Information Resilience through User-Assisted Caching in Disruptive Content-Centric Networks

Vasilis Sourlas
UCL, UK.
v.sourlas@ucl.ac.uk

Leandros Tassiulas
Yale, USA.
leandros.tassiulas@yale.edu

Ioannis Psaras
UCL, UK.
i.psaras@ucl.ac.uk

George Pavlou
UCL, UK.
g.pavlou@ucl.ac.uk

***Abstract*—We investigate an information-resilience scheme in the context of Content-Centric Networks (CCN) for the retrieval of content in disruptive, fragmented networks cases. To resolve and fetch content when the origin is not available due to fragmentation, we exploit content cached both in in-network caches and in end-users' devices. Initially, we present the required modifications in the CCN architecture to support the proposed resilience scheme. We also present the family of policies that enable the retrieval of cached content and we derive an analytical expression/lower bound of the probability that an information item will disappear from the network (be absorbed) and the time to absorption when the origin of the item is not reachable. Extensive simulations indicate that the proposed resilience scheme is a valid tool for the retrieval of cached content in disruptive scenarios, since it allows the retrieval of content for a long period after the fragmentation of the network and the "disappearance" of the content origin.**

***Index Terms*—Content-Centric Networks, in-network caching, disaster scenarios, Markov processes.**

I. Introduction

Content-Centric Networking, a.k.a. Information-Centric Networking (ICN), is emerging as one of the main future networking environments, given that the vast majority of Internet activities are related to information retrieval, while the location of the requested information is of less importance. Several ICN architectures [1] have been proposed for allowing information access and delivery based on network location-independent names, instead of endpoint addresses. We use the abbreviations CCN/NDN to refer to the main architecture proposed in [2], whereas the term ICN is used to describe the generic information-centric architectural paradigm. Stemming from its named-data networking nature, ICN has the potential of flexibly adapting to emerging service requirements through its native support to caching, mobility and multicast. In ICN each information item can now be uniquely identified and authenticated without being associated to a specific host, making in-network caching one of the salient characteristics of information-centric architectures. Every cache enabled node of the network caches every item that traverses it, and hence may serve future matching interests, avoiding going all the way to the content origin/server.

According to this line of thought, research has recently focused on the optimization of the in-network caching system performance in order to improve efficiency and reduce delivery delay. In this paper, we take a slightly different stance. In particular, we investigate the potential of the in-network caching system to prolong information/content lifetime through maintaining content in caches and serving it from there, when fragmentation happens and the origin server is not reachable. For instance, in a dynamic/disruptive environment, such as the aftermath of a disaster scenario like hurricane, earthquake, tsunami or a human-generated network breakdown, both users and content servers may dynamically join and leave the network (due to mobility or network fragmentation). Thus, users might join the network and request content when the network is fragmented and the corresponding content origin is not reachable. In such a scenario it is impossible for a user to retrieve content that match his/her interest.

In this work, we propose a simple, but novel and efficient scheme for realizing an in-network caching mechanism, whose focus is to preserve content over time. In other words, we do not only cache content to improve response time, but also to make information available to future users, since no persistent network connectivity or content origin availability is assumed.

Such an information resilience scheme is useful in networks set after a disaster scenario, where the network infrastructure is partially (temporal or spatial) available and the reachability of the content origin is not guaranteed. One of the main technical challenges according to the IETF ICNRG working group [3], regarding the usage of ICN is disaster scenarios, is to enable the usage of functional parts of the infrastructure, even when these are disconnected from the rest of the network. We also assume as in [3], that parts of the network infrastructure are functional after a disaster has taken place and it is desirable to be able to continue using such components for communication as much as possible. Unfortunately, this is especially difficult for today's mobile and fixed networks which are comprised of a centralized architecture, mandating connectivity to central entities for communication. The desired functionality of an ICN set in a disaster field is to allow the delivery of messages between relatives and friends, enable the spreading of crucial information to citizens and enable crucial content from legal authorities to reach all users in time. In such a deployment, where the content origin is not always reachable, a user can take advantage of similar interests issued by neighboring users and their cached content in order to retrieve the requested data.

In this paper we:

- Enhance the CCN/NDN [2] router architecture with a

ISBN 978-3-901882-68-5

new component called "*Satisfied Interest Table*" (*SIT*), so that users can retrieve cached content when the network is fragmented and the content origin is temporarily not reachable.
- Enhance the Interest packet forwarding mechanism of the CCN/NDN model so that interests could be forwarded to neighboring users with similar interests, upon the fragmentation of the network.
- Decompose the proposed information resilience scheme in a set of basic policies/strategies and examine various combinations of the proposed policies.
- Provide an analytical expression with the usage of continuous time, discrete state, Markov processes for the computation of the probability that an information item will disappear (be absorbed) from the network and the corresponding time to absorption upon the "disappearance" of the content origin, when the caching capacity of each router is equal to zero. These expressions help as the corresponding lower bounds for each of the examined caching mechanisms.
- Validate and evaluate the proposed resilience scheme and the corresponding caching mechanisms through extensive simulations for various system parameters.

The rest of the paper is organized as follows. In Section II we survey related work, whereas in Section III we present the functionality of the proposed information resilience scheme and the necessary augmentations to the original CCN/NDN router design to support it. Section IV, is devoted in the policies/strategies that compose the various caching mechanisms for the retrieval of cached content. In Section V we derive the analytical expressions for the absorption probability and the time to absorption of an information item. Finally, in Section VI we evaluate through simulations the performance of the proposed information resilience scheme, while we conclude the paper in Section VII.

II. Related Work

In-network packet-level caching in CCN/NDN has been extensively studied from different aspects, such as content popularity estimation or criteria for determining the probability of performing local caching. In general, CCN/NDN enables caching of addressable information items, in every cache-equipped node. However, this cache-everything-everywhere scheme [2] has already raised doubts and some authors have already questioned this aggressive strategy [4]. In that direction, a plethora of caching algorithms have been proposed according to which a NDN router that lies on the delivery path of an item decides, based on a probability, whether or not to cache passing-by content [5]. Also, various graph-based metrics have been examined for deciding the nodes to which caching should take place (e.g. [6]) , where concepts such as the betweenness centrality of the node is used for caching only in a subset of nodes to achieve better performance in terms of cache and server hit ratios.

