SIEĆ DHT – SZYBKO I ANONIMOWO?

17 marca 2009

Adam Kubiaczyk

MOTYWACJA

- Web x.0
- Ciągły wzrost zainteresowania aplikacjami P2P
- Decentralizacja
- Cloud - computing

- Brak efektywnych mechanizmów zapewniających anonimowość i prywatność

PLAN PREZENTACJI

1. DHT – Distributed Hash Table
2. KAD – implementacja DHT
3. Anonimość i prywatność
4. Koncepcja P2PRIV
5. Interfejs KAD – P2PRIV
6. Miara anonimowości

PRZESZUKIWANIE ZASOBÓW

- Podstawowe wyzwania decentralizacji
  - lokalizacja zasobów
  - Przeszukiwanie
  - Indeksowanie
- Drogi do rozwiązania
  - Centralny serwer indeksujący – single point of failure
  - Flooding – nie ustrukturalizowana sieć
  - DHT – ustrukturalizowana sieć
- Dopiero DHT w pełni i skuteczny sposób realizuje to zadanie

DHT – DISTRIBUTED HASH TABLE

- Proste założenia
  - Duża przestrzeń kluczy
  - Ustalony algorytm obliczania klucza
  - Każdy węzeł ma swój ID
  - Przestrzeń kluczy podzielona między węzły
  - Sieć łącząca wszystkie węzły

DHT – DISTRIBUTED HASH TABLE

- Struktura
  - Przechowywane są pary (klucz,wartość)
  - Każdy węzeł odpowiada za część przestrzeni kluczy
- Operacje
  - Put(klucz,wartość) – zapisanie klucza z wartością
  - Get(klucz) – odzyskiwana jest wartość

DHT – DISTRIBUTED HASH TABLE

- Efektywny lookup
- Zdecentralizowana struktura
- Skalowalność
- Niezawodność

Chord, CAN, Tapestry, Pastry, Kademlia

....

DHT – DISTRIBUTED HASH TABLE

- Wymagania
  - Dobrze zbadana (teoretycznie)
  - Sprawdzona w praktyce
  - Wciąż doskonalona
  - W każdym aspekcie zdecentralizowana
  - Rzetelnie opisana

- Lista bardzo krótka…

KAD

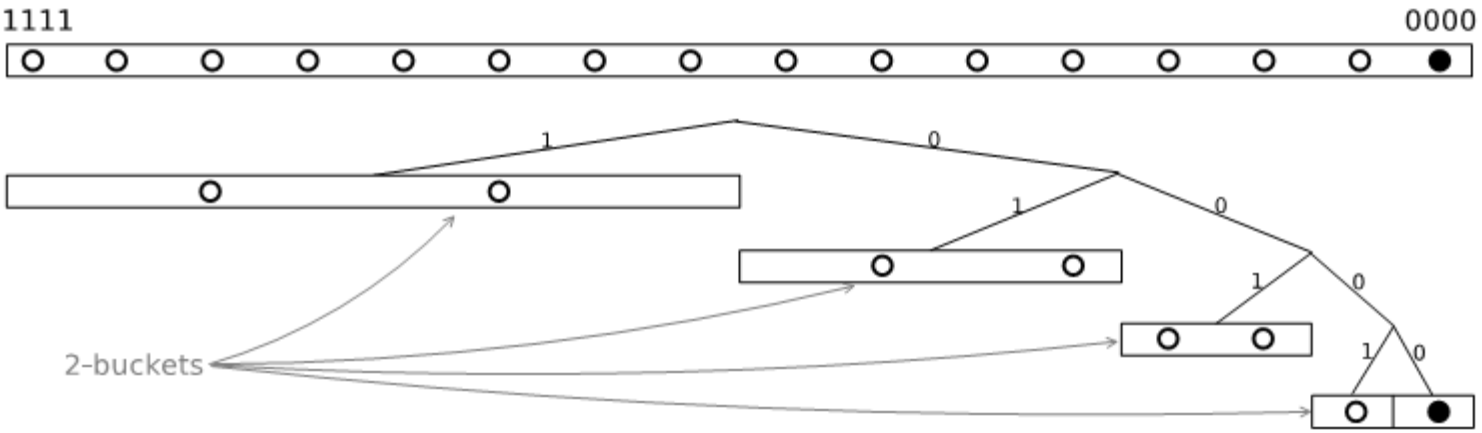
- Jedyna kompletna sieć DHT na większą skalę
  - 2 000 000 użytkowników
- Ciągle rozwijana
  - eDonkey -> eMule, aMule itd…
- DHT Kademlia
- e2dk hash
- Brak oficjalnej dokumentacji protokołu

