An Algorithm to Arrange Hidden Content to Distinguish Between Good and Malicious Content in the Information Central Network (ICN)

Dr. Ahmad Saker Ahmad*
Dr. Talal Alataki**
Ahmad Akel***

(Received 7 / 7 / 2019. Accepted 10 / 9 / 2019)

□ ABSTRACT □

Named Data Network is an ICN (Information Central Network) architecture designed to be an Internet architecture that is designed to address some of the limitations caused by the use of IP.

NDNs use the drag form to distribute content where content is first explicitly requested before you deliver it, showing the effectiveness of this form by aggregating the conflicting requests for the same content and temporarily storing that content on the router. Although it reduces latency and increases bandwidth usage, router caching makes the network vulnerable to new attacks on the cache by injecting and counterfeiting malicious content to replace good content and placing it in the router cache to be a candidate for customer requests.

Our research aims to provide a mechanism to protect content hidden in the cache memory of routers so as to prevent malicious content poisoning attacks by malicious providers. This mechanism relies on the order of copying the content in a priority that increases and decreases according to users' opinions.

We have conducted a series of experiments on the NDNSIM emulator within the NS3 environment, and these experiments have shown that our algorithm is capable of preventing these attacks.

Keywords: Named Data Networks, Caching, Good And Malicious Content.

***Professor, Faculty of Informatics, Engineering, Tishreen University, Lattakia, Syria.**
****Assistant Professor, Faculty of Informatics, Engineering, Tishreen University, Lattakia, Syria.**
*****Postgraduate Student (PhD), Faculty of Informatics Engineering, Tishreen University, Lattakia, Syria.**

خوارزمية لترتيب المحتوى المخبأ للتمييز بين المحتوى الجيد والخبيث في شبكات المعلومات المركزية(ICN)

د. أحمد صقر أحمد*

د. طلال العاتكي**

أحمد عاقل***

(تاريخ الإيداع 7 / 7 / 2019. قُبِل للنشر في 10/ 9 / 2019)

□ ملخّص □

تعتبر شبكات البيانات المسمّاة NDN (Named Data Network) إحدى معماريات شبكات المعلومات المركزية ICN (Information Central Network) ، والتي صممت لتكون بنية انترنت مرشحة لمعالجة بعض القيود التي يسببها استخدام بروتوكول الانترنت IP .

تستخدم شبكات الـ NDN نموذج السّحب (Pull Mode) لتوزيع المحتوى، حيث يتم أولاً طلب المحتوى بشكل صريح قبل أن تقوم بتسليمه، تظهر فعالية هذا النّموذج من خلال تجميع الطّلبات المتباعدة لنفس المحتوى وتخزين هذا المحتوى بشكل مؤقت على الموجّه.

على الرّغم من أنّ التّخزين المؤقت في أجهزة التّوجيه يقلل زمن الوصول ويزيد من استخدام النّطاق الترددي إلا أنّه يجعل الشّبكة عرضة لهجمات جديدة على ذاكرة التّخزين المؤقت من خلال حقن محتويات خبيثة وتزييفها لتكون بديلاً للمحتويات الجيّدة، ووضعها ضمن ذاكرة التّخزين المؤقت للموجّه لتكون مرشحة للطّلبات القادمة من الزّبائن.

يهدف بحثنا إلى تقديم آليّة لحماية المحتوى المخبأ في الذّاكرة المؤقتة لأجهزة التّوجيه بحيث نمنع هجمات تسميم المحتوى بنسخ خبيثة من قبل المزودين السّيئين، تعتمد هذه الآلية على ترتيب نسخ المحتوى وفق أولويّة تزيد وتنقص حسب آراء المستخدمين.

قمنا بإجراء مجموعة من التّجارب على المحاكي NDNSIM ضمن بيئة الـ NS3 ، وقد بيّنت هذه التّجارب أنّ الخوارزمية المقترحة قادرة على منع هذه الهجمات.

الكلمات المفتاحية: شبكات البيانات المسمّاة، التخزين المؤقت، المحتوى الجيد والخبيث.

* **أستاذ – قسم النظم والشبكات الحاسوبية – كلية الهندسة المعلوماتية – جامعة تشرين – اللاذقية – سورية.**

** **مدرس – قسم النظم والشبكات الحاسوبية – كلية الهندسة المعلوماتية– جامعة تشرين – اللاذقية – سورية.**

*** **طالب دراسات عليا(دكتوراه)–قسم النظم والشبكات الحاسوبية–كلية الهندسة المعلوماتية–جامعة تشرين– اللاذقية – سورية.**

مقدمة:

تعتبر شبكات البيانات المسمّاة NDN واحدة من العديد من الجّهود البحثية التي تهدف الى تصميم بنية انترنت من الجّيل الحديث ضمن شبكات المعلومات المركزية ICN ، وهي على عكس الشّبكات التّقليدية القائمة على بروتوكول الانترنت IP، والتي تعيّن العناوين للمضيفين والوسطاء، فإن NDN يعالج المحتوى من خلال إعطائه اسماً قابلاً للتّسجيل ، حيث يصدر الزّبون اهتماماً بمحتوى معين ويطلبه بالاسم.

يمكن أن تلبي الشّبكة اهتماماً بالمحتوى من أي مضيف أو ذاكرة تخزين مؤقت للموجّه طالما يطابق اسم المحتوى ذلك الاهتمام، حيث تتبع شبكة المحتوى NDN بشكل عكسي المسار الدّقيق للمحتوى السّابق على طول الطّريق الى الزّبون، لضمان الموثوقية تنص الـ NDN على أن كل محتوى يجب أن يوقّع عليه مقدّم المحتوى ويطلب من الزّبون التّحقق من هذا التّوقيع، إلا أن التّحقق من توقيع المحتوى يكون اختيارياً لأجهزة التّوجيه وذلك نظراً للتّكاليف المرتبطة بذلك [1] .

لتسهيل التّوزيع الفّعال للمحتوى المطلوب بكثرة فإن الموجّهات في شبكات NDN تحافظ على ما يسمى بمخازن المحتوى حيث يتم تخبئة المحتويات ، تعتبر هذه المخازن ميزة NDN الأساسية لكونها تحسن الاستفادة من عرض النّطاق التّرددي للمحتويات المطلوبة بكثرة ، وعلى الرّغم من فوائدها الواضحة إلا أنّها تفتح الباب لهجمات المحتوى المزيف ، حيث يقوم أحد المزودين بالخدمات بضخ محتوى مزيف في مخابئ أجهزة التوجيه مما يؤدي الى توزيعه لاحقاً (وربما على نطاق واسع) للمستهلكين على الرغم من أن التّحقق من صحة التّوقيع من قبل المستهلكين يكشف المحتوى الغير صالح فإن الـ NDN ليس لديها وسيلة لإزالته من ذاكرة التخزين المؤقت بخلاف ذاكرة التخزين المؤقت العادية(على سبيل المثال LRU Based) [2].

الطّريقة الوحيدة التي تمكن المستهلك من تجنب الحصول على محتوى مزيّف هي عن طريق مرشّحات استبعاد صريحة (تحتوي على واحد أو أكثر من تجزيئات المحتوى الغير مرغوب به) في قائمة محتويات الـ NDN.

أهمية البحث وأهدافه:

1 – أهمية البحث

إن الموجّهات في شبكات NDN تقوم بتخبئة المحتويات ، تعتبر هذه المخازن ميزة NDN الأساسية لكونها تحسن الاستفادة من عرض النّطاق التّرددي للمحتويات المطلوبة بكثرة ، إلا أنّها معرضة لهجمات المحتوى المزيف ، حيث يقوم أحد المزودين بالخدمات بضخ محتوى مزيف في مخابئ أجهزة التوجيه مما يؤدي الى توزيعه لاحقاً (وربما على نطاق واسع) للمستهلكين ومن هنا تكمن أهمية بحثنا في إيجاد آلية ديناميكية لمنع هذه الهجمات وحماية المحتوى المخبأ.