Every research attempt regarding in-network caching in CCN/NDN, takes as granted the presence of the content origin for each information item and caching is used to boost the overall performance of the network. Furthermore, all the aforementioned research attempts did not explore the caching capabilities of the users of the network and the possibility of exploiting them to further improve content retrieval. Only in [7] authors propose the "*user-assisted in-network caching*" scheme, where similarly to our approach users who request, download, and keep the content may be able to contribute to in-network caching by sharing their downloaded content with other users in the same network domain. In contrast to our approach though, the work in [7] also assumes the presence of the content origin (uninterrupted connectivity) and user-assisted caching is used to improve network performance in terms of cache hit ratio and not as an information resilience scheme in disruptive scenarios.

In the area of ICN usage in disruptive scenarios the GreenICN EU-Japan project [8] is exploiting the ICN architectural paradigm to support the aftermath of a disaster. Particularly, a major part of the project's vision/objective is *"the aftermath of a disaster e.g., hurricane or tsunami, when communication resources are at a premium and it is critical to efficiently distribute disaster notification and critical rescue information. Key to this is the ability to exploit fragmented networks with only intermittent connectivity"*. Additionally, in [9] authors developed a distributed serverless social networking service based on NDN for sharing information among users before and after a disaster. Moreover, in [10] authors presented the "*Name-based Replication*" (NREP) system for scope-based prioritization of ICN messages in disasters, where ICN messages have attributes such as user-defined priority, space, and temporal-validity. These attributes are then taken into account when prioritizing messages. This system is orthogonal to our approach here and can be used as an alternative to determine information items' popularity.

Finally, in [11] authors propose an information resilience scheme for the PURSUIT architecture [12]. Particularly, they introduce a resilience management function that supports both link failure detection and recovery. Nodes publish periodically link state notifications and by their delivery or not, the network can detect link removals or additions. Also, upon the detection of a link failure the proposed resilience function identifies if any delivery tree was affected by the failure and establishes a new tree for any broken one. This work also assumes the presence of the content origin/publisher and aims at the re-establishment of the connectivity between the users and the origin. To the best of our knowledge, the work presented here is the first attempt that the inherent caching capabilities of the NDN routers and their attached users are exploited to support content retrieval in disruptive scenarios, where the network is fragmented and the content origin is not always reachable.

III. CCN/NDN Router Design and Packet Processing

In this section, we present the functionality of the proposed information resilience scheme. The rationale behind our design

is to increase information resilience, by leveraging the in-network caching capabilities of the network (and its attached users) and benefit from the name-oriented routing to retrieve content when the network is fragmented and the content origin is temporarily unreachable from parts or the entirety of the network.

A. CCN/NDN router

We augment the original NDN content router design presented in [2], with the *Satisfied Interest Table (SIT)*, as shown in Fig. 1. The functionality of the other router components, namely the Content Store (CS), the Pending Interest Table (PIT) and the Forwarding Information Base (FIB) remain the same as in NDN.

We introduce the SIT to reap the benefits of the name-based routing and search for available content whenever the content origin is not reachable. Specifically, SIT keeps track of the Data packets that are heading towards users. In the event that Interest packets cannot reach the content origin, they can be forwarded based on the SIT entries towards users that successfully issued similar interests in the past. SIT entries like PIT entries also allows for a list of outgoing faces, allowing multiple sources for data, which can be queried in parallel.

Unlike a PIT entry, a SIT entry is triggered by a returning Data packet and similarly to a PIT entry SIT entries comprise a trail of "*bread crumbs*" for a matching Interest packet to follow back to users with similar satisfied interests. SIT entries are erased only when the associated user has been disconnected or according to an expiration methodology (e.g. a time-to-live parameter).

The SIT entries are similar to the Persistent Interests (PIs) proposed in [13]. Here instead of using them to reduce the waste of uplink bandwidth by explicitly requesting each packet in a media stream, we use them to allow content retrieval whenever the network is fragmented. The introduction of the SIT table increases the amount of states that should be maintained in the routers, but in this paper we assume that there is always enough space for the SIT and we leave for future work the study of the impact of the proposed scheme on the nodal memory capacity. Indicatively, in [14] a semi-stateless forwarding scheme is proposed, which effectively reduces the forwarding state maintained at each router, while preserving the advantages of CCN forwarding.

B. Packet format

In the Interest packet, we introduce a *Destination flag (DF)* bit to distinguish whether the packet is heading towards the content origin (DF is set to zero), following a FIB entry, or is heading towards users with similar satisfied interests. In the second case (DF is set to one), the Interest packet follows matching entries in the SIT of each passing-by router. The Data packet is exactly the same as in the CCN/NDN.

C. Interest packet processing

Whenever a user issues an Interest packet the Destination Flag bit is by default set to zero. This means that when an

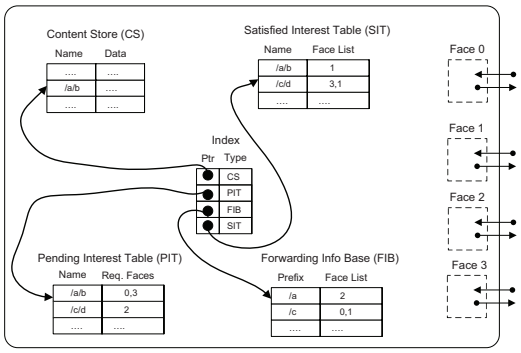

Fig. 1. Content Router design with the new Satisfied Interest Table (SIT).

