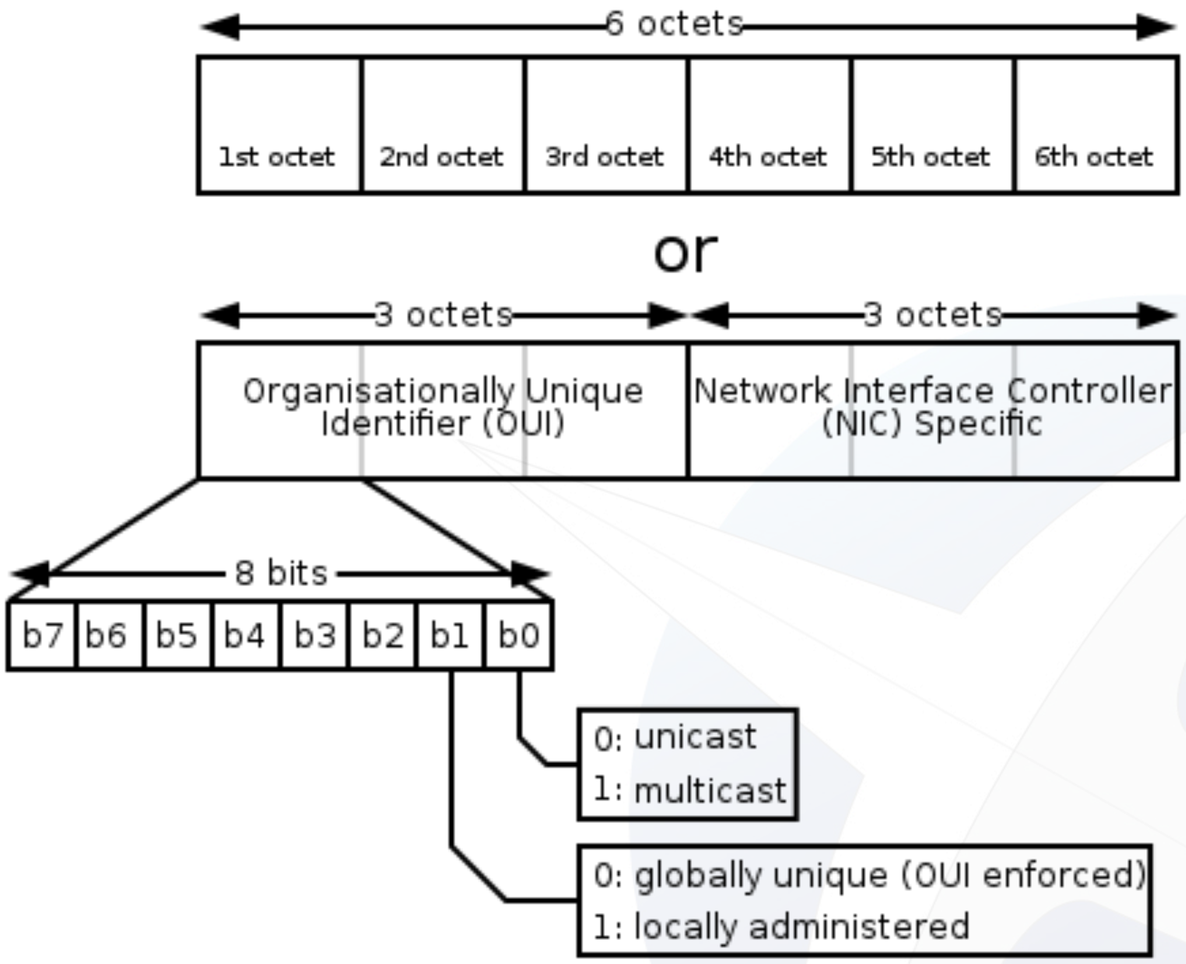


Image source/attribution: https://commons.wikimedia.org/wiki/File:MAC-48_Address.svg

MAC Address Allocation & Assignment

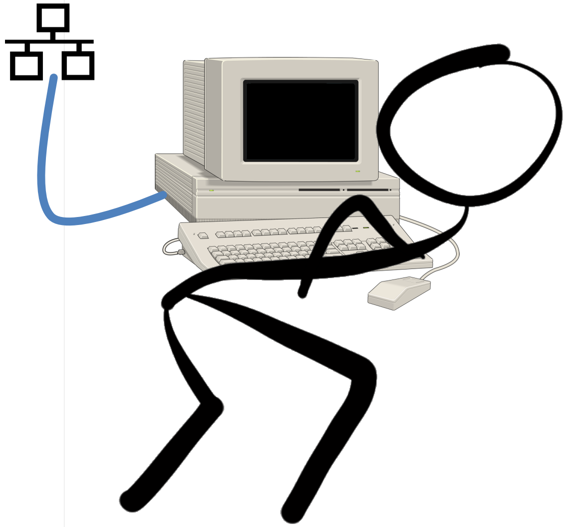
- Universal address, global uniqueness enforcement
  - – IEEE allocates “blocks” of address spaces to organizations
    - Typically hardware vendors
  - – Each organization to assign unique addresses from its “blocks”
    - Typically “burned” into hardware, read-only memory (ROM)

- Locally administered address
  - – Software/network administrator may assign other MAC addresses
  - – “Randomized addresses” must be “locally administered addresses”

Why Randomize MAC Addresses? Why now?

- **Privacy!**

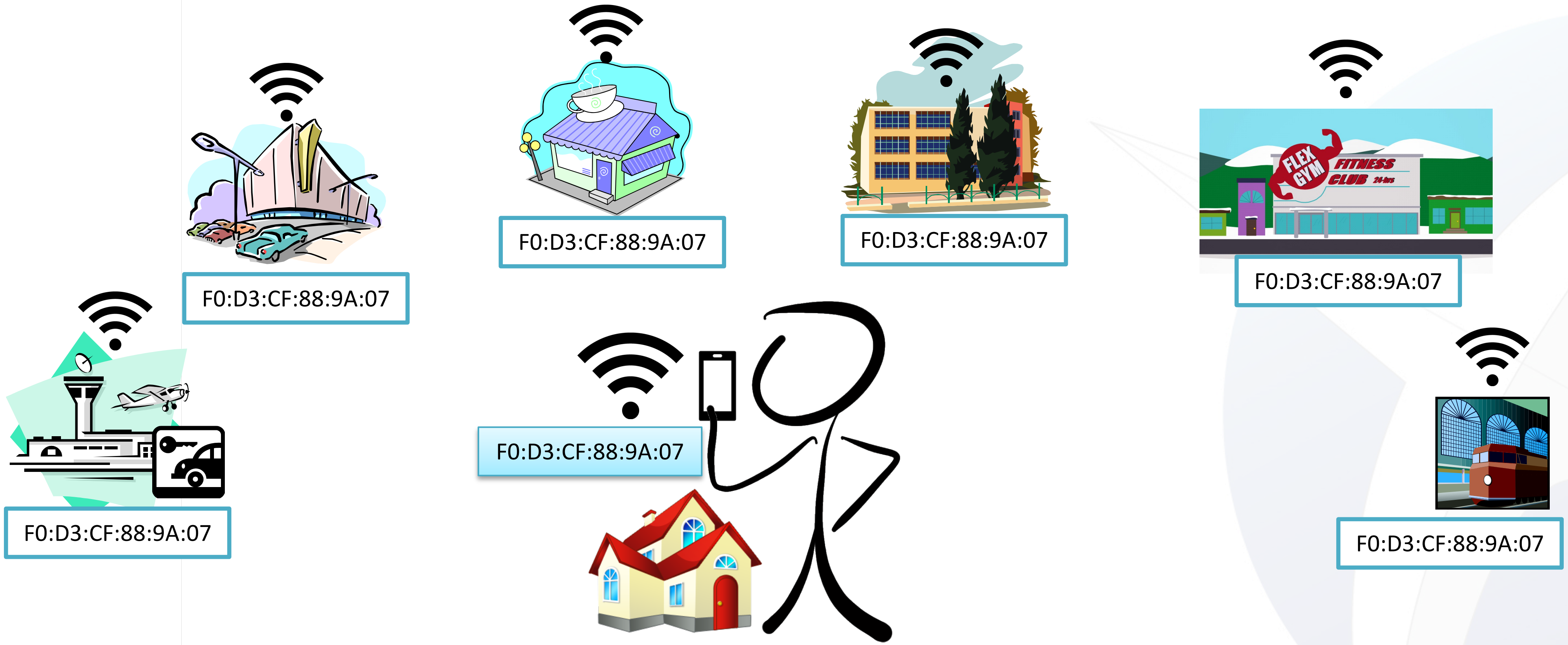
- We did not use to walk around with a MAC address