2 – أهداف البحث

الهدف الرّئيسي من البحث هو حماية المحتوى الصالح المخبأ ضمن ذاكرة التّخزين المؤقت لأجهزة التّوجيه ومنع هجمات المحتوى الخبيث، وذلك من خلال تطبيق خوارزمية لترتيب المحتوى المخبأ الذي يسمح لأجهزة التّوجيه التّمييز بين محتوى جيد وخبيث.

يعتمد التّرتيب على الإحصائيات التي تم جمعها من إجراءات المستخدمين بعد تسليم كائنات المحتوى.

تظهر تجاربنا على NDNSIM أن هذه الخوارزمية تحقق فعاليّة ضد هذه الهجمات.

طرائق البحث ومواده:

1- المحاكي المستخدم:

تم اجراء التجارب لهذه الخوارزمية ضمن طوبولوجيا شبكية مختلفة من خلال المحاكات ضمن بيئة ال NDNSIM ، والتي تم تنصيبها ضمن محاكي NS3 على نظام Ubuntu.

2- السيناريوهات المقترحة:

- **السيناريو الأول:**

تم اجراء التجارب على شبكة ذات طوبولوجيا شجرية وقمنا بقياس عدد المستهلكين الجيدين الذين يمكنهم استرداد محتوى صالح ومدة ذلك، حيث عمدنا الى تسميم ذاكرات أجهزة التوجيه بنسخ مزيفة للمحتوى المستهدف.

- **السيناريو الثاني:**

قمنا بتغيير طوبولوجيا الشبكة الى شبكة DFN (Deutsches Forschungs Netz) وهي شبكة ألمانية تم تطويرها لأغراض البحث والتعليم أكثر تعقيداً من الشبكة السابقة تتكون الشبكة من 300 جهاز توجيه و 80 مستهلك ولقد أجرينا عليها مجموعتين من التجارب وذلك لنرى مدى فعالية الخوارزمية المقترحة عند زيادة تعقيد طوبولوجيا الشبكة المستخدمة.

1- آلية عمل شبكات ال NDN :

يتميّز اتصال NDN بنموذج السّحب والذي يقوم بتسليم المحتوى للزبون بناءً على طلب واضح منه.

هناك نوعين من الحزم في شبكات NDN هما: حزم الاهتمام (interest) وحزم المحتوى (content)، يقوم الزّبون بطلب المحتوى عن طريق إرسال حزم اهتمام، إذا كان أحد المزودين يمكنه تلبية هذا الاهتمام يقوم بالرّد بإرسال حزمة البيانات المناسبة، وكل حزمة بيانات تم تسليمها في شبكة الـ NDN يجب أن تتطابق بدقة مع طلب الزبون، مثلاً إذا كان المحتوى (C) المطلوبة بالاسم (n) تم استلامها من الموجّه ولم يكن هناك طلب سابق لهذا الاسم، يتم عزل المحتوى وتجاهله .

تتضمن حزم محتوى الـ NDNعدة بارامترات ومنها [3]:

- الاسم: سلسلة من عدة أسماء صريحة يتبعها اسماً ضمنياً (hash) للمحتوى يعاد حسابه في كل قفزة، وهذا يعطي فعالية لربط كل محتوى باسم فريد غير مكرر ويضمن التّطابق الدّقيق مع طلب الزبون.

في معظم الحالات يكون عنصر الـ (hash) غير موجود في طلب الزّبون ولا تملك NDN أي آلية آمنة لتعلم المحتوى الضّمني (Content hash).

- التّوقيع: يولّد مقدم المحتوى مفتاحاً عاماً ليغطّي كل مكونات المحتوى الصّريحة ويتضمن مرجع (بالاسم) الى المفتاح المطلوب للتّحقق.

- الصّلاحية للمحتوى (Freshness): الزّمن الموصى به من قبل مقدم المحتوى لكائنات المحتوى ليتم تخزينها بشكل مؤقّت. يجب أن يكون لكل منتج مفتاح عام واحد على الأقل يمثل مع المحتوى ما يسمى بكائن المحتوى موقّع من قبل مزوّد المحتوى.

تتضمن حزم الاهتمام Interest البارامترات التالية:

- الاسم: اسم المحتوى المطلوب.
- Min Suffix Component: الحد الأدنى لعدد مكونات الاسم.
- Max Suffix Component: الحد الأعلى لعدد مكونات الاسم.
- Exclude: تحوي معلومات عن مكونات الاسم التي يجب أن يتم استبعادها لكي لا تكون مرشحة كمحتوى معاد لاهتمام الزّبون.

2- كيانات NDN:

- المستهلك: الكيان الذي يصدر طلباً للمحتوى.
- المنتج: الكيان الذي يقدم وينشر المحتوى (بالإضافة للتّوقيع).
- الموجّه: يقوم بتوجيه حزم الطلب والمحتوى المناسب لها.

وكل من هذه الكيانات يحوي ما يلي [4]:

- مخزن المحتوى (Content Store(CS : ذاكرة تستخدم لتخزين واسترجاع المحتوى بشكل مؤقت.
- قاعدة توجيه طلب الاهتمام(Forwarding Interest Base(FIB: جدول بأسماء المحتويات المقدمة تستخدم لتوجيه الطّلبات.
- جدول الطّلبات المعلقة (Pending Interest Table (PIT: جدول يحتوي طلبات الاهتمام المعلقة للزّبائن والمحتوى المناسب لها.

عندما يستلم موجه طلب اهتمام لاسم محتوى (n) ولا يوجد محتوى مقابل في الـ PIT يقوم بتوجيه الطّلب إلى الموجّه التّالي وفقاً لمعلومات FIB ، وعند توجيه الطّلب إلى الموجّه يقوم بتخزين معلومات عن حالة الطّلب تتضمن اسم الطّلب ورقم المنفذ الواردة منه ، في حال ورود محتوى موافق لطلب الزّبون تقوم الموجّهات بالاحتفاظ بمعلومات عنه في PIT دون الاهتمام بمصدر المحتوى .

يتم تحديد حجم الذّاكرة المؤقتة المتوفرة محلياً، و يقدّم جهاز التّوجيه حجم الذّاكرة لديه لمعرفة إذا كان مناسباً للمحتوى.

3- المحتوى المزيّف:

نصف في هذا القسم سيناريوهات الهجوم بتسميم المحتوى وهنا نعرف بعض المصطلحات:

- المحتوى المزيّف (Fake Content) يتميز بإحدى الصّفات التّالية [5]:
  - حقل التّوقيع: تم إنشاؤه بطريقة خاطئة وبالتّالي أصبح المحتوى منسق بشكل سيئ.
  - التّوقيع صحيح : لكن تم إنشاؤه بمفتاح خاطئ بمعنى آخر المفتاح ليس مفتاح المنتج المزعوم.
  - التّوقيع غير صحيح: عندما تعطينا خوارزمية التّحقق من التّوقيع بأنه خاطئ.

وهنا نستخدم المصطلح Fake للإشارة أن كائنات المحتوى المقدَمة قد أدخلت من قبل مهاجم سيئ، يكون كائن المحتوى صالح إذا احتوى توقيعاً قابل للتّحقق تم تشكيله بالمفتاح العام الصحيح.

- المهاجم (Adversary (Adv: هو أي كيان من كيانات الـ NDN (يمكن أن يكون أكثر من واحد) قادر على حقن المحتوى في الشّبكة.
- تسميم المحتوى (Content Poisoning) : هجوم يقوم فيه المهاجم Adv بحقن المحتوى المزيّف في مخابئ الموجّهات.