Interest packet arrives on some face of a router, the router processes the packet in the exact same way as in CCN (the packet heads towards the content origin). Particularly, if a matching content is found in the CS, the router initiates a Data packet. The router sends the Data packet to the face the interest arrived on and discards it (since it was satisfied). Also, the router creates a new SIT entry from the Interest packet and its arrival face. This entry points towards the user (the next hop of the path towards the user) that issued the responded interest. The SIT entry has the same format as a PIT entry, as shown in Fig. 1.

If the router does not find matching content in CS and there is an exact-match PIT entry, the interest's arrival face is added to the PIT entry's Requesting Faces list and the Interest packet is discarded. Otherwise, if there is a matching FIB entry, the Interest packet is sent upstream towards the content origin. In particular, the arrival face is removed from the face list of the matching FIB entry, and if the resulting list is not empty, the packet is sent out to all the remaining faces (in case of multiple content origins/replicas) and a new PIT entry is created from the interest and its arrival face.

The above procedure, with the exception of the SIT entry, is exactly the same to the functionality of the CCN/NDN [2]. Moreover, in CCN when an Interest packet does not find a match in any of the CS, PIT and FIB it is discarded, since this router does not have any matching data to respond (CS), store (PIT) or forward (FIB) the packet and does not know how to find any; the content origin is not reachable. In our scheme, when the router does not find a match in any of the CS, PIT and FIB it checks for an exact-match SIT entry; this implies that the network connectivity has been interrupted either due to the mobility of the content origin or the fragmentation of the network. If such a match is found, a new PIT entry is created from the interest and its arrival face and the packet is sent out to the corresponding faces with the Destination flag set to one. Note that the SIT entries point either to another router or to a user attached on a router. Below we describe both cases.

When an Interest packet arrives on some face of a router and its DF bit is set to one, the router checks for a matching content in its CS. If a matching content is found the router, as above, initiates a Data Packet, creates a new SIT entry and discards the interest. If the router does not find a matching

content in CS skips searching the PIT and FIB (the DF is set to one which means that the content origin is not reachable) and checks for an exact-match SIT entry, similarly to the procedure described above. If no such entry is found in SIT the router discards the Interest packet. This means that the user(s) who created the SIT entry that is followed by the Interest packet is no longer reachable (or the SIT entry towards that user has expired).

When an Interest packet arrives on some face of a user, the user initiates a Data packet with the requested cached content and discards the interest. Only when an Interest packet cannot be forwarded to the content origin, the proposed resilience scheme exploits the cache of other users, following SIT entries, to retrieve cached content. We assume that in the scenarios that this resilience scheme is enabled, users are willing to help in the dissemination of the data and respond to incoming interests.

Alternative mechanisms can also be considered, where Interest packets are heading both towards other users and content origin as in [7] in order to satisfy them faster. Also a router can add FIB entries (instead of SIT) whenever it caches an item or has attached users with satisfied interests for a corresponding item and play the role of a replication point (alternative content origin), but both scenarios are out of the scope of this paper and are left for future investigation. Finally, the procedure followed by the network operator in order to detect a network fragmentation and update the FIB entries (remove entries in order to enable the proposed resilience scheme) is also out of the scope of this paper, but a scheme similar to the one presented in [11] for the PURSUIT [12] architecture or the Named-data Link State Routing protocol (NLSR) presented in [15] for the NDN [2] architecture could be adopted.

D. Data packet processing

The Data packet processing is exactly the same to CCN/NDN [2]. Particularly, a Data packet is not routed but simply follows the chain of PIT entries back to the requesting user(s). A longest-match lookup of a Data packet's Content Name takes place upon the arrival of the packet at each router. A CS match means that the Data packet is a duplicate so it is discarded. A FIB match means that there are no matching PIT entries so the Data packet is unsolicited and it is discarded. A PIT match (there may be more than one) means that the Data packet was solicited by interest(s) sent by this node. A list is created, that is the union of the Requesting Faces list of each PIT match minus the arrival face of the Data packet. The Data packet is sent out each face on this list, the PIT entries are removed, and for each face a new SIT entry is created. The new SIT entries are exactly the same to the PIT entries matching the Data packet. Obviously, the new SIT entries might be identical with existing entries. In that case, a SIT entry points towards more than one users (faces to reach those users), similarly to a PIT entry. Finally, the Data packet is (optionally; see Section IV-B) cached to the Content Store of the router (in [2] a Data packet is always cached to the CS of every router along the delivery path).

IV. Strategies/Policies of the Caching Mechanism

In this section, we present the family of policies/strategies that are useful to enable the retrieval of cached content when the network is fragmented and the content origin is not reachable.

A. Interest forwarding policies

The *Interest forwarding policy* dictates how the Interest packet is propagated in the network when the content origin is not reachable. According to Section III-C, an Interest packet that has its Destination flag bit set to one is propagated following the entries found in the SIT of each router, until a matching Data packet is found (either in a router or an attached user). On the other hand, it is possible to flood the network with Interest packets in order to make sure that a Data packet is retrieved at the cost of higher overhead. Particularly, we will investigate the following forwarding policies:

- Interest forwarding based on SIT entries – *SIT-based forwarding policy, STB*.
- Interest flooded to the network – *flooding forwarding policy, FLD*.

The role of the interest forwarding policy is to determine the system's tolerance to interest overhead and the content retrieval efficiency.

B. Caching policies

A *caching policy* dictates where a Data packet will be cached. In principle, all routers have a cache (Content Store), nevertheless we may restrict our interest to those routers (Edge routers) with attached users interested to the Data packet. Those routers have in their PIT matching entries that point to attached users. A router with attached users interested in the Data packet is always present in a sequence of SIT entries (we call those routers "*interested routers*"). Also, it could be possible that every router along the path/route from the content origin/responding router towards the user could cache the passing-by Data packet. Of course, a router may choose not to cache any passing-by Data packets upon the detection that the content origin is not reachable. In that case the router uses its cache as a replication device and it becomes a replication point for the Data packets already cached. We will therefore investigate the following caching policies:

- No cache after the fragmentation of the network – *No caching policy, NCP*.
- Cache only in interested edge routers – *Edge caching policy, EDG*.
- Cache in all routers along the path/route – *En-route caching policy, NRT*.