- Now we do …

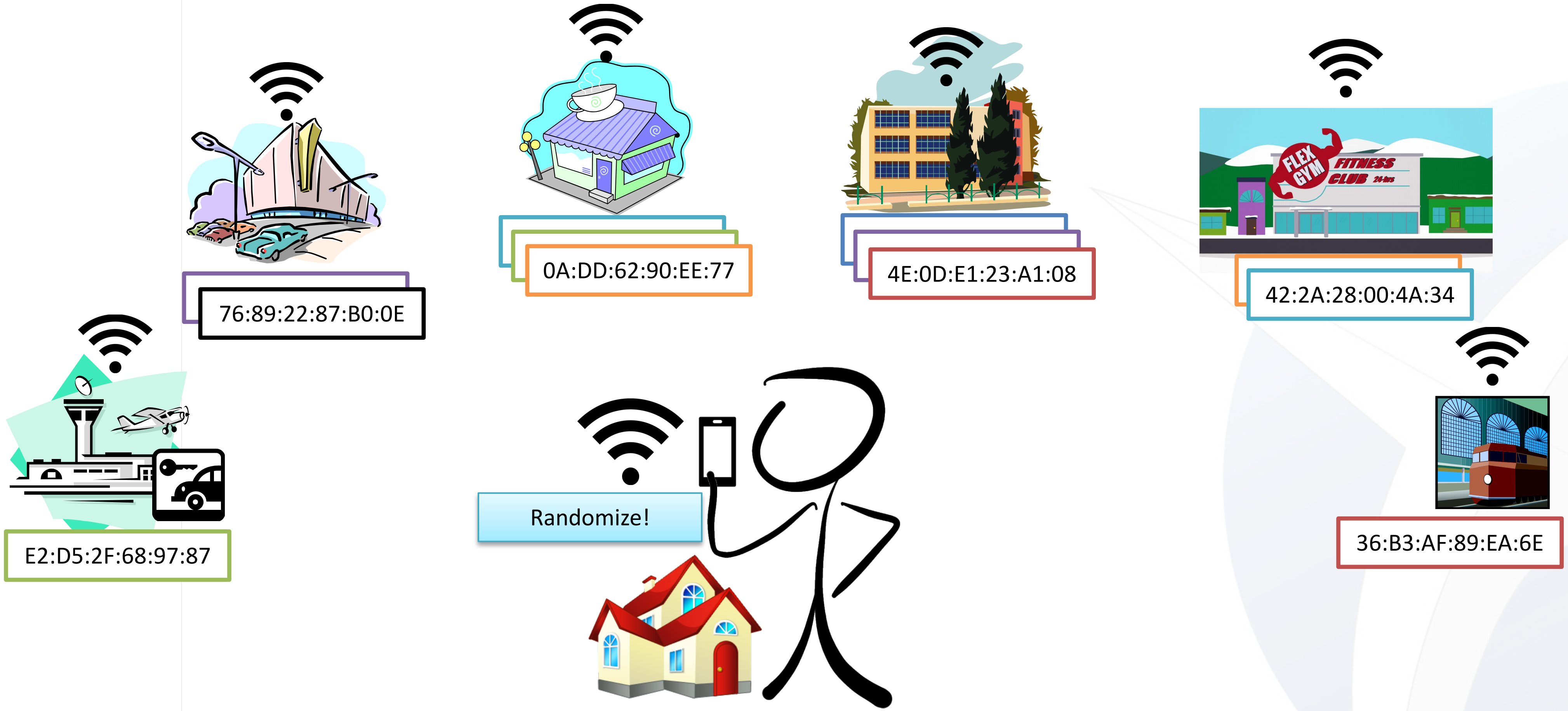
Tracking with MAC Address

- Using the same MAC address all the time allows tracking (geographical and temporal)



Prevent Tracking by Randomization

- Randomizing MAC address prevents tracking



Sidebar

How to tell if a MAC address is randomized?

- Cannot definitively tell if one is “randomized”
- Can only determine if one is “locally administered”
  – “randomized” MAC addresses must be “locally administered”

- First octet ends with { 2, 6, A, E }

#2:##:##:##:##:##	#6:##:##:##:##:##
#A:##:##:##:##:##	#E:##:##:##:##:##

BAWMAN

How are MAC Addresses Randomized?

How are MAC Addresses Randomized? (1)

- Randomization mechanism not prescribed by standards
  - No randomization method or distribution model specified in standards

- Precise randomization method, model, and parameters left to implementers:
  - May randomize multiple times for different reasons
    - e.g., discovery vs. attachment
  - May differ from one manufacturer to another
  - May differ from one device model to another
  - May differ from one operating system (version) to another

How are MAC Addresses Randomized? (2)

- Comprehensive surveys on how different devices and different operating systems randomize MAC addresses:

  - Martin, Jeremy, et al., “A Study of MAC Address Randomization in Mobile Devices and When it Fails”, *Proceedings on Privacy Enhancing Technologies* ; 2017 (4):268–286

  - Fenske, Ellis, et al., “Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds”, *Proceedings on Privacy Enhancing Technologies* ; 2021 (3):164–181

How are MAC Addresses Randomized? (3)

- Trend is to move towards 46-bit randomization
  - All bits in EUI-48 randomized excluding only
    - Unicast/Multicast bit
    - Universal (Globally Unique)/Locally Administered bit
- Baseline assumptions for subsequence analyses and discussion
  - 46-bit randomization
  - Uniform distribution