KAD - SZCZEGÓŁY

- 128-bitowa przestrzeń kluczy
- Metryka XOR
- Jeden algorytm routingu w ciągu całego procesu wyszukiwania klucza
- Organizacja w drzewo binarne
- Implementacja Kademlii

KAD-TABLICA ROUTINGU

- Drzewo binarne
  - Liście drzewa – „K-buckets”
  - K-bucket – reprezentuje poddrzewo
  - Drzewo zorganizowane według „odległosci”



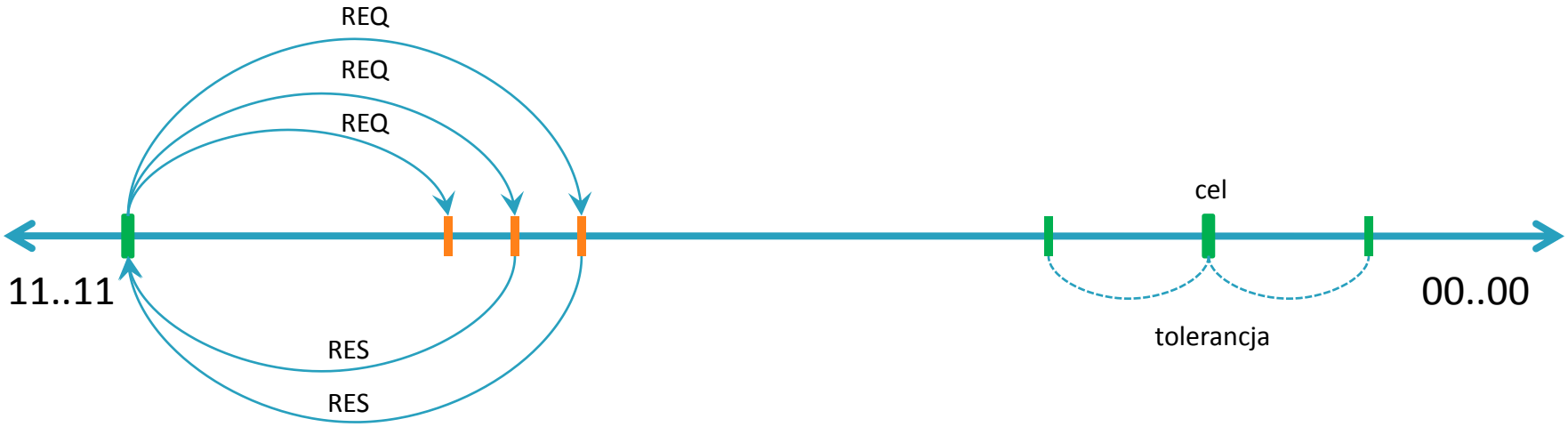
Źródło: Brunner R. A performance evaluation of the Kad-protocol.

KAD – WYSZUKIWANIE KLUCZA

- Węzeł wyszukuje w tablicy routingu węzły znajdujące się najbliżej

Źródło: Brunner R. A performance evaluation of the Kad-protocol.

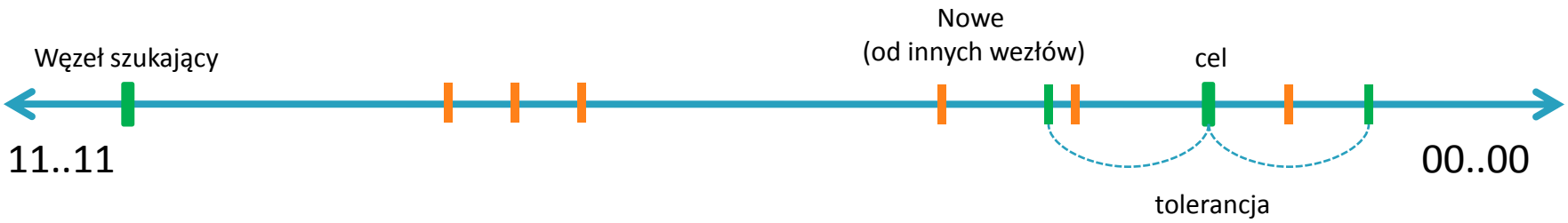
KAD – WYSZUKIWANIE KLUCZA



- Węzeł pyta wszystkich sąsiadów o dany klucz
- Zapytani odpowiadają nowymi kontaktami (zawsze bliżej celu niż oni)

Źródło: Brunner R. A performance evaluation of the Kad-protocol.