C. Placement/Replacement policies

The *placement/replacement policy* decides a position in the Content Store where a Data packet will be inserted and which packet will be discarded in case of an overflow. Due to space limitations and the objectives of this paper, we will only examine the *Least Recently Used, LRU* placement/replacement

TABLE I
STRATEGIES/POLICIES.

Policies		
Forwarding	Caching	Plac./Repl.
1. STB 2. FLD	1. NCP 2. EDG 3. NRT	1. LRU

policy. In literature (e.g. [16]-[17]) there exist more placement/replacement policies, but as shown in [17] their additive impact in the overall performance is negligible or require additional functionality for their application, not supported by a typical CCN/NDN router.

D. Caching mechanisms

Table I depicts the whole spectrum of the proposed policies, the combinations of which result to different information resilience caching mechanisms for the retrieval of cached content when the network is fragmented and the content origin is not reachable by all routers. From Table I, there exist six different combinations of caching mechanisms and we evaluate all of them in the performance evaluation section. Particularly, we evaluate the following combinations: 1) STB-NCP, 2) STB-EDG-LRU, 3) STB-NRT-LRU, 4) FLD-NCP, 5) FLD-EDG-LRU and 6) FLD-NRT-LRU.

In this paper we assume that during the normal operation of the network, when it is not yet fragmented (i.e. FIB entries exist) Interest packets are forwarded according to the CCN/NDN paradigm. Only when the content origin is no longer reachable the part of the network that cannot access it enables one of the six combinations for the retrieval of cached content.

V. PROBLEM FORMULATION AND PERFORMANCE BOUNDS

In this section we provide an analytical expression with the usage of continuous time, discrete state, Markov processes for the computation of the probability that an information item will disappear from the network (be absorbed) and the corresponding time to absorption, when the caching capacity of each router is equal to zero. This is also the case where the entire CS capacity is used for FIB, PIT and SIT entries or the no caching policy (see Section IV-B) is enabled and an information item is not cached at the CS of any router. The computed absorption time is also the lower bound for each one of the above caching mechanisms and depicts their capability to sustain and deliver content when the network is fragmented and only users are enabled to respond in interests for content.

A. System model

We consider a network of arbitrary topology, where $\mathcal{V}$ denotes the set of cache-enabled routers/nodes in the network. Generally, we are using the calligraphic letters to denote sets and the corresponding capitals for cardinality (e.g. $|\mathcal{V}| = V$).

We denote with $\mathcal{M}$ a set of M information items that can be requested by the network and with s^m the size (in bits) of item m. Without loss of generality we assume that all items are of the same size and equal to one ($s^m = s = 1$, $\forall m \in \mathcal{M}$). Each one of these items is stored permanently in a content server/origin and we assume that all items are stored in the same server. Also, every node $v \in \mathcal{V}$ has a cache of size C_v slots. This cache corresponds to the Content Store component of the NDN router described in Section III-A.

Requests for content access are generated by mobile users that connect and disconnect to the network nodes. We assume that new users connect at each node of the network with rate ζ_v, $v \in \mathcal{V}$, and each user connected at a node v disconnects from the network with rate ϕ_v, $v \in \mathcal{V}$. Without loss of generality we assume $\zeta_v = \zeta$ and $\phi_v = \phi$, $\forall v \in \mathcal{V}$.

At each node $v \in \mathcal{V}$ requests are generated with rate $r_v = \{r_v^1, \ldots, r_v^M\}$, where r_v^m denotes the aggregate incoming request rate (in requests per second) at node v for an information item $m \in \mathcal{M}$. The request rate for each item at each node is determined by its *popularity*. Here we approximate the popularity of the items at node v, $v \in \mathcal{V}$ by a Zipf law of exponents z_v. Literature provides ample evidence that the file popularity in the Internet follows such a distribution [18]. We denote by $\vartheta_v = \{\vartheta_v^m : m \in \mathcal{M}, v \in \mathcal{V}\}$ the popularity of each item m at node v. In that way the aggregate incoming request rate (in requests per second) at node v for an information item $m \in \mathcal{M}$ is given by:

$$r_v^m = \zeta_v \cdot \vartheta_v^m = \zeta \cdot \vartheta_v^m = \zeta \cdot \frac{1/k^{z_v}}{\sum_{i=1}^{M} 1/i^{z_v}}, \quad (1)$$

assuming that the particular item is ranked k-th out of the M information items within the Zipf distribution.

B. Absorption time and absorption probability

From Section III-C we recall that if the cache capacity of each router is equal to zero and the content origin is not reachable, the retrieval of a requested information item is based only on the SIT entries, or in other words in the newly proposed resilience scheme for retrieving content exploiting users connected in the network. The probability of retrieving a requested item m at time $t > 0$, assuming that at time $t = 0$ the network fragments and the content origin for that particular item is not reachable, depends only on the probability that another user has already retrieved that item in the past and is still connected in the network.

We define as $\{X_m(t),\ 0 \leq t < \infty\}$ the Markov process with stationary transition probabilities (where the possible values of $X_m(t)$ are nonnegative integers), that depicts the number of users (*population*) which have already retrieved item m and are connected in the network at time t. Clearly, if at any time instance $t' > 0$, $X_m(t') = 0$ the requested information item can no longer retrieved, since (i) it is not cached in the network (zero CS capacity), (ii) the content origin is not reachable and (iii) there are no connected users who have already retrieved the item and can be exploited for possible retrieval.

From the stochastic modeling theory $X_m(t)$ is a birth and death process with one absorbing state. We define as the zero state, the state at which $X_m(t) = 0$. This is an absorbing state (no user with a cached copy of item m is connected in the network), since after that state the requested item cannot be retrieved. Of course, new users can arrive at a network node but they cannot retrieve the requested item, until the network connectivity is "back" and the content origin is reachable through the FIB entries.