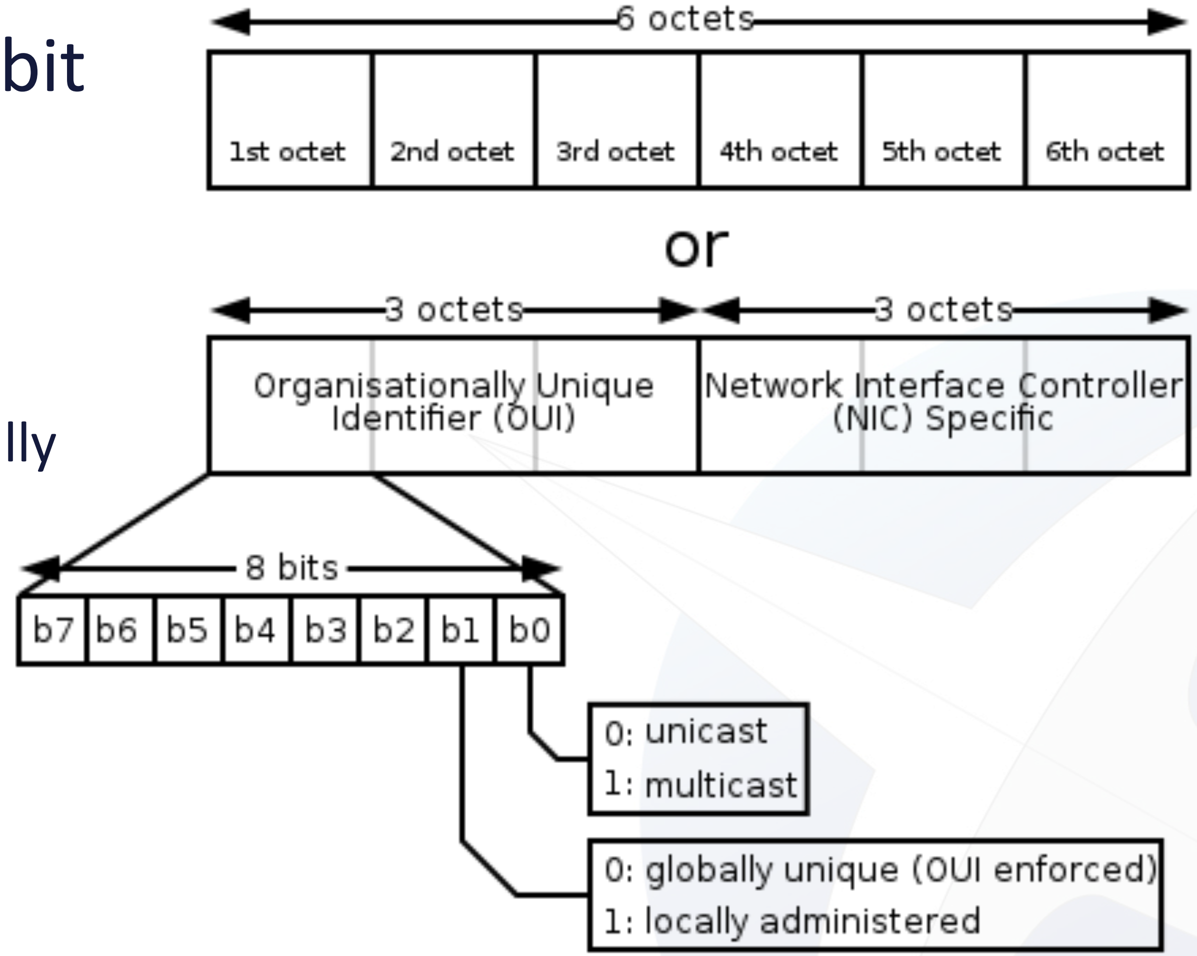


Image source/attribution: https://commons.wikimedia.org/wiki/File:MAC-48_Address.svg

Issues from MAC Address Randomization

Analyses, Discussion of Mitigation Options

Issues from MAC Address Randomizations (1)

- The obvious, consumer-facing issues …
  - – MAC address based access control
    - “Automatic” joining of Wi-Fi network
    - Billing based on MAC address
    - Throttling or usage cap based on MAC address

- The less obvious, back-end system issues …
  - – Analytics, AI/ML that use MAC addresses (marketing, advertising)
  - – Databases that use MAC addresses as indices (scale, duplicates)
  - – Resources mapped to MAC addresses (e.g., DHCP leases)

Issues from MAC Address Randomization (2)

- MAC address conflict
  - Two or more devices/network interfaces may end up with the same MAC address through independent randomizations
  - Networking protocols typically assume unique MAC addresses
  - *When does this become a problem? How big a problem is this?*

MAC Address Conflict Analysis (1)

- How likely is it for any two device interfaces to have the same MAC address?
  - Randomizable address space: 46 bits, $2^{46} \approx 70\ trillion$
  - Assume uniform distribution
  - Likelihood of conflict is a function of the number of device interfaces

- Model after the *Birthday Problem*
  - *"Given a fixed number of people in a room, how likely is it for any two people to have the same birthday?"*
    - $n =$*"number of people in a room"* ➔ number of device interfaces, variable
    - $d =$*"number of days in a year"* ➔ size of randomizable address space, $d = 2^{46}$

Sidebar

MAC Address Conflict Analysis (2)

- MAC address conflict modeled after the *Birthday Problem*
  - n =*"number of people in a room"* ➔ number of device interfaces, variable
  - d =*"number of days in a year"* ➔ size of randomizable address space, $d = 2^{46}$
- Probability of "no conflict" is
  - $\bar{P}(n,d) = 1 \times \left(1 - \frac{1}{d}\right) \times \left(1 - \frac{2}{d}\right) \times \cdots \times \left(1 - \left(\frac{n-1}{d}\right)\right) = \frac{n!}{d^n (d-n)!}$
- Probability of "conflict" is
  - $P(n,d) = 1 - \bar{P}(n,d) = 1 - \frac{n!}{d^n (d-n)!}$
- Hard to compute $\frac{n!}{d^n (d-n)!}$ when n and d are large
  - ➢ Resort to approximation

Sidebar

MAC Address Conflict Analysis (3)

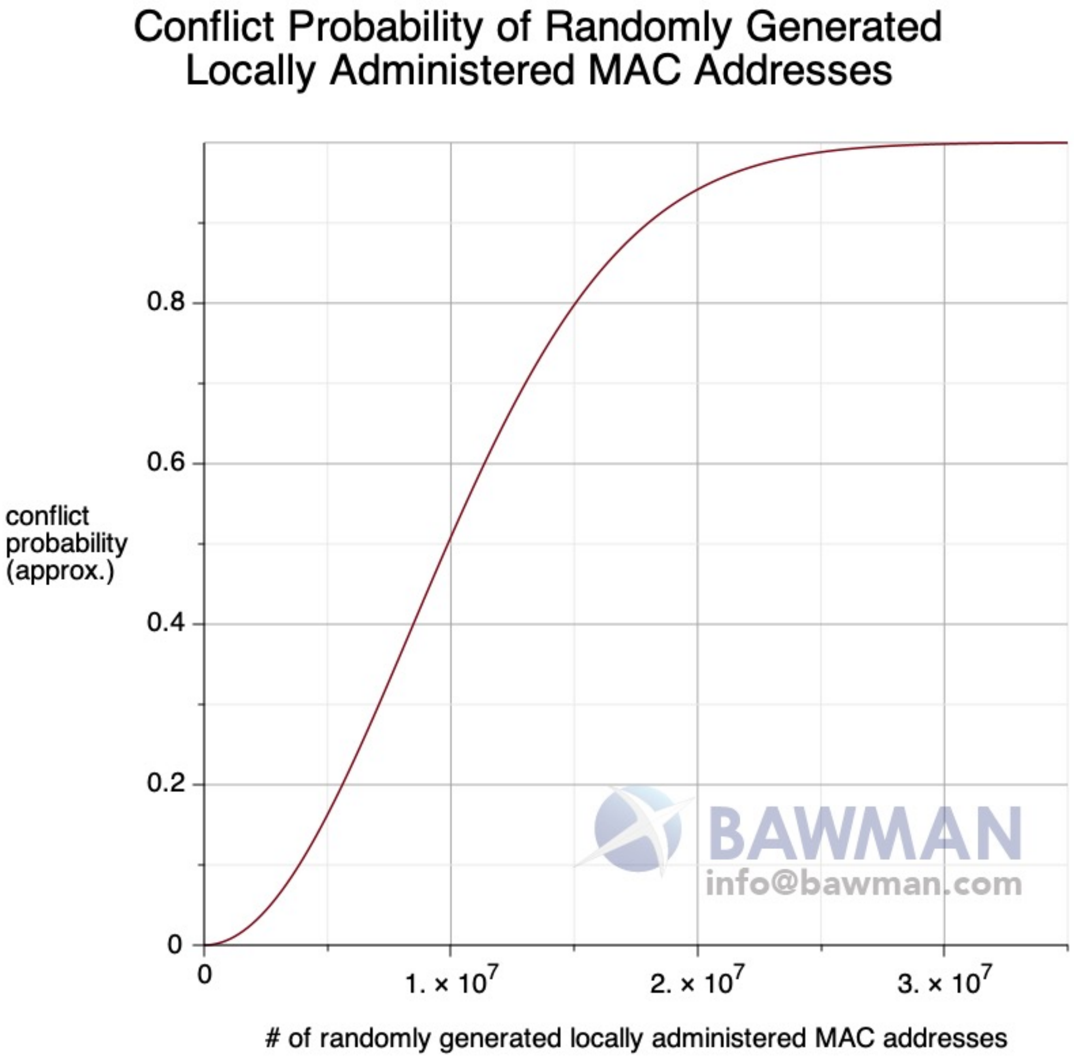
- Approximate $\bar{P}(n,d) = 1 \times \left(1 - \frac{1}{d}\right) \times \left(1 - \frac{2}{d}\right) \times \cdots \times \left(1 - \left(\frac{n-1}{d}\right)\right)$
  - Take advantage of Taylor series expansion of the exponential function
    - $e^x = 1 + x + \frac{x^2}{2!} + \cdots \approx 1 + x$ when x is small, i.e., $|x| \ll 1$
    - Set $x = -\frac{a}{d}$ to get $e^{\frac{-a}{d}} \approx 1 - \frac{a}{d}$
    - $|x| \ll 1$ is satisfied because $d = 2^{46} \approx 70\ trillion$
- $\bar{P}(n,d) \approx 1 \cdot e^{\frac{-1}{d}} \cdot e^{\frac{-2}{d}} \ldots e^{\frac{-(n-1)}{d}} = e^{\frac{-n(n-1)}{2d}}$
- $P(n,d) = 1 - \bar{P}(n,d) \approx 1 - e^{\frac{-n(n-1)}{2d}}$