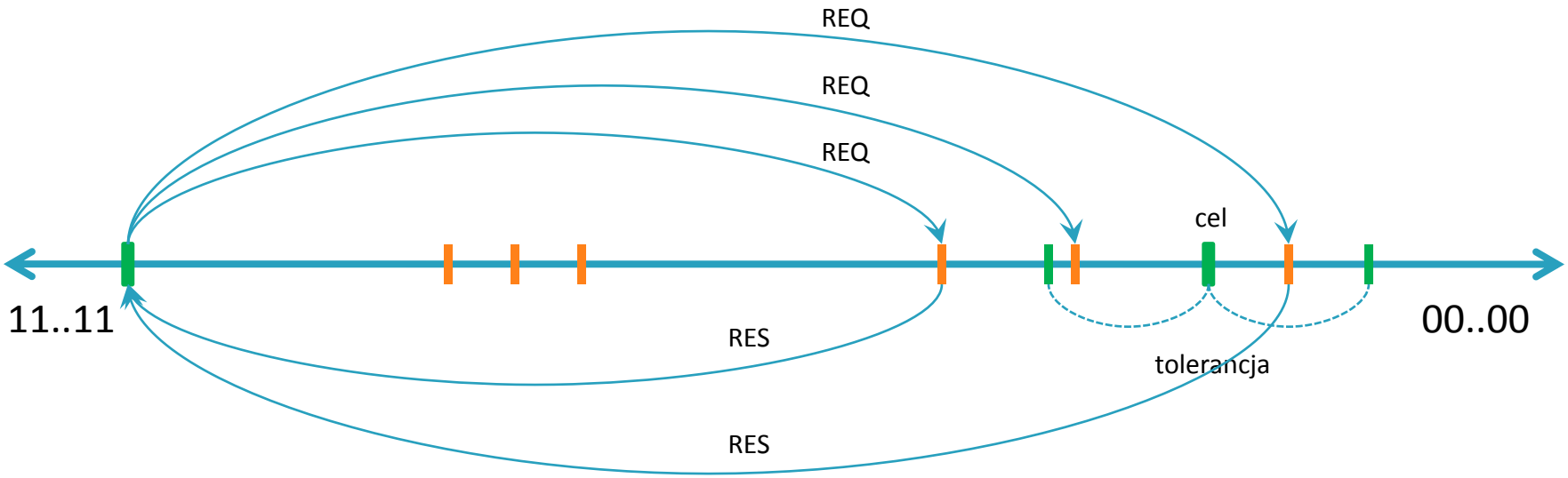
KAD – WYSZUKIWANIE KLUCZA



- Węzeł dostaje nowe kontakty

Źródło: Brunner R. A performance evaluation of the Kad-protocol.

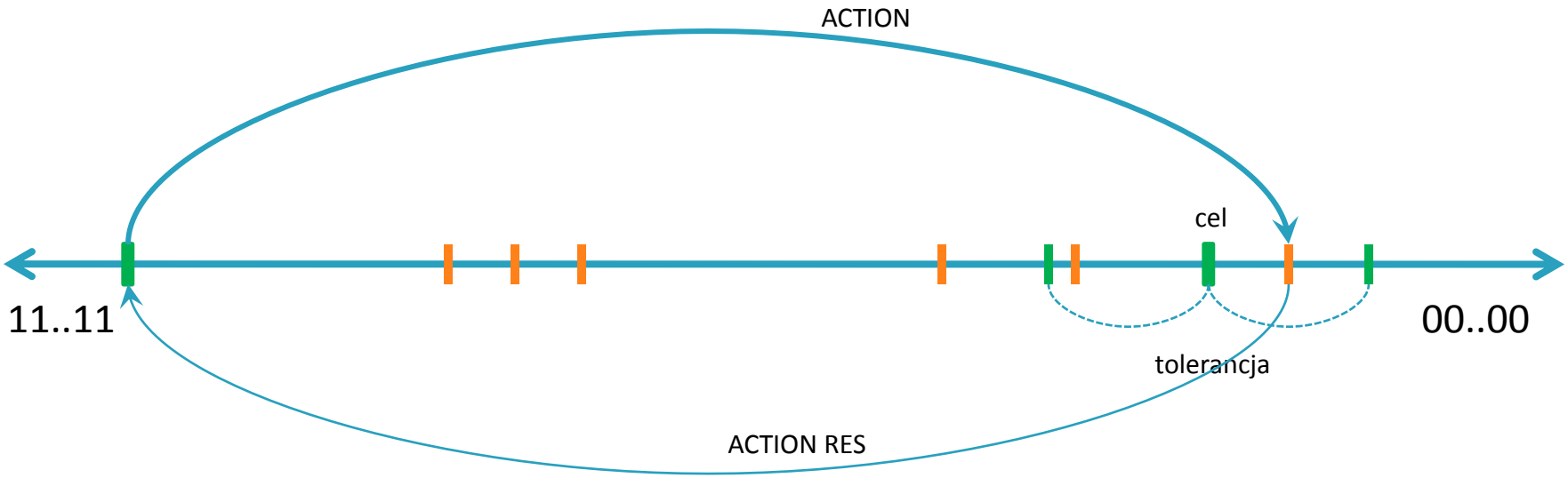
KAD – WYSZUKIWANIE KLUCZA



- Węzeł pyta nowe kontakty o klucz
- W najgorszym przypadku O(logn) takich kroków

Źródło: Brunner R. A performance evaluation of the Kad-protocol.