We define as λ_n^m the birth rate of the process when the process is at state n (n connected users in the network who have the information item m) and as μ_n^m the death rate of the same process. Clearly $\lambda_0^m = 0$. In our case we have for λ_n^m:

$$\lambda_n^m = \begin{cases} 0 & \text{if } n = 0, \\ \sum\limits_{v \in \mathcal{V}} r_v^m = \sum\limits_{v \in \mathcal{V}} \zeta \cdot \vartheta_v^m & \text{if } n > 0, \end{cases} \tag{2}$$

where r_v^m is given by Eq.(1). Note that the birth rate of the process is independent of its actual state when $n > 0$ ($\lambda_n^m = \lambda^m$) and it depends only on the popularity of the corresponding information item.

For the death rate of the process we have:

$$\mu_n^m = n \cdot \phi_v = n \cdot \phi. \tag{3}$$

According to the stochastic modeling theory we derive the following theorem:

THEOREM 1. *Consider the birth and death process that depicts the number of users (population) which have already retrieved item m and are connected in the network with birth and death parameters λ_n^m and μ_n^m, where $\lambda_0^m = 0$ so that 0 is an absorbing state. The probability of absorption into state 0 from the initial state $s > 0$ is:*

$$u_s^m = \begin{cases} 1 & \text{if } \sum\limits_{i=1}^{\infty} \rho_i^m = \infty, \\ \dfrac{\sum_{i=s}^{\infty} \rho_i^m}{1 + \sum_{i=1}^{\infty} \rho_i^m} & \text{if } \sum\limits_{i=1}^{\infty} \rho_i^m < \infty. \end{cases} \tag{4}$$

The mean time to absorption is:

$$T_s^m = \begin{cases} \infty & \text{if } \sum\limits_{i=1}^{\infty} \dfrac{1}{\lambda_i^m \cdot \rho_i^m} = \infty, \\ \sum\limits_{i=1}^{\infty} \dfrac{1}{\lambda_i^m \cdot \rho_i^m} + \sum\limits_{k=1}^{s-1} \rho_k^m \sum\limits_{j=k+1}^{\infty} \dfrac{1}{\lambda_j^m \cdot \rho_j^m} & \text{if } \sum\limits_{i=1}^{\infty} \dfrac{1}{\lambda_i^m \cdot \rho_i^m} < \infty. \end{cases} \tag{5}$$

where

$$\rho_i^m = \begin{cases} 1 & \text{if } i = 0, \\ \dfrac{\mu_1^m \mu_2^m \cdots \mu_i^m}{\lambda_1^m \lambda_2^m \cdots \lambda_i^m} = \dfrac{\phi \cdot 2\phi \cdots i\phi}{\lambda^m \lambda^m \cdots \lambda^m} = \left(\dfrac{\phi}{\lambda^m}\right)^i \cdot i! & \text{if } i > 0. \end{cases} \tag{6}$$

Proof. Due to limited space the detailed derivation of Eq.(4) and Eq.(5) is omitted. The proof is similar in rationale to [19]. ■

From the above theorem, we observe that when the rate ϕ at which users possessing item m disconnect from the network is larger than the rate λ^m at which new users interested in item m arrive to the network, the corresponding item will finally get absorbed. The time and the probability of absorption depends on the initial state; population of users that possess item m and are connected in the network at the moment when the network fragments and the content origin "disappears". Obviously, when new users arrive faster than those disconnecting, an item never gets absorbed and each one of the above proposed caching mechanisms allow the retrieval of content infinitely. In the following section we analyze the performance of the proposed caching mechanisms when all items are to be finally absorbed ($\lambda^m < \phi$, $\forall m \in \mathcal{M}$) and examine the effect of the caching capabilities of each router to further allow information resilience.

VI. Performance Evaluation

A. Evaluation setup

In this section, we use a custom built discrete event simulator to evaluate the performance of the proposed information resilience scheme and the various caching mechanisms. Initially, we compare the performance of the proposed scheme against the theoretical lower bounds derived in Section V-B. Then we compare the information resilience caching mechanisms presented in Section IV-D for various parameters (caching capacity of the routers and users' disconnection rate).

We assumed that item popularity is given by a Zipf law distribution of exponent z_v, $v \in \mathcal{V}$. In particular, we consider various values for z_v in the range from 0 to 1, and we assume that in each router users are connected with rate $\zeta_v = \zeta = 1$ user per second. Thus, the request rate for each item at each router varies from 0-1 reqs/sec depending on item's popularity (see Eq. (1)).

Generally, the popularity of each item may differ from place to place, a phenomenon that is referred to as locality of interest (*spatial skew* in [20]). In our experiments, the workload brought to the network is based on a localized request/interest model, where there are regional differences across requests at different locations. We assume ten different regions and the size of all regions are the same and equal to $(V/10)$ nodes each. Each region corresponds to a specific value for the Zipf distribution exponent characterizing its local popularity of items. Also at each region the ranking/order of the items within the Zipf distribution is different to the ranking followed in the rest of the regions. The set of nodes that constitute a region is computed by choosing randomly a central node and its $V/10-1$ closest neighbors, by executing a breadth-first search algorithm (as long as a node has not been already assigned to another region).

We use a network topology with $V = 50$ nodes from the Internet Topology Zoo dataset [21]. We also consider a disaster scenario where the content item population is $M = |\mathcal{M}| = 10^3$. Although 10^3 items may not seem representative of the current Internet content space, here we focus on a disaster case, where information regarding the state of the disaster is distributed by first responders and users request for updates. That is, first responders (e.g., fire brigade) is publishing information in specific places (e.g., neighborhoods), utilizing mobile data mules (e.g., ambulances). As the authorities move in the disaster area, the origin server (here represented by a mobile data mule) becomes inaccessible. In turn, users asynchronously request for updates on the state of the emergency. In this case, information has to be retrieved from either in-network caches, or from other users who have already downloaded the updated information. Given that authorities publish new content/information every some tens of minutes [10], we ex-

periment with a small content population and evaluate whether users can get access to important updates.