نطبق في هذا البحث هجوم تسميم المحتوى ، حيث نقوم بحقن المحتوى المزيّف C الذَي يحمل الاسم n الى ذاكرة التّخزين المؤقت لجهاز التوجيه ، نستطيع حقن هذا المحتوى الخبيث إما عن طريق الموجّهات أو العقد الطّرفية، حيث أن المهاجم يمكن أن يكون إما مقدم خبيث للمنتجات malicious producer (Pm) أو مستهلك خبيث consisting of malicious consumer (Crm)، وعلى افتراض أن Pm و Crm متصلة بموجهات مختلفة، يرسل Crm طلب اهتمام للمحتوى n وبمجرد أن يقوم الموجّه باستقبال الطّلب فيقوم ال Pm بإرسال المحتوى المزيّف للموجّه فيقوم الموجّه بتخزينه بالذّاكرة المؤقتة، وبالتّالي يصبح الموجّه ملوّث بمحتوى مزيّف ومعتقداً بأنه يملك المحتوى المناسب لطلب الزبائن ولزيادة مدّة الهجوم يقوم برفع حقل الصّلاحية للمنتج (Freshness) إلى أعلى قيمة.

4- ترتيب المحتوى

من حيث المبدأ إن منع التّسمم للمحتوى ليس صعباً إذا تحقق كل جهاز توجيه من توقيع كل محتوى قبل تقديمه أو تخزينه في الذّاكرة المؤقتة، لكن المشكلة تكمن في أن عملية التّحقق من التّوقيعات للمحتوى تتطلب آلية لجلب المفاتيح العامة والتّحقق منها لمنح الثّقة للتّطبيقات وهي عملية صعبة كثيراً ومكلفة، وهناك طريقة أخرى تقتضي فرض استخدام أسماء محددة للمحتويات جرى التّحقق منها سابقاً وهي عملية سهلة يقوم فيها الموجّه بتقديم دليل بأسماء المحتويات ليختار المستهلك منها الاسم المطلوب .

تحوي هذه الأدلّة أسماء المحتويات ونوع المحتوى المتعلق بتلك الأسماء ، إلا أنّه هناك مشكلة تتلخص بضرورة حصول المستهلك على هذه الأدلّة قبل طلب المحتوى وكذلك بالنّسبة للمحتويات الدّيناميكية التي تتغير أسماؤها باستمرار فإن هذه الطّريقة صعبة التحقيق [6].

في هذا البحث نقدّم خوارزمية ترتيب لكائنات المحتويات المخزنة مؤقتاً هدفها التّمييز الاحتمالي بين المحتوى الصّحيح والمزيّف استناداً إلى سلوك المستهلك وإعطاء أولويّة للمحتوى الصّالح أعلى من المحتوى المزيّف لترشيحه كاستجابة لطلب الزّبائن.

تقوم فكرة الخوارزمية على حقيقة أن المستهلكين للمحتوى الذين قاموا بالتّحقق من التّوقيعات واكتشاف المحتوى المزيّف سيقومون بإصدار اهتمام جديد للمحتوى يستثني المحتوى المزيّف الذي تم استلامه سابقاً ، إنَ تحليل معلومات الاستبعاد يمكننا من السّماح لأجهزة التّوجيه بترتيب المحتوى المخزن مؤقتاً عن طريق إعطاء مرتبة أعلى للمحتوى الصّالح من المحتوى المزيّف ، يكون بعد ذلك المحتوى الذي يملك أعلى ترتيب هو المرشح ليشكل استجابة لاهتمام المستهلكين وتكون قيم التّرتيب متدرجة بين (0) و (1) ، يعطى المحتوى الجديد الوارد الى الذّاكرة المؤقتة قيمة (1) ليمنح فرصة أن يكون استجابة لاهتمام المستهلكين ومع مرور الوقت تنخفض هذه القيمة تدريجياً حسب رأي المستهلكين ، كما قد تتغير هذه القيمة لأن المحتوى قد لا يلبي احتياجات المستخدمين وليس بالضرورة أن يكون المحتوى مزيّف ، بذلك نكون أعطينا الفرصة للمحتويات الحديثة ، أما المحتويات التي تصبح قيمتها (0) يتم استبعادها نظراً لعدم قبول المستخدمين بها سواءً أكانت مزيّفة أو لا تلبي متطلبات المستهلكين واحتياجاتهم.

4-1 الاستبعاد:

يعتمد هذا المعيار على عدد المرّات التي يتم فيها استبعاد كائن المحتوى c المعطى اسماً n ، كما قد تكون هناك إصدارات مختلفة من الكائن c تحتوي بيانات مختلفة ، الأمر الذي يؤدي إلى توقيعات مختلفة على الرّغم من أنها تحمل نفس الاسم n .

للتّمييز بين هذه الإصدارات نرمز الى كل إصدار باسم njH(c) ، يشير معدل الاستبعاد RnjH(c) إلى نسبة عدد الاستبعادات لـ njH(c) ويرمَز بالرّمز EnjH(c) ، مقسوماً على العدد الإجمالي للطّلبات للمحتوى C ويرمَز بالرّمز Qn بحيث [7]:

$$RnjH(c) = EnjH(c) / Qn$$

إنّ تغيير النّسبة أكثر أهمية من عدد الاستبعادات ، السّبب هو أن المستهلكين لا يقومون باستبعاد المحتوى المزيّف فقط وإنما المحتويات غير الصّحيحة أيضاً ، حيث تقوم فرضيتنا على أساس أن الكائنات للمحتوى المزيّف تميل الى الاستبعاد بدرجة معدل أكبر ، ويمكننا نمذجة كائنات المحتوى التي يتم استبعادها من خلال المعادلة:

$$r_{n|H(c)}(t) = e^{\frac{-t}{a}} \qquad (1)$$

حيث أن: t تمثل زمن n|H(c) و ($t \in [0, t_{t0}]$) حيث أن t_{t0} هي الزّمن الموصى به لتخزين المحتوى في ذاكرة التّخزين المؤقت و a: عامل يمثّل معدّل تدهور سرعة المحتوى.

نلاحظ أن $r_{n|H(c)} \in [r_{n|H(c)}(t0),1]$ أعلى قيمة يأخذها العامل a تمثل تغيّر لمعدّل تدهور سرعة المحتوى حتى نستطيع معرفة المعامل a .

لتكن r_{t0} هي معدّل التّخزين بالذّاكرة المؤقتة عند تخزين المحتوى $n|H(c)$ ، عندما تنتهي الصّلاحية(الزّمن الموصى به من قبل مزوّد المنتج Freshness) بشرط ألّا يتم استبعاد $n|H(c)$ من الذّاكرة خلال هذا الوقت وهي تحدد من قبل مدير الشّبكة.

$t0$: هي القيمة التي تجعل المحتوى غير مستبعد من الذّاكرة المؤقتة وبالتالي تكون قيمتها كبيرة للكائنات الأقل استبعاداً من الذّاكرة ، وكلما زاد عدد الاستبعادات للمحتوى $n|H(c)$ تجعل معامل التّدهور أكثر وضوحاً ، واذا لم يتم استبعاد المحتوى أبداً تأخذ نسبة التّدهور القيمة الأبطأ ضمن المجال $[1 , r_{t0}]$ ، بناءً على ذلك فإن a يعتمد على قيمة a_{t0} ويتأثر بشكل سلبي بقيمة $R_{n|H(c)}$:

$$a = a_{t0} - (R_{n|H(c)} * a_{t0})$$

وبالتالي يمكن إعادة صياغة المعادلة (1) لتصبح:

$$r_{n|H(c)}(t) = \frac{-t}{a_{t0} - (R_{n|H(c)} * a_{t0})} \qquad (2)$$

4-2 التّرتيب الموزع للاستبعادات:

لدينا عامل في خوارزمية التّرتيب هو زمن الاستبعاد، والهدف منه إعطاء مزيد من الوقت للاستبعاد الأحدث.

نعرف تأثير الاستبعاد $i_{n|H(c)}$ كمتغير يعكس قيمة الوقت الذي يجب أن ينتظره الموجّه قبل التّأثر بمحاولة استبعاد للمحتوى $n|H(c)$ ، والذي تم تعيينه بالحد الأدنى خلال حساب التّرتيب، ونستطيع حسابه من خلال العلاقة:

$$i_{n|H(c)}(te) = 1 - e^{\frac{-te}{\beta}} \qquad (3)$$

حيث أن te هي المدة الزّمنية المنقضية منذ آخر استبعاد.