MAC Address Conflict Analysis (4)

Approximated conflict probability

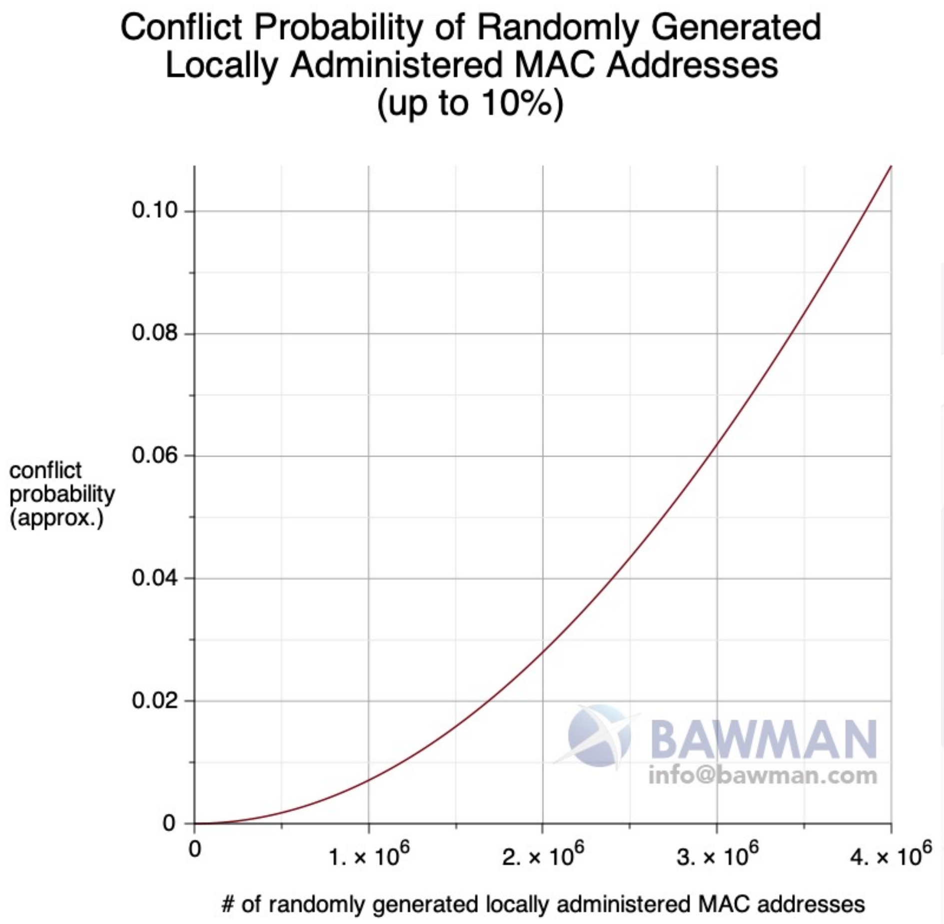
$$P(n, 2^{46}) \approx 1 - e^{\frac{-n(n-1)}{2 \cdot 2^{46}}}$$

where n is the number of device interfaces with randomized MAC addresses



MAC Address Conflict Analysis (5)

Number of randomly generated MAC addresses	Probability of MAC address conflict in a system
Roughly 38 thousand	0.001% (OK "five 9's" of the time)
Roughly 120 thousand	0.01%
Roughly 380 thousand	0.1%
Roughly 1.2 million	1%
Slightly below 4 million	10%
Slightly below 10 million	50%
Roughly 30 million	Near 100% (statistical certainty)



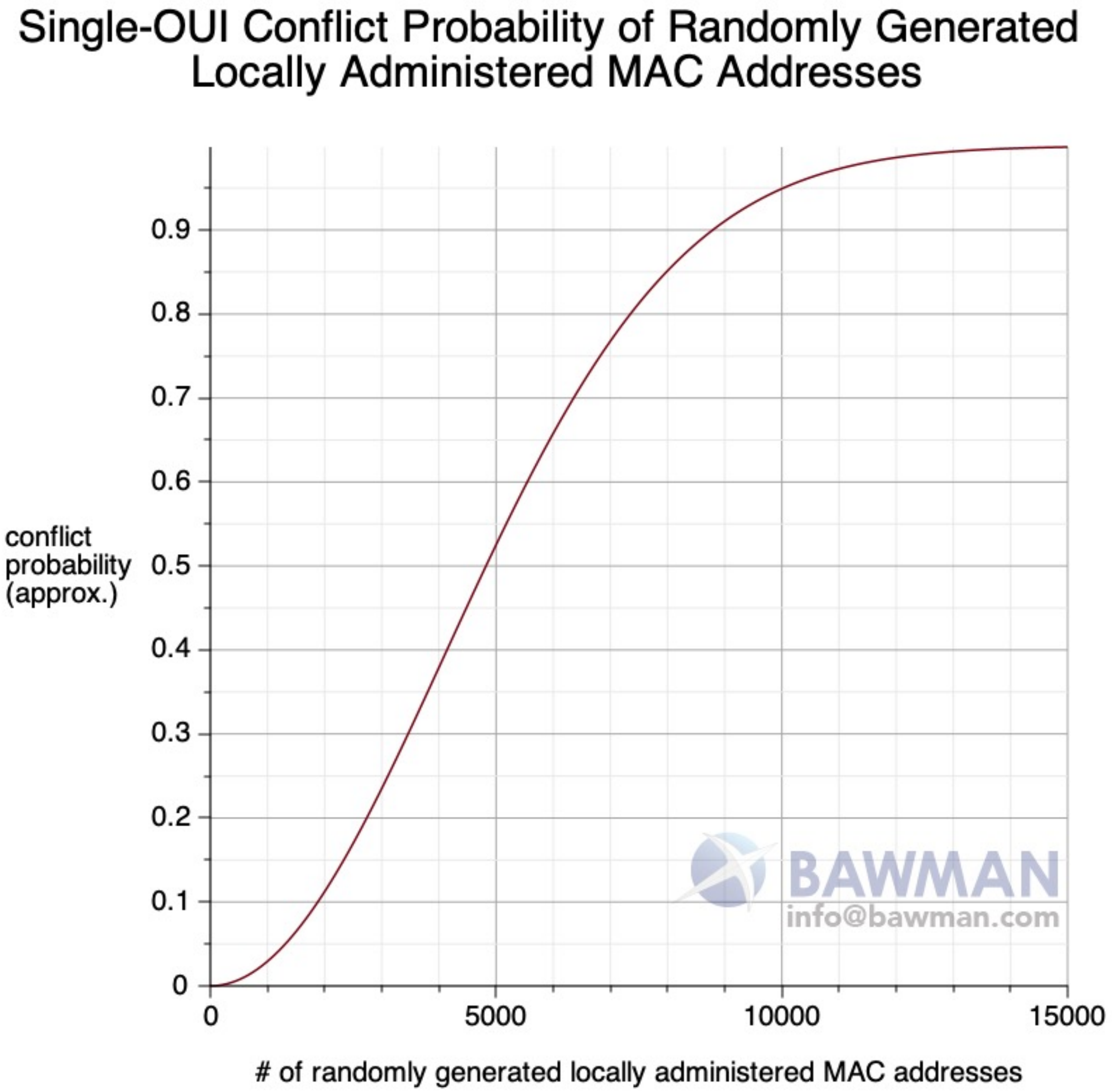
Sidebar

MAC Address Conflict Analysis (6)