For the evaluation of the proposed information resilience scheme, we initially assume that for a period of 3600sec = 1hour the content origin of every item is reachable (we call this period of time *"initialization period"*). During this period the network functions similarly to the CCN/NDN model (Interest packets follow the FIB entries towards the content origin and the routers follow the NRT caching policy and the LRU placement/replacement policy) and we assume that $\zeta = 1$ and $\phi = 0.1$ for the connection and disconnection rate of users to each node respectively. After the initialization period we assume that the network fragments, due to a disaster event, and the content origin of every item "disappears"[1]. We monitor the performance of the caching mechanisms and the proposed information resilience scheme for a period of $T = 10800\text{sec} = 3\text{hours}$ (we call this period of time *"observation period"*). We assume that during the observation period the popularity vector r_v at each node $v \in \mathcal{V}$ is stable and we leave for future work the evaluation of scenarios where popularity alterations exist after the fragmentation of the network.

Our evaluation is based on the following metrics:

- *Satisfaction* (in % of issued interests): It is the percentage of the interest that have been satisfied (found the requested item) during the observation period.
- *Absorbed Items* (in % of information items): It is the percentage of the M items that have been absorbed during the observation period.
- *Mean Absorption Time* (in *sec*) of an absorbed item: It is the mean time passed from the moment the network fragments until the item gets absorbed.
- *User Responses* (in % of satisfied interests): It is the percentage of the satisfied interests that were satisfied by a user and not from the cache CS of a router during the observation period.
- The *Minimum Hop Distance* (in *hops*): It is the minimum number of hops between a responding router/user and the user issued the Interest packet. This metric is indicative of the transfer delay as a function of hops in the network.
- The *Traffic Overhead* (in *hops*): It is the total number of hops that the duplicate initiated Data packets travel in the network, until they are discarded.

B. Model validation

In Fig. 2 we depict the actual absorption time for each information item computed using Eq.(5) (straight line) and using the simulator (crosses) described above. We observe that the theoretical results are inline with the output of the simulator, which further implies the validity of the model presented in Section V-B. In more details, we observe that the vast majority of the items $\approx 93\%$ are absorbed in $13-19$sec after the fragmentation of the network when routers have zero caching capacity and for the given users' connection and disconnection rates. Also, there is a significant group of items that remain in the network for a large period of time (e.g. four items can be retrieved for more that 3000sec), but due to the selected parameters all items are eventually absorbed.

[1]We assume that the examined network is the fragment which cannot reach the content origin of any information item e.g. an area that was cut off by an earthquake.

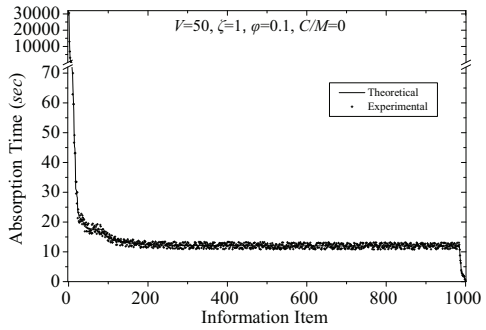

Fig. 2. Experimental and theoretical absorption time for the case where each router has zero cache capacity.

We should mention here that items are prioritized based only on their popularity. A scope-based scheme similar to the "NREP" [10] prioritization system can also be applied here and is on our imminent plans. Nevertheless, the proposed resilience scheme manages to retain and retrieve content for a significant amount of time, therefore content that is considered of high importance and of high popularity (e.g. a warning or directions from the authorities in a disaster scenario) will reach its recipients despite the lack of cache in the routers of the network and the intermittent connectivity with the content origin.

C. Impact of the cache size

In Fig 3 we depict the impact of the cache capacity, expressed as the fraction of the items population that can be stored in the CS of a router. We observe that the three caching mechanisms that incorporate the flooding request policy result in larger satisfaction ratios than the same mechanisms using the STB policy. Also, those mechanisms that use the STB policy perform almost identical, regarding the satisfaction metric, regardless of the cache capacity of each router. The two mechanisms that do not copy new items during the observation period (STB-NCP and FLD-NCP) perform significantly better compared to the other mechanisms, regarding the percentage of the absorbed items. In the cases where the cache capacity C/M is larger that 25% those two mechanisms can retrieve every information item infinitely; no item is getting absorbed. Regarding the mechanisms that continue to copy items after the fragmentation of the network, we observe that the Mean Absorption Time increases linearly with the caching capacity of the routers, since the less popular content that is frequently evicted can be cached for a larger period of time in larger Content Stores. This means that even if those mechanisms cannot retrieve all the items for the whole observation period, they can maintain items for a large period of time, $1000-6500$sec, after the fragmentation of the network.

For small values of the cache capacity ($2\%-5\%$ of the total information space), we observe that $12\%-45\%$ of the satisfied interests are served from users attached to the network using