β : عامل يحدد نمط التأثير فهو يعكس مدى السّرعة التي يتم بها تأثير أخر استبعاد بالحد الأدنى ، نلاحظ ان $i_{n|H(c)}(te) \in [0 , 1]$ حيث أن $i_{n|H(c)}(te) = 1$ تعني أن آخر استبعاد له الحد الأدنى من التّأثير في التّرتيب.

كلما كانت β أكبر يعني انه يجب أن يقضي المزيد من الوقت قبل الوصول للحد الأدنى والمرشح لآخر استبعاد ، هذا الوقت يرمز له بـ t_{mw} إعطاء قيمة لـ t_{mw} يتم من قبل مدير الشّبكة .

وعندما تكون $i_{n|H(c)}(te)=1$ فإن قيمة β المقابلة يمكن حسابها من قبل المعادلة (3) ونعطيها الرّمز β_{mw} ، وبذلك نستطيع تعديل المعادلة (2) لتشمل معامل التأثير للاستبعاد كالتالي:

$$r_{n|H(c)}(t)=\frac{-t}{i_{n|H(c)}(te)*[at0-(R_{n|H(c)}*at0)]} \qquad (4)$$

إذا كان 1= $i_{n|H(c)}(te)$ فإن معدّل التّرتيب لـ $n|H(c)$ يتأثر بمعدّل الاستبعاد $R_{n|H(c)}$.

3-4 نسبة منافذ الاستبعاد

يعتمد هذا المعيار على عدد المنافذ التي وصل إليها استبعاد للمحتوى، إذا كانت قيمته عالية فإن ذلك يشير إلى عدم رضا أو قبول للمحتوى من قبل الزّبائن ، يمكن استغلال هذا المعيار لتقليل ترتيب هذا المحتوى كعقاب له ، ويأخذ بعين الاعتبار الأمور التالية [8]:

fn : إجمالي عدد المنافذ الذي يقدمها جهاز توجيه معين .

$fe \in [0\ ,\ fn]$: عدد المنافذ التي جاء اليها محتوى يستبعد المحتوى السّابق $n|H(c)$.

$fs \in [1\ ,\ fn]$: عدد المنافذ على الموجّه التي قامت سابقاً بتخديم المحتوى $n|H(c)$.

(نلاحظ أن قيمة fs يجب ألا تساوي 0 باعتباره موجود ضمن التّرتيب ، أي أن المحتوى تم طلبه و أعطي نتيجة للطلب على الأقل من قبل منفذ واحد).

$e_{n|H(c)}$: نسبة عدد المنافذ في الموجّه قامت بتخديم المحتوى $n|H(c)$ و لم يصل له أي استبعاد الى fs :

$$e_{n|H(c)}=\begin{cases}\frac{fs-fe}{fs} & if\ fs\geq fe\\ 1 & Otherwise\end{cases} \qquad (5)$$

$e_{n|H(c)} \in [0\ ,\ 1]$ وعندما تأخذ القيمة (1) لن يتم استبعاد المحتوى $e_{n|H(c)}$ على الاطلاق.

يمكن ل fe أن تتجاوز fs ، أي يمكن لموجه أن يستقبل طلب اهتمام بمحتوى لكنه يستبعد $n|H(c)$ وذلك على منفذ لم يقوم بتخديم هذا المحتوى سابقاً وهذا يحدث لثلاثة أسباب :

1- تغييرات التّوجيه.
2- تنقل الزّبائن (مركزية المستهلك).
3- استبدال ذاكرة التّخزين المؤقت.

إن السبب الأول والثّاني واضحين أمّا الثّالث يشير الى الحالة التي يكون فيها المحتوى قد تم طلبه سابقاً وتم تخديمه وتخزينه بشكل مؤقت ثم تم مسحه من ذاكرة التّخزين المؤقت لأي سبب، بما فيها انتهاء المدة المسموحة لبقائه ، واستناداً للتّعريف السّابق يمكن إعادة صياغة المعادلة (4) لتصبح:

$$r_{n|H(c)}(t)=\frac{-t}{e^{e_{n|H(c)}*i_{n|H(c)}(te)*[ato-(R_{n|H(c)}*ato)]}} \qquad (6)$$

تعكس المعادلة (6) مرتبة كل كائن محتوى مخزن في ذاكرة التّخزين المؤقت في كل موجه ، ويستند هذا التّرتيب إلى ثلاثة معايير :(عدد الاستبعادات – الزّمن الموزع لمحاولة الاستبعاد – نسبة الاستبعاد على المنافذ)

يبين الجّدول (1) مقارنة بين أنماط تدهور التّرتيب لخمس كائنات محتوى مخزنة مؤقتاً، المعاملات في المعادلة(6) تختلف من كائن محتوى لآخر وهي ملخصة في الجّدول(1)كالتالي[9]:

الجَدول (1) معاملات كائنات المحتوى

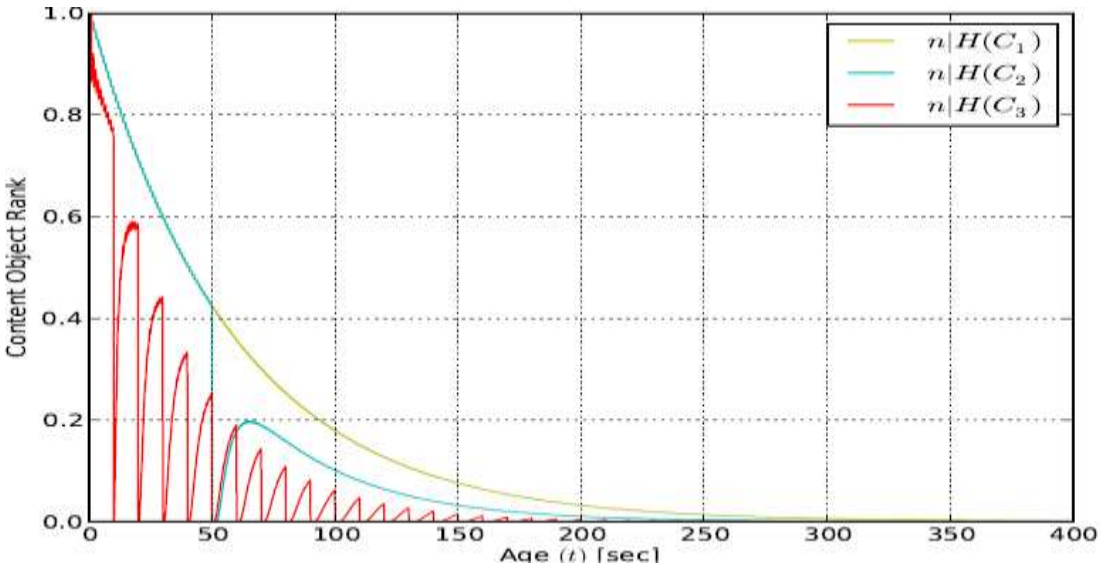
Parameter	$n\|H(c1)$	$n\|H(c2)$	$n\|H(c3)$	$n\|H(c4)$	$n\|H(c5)$
Content	C1	C2	C3	C4	C5
Name	N	N	N	N	N
Digest	H(C1)	H(C2)	H(C3)	H(C4)	H(C5)
T	[0,400], one sample every 100 [m sec]				
Freshness	400	400	400	400	400
r_{t0}	0.001	0.001	0.001	0.001	0.001
Qn	1	1 when t ∈ [0,50] and 2 when t ∈ [50,400]	Increased by 1 every 10 sec		
$e_{n\|H(c)}$	0	0 when t ∈ [0,50] and 1 when t ∈ [50,400]	Increased by 1 every 10 sec		
t_{mw}	400	400	400	400	400
te	تقارب إلى ∞	∞when t∈ [0,50] and Increased by 1 every 1 sec when t > 50	[0.10] Increased by 1 every 1 sec And reset every 10 sec		
fn	4	4	4	4	4
fe	0	0when t∈ [0,50] and 1 when t ∈ [50,400]	1	2	3

1- كائن المحتوى $n|H(c1)$ تم طلبه مرة واحدة فقط ولم يتم استبعاده أبداً طوال عمره في ذاكرة التّخزين المؤقت في جهاز التّوجيه.