KAD – WYSZUKIWANIE KLUCZA



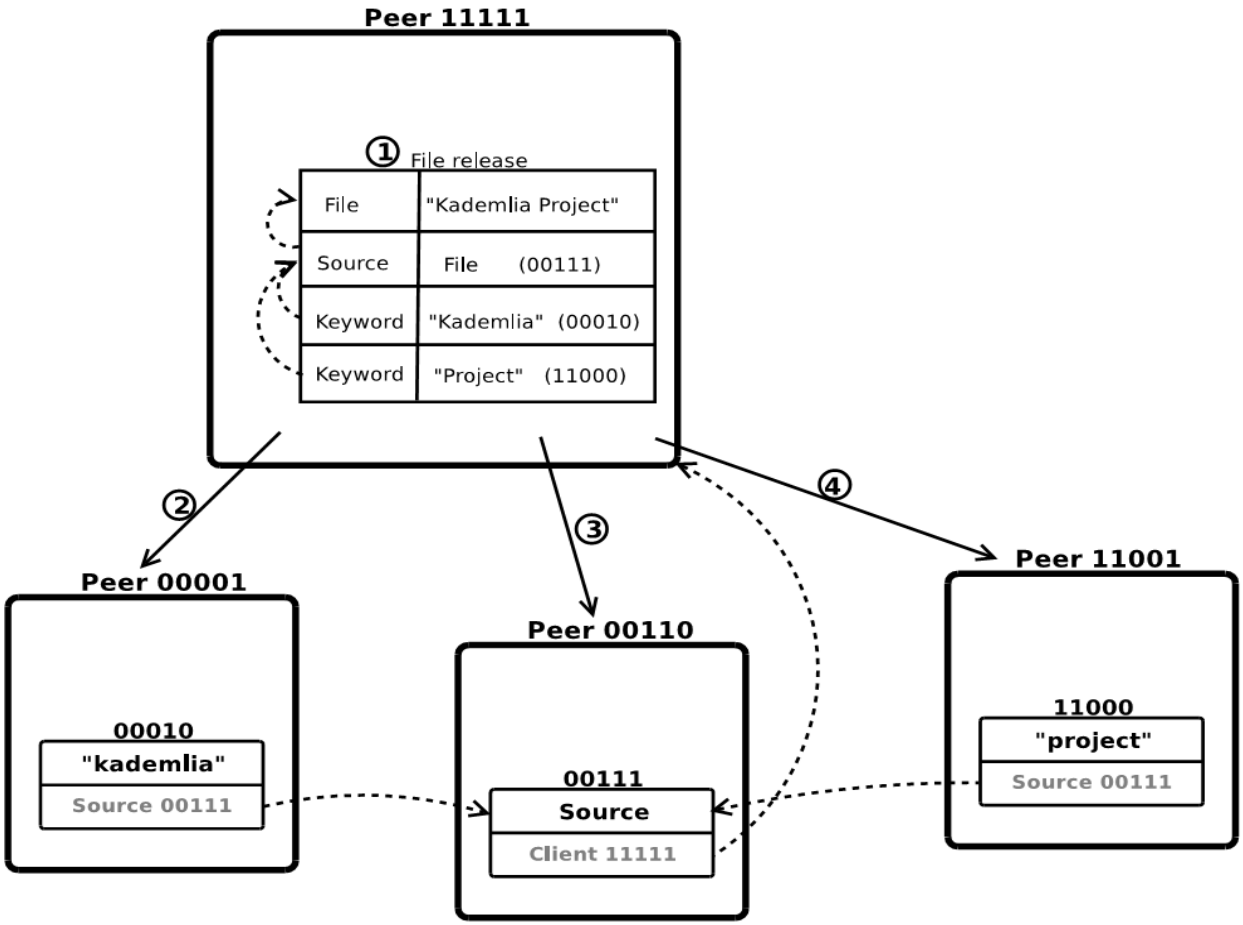
- Jeżeli kontakt jest podłączony wykonuje odpowiednią akcję (publikowanie/wyszukanie)

Źródło: Brunner R. A performance evaluation of the Kad-protocol.

KAD-PUBLIKACJA ZASOBU

- Informacje o pliku podzielone na dwie grupy
  - Metadane
    - nazwa
    - wielkość
    - itp.
  - Lokalizacja
    - IP węzła posiadającego plik
    - TCP port
    - itp.