- What if devices randomize only within assigned OUI?
  - i.e., first three octets "fixed"?
- OUIs are administratively distributed, but "clumping" likely because not all devices are equally popular in all places.

$$P(n, 2^{24}) \approx 1 - e^{\frac{-n(n-1)}{2 \cdot 2^{24}}}$$

- Better off randomizing over all 46 bits.



MAC Address Conflict Analysis (8)

- Beyond "30 million," it becomes more meaningful to estimate the number of device interfaces with conflicting addresses

$$N = n - n\left(\frac{2^{46} - 1}{2^{46}}\right)^{n-1}$$

Number of randomly generated locally administered MAC addresses in a system	Estimated number of non-unique MAC addresses in the system
50 million	36
75 million	80
100 million	142
125 million	222
150 million	320
200 million	568
250 million	888
500 million	3553
1 billion	14211

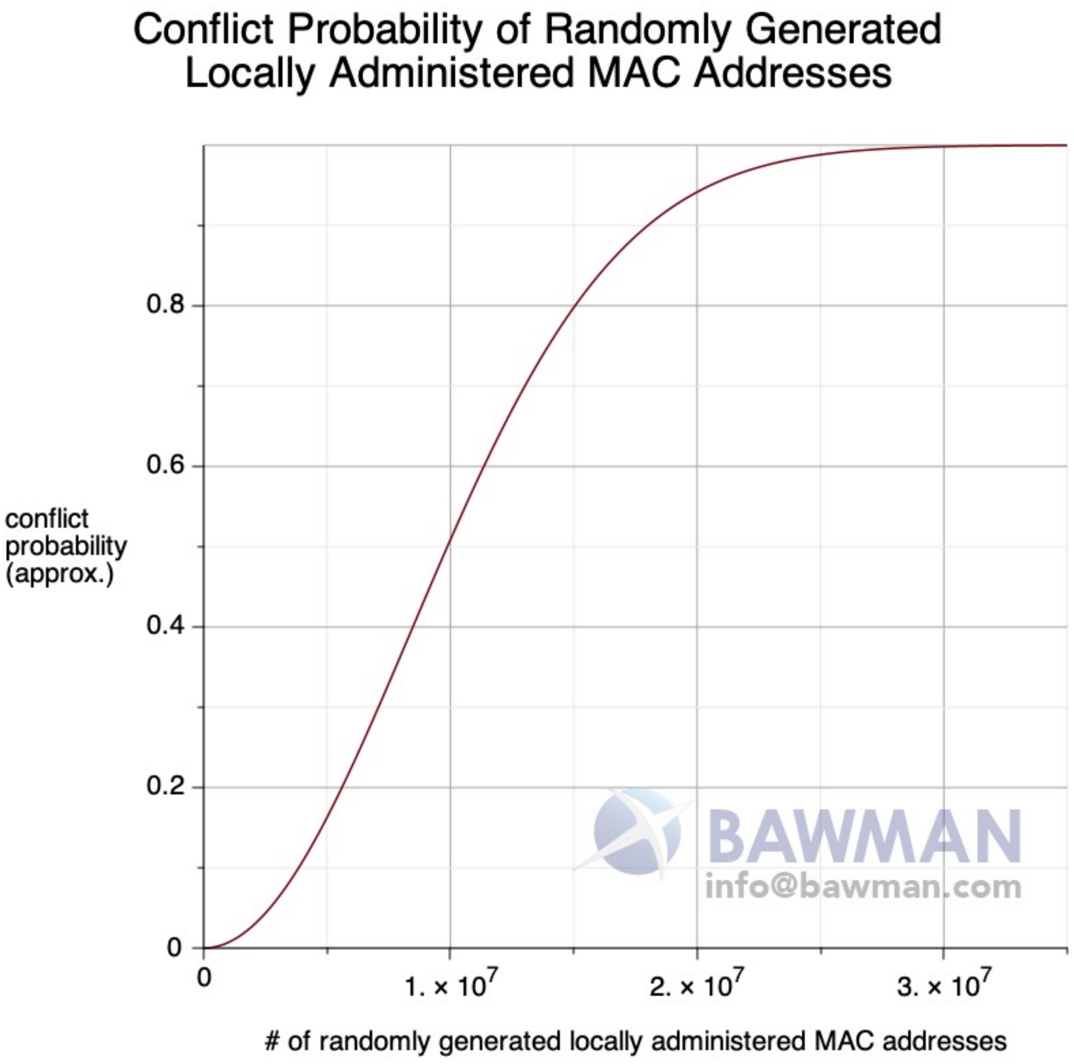
MAC Address Randomization Attendant Issues (1)

- Issues independent of system size (# of device interfaces)
  - MAC address based access control
    - Mitigation 1: disable MAC address randomization on the device
      - Requires user’s trust and intervention
      - Variant: disable MAC address randomization for select systems
        - » Requires some level of user sophistication (e.g., disable for which systems?)
    - Mitigation 2: reuse the same “randomized” address for the same system
      - Complications: how to determine if two networks are “the same system”?
        - » Easy if “the same system” is defined by “the same SSID”
        - » More complications otherwise
  - “Access control” also affects automatic system selection
  - Same issues extend to MAC address based billing and speed throttling

MAC Address Randomization Attendant Issues (2)

- Issues dependent on system size
  - MAC address conflict
  - High turn-over rate of MAC addresses

- Size scales to consider
  - Home, “small office”
  - Campus
  - Public venue
  - Multi-site, chain business
  - Infrastructure, public utility



MAC Address Randomization Attendant Issues (3)

- MAC address conflict
  - Unlikely to impact small systems
  - Unlikely to impact most campus systems

- High turn-over rate of MAC addresses
  - DHCP leases, (IPv4) address pool limitations
    - IPv4 addresses mapped to MAC addresses one-to-one
  - May impact even small systems (home, small office)
    - Consumer grade implementations may have a lower limit on the size of IPv4 address pool reservable via DHCP, a lower limit on the number of simultaneous DHCP leases that can be maintained at any one time
  - Mitigation (network side): shorten DHCP lease time
  - Mitigation (device side):
    - disable MAC address randomization
    - reuse the same “randomized” MAC address for the same network