2- كائن المحتوى $n|H(c2)$ تم طلبه مرة واحدة فقط وفي الثانية (50) تم استبعادها مرة واحدة.

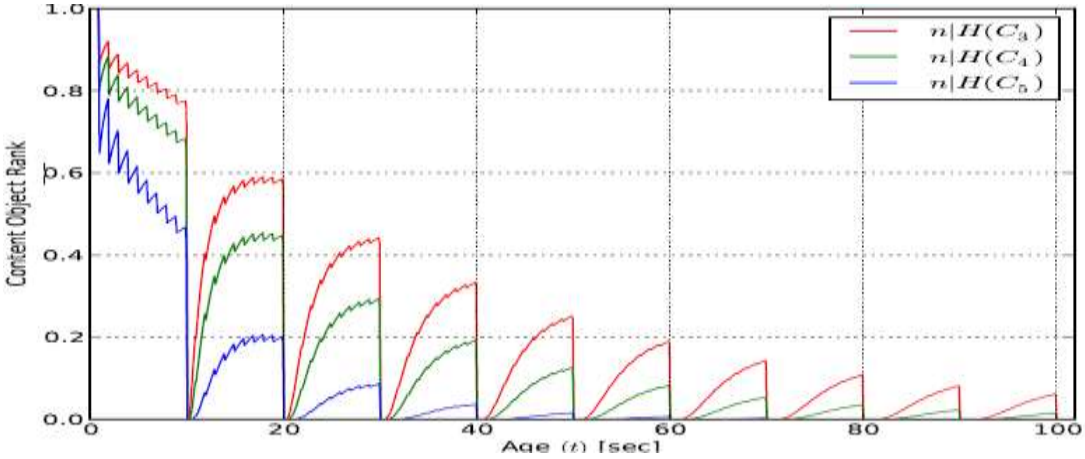
3- كائنات المحتوى ($n|H(c5)$ - $n|H(c4)$ - $n|H(c3)$) طلبت مرة واحدة وتم استبعادها كل (10) ثانية.

الفرق بين هذه الكائنات الثّلاثة هو نسبة استبعاد التوجيه ، بناءً على المعادلة(6) نفترض أن $n|H(c1)$ يجب أن يكون له ترتيب أعلى في جميع الأوقات متبوعاً بـ $n|H(c2)$ ثم الكائنات ($n|H(c5)$- $n|H(c4)$ - $n|H(c3)$)، يوضح الشكلان (1) و (2) أنماط التّدهور لكائنات المحتوى [10].



الشّكل (1) أنماط التدهور المتغيرة لكائنات المحتوى (njH(C1) ، njH(C2)، njH(C3)).

يبين الشكل الشكل (1) أنه حتى الثانية 50 يكون لكل من الكائنات ($n|H(c2) - n|H(c1)$) قيم متساوية وهي أعلى من الكائنات الثّلاثة الأخرى المستبعدة ثم نلاحظ أنه ينخفض ترتيب $n|H(c2)$ بعد استبعاده عند الثانية 50 ، كما نلاحظ أن النّمط التّكراري لـ $n|H(c3)$ تم توضيحه بسبب استبعاد كائن المحتوى كل 10 ثانية وبمجرد حدوث الاستبعاد ينخفض التّرتيب ليقترب من (0) ويبدأ في الزّيادة مرة أخرى حسب المعادلة (3) . من ناحية أخرى يقارن الشكل (2) بين كائنات المحتوى ($n|H(c5) - n|H(c4) - n|H(c3)$) في مدة زمنية أقصر من (100) ثانية نلاحظ أن تأثير fe مختلف ، على سبيل المثال يتم استبعاد $n|H(c5)$ على (3) منافذ توجيه مختلفة بينما يتم استبعاد $n|H(c3)$ على منفذ واحد فقط وبالتالي فإن $n|H(c5)$ لديه قيمة ترتيب أقل .



الشّكل (2) أنماط التدهور المتغيرة لكائنات المحتوى (njH(C5) ، njH(C4)، njH(C3)

بناءً على تحليل خوارزمية ترتيب المحتوى نستنتج أن كائنات المحتوى الأحدث لها ترتيب أعلى من تلك القديمة وهذه ميزة تصميم معتمد لمنح أولوية للكائنات الجديدة وفرصة لنشرها.

النتائج والمناقشة:

NDNSIM:هو تنفيذ بسيط لبنية NDN كوحدة نمطية لأغراض المحاكاة للتّحقق من صحّة وفعالية خوارزمية ترتيب المحتوى المقترحة ، حيث قمنا بتضمين هذه الخوارزمية في مخابئ التوجيه وأجرينا التّجارب ضمن بيئة الـ NS3 وقبل البدء بشرح التّجارب نعرف بعض المصطلحات [11]:

المستهلكين الجيدين: لن يكونوا راضين إذا أعاد الاهتمام الذي طلبوه محتوى مزيّف ، حيث يقومون باستبعاده في اهتماماتهم اللاحقة لنفس الاسم، ويمكن أن يتوقف الزّبون عن إرسال الاهتمام بعد تلقي محتوى صالح.

المستهلكين الخبيثين: يتصرفون بطريقة معاكسة إذا عرضت إحدى الرّسائل المعادة لهم محتوى صالح يقومون باستبعاده في جميع الاهتمامات اللاحقة لنفس الاسم ، والهدف من ذلك هو تغيير الإحصائيات التي تم جمعها من الاستبعادات وذلك لتفضيل المحتوى المزيّف.

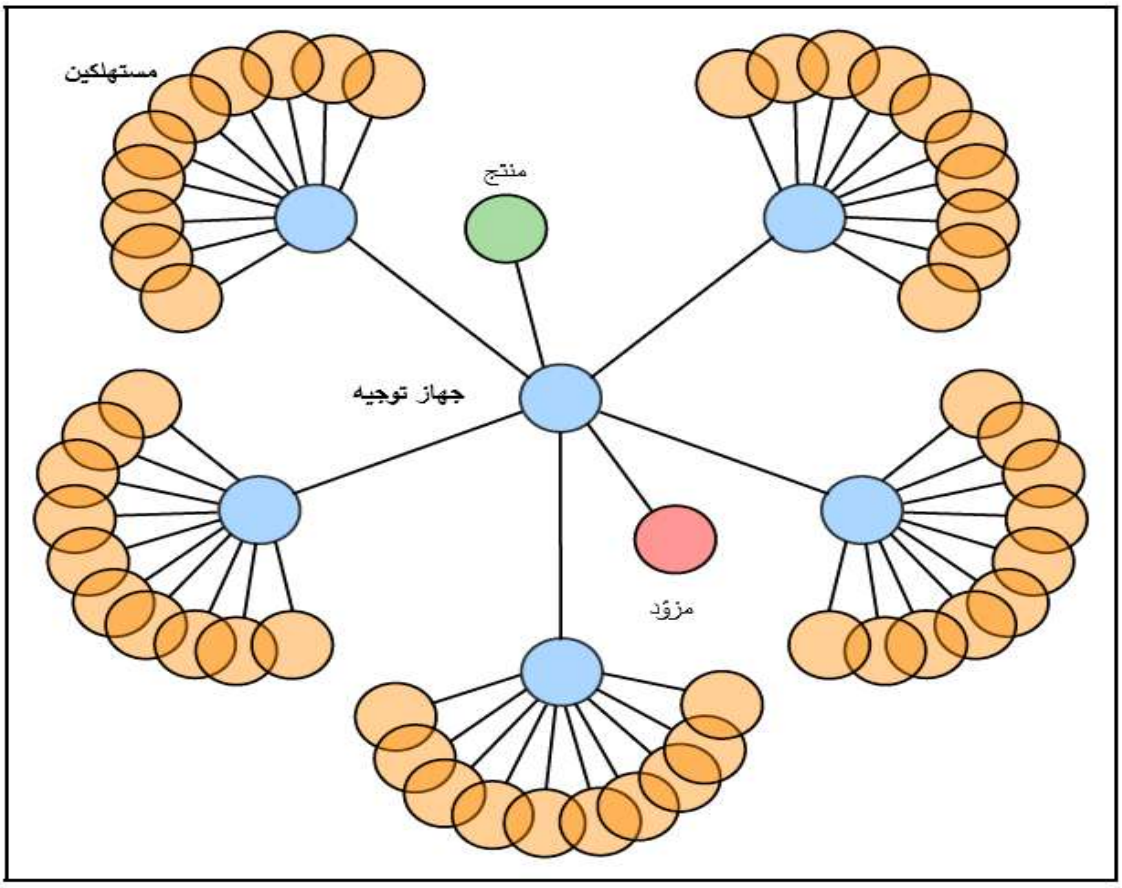
نحن حاولنا في تجاربنا قياس عدد المستهلكين الجيدين الذين يمكنهم استرداد محتوى صالح والمدة اللازمة لذلك، حيث قمنا بتسميم ذاكرات أجهزة التوجيه بنسخ مزيّفة للمحتوى المستهدف وضمن طوبولوجيا شبكية مختلفة كما يوضح الجدول (2) [12].