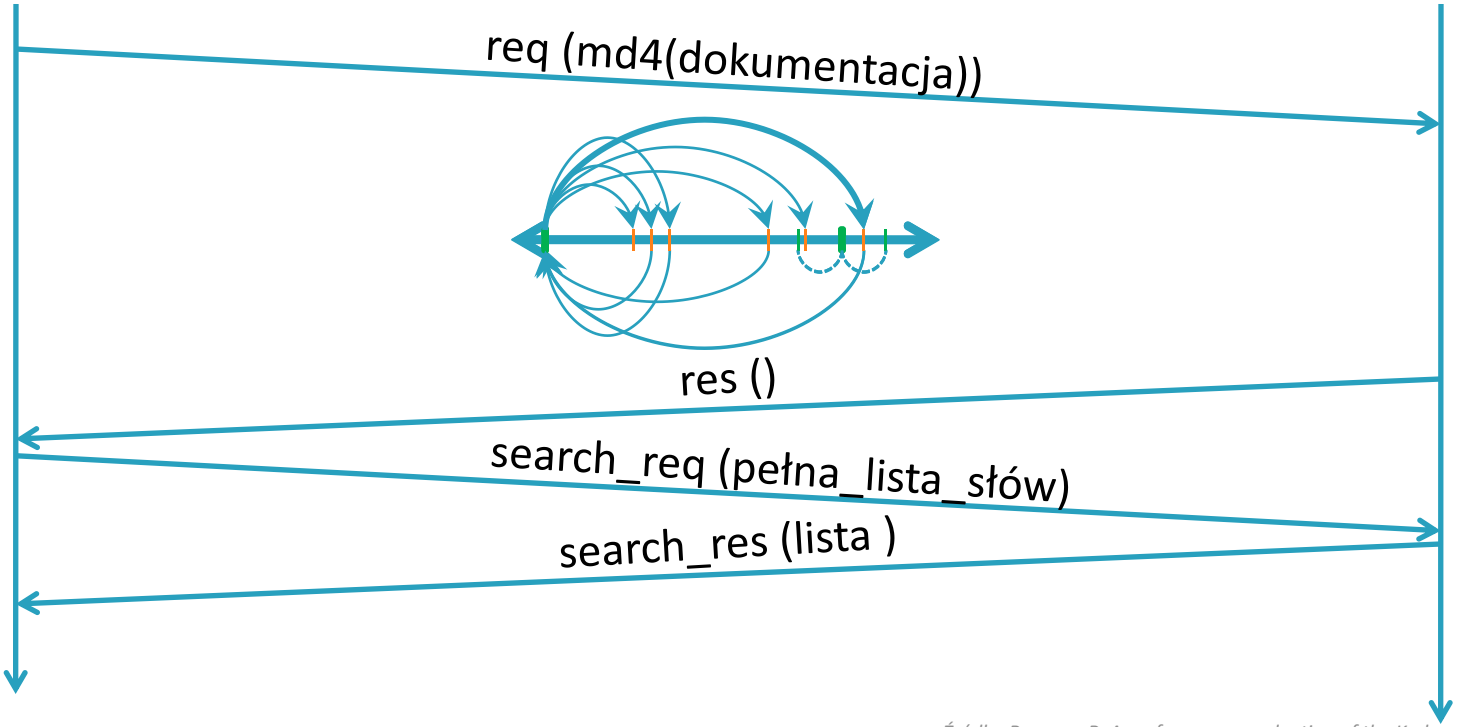
KAD-PUBLIKACJA ZASOBU



Źródło: Brunner R. A performance evaluation of the Kad-protocol.