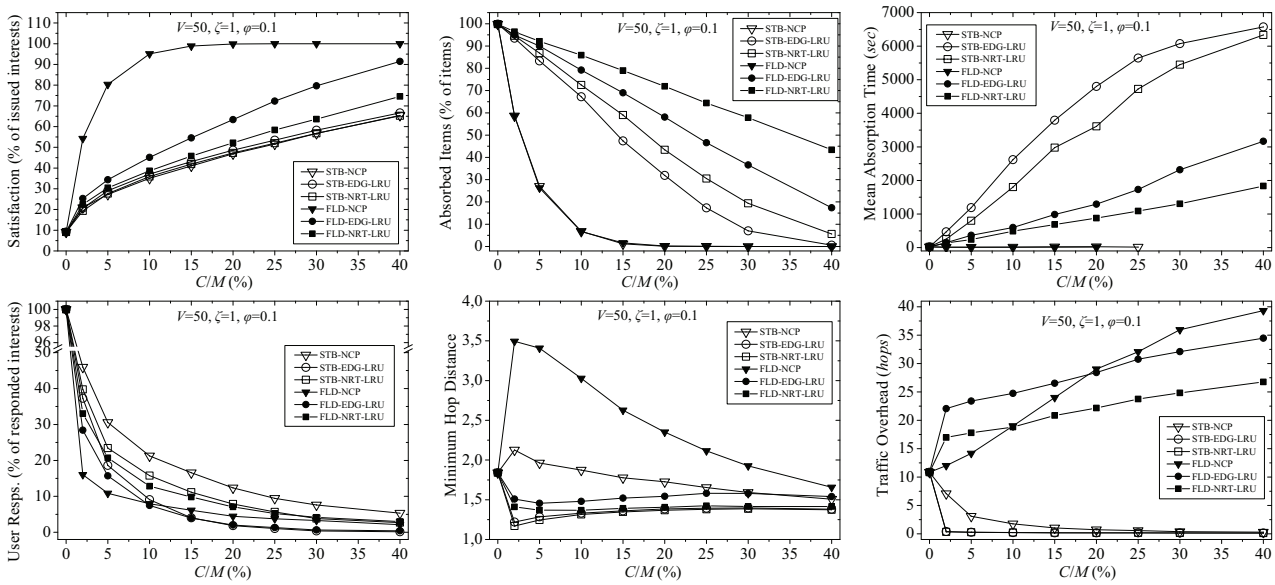

Fig. 3. The performance of the proposed caching mechanisms vs. the caching capacity of each router.

the proposed resilience scheme. This indicates that not only the introduction of the SIT table, but also the exploitation of the cache of the users attached to the network is important, since otherwise this amount of interests wouldn't have been satisfied. Of course, as we relax the storage capacity constraint and allow more items to fit in each router the majority of the interests are satisfied by its CS.

Regarding the hop distance metric, we observe that the mechanisms that use the STB forwarding policy perform slightly better than the same mechanisms that flood the interest to the network. Also, this metric is improved but not significantly by the increase of the cache capacity of each router, since this metric is mainly affected by the size of the network (number of routers and topology). As regards the Traffic Overhead metric, we observe that the mechanisms that use the STB forwarding policy are not affected by the increase of the caching capability of the routers. On the other hand, the mechanisms that use the flooding forwarding policy present an almost linear increase in the traffic overhead, and they perform up to 35 times worse than the mechanisms using the STB policy. Note that in Fig. 3 the values that correspond to $C/M = 0$ are inline and reproducible by the analysis of Section V-B.

D. Impact of the users' disconnection rate

Fig. 4 depicts the impact of users' disconnection rate on the performance of the proposed caching mechanisms. Starting from the user responses metric we observe that for values of $\phi > 0.2$ less than 5% of the satisfied interests are served from users attached to the network. This means that the 95% of the satisfied interests are served from the CS of the routers and the proposed resilience scheme (SIT entries) is used only for the forwarding of the Interest packets. Regarding the satisfaction metric we observe that the caching mechanisms are only slightly affected by the disconnection rate. Of course the mechanisms that use the flooding forwarding policy perform, as previously, better than the mechanisms using the STB policy, but at the cost of increased traffic overhead. Generally, only the two mechanisms that use the STB forwarding policy and continue to copy items after the "disappearance" of the content origin (STB-EDG-LRU and STB-NRT-LRU) are affected by the increase of the disconnection rate ϕ. In particular, the increase of the disconnection rate allows them to discard the less popular items in a short period of time and maintain the rest items (more popular) for a longer period.

From the above analysis we observe that each caching mechanism has its own pros and cons and it is up to the operator/manager of the network which one to enforce for the retrieval of cached content after the fragmentation of the network. Also, in the whole performance analysis we assumed that all the items' origins/servers "disappear" simultaneously. This assumption was made in order to examine the performance of the resilience scheme at an extreme disaster scenario. In a real setup it is more likely that only a portion of the content origins to "disappear" each time a network fragmentation occurs or the existence of replication points within the network would minimize the fragments of it that doesn't have access to the content origin (or a replica).

VII. Conclusions and Future Work

We put forward a new information resilience scheme in CCN/NDN for the retrieval of content in disruptive, fragmented networks cases. Particularly, we enhanced the NDN router design, as well as its Interest forwarding scheme so that users can retrieve cached content when the network is fragmented and the content origin is not reachable. We also derived a closed form lower bound for the computation of the probability that an information item will disappear from the network (be absorbed) and the corresponding time to absorption. Our extensive simulations show that the proposed resilience scheme is a valid tool for the retrieval of cached content in disruptive cases, where the content origin is not