الجدول (2) بارامترات المحاكاة

Parameter	Tree-based Topology	DFN Topology	
Consumers	50	80	80
Routers	6	30	30
Producers	0	0	0
Cache Replacement Policy	LRU	LRU	LRU
Simulation Time [sec]	400	400	400
Pre-Populated fake Content Object Rate	99.9%	80% - 90%- 99% and 99.9%	99% and 99.9%
Pre-Populated fake Content Object Freshness [sec]	400	400	400
Malicious Consumer Rate	0%,2%,4%,6% and 10%	0%	0%,1%,3%,5% and 10%
Interest Interval [millisecond]	[100,300]	[100,300]	[100,300]

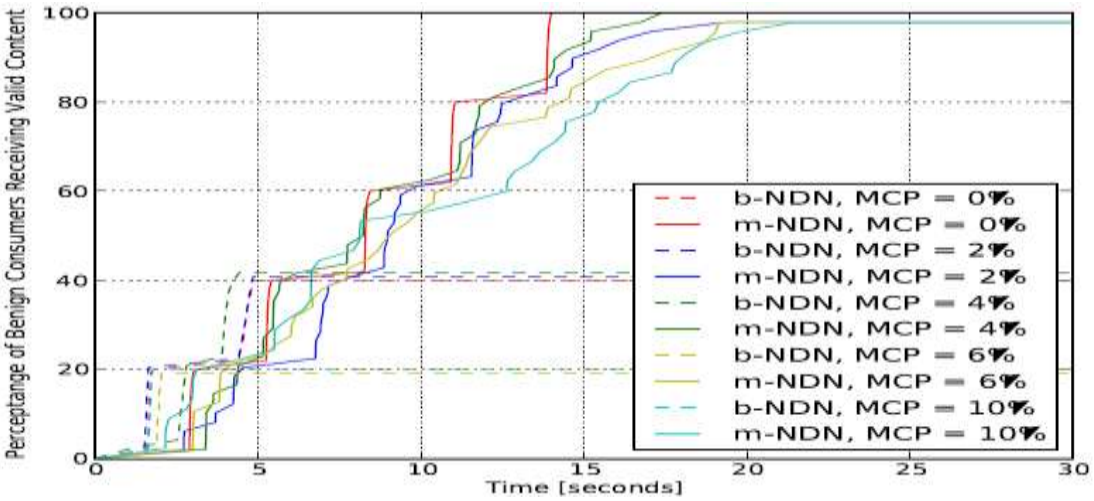
1- طوبولوجيا الشّبكة الشّجرية (Tree-based Topology):

يوضح الشّكل (3) الطّوبولوجيا ذات البنية الشجرية وتتكون من 5 أجهزة توجيه (كل منها يتصل بـ 10 مستهلكين) متصلة بجهاز توجيه مركزي وتحتوي على منتج واحد ومزود واحد متصل بجهاز التّوجيه المركزي [13].



الشّكل (3) طوبولوجيا شبكية شجرية

قمنا في البداية بتعبئة الذّاكرات المخبئة لأجهزة التوجيه بـ 1000 نسخة مختلفة لنفس المحتوى واحد منها صالح ، بمعنى أن معدل كائنات المحتوى المزيّف السّابق بنسبة 99.9% بالإضافة الى ذلك نطبق معدلات مختلفة من 50 مستهلك على أنها خبيثة (0% ، 2 % ، 4% ، 6 % ، 10%).

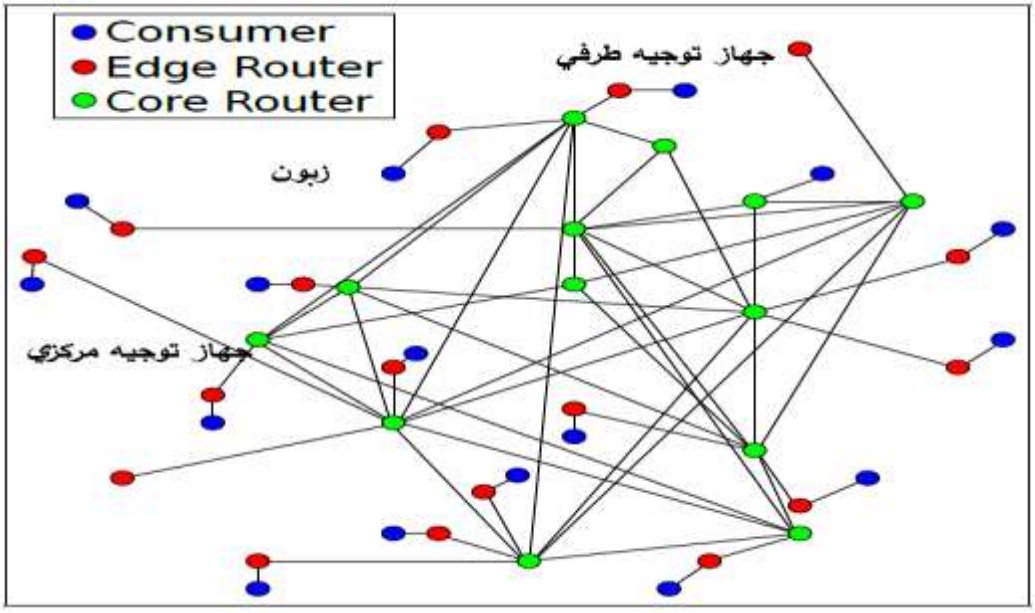


الشّكل (4) طوبولوجيا شبكية شجرية مع معدلات استهلاك نسخ خبيثة مختلفة.

قمنا بتطبيق معدلات استهلاك لنسخ خبيثة مختلفة (b–NDN : NDN الأساسية مع استبدال ذاكرة LRU مؤقتًا ، m–NDN : NDN المعدلّة مع أجهزة التوجيه التي تنفذ خوارزمية الترتيب ، MCP: النسبة المئوية للعقد الخبيثة).
يوضح الشّكل (4) نتائج التّجربة ويمكننا ملاحظة أنه في الحالات التي لا يتم فيها استخدام الترتيب لم تكن الشبكة قادرة على الوصول الى حالة يتلقى فيها جميع المستهلكين الضعفاء محتوى صالح ، وعندما يتلقى جميع المستهلكين الضعفاء محتوى صالح نقول أن الشبكة تتقارب الى حالة تسمى **بـ (التّقارب الكامل)** والسبب في عدم قدرة الشبكة على الوصول للتقارب الكامل بسرعة هو أن المستهلكين يمكنهم فقط استبعاد عدد معين من كائنات المحتوى المزيّف وهو (100) في تجربتنا، وبذلك فإن نسبة المستهلكين التي تم محاكاتها تؤثر على الترتيب وتؤخره إلا أنا خوارزميتنا تؤدي الى توجيه الشبكة للوصول لحالة التقارب الكامل على الرغم من أن ذلك يستغرق وقت أطول بالنسبة لمعدلات المستهلكين الخبيثين العالية.