KAD – WYSZUKIWANIE PLIKU

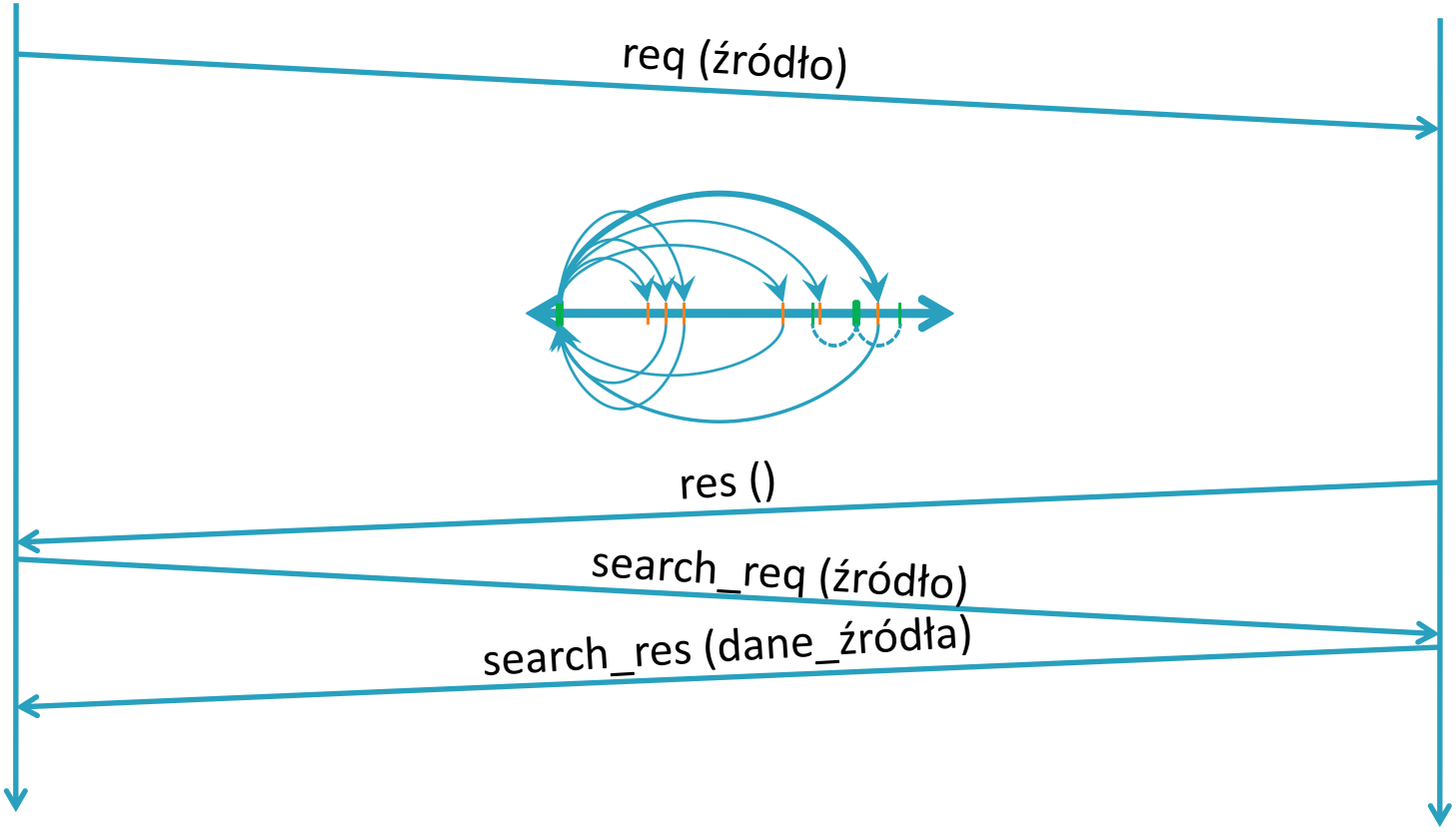
- Plik „Dokumentacja sieci KAD.pdf”
  - słowa kluczowe „dokumentacja” , „sieci”, „kad”



Źródło: Brunner R. A performance evaluation of the Kad-protocol.

KAD – WYSZUKIWANIE PLIKU

- Wyszukanie źródła



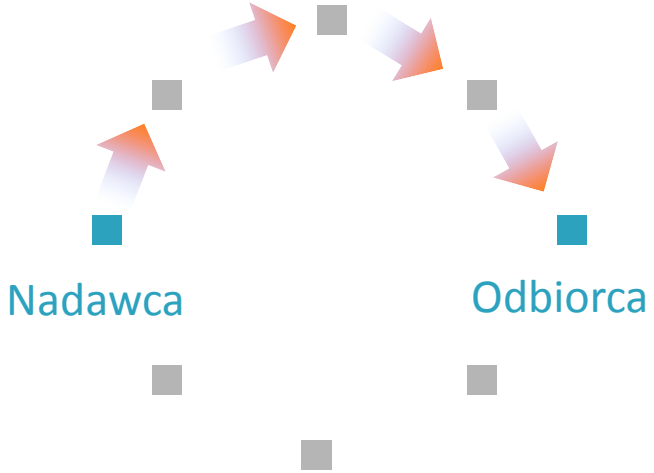
Źródło: Brunner R. A performance evaluation of the Kad-protocol.