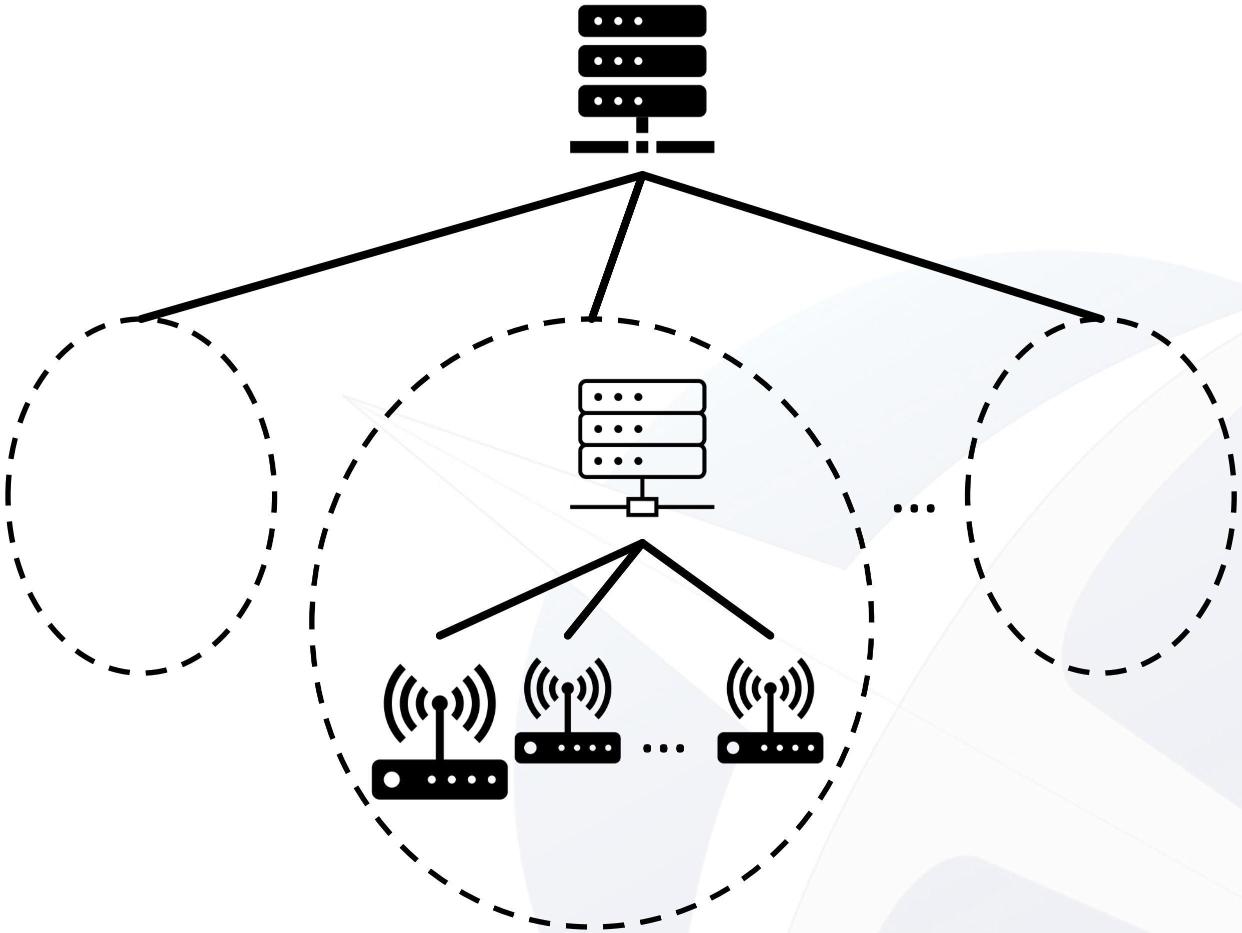
MAC Address Randomization Attendant Issues (4)

- MAC address conflict at scale
  - Impact larger systems
    - How "large" is "large"?
  - What sorts of impacts?
    - Basic data forwarding
    - Back-end analytics
      - Analytics for what purposes?

Number of randomly generated MAC addresses	Probability of MAC address conflict in a system
Roughly 38 thousand	0.001%
Roughly 120 thousand	0.01%
Roughly 380 thousand	0.1%
Roughly 1.2 million	1%
Slightly below 4 million	10%
Slightly below 10 million	50%
Roughly 30 million	Near 100% (statistical certainty)

MAC Address Randomization Attendant Issues (5)

- MAC address conflict, focus on
  - "Larger systems"
  - Basic data forwarding
    - Largely governed by standards
- "Large system"
  - Aggregates "small systems"
  - Built in "layers"



MAC Address Randomization Attendant Issues (6)

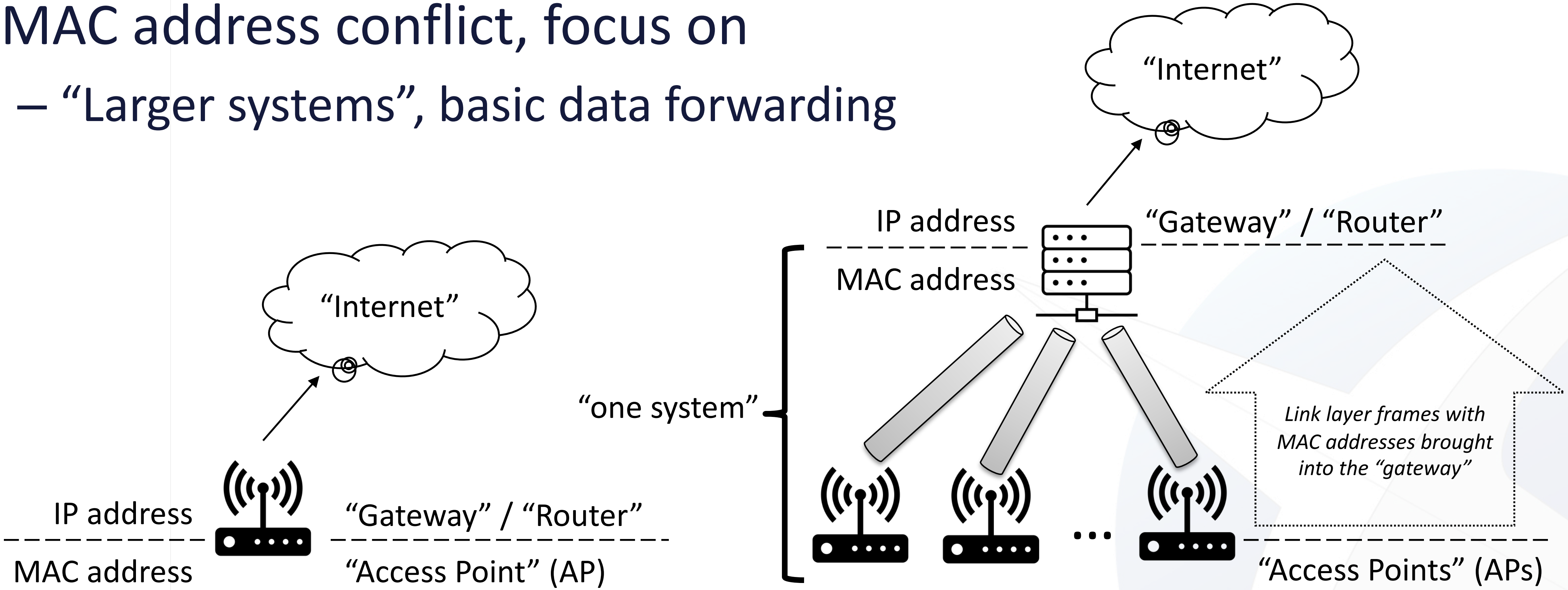
- MAC address conflict, focus on
  - “Larger systems”
  - Basic data forwarding
    - Largely governed by standards
- MAC address
  - Link layer concept
  - Conflict affects only link layer
  - Non-issue above link layer

Open Systems Interconnection
“OSI” reference model

Application Layer
Presentation Layer
Session Layer
Transport Layer
Network Layer
Data Link Layer
Physical Layer

MAC Address Randomization Attendant Issues (7)