2– طوبولوجيا الشّبكة DFN (Deutsches Forschungs Netz):

بعد التأكّد من السّلوك الصّحيح لخوارزمية التّرتيب باستخدام الطوبولوجيا الشّجرية نستخدم شبكة أكثر تعقيداً مثل شبكة DFN : وهي شبكة ألمانية تم تطويرها لأغراض البحث والتعليم تتكون من عدة أجهزة توجيه متصلة في مناطق مختلفة كما هو موضح بالشكل (5) يتم توصيل كل موجِّه طرفي بـ 5 مستهلكين ضمن شبكة NDN [14] وذلك لنرى مدى فعالية الخوارزمية المقترحة عند زيادة تعقيد طوبولوجيا الشبكة المستخدمة.

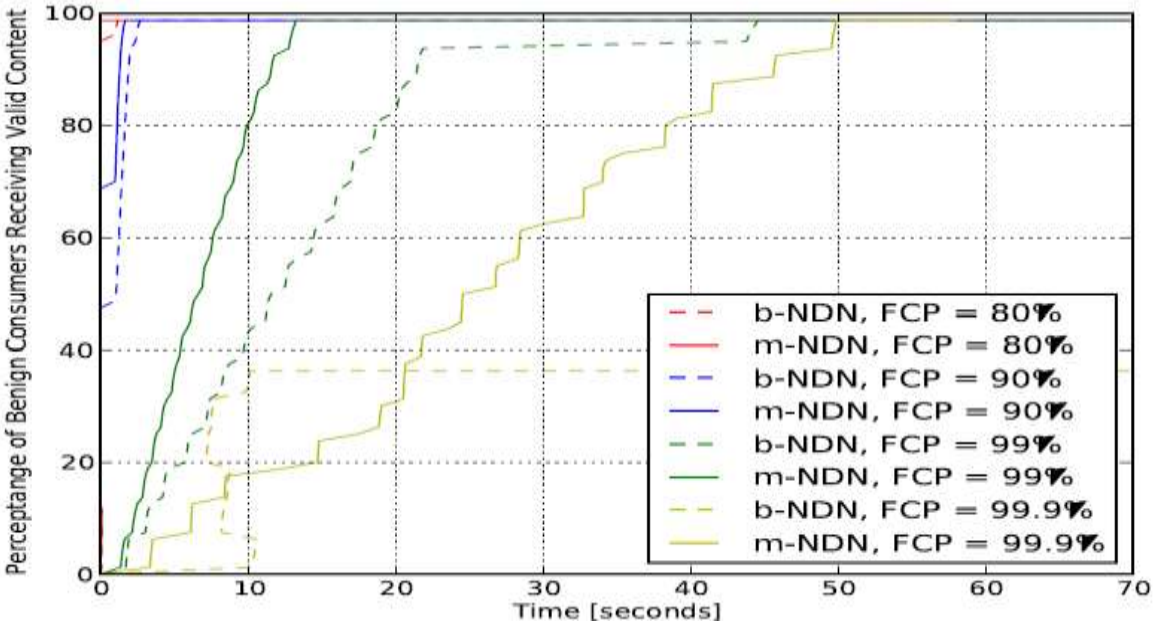


الشّكل (5) طوبولوجيا – DFN

تتكون الشبكة من 300 جهاز توجيه و 80 مستهلك ، أجرينا مجموعتين من التجارب باستخدام طوبولوجيا DFN.
يتم توجيه جميع أجهزة التوجيه في الشبكة مسبقاً بمعدلات مختلفة من كائنات المحتوى المزيّف :

1– 80% (1 كائن صالح و 4 كائنات مزيّفة).
2– 90% (1 كائن صالح و 9 كائنات مزيّفة).
3– 99% (1 كائن صالح و 99 كائنات مزيّفة).
4– 99.9% (1 كائن صالح و 999 كائنات مزيّفة).

يبين الشكل (6) نتائج التجربة:

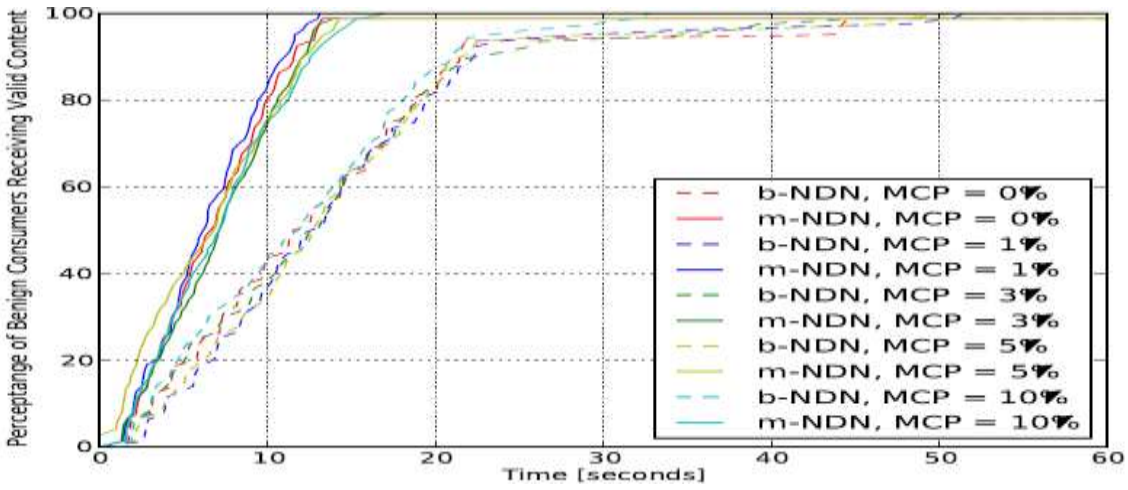


الشّكل (6) طوبولوجيا DFN بمعدلات مختلفة من كائنات المحتوى المزيف.

قمنا بتطبيق معدلات مختلفة من كائنات المحتوى المزيّف (b–NDN : NDN الأساسي مع استبدال LRU مؤقت ، m–NDN : NDNالمعدلة مع أجهزة التوجيه التي تنفذ خوارزمية الترتيب ، FCP : النسبة المئوية لكائنات المحتوى المزيّف).

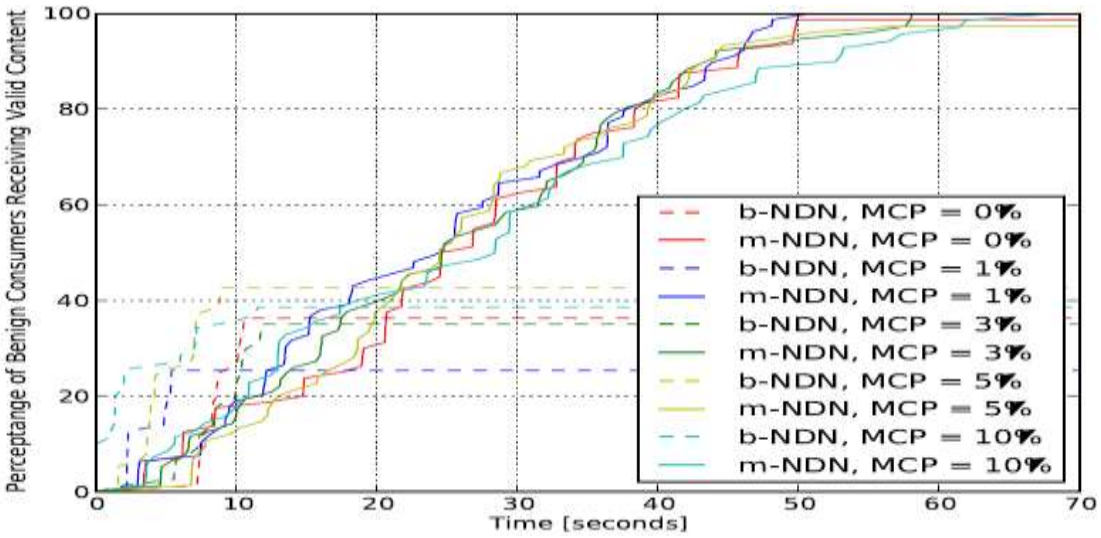
نلاحظ أن الشّبكة تصل الى التّقارب الكامل دائماً باستثناء ما إذا كان التّرتيب لا يتم تطبيقه على معدل كائنات المحتوى المزيّف الذي تم إعداده مسبقاً بنسبة 99.9 % والسّبب يعود الى حجم الاستبعاد.