ANONIMOWOŚĆ

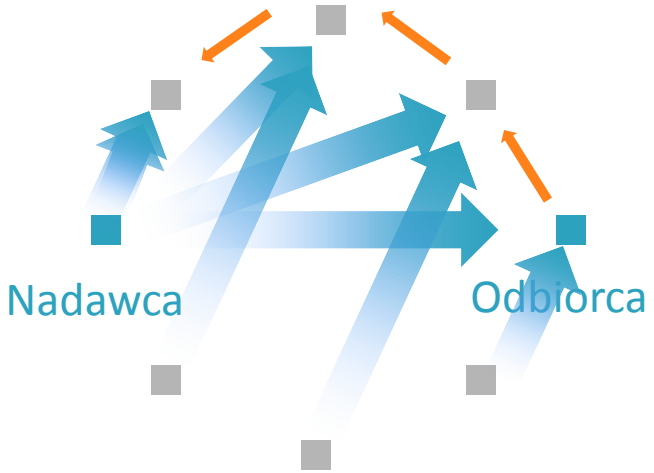
- Prywatność
  - może być zapewniana poprzez anonimowość
- Anonimowość
  - Anonimowość nadawcy
    - nadawca nie może zostać wskazany w zbiorze wszystkich nadawców
  - Anonimowość odbiorcy
    - odbiorca nie może zostać wskazany w zbiorze wszystkich odbiorców
  - Anonimowość przekazu
    - nie jest możliwe wskazanie związku między stronami

JAK DZIAŁA P2PRIV?

Tradycyjna sieć anonimizująca



P2PRIV

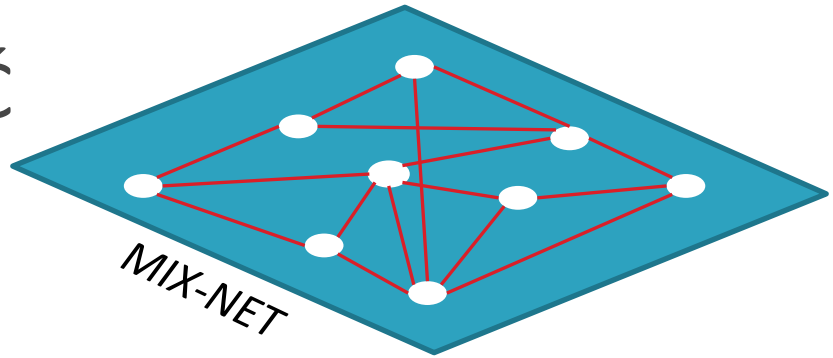


Źródło: www.p2priv.org

NA CZYM POLEGA RÓŻNICA?

- Mix-net
  - duże opóźnienia
  - duży narzut protokołu
  - mała przepustowość

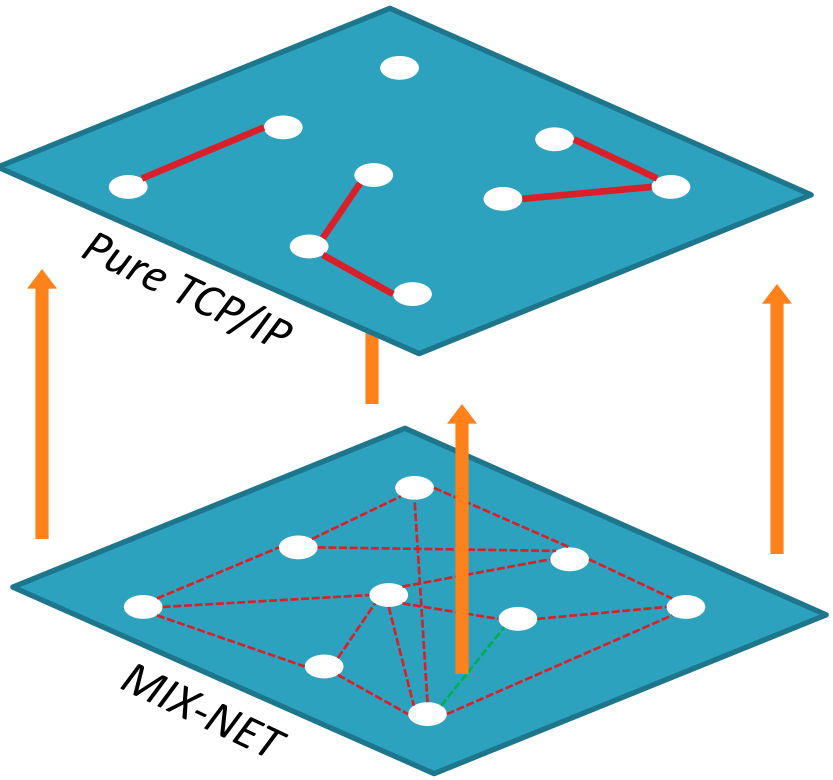
NISKA EFEKTYWNOŚĆ



NA CZYM POLEGA RÓŻNICA?