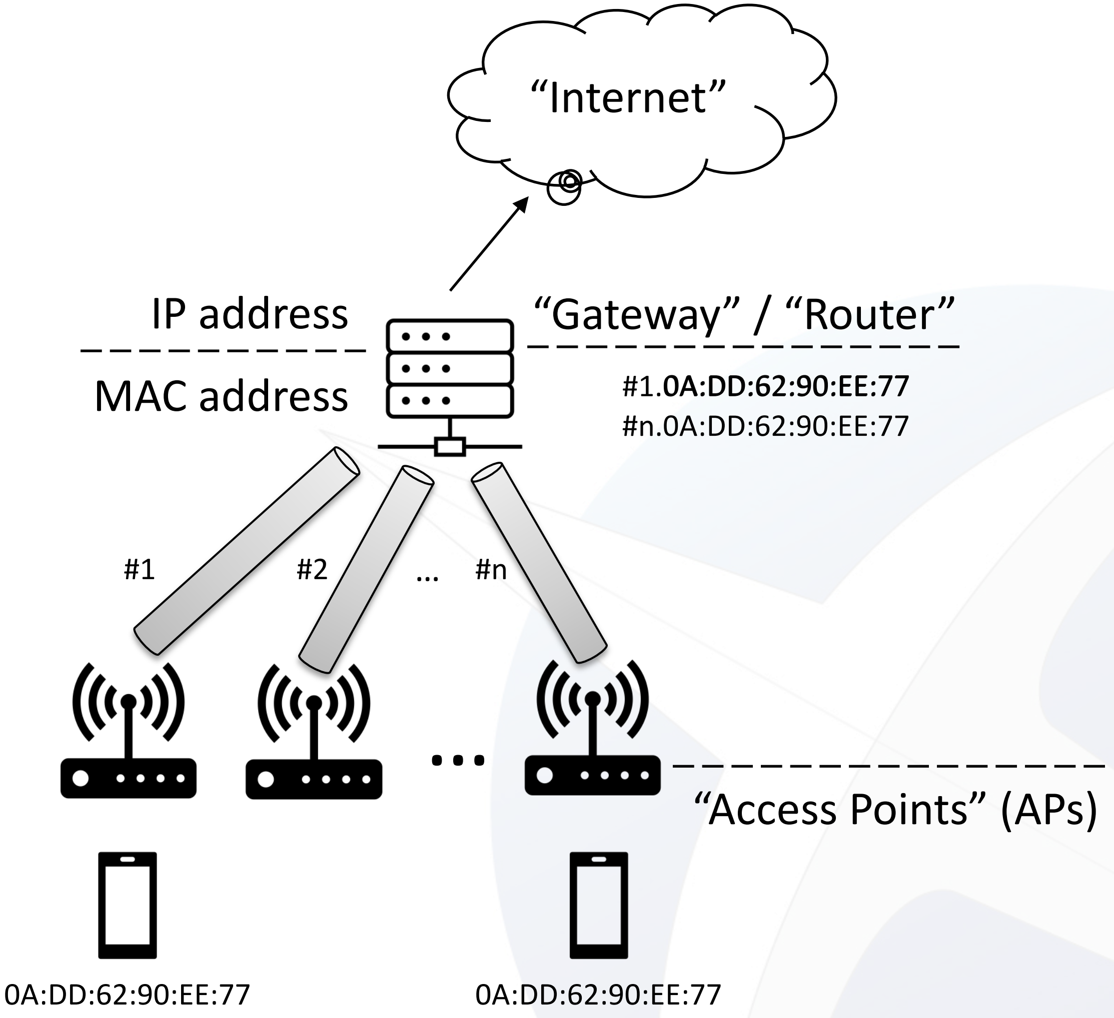
- MAC address conflict, focus on
  - "Larger systems", basic data forwarding



MAC Address Randomization Attendant Issues (8)

- MAC address conflict, focus on
  - "Larger systems"
  - Basic data forwarding

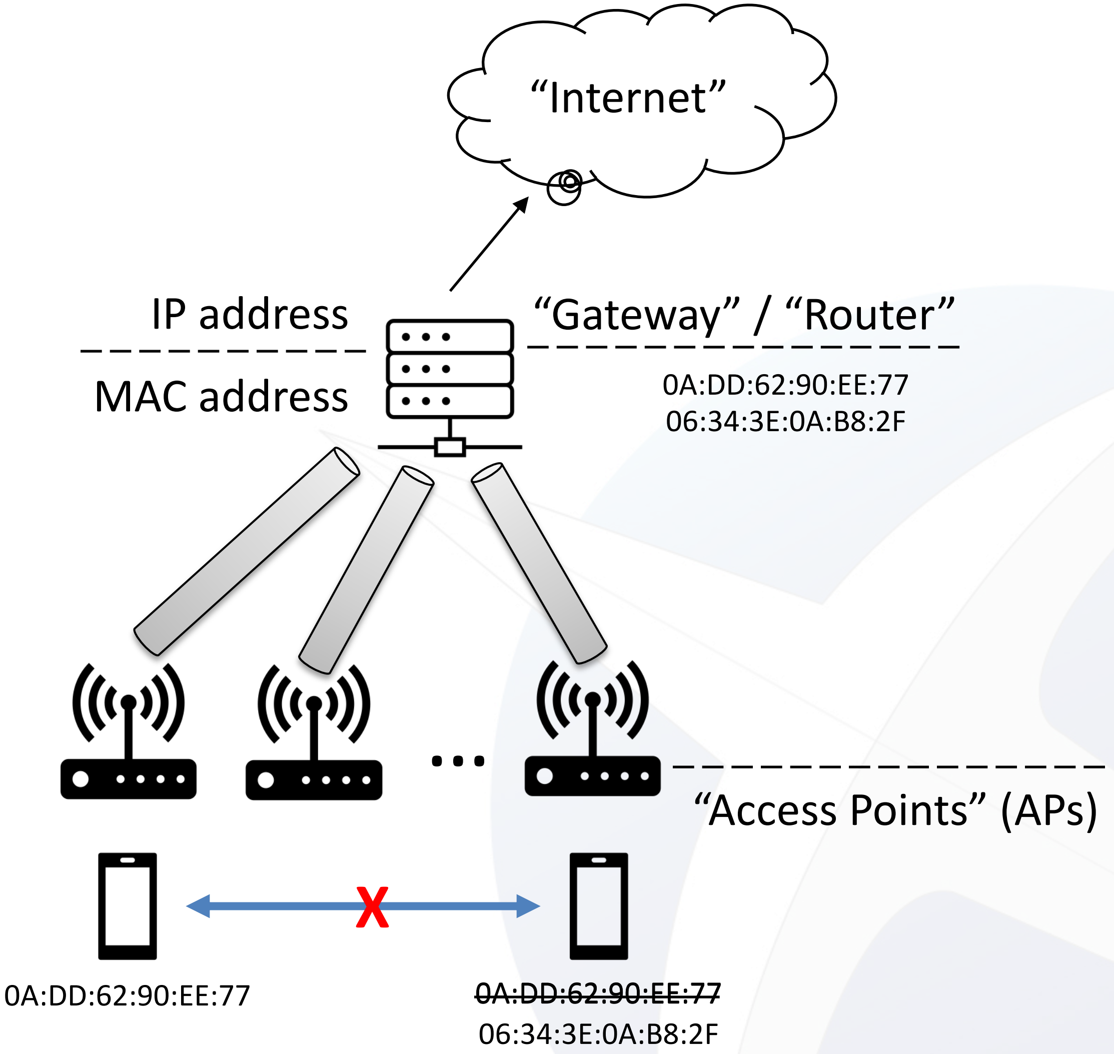
- Possible mitigation (1)
  - "Gateway" to enrich MAC address with tunnel identifier or access point identifier
    - Devices not required to change
    - Conceptually simple, implementation may be very complicated, very tedious for the "gateway"
    - Not "full proof"
      - Conflict may still occur within the same tunnel or the same access point



MAC Address Randomization Attendant Issues (9)