في هذه التّجربة نقوم بتغيير معدل المستهلكين الخبيثين (0% ، 1% ، 3% ، 5% ، 10%) وذلك لمعدلين من كائنات المحتوى المزيّف (99% ، 99.9%) يبين الشّكلين (7) و (8) النّتائج التي تم الحصول عليها:



الشّكل (7) طبولوجيا DFN بمعدلات مختلفة من المستهلكين الخبيثين و99 % من كائنات المحتوى المزيّف.

قمنا بتطبيق معدلات مختلفة من المستهلكين الخبيثين و99 % من كائنات المحتوى المزيّف (b–NDN : NDN الأساسي مع استبدال LRU للذاكرة ، m–NDN: NDN المعدّلة مع أجهزة التوجيه التي تنفذ خوارزمية ترتيبنا ، MCP : النسبة المئوية للعقد الخبيثة عند المستهلكين) ، يمكننا في الشّكل (7) ملاحظة أن خوارزمية التّرتيب تسمح بالتّقارب الكامل بشكل أسرع .



الشكل(8) نتائج طبولوجيا DFN بمعدلات مختلفة من المستهلكين الخبيثين و99.9% من كائنات المحتوى المزيَف

قمنا بعد ذلك بزيادة نسبة كائنات المحتوى المزيّف لتصل لـ 99.9 % مع معدلات مختلفة من المستهلكين (b–NDN : NDN الأساسي مع استبدال LRU للذاكرة ، m–NDN: NDN المعدلة مع أجهزة التوجيه التي تنفذ خوارزمية ترتيبنا ، MCP : النسبة المئوية للعقد الخبيثة عند المستهلكين).

إن خوارزمية الترتيب المقترحة تعمل على تحسين قدرة الشّبكة على مواجهة هجمات تسميم المحتوى حتى إذا تم تسميم جميع مخابئ أجهزة التوجيه وضمن طوبولوجيا كبيرة نسبياً.

الاستنتاجات والتوصيات:

الاستنتاجات:

- إن شبكات الـ NDN هي واحدة من عدة شبكات مرشحة للجيل القادم في هندسة الانترنت.
- على الرغم من العديد من ميزات الأمان المضمنة، فإنه لا يزال عرضة لبعض التّهديدات الجديدة، مثل التّسمم بالمحتوى، حيث يقوم مزوّد المحتوى الخبيث بضخ محتوى مزيّف في ذاكرة التّخزين المؤقت لجهاز التّوجيه. يتم تقديم هذه الأشياء لاحقًا للمستهلكين استجابةً لحزم الاهتمام.

التوصيات:

- في هذا البحث، اقترحنا خوارزمية ترتيب المحتوى لاكتشاف وتخفيف هجمات التسمم بالمحتوى في شبكة الـ NDN وهذه الخوارزمية تعتمد على إجراءات المستهلك عند تلقي محتوى مزيّف. تجمع عقد الـ NDN إحصائيات حول كائنات المحتوى المستبعدة وتقوم بتعيين قيمة رقمية لكل عنصر تم تخزينه مؤقتًا وترتيبها، يتم تحديد الكائن المخزن مؤقتًا الأعلى ترتيباً كمرشّح لتلبية جميع طلبات المستهلكين اللاحقة.

- إن نتائج التجارب تدعم تأكيدنا على ضرورة استخدام خوارزمية الترتيب المقترح لأنّها تكتشف وتخفف من هجمات التسمم بالمحتوى.

المراجع:

[1] B. AHLGREN, C. DANNEWITZ, C. IMBRENDA, D. KUTSCHER, and B. OHLMAN,"*A survey of information-centric networking*," Communications Magazine,IEEE, vol. 50, no. 7, pp. 26–36, 2018.

[2] V. JACOBSON, D. K. SMETTERS, J. D. THORNTON, M. F. PLASS, N. H. BRIGGS, and R. L. BRAYNARD, "*Networking named content*," in Proceedings Of the 5th International Conference on Emerging Networking Experiments and Technologies. ACM, 2017, pp. 1–12.

[3] P. GASTI, G. TSUDIK, E. UZUN, and L.ZHANG, "*DoS & DDoS in Nameddata networking*," in Proceedings of the 22nd International Conference on Computing Communications and Networks. IEEE, 2017.

[4] A. CHAABANE, E. De CRISTOFARO, M. A. KAAFAR, and E. UZUN, "*Privacy in content-oriented networking: Threats and countermeasures*," SIGCOMM Comput. Commun. Rev., vol. 43, no. 3, pp. 25–33, Jul. 2018.

[5] S. DIBENEDETTO, P. GASTI, G. TSUDIK, and E. UZUN, "*Andana: Anonymous named data networking application*," in NDSS.The Internet Society,2017.

[6] G. ACS, M. CONTI, P. GASTI, C. GHALI, and G. TSUDIK, "*Cache privacy in named-data networking*," in Proceedings of the 33rd International Conference on Distributed Computing Systems. IEEE, 2018.

[7] L. ZHANG, D. ESTRIN, J. BURKE, V. JACOBSON, J. D. THORNTON, D. K. SMETTERS, B. ZHANG, G. TSUDIK, D. MASSEY, C. PAPADOPOULOS., "*Named data networking (ndn) project*," Relat´orio T´ecnico NDN-0001,Xerox Palo Alto Research Center-PARC, 2017.

[8] A. AFANAYEV, I. MOISEENKO, and L. ZHANG, "*ndnsim: NDN simulator For NS-3*," University of California, Los Angeles, Technical Report, 2017.

[9] A. COMPAGNO, M. CONTI, P. GASTI, and G. TSUDIK, "*Poseidon:Mitigating interest flooding DDoS attacks in named data networking*," arXivpreprint arXiv:1303.4823, 2018.

[10] C. E. LEISERSON, R. L. RIVEST, C. STEIN, and T. H. CORMEN, Introduction to algorithms. The MIT press, 2016.

[11] M. D. ATKINSON, J.-R. Sack, N. SANTORO, and T. STROTHOTTE, "*Min-max heaps and generalized priority queues*," Communications of the ACM, vol. 29, no. 10, pp. 996–1000, 2013.

[12] A. AFANASYEV, P. MAHADEVAN, I. MOISEENKO, E. UZUN, and L. ZHANG, "Interest flooding attack and countermeasures in named data networking," in Proceedings of the IFIP Networking Conference, 2016.

[13] S. Y. NAM, D. KIM, AND J. KIM, "Enhanced ARP: preventing ARPpoisoning-based man-in-the-middle attacks," Communications Letters, IEEE, vol. 14, no. 2, pp. 187–189, 2015.

[14] Z. TRABELSI AND W. El-HAJJ, "Preventing ARP attacks using a Fuzzybased stateful ARP cache," in Proceedings of the IEEE International Conference on Communications. IEEE, 2017, pp. 1355–1360.