- Bezpośredni transport
  - małe opóźnienia
  - duża przepustowość

WYSOKA EFEKTYWNOŚĆ



JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
  - wynik: długa lista plików
  - do pobrania tylko kilka
- Wyszukiwanie źródeł
  - wynik: długa lista węzłów
  - do wybrania tylko kilka
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
  - wolniejsze
  - „bezpieczne”
- zapytanie via kaskada klonująca
  - szybsze
  - „mniej bezpieczne”

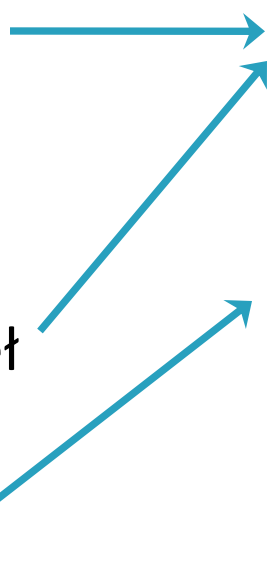
JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca



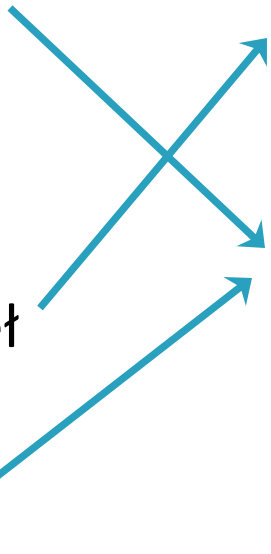
JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca



JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca

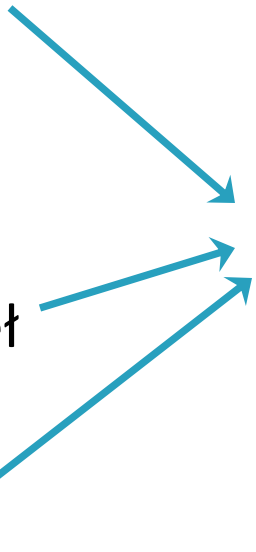
JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca



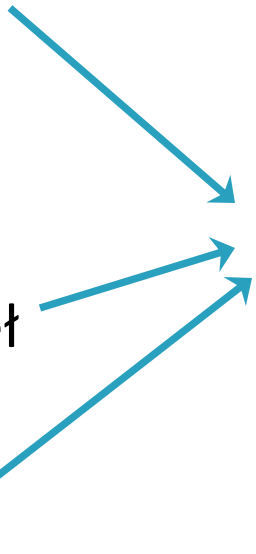
JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca



JAK POŁĄCZYĆ KAD Z P2PRIV?

KAD

- Wyszukiwanie słów kluczowych
- Wyszukiwanie źródeł
- Pobranie pliku

P2PRIV

- zapytanie via mix-net
- zapytanie via kaskada klonująca

Trzeba to policzyć!

MIARA ANONIMOWOŚCI

- Miara anonimowości
  - teoria informacji Shanonna
- Zmienna losowa

$$p_i = P(X = i)$$

prawdop. że *i*-ty węzeł jest inicjatorem

- Entropia

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

entropia systemu *N* węzłów, każdy z opowiednim prawdop. jest inicjatorem

MIARA ANONIMOWOŚCI

- Maksymalna entropia, gdy każdy z węzłów tak samo prawdop. jest inicjatorem

$$H_{max} = \log_2(N)$$

- Znormalizowana entropia

$$d = -\frac{\sum_{i=1}^{N} p_i \log_2(p_i)}{\log_2(N)}$$

- miara anonimowości

WNIOSKI

- Działają sieci w całości oparte na DHT – są skuteczne
- Możliwe jest praktyczne połączenie modelu P2PRIV z działającą siecią DHT
- Zaproponowane scenariusze mogą stanowić skuteczne metody anonimizacji użytkowników sieci służących do wymiany plików

PYTANIA?

Privacy has a *”ramified and intimate relation to the whole structure of human interaction and values, and to the nature of individual personality [..] If privacy changes much else will change.”*

A. Simmel

BIBLIOGRAFIA

- Brunner R. *A performance evaluation of the Kad-protocol*. Master Thesis at Institut Eurécom and Universität Mannheim. 2006
- Margasinski, I.; Pioro, M. *A Concept of an Anonymous Direct P2P Distribution Overlay System.* IEEE 22nd International Conference on Advanced Information Networking and Applications (AINA). 2008
- Diaz, C.; Seys, S.; Claessens, J. ;Preneel, B. *Towards measuring anonymity*. In the Proceedings of Privacy Enhancing Technologies Workshop (PET 2002). 2002
- Serjantov, A. ; Danezis, G. *Towards an Information Theoretic Metric for Anonymity*. In the Proceedings of Privacy Enhancing Technologies Workshop (PET 2002), April 2002.