- MAC address conflict, focus on
  - "Larger systems"
  - Basic data forwarding

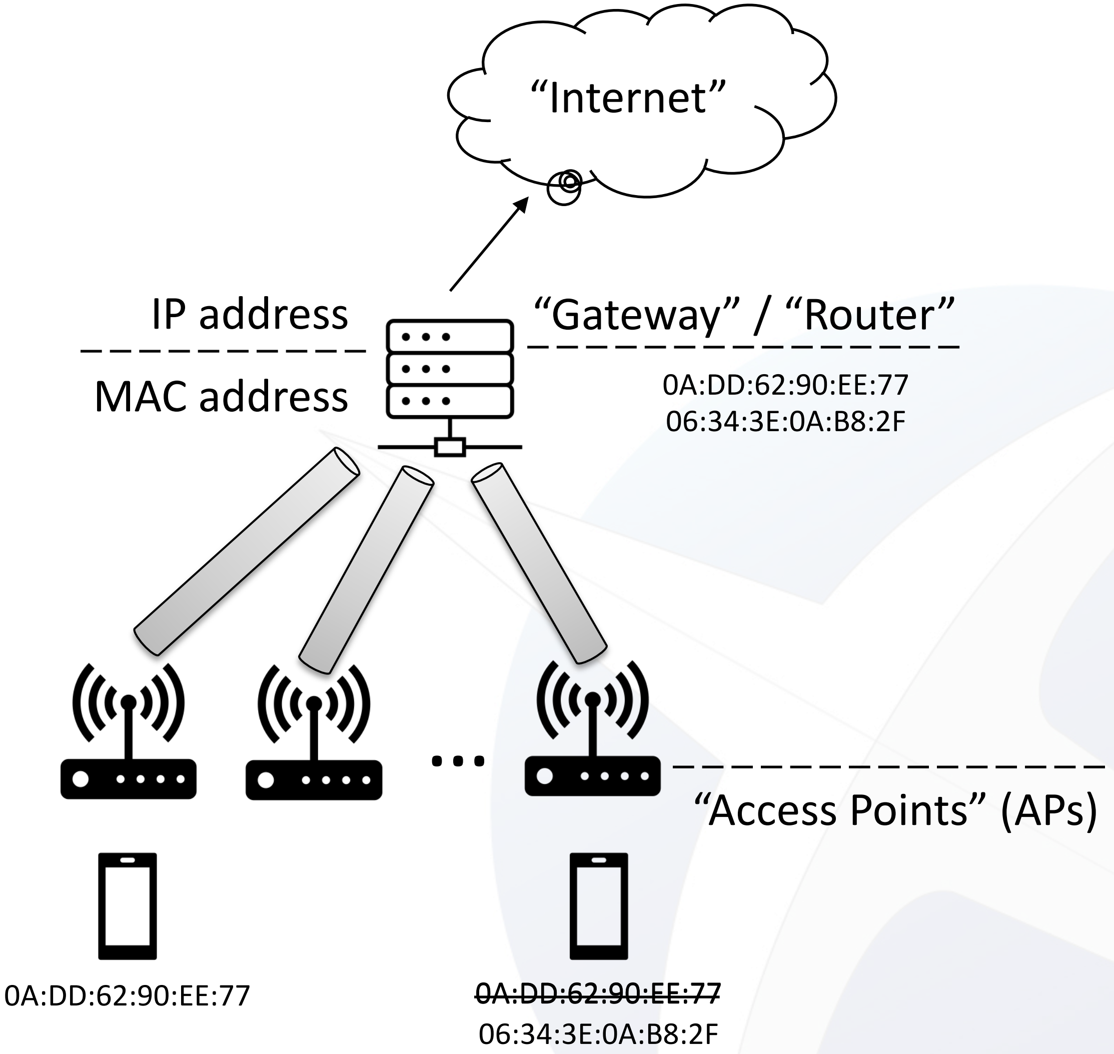
- Possible mitigation (2)
  - Devices, upon detecting that another device is using the same randomized MAC address, to disconnect and then rejoin the network with a different randomized address
    - Network (AP/gateway) not required to change
    - Simple implementation in devices
    - Not "full proof"
    - Works within the same broadcasting domain
    - Will not work across broadcasting domains where devices cannot naturally "hear" each other



MAC Address Randomization Attendant Issues (10)

- MAC address conflict, focus on
  - "Larger systems"
  - Basic data forwarding

- Mitigate by modifying standards
  - Allow the network to instruct a device to select another locally administered MAC address
    - Can be made "full proof"
    - Takes time to work through standards development, ratification, & adoption
      - Potential challenge to distinguish "new devices" from "transitioning/returning devices"
      - Potential security/privacy concerns with MAC address information leakage
    - Simple implementations for devices and single access points
    - Slightly more complicated implementations for hierarchical systems with controllers/gateways
    - Complete field adoption may take "forever"



MAC Address Randomization Attendant Issues (11)

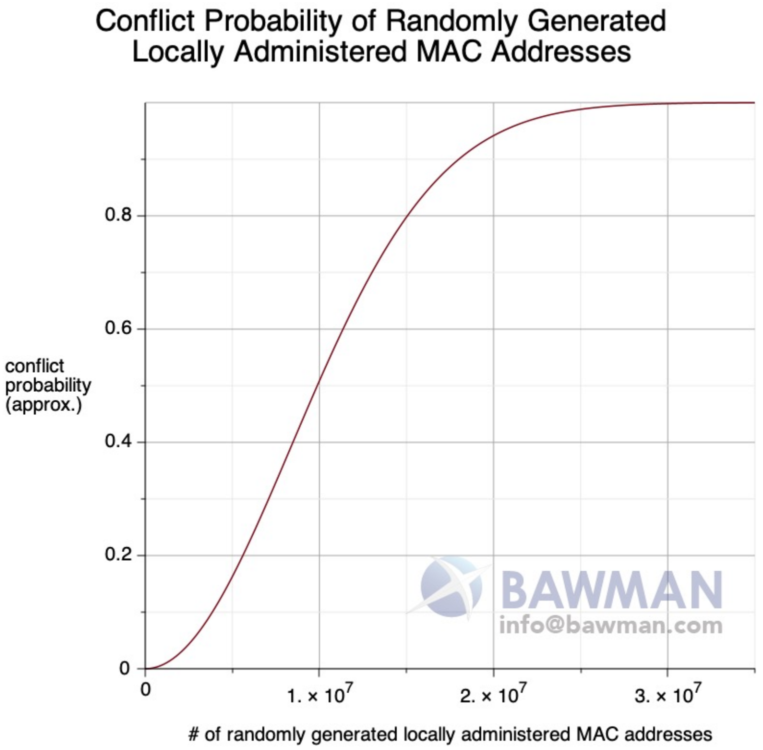
- Impacts on functions *other than data forwarding*
  - Use of MAC addresses not governed by standards
  - Harder to make general statements due to lack of standards
  - Nature and severity of impacts vary
  - MAC address is only “one of many” information elements

- Examples (for free form discussion)
  - Marketing/advertising databases and analytics
  - Traffic pattern analyses for capacity planning or academic research
  - Law enforcement surveillances

Randomized MAC Addresses Summary

Randomized MAC Addresses: Summary

- MAC address randomization are increasingly prevalent
  - Driven by privacy concerns, randomization enabled by default
  - Trending towards 46-bit randomization
- MAC address conflict analyses modeled after the *Birthday Problem*
  - Likelihood of data forwarding impact is dependent on system size
- Mitigating the data forwarding issues of MAC address randomization
  - Without modifying standards (implementation-only workarounds)
    - Not “full proof”
    - Implementation may be very hard, very tedious for the “gateway”
  - By modifying standards
    - Can be made “full proof”, implementations may be simpler
    - Long time frame from standards development to broad market adoption
- Impacts on functions *other than data forwarding*
  - Harder to make general statements on this due to lack of standards
  - Needs to be examined on a case-by-case basis



Number of randomized MAC addresses	Estimated number of non-unique MAC addresses
50 million	36
100 million	142
250 million	888
500 million	3553
1 billion	14211

Randomized MAC Addresses

Thank you!

Questions and Answers, Discussion …

Where to find this presentation and its corresponding paper:
https://bawman.com/BAWMAN/articles/RandomMAC/