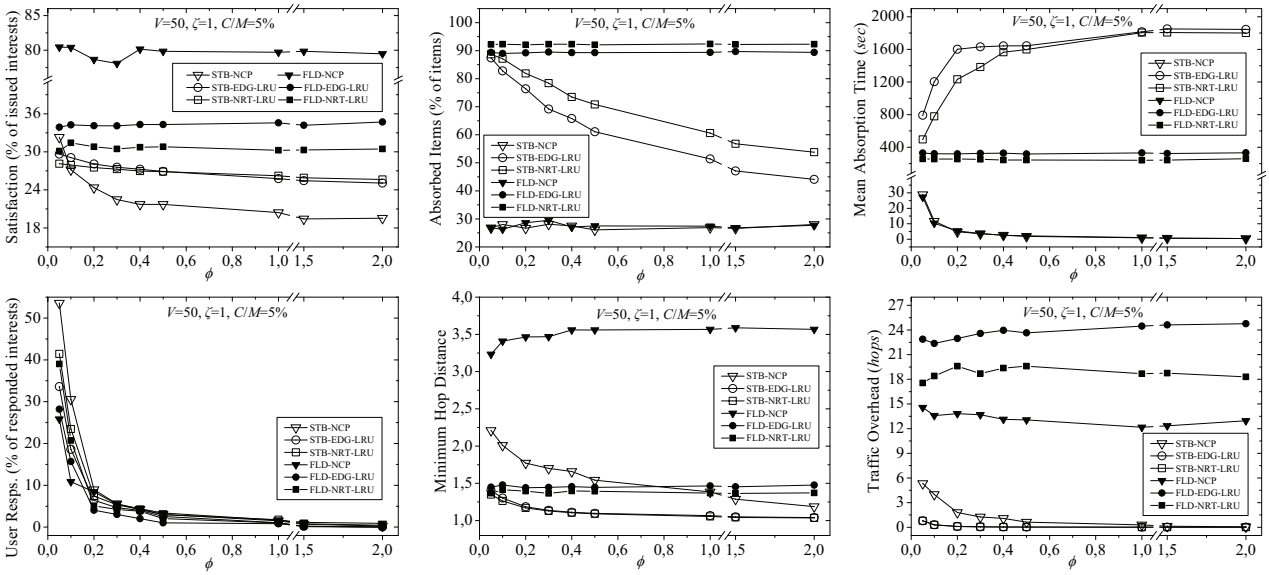

Fig. 4. The performance of the proposed caching mechanisms vs. the disconnection rate of the users from the network.

always present due to its own mobility or to network node/link failures.

Apart from the future work hints given throughout the paper, this work can also be extended in the following directions as well. Firstly, the detailed study of the impact of the proposed scheme in the memory of the routers. The proposed SIT table inflates the number of data that should be kept in a router by the number of users attached to the network, therefore novel mechanisms are required to guarantee scalability and nodal memory capacity. Additionally, the proposed scheme is examined only when the content origin is not reachable, but the exploitation of the cache of the users during the normal operation of the network can significantly decrease the traffic and increase the QoS perceived by them. Finally, the integration of a scope-based content prioritization scheme within our information resilience scheme is on our future plans.

ACKNOWLEDGMENTS

This work was conducted while V. Sourlas was a research associate at the University of Thessaly, Greece. V. Sourlas work is supported by the European Commission through the FP7-PEOPLE-IEF INTENT project, Grant Agreement no. 628360. I. Psaras and G. Pavlou work is supported by the EU-Japan initiative under EC FP7 Grant Agreement no. 608518 and NICT Contract no. 167 (the GreenICN project).

REFERENCES

[1] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros and G. C. Polyzos, "A Survey of Information-Centric Networking Research," IEEE Com. Surv, 2013.

[2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. Briggs, R. Braynard, "Networking named content," ACM CoNEXT, 2009.

[3] J. Seedorf, M. Arumaithurai, A. Tagami, K. Ramakrishnan and N. Blefari Melazzi, "Using ICN in disaster scenarios," IETF ICNRG Internet draft, 2014. http://tools.ietf.org/html/draft-seedorf-icn-disaster-02

[4] Al. Ghodsi, T. Koponen, B. Raghavan, S. Shenker, An. Singla and J. Wilcox, "Information-centric Networking: Seeing the forest for the trees", ACM HotNets, 2011.

[5] An. Ioannou and S. Weber, "Towards On-path Caching Alternatives in Information-Centric Networks," Tech. Report, Trinity College, Dublin, Ireland, 2014.

[6] W. K. Chai, I. Psaras and G. Pavlou, "Cache Less for More In Information-Centric Networks," IFIP NETWORKING, 2012.

[7] H. Lee and A. Nakao, "User-assisted in-network caching in information-centric networking," Computer Networks - Elsevier, 2013.

[8] Architecture and Applications of Green Information Centric Networking (GreenICN), http://www.greenicn.org/

[9] T. Ogawara, Y. Kawahara and T. Asami, "Information Dissemination Performance of Disaster Tolerant NDN-based Distributed Application over Disrupted Cellular Networks," IEEE Peer-to-Peer Computing (P2P) Conference, 2013.

[10] I. Psaras, L. Saino, M. Arumaithurai, K. Ramakrishnan and G. Pavlou, "Name-Based Replication Priorities in Disaster Cases," IEEE INFOCOM NOM, 2014.

[11] M. Al-Naday, M. Reed, D. Trossen, Kun Yang, "Information resilience: source recovery in an information-centric network," IEEE Network, 2014.

[12] D. Trossen and G. Parisis, "Designing and Realizing an Information-Centric Internet", IEEE Commun. Mag., 2012.

[13] C. Tsilopoulos and G. Xylomenos, "Supporting Diverse Traffic Types in Information-centric Networks," ACM SIGCOMM ICN workshop, 2011.

[14] C. Tsilopoulos, G. Xylomenos and Y. Thomas, "Reducing Forwarding State in Content-Centric Networks with Semi-Stateless Forwarding," IEEE INFOCOM, 2014.

[15] A. K. M. Mahmudul Hoque, S. O. Amin, A. Alyyan, B. Zhang, L. Zhang, L. Wang, "LSR: Named-data Link State Routing Protocol," ACM SIGCOMM ICN workshop, 2013.

[16] S. U. Khan and I. Ahmad, "Comparison and analysis of ten static heuristics-based Internet data replication techniques," Journal of Parallel and Distributed Computing, 2008.

[17] M. Diallo, V. Sourlas, P. Flegkas, S. Fdida, and L. Tassiulas, "A Content-Based Publish/Subscribe framework for Large-scale Content Delivery," Computer Networks - Elsevier, 2013.

[18] G. Dán and N. Carlsson, "Power-law Revisited: Large Scale Measurement Study of P2P Content Popularity," in 9th IPTPS, 2010.

[19] H. M. Taylor and S. Karlin, "An Introduction to Stochastic Modeling, 3rd edition", Academic Press, 1998.

[20] S. K. Fayazbakhsh, Y. Lin, A. Tootoonchian, A. Ghodsi, T. Koponen, B. Maggs, K. C. Ng, V. Sekar, S. Shenker, "Less Pain, Most of the Gain: Incrementally Deployable ICN," ACM SIGCOMM, 2013

[21] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden and M. Roughan, "The Internet Topology Zoo," IEEE JSAC, 2011.